On Technological Change in Crop Yields

Tor Tolhurst & Alan P. Ker

University of Guelph
Department of Food, Agricultural and Resource Economics

Havlicek Memorial Seminar
Department of Agricultural, Environmental, and Development Economics
The Ohio State University
September 9, 2013
Technological Change in Agriculture

Invention, diffusion and adoption of new technology
- Crops and crop varieties
- Tools (irrigation, fertilizer)
- Methods (tillage, rotation)

Impacts of technological change
- Food sustainability
- Economic growth and development
- World hunger
- Energy (i.e. biofuels)
- Mitigation of potential climate change effects
Measuring Technological Change in Agriculture

Chatnam–Kent Corn

Year (bu/ac)

Linear
3-degree Polynomial
2-knot Linear Spline

60 80 100 120 140 160 180

On Technological Change in Crop Yields

Tolhurst & Ker (University of Guelph)
Technologies often target subsets of the yield distribution


Some examples:

1. *Triple-stacked* seeds: increase resilience to pests and high winds (Edgerton *et al.* 2012)

2. *Racehorse* seeds: increase top-end yield under near-optimal conditions (Lauer and Hicks 2005)

Not likely to result in a shift of the mean
Possible that rates of technological change are different across different subsets of the yield distribution

- i.e. top subset could increase faster than bottom subset, or vice versa

Technological change could also change the probability of different subsets occurring

- i.e. ideally top subset would become more likely than bottom subset

Neither would be captured by estimating a trend only at the mean (i.e. a shift of a location in the mean)
An Illustrative Example: Adams County, IL Corn

Estimated Technological Trend

Linear Trend at the Mean

Two Trend Approach
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)

(But all these approaches estimate the trend at the mean)
**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)

(But all these approaches assume distribution constant over time)
Our approach:

- Empirical model to simultaneously estimate “states” and their parameters (which includes rate of technological change)
- States are weakly-identified subsets of the yield distribution
- Then we can:
  1. test for unique rates of technological change across states
  2. test to see if probability of states is constant over time
Estimate rate of technological change in two states ("upper state" and "lower state") using a mixture model.

Test to see if the rates of technological change are different.

Find **78%** of cases have statistically different rates of change.

Test to see if the probability of a state is constant but find inconclusive results (specification challenging).
Model Intuition

- Producer’s output is determined by state of the world
- There are $J \in \mathbb{N}$ finite possible states
- Producer’s output (i.e. yield) determined by a two-step process:
  1. Realized state $j \in J$ is determined by a random i.i.d. draw
  2. Realized yield $y_t \sim \phi(\theta_j | j)$ in another random i.i.d. draw
- Parameters $\theta_j$ may change over time (technological change)

In different states there are potentially different trajectories of technological change
An Illustrative Example: Adams County, IL Corn

Estimated Technological Trend

Linear Trend at the Mean

Two Trend Approach

Tolhurst & Ker (University of Guelph)  On Technological Change in Crop Yields  September 9, 2013 12 / 55
An Illustrative Example: Adams County, IL Corn

Estimated Conditional Yield Densities

**Linear Trend at the Mean**

**Two Trend Approach**
An Economic Application
Estimated Crop Insurance Premium Rates

<table>
<thead>
<tr>
<th>Method</th>
<th>Coverage Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional</td>
<td>75% 6.54%</td>
</tr>
<tr>
<td>Mixture</td>
<td>75% 3.46%</td>
</tr>
</tbody>
</table>

- Traditional rates and two-trend rates are quite different
- Differences of this magnitude would have significant economic consequences for the actuarial soundness of a crop insurance program
Two-Trend Mixture Model

Model to be Estimated:

\[ y_t \sim (1 - \lambda)N(\alpha^\ell + \beta^\ell t, \sigma^2_\ell) + \lambda N(\alpha^u + \beta^u t, \sigma^2_u) \]  

- \( y_t \): observed crop yields \( y \) over time \( t \)
- \( \lambda \): probability of the upper state
- \( \alpha^\ell \): intercept for lower state technological trend
- \( \beta^\ell \): slope for lower state technological trend
- \( \sigma^2_\ell \): lower state homoscedastic component variance
- \( \alpha^u \): intercept for upper state technological trend
- \( \beta^u \): slope for upper state technological trend
- \( \sigma^2_u \): upper state homoscedastic component variance
Normal Mixture Model
Why Choose Two Components?

- Mixture model of two normals is quite flexible and can accommodate:
  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

Brant Wheat

Chatnam-Kent Wheat

Full lines illustrate EM-estimated normal mixture model
Dashed lines illustrate nonparametric kernel density estimate for comparison
Realized yields are a function of adopted technologies \textit{not necessarily} set of possible technologies

\begin{itemize}
\item Therefore our conclusions concern rate of \textit{adopted} rather than \textit{possible} technological change
\end{itemize}

Must use county-level data (farm-level data would be ideal)

\begin{itemize}
\item Relevant for area-yield insurance programs (GRP, GRIP, GRIPH, proposed shallow loss programs)
\end{itemize}
Crop-county Combinations

<table>
<thead>
<tr>
<th>Region</th>
<th>Source</th>
<th>Period</th>
<th>Corn</th>
<th>Soybean</th>
<th>Wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontario</td>
<td>OMAF</td>
<td>1949 - 2010</td>
<td>32</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>Illinois</td>
<td>NASS</td>
<td>1955 - 2011</td>
<td>97</td>
<td>97</td>
<td>-</td>
</tr>
<tr>
<td>Indiana</td>
<td>“”</td>
<td>“”</td>
<td>79</td>
<td>82</td>
<td>-</td>
</tr>
<tr>
<td>Iowa</td>
<td>“”</td>
<td>“”</td>
<td>99</td>
<td>98</td>
<td>-</td>
</tr>
<tr>
<td>Kansas</td>
<td>NASS</td>
<td>1968 - 2011</td>
<td>-</td>
<td>-</td>
<td>93</td>
</tr>
<tr>
<td>Nebraska</td>
<td>“”</td>
<td>“”</td>
<td>-</td>
<td>-</td>
<td>50</td>
</tr>
<tr>
<td>Texas</td>
<td>“”</td>
<td>“”</td>
<td>-</td>
<td>-</td>
<td>96</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>307</td>
<td>283</td>
<td>264</td>
</tr>
</tbody>
</table>

- 854 total crop-county combinations
- All counties with incomplete yield histories excluded
Importance of Selected State-Crops

U.S. Data: Share and Rank of National Production

<table>
<thead>
<tr>
<th>State</th>
<th>Corn Share</th>
<th>Corn Rank</th>
<th>Soybean Share</th>
<th>Soybean Rank</th>
<th>Wheat Share</th>
<th>Wheat Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illinois</td>
<td>15.6%</td>
<td>2</td>
<td>13.7%</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indiana</td>
<td>7.2%</td>
<td>5</td>
<td>7.8%</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iowa</td>
<td>17.3%</td>
<td>1</td>
<td>15.4%</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kansas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24.2%</td>
<td>1</td>
</tr>
<tr>
<td>Nebraska</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.3%</td>
<td>6</td>
</tr>
<tr>
<td>Texas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.6%</td>
<td>2</td>
</tr>
</tbody>
</table>

In corresponding data set’s most recent reporting year.

Ontario: three most important field crops (> $2 billion in 2010)
Two Empirical Questions

**Question**

*Is there a statistically significant difference between the rate of technological change in the upper and lower states?*

**Question**

*Is the probability of being in the upper state stable over time?*
Operationalizing the Model

Maximum likelihood approach to incomplete data
(Dempster, Laird & Rubin, 1977)

- We want to know the parameters of each state
- Assume there exists an identity variable vector $Z$ that identifies the states
- If we knew $Z$ it would be easy to estimate $\theta_j$
- However $Z$ is latent and we only observe yields $y_t$
- Let $\gamma \in [0, 1]$ be a weakly-assigned estimate of $Z$ called the “expectations vector”
- We can use $y_t$ to estimate $\gamma$ and the respective parameters of each state in an iterative algorithm (the EM algorithm)
Expectation-Maximization Algorithm
Overview of the EM Algorithm

- Name from its two steps
  - E Expectation Step
  - M Maximization Step

- Convergence problems of direct likelihood maximization with mixture models and therefore must use EM algorithm

- EM Algorithm is heuristic
  - parameter estimates improve at each iteration

- Limitation: may converge on local maxima
Expectation-Maximization Algorithm
The EM Algorithm

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

- $\gamma$ is the vector of expectations calculated in the E-step
- $\lambda$ is the scalar mean of $\gamma$
- Given current iteration’s parameter estimates E-step calculates probability of being in the upper distribution
- Therefore “lower” and “upper” years are not chosen but relatively defined estimation parameters of the model (and hence the quotation marks)
A Closer Look at the E-Step
A Closer Look at the E-Step

![Graph showing yield over years with gamma values and year scale from 1960 to 2000.]
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

The diagram shows a scatter plot with data points representing crop yields over years. The x-axis represents the year, ranging from 1960 to 2000, and the y-axis represents yield. The data points are color-coded and labeled with gamma values, indicating different technological change phases.
A Closer Look at the E-Step
A Closer Look at the E-Step
Representative Estimate
Clinton IA Corn

![Graph showing yield over time]
Atypical Estimate
Cherokee IA Soybean

Tolhurst & Ker (University of Guelph)  On Technological Change in Crop Yields  September 9, 2013  37 / 55
Atypical Estimate
Hodgeman KS Wheat

![Graph showing yield (bu/ac) over years for Hodgeman KS Wheat. The graph includes two trend lines, one blue and one red, illustrating technological change in crop yields.]
Different Trends in Two States?

Evidence

- 8.6% have faster rate of technological change in *lower state*
- 4.6% have equal rate of technological change in *upper state*
- 81.5% *upper state* trend is greater 110% of *lower state* trend
- 68.4% *upper state* trend is greater 125% of *lower state* trend
- 37.9% *upper state* trend is greater 150% of *lower state* trend
- 16.7% *upper state* trend is **double** the *lower state* trend
Different Trends in Two States?

Hypothesis Test One:

\[ H_0^1 : \beta_\ell = \beta_u \]
\[ H_a^1 : \beta_\ell \neq \beta_u \]

Using likelihood ratio test.

Rejection rates (5% significance level):

<table>
<thead>
<tr>
<th></th>
<th>Corn</th>
<th>Soybean</th>
<th>Wheat</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>84.4%</td>
<td>82.7%</td>
<td>65.5%</td>
<td>78.0%</td>
</tr>
</tbody>
</table>

Majority **reject** equivalent rate of change in two states
Stable Probability of “Belonging to” Upper State?

Hypothesis Test Two:
Where $\delta$ is the estimated slope coefficient from a linear regression of $\gamma = h(t)$

$$H_0^2 : \delta = 0$$
$$H_{a}^2 : \delta \neq 0$$

Using a $t$-test with robust standard errors.

Rejection rates (5% significance level):

<table>
<thead>
<tr>
<th></th>
<th>Corn</th>
<th>Soybean</th>
<th>Wheat</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rejection Rate</td>
<td>10.8%</td>
<td>5.3%</td>
<td>6.0%</td>
<td>7.5%</td>
</tr>
</tbody>
</table>

Majority **fail to reject** stable expectations vector
## Corn Hypothesis Rejection Rates (5% Significance Level)

<table>
<thead>
<tr>
<th>Counties</th>
<th>Number of Counties</th>
<th>One</th>
<th>Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontario</td>
<td>32</td>
<td>87.50%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Illinois</td>
<td>97</td>
<td>81.40%</td>
<td>16.50%</td>
</tr>
<tr>
<td>Indiana</td>
<td>79</td>
<td>84.80%</td>
<td>15.20%</td>
</tr>
<tr>
<td>Iowa</td>
<td>99</td>
<td>85.90%</td>
<td>5.10%</td>
</tr>
<tr>
<td>Sub-total</td>
<td>307</td>
<td>84.36%</td>
<td>10.77%</td>
</tr>
<tr>
<td>Total</td>
<td>854</td>
<td>77.99%</td>
<td>7.49%</td>
</tr>
</tbody>
</table>

Note: $H^1_0$ evaluated using a likelihood ratio test. $H^2_0$ evaluated using a $t$-test with robust standard errors where $\delta$ is the estimated slope coefficient of a linear regression $\hat{\gamma}_i = f(t)$. 
## Soybean Hypothesis Rejection Rates (5% Sign. Level)

<table>
<thead>
<tr>
<th>Number of Counties</th>
<th>Null Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_P = \hat{\beta}_R$</td>
</tr>
<tr>
<td>Ontario</td>
<td>6</td>
</tr>
<tr>
<td>Illinois</td>
<td>97</td>
</tr>
<tr>
<td>Indiana</td>
<td>82</td>
</tr>
<tr>
<td>Iowa</td>
<td>98</td>
</tr>
<tr>
<td>Sub-total</td>
<td>283</td>
</tr>
<tr>
<td>Total</td>
<td>854</td>
</tr>
</tbody>
</table>

Note: $H_0^1$ evaluated using a likelihood ratio test. $H_0^2$ evaluated using a $t$-test with robust standard errors where $\delta$ is the estimated slope coefficient of a linear regression $\hat{\gamma}_i = f(t)$. 
## Wheat Hypothesis Rejection Rates (5% Significance Level)

<table>
<thead>
<tr>
<th>State</th>
<th>Number of Counties</th>
<th>One $\hat{\beta}_P = \hat{\beta}_R$</th>
<th>Two $\delta = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontario</td>
<td>25</td>
<td>80.00%</td>
<td>20.00%</td>
</tr>
<tr>
<td>Kansas</td>
<td>93</td>
<td>79.60%</td>
<td>1.10%</td>
</tr>
<tr>
<td>Nebraska</td>
<td>50</td>
<td>56.00%</td>
<td>8.00%</td>
</tr>
<tr>
<td>Texas</td>
<td>96</td>
<td>53.10%</td>
<td>6.20%</td>
</tr>
<tr>
<td>Sub-total</td>
<td>264</td>
<td>65.53%</td>
<td>6.05%</td>
</tr>
<tr>
<td>Total</td>
<td>854</td>
<td>77.99%</td>
<td>7.49%</td>
</tr>
</tbody>
</table>

Note: $H_o^1$ evaluated using a likelihood ratio test. $H_o^2$ evaluated using a $t$-test with robust standard errors where $\delta$ is the estimated slope coefficient of a linear regression $\hat{\gamma}_i = f(t)$. 
Stable Probability of “Belonging to” Upper State?

Possible determinants of $\delta$? (technology, weather, etc.)

Might be (a) no effect or (b) no net effect
An Application: Crop Insurance Rates

Out-of-Sample Simulated Game

- Common technique in crop insurance literature for comparing two rate-setting techniques (Ker & McGowan 2000; Ker & Coble 2003; Racine & Ker 2006; Harri et al. 2011.)
- Out-of-sample: mimics real life
- If new method has no advantage, will not perform better

Number of “winning” states:

<table>
<thead>
<tr>
<th></th>
<th>Corn</th>
<th>Soybean</th>
<th>Wheat</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/6 (1)</td>
<td>4/6 (0)</td>
<td>5/6 (4)</td>
<td>13/18 (5)</td>
<td></td>
</tr>
</tbody>
</table>

Statistically significant wins at the 5% level in brackets
An Application: Crop Insurance Rates

Some Minor Caveats

- Comparison at lower-end of the density tail only
- Not quite an apples-to-apples comparison
  - Two component trend method does not exhibit heteroscedasticity
- Two trend method seems to perform better
- Rate improvement in 13 of 18 crop-state combinations
- Statistically significant improvement in 5 of 18 states
- No statistically significant wins in opposite direction
- Higher bar to exceed than earlier in-sample likelihood-ratio test - power of the test is weaker
## Out-of-Sample Rating Game Results: Corn

<table>
<thead>
<tr>
<th>Set</th>
<th>Coverage Level</th>
<th>Coverage Retained by Private</th>
<th>Psuedo Loss Ratio Private</th>
<th>Psuedo Loss Ratio Government</th>
<th>p-value</th>
<th>% Payout</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL</td>
<td>75%</td>
<td>85.9%</td>
<td>0.092</td>
<td>0.026</td>
<td>0.787</td>
<td>3.0%</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>87.8%</td>
<td>0.287</td>
<td>0.465</td>
<td>0.012</td>
<td>18.6%</td>
</tr>
<tr>
<td>IN</td>
<td>75%</td>
<td>81.1%</td>
<td>0.164</td>
<td>0.134</td>
<td>0.604</td>
<td>3.7%</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>82.4%</td>
<td>0.395</td>
<td>0.434</td>
<td>0.373</td>
<td>19.8%</td>
</tr>
<tr>
<td>IA</td>
<td>75%</td>
<td>83.9%</td>
<td>0.357</td>
<td>0.409</td>
<td>0.361</td>
<td>6.6%</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>86.8%</td>
<td>0.406</td>
<td>0.444</td>
<td>0.291</td>
<td>12.2%</td>
</tr>
</tbody>
</table>

Note: Winner and $p < 0.05$ or $p < 0.95$
## Out-of-Sample Rating Game Results: Soybean

<table>
<thead>
<tr>
<th>Set</th>
<th>Level</th>
<th>Coverage by Private</th>
<th>Retained</th>
<th>Psuedo Loss Ratio</th>
<th>p-value</th>
<th>% Payout</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Private</td>
<td>Government</td>
<td></td>
</tr>
<tr>
<td>IL</td>
<td>75%</td>
<td>77.7%</td>
<td>0.366</td>
<td>0.374</td>
<td>0.431</td>
<td>3.0%</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>67.4%</td>
<td>0.540</td>
<td>0.474</td>
<td>0.695</td>
<td>12.9%</td>
</tr>
<tr>
<td>IN</td>
<td>75%</td>
<td>86.9%</td>
<td>0.257</td>
<td>0.258</td>
<td>0.479</td>
<td>3.4%</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>78.4%</td>
<td>0.611</td>
<td>0.746</td>
<td>0.201</td>
<td>19.6%</td>
</tr>
<tr>
<td>IA</td>
<td>75%</td>
<td>80.6%</td>
<td>0.809</td>
<td>0.467</td>
<td>0.757</td>
<td>6.1%</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>78.7%</td>
<td>0.751</td>
<td>0.911</td>
<td>0.076</td>
<td>16.1%</td>
</tr>
</tbody>
</table>

Note: Winner and \( p < 0.05 \) or \( p < 0.95 \)
## Out-of-Sample Rating Game Results: Wheat

<table>
<thead>
<tr>
<th>Set</th>
<th>Coverage</th>
<th>Retained by Private</th>
<th>Psuedo Loss Ratio</th>
<th>p-value</th>
<th>% Payout</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Private</td>
<td>Government</td>
<td></td>
</tr>
<tr>
<td>KS</td>
<td>75%</td>
<td>52.6%</td>
<td>1.297</td>
<td>1.921</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>41.9%</td>
<td>1.268</td>
<td>1.399</td>
<td>0.217</td>
</tr>
<tr>
<td>NE</td>
<td>75%</td>
<td>40.8%</td>
<td>0.309</td>
<td>0.747</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>45.3%</td>
<td>0.652</td>
<td>0.611</td>
<td>0.587</td>
</tr>
<tr>
<td>TX</td>
<td>75%</td>
<td>73.5%</td>
<td>0.957</td>
<td>2.096</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>69.1%</td>
<td>1.112</td>
<td>1.593</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note: Winner and \( p < 0.05 \) or \( p < 0.95 \)
Crop yields are important measuring stick of agricultural productivity

Crop science literature suggests yields have distinct subsets of the yield distribution—are the rates of technological change equal across these subsets?

We propose an empirical model that explicitly allows for unique trends in two states
Conclusions

1. Trajectory of technological change
   - different trajectories in different states
   - statistically significant difference in 78% of cases
   - > 85% slower rate of technological change in lower state

2. Inconclusive results w.r.t. stable probability of the upper state

3. Mixture model opens the door for a lot of interesting questions.

4. Results consistent with plant science research expenditures (corn versus beans, racehorse versus suboptimal)

5. Results consistent with stylized facts regarding heteroscedasticity

6. Results suggest little effect of climate change.

7. Results inconsistent with what you would find looking at moments from single technology estimation

8. Results are fairly robust within a crop
Going Forward

- Can we isolate the underlying cause(s) of different trends in the two states?

- How different are the results at the farm-level?
  - Trial farm project in initial stages

- What are the determinants of yields “belonging to” the upper state?
  - What climatic/weather conditions can we isolate as important?
Questions?

Funding for this research was generously provided by the Ontario Ministry of Food and Agriculture (OMAF) and the Institute for the Advanced Study of Food and Agricultural Policy.
Operationalizing the Model

Maximum likelihood approach to incomplete data
(Dempster, Laird & Rubin, 1977)

- We want to know the parameters of each state
- Assume there exists an identity variable vector $Z$ that identifies the states
Operationalizing the Model

Maximum likelihood approach to incomplete data
(Dempster, Laird & Rubin, 1977)

- We want to know the parameters of each state
- Assume there exists an identity variable vector $Z$ that identifies the states
- If we knew $Z$ it would be easy to estimate $\theta_j$
- However $Z$ is latent and we only observe yields $y_t$
Operationalizing the Model

Maximum likelihood approach to incomplete data
(Dempster, Laird & Rubin, 1977)

- We want to know the parameters of each state
- Assume there exists an identity variable vector $Z$ that identifies the states
- If we knew $Z$ it would be easy to estimate $\theta_j$
- However $Z$ is latent and we only observe yields $y_t$
- Let $\gamma \in [0, 1]$ be a weakly-assigned estimate of $Z$ called the “expectations vector”
- We can use $y_t$ to estimate $\gamma$ and the respective parameters of each state in an iterative algorithm (the EM algorithm)
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)

(But all these approaches estimate the trend at the mean)
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)

(But all these approaches assume distribution constant over time)
Hypothesis Rejection Rates (5% Significance Level)

<table>
<thead>
<tr>
<th>Number of Counties</th>
<th>Null Hypothesis One $\hat{\beta}_P = \hat{\beta}_R$</th>
<th>Null Hypothesis Two $\delta = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn</td>
<td>307</td>
<td>84.36%</td>
</tr>
<tr>
<td>Soybean</td>
<td>283</td>
<td>82.70%</td>
</tr>
<tr>
<td>Wheat</td>
<td>264</td>
<td>65.53%</td>
</tr>
<tr>
<td>Total</td>
<td>854</td>
<td>77.99%</td>
</tr>
</tbody>
</table>

Note: $H^1_o$ evaluated using a likelihood ratio test. $H^2_o$ evaluated using a $t$-test with robust standard errors where $\delta$ is the estimated slope coefficient of a linear regression $\hat{\gamma}_i = f(t)$. 
<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>One ( \hat{\beta}_P = \hat{\beta}_R )</th>
<th>Two ( \delta = 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Counties</td>
<td>Ontario 32</td>
<td>Illinois 97</td>
</tr>
<tr>
<td></td>
<td>87.50%</td>
<td>81.40%</td>
</tr>
<tr>
<td></td>
<td>0.00%</td>
<td>16.50%</td>
</tr>
</tbody>
</table>

Note: \( H^1_o \) evaluated using a likelihood ratio test. \( H^2_o \) evaluated using a \( t \)-test with robust standard errors where \( \delta \) is the estimated slope coefficient of a linear regression \( \hat{\gamma}_i = f(t) \).
## Soybean Hypothesis Rejection Rates (5% Sign. Level)

<table>
<thead>
<tr>
<th>Number of Counties</th>
<th>Null Hypothesis</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One</td>
<td>Two</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_P = \hat{\beta}_R$</td>
<td>$\delta = 0$</td>
<td></td>
</tr>
<tr>
<td>Ontario</td>
<td>100.00%</td>
<td>0.00%</td>
<td></td>
</tr>
<tr>
<td>Illinois</td>
<td>83.50%</td>
<td>4.10%</td>
<td></td>
</tr>
<tr>
<td>Indiana</td>
<td>80.50%</td>
<td>4.90%</td>
<td></td>
</tr>
<tr>
<td>Iowa</td>
<td>82.70%</td>
<td>7.10%</td>
<td></td>
</tr>
<tr>
<td>Sub-total</td>
<td>82.70%</td>
<td>5.28%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>77.99%</td>
<td>7.49%</td>
<td></td>
</tr>
</tbody>
</table>

Note: $H_0^1$ evaluated using a likelihood ratio test. $H_0^2$ evaluated using a $t$-test with robust standard errors where $\delta$ is the estimated slope coefficient of a linear regression $\hat{\gamma}_i = f(t)$. 

Tolhurst & Ker (University of Guelph)  
On Technological Change in Crop Yields  
September 9, 2013  
61 / 55
### Wheat Hypothesis Rejection Rates (5% Significance Level)

<table>
<thead>
<tr>
<th>Number of Counties</th>
<th>Null Hypothesis</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One</td>
<td>Two</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_P = \hat{\beta}_R$</td>
<td>$\delta = 0$</td>
<td></td>
</tr>
<tr>
<td>Ontario</td>
<td>25</td>
<td>80.00%</td>
<td>20.00%</td>
</tr>
<tr>
<td>Kansas</td>
<td>93</td>
<td>79.60%</td>
<td>1.10%</td>
</tr>
<tr>
<td>Nebraska</td>
<td>50</td>
<td>56.00%</td>
<td>8.00%</td>
</tr>
<tr>
<td>Texas</td>
<td>96</td>
<td>53.10%</td>
<td>6.20%</td>
</tr>
<tr>
<td>Sub-total</td>
<td>264</td>
<td>65.53%</td>
<td>6.05%</td>
</tr>
<tr>
<td>Total</td>
<td>854</td>
<td>77.99%</td>
<td>7.49%</td>
</tr>
</tbody>
</table>

Note: $H_0^1$ evaluated using a likelihood ratio test. $H_0^2$ evaluated using a t-test with robust standard errors where $\delta$ is the estimated slope coefficient of a linear regression $\hat{\gamma}_i = f(t)$. 

Tolhurst & Ker (University of Guelph) | On Technological Change in Crop Yields | September 9, 2013
Out-of-Sample Simulated Game

- Common technique in crop insurance literature for comparing two rate-setting techniques (Ker & McGowan 2000; Ker & Coble 2003; Racine & Ker 2006; Harri et al. 2011.)
- Out-of-sample: mimics real life
- If new method has no advantage, will not perform better

Number of “winning” states:

<table>
<thead>
<tr>
<th>Crop</th>
<th>4/6 (1)</th>
<th>4/6 (0)</th>
<th>5/6 (4)</th>
<th>13/18 (5)</th>
</tr>
</thead>
</table>

Statistically significant wins at the 5% level in brackets
Some Minor Caveats

- Comparison at lower-end of the density tail
- Does not reflect ability of the model to fit all the data
- *And* not quite an apples-to-apples comparison
  - Two component trend method more constrained in its heteroscedasticity treatment (minor disadvantage)
An Application: Crop Insurance Rates

Some Minor Caveats

- Comparison at lower-end of the density tail
- Does not reflect ability of the model to fit all the data
- *And* not quite an apples-to-apples comparison
  - Two component trend method more constrained in its heteroscedasticity treatment (minor disadvantage)

- Two trend method performs better (despite minor disadvantage)
- Rate improvement in 13 of 18 crop-state combinations
- Statistically significant improvement in 5 of 18 states
- No statistically significant wins in opposite direction
- Higher bar to exceed than earlier in-sample likelihood-ratio test

Tolhurst & Ker (University of Guelph)  On Technological Change in Crop Yields  September 9, 2013
Out-of-Sample Rating Game Results: Corn

<table>
<thead>
<tr>
<th>Set</th>
<th>Coverage</th>
<th>Retained by Private</th>
<th>Psuedo Loss Ratio</th>
<th>p-value</th>
<th>% Payout</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Private</td>
<td>Government</td>
<td></td>
</tr>
<tr>
<td>IL</td>
<td>75%</td>
<td>85.9%</td>
<td>0.092</td>
<td>0.026</td>
<td>0.787</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>87.8%</td>
<td>0.287</td>
<td>0.465</td>
<td><strong>0.012</strong></td>
</tr>
<tr>
<td>IN</td>
<td>75%</td>
<td>81.1%</td>
<td>0.164</td>
<td>0.134</td>
<td>0.604</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>82.4%</td>
<td>0.395</td>
<td>0.434</td>
<td>0.373</td>
</tr>
<tr>
<td>IA</td>
<td>75%</td>
<td>83.9%</td>
<td>0.357</td>
<td>0.409</td>
<td>0.361</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>86.8%</td>
<td>0.406</td>
<td>0.444</td>
<td>0.291</td>
</tr>
</tbody>
</table>

Note: Winner and $p < 0.05$ or $p < 0.95$
### Out-of-Sample Rating Game Results: Soybean

<table>
<thead>
<tr>
<th>Set</th>
<th>Coverage Level</th>
<th>Retained by Private</th>
<th>Psuedo Loss Ratio Private</th>
<th>Psuedo Loss Ratio Government</th>
<th>p-value</th>
<th>% Payout</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL</td>
<td>75%</td>
<td>77.7%</td>
<td>0.366</td>
<td>0.374</td>
<td>0.431</td>
<td>3.0%</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>67.4%</td>
<td>0.540</td>
<td>0.474</td>
<td>0.695</td>
<td>12.9%</td>
</tr>
<tr>
<td>IN</td>
<td>75%</td>
<td>86.9%</td>
<td>0.257</td>
<td>0.258</td>
<td>0.479</td>
<td>3.4%</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>78.4%</td>
<td>0.611</td>
<td>0.746</td>
<td>0.201</td>
<td>19.6%</td>
</tr>
<tr>
<td>IA</td>
<td>75%</td>
<td>80.6%</td>
<td>0.809</td>
<td>0.467</td>
<td>0.757</td>
<td>6.1%</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>78.7%</td>
<td>0.751</td>
<td>0.911</td>
<td>0.076</td>
<td>16.1%</td>
</tr>
</tbody>
</table>

Note: Winner and $p < 0.05$ or $p < 0.95$
Out-of-Sample Rating Game Results: Wheat

<table>
<thead>
<tr>
<th>Set</th>
<th>Coverage</th>
<th>Retained by Private</th>
<th>Psuedo Loss Ratio</th>
<th>p-value</th>
<th>% Payout</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Private</td>
<td>Government</td>
<td></td>
</tr>
<tr>
<td>KS</td>
<td>75%</td>
<td>52.6%</td>
<td>1.297</td>
<td>1.921</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>41.9%</td>
<td>1.268</td>
<td>1.399</td>
<td>0.217</td>
</tr>
<tr>
<td>NE</td>
<td>75%</td>
<td>40.8%</td>
<td>0.309</td>
<td>0.747</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>45.3%</td>
<td>0.652</td>
<td>0.611</td>
<td>0.587</td>
</tr>
<tr>
<td>TX</td>
<td>75%</td>
<td>73.5%</td>
<td>0.957</td>
<td>2.096</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>69.1%</td>
<td>1.112</td>
<td>1.593</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note: Winner and $p < 0.05$ or $p < 0.95$