New developments in modeling hogs production and growth at a finer temporal resolution

Luca Sartore\textsuperscript{1,2}, Yijun Wei\textsuperscript{1,2}, Emilola Abayomi\textsuperscript{2}, Seth Riggins\textsuperscript{2}, Gavin Corral\textsuperscript{2}, Nell Sedransk\textsuperscript{1,2}

\textsuperscript{1}National Institute of Statistical Science (NISS)
\textsuperscript{2}United States Department of Agriculture
National Agricultural Statistics Service (USDA NASS)

lsartore@niss.org

QPRC 2019
Applications of data science in diverse fields
June 12, 2019
The findings and conclusions in this presentation are those of the author(s) and should not be construed to represent any official USDA of U.S. Government determination or policy.
Presentation outline

1. Swine survey at NASS
2. Swine industry characterization
3. Inventory equations
4. Case study
5. Conclusion
**Swine survey at NASS**

NASS conducts quarterly surveys to evaluate status of US swine industry

Reported quantities over previous three months:
- Monthly **pig crop**
- Monthly **sows farrowed**

Reported inventories at the first day of the surveyed quarter:
- **Breeding herd**
- **Market hogs** distinct in weight groups:
  1. Less than 50 lbs
  2. 50-119 lbs
  3. 120-179 lbs
  4. More than 180 lbs
Pros and cons of current modeling approaches

**Constrained Kalman filter model**
- Reliable during periods of equilibrium
- Too rigid during periods of shocks

**Sequential generalized linear models**
- Very flexible
- Ignoring biological constraints
Objectives of the new model

1. Satisfy constraints without forcing them
2. Produce stable estimates during equilibrium periods
3. Flexible formulation to quickly adapt estimates during shocks
4. Predict national inventories at a finer temporal resolution
Swine population dynamics

- **Hogs to market**
  - Group 4
  - Group 3
  - Group 2
  - Group 1

- **End market**
  - Pork

- **Hog Production**
  - Breeding Herd
    - Sows Bred
  - Sows Farrowed
  - Pig Crop

Symbols:
- $y_{1t}$
- $y_{2t}$
- $y_{3t}$
- $y_{4t}$
- $y_{5t}$
- $y_{6t}$
- $y_{7t}$
Breeding dynamics

\[
\begin{align*}
z_{1,t} &= \rho_t z_{2,t} \\
z_{2,t} &= \varphi y_{7,t-2} + \varepsilon_{2,t}
\end{align*}
\]

- $z_{1,t}$ monthly pig-crop at time $t$
- $z_{2,t}$ monthly sows farrowed at time $t$
- $y_{7,t-2}$ size of breeding herd (inventory) at time $t - 2$
- $\rho_t$ piglets per sows at time $t$
- $\varphi$ breeding/farrowing rate
- $\varepsilon_{2,t}$ error in modeling sows farrowed

\[
y_{7,t} = \varphi^{-1} \left( z_{2,t+2} + \varepsilon_{2,t+2} \right)
\]
Time series models

Account for trend, seasonality, and auto-correlations

Dynamics of logarithms of pig crop and sows farrowed are modeled by SARIMA model (Box et al., 2015)

The model choice for $\log(z_{1,t})$ and $\log(z_{2,t})$ is

$$\text{SARIMA}(2, 1, 2) \times (2, 1, 2)_{12}$$

Model fitting via LASSO (Tibshirani, 1996) for automatic selection of a sub-model
Inventory equations

\[
\begin{align*}
\begin{cases}
  y_{3,t} &= \zeta_{t-1} z_{1,t-1} + \zeta_{t-2} z_{1,t-2} + \zeta_{t-3} \alpha_1 z_{1,t-3} + \varepsilon_{3,t}, \\
  y_{4,t} &= \zeta_{t-3} (1 - \alpha_1) z_{1,t-3} + \zeta_{t-4} z_{1,t-4} + \zeta_{t-5} \alpha_2 z_{1,t-5} + \varepsilon_{4,t}, \\
  y_{5,t} &= \zeta_{t-5} (1 - \alpha_2) z_{1,t-5} + \zeta_{t-6} \alpha_3 z_{1,t-6} + \varepsilon_{5,t}, \\
  y_{6,t} &= \zeta_{t-6} (1 - \alpha_3) z_{1,t-6} + \zeta_{t-7} \alpha_4 z_{1,t-7} + \varepsilon_{6,t},
\end{cases}
\end{align*}
\]

- $z_{1,t}$ monthly pig-crop at time $t$
- $y_{2+i,t}$ market hogs, for weight group $i = 1, \ldots, 4$
- $\varepsilon_{2+i,t}$ errors in modeling market hogs
- $\alpha_i$ projection parameters, for $i = 1, \ldots, 4$
- $\zeta_t$ survival rates at time $t$
Inventory equations (cartoon interpretation)

Death rates and allocation percentages are not realistic!

- **Deaths**
- **Survivors to y₃**
- **Survivors to y₄**
- **Survivors to y₅**
- **Survivors to y₆**
- **Slaughtered from survivors**

<table>
<thead>
<tr>
<th>Time</th>
<th>Pig Crop</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>Z₁</td>
</tr>
<tr>
<td>t-1</td>
<td>Y₆ &gt; 180 lbs</td>
</tr>
<tr>
<td>t-2</td>
<td>Y₅ 120-179 lbs</td>
</tr>
<tr>
<td>t-3</td>
<td>Y₄ 50-119 lbs</td>
</tr>
<tr>
<td>t-4</td>
<td>Y₃ &lt; 50 lbs</td>
</tr>
<tr>
<td>t-5</td>
<td></td>
</tr>
<tr>
<td>t-6</td>
<td></td>
</tr>
<tr>
<td>t-7</td>
<td></td>
</tr>
</tbody>
</table>
Case study

Historical estimates from 2008 are combined with survey data up to 2017

The proposed model (SWARCS) and the Kalman filter (KFM) are both used to estimate inventory numbers between 2013 and 2017

The results from the two models are compared against

- Final estimates\(^1\) (NASS USDA, 2019)
- Root mean squared error (RMSE) as selection criterion

\(^1\)https://downloads.usda.library.cornell.edu/usda-esmis/files/jd472w45t/h128nn160/m613n493n/hgpgsb19.pdf
Model comparisons via RMSE

- Pig Crop
- Sows Farrowed
- Mkt >180 lbs
- Mkt 120–179 lbs
- Mkt 50–119 lbs
- Mkt <50 lbs
- Total Mkt Hogs
- Breeding Herd
- Total Hogs

Hogs

<table>
<thead>
<tr>
<th>SWARCS</th>
<th>KFM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>400</td>
<td>600</td>
</tr>
<tr>
<td>800</td>
<td>800</td>
</tr>
<tr>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>1200</td>
<td>1200</td>
</tr>
</tbody>
</table>

QPRC 2019 – Applications of data science in diverse fields – Luca Sartore
Conclusion

For the new model

1. Biological dynamics are satisfied without forcing constraints
2. Stable estimates are produced by a flexible formulation of survival rates
3. Estimates can be produced on a monthly basis at US level

Future work

- Improving breeding herd estimates
- Modeling spatial relationships among states
Selected References I


Selected References II


Thank you!

Questions?

Luca Sartore, PhD  lsartore@niss.org
Yijun Wei  ywei@niss.org
Emilola Abayomi, PhD  emilola.abayomi@usda.gov
Seth Riggins  seth.riggins@usda.gov
Gavin Corral, PhD  gavin.corral@usda.gov
Nell Sedransk, PhD  nsedransk@niss.org