Farmers credit risk modelling under climate uncertainty

Matthew Ames[4] Guillaume Bagnarosa[1,5] Suikai Gao[1]
Tomoko Matsui[4] Gareth W.Peters[2,3]

[1]Rennes School of Business, France
[2]Heriot Watt, UK
[3]Oxford-Man Institute, Oxford University, UK
[4]ISM, Japan
[5]University College London, UK

INRIA November 2018

Outline

Introduction

- Weather is inherently one of the most important factors deciding farming practices related to:

Introduction

- Weather is inherently one of the most important factors deciding farming practices related to:
  - Land use

Introduction

- Weather is inherently one of the most important factors deciding farming practices related to:
  - Land use
  - Yield improvement or hedging measures

Introduction

- Weather is inherently one of the most important factors deciding farming practices related to:
  - Land use
  - Yield improvement or hedging measures
  - Economic policies in agricultural sectors

Introduction

- Weather is inherently one of the most important factors deciding farming practices related to:
  - Land use
  - Yield improvement or hedging measures
  - Economic policies in agricultural sectors
- While the hedging solutions against climatic hazard remain under developed outside of US and Canada (**Paulson et al. 2010**)

Introduction

- Weather is inherently one of the most important factors deciding farming practices related to:
  - Land use
  - Yield improvement or hedging measures
  - Economic policies in agricultural sectors
- While the hedging solutions against climatic hazard remain under developed outside of US and Canada (**Paulson et al. 2010**)

  ⇒ According to the CRED, in 2016, the economic losses due to floods and droughts continue to increase, making them one of the most damaging natural disasters for our economies worldwide.

Commodities and Weather Risk

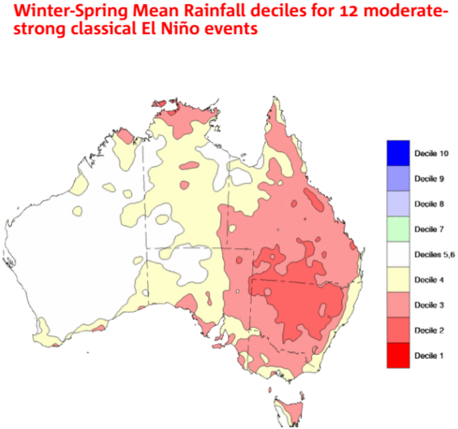


Figure: El Niño background, source: nab

.

Impact of El Niño

- El Niño affects temperature and rainfall in North and South America, Africa, East and Southeast Asia, the Indian subcontinent, Australia and the Pacific.

Impact of El Niño

- El Niño affects temperature and rainfall in North and South America, Africa, East and Southeast Asia, the Indian subcontinent, Australia and the Pacific.
- Locally, the phenomenon causes generally lower winter and spring rainfall in northern and eastern Australia.

Impact of El Niño

- El Niño affects temperature and rainfall in North and South America, Africa, East and Southeast Asia, the Indian subcontinent, Australia and the Pacific.
- Locally, the phenomenon causes generally lower winter and spring rainfall in northern and eastern Australia.
- The impact of any given El Niño event is highly variable. Many previous El Niño events have been associated with lower farm GDP. Real farm GDP declined between 0.7% and 25.4% during the last five El Niños, with an average decline of 12.6%.

Impact of El Niño

- El Niño affects temperature and rainfall in North and South America, Africa, East and Southeast Asia, the Indian subcontinent, Australia and the Pacific.
- Locally, the phenomenon causes generally lower winter and spring rainfall in northern and eastern Australia.
- The impact of any given El Niño event is highly variable. Many previous El Niño events have been associated with lower farm GDP. Real farm GDP declined between 0.7% and 25.4% during the last five El Niños, with an average decline of 12.6%.
- "Pressing need for an adapted, durable and scalable hedging solution" (OECD)

Impact of El Niño

- El Niño affects temperature and rainfall in North and South America, Africa, East and Southeast Asia, the Indian subcontinent, Australia and the Pacific.
- Locally, the phenomenon causes generally lower winter and spring rainfall in northern and eastern Australia.
- The impact of any given El Niño event is highly variable. Many previous El Niño events have been associated with lower farm GDP. Real farm GDP declined between 0.7% and 25.4% during the last five El Niños, with an average decline of 12.6%.
- "Pressing need for an adapted, durable and scalable hedging solution" (OECD)

$\Rightarrow$ who fosters the development of insurance products such as weather, crop or revenue insurance policies which could help the agricultural businesses to overcome the more frequent and damaging weather events.

Outline

The Balance Sheet

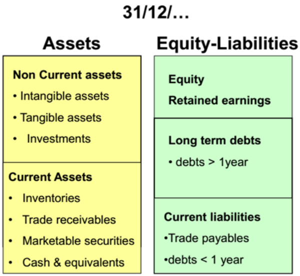


Figure: Balance Sheet

- Double entry principle means that assets always equal liabilities

The Balance Sheet

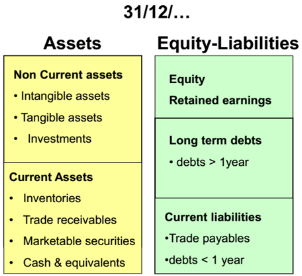


Figure: Balance Sheet

- Double entry principle means that assets always equal liabilities
- In our case we are going to model the dynamic of the assets $\Leftrightarrow$ dynamic of the liabilities $\Longrightarrow$ dynamic of the profits

The Balance Sheet

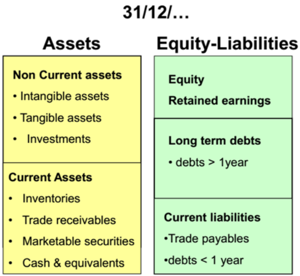


Figure: Balance Sheet

- Double entry principle means that assets always equal liabilities
- In our case we are going to model the dynamic of the assets $\Leftrightarrow$ dynamic of the liabilities $\Longrightarrow$ dynamic of the profits
- The profits are thus partially explaining the dynamic of the assets

Merton's Main Assumptions

- No transaction costs nor taxes.

Merton's Main Assumptions

- No transaction costs nor taxes.
- We can borrow and lend at the same rate of interest

Merton's Main Assumptions

- No transaction costs nor taxes.
- We can borrow and lend at the same rate of interest
- Short sales of all assets are allowed

Merton's Main Assumptions

- No transaction costs nor taxes.
- We can borrow and lend at the same rate of interest
- Short sales of all assets are allowed
- The MM proposition I is holding

Merton's Main Assumptions

- No transaction costs nor taxes.
- We can borrow and lend at the same rate of interest
- Short sales of all assets are allowed
- The MM proposition I is holding
- The term structure is flat and known with certainty which means that at time t a $1 nominal bond value of maturity T equals:

$$P(t, T) = e^{-r(T-t)} \tag{1}$$

Where r is the risk free rate.

Merton's Main Assumptions

- No transaction costs nor taxes.
- We can borrow and lend at the same rate of interest
- Short sales of all assets are allowed
- The MM proposition I is holding
- The term structure is flat and known with certainty which means that at time t a \$1 nominal bond value of maturity T equals:

$$P(t, T) = e^{-r(T-t)} \tag{1}$$

  Where r is the risk free rate.
- We can describe the value of the firm, V, with a diffusion type stochastic process

Merton's Main Assumptions

- No transaction costs nor taxes.
- We can borrow and lend at the same rate of interest
- Short sales of all assets are allowed
- The MM proposition I is holding
- The term structure is flat and known with certainty which means that at time t a $1 nominal bond value of maturity T equals:

$$P(t, T) = e^{-r(T-t)} \tag{1}$$

  Where r is the risk free rate.
- We can describe the value of the firm, V, with a diffusion type stochastic process
- Value of the assets follows a Geometric Brownian Motion: Merton derived the value of three assets among which the zero-coupon (but also the coupon-bearing and callable bonds)

$$dV_t = V_t([r - \delta]dt + \sigma dW_t) \tag{2}$$

Merton's Model Breakthroughs

1. The company issue a zero-coupon bond with face value B and maturity T.

Merton's Model Breakthroughs

1. The company issue a zero-coupon bond with face value B and maturity T.
2. Default occurs when the value V_T of the asset is below the level of the debt B.

Merton's Model Breakthroughs

1. The company issue a zero-coupon bond with face value B and maturity T.
2. Default occurs when the value V_T of the asset is below the level of the debt B.
3. Default may occur only at date T in which case the creditors take over the firm without incurring any distress costs and realize the amount V_T, so the payoff to the creditor at time T is:

$$D(V_T, T) = min(V_T, B) = B - max(B - V_T, 0)$$

Merton's Model Breakthroughs

1. The company issue a zero-coupon bond with face value B and maturity T.
2. Default occurs when the value V_T of the asset is below the level of the debt B.
3. Default may occur only at date T in which case the creditors take over the firm without incurring any distress costs and realize the amount V_T, so the payoff to the creditor at time T is:

$$D(V_T, T) = min(V_T, B) = B - max(B - V_T, 0)$$

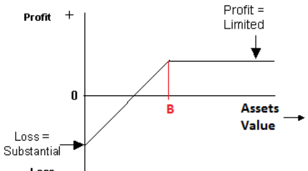
$max(B - V_T, 0)$ is nothing else than a Short Put option on the Assets of the company, with strike B

Merton's Model Breakthroughs

1. The company issue a zero-coupon bond with face value B and maturity T.
2. Default occurs when the value V_T of the asset is below the level of the debt B.
3. Default may occur only at date T in which case the creditors take over the firm without incurring any distress costs and realize the amount V_T, so the payoff to the creditor at time T is:

$$D(V_T, T) = min(V_T, B) = B - max(B - V_T, 0)$$

$max(B - V_T, 0)$ is nothing else than a Short Put option on the Assets of the company, with strike B



Merton's Model Breakthroughs (cont'd)

- On the contrary, the shareholder:

Merton's Model Breakthroughs (cont'd)

- On the contrary, the shareholder:
  1. Owns the company V_t

Merton's Model Breakthroughs (cont'd)

- On the contrary, the shareholder:
  1. Owns the company V_t
  2. Issued (sold) the bond B

Merton’s Model Breakthroughs (cont’d)

- On the contrary, the shareholder:
  1. Owns the company V_t
  2. Issued (sold) the bond B
  3. Owns the put option on the assets with strike B??

Merton's Model Breakthroughs (cont'd)

- On the contrary, the shareholder:
  1. Owns the company V_t
  2. Issued (sold) the bond B
  3. Owns the put option on the assets with strike B??
- $\Longrightarrow$ Put-call parity tells us that the shareholder holds a call:

Merton's Model Breakthroughs (cont'd)

- On the contrary, the shareholder:
  1. Owns the company V_t
  2. Issued (sold) the bond B
  3. Owns the put option on the assets with strike B??
- $\Longrightarrow$ Put-call parity tells us that the shareholder holds a call:

$$E(V_t, T) = V_T - B + max(B - V_T, 0) = max(V_T - B, 0) \quad (3)$$

Merton's Model Breakthroughs (cont'd)

- On the contrary, the shareholder:
  1. Owns the company V_t
  2. Issued (sold) the bond B
  3. Owns the put option on the assets with strike B??
- $\Longrightarrow$ Put-call parity tells us that the shareholder holds a call:

$$E(V_t, T) = V_T - B + max(B - V_T, 0) = max(V_T - B, 0) \tag{3}$$

So at time t the value of the bond and the stock are:

Merton's Model Breakthroughs (cont'd)

- On the contrary, the shareholder:
  1. Owns the company V_t
  2. Issued (sold) the bond B
  3. Owns the put option on the assets with strike B??
- $\Longrightarrow$ Put-call parity tells us that the shareholder holds a call:

$$E(V_t, T) = V_T - B + max(B - V_T, 0) = max(V_T - B, 0) \tag{3}$$

So at time t the value of the bond and the stock are:

$$\begin{aligned} D(V_t, T) &= B - Put(V_t, B, r, R - t, \sigma) \\ E(V_t, t) &= Call(V_t, B, r, T - t, \sigma) \end{aligned} \tag{4}$$

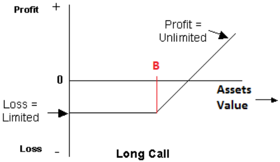
Merton's Model Breakthroughs (cont'd)

- On the contrary, the shareholder:
  1. Owns the company V_t
  2. Issued (sold) the bond B
  3. Owns the put option on the assets with strike B??
- $\Longrightarrow$ Put-call parity tells us that the shareholder holds a call:

$$E(V_t, T) = V_T - B + max(B - V_T, 0) = max(V_T - B, 0) \qquad (3)$$

So at time t the value of the bond and the stock are:

$$\begin{aligned} D(V_t, T) &= B - Put(V_t, B, r, R - t, \sigma) \\ E(V_t, t) &= Call(V_t, B, r, T - t, \sigma) \end{aligned} \qquad (4)$$



Whys and Wherefores of the Merton's Model

- As far as we know the diffusion process associated to the assets we can obtain the value of the debt and the equity of a given company through the Black, Scholes and Merton (1973) formula:

$$\begin{aligned} C\left[V_t, B, r, T-t, \sigma\right] &= C_t = N(d_1)V_t - N(d_2)Be^{-r(T-t)} \\ P\left[V_t, B, r, T-t, \sigma\right] &= P_t = C_t + Be^{-r(T-t)-V_t} \end{aligned} \tag{5}$$

where:

$$\begin{aligned} d_1 &= \frac{ln\left(\frac{V_t}{B}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \\ d_2 &= d_1 - \sigma\sqrt{T-t} \end{aligned} \tag{6}$$

Whys and Wherefores of the Merton's Model (cont'd)

- We then get the probability of default PD_i for any company i as far as we manage to model the asset dynamic A_t.

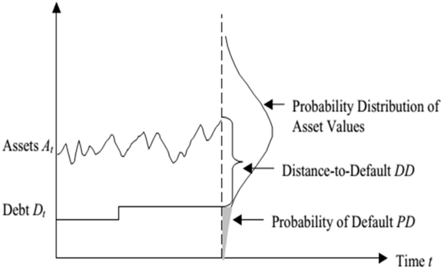


Figure: Graph representation of the Merton's Theory.

Whys and Wherefores of the Merton's Model (cont'd)

- We then get the probability of default PD_i for any company i as far as we manage to model the asset dynamic A_t.

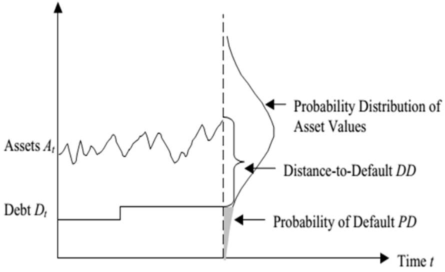


Figure: Graph representation of the Merton's Theory.

- Contribution of our work: How can we model farms assets dynamic using price of the commodities and weather conditions?

Assumptions

- For each farmer we know the breakdown of land use per type of crop

Assumptions

- For each farmer we know the breakdown of land use per type of crop
- The asset value of a farm is a cumulative function of the farm profits (under the retained earnings mechanism)

Assumptions

- For each farmer we know the breakdown of land use per type of crop
- The asset value of a farm is a cumulative function of the farm profits (under the retained earnings mechanism)
- Conditional on the global filtration, we define an adapted process for the asset value of farm i at time t:

$$A_t^i|\mathcal{F}_{t-1} = [A_{t-1}^i + R_t^i + \triangle E_t^i + \triangle D_t^i, |\mathcal{F}_{t-1}] \tag{7}$$

If we assume that the farmer will not issue debt or equity from one year to another $\triangle E_t^i = 0$ and $\triangle D_t^i = 0$:

$$A_t^i|\mathcal{F}_{t-1} = \left[A_{t-1}^i\left(1 + \frac{R_t^i}{A_{t-1}^i}\right)\middle|\mathcal{F}_{t-1}\right], \tag{8}$$

where:

$$\left[\frac{R_t^i}{A_{t-1}^i}\middle|\mathcal{F}_{t-1}\right] = \sum_{k=1}^{K} \triangle_{k,t}^i [y_{k,t}^i(\tilde{\omega}_{\tau_t^k}^i) \cdot \tilde{C}_{k,t}|\mathcal{F}_{t-1}] - F_t^i, \tag{9}$$

Assumptions

- The vector $\omega_t = \{\mathcal{T}_t, \mathcal{P}_t\} \in \mathbb{R}^S \times \mathbb{R}^S$ represents the information about weather conditions over time. Where $\mathcal{T}_t$ and $\mathcal{P}_t$ stand respectively for the temperature and the log-precipitation random variables associated to a set of S meteorological stations non-equally spread over a given territory.

Assumptions

- The vector $\omega_t = \{\mathcal{T}_t, \mathcal{P}_t\} \in \mathbb{R}^S \times \mathbb{R}^S$ represents the information about weather conditions over time. Where $\mathcal{T}_t$ and $\mathcal{P}_t$ stand respectively for the temperature and the log-precipitation random variables associated to a set of S meteorological stations non-equally spread over a given territory.
- The filtration generated by the weather conditions $\{\omega_t\}_{t\geq 0}$ is denoted $\mathcal{H}_t$ while $\mathcal{Y}_t$ represents the filtration generated by the crop yields and we finally denoted the commodity prices vector $\{C_t\}_{t\geq 0} \in \mathbb{R}^K$ and its associated filtration $\mathcal{C}_t$ such that $\mathcal{F}_t = \mathcal{H}_t \vee \mathcal{C}_t \vee \mathcal{Y}_t$.

Assumptions

- The vector $\omega_t = \{\mathcal{T}_t, \mathcal{P}_t\} \in \mathbb{R}^S \times \mathbb{R}^S$ represents the information about weather conditions over time. Where $\mathcal{T}_t$ and $\mathcal{P}_t$ stand respectively for the temperature and the log-precipitation random variables associated to a set of S meteorological stations non-equally spread over a given territory.
- The filtration generated by the weather conditions $\{\omega_t\}_{t\geq 0}$ is denoted $\mathcal{H}_t$ while $\mathcal{Y}_t$ represents the filtration generated by the crop yields and we finally denoted the commodity prices vector $\{C_t\}_{t\geq 0} \in \mathbb{R}^K$ and its associated filtration $\mathcal{C}_t$ such that $\mathcal{F}_t = \mathcal{H}_t \vee \mathcal{C}_t \vee \mathcal{Y}_t$.
- the $\mathcal{F}_t$-measurable random variable $R_{i,t}$ embodies the retained earning generated over the ending year by the farmer i and is function both of the weather conditions ω_t through the K crop yields generated by the farmer i at time t and C_t, the agricultural commodity price at which he sold his harvested or not yet harvested crops.

Farms Profits Dynamic

The conditional dynamic retained earnings process for farm *i* at time *t* is written as:

$$\left[\frac{R_t^i}{A_{t-1}^i}\middle|\mathcal{F}_{t-1}\right] = \sum_{k=1}^{K} \triangle_{k,t}^i [y_{k,t}^i(\tilde{\omega}_{\tau_t^k}^i)\cdot \tilde{C}_{k,t}|\mathcal{F}_{t-1}] - F_t^i, \tag{10}$$

Farms Profits Dynamic

The conditional dynamic retained earnings process for farm i at time t is written as:

$$\left[\frac{R^i_t}{A^i_{t-1}}\middle|\mathcal{F}_{t-1}\right] = \sum_{k=1}^{K} \triangle^i_{k,t}[y^i_{k,t}(\tilde{\omega}^i_{\tau^k_t})\cdot \tilde{C}_{k,t}|\mathcal{F}_{t-1}] - F^i_t, \quad (10)$$

- $\triangle^{(i)}_{k,t} = \frac{\delta^i_{k,t}}{A^i_{t-1}}$ where $\delta^i_{k,t}$ the hectares allocated by the farmer i to the crop k

Farms Profits Dynamic

The conditional dynamic retained earnings process for farm i at time t is written as:

$$\left[\frac{R_t^i}{A_{t-1}^i}\middle|\mathcal{F}_{t-1}\right] = \sum_{k=1}^{K} \triangle_{k,t}^i [y_{k,t}^i(\tilde{\omega}_{\tau_t^k}^i)\cdot \tilde{C}_{k,t}|\mathcal{F}_{t-1}] - F_t^i, \qquad (10)$$

- $\triangle_{k,t}^{(i)} = \frac{\delta_{k,t}^i}{A_{t-1}^i}$ where $\delta_{k,t}^i$ the hectares allocated by the farmer i to the crop k
- $y_{k,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right)$ denotes the yield per hectares for a given crops and under given weather condition $\tilde{\omega}_t$ for the period of time τ_t^k

Farms Profits Dynamic

The conditional dynamic retained earnings process for farm i at time t is written as:

$$\left[\frac{R_t^i}{A_{t-1}^i}\middle|\mathcal{F}_{t-1}\right] = \sum_{k=1}^{K} \triangle_{k,t}^i[y_{k,t}^i(\tilde{\omega}_{\tau_t^k}^i)\cdot \tilde{C}_{k,t}|\mathcal{F}_{t-1}] - F_t^i, \qquad (10)$$

- $\triangle_{k,t}^{(i)} = \frac{\delta_{k,t}^i}{A_{t-1}^i}$ where $\delta_{k,t}^i$ the hectares allocated by the farmer i to the crop k
- $y_{k,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right)$ denotes the yield per hectares for a given crops and under given weather condition $\tilde{\omega}_t$ for the period of time τ_t^k
- $\tilde{C}_{k,t} = (\tilde{c}_{k,t} - \upsilon_{k,t}^i)$ represents the random price of a specific commodity k on the market at a given time t

Farms Profits Dynamic

The conditional dynamic retained earnings process for farm i at time t is written as:

$$\left[\frac{R_t^i}{A_{t-1}^i}\middle|\mathcal{F}_{t-1}\right] = \sum_{k=1}^{K} \triangle_{k,t}^i[y_{k,t}^i(\tilde{\omega}_{\tau_t^k}^i)\cdot \tilde{C}_{k,t}|\mathcal{F}_{t-1}] - F_t^i, \qquad (10)$$

- $\triangle_{k,t}^{(i)} = \frac{\delta_{k,t}^i}{A_{t-1}^i}$ where $\delta_{k,t}^i$ the hectares allocated by the farmer i to the crop k
- $y_{k,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right)$ denotes the yield per hectares for a given crops and under given weather condition $\tilde{\omega}_t$ for the period of time τ_t^k
- $\tilde{C}_{k,t} = (\tilde{c}_{k,t} - \upsilon_{k,t}^i)$ represents the random price of a specific commodity k on the market at a given time t
- $\upsilon_{k,t}^i$ represent the variable cost associated to the crop k

Farms Profits Dynamic

The conditional dynamic retained earnings process for farm i at time t is written as:

$$\left[\frac{R_t^i}{A_{t-1}^i}\middle|\mathcal{F}_{t-1}\right] = \sum_{k=1}^{K} \triangle_{k,t}^i [y_{k,t}^i(\tilde{\omega}_{\tau_t^k}^i)\cdot \tilde{C}_{k,t}|\mathcal{F}_{t-1}] - F_t^i, \qquad (10)$$

- $\triangle_{k,t}^{(i)} = \frac{\delta_{k,t}^i}{A_{t-1}^i}$ where $\delta_{k,t}^i$ the hectares allocated by the farmer i to the crop k
- $y_{k,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right)$ denotes the yield per hectares for a given crops and under given weather condition $\tilde{\omega}_t$ for the period of time τ_t^k
- $\tilde{C}_{k,t} = (\tilde{c}_{k,t} - \upsilon_{k,t}^i)$ represents the random price of a specific commodity k on the market at a given time t
- $\upsilon_{k,t}^i$ represent the variable cost associated to the crop k
- $F_t^i = \frac{f_t^i}{A_{t-1}^i}$ for the fixed costs independent from the type of crop.

Farms Profits Dynamic

We can distinguish two sources of uncertainty:

Farms Profits Dynamic

We can distinguish two sources of uncertainty:

- A local risk related to weather conditions:
  - due to the relation between weather conditions and crops yield
  - bad weather conditions in a specific region doesn't necessarily impact other region or other countries.
  - leads to a local dependence among the farmers.

Farms Profits Dynamic

We can distinguish two sources of uncertainty:

- A local risk related to weather conditions:
  - due to the relation between weather conditions and crops yield
  - bad weather conditions in a specific region doesn't necessarily impact other region or other countries.
  - leads to a local dependence among the farmers.
- A global risk related to market prices of the agricultural commodities:
  - due to the relation between these prices and the profits generated by the farmers
  - commodity markets globalisation and transportation networks development linked the local prices to international market prices
  - generates a global dependence: a large price decrease of a given commodity may impact both the Romanian and the American farmers

Outline

Commodities Conditional Loss Distribution

- We first analyze the conditional loss distribution given the filtration of the weather conditions $\mathcal{H}_t$ and the yield $\mathcal{Y}_t$:
    - We assume as known and unchanged the yields associated to each farm

Commodities Conditional Loss Distribution

- We first analyze the conditional loss distribution given the filtration of the weather conditions $\mathcal{H}_t$ and the yield $\mathcal{Y}_t$:
  - We assume as known and unchanged the yields associated to each farm
  - We only consider uncertainty about commodity prices

Commodities Conditional Loss Distribution

- We first analyze the conditional loss distribution given the filtration of the weather conditions $\mathcal{H}_t$ and the yield $\mathcal{Y}_t$:
  - We assume as known and unchanged the yields associated to each farm
  - We only consider uncertainty about commodity prices
  - We assume the following dynamic for the commodities market prices.

$$d\tilde{\boldsymbol{C}}_t = \tilde{\mu}_t dt + \tilde{\Omega}_t d\boldsymbol{W}_t$$

where $d\boldsymbol{W}_t$ is the vector of dW_t^k associated to the K $\mathcal{F}_t$-standard Brownian motion $\{W_t^k\}_{t\geq 0}$, for $k = 1 \ldots K$. The matrix of variance covariance at time t is then equal to $\tilde{\Omega}_t^\top \tilde{\Omega}_t$ and

$$d\tilde{C}_t \sim N(\tilde{\boldsymbol{\mu}}_{\boldsymbol{t}} dt\,, \tilde{\boldsymbol{\Omega}}_{\boldsymbol{t}}^\top \tilde{\boldsymbol{\Omega}}_{\boldsymbol{t}}\, dt)$$

Commodities Conditional Loss Distribution

- Assuming that the yield vector $\mathcal{Y}_t = y^i_{k,t}(\omega_t)\ \forall i, k$ is known, that $\tilde{C}_{k,t}$ is independent from the local weather conditions, we can then rewrite the previous equation given

$$\mathbb{E}\left(\frac{R^i_t}{A^i_{t-1}} | \mathcal{H}_t, \mathcal{Y}_t\right) = \mathbf{H_{i,t}}\left(\omega_\mathbf{t}\right)\left(\tilde{\boldsymbol{C}}_{t-1} + \tilde{\boldsymbol{\mu}}_{\boldsymbol{t}}\Delta t\right) - F^i_t$$

with:

$$\mathbf{H_{i,t}}\left(\omega_\mathbf{t}\right) = \begin{bmatrix} \Delta^i_{1,t} y^i_{1,t}\left(\omega_t\right) \\ \vdots \\ \Delta^i_{K,t} y^i_{K,t}\left(\omega_t\right) \end{bmatrix}$$

and:

$$\tilde{C}_{t-1} = \left(\tilde{C}_{1,t-1}, \ldots, \tilde{C}_{K,t-1}\right)^\top$$

Commodities Conditional Loss Distribution

- Assuming that the yield vector $\mathcal{Y}_t = y^i_{k,t}(\omega_t)\ \forall i, k$ is known, that $\tilde{C}_{k,t}$ is independent from the local weather conditions, we can then rewrite the previous equation given

$$\mathbb{E}\left(\frac{R^i_t}{A^i_{t-1}}|\mathcal{H}_t, \mathcal{Y}_t\right) = \mathbf{H_{i,t}}\left(\omega_\mathbf{t}\right)\left(\tilde{\boldsymbol{C}}_{t-1} + \tilde{\boldsymbol{\mu}}_\mathbf{t}\Delta t\right) - F^i_t$$

with:

$$\mathbf{H_{i,t}}\left(\omega_\mathbf{t}\right) = \begin{bmatrix} \Delta^i_{1,t} y^i_{1,t}\left(\omega_t\right) \\ \vdots \\ \Delta^i_{K,t} y^i_{K,t}\left(\omega_t\right) \end{bmatrix}$$

and:

$$\tilde{C}_{t-1} = \left(\tilde{C}_{1,t-1}, \ldots, \tilde{C}_{K,t-1}\right)^\top$$

- We can also express the profits conditional variance as follows:

$$\sigma^2_{i,t}|\mathcal{H}_t, \mathcal{Y}_t = \mathbb{V}\left(\frac{R_{i,t}}{A^i_{t-1}}|\mathcal{H}_t, \mathcal{Y}_t\right) = \mathbf{H_{i,t}}\left(\omega_\mathbf{t}\right)\tilde{\boldsymbol{\Omega}}^\top_t\tilde{\boldsymbol{\Omega}}_t\mathbf{H_{i,t}}\left(\omega_\mathbf{t}\right)^\top$$

- The risk of default of the farmer i is then express as follows:

$$\begin{aligned} PD_i|\mathcal{H}_t,\mathcal{Y}_t &= Pr\left[A_{i,t} \leq D_{i,t}|\mathcal{H}_t,\mathcal{Y}_t\right] \\ &= Pr\left[\frac{R_{i,t}}{A^i_{t-1}} \leq \frac{D_{i,t}}{A^i_{t-1}} - 1|\mathcal{H}_t,\mathcal{Y}_t\right] \\ &= \Phi\left[\frac{\left(\frac{D_{i,t}}{A^i_{t-1}}-1\right)-\mathbf{H_{i,t}}(\omega_\mathbf{t})\left(\tilde{\boldsymbol{C}}_{t-1}+\tilde{\boldsymbol{\mu}}_{\boldsymbol{t}}\right)+F^i_t}{\sqrt{\mathbf{H_{i,t}}(\omega_\mathbf{t})\tilde{\boldsymbol{\Omega}}_t^\top\tilde{\boldsymbol{\Omega}}_t\mathbf{H_{i,t}}(\omega_\mathbf{t})^\top}}|\mathcal{H}_t,\mathcal{Y}_t\right] \end{aligned}$$

- The risk of default of the farmer i is then express as follows:

$$\begin{aligned} PD_i|\mathcal{H}_t, \mathcal{Y}_t &= Pr\left[A_{i,t} \leq D_{i,t}|\mathcal{H}_t, \mathcal{Y}_t\right] \\ &= Pr\left[\frac{R_{i,t}}{A^i_{t-1}} \leq \frac{D_{i,t}}{A^i_{t-1}} - 1|\mathcal{H}_t, \mathcal{Y}_t\right] \\ &= \Phi\left[\frac{\left(\frac{D_{i,t}}{A^i_{t-1}} - 1\right) - \mathbf{H_{i,t}}(\omega_\mathbf{t})\left(\tilde{\boldsymbol{C}}_{t-1} + \tilde{\boldsymbol{\mu}}_t\right) + F^i_t}{\sqrt{\mathbf{H_{i,t}}(\omega_\mathbf{t})\tilde{\boldsymbol{\Omega}}_t^\top \tilde{\boldsymbol{\Omega}}_t \mathbf{H_{i,t}}(\omega_\mathbf{t})^\top}}|\mathcal{H}_t, \mathcal{Y}_t\right] \end{aligned}$$

- While the default correlation between the farmer i and farmer j can be computed under the assumption of Gaussian joint distribution such as:

$$\rho_{ij}|\mathcal{H}_t, \mathcal{Y}_t = \frac{Pr\left(A_{i,t} \leq D_{i,t}, A_{j,t} \leq D_{j,t}|\mathcal{H}_t, \mathcal{Y}_t\right) - PD_i|\mathcal{H}_t, \mathcal{Y}_t PD_j|\mathcal{H}_t, \mathcal{Y}_t}{\sqrt{PD_i|\mathcal{H}_t, \mathcal{Y}_t\left(1 - PD_i|\mathcal{H}_t, \mathcal{Y}_t\right) PD_j|\mathcal{H}_t, \mathcal{Y}_t\left(1 - PD_j|\mathcal{H}_t, \mathcal{Y}_t\right)}}$$

where:

$$\begin{aligned} & Pr\left(\frac{R_{i,t}}{A^i_{t-1}} \leq \frac{D_{i,t}}{A^i_{t-1}} - 1, \frac{R_{j,t}}{A^j_{t-1}} \leq \frac{D_{j,t}}{A^j_{t-1}} - 1|\mathcal{H}_t, \mathcal{Y}_t\right) \\ &= \int_0^{\frac{D_{i,t}}{A^i_{t-1}} - 1} \int_0^{\frac{D_{j,t}}{A^j_{t-1}} - 1} MVN\left(\frac{R_{i,t}}{A^i_{t-1}}, \frac{R_{j,t}}{A^j_{t-1}}, \theta_{ij}|\mathcal{H}_t, \mathcal{Y}_t\right) dR_{j,t} dR_{i,t} \end{aligned}$$

with:

$$\theta_{ij}|\mathcal{H}_t, \mathcal{Y}_t = \frac{cov(R_i, R_j|\mathcal{H}_t, \mathcal{Y}_t)}{\sqrt{\mathbb{V}\left(R_{i,t}|\mathcal{H}_t, \mathcal{Y}_t\right)\mathbb{V}\left(R_{j,t}|\mathcal{H}_t, \mathcal{Y}_t\right)}}$$

while:

$$cov(R_i, R_j|\mathcal{H}_t, \mathcal{Y}_t) = \mathbf{H_{i,t}}\left(\omega_\mathbf{t}\right)\tilde{\boldsymbol{\Omega}}_t^\top \tilde{\boldsymbol{\Omega}}_t \mathbf{H_{j,t}}\left(\omega_\mathbf{t}\right)^\top$$

and:

$$MVN\left(R_{i,t}, R_{j,t}, \theta_{ij}|\mathcal{H}_t, \mathcal{Y}_t\right) = \frac{1}{2\pi\sigma_{i,t}\sigma_{j,t}\sqrt{1-(\theta_{ij}|\mathcal{H}_t,\mathcal{Y}_t)^2}}$$
$$\times exp\left\{\frac{-1}{2\left(1-(\theta_{ij}|\mathcal{H}_t,\mathcal{Y}_t)^2\right)}\left(\frac{\left(R_{i,t}-\mathbf{H_{i,t}}(\omega_\mathbf{t})\left(\tilde{\boldsymbol{c}}_{t-1}+\tilde{\boldsymbol{\mu}}_t\right)+F_t^i\right)^2}{\mathbb{V}\left(R_{i,t}|\mathcal{H}_t,\mathcal{Y}_t\right)} + \frac{\left(R_{j,t}-\mathbf{H_{j,t}}(\omega_\mathbf{t})\left(\tilde{\boldsymbol{c}}_{t-1}+\tilde{\boldsymbol{\mu}}_t\right)+F_t^j\right)^2}{\mathbb{V}\left(R_{j,t}|\mathcal{H}_t,\mathcal{Y}_t\right)}\right)\right\}$$
$$\times exp\left\{\frac{-1}{2\left(1-(\theta_{ij}|\mathcal{H}_t,\mathcal{Y}_t)^2\right)}\left(-\frac{2\theta_{ij}|\mathcal{H}_t,\mathcal{Y}_t\left(R_{i,t}-\mathbf{H_{i,t}}(\omega_\mathbf{t})\left(\tilde{\boldsymbol{c}}_{t-1}+\tilde{\boldsymbol{\mu}}_t\right)+F_t^i\right)\left(R_{j,t}-\mathbf{H_{j,t}}(\omega_\mathbf{t})\left(\tilde{\boldsymbol{c}}_{t-1}+\tilde{\boldsymbol{\mu}}_t\right)+F_t^j\right)}{\sqrt{\mathbb{V}\left(R_{i,t}|\mathcal{H}_t,\mathcal{Y}_t\right)\mathbb{V}\left(R_{j,t}|\mathcal{H}_t,\mathcal{Y}_t\right)}}\right)\right\}$$

- We can then compute respectively the portfolio loss function L of the farms' creditor, the expected loss EL and the unexpected loss UL which respectively represents the expected value and the variance of the loss function associated to this portfolio of loans.

$$L|\mathcal{H}_t, \mathcal{Y}_t = \sum_{n=1}^{N} EAD_n LGD_n D_n|\mathcal{H}_t, \mathcal{Y}_t$$

where $D_n|\mathcal{H}_t, \mathcal{Y}_t \sim Bernoulli\,(PD_n|\mathcal{H}_t, \mathcal{Y}_t)$. In order to simplify the forthcoming notations we use $PD_n^\star = PD_n|\mathcal{H}_t, \mathcal{Y}_t$

- We can then compute respectively the portfolio loss function L of the farms' creditor, the expected loss EL and the unexpected loss UL which respectively represents the expected value and the variance of the loss function associated to this portfolio of loans.

$$L|\mathcal{H}_t, \mathcal{Y}_t = \sum_{n=1}^{N} EAD_n LGD_n D_n|\mathcal{H}_t, \mathcal{Y}_t$$

where $D_n|\mathcal{H}_t, \mathcal{Y}_t \sim Bernoulli\,(PD_n|\mathcal{H}_t, \mathcal{Y}_t)$. In order to simplify the forthcoming notations we use $PD_n^\star = PD_n|\mathcal{H}_t, \mathcal{Y}_t$

$$EL|\mathcal{H}_t, \mathcal{Y}_t = \sum_{n=1}^{N} EAD_n ELGD_n PD_n^\star$$

- We can then compute respectively the portfolio loss function L of the farms' creditor, the expected loss EL and the unexpected loss UL which respectively represents the expected value and the variance of the loss function associated to this portfolio of loans.

$$L|\mathcal{H}_t, \mathcal{Y}_t = \sum_{n=1}^{N} EAD_n LGD_n D_n|\mathcal{H}_t, \mathcal{Y}_t$$

where $D_n|\mathcal{H}_t, \mathcal{Y}_t \sim Bernoulli\,(PD_n|\mathcal{H}_t, \mathcal{Y}_t)$. In order to simplify the forthcoming notations we use $PD_n^\star = PD_n|\mathcal{H}_t, \mathcal{Y}_t$

$$EL|\mathcal{H}_t, \mathcal{Y}_t = \sum_{n=1}^{N} EAD_n ELGD_n PD_n^\star$$

$$UL|\mathcal{H}_t, \mathcal{Y}_t = \sqrt{V\left(L|\mathcal{H}_t, \mathcal{Y}_t\right)}$$
$$= \sqrt{\sum_{n,k=1}^{N} EAD_n EAD_k ELGD_n ELGD_k \rho_{nk} \sqrt{PD_n^\star\left(1 - PD_n^\star\right) PD_n^\star\left(1 - PD_n^\star\right)}}$$

Outline

Weather Conditional Loss Distribution

Roadmap:

- To take into consideration the impact of the weather condition on the profit of the farmers and thus on their probability of default we consider the yield associated to each farm as a function of local weather conditions.

Weather Conditional Loss Distribution

Roadmap:

- To take into consideration the impact of the weather condition on the profit of the farmers and thus on their probability of default we consider the yield associated to each farm as a function of local weather conditions.
- We model the yield of each farm as a linear function of non linear estimator of temperatures and precipitations associated to this region.

Weather Conditional Loss Distribution

Roadmap:

- To take into consideration the impact of the weather condition on the profit of the farmers and thus on their probability of default we consider the yield associated to each farm as a function of local weather conditions.
- We model the yield of each farm as a linear function of non linear estimator of temperatures and precipitations associated to this region.
- In order to obtain the temperature and precipitation estimators for all the farms according to their respective longitude and latitude we consider a gaussian process model with as design points for the input space a set of weather stations records

Crop yield distribution modelling for calculating an insurance premium.

- Parametric: Nelson and Preckel, 1989 applied a conditional beta distribution to study corn yields modelling notwithstanding but difficult to obtain the standard errors of moment elasticities.

Crop yield distribution modelling for calculating an insurance premium.

- Parametric: Nelson and Preckel, 1989 applied a conditional beta distribution to study corn yields modelling notwithstanding but difficult to obtain the standard errors of moment elasticities.
- Non-parametric: Goodwin and Ker 1998 proposed to consider non-parametric density estimation techniques to work out county-level wheat and barley area-yield distribution estimation.

Crop yield distribution modelling for calculating an insurance premium.

- Parametric: Nelson and Preckel, 1989 applied a conditional beta distribution to study corn yields modelling notwithstanding but difficult to obtain the standard errors of moment elasticities.
- Non-parametric: Goodwin and Ker 1998 proposed to consider non-parametric density estimation techniques to work out county-level wheat and barley area-yield distribution estimation.
- non-parametric methods offer an appealing flexibility since they heavily rely on the data sample to determine the most appropriate density representation avoiding thus the restraining choice of a specific parametric probability distribution, their rate of convergence to the true distribution might be relatively slow and consequently makes those methods data-intensive. (Sherrick et al., 2014)

Crop yield distribution modelling for calculating an insurance premium.

- Parametric: Nelson and Preckel, 1989 applied a conditional beta distribution to study corn yields modelling notwithstanding but difficult to obtain the standard errors of moment elasticities.
- Non-parametric: Goodwin and Ker 1998 proposed to consider non-parametric density estimation techniques to work out county-level wheat and barley area-yield distribution estimation.
- non-parametric methods offer an appealing flexibility since they heavily rely on the data sample to determine the most appropriate density representation avoiding thus the restraining choice of a specific parametric probability distribution, their rate of convergence to the true distribution might be relatively slow and consequently makes those methods data-intensive. (Sherrick et al., 2014)
- the choice of the crop yields distribution or the non-parametric estimation approach leads to out-of-sample performances and large differences in expected payouts (Sherrick et al,2014,Woodard and Sherrick, 2011,Sherrick et al., 2004).

Crop yield distribution modelling for calculating an insurance premium.

- Parametric: Nelson and Preckel, 1989 applied a conditional beta distribution to study corn yields modelling notwithstanding but difficult to obtain the standard errors of moment elasticities.
- Non-parametric: Goodwin and Ker 1998 proposed to consider non-parametric density estimation techniques to work out county-level wheat and barley area-yield distribution estimation.
- non-parametric methods offer an appealing flexibility since they heavily rely on the data sample to determine the most appropriate density representation avoiding thus the restraining choice of a specific parametric probability distribution, their rate of convergence to the true distribution might be relatively slow and consequently makes those methods data-intensive. (Sherrick et al., 2014)
- the choice of the crop yields distribution or the non-parametric estimation approach leads to out-of-sample performances and large differences in expected payouts (Sherrick et al,2014,Woodard and Sherrick, 2011,Sherrick et al., 2004).
- More recently, a couple of articles investigate instead the yield distribution at the farm level in order to get a better grasp on the aggregation process to the county level (Gerlt et al. 2014, Claassen and Just 2011).

Time and Space Decomposition

- Multivariate weather time series associated to a set of weather stations geolocalisations not only encompasses spatial dependance structure but also temporal relationships

- Multivariate weather time series associated to a set of weather stations geolocalisations not only encompasses spatial dependance structure but also temporal relationships
- The seasonality in daily temperatures and precipitations constitutes for instance one of the main sources of data auto-correlation

- Multivariate weather time series associated to a set of weather stations geolocalisations not only encompasses spatial dependance structure but also temporal relationships
- The seasonality in daily temperatures and precipitations constitutes for instance one of the main sources of data auto-correlation
- The trend is modelled through a standardized time series model, namely the Seasonal Autoregressive Integrated Moving Average (SARIMA) (Sumer et al., 2009, Ediger et al., 2006),(Brandao and Nova, 2012, (Mills, 2014)

- Multivariate weather time series associated to a set of weather stations geolocalisations not only encompasses spatial dependance structure but also temporal relationships
- The seasonality in daily temperatures and precipitations constitutes for instance one of the main sources of data auto-correlation
- The trend is modelled through a standardized time series model, namely the Seasonal Autoregressive Integrated Moving Average (SARIMA) (Sumer et al., 2009, Ediger et al., 2006),(Brandao and Nova, 2012, (Mills, 2014)
- Each weather station observed temperatures $\mathcal{T}_t$ and observed precipitations $\mathcal{P}_t$ cross-sectional data are expressed as a combination of:

- Multivariate weather time series associated to a set of weather stations geolocalisations not only encompasses spatial dependance structure but also temporal relationships
- The seasonality in daily temperatures and precipitations constitutes for instance one of the main sources of data auto-correlation
- The trend is modelled through a standardized time series model, namely the Seasonal Autoregressive Integrated Moving Average (SARIMA) (Sumer et al., 2009, Ediger et al., 2006),(Brandao and Nova, 2012, (Mills, 2014)
- Each weather station observed temperatures $\mathcal{T}_t$ and observed precipitations $\mathcal{P}_t$ cross-sectional data are expressed as a combination of:
  - a national (or a county) global seasonality-adapted trend $\bar{w}_t$ evenly impacting all the country regions and model through a SARIMA model

- Multivariate weather time series associated to a set of weather stations geolocalisations not only encompasses spatial dependance structure but also temporal relationships
- The seasonality in daily temperatures and precipitations constitutes for instance one of the main sources of data auto-correlation
- The trend is modelled through a standardized time series model, namely the Seasonal Autoregressive Integrated Moving Average (SARIMA) (Sumer et al., 2009, Ediger et al., 2006),(Brandao and Nova, 2012, (Mills, 2014)
- Each weather station observed temperatures $\mathcal{T}_t$ and observed precipitations $\mathcal{P}_t$ cross-sectional data are expressed as a combination of:
  - a national (or a county) global seasonality-adapted trend $\bar{w}_t$ evenly impacting all the country regions and model through a SARIMA model
  - a spatial dependence structure which furnishes a local adjustment for each weather station through a spatial Gaussian Process $f^w(\mathbf{x})$

- $\mathbf{W}_t^{\mathcal{T}}(\mathbf{x})$ representing the observed precipitation $\mathcal{P}_t$ or temperature $\mathcal{T}_t$ are defined as follows:

$$\begin{aligned}\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x}) &= \mathbf{W}_t^{\mathcal{T}}(\mathbf{x}) - \mathbb{E}_t\left[\mathbb{E}_{\mathbf{x}}\left(\mathbf{W}_t^{\mathcal{T}}(\mathbf{x})\right)|\mathbf{w}_{t-1}^{\mathcal{T}}, \mathbf{w}_{t-2}^{\mathcal{T}}, \dots\right]\mathbf{1}_n \qquad (11)\\ &= f^{\mathcal{T}}(\mathbf{x}, t) + \epsilon_{t,\mathbf{x}}\end{aligned}$$

$$\text{with } \epsilon_{t,\mathbf{x}} \sim \mathcal{N}\left(0, \sigma_t^2 \mathbf{I}_n\right) \qquad (12)$$

- $\mathbf{W}_t^{\mathcal{T}}(\mathbf{x})$ representing the observed precipitation $\mathcal{P}_t$ or temperature $\mathcal{T}_t$ are defined as follows:

$$\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x}) = \mathbf{W}_t^{\mathcal{T}}(\mathbf{x}) - \mathbb{E}_t\left[\mathbb{E}_{\mathbf{x}}\left(\mathbf{W}_t^{\mathcal{T}}(\mathbf{x})\right)|\mathbf{w}_{t-1}^{\mathcal{T}}, \mathbf{w}_{t-2}^{\mathcal{T}}, \dots\right]\mathbf{1}_n \qquad (11)$$

$$= f^{\mathcal{T}}(\mathbf{x}, t) + \epsilon_{t,\mathbf{x}}$$

$$\text{with } \epsilon_{t,\mathbf{x}} \sim \mathcal{N}\left(0, \sigma_t^2 \mathbf{I}_n\right) \qquad (12)$$

- where the integrated and seasonally adjusted conditional mean operator $\mathbb{E}_t$ copes with the serial correlation observed in temperature and precipitation data at the level of the country through a SARIMA$(p, d, q)(P, D, Q)_s$ trend formulation:

$$\Phi_P(B^s)\cdot \phi(B)\cdot \nabla_s^D\cdot \nabla^d\cdot \mathrm{X}_t = c + \Theta_Q(B^s)\cdot \theta(B)\cdot \varepsilon_t. \qquad (13)$$

where the integer p, d and q is refereed respectively to the order of autoregression, of integration and the number of moving average lags. $B^k\mathrm{X}_t = \mathrm{X}_{t-k}$ represents the backshift operator. While:

$$\tilde{\varepsilon}_t \overset{\text{iid}}{\sim} N(0, \sigma_{\tilde{\varepsilon}}^2) \qquad (14)$$

- To impose weak stationarity to the discrete-time stochastic process required that $\phi(B) \neq 0$ and $\phi(B)$ has all roots outside unit disc. ∇_s^D and ∇^d denote respectively the seasonal difference and non-seasonal difference components.

- To impose weak stationarity to the discrete-time stochastic process required that $\phi(B) \neq 0$ and $\phi(B)$ has all roots outside unit disc. ∇_s^D and ∇^d denote respectively the seasonal difference and non-seasonal difference components.
- Furthermore, the spatial conditional expected value $\mathbb{E}_x$ which corresponds to the average value of the temperatures collected by all the weather stations on a given date t:

$$\mathbb{E}_x \left(\mathbf{W}_t^{\mathcal{T}} (\mathbf{x}) \right) = n^{-1} \sum_{i=1}^{n} w_t^{\mathcal{T}} (\mathbf{x_i})$$

- To impose weak stationarity to the discrete-time stochastic process required that $\phi(B) \neq 0$ and $\phi(B)$ has all roots outside unit disc. ∇_s^D and ∇^d denote respectively the seasonal difference and non-seasonal difference components.
- Furthermore, the spatial conditional expected value $\mathbb{E}_x$ which corresponds to the average value of the temperatures collected by all the weather stations on a given date t:

$$\mathbb{E}_x \left(\mathbf{W}_t^{\mathcal{T}}(\mathbf{x}) \right) = n^{-1} \sum_{i=1}^{n} w_t^{\mathcal{T}}(\mathbf{x_i})$$

- while the spatial Gaussian Process $f^w(\mathbf{x})$ is defined such as:

$$f^{\mathcal{T}}(\mathbf{x}, t) \sim \mathcal{GP}\left(0, k(\mathbf{x}, \mathbf{x}^*; t, t^*)\right)$$

with $\mathbf{w}_t^{\mathcal{T}} = \{w_t^{\mathcal{T}}(\mathbf{x_1}), w_t^{\mathcal{T}}(\mathbf{x_2}), \ldots, w_t^{\mathcal{T}}(\mathbf{x_n})\}$ denotes the vector of the temperature observed for the n weather stations with the associated n locations vectors written as $\mathbf{x} = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}\}$.

Spatial Kernel

- The Gaussian Process is fully specified by a conditional mean function $\mu(\mathbf{x})$ and a conditional covariance function which we consider time independent $k(\mathbf{x}, \mathbf{x}^*)$ such that:

$$\begin{aligned} f^{\mathcal{T}}(\mathbf{x}, t) &\sim \mathcal{GP}\left(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}^*)\right) \\ \mu(\mathbf{x}) &= \mathbb{E}\left[f(\mathbf{x})\right] \\ k(\mathbf{x}, \mathbf{x}^*) &= \mathbb{E}\left[\left(f(\mathbf{x}) - \mu(\mathbf{x})\right)\left(f(\mathbf{x}^*) - \mu(\mathbf{x}^*)\right)\right] \end{aligned}$$

where $\mathbf{x}$ and $\mathbf{x}^*$ represent two different location vectors.

Spatial Kernel

- The Gaussian Process is fully specified by a conditional mean function $\mu(\mathbf{x})$ and a conditional covariance function which we consider time independent $k(\mathbf{x}, \mathbf{x}^*)$ such that:

$$\begin{aligned} f^{\mathcal{T}}(\mathbf{x}, t) &\sim \mathcal{GP}\left(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}^*)\right) \\ \mu(\mathbf{x}) &= \mathbb{E}\left[f(\mathbf{x})\right] \\ k(\mathbf{x}, \mathbf{x}^*) &= \mathbb{E}\left[\left(f(\mathbf{x}) - \mu(\mathbf{x})\right)\left(f(\mathbf{x}^*) - \mu(\mathbf{x}^*)\right)\right] \end{aligned}$$

  where $\mathbf{x}$ and $\mathbf{x}^*$ represent two different location vectors.

- For the purpose of this paper, we assume the random variables $f_t(\mathbf{x})$ associated to the location vector $\mathbf{x} = \{x^{Lg}, x^{Lt}\}$ to be characterised by a zero-mean and the following covariance function:

$$k(\mathbf{x}, \mathbf{x}^*) = \mathrm{cov}(f_t(\mathbf{x}), f_t(\mathbf{x}^*)) = \left(\sigma_t^f\right)^2 \exp\left[-\frac{1}{2}\left(\mathbf{x} - \mathbf{x}^*\right)^\top \mathbf{M}_t \left(\mathbf{x} - \mathbf{x}^*\right)\right]$$

  which corresponds to the squared exponential covariance function and is fully specified by the hyperparameter σ_t^f and the symmetric matrix $\mathbf{M}_t = diag(\boldsymbol{\theta}_t)^{-1}$, where $\boldsymbol{\theta}_t = \{\theta_t^{Lg}, \theta_t^{Lt}\}$ corresponds to the vector of the longitude and latitude scaling hyperparameters.

Estimation Procedure

- If we assume that the gaussian process as zero-mean $\mu(\mathbf{x}) = 0$ so that $\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x}) \sim \mathcal{N}\left(0, K + \sigma_t^2 \mathbf{I}_n\right)$ where $K = \left(\sigma_t^f\right)^2 K'$ and the correlation K' having elements $k(\mathbf{x_i}, \mathbf{x_j})$ we can write then the marginal likelihood such as:

$$\log p(\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x})|X) = -\frac{1}{2}\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x})^{\top}(K+\sigma_t^2\mathbf{I}_n)^{-1}\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x}) - \frac{1}{2}log|K+\sigma_t^2\mathbf{I}_n| - \frac{n}{2}log2\pi$$

Estimation Procedure

- If we assume that the gaussian process as zero-mean $\mu(\mathbf{x}) = 0$ so that $\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x}) \sim \mathcal{N}\left(0, K + \sigma_t^2 \mathbf{I}_n\right)$ where $K = \left(\sigma_t^f\right)^2 K'$ and the correlation K' having elements $k(\mathbf{x_i}, \mathbf{x_j})$ we can write then the marginal likelihood such as:

$$\log p(\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x})|X) = -\frac{1}{2}\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x})^\top (K+\sigma_t^2\mathbf{I}_n)^{-1}\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x}) - \frac{1}{2}log|K+\sigma_t^2\mathbf{I}_n| - \frac{n}{2}log2\pi$$

- to set the hyperparameters by maximizing the marginal likelihood, we seek the partial derivatives of the marginal likelihood w.r.t. the hyperparameters such that:

$$\begin{aligned}\frac{\partial}{\partial\theta_j}\log p(\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x})|X,\theta) &= \frac{1}{2}\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x})^\top K_{\widetilde{\mathbf{W}}}^{-1}\frac{\partial K}{\partial\theta_j}K_{\widetilde{\mathbf{W}}}^{-1}\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x}) - \frac{1}{2}tr(K_{\widetilde{\mathbf{W}}}\frac{\partial K}{\partial\theta_j}) \\ &= \frac{1}{2}tr((\alpha\alpha^\top - K_{\widetilde{\mathbf{W}}}^{-1})\frac{\partial K}{\partial\theta_j})\end{aligned}$$

where $\alpha = K_{\widetilde{\mathbf{W}}}^{-1}\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x})$ and $K_{\widetilde{\mathbf{W}}} = K + \sigma_t^2\mathbf{I}_n$

Local Approximation GP

- A laGP is a localized approximated emulation by means of a fast sequential updating greedy search algorithm in order to minimize the mean-squared prediction error (MSPE).

Local Approximation GP

- A laGP is a localized approximated emulation by means of a fast sequential updating greedy search algorithm in order to minimize the mean-squared prediction error (MSPE).
- The idea is to remove some vanishingly low impact observed sites while maintain the rest of the reference points under certain criteria, including active learning Cohn (ALC) and MSPE.

Local Approximation GP

- A laGP is a localized approximated emulation by means of a fast sequential updating greedy search algorithm in order to minimize the mean-squared prediction error (MSPE).
- The idea is to remove some vanishingly low impact observed sites while maintain the rest of the reference points under certain criteria, including active learning Cohn (ALC) and MSPE.
- The iterative estimation starts from a small subset $D_{n_0}(x) = (X_{n_0}(x), Y_{n_0}(x))$ close to x and to choose x_{j+1} to augment $X_j(x)$ and thus form a new subset $D_{j+1}(x)$ according to the MSPE objective criteria to minimize which is defined as:

$$\mathrm{J}\left(\mathbf{x}_{j+1}, \mathbf{x}\right) = \mathbb{E}\left\{\left[\mathbf{Y}(x) - \mu_{j+1}\left(\mathbf{x}; D_{j+1}(x), \hat{\theta}_{j+1}\right)\right]^2 \mid D_j(x)\right\}$$

which can be approximated by:

$$\mathrm{J}\left(\mathbf{x}_{j+1}, \mathbf{x}\right) \approx \mathrm{V}_j\left(\mathbf{x}|\mathbf{x}_{j+1}; \hat{\theta}_j\right) + \left(\frac{\partial \mu_j\left(\mathbf{x}; \theta\right)}{\partial \theta}\Big|_{\theta=\hat{\theta}_j}\right)^2 / \mathrm{I}_{j+1}\left(\hat{\theta}_j\right),$$

where I is the expected Fisher information.

Local Approximation GP

- Which is equivalent to:

$$\underset{\mathbf{x}_{j+1} \in \mathbf{x} \setminus \mathbf{x}_j}{\operatorname{argmax}} \left\{ \mathrm{V}_j \left(\mathbf{x}; \theta \right) - \mathrm{V}_{j+1} \left(\mathbf{x}; \theta \right) \right\}. \quad (15)$$

Local Approximation GP

- Which is equivalent to:

$$\underset{\mathbf{x}_{j+1} \in \mathbf{x} \setminus \mathbf{x}_j}{\operatorname{argmax}} \left\{ \mathrm{V}_j\left(\mathbf{x};\theta\right) - \mathrm{V}_{j+1}\left(\mathbf{x};\theta\right) \right\}. \tag{15}$$

- with :

$$\mu(\mathbf{x}) = \left[\frac{\left(1 - r^\top R^{-1} \mathbf{1_n}\right)}{\mathbf{1_n}^\top R^{-1} \mathbf{1_n}} \mathbf{1_n}^\top + r^\top \right] R^{-1} \mathbf{y_t},$$

where r is the vector of correlations between the input $\mathbf{x}$ and $\mathbf{x_{i=1,\ldots,n}}$ at the n design sites, $r = \left[\mathrm{cor}\left(f(\mathbf{x}_1), f(\mathbf{x})\right), \ldots, \mathrm{cor}\left(f(\mathbf{x}_n), f(\mathbf{x})\right)\right]$. While the mean squared error (MSE) is expressed such as follows:

$$\mathrm{V}_j\left(\mathbf{x};\theta\right) = (\hat{\sigma}_t^f)^2 \cdot \left(1 - r^\top R^{-1} r + \frac{\left(1 - \mathbf{1_n}^\top R^{-1} r\right)^2}{\mathbf{1_n}^\top R^{-1} \mathbf{1_n}} \right).$$

Local Approximation GP

- Which is equivalent to:

$$\underset{\mathbf{x}_{j+1} \in \mathbf{x} \setminus \mathbf{x}_j}{\operatorname{argmax}} \left\{ \mathrm{V}_j (\mathbf{x}; \theta) - \mathrm{V}_{j+1} (\mathbf{x}; \theta) \right\}. \tag{15}$$

- with :

$$\mu(\mathbf{x}) = \left[\frac{\left(1 - r^\top R^{-1} \mathbf{1_n}\right)}{\mathbf{1_n}^\top R^{-1} \mathbf{1_n}} \mathbf{1_n}^\top + r^\top \right] R^{-1} \mathbf{y_t},$$

where r is the vector of correlations between the input $\mathbf{x}$ and $\mathbf{x_{i=1,\ldots,n}}$ at the n design sites, $r = [\mathrm{cor}(f(\mathbf{x}_1), f(\mathbf{x})), \ldots, \mathrm{cor}(f(\mathbf{x}_n), f(\mathbf{x}))]$. While the mean squared error (MSE) is expressed such as follows:

$$\mathrm{V}_j (\mathbf{x}; \theta) = (\hat{\sigma}_t^f)^2 \cdot \left(1 - r^\top R^{-1} r + \frac{\left(1 - \mathbf{1_n}^\top R^{-1} r\right)^2}{\mathbf{1_n}^\top R^{-1} \mathbf{1_n}} \right).$$

- We then update the subset to $D_{j+1}(x)$ meanwhile independently compute the hyper-parameter $\hat{\theta}_j(\mathbf{x}) \mid D_j(x)$ by maximizing the likelihood which possibly could smooth spatially over all the locations.

Yield Model

- We denote $\hat{y}^i_{k,t}\left(\tilde{\omega}^i_{\tau^k_t}\right)$ as the best linear unbiased predictor of the yield per hectares at time t of the farm i for the k-th crop and function of the random variable $\tilde{\omega}^i_{\tau^k_t}$ which embodies the precipitation and temperature over the period τ^k_t (Bokusheva, 2014; Roberts and al., 2012) as:

$$\hat{y}^i_{k,t}\left(\tilde{\omega}^i_{\tau^k_t}\right) = \alpha_0 + \alpha_{\mathcal{P},k} \cdot \hat{\mathbf{W}}^{\mathcal{P},i}_{k,\tau^k_t} + \alpha_{\mathcal{T},k} \cdot \hat{\mathbf{W}}^{\mathcal{T},i}_{k,\tau^k_t}, \quad (16)$$

Yield Model

- We denote $\hat{y}^i_{k,t}\left(\tilde{\omega}^i_{\tau^k_t}\right)$ as the best linear unbiased predictor of the yield per hectares at time t of the farm i for the k-th crop and function of the random variable $\tilde{\omega}^i_{\tau^k_t}$ which embodies the precipitation and temperature over the period τ^k_t (Bokusheva, 2014; Roberts and al., 2012) as:

$$\hat{y}^i_{k,t}\left(\tilde{\omega}^i_{\tau^k_t}\right) = \alpha_0 + \alpha_{\mathcal{P},k} \cdot \hat{\mathbf{W}}^{\mathcal{P},i}_{k,\tau^k_t} + \alpha_{\mathcal{T},k} \cdot \hat{\mathbf{W}}^{\mathcal{T},i}_{k,\tau^k_t}, \tag{16}$$

- If we substitute the unbiased out-of-sample predictive value of weather random variables, we will then have:

$$\hat{y}^i_{k,t}\left(\tilde{\omega}^i_{\tau^k_t}\right) = \alpha_0 + \alpha_{\mathcal{P},k}.\left(\bar{w}^{\mathcal{P}}_{\tau^k_t} + \hat{f}^{\mathcal{P}}(\mathbf{x_i})\right) + \alpha_{\mathcal{T},k}.\left(\bar{w}^{\mathcal{T}}_{\tau^k_t} + \hat{f}^{\mathcal{T}}(\mathbf{x_i})\right)$$

where α_0 is constant and $\left(y^i_{k,t} - \hat{y}^i_{k,t}\left(\tilde{\omega}^i_{\tau^k_t}\right)\right) \sim N\left(0, \Psi^i_{\tau^k_t}\right)$ while:

$$\bar{w}^{\mathcal{T}}_t = \mathbb{E}_t\left[\mathbb{E}_{\mathbf{x}}\left(\mathbf{W}^{\mathcal{T}}_t(\mathbf{x})\right)|\mathbf{w}^{\mathcal{T}}_{t-1}, \mathbf{w}^{\mathcal{T}}_{t-2}, \ldots\right]$$

corresponds to the SARIMA expected temperature at the country level. While for the precipitations we have the same expression:

$$\bar{w}^{\mathcal{P}}_t = \mathbb{E}_t\left[\mathbb{E}_{\mathbf{x}}\left(\mathbf{W}^{\mathcal{P}}_t(\mathbf{x})\right)|\mathbf{w}^{\mathcal{P}}_{t-1}, \mathbf{w}^{\mathcal{P}}_{t-2}, \ldots\right]$$

Weather Conditional Loss Distribution

- Assuming that $\tilde{\omega}_t^i = \{\widetilde{\mathbf{W}}_{k,t}^{\mathcal{T},i}, \widetilde{\mathbf{W}}_{k,t}^{\mathcal{P},i}\}$ are both independently and identically normally distributed leads to the farm-i's expected yield $\hat{y}_{k,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right)$ with a variance equals to the i-th element on the diagonal of the variance covariance matrix:
$$\mathbb{V}\left(\mathbf{y}_{k,t}\left(\tilde{\omega}_{\tau_t^k}\right)\right) = \alpha_{\mathcal{P},k}^2 . \mathbb{V}\left(\widetilde{\mathbf{W}}_t^{\mathcal{P}}(\mathbf{x})\right) + \alpha_{\mathcal{T},k}^2 . \mathbb{V}\left(\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x})\right) + \Psi_{\tau_t^k}^i$$
where:
$$\mathbb{V}\left(\widetilde{\mathbf{W}}_t^{\mathcal{T}}(\mathbf{x})\right) = \mathbb{V}\left[\mathbb{E}_{\mathbf{x}}\left(\mathbf{W}_t^{\mathcal{T}}(\mathbf{x})\right)|\mathbf{w}_{t-1}^{\mathcal{T}}, \mathbf{w}_{t-2}^{\mathcal{T}}, \dots\right] I_n + K(\mathbf{x},\mathbf{x}) + \sigma_t^2 I_n,$$
While $\mathbb{V}\left[\mathbb{E}_{\mathbf{x}}\left(\mathbf{W}_t^{\mathcal{T}}(\mathbf{x})\right)|\mathbf{w}_{t-1}^{\mathcal{T}}, \mathbf{w}_{t-2}^{\mathcal{T}}, \dots\right]$ can be derived from $\Gamma(x)$, the autocovariance generating function (AGF) which for summable autocovariance functions $\sum_{h=-\infty}^{\infty}\gamma(h) < \infty$ is defined such that:
$$\Gamma(x) = \sum_{h=-\infty}^{\infty}\gamma(h)x^h \tag{17}$$
where $\gamma(h)$ is the process autocavariance between x_t and x_{t+h}.

Weather Conditional Loss Distribution

- We can then determine another conditional probability of default which is linked now to the weather conditions ω_t and their local impact conditionally on the K net variable prices of the commodities $\tilde{C}_t = \left(\tilde{C}_t^1, \ldots, \tilde{C}_t^K\right)$.

Weather Conditional Loss Distribution

- We can then determine another conditional probability of default which is linked now to the weather conditions ω_t and their local impact conditionally on the K net variable prices of the commodities $\tilde{C}_t = \left(\tilde{C}_t^1, \ldots, \tilde{C}_t^K\right)$.
- Using the properties of the Gaussian process we can then write the value of the conditional expected returns conditionally on the commodity prices filtration $\mathcal{C}_t$:

$$\begin{aligned}
\mathbb{E}\left(\frac{R_{i,t}}{A_{t-1}^i}|\mathcal{H}_t, \mathcal{C}_t\right) &= \sum_{k=1}^K \Delta_{k,t}^i \mathbb{E}\left[y_{k,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right)\tilde{C}_{k,t}|\mathcal{H}_t, \mathcal{C}_t\right] - F_t^i \\
&= \sum_{k=1}^K \Delta_{k,t}^i \mathbb{E}\left[y_{k,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right)|\mathcal{H}_t, \mathcal{C}_t\right]\tilde{C}_{k,t} - F_t^i \\
&= \tilde{C}_t \begin{bmatrix} \Delta_{1,t}^i \mathbb{E}\left[y_{1,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right)|\mathcal{H}_t,\right] \\ \vdots \\ \Delta_{K,t}^i \mathbb{E}\left[y_{K,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right)|\mathcal{H}_t,\right] \end{bmatrix} - F_t^i
\end{aligned}$$

While the log-return variance conditionally on the K net variable prices of the commodities $\tilde{C}_t = \left(\tilde{C}_t^1, \ldots, \tilde{C}_t^K\right)$ is given by:

$$\mathbb{V}\left(\frac{R_{i,t}}{A_{t-1}^i} | \mathcal{H}_t, \mathcal{C}_t\right) = \tilde{C}_t^\top \begin{bmatrix} \left(\Delta_{1,t}^i\right)^2 \mathbb{V}\left[y_{1,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right) | \mathcal{H}_t\right] \\ \vdots \\ \left(\Delta_{K,t}^i\right)^2 \mathbb{V}\left[y_{K,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right) | \mathcal{H}_t\right] \end{bmatrix} \tilde{C}_t$$

While the log-return variance conditionally on the K net variable prices of the commodities $\tilde{C}_t = \left(\tilde{C}_t^1, \dots, \tilde{C}_t^K\right)$ is given by:

$$\mathbb{V}\left(\frac{R_{i,t}}{A_{t-1}^i}|\mathcal{H}_t, \mathcal{C}_t\right) = \tilde{C}_t^\top \begin{bmatrix} \left(\Delta_{1,t}^i\right)^2 \mathbb{V}\left[y_{1,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right)|\mathcal{H}_t\right] \\ \vdots \\ \left(\Delta_{K,t}^i\right)^2 \mathbb{V}\left[y_{K,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right)|\mathcal{H}_t\right] \end{bmatrix} \tilde{C}_t$$

- Eventually the local probability of default of the farmer i can be expressed such as:

$$\begin{aligned} PD_i|\mathcal{H}_t, \mathcal{C}_t &= Pr\left[A_{i,t} \leq D_{i,t}|\mathcal{H}_t, \mathcal{C}_t\right] \\ &= Pr\left[\frac{R_t^i}{A_{t-1}^i} \leq \frac{D_{i,t}}{A_{i,t-1}} - 1|\mathcal{H}_t, \mathcal{C}_t\right] \\ &= \Phi\left[\frac{\left(\frac{D_{i,t}}{A_{i,t-1}} - 1\right) - \tilde{C}_t \begin{bmatrix} \Delta_{1,t}^i \mathbb{E}\left[y_{1,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right)\middle| \mathcal{H}_t\right] \\ \vdots \\ \Delta_{K,t}^i \mathbb{E}\left[y_{K,t}^i\left(\tilde{\omega}_{\tau_t^k}^i\right)\middle| \mathcal{H}_t\right] \end{bmatrix} + F_t^i}{\sqrt{\mathbb{V}\left(\frac{R_{i,t}}{A_{t-1}^i}|\mathcal{H}_t, \mathcal{C}_t\right)}}\right. \end{aligned}$$

Weather Conditional Loss Distribution

The default correlation between the farmer i and farmer j can naturally be computed under the assumption of Gaussian joint distribution such as:

$$\rho_{ij}|\mathcal{H}_t,\mathcal{C}_t = \frac{Pr\left(A_{i,t} \leq D_{i,t}, A_{j,t} \leq D_{j,t}|\mathcal{H}_t,\mathcal{C}_t\right) - PD_i|\mathcal{H}_t,\mathcal{C}_t PD_j|\mathcal{H}_t,\mathcal{C}_t}{\sqrt{PD_i|\mathcal{H}_t,\mathcal{C}_t\left(1 - PD_i|\mathcal{H}_t,\mathcal{C}_t\right) PD_j|\mathcal{H}_t,\mathcal{C}_t\left(1 - PD_j|\mathcal{H}_t,\mathcal{C}_t\right)}}$$

where:

$$Pr\left(\frac{R_t^i}{A_{t-1}^i} \leq \frac{D_{i,t}}{A_{i,t-1}} - 1, \frac{R_t^j}{A_{t-1}^j} \leq \frac{D_{j,t}}{A_{j,t-1}} - 1|\mathcal{H}_t,\mathcal{C}_t\right)$$
$$= \int_0^{\frac{D_{i,t}}{A_{i,t-1}}-1}\int_0^{\frac{D_{j,t}}{A_{j,t-1}}-1} MVN\left(\frac{R_{i,t}}{A_{t-1}^i}, \frac{R_{j,t}}{A_{t-1}^j}, \theta_{ij}|\mathcal{H}_t,\mathcal{C}_t\right) dR_{j,t}dR_{i,t}$$

with:

$$\theta_{ij}|\mathcal{H}_t,\mathcal{C}_t = \frac{Cov(R_{i,t}, R_{j,t}|\mathcal{H}_t,\mathcal{C}_t)}{\sqrt{\mathbb{V}\left(R_{i,t}|\mathcal{H}_t,\mathcal{C}_t\right)\mathbb{V}\left(R_{j,t}|\mathcal{H}_t,\mathcal{C}_t\right)}}$$

Weather Conditional Loss Distribution

while:

$$Cov(R_i, R_j|\mathcal{H}_t, \mathcal{C}_t) = \tilde{C}_t^\top \begin{bmatrix} \Delta^i_{1,t} Cov\left[y^i_{1,t}\left(\tilde{\omega}^i_{\tau^k_t}\right), y^j_{1,t}\left(\tilde{\omega}^j_{\tau^k_t}\right)|\mathcal{H}_t,\right]\Delta^j_{1,t} \\ \vdots \\ \Delta^i_{K,t}\mathrm{Cov}\left[y^i_{K,t}\left(\tilde{\omega}^i_{\tau^k_t}\right), y^j_{K,t}\left(\tilde{\omega}^j_{\tau^k_t}\right)|\mathcal{H}_t,\right]\Delta^j_{K,t} \end{bmatrix} \tilde{C}_t$$

where:

$$Cov\left[y^i_{1,t}\left(\tilde{\omega}^i_{\tau^k_t}\right), y^j_{1,t}\left(\tilde{\omega}^j_{\tau^k_t}\right)|\mathcal{H}_t,\right] = \mathbb{V}\left(\mathbf{y}_{k,t}\left(\tilde{\omega}_{\tau^k_t}\right)\right)_{i,j}$$

with

$$\mathbb{V}\left(\mathbf{y}_{k,t}\left(\tilde{\omega}_{\tau^k_t}\right)\right) = \alpha^2_{\mathcal{P},k}.\,\mathbb{V}\left(\widetilde{\mathbf{W}}^{\mathcal{P}}_t(\mathbf{x})\right) + \alpha^2_{\mathcal{T},k}.\,\mathbb{V}\left(\widetilde{\mathbf{W}}^{\mathcal{T}}_t(\mathbf{x})\right) + \Psi^i_{\tau^k_t}$$

and

$$\mathbb{V}\left(\widetilde{\mathbf{W}}^{\mathcal{T}}_t(\mathbf{x})\right) = \mathbb{V}\left[\mathbb{E}_{\mathbf{x}}\left(\mathbf{W}^{\mathcal{T}}_t(\mathbf{x})\right)|\mathbf{w}^{\mathcal{T}}_{t-1}, \mathbf{w}^{\mathcal{T}}_{t-2}, \ldots\right] I_n + K(\mathbf{x},\mathbf{x}) + \sigma^2_t I_n,$$

Weather Conditional Loss Distribution

and:

$$MVN\left(R_{i,t}, R_{j,t}, \theta_{ij}|\mathcal{H}_t, \mathcal{C}_t\right) = \frac{1}{2\pi\sqrt{\mathbb{V}(R_{i,t}|\mathcal{H}_t,\mathcal{C}_t)\mathbb{V}(R_{j,t}|\mathcal{H}_t,\mathcal{C}_t)\left(1-(\theta_{ij}|\mathcal{H}_t,\mathcal{C}_t)^2\right)}}$$
$$\times exp\left\{\frac{-1}{2\left(1-(\theta_{ij}|\mathcal{H}_t,\mathcal{C}_t)^2\right)}\left(\frac{(R_{i,t}-\mathbb{E}(R_{i,t}|\mathcal{H}_t,\mathcal{C}_t))^2}{\mathbb{V}(R_{i,t}|\mathcal{H}_t,\mathcal{C}_t)} + \frac{(R_{j,t}-\mathbb{E}(R_{j,t}|\mathcal{H}_t,\mathcal{C}_t))^2}{\mathbb{V}(R_{j,t}|\mathcal{H}_t,\mathcal{C}_t)}\right)\right.$$
$$\left.\times exp\left\{\frac{-1}{2\left(1-(\theta_{ij}|\mathcal{H}_t,\mathcal{C}_t)^2\right)}\left(-\frac{2(\theta_{ij}|\mathcal{H}_t,\mathcal{C}_t)(R_{i,t}-\mathbb{E}(R_{i,t}|\mathcal{H}_t,\mathcal{C}_t))(R_{j,t}-\mathbb{E}(R_{j,t}|\mathcal{H}_t,\mathcal{C}_t))}{\sqrt{\mathbb{V}(R_{i,t}|\mathcal{H}_t,\mathcal{C}_t)}\sqrt{\mathbb{V}(Rj,t|\mathcal{H}_t,\mathcal{C}_t)}}\right)\right\}\right.$$

we can then compute respectively the portfolio loss function L, the expected loss EL and the unexpected loss UL which respectively represents the expected value and the variance of the loss function associated to this portfolio of loans conditionally to the filtrations $\mathcal{C}_t$ and $\mathcal{H}_t$.

$$L|\mathcal{H}_t, \mathcal{C}_t = \sum_{n=1}^{N} EAD_n LGD_n D_n|\mathcal{H}_t, \mathcal{C}_t$$

where $D_n|\mathcal{H}_t, \mathcal{C}_t \sim Bernoulli\left(PD_n|\mathcal{H}_t, \mathcal{C}_t\right)$

$$EL|\mathcal{H}_t, \mathcal{C}_t = \sum^{N} EAD_n ELGD_n PD_n|\mathcal{H}_t, \mathcal{C}_t$$

Weather Conditional Loss Distribution

$$UL|\mathcal{H}_t, \mathcal{C}_t = \sqrt{V\left(L|\mathcal{H}_t, \mathcal{C}_t\right)}$$
$$= \sqrt{\sum_{n,k=1}^{N} EAD_n EAD_k ELGD_n ELGD_k \rho_{nk} \sqrt{PD_n|\mathcal{H}_t, \mathcal{C}_t\left(1 - PD_n|\mathcal{H}_t, \mathcal{C}_t\right) PD_k|\mathcal{H}_t, \mathcal{C}_t\left(1 - PD_k|\mathcal{H}_t, \mathcal{C}_t\right)}} \tag{18}$$

Farm's Return Distribution

Considering the initial equation in matrix form:

$$\left[R_t \circ A_{t-1}^{\circ -1} \middle| \mathcal{F}_{t-1}\right] = \left[\triangle_t \circ y_t(\tilde{\omega}) \cdot \tilde{C}_{k,t} | \mathcal{F}_{t-1}\right] - F_t, \quad (19)$$

The global risk being the product of two MVN distributions we obtain a unique MVN distribution with expected value:

$$\mu_R = \Sigma_R \left(\Sigma_y^{-1}\mu_y + \Sigma_C^{-1}\mu_C\right) \quad (20)$$

and a variance equals to:

$$\Sigma_R = \left(\Sigma_y^{-1} + \Sigma_C^{-1}\right)^{-1} \quad (21)$$

with a normalizing constant:

$$\Sigma_R = (2\pi)^{-n/2}|\Sigma_y + \Sigma_C|^{-1/2} exp\left(-\frac{1}{2}(\mu_y - \mu_R)^\top (\Sigma_y + \Sigma_C)(\mu_y - \mu_R)\right) \quad (22)$$

Outline

Data

- Real data obtained from a french fertilizer company, the Roullier Group.

Data

- Real data obtained from a french fertilizer company, the Roullier Group.
- 2014 clients database containing 11,982 farms located in 41 regions in Romania,

Data

- Real data obtained from a french fertilizer company, the Roullier Group.
- 2014 clients database containing 11,982 farms located in 41 regions in Romania,
- Data attributes include:
  - type of crops,
  - crops rotation,
  - number of hectares cultivated

Data

- Real data obtained from a french fertilizer company, the Roullier Group.
- 2014 clients database containing 11,982 farms located in 41 regions in Romania,
- Data attributes include:
    - type of crops,
    - crops rotation,
    - number of hectares cultivated
    - a precise geolocalisation of each farm.

Data

- Real data obtained from a french fertilizer company, the Roullier Group.
- 2014 clients database containing 11,982 farms located in 41 regions in Romania,
- Data attributes include:
  - type of crops,
  - crops rotation,
  - number of hectares cultivated
  - a precise geolocalisation of each farm.
- This farms sample adds up to 4.6 million hectares which occupied over one-third of the total Romanian utilized agricultural area (UAA)

Utilized Agricultural Area

Utilized agricultural area (UAA) in EU:

(40.0%) of the total land area of the EU-28 in 2013

Utilized Agricultural Area

Utilized agricultural area (UAA) in EU:

(40.0%) of the total land area of the EU-28 in 2013

1. France with 27,8 million hectares (16%)
2. Spain, with 23,75 million hectares (13,6%)
3. United Kingdom, with 16,88 million hectares (9,7%)
4. Germany, with 16,7 million hectares (9,6%)
5. Poland, with 14,4 million hectares (8,3%)

Utilized Agricultural Area

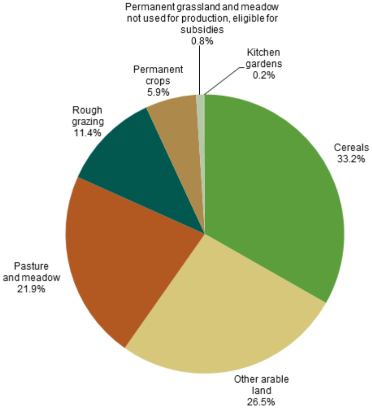
Utilized agricultural area (UAA) in EU:

(40.0%) of the total land area of the EU-28 in 2013

1. France with 27,8 million hectares (16%)
2. Spain, with 23,75 million hectares (13,6%)
3. United Kingdom, with 16,88 million hectares (9,7%)
4. Germany, with 16,7 million hectares (9,6%)
5. Poland, with 14,4 million hectares (8,3%)
6. Romania holds 7,6% of the , with 13,05 million hectares. . .

Utilized Agricultural Area

Utilised agricultural area by land use:



(*) Estimates.
Source: Eurostat (online data code: ef_oluft)

Agricultural Output Breakdown

Output components	2013	2014	2015		
	Million EUR		Million EUR	% of total	% of EU-28
Cereals:	**4501**	**3853**	**3184**	**33.7%**	**6.4%**
Wheat and spelt	1 351	1 252	1 295	13.7%	4.9%
Rye and meslin	4	4	4	0.0%	0.4%
Barley	357	341	317	3.4%	3.5%
Oats and summer cereal mixtures	93	81	65	0.7%	5.4%
Grain maize	2638	2125	1453	15.4%	16.5%
Rice	15	10	12	0.1%	1.4%
Other cereals	42	38	37	0.4%	1.9%
Industrial crops:	**1 238**	**1 143**	**1 109**	**11.7%**	**5.7%**
Oil seeds and oleaginous fruits	1125	1012	1002	10.6%	8.0%
Protein crops	29	32	34	0.4%	3.0%
Raw tobacco	2	1	1	0.0%	0.3%
Sugar beet	39	50	32	0.3%	0.9%
Other industrial crops	42	48	39	0.4%	2.4%
Forage plants	**1705**	**1465**	**1314**	**13.9%**	**5.4%**
Vegetables and horticultural products	**2 024**	**2 021**	**1 878**	**19.9%**	**3.5%**
Potatoes	**1289**	**1161**	**678**	**7.2%**	**7.1%**
Fruits	**1087**	**1137**	**1093**	**11.6%**	**4.3%**
Wine	**306**	**240**	**185**	**2.0%**	**0.8%**
Olive oil	**:**	**:**	**:**	**:**	**:**
Other crop products	**35**	**19**	**10**	**0.1%**	**0.3%**
Crop output	**12 185**	**11 040**	**9 450**	**70.2%**	**4.5%**
Animals:	**1 911**	**1 891**	**1 801**	**44.9%**	**1.9%**
Cattle	303	271	333	8.3%	1.0%
Pigs	965	896	779	19.4%	2.5%
Equines	22	19	22	0.6%	2.2%
Sheep and goats	196	253	225	5.6%	4.3%
Poultry	425	451	441	11.0%	2.1%
Other animals	1	1	1	0.0%	0.0%
Animal products:	**1 996**	**2 076**	**2 207**	**55.1%**	**3.4%**
Milk	1 012	1 106	1 111	27.7%	2.1%
Eggs	662	685	777	19.4%	2.1%
Other animal products	322	285	319	8.0%	2.1%
Animal output	**3 908**	**3 967**	**1 801**	**29.8%**	**2.5%**
Agricultural goods output	**16 092**	**15 007**	**13 458**	**100.0%**	**3.6%**
Gross value added at basic prices	**7 621**	**7 099**	**6 444**		**4.0%**

Source: Eurostat, Economic Accounts for Agriculture (values at current producer prices).
Updated: March 2016

Figure: Agricultural Output per Type, Romania, 2013 (% share of utilised agricultural area)

Farming Data

The Group Roullier provided us a part of their clients database with:

- Precise geolocalisation of each farm
- Accounting information about more than 12,000 farms located in Romania for the last 5 years
- Types of crop and number of hectares cultivated per farm
- Yields per hectare per crop per region since 1990

Farming Data

The Group Roullier provided us a part of their clients database with:

- Precise geolocalisation of each farm
- Accounting information about more than 12,000 farms located in Romania for the last 5 years
- Types of crop and number of hectares cultivated per farm
- Yields per hectare per crop per region since 1990

We collected the market prices time series for the five main crops:

- Wheat
- Corn
- Barley
- Sunflower
- Rapeseed

Farming Data

The Group Roullier provided us a part of their clients database with:

- Precise geolocalisation of each farm
- Accounting information about more than 12,000 farms located in Romania for the last 5 years
- Types of crop and number of hectares cultivated per farm
- Yields per hectare per crop per region since 1990

We collected the market prices time series for the five main crops:

- Wheat
- Corn
- Barley
- Sunflower
- Rapeseed

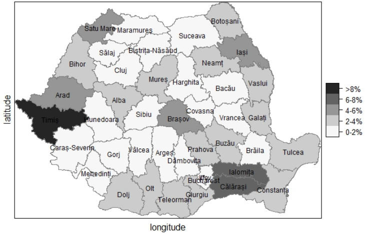
We got access to a European weather database with the following characteristics:

- Daily precipitations (over more than 20 years)
- Daily mean, max and min temperatures (over more than 20 years)
- For 40 different weather stations in Romania, Ukraine, Moldova, Hungary and Serbia

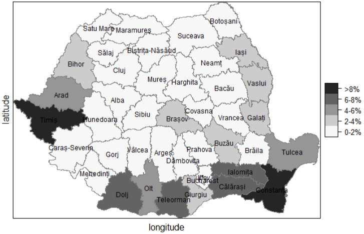
Data

Table: Crops growing seasons and corresponding critical growing period

2 Crops	Whole growing period	**critical growing period**
Wheat (W)	Sep/Oct - July/Aug	April - July
Corn (C)	April/May - Aug/Sep	June - Aug



(a) Corn arable hectares



(b) Wheat arable hectares

Data

- We compare several models:

Group	Model
Without GP	(1) Most nearest station(region central point)
	(2) Most nearest station(farm level simple average)
	(3) Weighted inverse distance
	(4) Weighted farm hectares
	(5) Weighted distance-hectares
With laGP	(6) Simple average
	(7) Weighted inverse distance
	(8) Weighted farm hectares
	(9) Weighted distance-hectares
With GPfit	(10) Simple average
	(11) Weighted inverse distance
	(12) Weighted farm hectares
	(13) Weighted distance-hectares

- Physical distance φ between any two locations given longitude λ_x and latitude ω_x is measured as in (Norton et al., 2012):

$$\varphi = R \cdot Cos^{-1}\left(Sin(\omega_1) \cdot Sin(\omega_2) + Cos(\omega_1) \cdot Cos(\omega_2) \cdot Cos(\lambda_2 - \lambda_1)\right)$$

where R is a constant stand for the radius of the sphere (3963.1 miles).

Model Fitting Quality

Region	Wheat α_0	$\alpha_{P,k}$	$\alpha_{T,k}$	Ad-R^2	LB-Q	Arch	KS	Region	Corn α_0	$\alpha_{P,k}$	$\alpha_{T,k}$	Ad-R^2	LB-Q	Arch	KS
Arad	462.63	76.93*	14.16	0.06	0	0	0	**Arad**	5019.51***	97.74***	−32.14**	0.34	1	0	0
Bihor	-329.15	87.55*	19.96*	0.10	0	0	0	**Bihor**	1298.85	115.13***	4.30	0.20	1	0	0
Covasna	2032.47**	18.47	9.76	-0.02	0	0	0	Covasna	6454.78***	26.60	−41.93**	0.16	1	0	0
Dolj	-88.96	114.79**	5.54	0.13	0	0	0	**Dolj**	6289.38**	124.47**	−56.85**	0.31	1	0	0
Giurgiu	308.03	94.27**	8.10	0.09	0	0	0	**Giurgiu**	5790.82**	130.11**	−50.03*	0.28	0	0	0
Gorj	1773.03	61.41	-3.24	0.04	0	0	0	Gorj	5324.84***	108.31***	−43.83**	0.37	0	0	0
Harghita	2112.35***	16.64	-1.78	-0.06	0	0	0	Harghita	3482.61*	34.25	-12.29	-0.03	1	1	0
Hunedoara	1744.69	35.08	3.94	-0.05	0	0	0	Hunedoara	3212.16**	62.54*	-11.71	0.13	0	1	0
Ilfov	-1472.03	140.70***	19.55**	0.19	1	0	0	Ilfov	3617.22**	159.99***	−31.27*	0.36	0	0	0
Alba	3189.39**	6.96	-6.36	-0.07	0	0	0	Alba	5557.07***	53.59	−33.41*	0.13	1	0	0
Arges	1797.51*	27.72	5.53	-0.04	0	0	0	Arges	2515.45**	125.17***	-12.25	0.26	0	0	0
Bacau	858.54	68.13*	6.95	0.06	0	0	0	Bacau	2648.02	111.43***	-13.52	0.23	0	0	0
Bistrita Nasaud	1892.33**	29.01	1.86	-0.04	0	0	0	Bistrita Nasaud	2309.51***	26.30	0.82	-0.01	0	0	0
Botosani	94.93	78.87	10.00	0.02	0	0	0	**Botosani**	5331.78***	24.26	−31.07*	0.04	1	1	0
Braila	-378.10	121.63***	12.38	0.19	0	0	0	**Braila**	7880.84***	99.95*	−59.98***	0.26	0	0	0
Brasov	2674.44***	22.81	-3.70	0.00	0	0	0	Brasov	4643.84***	1.26	−20.12**	0.11	0	1	0
Buzau	-276.11	137.12***	2.94	0.23	0	0	0	**Buzau**	4586.60**	129.65***	−40.18**	0.31	0	0	0
Calarasi	-119.54	157.75***	5.94	0.22	0	0	0	**Calarasi**	6238.00***	158.44***	−51.52**	0.41	0	0	0
Caras Severin	1262.07	19.28	11.90	-0.02	0	0	0	Caras Severin	1526.01	94.57***	-1.51	0.20	0	1	0
Cluj	954.70	56.58	9.42	0.01	0	0	0	Cluj	3985.07**	36.93	-15.92	0.00	1	0	0
Constanta	1049.41	90.49*	0.89	0.07	0	0	0	**Constanta**	3566.98*	154.42***	−36.00*	0.40	0	0	0
Dambovita	1231.34	44.89	5.49	-0.04	1	0	0	**Dambovita**	3440.14***	90.50***	−20.64*	0.28	0	0	0
Galati	2256.85	59.00	-11.48	0.14	0	0	0	**Galati**	4056.57**	135.34***	−37.08*	0.35	0	0	0
Ialomita	218.11	146.74**	0.72	0.18	0	0	0	**Ialomita**	2528.02	156.19**	-16.88	0.17	0	0	0
Iasi	131.84	122.79**	2.87	0.20	0	0	0	**Iasi**	5339.59***	84.01***	−41.99***	0.39	0	0	0
Maramures	1085.22	26.63	8.88	0.03	0	0	0	Maramures	2545.80**	5.34	1.15	-0.09	1	0	0
Mehedinti	283.94	103.88**	5.45	0.12	0	0	0	**Mehedinti**	4339.33*	161.33***	−45.73**	0.35	0	0	0
Mures	2157.15*	41.77	1.54	0.01	0	0	0	Mures	3129.21**	29.23	1.22	-0.04	0	0	0
Neamt	1542.09	79.67*	0.77	0.06	0	0	0	Neamt	1517.66	76.78	9.64	0.03	0	1	0
Salaj	1337.75	38.42	4.86	-0.04	0	0	0	Salaj	3550.41**	10.59	-9.42	-0.06	0	1	0
Sibiu	2882.86**	1.63	-3.56	-0.08	0	0	0	Sibiu	4570.50***	-1.15	-14.28	-0.05	1	0	0
Timis	-23.29	65.87	24.41*	0.09	0	0	0	**Timis**	2925.89*	108.74***	-12.82	0.28	0	0	0
Valcea	2297.65***	54.41	-2.81	0.12	0	0	0	Valcea	4750.23***	56.93	−17.28**	0.20	0	0	0
Olt	1599.65	84.96*	-3.21	0.11	0	0	0	**Olt**	4105.78**	124.14***	−33.32*	0.34	0	0	0
Prahova	405.87	103.14**	4.92	0.12	0	0	0	Prahova	6429.79***	117.27***	−52.26***	0.44	0	0	0
Satu Mare	2059.53**	45.36	4.09	-0.01	0	0	0	Satu Mare	2852.89*	41.74	0.46	-0.04	0	0	0
Suceava	1739.07	19.23	6.39	-0.07	0	0	0	Suceava	3679.85***	-14.99	-3.52	-0.07	1	0	0
Teleorman	2100.38	55.85	-2.49	0.02	0	0	0	**Teleorman**	2148.81	222.59***	−27.24*	0.65	0	0	0
Tulcea	733.47	80.72	-0.93	0.07	0	0	0	Tulcea	4492.78*	95.34	-33.92	0.09	0	0	0
Vaslui	-1147.34	124.76***	13.74	0.24	0	0	0	**Vaslui**	1716.83	123.26***	-16.66	0.32	0	0	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

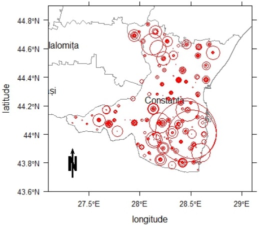
- The SARIMA time series coupled spatial Gaussian process model exhibits distinguishable superiority compared with nonGP approach

Confidence interval	Group	Model	Wheat		Corn	
			Whole regions(41)	Prolific zone(16)	Whole regions(41)	Prolific zone(18)
Panel : 5%	Without GP	(1)	41%	44%	68%	78%
		(2)	46%	69%	71%	94%
		(3)	51%	75%	73%	94%
		(4)	51%	69%	76%	94%
		(5)	54%	75%	76%	94%
	With laGP	(6)	68%	63%	93%	100%
		(7)	73%	75%	93%	100%
		(8)	71%	75%	93%	94%
		(9)	76%	75%	95%	100%
	With GPfit	(10)	73%	94%	98%	100%
		(11)	73%	94%	98%	100%
		(12)	68%	88%	98%	100%
		(13)	71%	100%	98%	100%

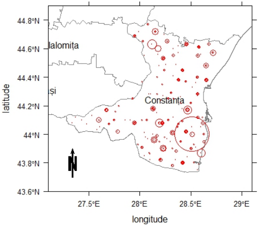
- Weighted distance-hectares ratio method (Model (5), Model (9) and Model (13)) provides us the best estimation results compared with the other weighting methods.

Farms Size Dispersion

- degree of dispersion of the farms crop size characterizing the region of Constanta, a highly productive area of the south east of Romania



(c) Corn arable hectares



(d) Wheat arable hectares

The Balance Sheet

Credit risk profile through Weather risk model at 0.95 quantile weather condition

Statistic	1987 sample			103 sample		
	Benchmark	Correl=0	Correl=1	Benchmark	Correl=0	Correl=1
Actual farm obs.	1442	1442	1442	63	63	63
Mean Prob. of default	5.79%	5.79%	5.79%	6.20%	6.20%	6.20%

Credit risk profile through Weather risk model under mean weather condition

Mean Prob. of default	4.69%	4.69%	4.69%	3.18%	3.18%	3.18%

Credit risk profile through commodity default risk model

Statistic	1987 sample			103 sample		
	Benchmark	Correl=0	Correl=1	Benchmark	Correl=0	Correl=1
Actual farm obs.	1442	1442	1442	63	63	63
Mean Prob. of default	4.66%	4.66%	4.66%	3.98%	3.98%	3.98%

More diversified than average for standard weather conditions

Less diversified than average for extreme weather conditions

- We proposed a credit risk model taking into account the impacts of the weather conditions upon farmers profits

- We proposed a credit risk model taking into account the impacts of the weather conditions upon farmers profits
- We combine GP with a SARIMA time-series model for handling spatio-temporal weather information.

- We proposed a credit risk model taking into account the impacts of the weather conditions upon farmers profits
- We combine GP with a SARIMA time-series model for handling spatio-temporal weather information.
- Through our model we distinguish a global and a local risk of credit dependence

- We proposed a credit risk model taking into account the impacts of the weather conditions upon farmers profits
- We combine GP with a SARIMA time-series model for handling spatio-temporal weather information.
- Through our model we distinguish a global and a local risk of credit dependence
- We treat separately the conditional loss distribution for the commodity risk and the conditional loss distribution associated to the weather risk

- We proposed a credit risk model taking into account the impacts of the weather conditions upon farmers profits
- We combine GP with a SARIMA time-series model for handling spatio-temporal weather information.
- Through our model we distinguish a global and a local risk of credit dependence
- We treat separately the conditional loss distribution for the commodity risk and the conditional loss distribution associated to the weather risk
- If we assume the two sources of risk as independent we also proposed a version where both risk sources are included

... Thank you!

Thank you