A Bayesian Hierarchical Model for Combining Several Crop Yield Indications

Nathan B. Cruze

National Agricultural Statistics Service (NASS)
United States Department of Agriculture
nathan.cruze@nass.usda.gov

FCSM 2015
Washington, D.C.
December 1, 2015
Goal and technical approach

- **Goal:** Model sequence of in-season forecasts and estimates of crop yield
  - NASS Crop Production Report–state and national yield estimates
  - Reproducibility with appropriate measures of uncertainty
- **Approach:** Bayesian hierarchical model–synthesis of data from several surveys
  - Enforce physical relationships at two spatial scales
  - Incorporate variety of auxiliary data types

**Challenge:** From data to publication in 3-4 days
NASS crop yield surveys and reports

Yield measures output per area harvested (bushels/acre)

Yield for state $j$: $\mu_j$, $j = 1, 2, \ldots, J$
Yield for speculative region: $\mu = \sum_{j=1}^{J} w_j \mu_j$
Weights $w_j \propto$ harvested acres for state $j$

NASS surveys: Objective Yield (OYS), Agricultural Yield (AYS), Acreage, Production, and Stocks (APS)

Survey and Publication Timeline for Winter Wheat

<table>
<thead>
<tr>
<th>Month</th>
<th>OYS</th>
<th>AYS</th>
<th>OYS</th>
<th>AYS</th>
<th>OYS</th>
<th>APS</th>
</tr>
</thead>
<tbody>
<tr>
<td>May</td>
<td>OYS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jun</td>
<td>OYS</td>
<td>AYS</td>
<td></td>
<td>AYS</td>
<td></td>
<td>APS</td>
</tr>
<tr>
<td>Jul</td>
<td>OYS</td>
<td>AYS</td>
<td></td>
<td>AYS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aug</td>
<td>OYS</td>
<td>AYS</td>
<td></td>
<td>AYS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep</td>
<td>OYS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FCSM 2015–A Bayesian Hierarchical Model for Combining Several Crop Yield Indications
Role of the Agricultural Statistics Board (ASB)

Expert panel of commodity specialists
- Current and historical survey ‘indications’
- Other information, e.g., weather, crop condition ratings
- Consensus on yield

Publish national and state estimates

*OMB Standard 4.1 (2006): “Agencies must use accepted theory and methods when deriving...projections that use survey data. Error estimates must be calculated and disseminated to support assessment of the appropriateness of the uses of the estimates or projections...”*

**Challenge:** Capture expert assessment in a manner that is 1) easily reproducible and 2) includes appropriate measures of uncertainty
Example survey data

NASS Yield Survey Indications: Example Winter Wheat State

- May
- Aug
- June
- Sept
- July

<table>
<thead>
<tr>
<th>Year</th>
<th>Yield (Bu/Acre)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
</tr>
</tbody>
</table>

- AYS
- OYS
- Sept. APS
Bayesian hierarchical model for speculative region

**Notation**

- $\mu_t$–true yield
- $y_{ktm}$–observed yield
- $k \in \{O, A, Q\}$–survey index
- $t \in \{1, \ldots, T\}$–year index
- $m \in \{months\}$–survey month
- $m^*$–forecast month

**Region data model**

\[
y_{ktm^\ast} | \mu_t \sim \text{indep } N \left( \mu_t + b_{km^\ast}, s_{ktm^\ast}^2 + \sigma_{km^\ast}^2 \right), \quad k = O, A \tag{1}
\]

\[
y_{Qt} | \mu_t \sim \text{indep } N \left( \mu_t, s_{Qt}^2 \right) \tag{2}
\]

**Region process model**

\[
\mu_t \sim \text{indep } N \left( z_t' \beta, \sigma_\eta^2 \right) \tag{3}
\]

**Diffuse prior distributions**

- Data model parameters: $\Theta_d \equiv (b_{km^\ast}, \sigma_{km^\ast}^2)$
- Process model parameters: $\Theta_p \equiv (\beta, \sigma_\eta^2)$
Bayesian hierarchical model for speculative region

**Likelihood function**—assuming conditional independence

\[
[y_O, y_A, y_Q | \mu_t, \Theta_d] = \prod_{k \in \{O, A, Q\}} [y_k | \mu_t, \Theta_d] \tag{4}
\]

**Posterior distribution**

\[
[\mu_t, \Theta_d, \Theta_p | y_O, y_A, y_Q] \propto \prod_{k \in \{O, A, Q\}} [y_k | \mu_t, \Theta_d][\mu | \Theta_p][\Theta_d][\Theta_p] \tag{5}
\]

**Full conditional of regional yield, \( \mu_t \)**

\[
[\mu_t | y_O, y_A, y_Q, \Theta_d, \Theta_p] \sim N \left( \frac{\Delta_2}{\Delta_1}, \frac{1}{\Delta_1} \right) \tag{6}
\]

\[
\Delta_1 = \sum_{k=O, A} \frac{1}{\sigma^2_{km^*} + s^2_{kTm^*}} + \frac{l_{\{Q\}}}{s^2_{QT}} + \frac{1}{\sigma^2_\eta} \tag{7}
\]

\[
\Delta_2 = \sum_{k=O, A} \frac{y_{ktm^*} - b_{km^*}}{\sigma^2_{km^*} + s^2_{kTm^*}} + \frac{l_{\{Q\}}y_{Qt}}{s^2_{QT}} + \frac{z_t^t\beta}{\sigma^2_\eta} \tag{8}
\]
Bayesian hierarchical model–state level yield

State-level counterparts indexed by \( j \in \{1, 2, \ldots, J\} \)

**Unconstrained State Model**–Define \( \mu_t \equiv (\mu_{t1}, \mu_{t2}, \ldots, \mu_{tJ}) \),

\[
\mu_t | y, \Theta_d, \Theta_p, \sim \text{indep MVN} \left( \text{vec} \left( \frac{\Delta_{2j}}{\Delta_{1j}} \right), \text{diag} \left( \frac{1}{\Delta_{1j}} \right) \right) \quad (9)
\]

**Constrained State Model**–Enforce constraint by conditioning (9) on \( \mu_t = \sum_j w_j \mu_{tj} \)

\[
(\mu_{t1}, \mu_{t2}, \ldots, \mu_{t(J-1)}) \sim \text{MVN}(\bar{\mu}, \bar{\Sigma}) \quad (10)
\]

\[
\mu_{tJ} = \mu_t - \frac{1}{w_{tJ}} \sum_{j=1}^{J-1} w_{tj} \mu_{tj} \quad (11)
\]
## Summary of model outputs

<table>
<thead>
<tr>
<th>Speculative Region Model</th>
<th>Constrained State Model</th>
<th>Unconstrained State Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region yield and error</td>
<td>Benchmark state yields and errors</td>
<td>State forecast decompositions and benchmarking adjustments</td>
</tr>
<tr>
<td>Region forecast decomposition</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Equation 12

\[
\hat{\mu}_{tj} \approx \sum_{k \in \{O, A, Q, \text{Covariates}\}} c_k(SOURCE)_k + d_j
\]

\[
c_k \propto (\text{variance})_{k}^{-1}
\]
Winter wheat speculative region

- 10 state region—some states geographically isolated
- Kansas has major share of harvested acres (Plotted: $w_j$, 2012)
- Four distinct types of winter wheat
- Differential planting and harvest
## Winter wheat speculative region—types of wheat

### States ‘specialize’
- Soft varieties associated with higher yield
- Washington, Missouri, Illinois, Ohio have higher yields
- Confounding with state

<table>
<thead>
<tr>
<th>State</th>
<th>Percent Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Washington</td>
<td>100</td>
</tr>
<tr>
<td>Montana</td>
<td>75</td>
</tr>
<tr>
<td>Colorado</td>
<td>50</td>
</tr>
<tr>
<td>Nebraska</td>
<td>25</td>
</tr>
<tr>
<td>Kansas</td>
<td>100</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>75</td>
</tr>
<tr>
<td>Texas</td>
<td>50</td>
</tr>
<tr>
<td>Missouri</td>
<td>25</td>
</tr>
<tr>
<td>Illinois</td>
<td>100</td>
</tr>
<tr>
<td>Ohio</td>
<td>75</td>
</tr>
</tbody>
</table>

**Type**
- Red Hard
- Red Soft
- White Hard
- White Soft

---

FCSM 2015–A Bayesian Hierarchical Model for Combining Several Crop Yield Indications
Winter wheat speculative region–differential harvest

- May OYS: only TX, OK, KS
- Southern states complete harvest before northern states begin
- Timing of covariates
- Deriving covariates for the region
Winter wheat model—covariates

Covariates reflect conditions approaching active harvest dates

\[ \mu_{tj} = \beta_{j1} + \beta_{j2}z_{j2} + \beta_{j3}z_{j3} + \beta_{j5}z_{j4} + \beta_{j5}z_{j5} \]

- State-specific constant
- \(z_{j2}\): Linear time trend
- \(z_{j3}\): Monthly precipitation (NOAA)
- \(z_{j4}\): Monthly avg. temperature (NOAA)
- \(z_{j5}\): Crop condition—% good + % excellent week # (NASS)
Comparing ASB estimates and model outputs–2012

Year 2012 Comparisons: Published Yield and Model–based Yield Indications

<table>
<thead>
<tr>
<th>State 1</th>
<th>State 2</th>
<th>State 3</th>
<th>State 4</th>
<th>State 5</th>
<th>State 6</th>
<th>State 7</th>
<th>State 8</th>
<th>State 9</th>
<th>State 10</th>
<th>Region</th>
</tr>
</thead>
</table>

Source
- ASB
- Model
- 95% CI.u
- 95% CI.l

FCSM 2015–A Bayesian Hierarchical Model for Combining Several Crop Yield Indications
Weights applied in wheat forecast decomposition

- Early season emphasis on covariates
- Increasing emphasis on OYS in July
- Heavy emphasis on last AYS in August
- Heavy emphasis on quarterly survey in September
Extensions and conclusions

1. NASS yield models (corn, soybeans, winter wheat) capture expert assessment in a manner which is reproducible and provide justifiable measures of uncertainty.

2. This methodology is flexible enough to accommodate many types of auxiliary data.

- Additional commodities
- Non-spec region states
- New technologies, e.g., soil moisture monitors
Select references


Thank you!
Questions?

nathan.cruze@nass.usda.gov