TPO: Current efforts, ideas and plans for annualization, sampling strategies
Participation and response by industry

John Coulston
USDA Forest Service
NCASI Biometrics Working Group
July 14, 2016 Atlanta, GA
Current efforts

• Wrap up legacy effort on national compilation system: led by FIA information management group

• Address action items from National TPO redesign meeting (Oct. 2015)
  • Revisit single Federal survey form.
  • Developing a national mill list.
  • Develop and test candidate sample designs for implementation.
  • Identify larger companies with the capacity and willingness to provide automated updates
  • Develop a marketing strategy for TPO.
  • Continue to develop interagency and external relationships.

• Harvested Wood Products Carbon estimation and projections

• Harvest probability modeling

• RPA land use change and forest dynamics forecasting
Ideas and plans for annualization

• Annual design will be sample-based
• Sample design will be flexible to allow different state-level industry dynamics (e.g. pulp mill vs small hardwood mills)
• Sample design will include a non-response plan
• Design needs to be operationally feasible
• Design will based compatible with small area estimation techniques
• Success of an annual design will be based on industry participation
• Need clear signals from industry partners on willingness to be ‘sampled with certainty’ – perhaps with automated data transfer
Sampling Strategies: A sampling frame is requisite

- A national mill list is needed for an annual sample design
- Mill list must have some measure of size associated with each mill (e.g. capacity)
- Mill list needs to be updated at some frequency
  - Out of date or incomplete mill list = frame error.
  - How much frame error is acceptable? Depends.
- Three key sources are being examined
  - FIA compiled mill list
  - Purchased mill list from University of Georgia
  - Working with US Endowment for Forestry & Communities on cost sharing updates to Wood2Energy mill list.
- Key issue regarding external sources is whether there will be sustained capacity to maintain/update mill lists
Sampling Strategies: Testing candidate sample designs

• Two designs are currently under examination
• In all cases the mills are sampled from the mill list
• Testing done based on southern TPO survey from 2011
• Simple Random Sample.
  • Equal probability
  • Serves as a benchmark
• Tille sampling
  • Unequal probability sampling
  • Requires a measure of size (MOS)
  • The MOS is used to determine inclusion probabilities. Larger mills have a greater inclusion probability (i.e., more likely to be sampled)
• Stratified 2 unit per strata sampling
  • Requires a MOS
  • Develop a large number of equal sized strata based on the MOS
  • Sample 2 units from each stratum
  • Common sampling approach used in Energy Information Administration surveys
## Sampling Strategies: example

<table>
<thead>
<tr>
<th>Measure of Size</th>
<th>Inclusion Probability</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.73</td>
<td>x</td>
</tr>
<tr>
<td>9</td>
<td>0.66</td>
<td>x</td>
</tr>
<tr>
<td>9</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.59</td>
<td>x</td>
</tr>
<tr>
<td>7</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.29</td>
<td>x</td>
</tr>
<tr>
<td>4</td>
<td>0.29</td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td></td>
</tr>
</tbody>
</table>

### Stratified nh=2

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Selection</th>
<th>Inclusion Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x</td>
<td>0.67</td>
</tr>
<tr>
<td>1</td>
<td>x</td>
<td>0.67</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td>2</td>
<td>x</td>
<td>0.40</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.40</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.40</td>
</tr>
<tr>
<td>2</td>
<td>x</td>
<td>0.40</td>
</tr>
<tr>
<td>3</td>
<td>x</td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td>x</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Tille Stratified nh=2
Sample Strategies testing

• Goal: Estimate removals going to products from each State and County in the South
• These are actually domain or small area estimates because not all mills draw from all counties
• Designs
  • SRS region wide (SRS)
  • Tille region wide (Tille)
  • Tille with separate population based on number of employees (Tille-size)
  • Tille with each State considered a separate population (Tille State)
  • Stratified nh=2 with each State considered a separate population (Stratified nh=2)
• Comparisons
  1. Based on 2011 Southern Mill Census (all mills)
  2. Two sampling fractions tested (0.25 and 0.5)
  3. For each design and sampling fraction
     • Draw sample
     • Estimate county and state cuft roundwood production totals
     • Quantify error as the difference between the sample-based estimate and the “true value” from the Census
  4. Replicate #3 1000 times
     • Calculate mean square error for each state and county
     • Calculate percent error 100*√mean square error/true value
Testing Sample Designs

- Example of observed vs estimated total production
- Based on 1 Monte Carlo replication
- 0.5 sampling fraction
Sample Design testing: Monte Carlo results

County Estimates

- Stratified nn=2
- srs
- Tille size
- Tille state

State Estimates

- Stratified nn=2
- srs
- Tille size
- Tille state

$100 \left( \frac{\text{MSE}(\bar{Z}^{0.5})}{Z} \right)$

- max=610%
- max=320%
- max=627%
- max=273%
- max=100%
Sample Strategy testing:

• Goal of Monte Carlo analysis is to understand true MSE
• How does Monte Carlo MSE compare to design-based sampling error for small area estimates?
Sample Strategy testing:

- County level estimate estimator bias.
- Variance estimator become ~ unbiased with groups of 5-10 counties.
Sample Design: Importance of domain estimation

• Mills are sampled by some mechanism but our interest is not only in making a population inference about mills (e.g. mill receipts).
• We are interested in product domains and spatial domains
  • Sawlogs, veneer logs, etc.
  • By state, county
• Estimates for counties would benefit from small area estimation techniques

• Both the Tille approach and the stratified nh=2 approach can be stratified by State and primary mill product.
  • This is helpful for some products
• County-level estimates are small area estimates
• Current effort suggests that the design-based variance of the estimate is underestimated.
• Particular issue when a small number of mills draw a portion of their receipts from a county
Sample Design: small area estimation

• Small area estimation / synthetic estimation techniques are applicable to the TPO design.
• Regression type estimators (blup) that rely on ancillary data are being tested.
• Synthetic estimators that leverage surround area are being tested.
• Ancillary data streams need to be expanded.
  • Remotely sensed ‘current’ harvest area predictions by county
  • Harvest probability models tied to FIA plot data
  • Others
• Some challenges is correctly estimating sampling error
  • Occurs because some number of mills (which are the observation) receive a portion or all of their wood from a particular county (domain of interest)
Sampling Strategy

• Initial results suggest that the stratified approach fits our design criteria
  • Relatively straightforward to implement / operationally feasible
  • Generally as precise as PPS and far superior to SRS
  • Non response can be more easily addressed by collapsing strata rather than recalibrating inclusion probabilities
  • Compatible with small area estimation techniques
Ideal annualized scheduling

- Have updated mill list in Fall
- Develop sample from the frame
- Send out surveys in January
- Surveys returned by May
- Follow-up on non-response Summer
- Load and compile data, construct estimates Fall
- This schedule would provide for estimates being available with a 1 year lag (e.g. 2017 estimates available in 2018)
Participation and Response by Industry

• Crucial for current periodic design and any annual design.

• We are pursuing automated data transfer approaches for larger companies.

• We will work with NCASI and AFPA on encouraging industry to respond.
  • Issue: some companies that have shown support don’t have all their mills responding.

• Need to build relationships with certification groups (e.g. SFI, FSC) to understand their position and opportunities promote response.
Discussion