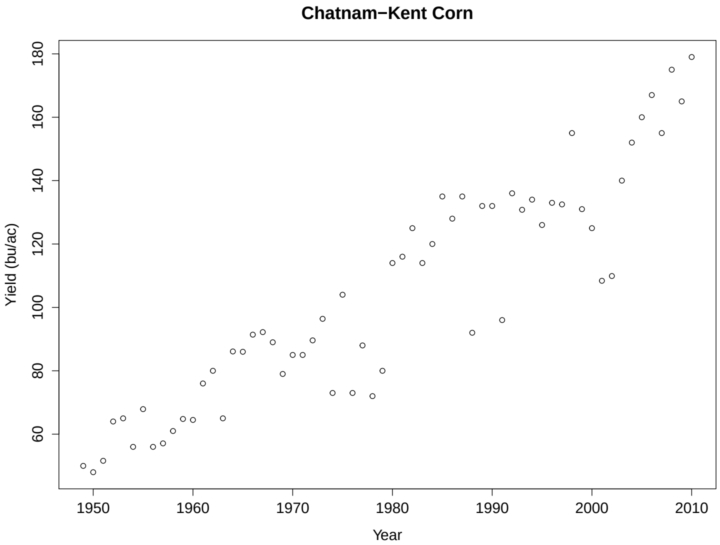
The Effect of Technological Change on Crop Yields in Regular and Catastrophic Years

Tor Tolhurst & Alan P. Ker

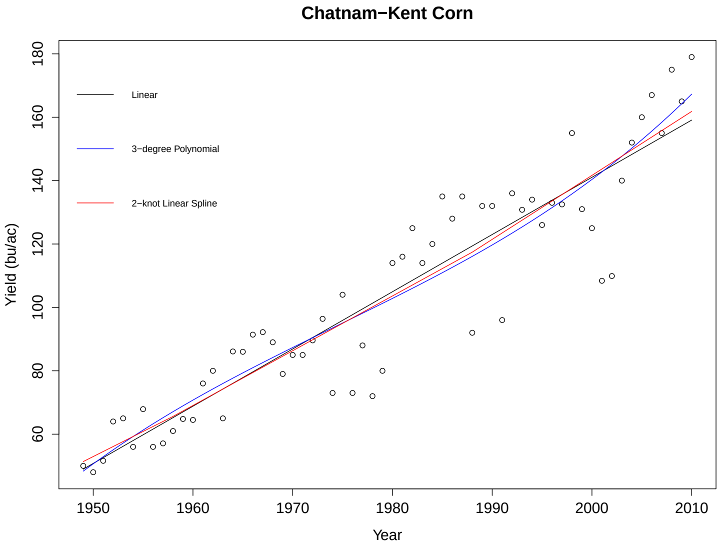
University of Guelph
Department of Food, Agricultural and Resource Economics

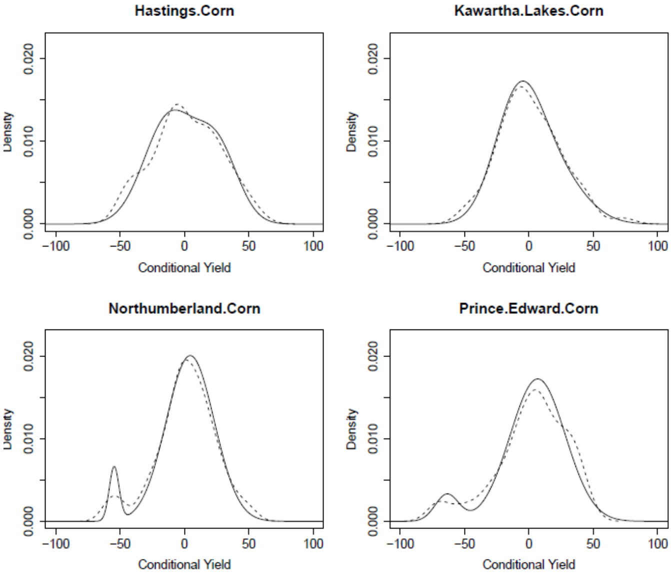
Annual Workshop of the ERCA Research Network on the Structure and Performance of Agriculture and Agri-products Industry
March 8, 2013

OAC Food Agriculture Communities Environment

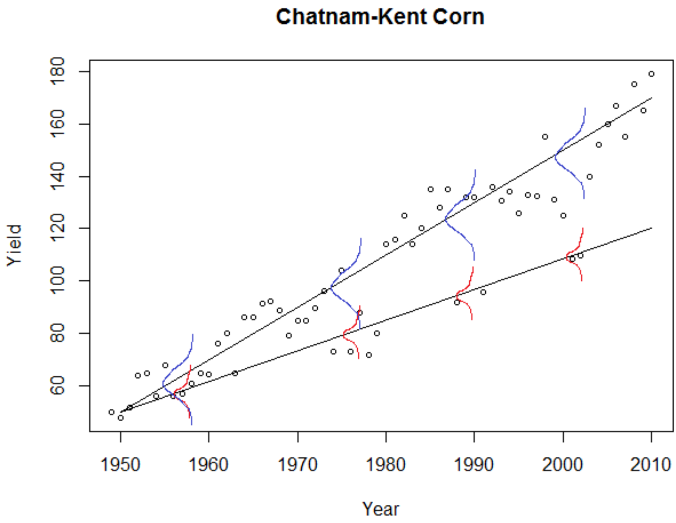


Chatnam−Kent Corn
Yield (bu/ac)
60
80
100
120
140
160
180
1950
1960
1970
1980
1990
2000
2010
Year


Chatnam−Kent Corn
Linear
3−degree Polynomial
2−knot Linear Spline
Yield (bu/ac)
60
80
100
120
140
160
180
1950
1960
1970
1980
1990
2000
2010
Year



OAC Food Agriculture Communities Environment


Chatnam-Kent Corn
Yield
Year
60
80
100
120
140
160
180
1950
1960
1970
1980
1990
2000
2010

Outline

Introduction
- Modelling Yield Distributions
- Technological Trend

Empirical Framework
- Normal Mixture Model
- Expectation-Maximization Algorithm

Estimation
- Ontario County-Level Field Crop Yield Data
- Preliminary Estimation Results
- Hypothesis Test Results

Conclusions and Caveats

Modelling Conditional Crop Yield Distributions

Overview of Process

1. Condition observed yields for technological trend
2. Test and (possibly) adjust for heteroscedasticity [1]
3. Estimate conditional crop yield distribution with time-conditioned crop yields

[1] Forthcoming manuscript

Approaches from the Literature

Conditional Yield Density Models

Some examples:

- Normal (Botts & Boles 1958; Just & Weninger 1999)
- Lognormal (Day 1965)
- Gamma (Gallagher 1987)
- Beta (Nelson & Preckel 1989)
- Mixture of two normals (Ker 1996)
- Nonparametric kernel densities (Goodwin & Ker 1998)
- Semiparametric (Ker and Coble 2003)
- Logistic (Atwood, Shaik & Watts 2003)
- Weibull (Sherrick et al 2004)

Approaches from the Literature

Technological Trend

Some examples:

- Simple linear trend
- Piecewise linear splines (Skees & Reed 1986)
- Stochastic Kalman filter (Kaylen & Koroma 1991)
- ARIMA (Goodwin & Ker 1998)
- Polynomial trend (Just & Weninger 1999)
- Spatio-temporal approach (Ozaki & Silva 2009)

An Alternative Approach

Using normal mixture model to estimate density allows estimation of unique trend in the individual components

Two questions:

1. Is technology effecting “regular” and “catastrophic” year yields at the same rate?
   - Estimate a technological trend coefficient β specific to each component of the mixture model
2. Has technology effected the probability of a “regular” versus “catastrophic” year?

Normal Mixture Model

Definition

Finite mixture of normal distributions

$$x \sim \sum_{j=1}^{J} \lambda_j N(\mu_j, \sigma_j^2) \quad (1)$$

where

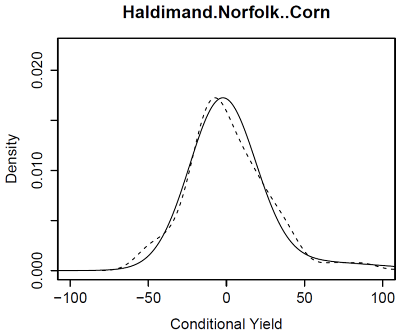
- $J \in \mathbb{N}$ is the finite number of mixture components
- λ_j is the mixture weight for the j^{th} component
- $N(\mu_j, \sigma_j^2)$ is the normal density function parameterized for each j^{th} component by θ_j evaluated at x

Here we select two components: $J = 2$

- Mixture model of two normals is quite flexible and can accomodate:
  1. Symmetrical unimodal densities
  2. Skewed unimodal densities
  3. Bimodal densities
- Typically the literature considers only (1) and (2)
- Some examples with real data

Normal Mixture Model Fitting Examples: Corn

Estimated Conditional Crop Yield Densities

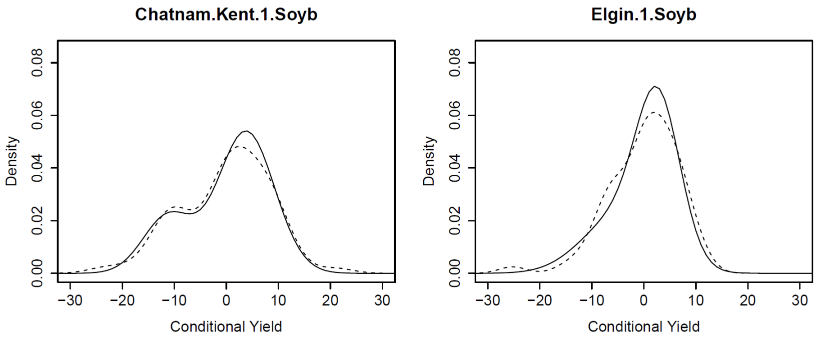


Full lines illustrate EM-estimated normal mixture model

Dashed lines illustrate nonparametric kernel density estimate for comparison

Normal Mixture Model Fitting Examples: Soybeans

Estimated Conditional Crop Yield Densities

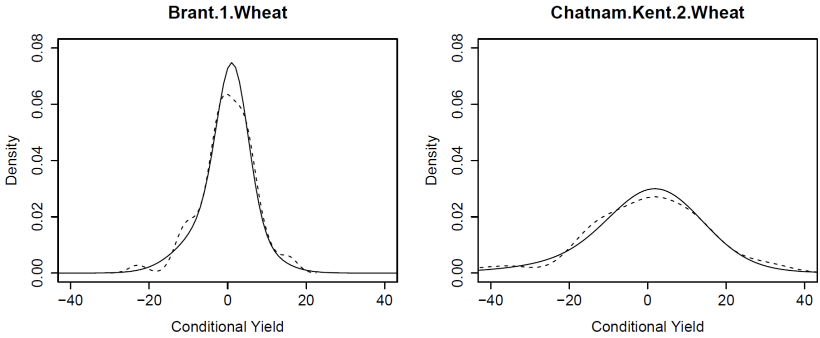


Full lines illustrate EM-estimated normal mixture model

Dashed lines illustrate nonparametric kernel density estimate for comparison

Normal Mixture Model Fitting Examples: Wheat

Estimated Conditional Crop Yield Densities



Full lines illustrate EM-estimated normal mixture model

Dashed lines illustrate nonparametric kernel density estimate for comparison

Normal Mixture Model

Definition

Interested in the rate of technological change in the component means

- Model $\mu_j = f(t) = \alpha_j + \beta_j \cdot t$

Further let $\mu_2 > \mu_1$ then

- component 1 is a “catastrophic” year denoted by c subscript
- component 2 is a “regular” year denoted by r subscript

Normal Mixture Model

Model to be Estimated

$$x_t \sim (1-\lambda)N(\alpha_c + \beta_c \cdot t, \sigma_c^2) + \lambda N(\alpha_r + \beta_r \cdot t, \sigma_r^2) \qquad (2)$$

- x_t is a vector of observed crop yields corresponding to t
- t is a vector of years corresponding to x_t
- λ is the probability of a “regular” year
- α_c is the intercept coefficient and β_c is the slope coefficient for technological trend in the “catastrophic” year component
- σ_c^2 is the variance of the “catastrophic” year component
- α_r is the intercept coefficient and β_r is the slope coefficient for technological trend in the “regular” year component
- σ_r^2 is the variance of the “regular” year component

Expectation-Maximization Algorithm

Overview of the EM Algorithm

- Name from its two steps
  - E Expectation Step
  - M Maximization Step
- Maximum-likelihood approach
- Likelihood of the Normal Mixture Model is:
  - unbounded (likelihood goes to ∞ when $\sigma_j^2 \to 0$)
  - no analytical solution
- Therefore must use EM algorithm
- EM Algorithm is heuristic
  - parameter estimates improve at each iteration
- Limitation: may converge on local maxima

OAC Food Agriculture Communities Environment

1. Expectation (E-)Step
   - Estimate the expectations
2. Maximization (M-)Step
   - Use expectations to analytically update parameter estimates

With updated parameter estimates repeat E-step to calculate new expectations vector, and so on, until convergence criteria are fulfilled

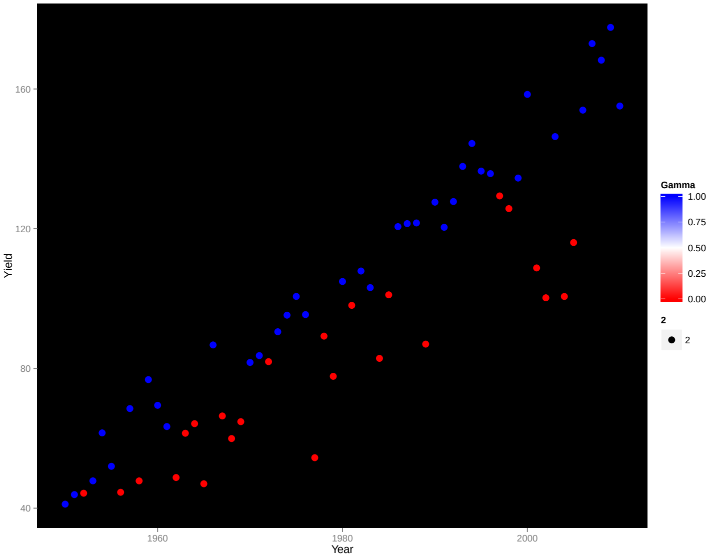
Expectation-Maximization Algorithm

A Closer Look at the E-Step

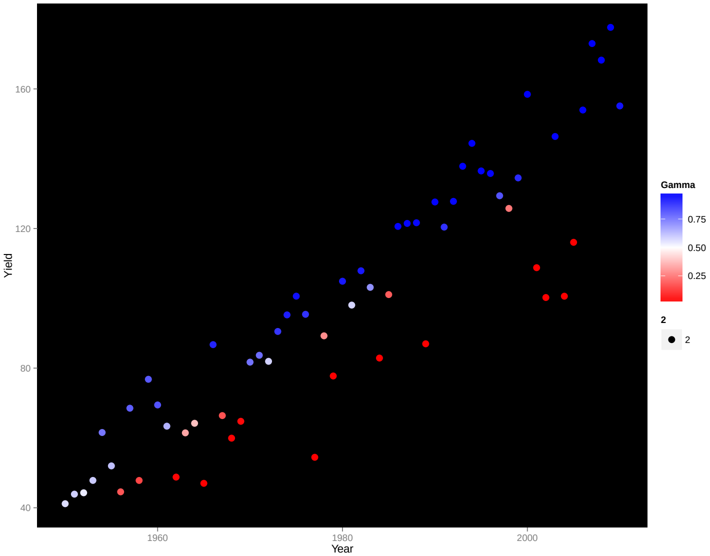
- γ is the vector of expectations calculated in the E-step
- λ is the scalar mean of γ
- Critical mechanism: performs the clustering
- Given current iteration's parameter estimates
  - E-step calculates probability of being in the upper distribution
  - i.e. expectation an observation is a “regular” year
- Therefore “regular” and “catastrophic” years are not chosen but relatively defined estimation parameters of the model (and hence the quotation marks)

OAC Food Agriculture Communities Environment

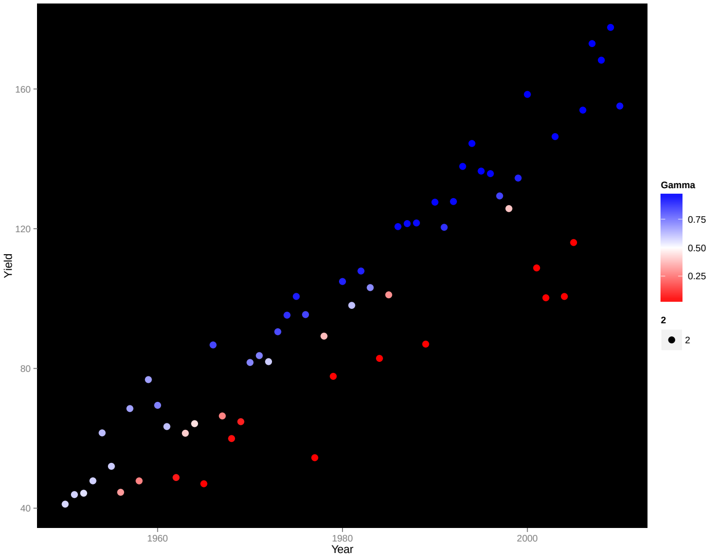
A Closer Look at the E-Step



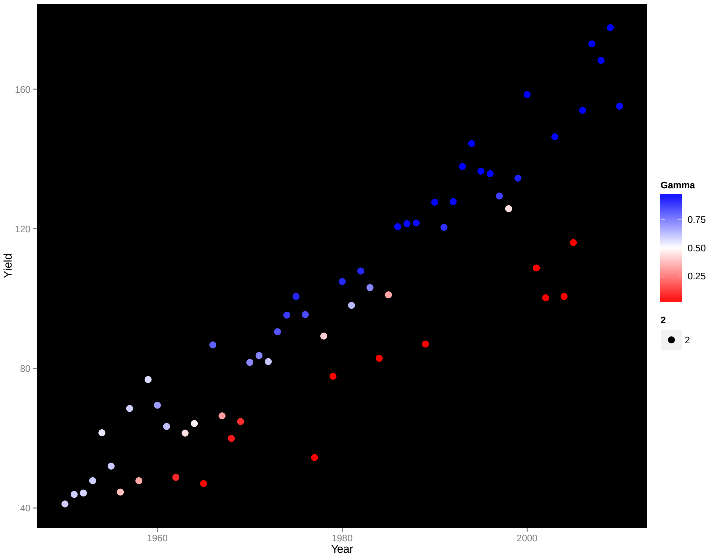
A Closer Look at the E-Step



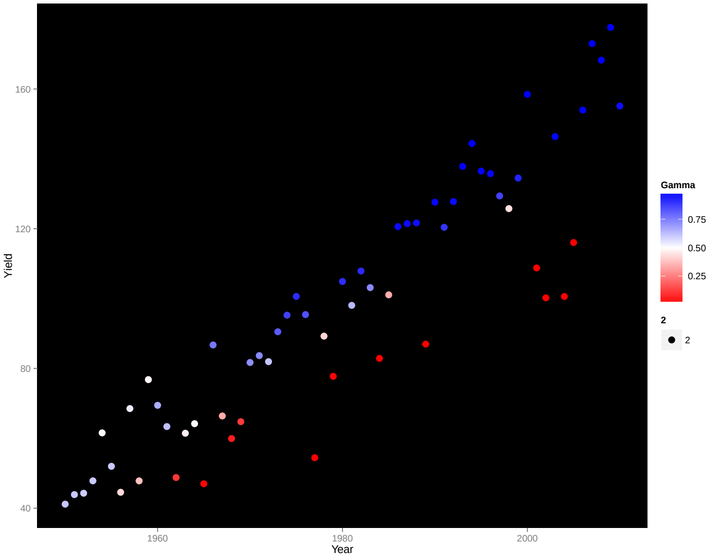
A Closer Look at the E-Step



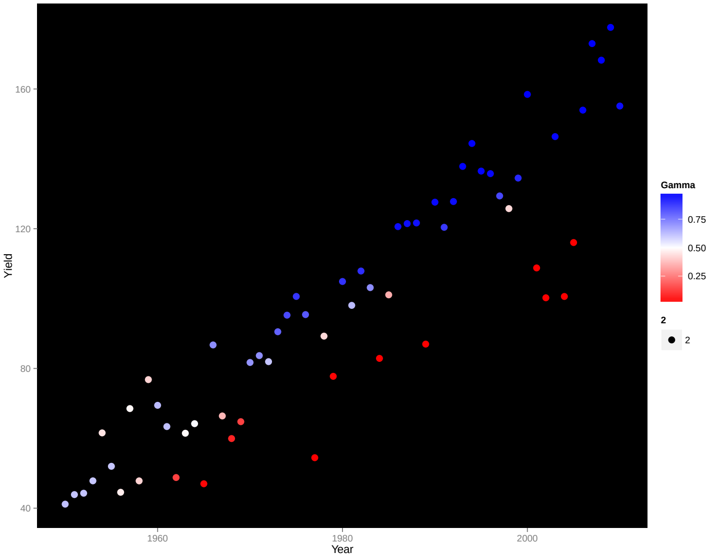
A Closer Look at the E-Step



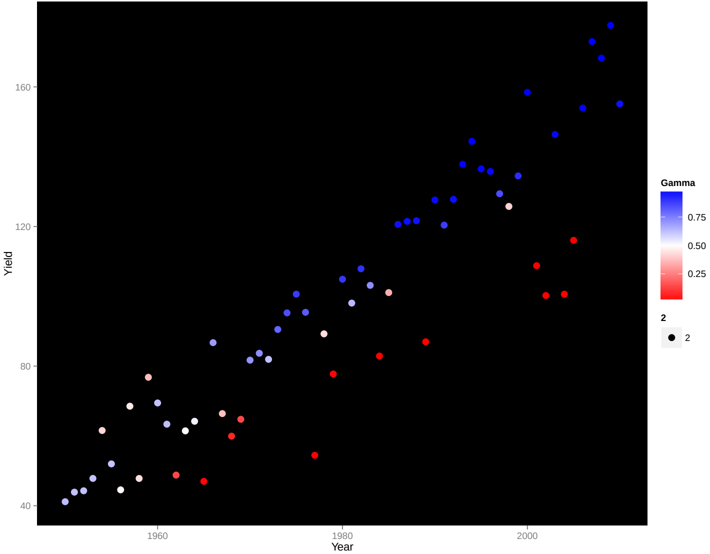
A Closer Look at the E-Step



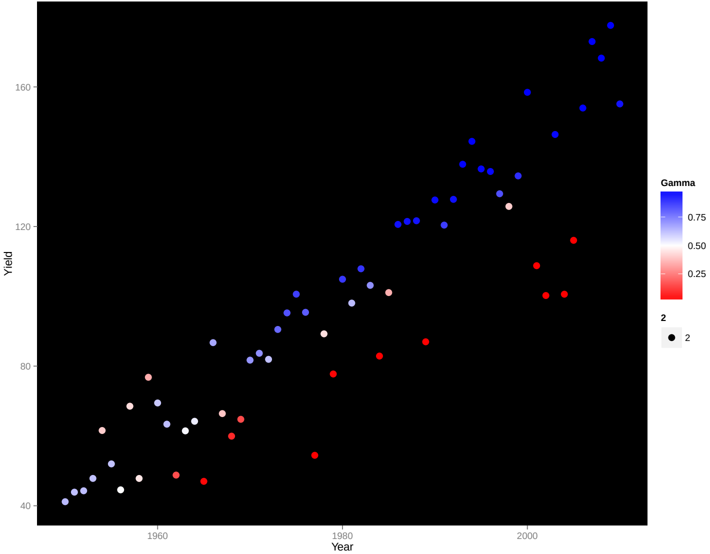
A Closer Look at the E-Step



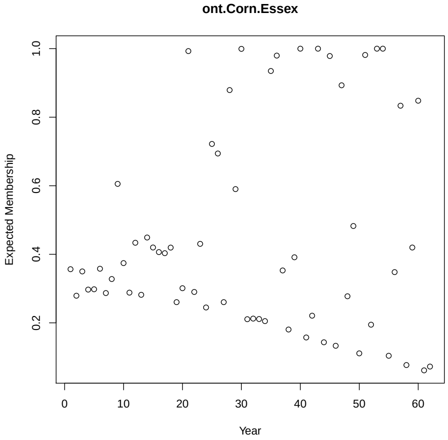
A Closer Look at the E-Step



A Closer Look at the E-Step



Probability of a "Regular" Year Over Time: $\gamma = f(t)$



Recalling Our Empirical Questions

Using the Normal Mixture Model to Test Our Questions:

1. Is technology effecting "regular" and "catastrophic" year yields at the same rate?

$$H_o^1 : \beta_c = \beta_r$$
$$H_a^1 : \beta_c \neq \beta_r$$

2. Has technology effected the probability of a "regular" versus "catastrophic" year?
   - Test to see if time coefficient on $\gamma = f(t)$ is significantly different from zero

Data Overview

- County-level yields for Ontario[2]
- Three most important field crops
- Data collected from annual provincial agriculture reports
- Collection of annual reports stored at OMAFRA
- Long data set: 1949 - 2010
- Excludes Northern Ontario

[2] Technically defined as Census Statistics Division

Ontario County-Level Field Crop Yield Data

Three most important field crops in Ontario

Crop	Value Millions, 2010	% Complete (1949 - 2010)	% Prov. Prod. Total, 2010
Soybeans*	$ 1 243	15.8%	45.5%
Corn	$ 1 602	84.2%	96.3%
Wheat**	$ 340	65.8%	97.8%

* Some notes on next slide

**Barley is the next most important, but is only 11% of winter wheat's acreage and 10% of its value.

- Unfortunately county-level data for soybeans is sparse
- Material production levels outside of southern Ontario weren't attained until post-1985
- Could add eight counties that have complete yield series beginning in the late-1980s
- Would increase total of 2010 production to 73.8%
- To study county-level soybean trends we will have to use US data

Overview of Preliminary Results

- Average $\frac{\beta_r}{\beta_c} = 1.70$
  - Corn 1.74
  - Soybean 1.76
  - Wheat 1.63
- No obvious patterns across regions or crops (surprising)
- $\beta_r \nless \beta_c$ in any case (esp. corn)

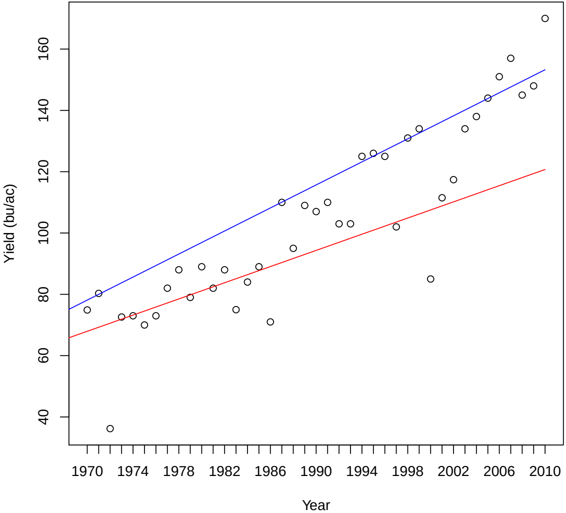
In practically all crop county combinations β_r and β_c are diverging

- **suggests a prevalent heterogeneous trend in the component means**

OAC Food Agriculture Communities Environment

Ottawa Corn

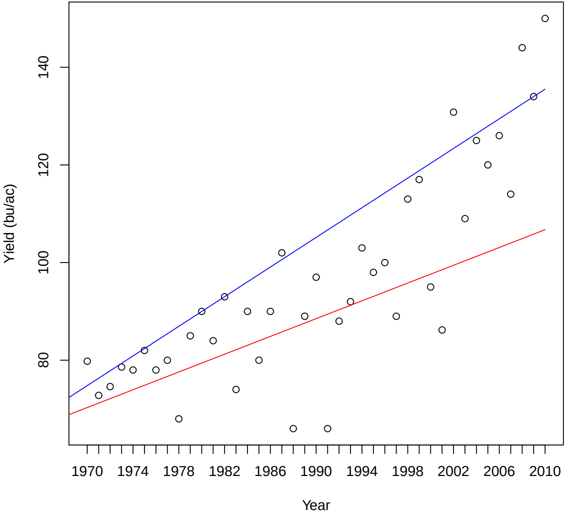
Below Average $\frac{\beta_r}{\beta_c}$



$$\frac{\beta_r}{\beta_c} = 1.42$$

Northumberland Corn

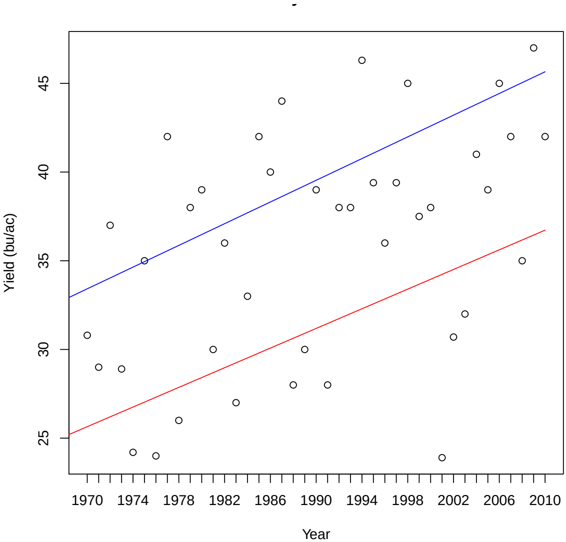
Average $\frac{\beta_r}{\beta_c}$



$$\frac{\beta_r}{\beta_c} = 1.69$$

Essex Soybean

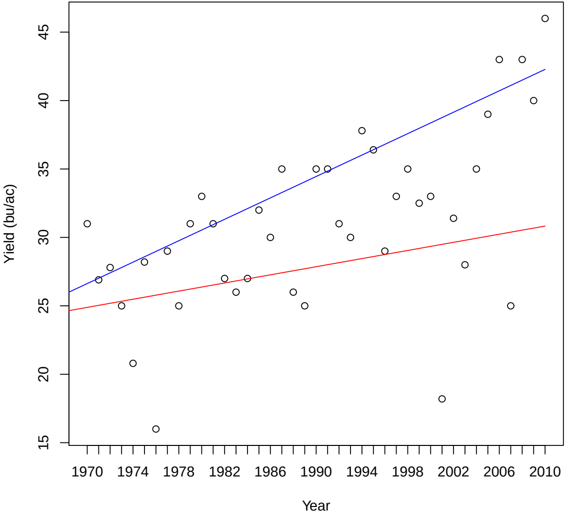
Below Average $\frac{\beta_r}{\beta_c}$



$$\frac{\beta_r}{\beta_c} = 1.10$$

Haldimand-Norfolk Soybean

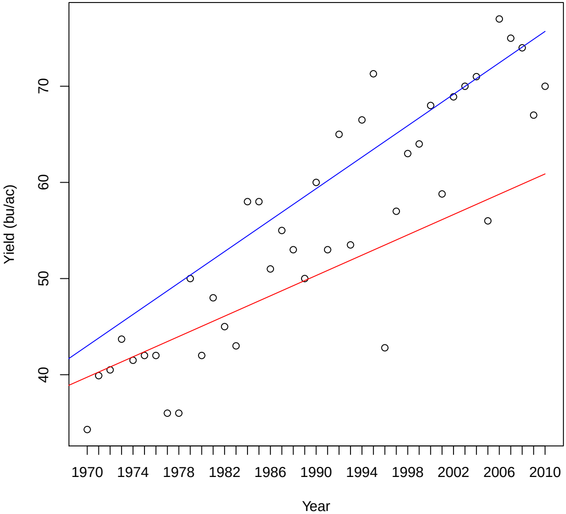
Above Average $\frac{\beta_r}{\beta_c}$



$$\frac{\beta_r}{\beta_c} = 2.62$$

Simcoe Wheat

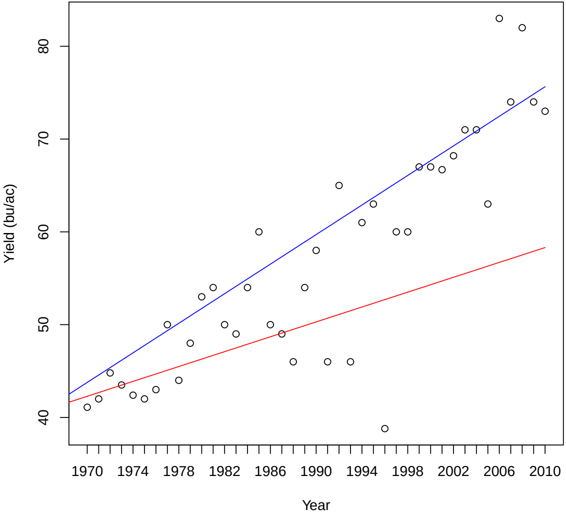
Average $\frac{\beta_r}{\beta_c}$



$$\frac{\beta_r}{\beta_c} = 1.55$$

Waterloo Wheat

Above Average $\frac{\beta_r}{\beta_c}$



$$\frac{\beta_r}{\beta_c} = 2.00$$

Is Trend in Component Means Homogeneous?

Hypothesis Test One

Rejection rate* at 5% significance:

Region	Corn	Soybean	Wheat
Southern	87.5	50.0	77.8
Western	87.5	–	100
Central	100.0	–	80.0
Eastern	75.0	–	–
Total	**87.5**	**50.0**	**84.2**

* *p*-values generated through likelihood-ratio test

Is Probability of “Regular” Year Changing?

Hypothesis Test Two

- Ordinary least squares regression with robust standard errors
  - Two-sided test: reject in 20.6% crop-county combinations at 5% significance level
- Ordinal ranked regression with robust standard errors
  - Nearly identical results to OLS
- Interestingly vast majority of counties that reject are:
  - Wheat counties
  - Negative coefficients
  - Suggests probability of “catastrophic” year is increasing for wheat

Summary of Hypothesis Test Results

Ontario County-Level Yields

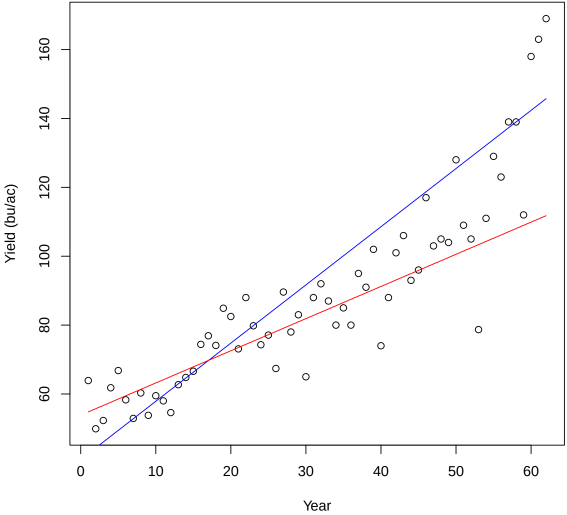
1. The vast majority of county-crop combinations reject the assumption in the literature of homogeneous component trends
2. Fail to reject a significant effect of technology on the probability of "regular" year for majority (> 90%) of corn and soybean counties
3. In constrast, the majority of wheat counties (> 50%) significantly reject and suggest technology has decreased the probability of "regular" year

Caveats

- Results are in early preliminary stages
- Slopes cross in many cases (problem?)

Caveats

Example of Slopes Crossing—Problem?



$$\frac{\beta_r}{\beta_c} = 1.82$$

Policy Implications

- Seems improvement in crop yields has focused on improving yields under ideal conditions
- Suggest yield improvements have been less spectacular under sub-optimal conditions
- Should crop science research focus on robustness of crops to sub-optimal conditions?
  - Climate change: sub-optimal conditions more prevalent?
  - Economics: decent yields under sub-optimal conditions highly profitable (i.e. 2012)

Summary

- Agricultural economics literature has modelled technological change at the mean
- Ignores effect of technology on subpopulations of yields: “regular” versus “catastrophic” years
- Findings suggest a prevalent heterogeneous trend across “regular” and “catastrophic” years
- Findings have implications for models of crop yields (insurance, climate change, agricultural production, etc.) and potentially R&D policy

Thank You

Generous support for this research provided by
The Ontario Ministry of Agricultural, Food and Rural Affairs

Thank You

QUESTIONS AND FEEDBACK?

Additional Slides

- Presume existence of an identity variable $\mathbb{Z}$ that identifies the membership of each observation
  - $\mathbb{Z}_i = 1$ when observation x_i was drawn from the upper cluster and $\mathbb{Z}_i = 0$ otherwise
- If $\mathbb{Z}$ were known log-likelihood could be calculated analytically
- $\mathbb{Z}$ is unknown and latent
- EM approach: estimate $\mathbb{Z}$'s **expectations** (E-step)

Define γ as vector of expectation probabilities corresponding to each observation

$\hat{\gamma}_i$ is the expectation x_i was drawn from the upper cluster

$$x = \begin{pmatrix} 100 \\ 40 \\ 85 \\ 96 \\ 45 \end{pmatrix} \quad \mathbb{Z} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \quad \hat{\gamma} = \begin{pmatrix} 0.95 \\ 0.21 \\ 0.70 \\ 0.85 \\ 0.40 \end{pmatrix}$$

i.e. estimated probability of the true but unobservable $\mathbb{Z}_i$ being one

Expectation-Maximization Algorithm

Expectation Vector and the Probability of a “Good Year”

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_i$$

- λ is the probability of a “good year”
- The probability of a “good year” is the mean of γ
- As the expectation vector changes across algorithm iterations so too does λ

The probability of a “good year” is not fixed across counties or crops but an estimated parameter of the model

Expectation-Maximization Algorithm

A Closer Look at the E-Step

$$\gamma_i = \frac{\lambda N(x_i : \theta_g)}{(1-\lambda)N(x_i : \theta_b) + \lambda N(x_i : \theta_g)}$$

- Critical mechanism: performs the clustering
- Given current iteration's parameter estimates
  - E-step calculates probability of being in the upper ("good") distribution
- Therefore "good" and "bad" years are not chosen but estimated parameters of the model
- They are defined relatively for each data set (hence the quotation marks around "good" and "bad")

Expectation-Maximization Algorithm

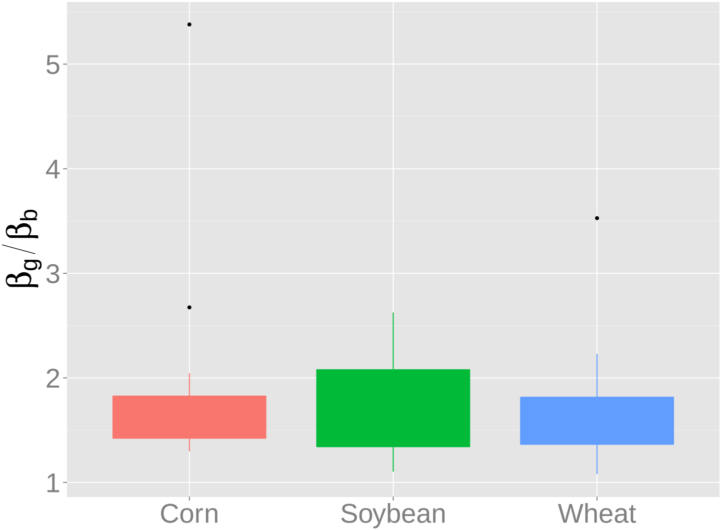
Updating Parameter Estimates: The M-Step

Parameter estimates are updated analytically

- Estimate component mean trend coefficients instead of traditional M-step scalar mean
  - Component mean trend coefficients from weighted least squares
  - Weights are expected memberships

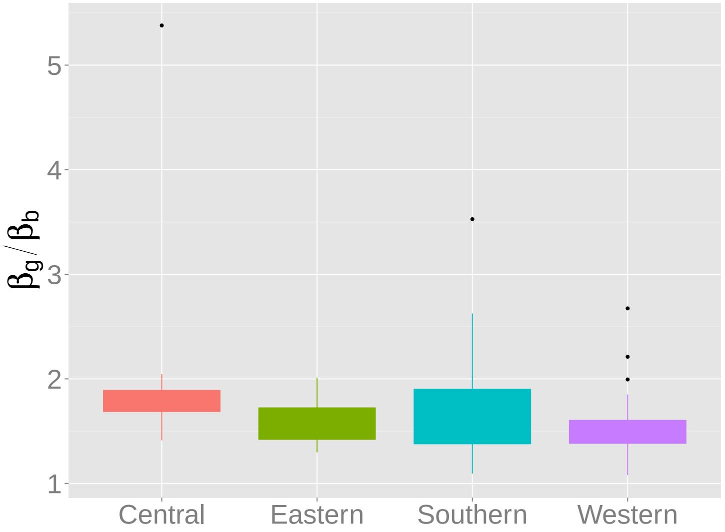
- Scalar component variances as per traditional M-step

Preliminary Estimation Results[3]



[3] 1 = assumption in literature

Preliminary Estimation Results[4]



[4] 1 = assumption in literature

Preliminary Estimation Results

Ontario County-Level Yields

Summary Statistic Breakdown of Slope Ratio $\frac{\beta_g}{\beta_b}$

Corn	Min	Mean	Max	CV
Southern	1.32	1.53	1.93	0.13
Western	1.31	1.65	2.67	0.24
Central	1.67	2.56	5.38	0.62
Eastern	1.30	1.59	2.01	0.16

Summary Statistic Breakdown of Slope Ratio $\frac{\beta_g}{\beta_b}$

Soybeans	Min	Mean	Max	CV
Southern	1.10	1.76	2.63	0.33

Wheat	Min	Mean	Max	CV
Southern	1.09	1.73	3.53	0.42
Western	1.08	1.52	2.21	0.24
Central	1.41	1.68	1.88	0.12