Routledge
Taylor & Francis Group

North American Actuarial Journal

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uaaj20

Cointegration Analysis of Crop Yields and Extreme Weather Factors Using Actuaries Climate Index with Application of Bonus–Malus System

Eric C. K. Cheung, Ryan H. L. Ip, Ho On Tam & Jae-Kyung Woo

To cite this article: Eric C. K. Cheung, Ryan H. L. Ip, Ho On Tam & Jae-Kyung Woo (29 Oct 2024): Cointegration Analysis of Crop Yields and Extreme Weather Factors Using Actuaries Climate Index with Application of Bonus–Malus System, North American Actuarial Journal, DOI: 10.1080/10920277.2024.2410822

To link to this article: https://doi.org/10.1080/10920277.2024.2410822

North American Actuarial Journal, 0(0), 1–19, 2024
© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC.
ISSN: 1092-0277 print / 2325-0453 online
DOI: 10.1080/10920277.2024.2410822

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Cointegration Analysis of Crop Yields and Extreme Weather Factors Using Actuaries Climate Index with Application of Bonus–Malus System

Eric C. K. Cheung,[1] Ryan H. L. Ip,[2] Ho On Tam,[3] and Jae-Kyung Woo[1]
[1]*School of Risk and Actuarial Studies, University of New South Wales, Sydney, New South Wales, Australia*
[2]*Department of Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand*
[3]*Actuarial Consulting, Taylor Fry Consulting Actuaries, Sydney, New South Wales, Australia*

This article analyzes the long-term temporal co-movement of the extreme weather variables in the Actuaries Climate Index (ACI) and crop yields, modeling their relationship using an error correction model (ECM). The analysis suggests that significant weather variables can serve as trigger parameters in the pricing framework of weather index crop insurance. To address the challenge of weather index crop insurance while preserving the advantages of a bonus–malus system (BMS), we propose a transition rule that distinguishes between damages caused by severe weather and those resulting from the policyholder's decisions. Subsequently, we also explore the challenges of implementing such a new hybrid BMS for crop insurance where extreme weather outcomes are integrated into the classical BMS.

1. INTRODUCTION

Crop insurance is a major instrument for farmers around the world to limit their losses in adverse situations that affect the next harvest. However, there are fundamental challenges in applying traditional indemnity plans to crop insurance. According to Woodard and Garcia (2008), these challenges can be classified into two categories identified in many research papers: the systematic nature of the risk and asymmetric information. These issues significantly affect the crop insurance industry, leading to high loss ratios, low participation rates, and the reluctance of private insurers to offer services in the agricultural sector.

Asymmetric information remains one of the principal challenges in crop insurance. By linking benefits paid to a variable independent of the farmer's control, such issues can potentially be addressed. This has led to the development of index-based crop insurance, where indemnity payments are not based on the individual's losses but on a predefined index that is believed to correlate with the farmer's experience (Deng et al. 2007). However, as suggested by Elabed et al. (2013), basis risk might exist in weather index-based insurance where the payouts are determined by a weather index that poorly predicts the individual farmer's loss.

This motivates us to propose a modified ratemaking framework based on the individual's loss experience by incorporating the external weather factor that significantly affects the crop yield in a long term. The bonus–malus system (BMS), commonly applied to auto insurance, differentiates between high-risk and low-risk policyholders, charging each policyholder a premium according to his or her claim history. Implementation of the BMS in this context could better account for the individual's loss experience, resulting in more equitable calculation of the premium.

This article begins the statistical analysis with cointegration tests between weather variables in the Actuaries Climate Index (ACI 2021) and wheat yield in six regions in the United States to identify an extreme weather variable having a significant long-term relationship with the crop yield. We subsequently utilize the error correction model (ECM) to model the relationship with the extreme weather variable, which assists with long-term forecasting of crop production. Then, we propose a hybrid

Address correspondence to Jae-Kyung Woo, School of Risk and Actuarial Studies, Level 6 East Wing, UNSW Business School, UNSW Sydney, Sydney, NSW 2052, Australia. E-mail: j.k.woo@unsw.edu.au

BMS framework that incorporates the weather variable cointegrated with crop yield in the design of the BMS transition rule, accounting for losses caused by weather events uncontrollable by the individual farmer. Consequently, individual farmers are not penalized by these types of losses caused by systematic risks beyond their control. This study innovatively incorporates extreme weather data, which exhibit a significant long-term cointegration relationship with crop yields, as a trigger parameter in a hybrid BMS framework, an approach not previously explored in the literature. Leveraging the rich data from the ACI, this model addresses key issues in crop insurance, such as systematic risk and asymmetric information, offering an idea of a more appropriate and equitable solution.

The rest of article is organized as follows. In Section 2, we first give further background on our research problem. Section 3 provides general descriptions of the methodology concerning cointegrated time series model and the ECM. In Section 4, we implement these models to analyze the extreme weather factors from the ACI and identify that sea levels have a significant long-term relationship with crop yield. Then, we estimate the parameters of the ECM for the cointegrated time series and subsequently predict the crop yield. In Section 5, we briefly introduce a classical BMS and explain how to modify the framework by incorporating the weather variable in the design of the transition rule. We utilize different transition rules to illustrate how the system works, and discuss several challenges in implementing a hybrid BMS within the proposed framework in crop insurance. The article concludes with some remarks in Section 6.

2. BACKGROUND

First, crop insurance serves as a risk management tool against crop losses from perils beyond the farmer's control and supports farm survival, as studied by Kim, Yu, and Pendell (2020). Unlike other areas of insurance, the agricultural industry exhibits inherent characteristics that reduce the functionality of insurance policies. As a result, a direct application of traditional insurance products is difficult in crop insurance, especially due to two key aspects: the systematic nature of the risk (e.g., Miranda and Glauber 1997; Mason, Hayes, and Lence 2003) and asymmetric information (e.g., Skees and Reed 1986; Nelson and Loehman 1987; Chambers 1989). Both are explained in more detail below:

- *Correlated risks*: Traditional crop insurance provides protection against one or multiple natural perils, such as pest damage, drought, and other weather conditions. The challenge with this approach is that the risks due to natural disasters exhibit high spatial correlation. Events such as droughts and floods often affect a wide area across the country, meaning that individual farm-level losses tend to mirror the broader regional experience (Miranda and Glauber 1997).
- *Asymmetric information*: Another challenge for insurers is asymmetric information. Moral hazard arises where the insured alters their optimal decisions owing to the presence of an insurance policy. On the other hand, adverse selection occurs when individuals most at risk are more likely to purchase insurance (Quiggin, Karagiannis, and Stanton 1993). In practice, Smith and Goodwin (1996) found that insured farmers tend to use less fertilizers and chemical inputs, making their crops more vulnerable to weather changes compared to those of noninsured farmers.

To address the first issue, weather index crop insurance has been developed. In much of the research on weather index insurance, rainfall is often used as a trigger parameter to determine whether weather conditions are favorable. For an overview of operational index insurances for grasslands, we refer interested readers to Table 1 in Vroege, Dalhaus, and Finger (2019). It is essential to ensure a strong correlation between the losses experienced by the insured and the defined adverse weather events; otherwise, the protection provided by crop insurance is limited. However, much of exisiting research focuses on correlation, which measures the strength of a linear relationship between variables, rather than examining the long-term stochastic trend shared by time series variables. Because the concept of cointegration is better suited to analyzing long-term relationships between time series, in this article we make use of the analysis of cointegrated time series of crop yield and extreme weather variables. These weather variables can serve as trigger parameters owing to their significant long-term relationships with crop yields. Cointegrated time series are commonly used in economics and finance when considering the long-term relationship between variables. For example, Bishai (1995) studied the cointegrated relationship between infant mortality and gross national product per capita in Sweden, the United Kingdom, and the United States. This approach reduces the possibility of spurious correlation that may arise when using linear regressions to examine relationships.

In our study, we utilize the ACI to extract extreme weather records for analysis. This index was jointly developed by the American Academy of Actuaries, the Casualty Actuarial Society, the Canadian Institute of Actuaries, and the Society of Actuaries. It provides an objective indicator of the frequency of extreme weather events and the extent of sea level change in the United States and Canada. The joint analysis of extreme weather variables and contingency events can be found in various actuarial studies. For example, Li and Tang (2022) conducted an empirical study on the dependency between cold weather and old-age death counts in the United States, and Pan, Porth, and Li (2022) applied various regression models to the weather variables from the ACI and corn yields in Midwestern states in the United States. Because weather is a significant factor influencing crop productivity, there are other research works related to the weather effects in agriculture. In particular, Woodard (2014) developed a conditional Weibull distribution approach to model the impact of several variables including weather on crop yield, and Nadolnyak, Vedenov, and Novak (2008) studied the dependence between the yield distributions for corn, cotton, and peanuts and long-term climate forecast information from the El Niño-Southern Oscillation in the Southeastern United States.

To tackle the issue of asymmetric information, we adopt the BMS, which is an experience rating method broadly used in auto insurance. With regard to the experience rating system in agricultural insurance, Vilar-Zanón, Heras, and de Frutos (2020) outlined the application of BMS in the Spanish market, and Rejesus et al. (2006) conducted an empirical analysis of a premium rate discount system (bonus–malus incentives) in US. crop insurance. Incorporating external weather factors into the design of the BMS can help mitigate an issue of the premium surcharge for unfavorable years caused by adverse weather conditions. For a comprehensive overview, see Lemaire (1995). Biffis et al. (2022) studied the bundling of parametric insurance with loans in low-income countries, where assessing indemnity insurance is challenging. Their study addressed the basis risk of weather index insurance and moral hazard issues.

In the next section, we lay the methodological foundation for the subsequent statistical analysis in this article.

3. METHODOLOGY

3.1. Cointegrated Time Series

Cointegration refers to the concept where a combination of two nonstationary time series forms a stationary series, indicating a long-run relationship between the two; that is, they move jointly. In the following, $I(d)$ denotes a time series that is integrated of order d. By definition, a stationary series is $I(0)$. If the differenced time series is stationary, the data are said to be integrated of order one or $I(1)$. According to the definition provided by Engle and Granger (1987), in mathematical terms, the components of the row vector $\boldsymbol{w}_t$ are said to be cointegrated of orders d and c, denoted as $\boldsymbol{w}_t \sim CI(d, c),$ if the following conditions hold:

- All components of $\boldsymbol{w}_t$ are $I(d)$.
- There exists a row vector $\boldsymbol{\eta}$ with nonzero elements such that $z_t = \boldsymbol{\eta}\boldsymbol{w}_t^\top \sim I(d-c)$ for $0 < c \leq d,$ where $\boldsymbol{\eta}$ is called the cointegrating vector.

With the focus on the case $d = c = 1,$ cointegration for $\boldsymbol{w}_t = (x_t, y_t)$ means that $x_t \sim I(1)$ and $y_t \sim I(1),$ and there exists a parameter b such that $z_t = y_t - bx_t$ is a stationary process. That is, $(x_t, y_t) \sim CI(1, 1)$ and $z_t \sim I(0)$. The above condition implies that the residuals from the ordinary least squares (OLS) estimates of the linear relationship between x_t and y_t are stationary. That is,

$$e_t = y_t - a - bx_t \sim I(0). \tag{1}$$

The nonstationarity of the time series can be determined by performing the augmented Dicky-Fuller (ADF) test; see Maddala and Kim (1998). Several conventional approaches exist for testing cointegration, including the Philip-Ouliaris (PO) cointegration test and the Engle-Granger procedure (Engle and Granger 1987), which utilize the Philips-Perron unit root test and the ADF test, respectively. Critical values for the Philips-Perron test were simulated by Phillips and Ouliaris (1990). The null hypothesis of these tests is non-cointegration; thus, cointegration is confirmed if the null hypothesis is rejected. Within the Engle-Granger procedure, the parameters are estimated to allow a visual representation of the relationship. This is also used in the ECM, which is described in the next section. Moreover, a trend must be carefully included in the cointegrating regression if the time series contains a trend. Otherwise, the asymptotic critical values will be different. Hence, for each weather variable in a specific region, significance of the coefficients to the model components such as a drift and linear trend is statistically checked together with the test statistics for the hypothesis.

Furthermore, to test for cointegration among multiple (more than two) time series, we can apply the Johansen test (Johansen 1991). This test overcomes the limitation of the aforementioned cointegration ADF test, which can only assess cointegration between two time series. The Johansen test sequentially examines how many times series under test are cointegrated. The null hypothesis of $r = 0$ indicates no cointegration, and the test runs from $r = 0$ to $r = n-1,$ where n represents the number of time series being tested. However, a drawback of this method is its lower statistical power compared to the ADF cointegration test (Gonzalo and Lee 1998).

3.2. Error Correction Model

The ECM is an effective tool for quantifying the magnitude and direction of relationships between variables in a cointegrated time series. Statistical analysis and inference of cointegrated time series and the ECM were first investigated in the works by Granger (1981), Granger and Weiss (1983), and Engle and Granger (1987). The ECM provides a systematic framework for jointly analyzing short-run dynamics and long-run relationship of the cointegrated time series. Whereas the cointegrating equation captures the the long-run relationship between the variables, the ECM allows for modeling the short-run dynamics. Specifically, the ECM reflects the long-run equilibrium relationships between variables and includes a short-run dynamic adjustment factor, which adjusts the variables when they deviate from the long-run equilibrium. To apply the ECM to the variables, the following conditions must first be satisfied:

- All variables are integrated to the order of 1.
- The variables are cointegrated.

Then, ECM is applied to forecast the time series of our interest. In Section 4, under the ECM, we estimate the future change in wheat yield by considering the current deviation from the long-run equilibrium estimated from past experience. We adopt a two-step approach to find the ECM as follows:

- *OLS estimation of the long-term relationship*: Run the OLS regression between the two variables (called the cointegrating regression) so that

$$y_t = \hat{a} + \hat{b}x_t + e_t, \tag{2}$$

  and the residuals e_t's are tested for stationarity. We test the hypothesis that e_t's are not stationary by running an ADF test on the residual series. If it is stationary, we proceed to the next step.

- *Estimating ECM parameters for the short-term dynamics*: Consider the ECM

$$\Delta y_t = \hat{\alpha} + \hat{\beta}\Delta x_t + \hat{\gamma}e_{t-1} + \epsilon_t, \qquad \hat{\gamma} < 0, \tag{3}$$

  where the lagged residual from (2), namely, $e_{t-1} = y_{t-1} - \hat{a} - \hat{b}x_{t-1},$ is the error correction term. Lagged values of the error correction term are useful for predicting Δx_t and/or Δy_t. We estimate the parameters $\hat{\alpha},\ \hat{\beta},$ and $\hat{\gamma}$ in the ECM (3) by OLS.

Note that (3) only considers the experience from one lag behind (i.e., only the current year's data are used to predict next year's value). Once all the coefficients have been estimated, the model can be employed to forecast future values. The predictive power can be validated by withholding recent periods and comparing the actual experience with the predicted values. Furthermore, the residuals obtained after fitting the ECM are tested for the presence of autocorrelation and heteroscedasticity.

For multiple time series that may have cointegrating relationships, a vector error correction model, which extends the ECM in (3), can be applied. A vector error correction model incorporates error correction features in a vector autoregression model. The Johansen test, discussed in Section 3.1, can be used to assess the cointegration among multiple time series.

Equipped with the necessary tools, we shall proceed to the data analysis in the next section.

4. DATA ANALYSIS

We obtain the time series of wheat yield and weather variables in six regions in the United States as follows.

4.1. Data Description

Region Definition: To reduce the complexity of this data analysis and to allow the direct application of data from the ACI produced for the United States and Canada, we group the 50 states in the United States into six regions as defined by ACI Executive Summary (ACI 2021); see Figure 1. The states within each region are provided in Appendix A.

Wheat Yield: We collate the wheat yield data from the US. Department of Agriculture for acres planted and production in 2020 concerning the total annual yield for both spring and winter wheat. The data are available from 1895 to 2020 for most states (excluding Alaska, District of Columbia, Hawaii, and Massachusetts where wheat is not grown) and are measured in bushels per acre. It is acknowledged that for the early 1900s there are missing data in area planted, leading to high yields. However, this has no impact on our study because we only use data after 1961.

To reclassify the yield data to the appropriate regions defined as above, we use acres and bushels as the units for the land planted and the total production, respectively, to calculate the yield in each region, rather than directly using the yield data made available by the U.S. Department of Agriculture. That is, the aggregate production in the region per year is divided by the total land planted in the region. Although the standard unit for wheat yield is tonnes (1000 kg) per hectare, no conversion is necessary because our research focuses on analyzing the relationships between weather variables and the yield. The time series plots of wheat yield (in bushels per acre) in each region from 1961 to 2020 are provided in Figure 2. From these plots, wheat yields show an upward trend over time in most regions except for the case of the Southwest Pacific (SWP), which is not quite obvious. We also plot the detrended time series data after differencing in Figure B.2 (see Appendix B).

Weather Variables: We extract the weather data for the period 1961 to 2020 from the unsmoothed, unstandardized data made available by the ACI. This is done to allow the use of simple averaging when converting monthly data to yearly observations to match the wheat yield data. From the ACI in 2021, the variables include the following:

- SL: Monthly sea level measurements via tide gauges measure in millimetres.
- CDD: Maximum number of consecutive days with precipitation less than 1 mm in a year
- RX: Maximum consecutive 5-day precipitation recorded in a month
- T10: Percentage of days in a month in which the lowest temperature recorded is below the 10th percentile
- T90: Percentage of days in a month in which the highest temperature recorded is above the 90th percentile
- WP90: Percentage of days in a month where mean wind power exceeds the 90th percentile

All values are averaged across a 12-month period to find the annual values, with the assumption that all months have equal days. No sea level values are recorded for the Midwest (MID) for geographical reasons. Details on how the weather components contributing the ACI are recorded, calculated, and combined can be found in the ACI Development and Design (ACI 2021).

We note from Figure B.1 (see Appendix B) that sea levels (SL) and T90 have increased across the reference period with T10 decreasing over the period. Some of the observations are in line with the effects of climate change investigated in prior research, where the frequency of high temperatures is increasing, and that of low temperatures decreasing (e.g., Special report: Global warming of 1.5° C by IPCC 2021; Lindsey 2023). The trends for CDD, RX, and WP90 are rather unclear, with major year-on-year fluctuations. Hence, the graphs suggest that SL and T90 may share a long-term relationship with wheat yield in some regions. In the following section, we perform the stationarity test for each weather variable and wheat yield and statistically investigate the cointegration between them.

FIGURE 1. Graphical Representation of the Six Regions

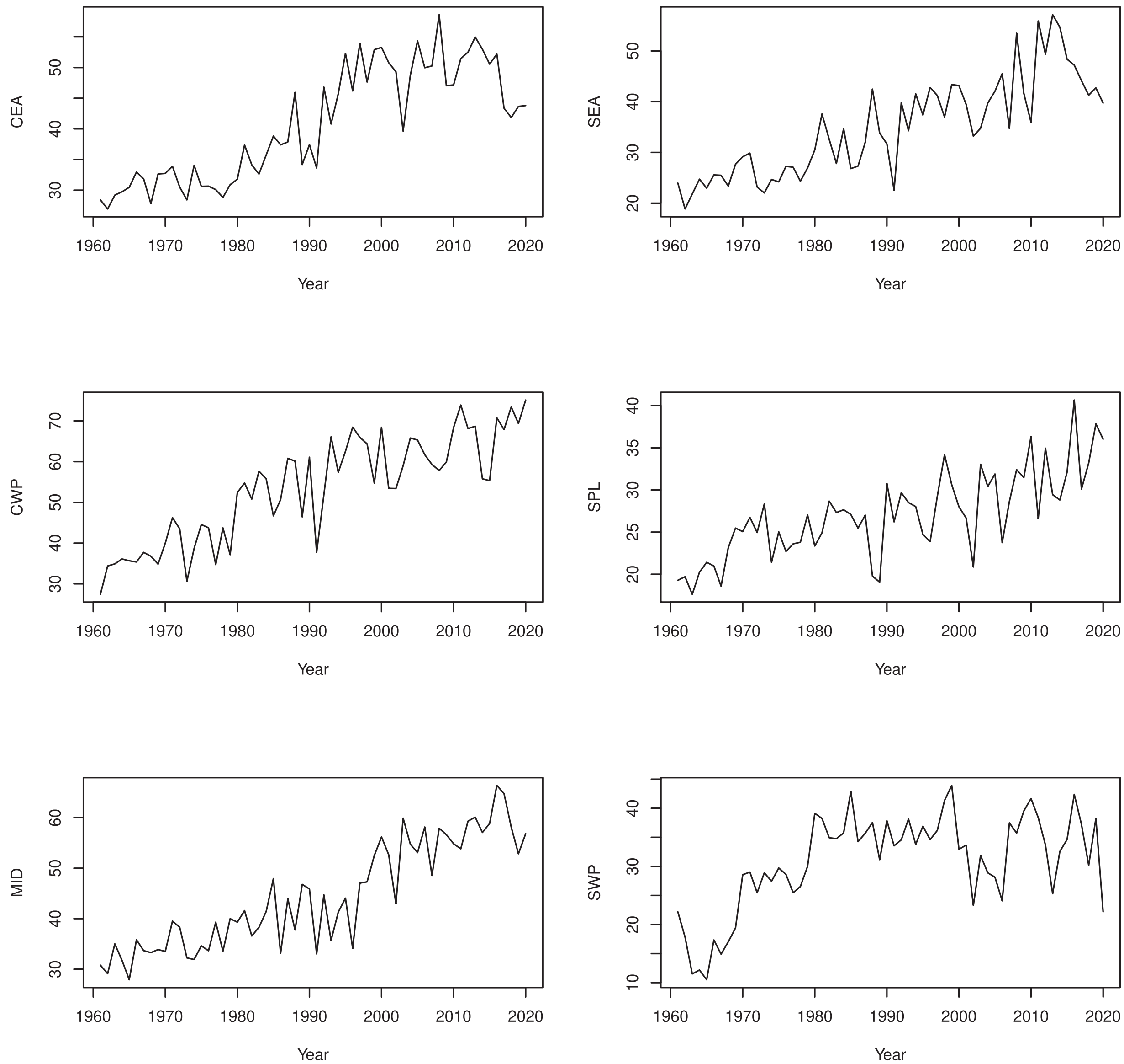

FIGURE 2. Time Series Plots of Wheat Yield in the Defined Regions from 1961 to 2020.

4.2. Stationarity

As described in Section 3.1, to apply the techniques of cointegration, the time series must first be tested for stationarity. If the lag is chosen as 3 with the general rule of thumb used in this test, with 60 observations in our dataset we find that the test statistics from the ADF test on the wheat yield are −1.3718, −3.7972, −3.0167, −3.5363, −4.1890, and −2.2797 for the Central East Atlantic (CEA), Central West Pacific (CWP), MID, Southeast Atlantic (SEA), Southern Plains (SPL), and SWP, respectively. In the following, with a critical value of 5%, we perform the ADF test to determine the order of integration of each time series, and Table 1 summarizes the number of differencing required for stationarity for each time series by region. A few weather variables in certain regions are nonstationary and are integrated to the same order, meeting the first condition of cointegration.

As graphically observed from Figures B.2 and B.3 (see Appendix B), we can now confirm that the differenced time series of yield in all regions and SL in CEA, SEA, and SPL are stationary. In some areas, T90 and WP90 time series exhibit nonstationarity as well.

Alternatively, we also test the individual time series for stationarity using the ADF unit root test. assuming there is a drift and the lag selection is according to the Akaike information criterion (AIC). We obtain the ADF unit root test result on regression based on the equation (there may be additional terms depending on the lags chosen):

TABLE 1
Order of Integration under the ADF Test at 5%

	Yield	SL	CDD	T10	T90	WP90	RX
CEA	1	1	0	0	0	1	0
CWP	1	0	0	0	0	0	0
MID	1	N/A	0	0	0	0	0
SEA	1	1	0	0	0	0	0
SPL	1	1	0	0	0	0	0
SWP	1	0	0	0	1	0	0

$$\Delta y_t = \theta + \omega y_{t-1} + e_t,$$

where e_t is the error term, presumed to be white noise, and y_{t-1} is the lag term. In other words, $y_t = \theta + (\omega + 1)y_{t-1} + e_t$. The associated hypotheses to be tested are as follows:

$$H_0 : \omega = 0, \text{there exists a unit root (i.e., non stationary)} \qquad H_1 : \omega \neq 0, \text{ there is no unit root} \tag{4}$$

The critical values for the ADF test with drift to the above H_0 are −3.51, −2.89, and −2.58 at 1%, 5%, and 10% significance levels, respectively. For the purpose of this step, if the test statistic is greater than −2.89, we fail to reject the null hypothesis at a 5% level, suggesting that the time series is nonstationary and thereby meeting the condition of the Engle-Granger cointegration procedure.

Table 2 shows the corresponding test statistics for all variables across all regions with the regression model after the lag is chosen based on the AIC. The results are consistent with those in Table 1. As an example, we provide detailed results of the ADF test for yield in SPL in Appendix C. Based on the results, wheat yield and SL (except for CWP and SWP regions) are generally nonstationary and have a unit root.

4.3. Cointegration

In the following, for those weather variables integrated to the same order of 1 as yield in Table 1 and being nonstationary as in Table 2, we adopt two approaches to ensure that cointegration is present.

PO test: First, we use a residual-based test (PO test) to test for cointegration. This approach tests the residuals from the linear regression (with an intercept) between the two variables for stationarity with the following null and alternative hypotheses:

$$H_0 : \text{there is no cointegration} \qquad H_1 : \text{there is cointegration} \tag{5}$$

In this case, we consider the wheat yield and the relevant weather factors (excluding CDD, T10, and RX, which have been shown to be stationary from Table 1 at a 5% significance level) when performing the PO test. A p-value less than 5% indicates that the residuals are stationary, and hence the time series are cointegrated. Table 3 shows the results of the PO test for cointegration.

TABLE 2
Test Satistics for the ADF Test

	Yield	SL	CDD	T10	T90	WP90	RX
CEA	−1.725	−0.991	−4.842*	−3.423*	−4.430*	−1.346	−3.275*
CWP	−1.787	−4.285*	−6.247*	−4.147*	−3.472*	−3.283*	−3.019*
MID	−1.533	N/A	−4.997*	−6.442*	−6.622*	−3.729*	−3.909*
SEA	−2.025	−0.079	−4.983*	−4.640*	−3.525*	−2.898*	−4.534*
SPL	−2.372	−0.309	−4.984*	−6.058*	−4.694*	−4.533*	−4.577*
SWP	−2.261	−3.187*	−4.199*	−2.927*	−2.239	−4.053*	−3.988*

Note: *The null hypothesis is rejected; that is, the time series is stationary.

TABLE 3
p-value under the PO Cointegration Test for Yield and Weather Factors

	SL	T90	WP90
CEA	0.011*	—	0.128
CWP	—	—	—
MID	N/A	—	—
SEA	0.01*	—	—
SPL	0.01*	—	—
SWP	—	0.072	—

Note: *The null hypothesis is rejected.

TABLE 4
Test Statistics for the ADF Test (Residuals)

	SL	T90	WP90
CEA	−2.839	—	−2.290
CWP	—	—	—
MID	N/A	—	—
SEA	−3.637*	—	—
SPL	−5.788*	—	—
SWP	—	−2.466	—

Note: *The null hypothesis is rejected.

With a significance level of 5%, SL in all tested regions shows a cointegration with wheat yield. Also, it shows that T90 at SWP and WP90 at CEA are not cointegrated with yield. There seems to be a long-run relationship between sea level increase and the yield observed for these regions. The relationship between sea levels and wheat yield has not been extensively studied by region in the literature; however, rising sea levels are attributed mostly to increasing global temperature, which has been widely studied by many researchers. See the comprehensive review by Kang, Khan, and Ma (2009) for the impact of climate change on crop yield.

The next step is to test the residual of the linear model between the two variables using the ADF test. If the residuals are stationary, then the series are cointegrated and the ECM specification is valid. The weather variables that exhibit nonstationarity at a 5% significance level are included in this test, that is; SL in CEA, SEA, SPL and T90 in SWP, and WP90 in CEA. For two time series to be cointegrated, it must satisfy the condition in (1), where the residuals of the linear regression between the two series are stationary. Hence, we perform the regression on (2) (between each weather variable and the yield) in the respective regions and then use the same test as in the previous step to test for stationarity. Table 4 shows the corresponding test statistics for the test of stationarity in the residuals at a 5% significance level.

With the same hypotheses as in (4), we apply the same critical values as before. That is, if the test statistic with drift is less than −3.51, −2.89, and −2.58 (the critical values without drift are −2.6, −1.95, and −1.61) for significance levels of 1%, 5%, and 10%, respectively, the null hypothesis is rejected; that is, the residuals are stationary. Combining these results, we conclude that wheat yield and SL are cointegrated in SEA and SPL. In other words, in each of these two regions the two time series of wheat yield and SL are jointly denoted as $CI(1,1)$. Thus, we shall obtain the ECM model for these cases in the following section.

Prior to constructing the ECM, we will further test multiple nonstationary time series of extreme weather variables and crop yield for cointegration to confirm the previous results.

Johansen test: From Tables 1 and 2, CEA is the only region with more than two variables having the same order of integration (i.e., yield, SL, and WP90). Although Table 4 indicates that neither SL nor WP90 is cointegrated with yield at CEA, we still conduct the Johansen test for confirmatory purposes. With three variables, we test the null hypotheses of $r = 0, r = 1$, and $r = 2$ sequentially and summarize the results in Table 5. Because the test statistic is smaller than the critical values at significance levels 1%, 5%, and 10%, there is no sufficient evidence to reject the null hypothesis that $r = 0$ (there is no cointegration), a result consistent with those presented in Table 4.

TABLE 5
Test Statistics for the Johansen Test for Yield, SL, and WP90 in CEA

	Test statistics	10%	5%	1%
$r \leq 2$	3.25	7.52	9.24	12.97
$r \leq 1$	10.96	17.85	19.96	24.60
$r = 0$	25.63	32.00	34.91	41.07

TABLE 6
Parameters Estimates in the ECM

Region	SPL			SEA		
Term	Estimate	SE	p-value	Estimate	SE	p-value
α_1	−218.3	46.7	< 0.001	−449.7	142.2	< 0.01
β	0.029	0.011	0.012	−0.037	0.030	0.225
γ	−0.972	0.153	< 0.001	−0.585	0.130	< 0.001
γ_1	0.034	0.007	< 0.001	0.066	0.021	< 0.01

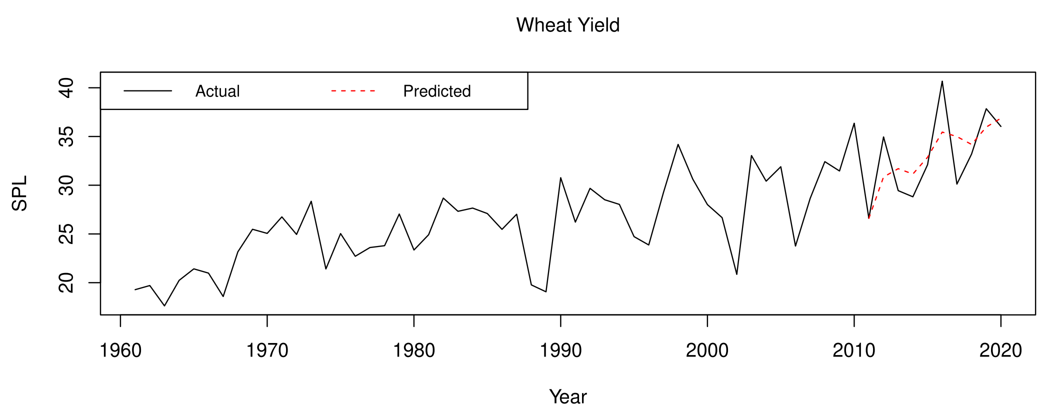

FIGURE 3. Time Series Plot of Actual and Predicted Wheat Yield in SPL.

4.4. Error Correction Model

In this section, for the weather variables satisfying the two conditions specified in the beginning of Section 3.2, we estimate the ECM parameters in (3) and utilize this fitted model for prediction. Based on the results in Sections 4.2 and 4.3, for $I(1)$ and cointegrated time series, we illustrate how to construct the ECM and perform the prediction. From Tables 1 to 4, we choose yield and SL in SEA and SPL for the following analysis. In particular, concerning the sea level SL (x_t) and wheat yield (y_t) for each region, we will construct the ECM model with the following form:

$$\Delta y_t = \hat{\alpha}_1 + \hat{\beta}\Delta x_t + \hat{\gamma} y_{t-1} + \hat{\gamma}_1 x_{t-1}, \qquad \hat{\gamma} < 0, \tag{6}$$

where $\hat{\alpha}_1 = \hat{\alpha} - \hat{\gamma}\hat{a}$ and $\hat{\gamma}_1 = -\hat{\gamma}\hat{b}$. Equation (6) is equivalent to Equation (3) but allows the use of OLS methods to estimate the model parameters. The estimated parameters are shown in Table 6. Because all parameters in the ECM model for SPL are statistically significant at the 5% level, we utilize it to predict future yields. For the time series in SPL, we consider the first 50 years (1961–2010) as training data and the last 10 years as testing data (from 2011 onwards). The prediction for wheat yield in SPL is graphically shown in Figure 3.

Furthermore, we perform diagnostic tests for ECM residuals to ensure the validity and robustness of the ECM. For each of the two regions SPL and SEA, we check the autocorrelation of the residuals using Durbin-Watson (DW) and Breusch-Godfrey

TABLE 7
Diagnostic Tests for the ECM Residuals

Region	SPL		SEA	
Test	Test statistics	*p*-value	Test statistics	*p*-value
DW	1.901	0.646	2.116	0.794
BG	0.845	0.358	1.299	0.254
BP	5.011	0.171	5.111	0.164
White	5.360	0.498	6.210	0.401

TABLE 8
Summary of Test Results for the Cointegration and ECM Specification

Subject	Test	Results
Stationarity[a–c]	ADF	$I(1)$, unit root exists: yield, SL (CEA, SEA, SPL), T90 (SWP), WP90 (CEA)
Cointegration[c]	PO	Exists: yield and SL (CEA, SEA, SPL)
	ADF	Exists: yield and SL (SEA, SPL)
	Johansen	No cointegration: yield, SL, and WP90 (CEA)
ECM residuals[d]	DW, BG	Autocorrelation: insignificant in both SEA and SPL
	BP, White	Heteroscedasticity: insignificant in both SEA and SPL

Notes: [a] The rest of the weather variables in each region are stationary without differencing.
[b] Unit root test shows that the rest of weather variables are stationary.
[c] 5% significance level.
[d] *p*-values for all tests are significantly large from Table 7.

(BG) tests, and we also test for heteroscedasticity using Breusch-Pagan (BP) and White tests. The test results suggest the absence of autocorrelation and heteroscedasticity in the ECM residuals.

We summarize the results of the diagnostic tests together with the key results from the previous parts in Table 8.

In the next section, we shall discuss how our statistical findings can potentially be applied to develop a novel hybrid BMS for crop insurance.

5. INCORPORATING EXTERNAL WEATHER FACTORS IN THE BMS FRAMEWORK

As discussed in Section 1, direct application of the BMS to agricultural insurance is not reasonable because farmers who have experienced losses from systematic risk sources, such as adverse weather events, should not face premium increases as a penalty. Although Rejesus et al. (2006) incorporated these factors as observable components, such an approach may fail to capture the significant long-term relationship between crop yield and weather variables. Moreover, the issue of moral hazard arises in parametric insurance, as studied in Biffis et al. (2022). To address these concerns, we incorporate indemnity crop insurance into the BMS framework while also utilizing information from the cointegrated time series of weather variable that influences crop yield on a long-term basis.

5.1. Illustrative Examples: Modified BMS Transition Rule

In the following, we explain how the external weather factor can be integrated into the design of the BMS framework. To begin, we recall the description of a BMS from Ahn et al. (2022). In a BMS, there are three major components: bonus–malus (BM) levels, transition rule, and BM relativities. Here we focus only on the design of the transition rule. At the beginning of the year, a policyholder is assigned a BM level associated with a BM relativity, such that the premium payable is a product of the BM relativity and the base premium. The base premium is determined based on the policyholder's observable risk characteristics; for example, farm characteristics and the farmer's attributes. At the end of the year, the policyholder's BM level changes according to a prescribed transition rule, leading to a change in the premium. Typically, the transition rule can be described as a discrete-time Markov chain. To incorporate the external weather factor affecting crop yield in the long term, the state space that underpins the transitions of BM levels is enlarged. By combining this weather information with the claim history of individual farmers in the design of the transition rule, premiums can be appropriately adjusted for each farmer to reflect farm-specific risk, without penalizing farmers for losses caused by extreme weather events.

Similar to the practice observed in the Swiss agricultural crop insurance system, our proposed transition rule adjusts premiums based on reported damages. If damages are reported, the BM level may increase, resulting in a subsequent rise in the premium for the following year (malus). On the other hand, if no damage is reported in the current year, the BM level may decrease in the next year (bonus). The maximum and minimum BM levels are fixed beforehand: once the BM level reaches the maximum (respectively minimum), it can no longer increase (respectively decrease) in the next year even if damages are reported (respectively even there is no damage). However, if losses are caused by extreme weather conditions (i.e., when the weather variable exceeds a certain threshold), reported damages will not affect the BM level, and next year's premium remains unchanged.

To explain how the transition rule in a BMS works, we first introduce some notations. Consider a particular policyholder, say the ith policyholder. For convenience, we omit the subscript i for quantities such as claim counts and claim amounts. In the tth policy year, let N_t be the number of claims and $\boldsymbol{Y}_t = (Y_{t1}, ..., Y_{tN_t})$ the vector of associated claim amounts, where Y_{tj} is the jth claim amount. Note that $\boldsymbol{Y}_t$ can be regarded as an empty vector when $N_t = 0$. Also, the aggregate claim amount is defined as $S_t = \sum_{j=1}^{N_t} Y_{tj}$ for $N_t > 0$ and 0 for $N_t = 0$. The BM level at time t is denoted by $L_t,$ which takes an integer value from 1 to z. The largest discount to the base premium is applied at the lowest BM level 1, and the highest premium (with the largest addition to the base premium) is charged at the highest BM level z. Based on the BM level L_{t-1} at time $t-1,$ the t-th policy year's premium is determined. Under the classic BMS design, at the end of the tth year (i.e., at the beginning of the $(t+1)$th year) the BM level increases by h for each reported claim in year t but decreases by 1 when there is no claim reported. Clearly, when the parameter h is larger, the model penalizes policyholders more severely for reporting claims. If only the number of claims in the tth policy year is considered for the transition of BM levels from year t to year $t+1,$ then L_t is defined as

$$L_t := \begin{cases} \min\{L_{t-1} + hN_t, z\}, & N_t > 0, \\ \max\{L_{t-1} - 1, 1\}, & N_t = 0, \end{cases} \tag{7}$$

for $t \in \{1, 2, ...\},$ where it is assumed that a new policyholder without any past history is assigned the BM level $L_0 = \ell_0$ in the beginning.

We now propose to incorporate two factors into the transition rule: (i) the aggregate claim amount S_t in the tth policy year; and (ii) the external weather variable (denoted by x_t in year t), which is cointegrated with the crop yield. It is important to note that Oh, Kim, and Ahn (2022) considered a transition rule based on the individual claim amounts (i.e., the Y_{tj} s) and distinguished between claims above or below a certain threshold level. This is different from our present model, which relies on aggregate (annual) data rather than microlevel claim data. Moreover, the novelty of our transition rule lies in the inclusion of the external weather variable, described as follows. Specifically, extreme weather condition is said to occur when x_t falls below a threshold level d. In principle, the threshold d can be set based on historical weather data and represents a key factor in ensuring that claims resulting from extreme weather conditions beyond the farmer's control do not lead to penalties. For a predetermined $s > 0,$ we classify a positive aggregate claim amount into two categories: small ($S_t < s$) and large ($S_t \geq s$). When $x_t < d$ (indicating severe weather condition), the BM level does not increase at the end of the tth year regardless of the size of S_t. When $x_t \geq d$ (indicating normal weather condition), the BM level increases by h_1 units for large aggregate claim ($S_t \geq s$) but only increases by h_2 units for small aggregate claim ($S_t < s$), where $h_2 < h_1$. Of course, if no claim is reported in the tth year, then the BM level decreases by 1. In our proposed model, instead of the classic transition rule (7), the BM level L_t^* at time t is determined via

$$L_t^* := \begin{cases} \min\{L_{t-1}^* + h_1, z\}, & \text{if} \quad S_t \geq s \text{ and } x_t \geq d, \\ \min\{L_{t-1}^* + h_2, z\}, & \text{if} \quad 0 < S_t < s \text{ and } x_t \geq d, \\ L_{t-1}^*, & \text{if} \quad S_t > 0 \text{ and } x_t < d, \\ \max\{L_{t-1}^* - 1, 1\}, & \text{if} \quad S_t = 0, \end{cases} \tag{8}$$

for $t \in \{1, 2, ...\},$ where a new policyholder without any history starts with $L_0^* = \ell_0$.

The transition rule can be modified in various ways by further incorporating more realistic and suitable conditions into the BMS design. In particular, if the severe weather condition persists for w consecutive years (i.e., $x_k < d$ for $k = t-w+1, ..., t-1, t$), then the policyholder retains the same BM level as in the previous year even if there are reported claims. In this case, we must redefine the state space of the BM levels in (8) to keep track of the information concerning the number of periods with severe weather, thereby ensuring that the transition rule is Markovian. A similar approach of enlarging the state space was adopted in Ahn et al. (2022), who studied a different problem of requiring the policyholder to have consecutive claim-free years to decrease his or her BM level. In our case, for each $\ell = 1, ..., z$ the BM level ℓ is augmented to

$$(\ell)_1, (\ell)_2, ..., (\ell)_w.$$

Here the subscript indicates the number of additional years needed to reach w consecutive years of severe weather condition by the end of the year. Hence, under the augmented system, the number of BM levels is increased to $z \times w$. It is important to point out that when a policyholder occupies an augmented BM level $(\ell)_j$, the associated BM relativity should only depend on ℓ but not j because the subscript j has been artificially introduced to make the transition rule Markovian. For example, when the augmented BM level is $(\ell)_1$ at time $t-1$, it means that the BM level is ℓ and there have already been at least $w-1$ consecutive periods of severe weather condition. We distinguish between the following cases:

i. If there is a claim in year t,
   a. if the severe weather condition continues in year t (i.e., $x_t < d$), regardless of the aggregate claim amount, the augmented BM level next year remains unchanged at $(\ell)_1$;
   b. if $x_t \geq d$, then the streak of severe weather condition is broken and the subscript of the augmented BM level becomes w, and the BM level ℓ increases by h_1 or h_2 units depending on whether the aggregate claim S_t is larger than or less than s.

ii. If there is no claim in year t, then the BM level shall decrease by one unit regardless of whether the severe weather condition continues, but the subscript of the augmented BM level remains at 1 if $x_t < d$ and becomes w if $x_t \geq d$.

Consolidating the above, we arrive at the result that, given the augmented BM level $L_{t-1}^{**} = (\ell)_1$ at time $t-1$, the augmented BM level L_t^{**} at time t is specified by

$$L_t^{**} := \begin{cases} (\min\{\ell + h_1, z\})_w, & \text{if} \quad S_t \geq s \text{ and } x_t \geq d, \\ (\min\{\ell + h_2, z\})_w, & \text{if} \quad 0 < S_t < s \text{ and } x_t \geq d, \\ (\ell)_1, & \text{if} \quad S_t > 0 \text{ and } x_t < d, \\ (\max\{\ell - 1, 1\})_w, & \text{if} \quad S_t = 0 \text{ and } x_t \geq d, \\ (\max\{\ell - 1, 1\})_1, & \text{if} \quad S_t = 0 \text{ and } x_t < d, \end{cases} \tag{9}$$

for $t \in \{1, 2, ...\}$. Similarly, given the augmented BM level $L_{t-1}^{**} = (\ell)_j$ for $j \in \{2, ..., w\}$ at time $t-1$, we obtain

$$L_t^{**} := \begin{cases} (\min\{\ell + h_1, z\})_w, & \text{if } S_t \geq s \text{ and } x_t \geq d, \\ (\min\{\ell + h_2, z\})_w, & \text{if } 0 < S_t < s \text{ and } x_t \geq d, \\ (\min\{\ell + h_1, z\})_{j-1}, & \text{if } S_t \geq s \text{ and } x_t < d, \\ (\min\{\ell + h_2, z\})_{j-1}, & \text{if } 0 < S_t < s \text{ and } x_t < d, \\ (\max\{\ell - 1, 1\})_w, & \text{if } S_t = 0 \text{ and } x_t \geq d, \\ (\max\{\ell - 1, 1\})_{j-1}, & \text{if } S_t = 0 \text{ and } x_t < d, \end{cases} \tag{10}$$

for $t \in \{1, 2, ...\}$. Note that when $x_t \geq d$, the counter that monitors the number of years required to achieve w consecutive years of severe weather condition restarts from w. On the other hand, when $x_t < d$, the counter is reduced by 1. When $w = 1$, the definition (10) is not needed (because $\{2, ..., w\}$ is an empty set) and therefore the proposed model is only defined via (9), which reduces to the previous model (8).

The following example illustrates the transition rules with a certain extreme weather information.

Example 1 Let us consider the BMS with two different transition rules: (a) the rule defined by (8) and (b) the rule specified via (9) and (10). We assume that $h_1 = 2$, $h_2 = 1$, $s = 100$, $d = 7$, and $z = 10$ in both cases. The sea level in year t is recorded as x_t, which is statistically cointegrated with the crop yield, and when it falls below $d = 7$ the weather condition is regarded as

TABLE 9
Evolution of L_t^* (Rule [8]) and L_t^{**} (Rules [9] and [10]) with $h_1 = 2, h_2 = 1$

t	0	1	2	3	4	5	6	7	8	9
S_{t+1}	70	50	90	102	104	60	30	0	0	
x_{t+1}	7.5	7.3	6.8	6.7	7.2	7.1	7.1	7.5	7.6	
L_t^*	5	6	7	7	7	9	10	10	9	8
L_t^{**}	$(5)_2$	$(6)_2$	$(7)_2$	$(8)_1$	$(8)_1$	$(10)_2$	$(10)_2$	$(10)_2$	$(9)_2$	$(8)_2$

severe. Thus, even when $S_t > 0,$ the BM level (and hence premium) is unchanged in the $(t+1)$th policy year if $x_t < 7$ in Case (a) and if $x_{t-1} < 7$ and $x_t < 7$ in Case (b) (that is, the sea level is below the threshold for $w = 2$ consecutive years).

When a claim is reported under normal weather condition $(x_t \geq 7),$ the BM level is to be increased by two units if the aggregate claim amount is large ($S_t \geq 100$) and by one unit if the aggregate claim is small ($S_t < 100$). Given the BM level (L_t^* in Case (a) and L_t^{**} in Case (b)) at the time t as well as the aggregate claim amount S_{t+1} and the sea level x_{t+1} in the $(t+1)$th year, the BM level at time $t+1$ is determined according to (8) in Case (a) and (9) and (10) in Case (b). We assume that a new policyholder without claim history starts at the BM level $L_0^* = 5$ in Case (a) and $L_0^{**} = (5)_2$ in Case (b). We summarize the evolution of the BM levels from time 0 to 9 in Table 9 and explain the first few transitions as follows:

i. Transition from $t = 0$ to $t = 1$:
   a. **Rule (8)**: The reported aggregate claim at $t = 1$ is small (i.e., $S_1 = 70 < 100$) and the weather condition was normal in the first year (i.e., $x_1 = 7.5 > 7$), and therefore the BM level moves up to $L_1^* = 6\ (= 5 + 1)$.
   b. **Rules (9) and (10)**: The BM level also increases by $h_2 = 1,$ and the augmented BM level becomes $L_1^{**} = (6)_2$ where the subscript remains at $w = 2$ as $x_1 > 7$.

ii. Transition from $t = 1$ to $t = 2$:
   a. **Rule (8)**: Again, the aggregate claim reported at $t = 2$ is small (i.e., $S_2 = 50 < 100$) and the weather condition was not severe in the second year (i.e. $x_2 = 7.3 > 7$). The BM level increases by $h_2 = 1$ to $L_2^* = 7$.
   b. **Rules (9) and (10)**: Same as the previous year, the BM level moves up to 7 and the subscript still remains at 2 because $x_2 = 7.3 > 7$.

iii. Transition from $t = 2$ to $t = 3$:
   a. **Rule (8)**: Because the sea level in the third year $x_3 = 6.8 < 7$ is below its threshold, the BM level remains unchanged even claims are reported as $S_3 = 90$. That is, $L_3^* = 7$.
   b. **Rules (9) and (10)**: Though the aggregate claim amount $S_3 = 90 < 100$ in the third year is small, it is the first time that there is severe weather condition in the year, with $x_3 < 7$ (and it will take another year of severe weather condition for a reported claim not to be penalized). Thus, the BM level increases by $h_2 = 1$ and the subscript decreases by 1 so that $L_3^{**} = (8)_1$.

iv. Transition from $t = 3$ to $t = 4$:
   a. **Rule (8)**: The sea level continues to be below the threshold in the fourth year as $x_4 = 6.7 < 7$. Therefore, the BM level remains the same although the aggregate claim is large at $S_4 = 102 > 100$.
   b. **Rules (9) and (10)**: The BM level remains unchanged as the weather condition has been severe for two consecutive years; thus, $L_4^{**} = (8)_1$.

v. Transition from $t = 4$ to $t = 5$:
   a. **Rule (8)**: Because the weather condition is not severe in the fifth year with $x_5 = 7.2 > 7,$ the change in BM level depends on the size of the aggregate claim. A large aggregate claim amount of $S_5 = 104 > 100$ is reported in the year, so the BM level moves up to $L_5^* = 9\ (= 7 + 2)$.
   b. **Rules (9) and (10)**: The BM increases by $h_1 = 2$ owing to the large aggregate claim amount, and the subscript restarts from $w = 2$. Hence, $L_5^{**} = (10)_2$.

Clearly, consecutive periods of severe weather conditions can have a significant impact on the transition between BM levels. A large w represents a stricter transition rule for the policyholder because his or her reported claims are attributed to systematic risk (and hence not penalized) only when there are w consecutive years of severe weather condition. The selection of the parameter w may depend on the specific situation, such as the degree to which the severe weather condition impacts crop yields in a given region. This impact may result from a single extreme event or persistent extreme events over multiple years that significantly affect crop yields and is beyond the control of farmers.

5.2. Challenges and Strategies for Implementing BMS in Crop Insurance

Implementation of a BMS within the proposed framework in crop insurance, which incorporates severe weather information into the classical transition rule, presents several challenges and requires specific strategies. In the following, we outline these challenges and provide a roadmap, highlighting key factors to consider when constructing this type of hybrid BMS.

Challenges:

- *Data availability and quality*: Inconsistent or incomplete data can impact the accuracy and reliability of the BMS. The ACI, which provides trends in aggregated extreme weather data, currently lacks granularity. The insignificance of weather variables might result from relying on regional average values rather than location-specific data.
- *Non-climate variables*: To accurately assess the impact of extreme weather variables on crop yields, non-climate variables affecting expected claims must be considered when determining premiums. For example, in auto insurance, initial ratemaking uses known variables, and residual unknown risks are accounted for in subsequent posteriori ratemaking.
- *System construction cost*: The proposed hybrid BMS requires inputs of trigger parameters and the development of a pricing framework for individual farmers. As a result, costs associated with building the system and collecting the necessary data must be anticipated.
- *Stakeholder resistance*: Though the BMS is widely adopted in auto insurance, it is a new concept for farmers and crop insurers. Resistance may arise due to unfamiliarity with the technology or skepticism about its benefits. Addressing concerns and building trust through education and demonstrating the value of the BMS will be essential.
- *Demand for income protection*: Farmers are increasingly adopting higher levels of risk management tools to prepare for large-scale catastrophic weather conditions, but price fluctuations in commodities remain a significant risk. For example, only a limited number of Australian farmers are able to fully leverage the liquidity of financial markets to protect their growth against such volatilities.

Because implementing the BMS in crop insurance is a long-term initiative, we focus on creating a roadmap for the pilot project below. A successful pilot will generate greater consensus and demand for broader implementation of this system in crop insurance.

A roadmap for implementing the proposed BMS:

- *Identify key stakeholders*: Identify and engage key stakeholders, including farmers, insurers, and regulators, to establish clear objectives for the BMS implementation.
- *Gather detailed data*: Collect granular data on premiums, crop yields, and relevant weather variables, matched to geographical areas of interest. Implement pilot studies in selected regions to gather initial data and test the feasibility of the BMS.
- *Consider aspects specific to different regions and crops*: The severity of extreme weather events varies by region, so the model must adjust thresholds based on region-specific climate data. In addition, different crops respond differently to weather conditions at various growth stages. For example, wheat is vulnerable to frost during early growth and drought during grain filling, and rice is susceptible to flooding during the vegetative stage. Tailoring the model for specific regions and crops enables insurers to set more accurate, risk-adjusted premiums and ensures equitable pricing. It also helps policymakers target subsidies for regions and crops most affected by extreme weather.
- *Develop the BMS structure*: Design the overall structure of the BMS, integrating various data sources. Further explore the use of explanatory variables and latent variables to create a more accurate ratemaking system in crop insurance, similar to the approach in auto insurance.
- *Run the BMS in pilot regions*: Implement and monitor the BMS in pilot regions over multiple seasons. Adjust the system based on performance data and user feedback. Ensure continuous monitoring and improvement to regularly update the system, such as transition rule and claim frequency model.

- *Enhance the BMS*: Develop a more comprehensive risk management tool to benefit farmers beyond merely updating their premiums based on yield experience and weather conditions.

By addressing these barriers and following a structured roadmap, the successful implementation of the BMS could enhance the current model of weather index crop insurance and improve long-term risk management together with agricultural productivity.

6. CONCLUDING REMARKS

Several studies have examined the correlation between crop yields and weather variables in weather index insurance. However, there has been a lack of research on joint models that analyze the co-movement of these variables over a long period of time. Given that severe weather conditions can have lasting and time-dependent impacts on crop yields, it is essential to conduct updated analyses to better understand their relationship. This understanding, in turn, will enable the appropriate selection of trigger parameter in weather index insurance.

We note that the cointegration model is a good candidate for capturing long-run connections between two nonstationary time series. The cointegration analysis conducted in this article provides evidence of cointegration between crop yield and sea level in some regions of the United States. Whereas previous studies have often focused on the impact of weather variables on specific crops in certain regions of the United States, this study aims to analyze wheat yield and extreme weather variables (rather than just averages) across all regions of the United States. We use extreme weather data from the ACI, which serves as a valuable source of such information. Although the ACI is designed to show trends in extreme weather rather than to analyze the potential effects of climate change, this study demonstrates an alternative application of the ACI. Specifically, it aids certain industries, such as crop insurance, in modeling the long-term relationship between weather metrics and a proxy for expected claims (in our case, wheat yield). It is important to highlight that this research differs from the Actuaries Climate Risk Index, which provides estimates of the relationship between the ACI's weather metrics and weather-related losses.

To address the issue of basis risk in weather index crop insurance, Biffis et al. (2022) proposed bundling insurance with debt to minimize moral hazard. This approach has shown success in developing countries. For developed countries where weather data are more widely available, the proposed BMS framework can be utilized as an indemnity insurance. By incorporating the extreme weather factors that are cointegrated with crop yield, this framework can design appropriate transition rules, thereby addressing issues related to systematic risk and asymmetric information in crop insurance. The proposed design of the transition rule still offers incentives in the form of reduced premiums for those who make no claims and penalizes those who file multiple and/or large claims, encouraging farmers to take better care of their farm and adopt risk-reducing strategies. At the same time, claims resulting from events beyond the farmer's control are treated fairly, allowing for appropriate adjustments in ratemaking under severe weather conditions. Despite the challenges related to limited data and other aspects discussed in Section 5.2, the concept of BMS can help incentivize responsible behavior, promote risk reduction, and foster equity in crop insurance pricing.

For future research, we plan to design a comprehensive BMS tailored for crop insurance based on the transition rules proposed in this article. In addition to data on crop yields and updated information on extreme weather variables available in the public domain, we will need information on individual farm losses and characteristics to accurately calculate crop insurance premiums and compare various ratemaking models discussed in the literature. In particular, determining the optimal BM relativities will be crucial to ensure fair pricing of insurance premiums. We hope this research will provide valuable insights into the impact of climate change on crop insurance.

ACKNOWLEDGMENTS

We sincerely thank the co-editor and the two anonymous reviewers for their constructive comments and suggestions, which have undoubtedly improved the original manuscript. EC and JKW acknowledge the support from the Australian Research Council's Discovery Project DP200100615.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

REFERENCES

Actuaries Climate Index. 2021. https://actuariesclimateindex.org/

Ahn, J. Y., E. C. K. Cheung, R. Oh, and J.-K. Woo. 2022. Optimal relativities in a modified bonus–malus system with long memory transition rules and frequency–severity dependence. *Variance* 15 (2).

Biffis, E., E. Chavez, A. Louaas, and P. Picard. 2022. Parametric insurance and technology adoption in developing countries. *Geneva Risk and Insurance Review* 47 (1):7–44. 10.1057/s10713-020-00061-0

Bishai, D. M. 1995. Infant mortality time series are random walks with drift: Are they cointegrated with socioeconomic variables? *Health Economics* 4 (3): 157–67. 10.1002/hec.4730040302

Chambers, R. G. 1989. Insurability and moral hazard in agricultural insurance markets. *American Journal of Agricultural Economics* 71 (3):604–16. 10.2307/1242016

Deng, X., B. J. Barnett, G. Hoogenboom, Y. Yu, and A. Garcia y Garcia. 2007. Alternative crop insurance indexes. *Journal of Agricultural and Applied Economics* 40 (1):223–37. 10.1017/S1074070800028078

Elabed, G., M. F. Bellemare, M. R. Carter, and C. Guirkinger. 2013. Managing basis risk with multiscale index insurance. *Agricultural Economics* 44 (4–5): 419–31. 10.1111/agec.12025

Engle, R. F., and C. W. J. Granger. 1987. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55 (2):251–76. 10.2307/1913236

Gonzalo, J., and T. Lee. 1998. Pitfalls in testing for long run relationships. *Journal of Econometrics* 86 (1):129–54. 10.1016/S0304-4076(97)00111-5

Granger, C. W. J. 1981. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16 (1):121–30. 10.1016/0304-4076(81)90079-8

Granger, C. W. J., and A. A. Weiss. 1983. Time series analysis of error-correction models. In *Studies in econometrics, time series, and multivariate statistics*, ed. S. Karlin, T. Amemiya, and L. A. Goodman, 255–78. New York: Academic Press.

Intergovernmental Panel on Climate Change. 2021. *Special report: Global warming of 1.5°C*. https://www.ipcc.ch/sr15/.

Johansen, S. 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59 (6):1551–80. 10.2307/2938278

Kang, Y., S. Khan, and X. Ma. 2009. Climate change impacts on crop yield, crop water productivity and food security—A review. *Progress in Natural Science* 19 (12):1665–74. 10.1016/j.pnsc.2009.08.001

Kim, Y., J. Yu, and D. L. Pendell. 2020. Effects of crop insurance on farm disinvestment and exit decisions. *European Review of Agricultural Economics* 47 (1):324–47.

Lemaire, J. 1995. *Bonus–malus systems in automobile insurance*. Dordrecht, Netherlands: Kluwer Academic.

Li, H., and Q. Tang. 2022. Joint extremes in temperature and mortality: A bivariate POT approach. *North American Actuarial Journal* 26 (1):43–63. 10.1080/10920277.2020.1823236

Lindsey, R. 2023. *Climate change: Global sea level*. The National Oceanic and Atmospheric Administration. https://www.climate.gov/.

Maddala, G. S., and I. Kim. 1998. *Unit roots, cointegration, and structural change*. Cambridge, UK: Cambridge University Press.

Mason, C., D. J. Hayes, and S. H. Lence. 2003. Systemic risk in U.S. crop reinsurance programs. *Agricultural Finance Review* 63 (1):23–39. 10.1108/00214980380001139

Miranda, M. J., and J. W. Glauber. 1997. Systemic risk, reinsurance, and the failure of crop insurance markets. *American Journal of Agricultural Economics* 79 (1):206–15. 10.2307/1243954

Nadolnyak, D., D. Vedenov, and J. Novak. 2008. Information value of climate-based yield forecasts in selecting optimal crop insurance coverage. *American Journal of Agricultural Economics* 90 (5):1248–55. 10.1111/j.1467-8276.2008.01212.x

Nelson, C. H., and E. T. Loehman. 1987. Further toward a theory of agricultural insurance. *American Journal of Agricultural Economics* 69 (3):523–31. 10.2307/1241688

Oh, R., J. H. T. Kim, and J. Y. Ahn. 2022. Designing a bonus–malus system reflecting the claim size under the dependent frequency-severity model. *Probability in the Engineering and Informational Sciences* 36 (4):963–87. 10.1017/S0269964821000188

Pan, Q., L. Porth, and H. Li. 2022. Assessing the effectiveness of the Actuaries Climate Index for estimating the impact of extreme weather on crop yield and insurance applications. *Sustainability* 14 (11):6916. 10.3390/su14116916

Phillips, P. C. B., and S. Ouliaris. 1990. Asymptotic properties of residual based tests for cointegration. *Econometrica* 58 (1):165–93. 10.2307/2938339

Quiggin, J., G. Karagiannis, and J. Stanton. 1993. Crop insurance and crop production: An empirical study of moral hazard and adverse selection. *Australian Journal of Agricultural Economics* 37 (2):95–113. 10.1111/j.1467-8489.1993.tb00531.x

Rejesus, M. R., K. H. Coble, T. O. Knight, and Y. Jin. 2006. Developing experience-based premium rate discounts in crop insurance. *American Journal of Agricultural Economics* 88 (2):409–19. 10.1111/j.1467-8276.2006.00867.x

Skees, J. R., and M. R. Reed. 1986. Rate making for farm-level crop insurance: Implications for adverse selection. *American Journal of Agricultural Economics* 68 (3):653–59. 10.2307/1241549

Smith, V. H., and B. K. Goodwin. 1996. Crop insurance, moral hazard, and agricultural chemical use. *American Journal of Agricultural Economics* 78 (2): 428–38. 10.2307/1243714

Vilar-Zanón, J. L., A. Heras, and E. de Frutos. 2020. An average model approach to experience based premium rates discounts: An application to Spanish agricultural insurance. *European Actuarial Journal* 10 (2):361–75. 10.1007/s13385-020-00234-1

Vroege, W., T. Dalhaus, and R. Finger. 2019. Index insurances for grasslands—A review for Europe and North-America. *Agricultural Systems* 168:101–11. 10.1016/j.agsy.2018.10.009

Woodard, J. D. 2014. Impacts of weather and time horizon selection on crop insurance ratemaking: A conditional distribution approach. *North American Actuarial Journal* 18 (2):279–93. 10.1080/10920277.2014.887476

Woodard, J. D., and P. Garcia. 2008. Weather derivatives, spatial aggregation, and systemic risk: Implications for reinsurance hedging. *American Journal of Agricultural Economics* 33 (1):34–51.

APPENDIX A. ACI REGION DEFINITIONS

The following were extracted from the ACI (2021):

- CEA: Central East Atlantic (CT, DC, DE, MA, MD, ME, NH, NJ, NY, PA, RI, VT, WV)
- CWP: Central West Pacific (WA, OR, ID)
- MID: Midwest (IA, IL, IN, MI, MN, MO, OH, WI)
- SEA: Southeast Atlantic (AL, AR, FL, GA, KY, LA, MS, NC, SC, TN, VA)
- SPL: Southern Plains (KS, MT, ND, NE, OK, SD, TX, WY)
- SWP: Southwest Pacific (AZ, CA, CO, NM, NV, UT)

APPENDIX B. TIME SERIES PLOTS

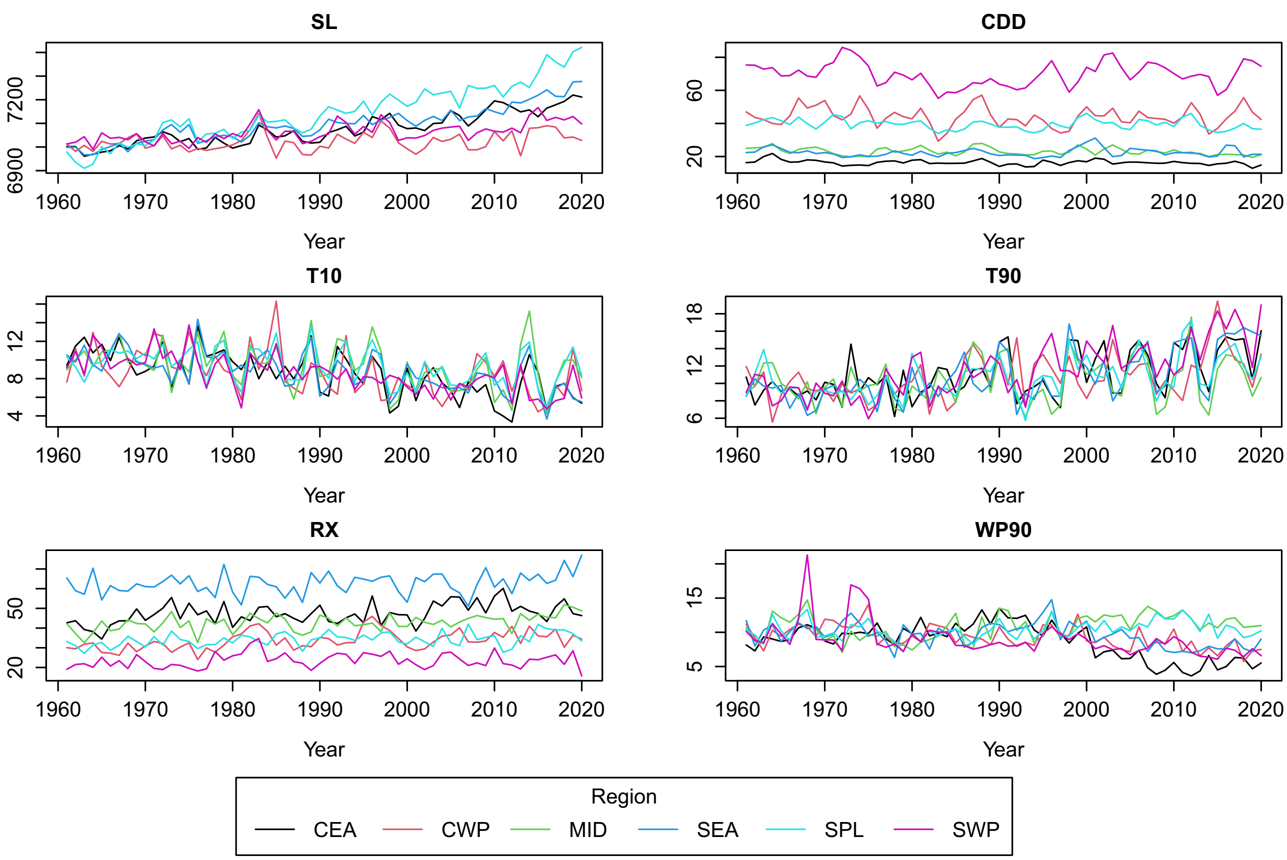

FIGURE B.1. Time Series Plots of Various Weather Variables in the Defined Regions from 1961 to 2020.

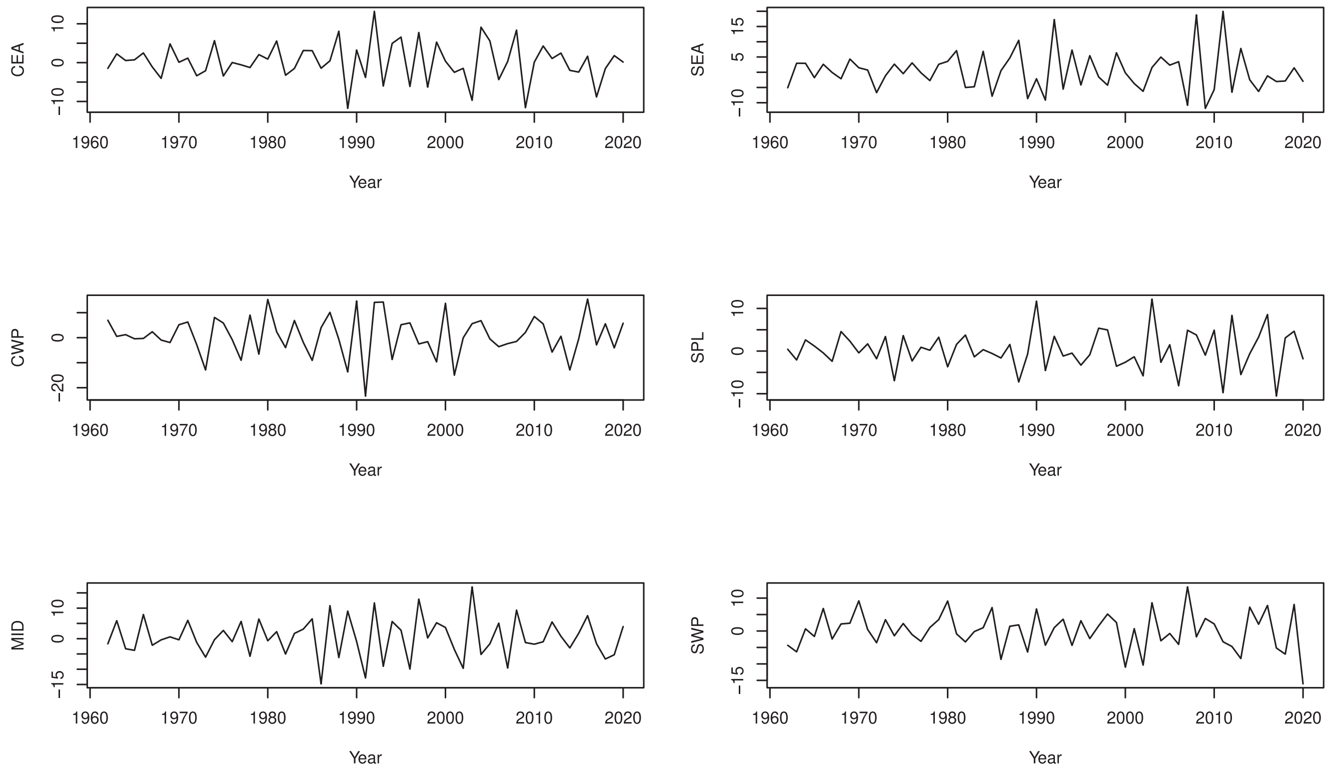

FIGURE B.2. Detrended Time Series Plots of Wheat Yield in the Defined Regions from 1961 to 2020.

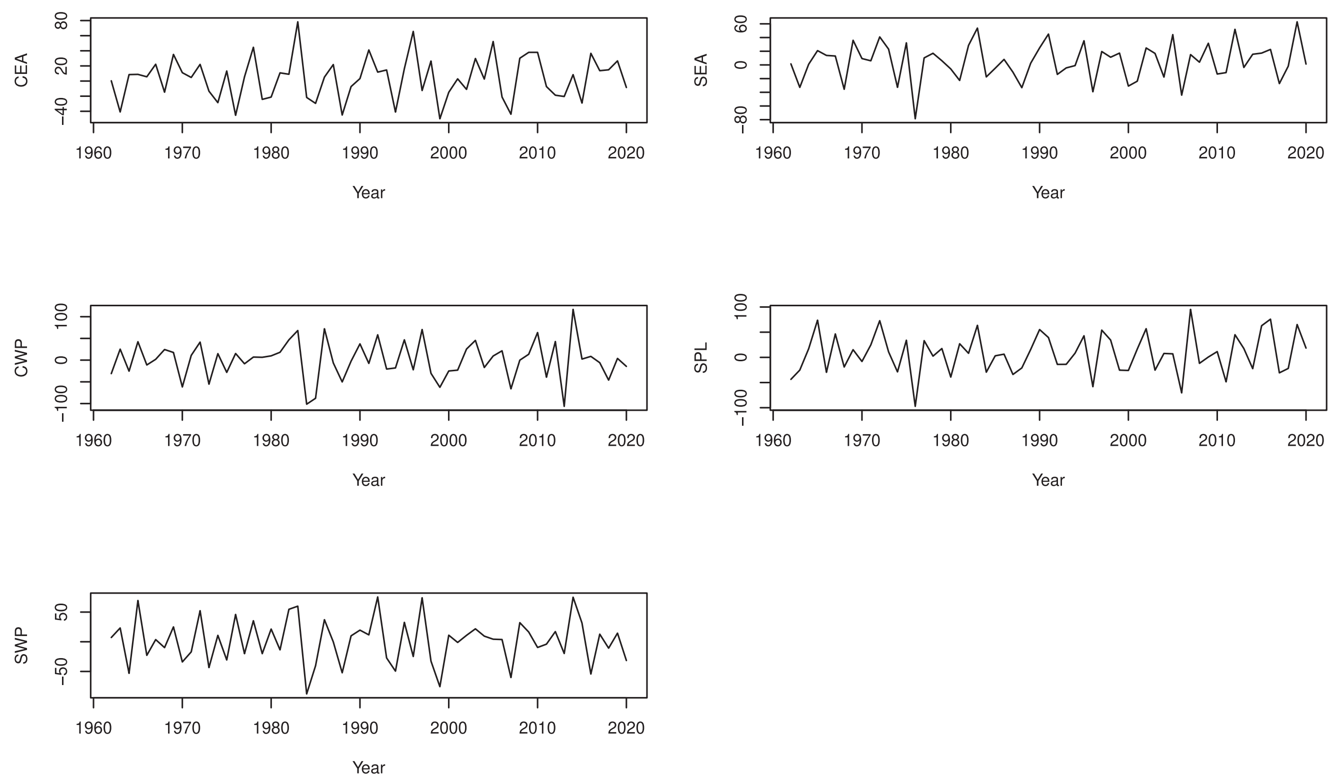

FIGURE B.3. Detrended Time Series Plots of Monthly SL in the Defined Regions from 1961 to 2020.

APPENDIX C. STATISTICAL TEST RESULTS

The ADF test example: For the Engle-Granger two-step procedure for cointegration in Section 4.2, the detailed results of the ADF test (unit root test) for weather variables in some regions are given in the following. In this example, the test is based on $\Delta y_t = \theta + \omega y_{t-1} + e_t$. The first statistic (τ_2) is to test $\omega = 0$ and the second statistic (ϕ_1) is to test $\omega = \theta = 0$.

For yield in `SPL`: `ur.df(ts.yield[,"SPL"],lags = 1,type="drift") returns`

```
Test regression trend
Call: lm(formula = z.diff z.lag.1 + 1 + z.diff.lag)
Coefficients:
```

	Estimate Std.	Error	t Value	Pr(> $\|t\|$)
(Intercept)	8.3037	3.3613	2.470	0.0166*
z.lag.1	−0.2928	0.1235	−2.372	0.0212*
z.diff.lag	−0.3077	0.1313	−2.343	0.0228*

```
-Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 4.001 on 55 degrees of freedom multiple R-squared: 0.2853,
Adjusted R-squared: 0.2593 F-statistic: 10.98 on 2 and 55 DF; p-value: 9.733e-05
value of test statistic is: −2.372 3.1453.
Critical values for test statistics:
```

	1pct	5pct	10pct
tau2	−3.51	−2.89	−2.58
phi1	6.70	4.71	3.86

Here, the first statistic for τ_2 (−2.372) is greater than the critical values at all significance levels, and we do not reject the null hypothesis of nonstationarity (i.e., $\omega = 0$), implying that yield time series on SPL is nonstationary. The second statistic for ϕ_1 (3.1453) is smaller than the critical values at all levels, and we fail to reject $\omega = \theta = 0$. The coefficient of intercept is shown to be statistically significant at the level 5%.