ALTERNATIVES TO THE MOVING AVERAGE[1]

Paul C. Van Deusen[2]

Abstract—There are many possible estimators that could be used with annual inventory data. The 5-year moving average has been selected as a default estimator to provide initial results for states having available annual inventory data. User objectives for these estimates are discussed. The characteristics of a moving average are outlined. It is shown that moving average characteristics don't always coincide with user objectives. Alternative estimators are proposed that may have more desirable characteristics than the simple moving average.

INTRODUCTION

The Forest Inventory and Analysis (FIA) program of the USDA Forest Service is shifting from doing periodic inventories in each state to implementing an annual forest inventory (USDA Forest Service 1999) where a percentage of the plots are measured each year. The design is intended to provide annual systematic coverage of each state and to generally provide the same amount of information each year. For the most part, the annual inventory uses the same plot system that existed under the periodic design, and one could argue that the annual system merely changes the timing of plot visits. In fact, the annual inventory is a response to changing user needs and therefore represents a major transition for FIA.

A widespread user desire for more timely data is arguably the driving force that led to the 1998 Farm Bill directive for annual inventories. This is documented in 2 Blue Ribbon Panel reports, BRP I and BRP II (American Forest Council 1992, American Forest and Paper Association 1998). BRP I called for shortening the cycle between periodic surveys from 10 to 5 years. This shortened cycle was never achieved and cycles averaged 10 years or more when BRP II convened in 1997. The BRP II call for an annual survey led to the 1998 Farm Bill legislative mandate for annual surveys.

WHAT THE USER WANTS

Most users want timely data and timely estimates. They want estimates that reflect current values in accordance with the current data that an annual inventory provides. It follows that estimates of per acre values are needed for year t, where t can denote any year beginning with annual inventory implementation up through the current year. Likewise, estimates of change between any 2 years should be available. It goes without saying that the user also wants current estimates of area by forest type, but that is a subject for another paper.

These needs can be quantified by letting the implementation year for a state be year 1 and the current year be T. Call the true per acre value $_t$. Therefore the user wants

- Current per acre estimates, $\hat{}_t$, for years $t=1,...,T$.
- The related variance estimate $V(\hat{}_t)$, for $t=1,...,T$.
- Annual change estimates, $\hat{}_t - \hat{}_{t-k}$ for $k=1,...,t-1$
- Estimated variance of change, $V(\hat{}_t - \hat{}_{t-k})$.

WHAT THE MOVING AVERAGE ESTIMATES

The 5-year moving average is equivalent to taking all plot measurements from the last 5 years in a state and averaging them together. For years t-4 through t this can be written as

$$MA_{t-4,t} = \sum_{j=t-4}^{t} w_j \bar{y}_j$$

where $\bar{y}_j$ is the average of all plot values measured in year j, and w_j is a weight such that $\sum w_j = 1$. The plan for the annual inventory is to assign plots to panels and to measure 1 panel per year. Therefore, $\bar{y}_j$ can also be called the panel mean. The weight, w_j, ensures that each panel is weighted according to the proportion of the total plots it contains. With an exact 20 percent sample, w_j =0.2.

The panel mean is unbiased for the true underlying value, $_t$, and we can write

$$\bar{y}_t = {}_t + e_t \qquad (1)$$

where e_t is a random error term. It follows that the expected value of the moving average is

$$E(MA_{t-4,t}) = \sum_{j=t-4}^{t} w_j \; {}_j \qquad (2)$$

Therefore $MA_{t-4,t}$ estimates the true average over the last 5 years and is not an unbiased estimate of the current value, $_t$. This isn't what most users want, but it is similar

[1] Paper presented at the Second Annual Forest Inventory and Analysis (FIA) Symposium, Salt Lake City, UT, October 17–18, 2000.

[2] Principal Research Scientist, National Council for Air and Stream Improvement, 600 Suffolk Street, Lowell, MA 01854.

to what was done under the old periodic design. It often took 3 or more years to complete the full periodic survey in a state, and state estimates were obtained by averaging all plots together as if they had been measured in the same year. This estimate was then assumed to represent the value at the final measurement year. Based on this precedent, one could conclude that the 5-year moving average is "good enough" even though it isn't estimating the current year value.

The variance of the moving average is easy to derive as

$$V(MA_{t-4,t}) = \sum_{j=t-4}^{t} w_j^2 \, V(\bar{y}_j) \tag{3}$$

where $V(\bar{y}_j) = {}_j^2/n_j$, ${}_j^2$ is estimated from the between plot variance within the panel and n_j is the number of plots in the panel measured in year *j*. Therefore, the expected value and the variance of the moving average are well-defined, and both are easy to estimate.

ESTIMATING CHANGE WITH THE MOVING AVERAGE

Change and trend are more important to many FIA users than current status. FIA is committed to producing official state-level reports every 5 years, but users will not wait for 10 years to assess trend. Since the moving average is currently considered to be the default estimator it makes sense to look at the difference between 2 moving average estimates. Suppose we are at year 6 of the annual survey and want an estimate of change since year 5. The difference between the year 6 and year 5 moving average is

$$MA_{2,6} - MA_{1,5} \begin{aligned} &= \frac{1}{5}\left((\bar{y}_2 + \bar{y}_3 + \bar{y}_4 + \bar{y}_5 + \bar{y}_6) - (\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4 + \bar{y}_5)\right) \\ &= \frac{1}{5}\left(\bar{y}_6 - \bar{y}_1\right) \end{aligned} \tag{6}$$

Equation (6) shows that taking the difference of consecutive moving averages cancels out much of the data. Remember that the goal of the annual survey is to cycle through the plots every 5 years, so the plots measured in year 1 will generally be remeasured in year 6. Therefore, equation (6) shows that simple 5-year moving average change estimates give the average of the 5-year change in the current panel. The other 4 panels (80 percent of the plots) measured over the past 5 years are ignored. This is clearly an undesirable situation and less variable estimators could be constructed by using more of the data.

The variance of the moving average change estimator (equation 6) is

$$V(MA_{2,6} - MA_{1,5}) = \frac{1}{25}\left(V(\bar{y}_1) + V(\bar{y}_6) - 2COV(\bar{y}_1, \bar{y}_6)\right)$$

so the covariance between remeasured plots will reduce the overall variance. However, this doesn't justify ignoring 80 percent of the plots.

ALTERNATIVES TO THE MOVING AVERAGE

Alternatives to the moving average are needed because (1) the moving average is not unbiased for current status and (2) moving average annual-change estimates ignore 80 percent of the plots under a 5-panel annual inventory design. There are many alternatives that could be considered. Here the focus will be on mixed-estimation methods (Theil 1971), but multiple imputation (Rubin 1987) and double sampling for regression will also be briefly discussed.

Multiple Imputation

Multiple imputation (Van Deusen 1997, Reams and Van Deusen 1999, Roesch and Reams 1999) uses the intuitively appealing approach of filling in values for unmeasured plots and then applying standard complete data analysis methods. Imputation can be performed by database lookup (hotdeck methods), with regression estimates, or with more elaborate modeling efforts. Single imputation is a special case where only one possible value is imputed for each missing value. Single imputation usually requires complex procedures to properly estimate variance. This makes it tempting to treat imputed values as if they are real which will lead to under-estimating the variance. Multiple imputation requires the imputer to incorporate variability into the imputations, which leads to a simplified variance estimation process for the analyst.

Multiple imputation can work for variables that are difficult to model but are amenable to database lookup. Examples of such variables include: number of snags, Red-cockaded Woodpecker (*Picoides borealis*) nests, or disturbance status. A disadvantage is that multiple datasets must be stored (say m) and each analysis must be repeated m times. Typical users might find this confusing, so multiply imputed datasets will probably not become an official FIA product in the near future.

Double Sampling for Regression

Double sampling for regression (DSR) can be viewed as a single imputation procedure. Intuitively, single imputation methods should place different weights on imputed values and real data. DSR (Cochran 1977, Fairweather and Turner 1983, and Hansen 1990) does this by incorporating predictions via the following regression equation

$$\hat{}_t = \bar{y}_t + a\left(\bar{X} - \bar{x}\right)$$

where $\bar{y}_t$ is the mean from the year t panel, a is a regression coefficient, $\bar{X}$ represents concomitant information from the 4 panels not measured in year t, and $\bar{x}$ is concomitant information from the year t panel. Usually one refers to a large and a small sample with DSR, where the small sample includes the hard-to-measure variable, y, and the easy-to-measure variable, x. Only the easy-to-measure variable is measured in the large sample. For the annual inventory application, the current panel is the small sample and the other 4 panels constitute the large sample.

A necessary assumption is that the large and small samples represent the same population, and therefore $E(\bar{X}) = E(\bar{x})$. However, this may not be true with the annual inventory. The values for $\bar{x}$ from the current panel must come from measurements made 5 years earlier, whereas $\bar{X}$ comes from measurements made 4, 3, 2, and 1 years earlier on the other 4 panels. Therefore, there are systematic differences between the small and large sample x's, and it is likely that $E(\bar{X}) \neq E(\bar{x})$. Regardless, double sampling for regression could be used as a single imputation technique, but some validation studies should be conducted first. Also, variance of DSR change estimates would be difficult to derive such that auto correlation is correctly handled.

Mixed Estimation Methods

Mixed estimation (Theil 1971) offers a flexible time series approach that leads to model-unbiased estimates of current status, change estimates over any time interval, and variance estimates. There are numerous variations that can be considered (Van Deusen 1996, 1999) and a subset of the possibilities is presented here.

Generally, a mixed estimator is defined by an observation equation and a transition equation, where the transition equation is analogous to the Bayesian prior distribution. Although mixed estimation has a Bayesian flavor, it is a cross between Bayesian and frequentist approaches. The observation equation used here is

$$\bar{y}_t = {}_t + e_t \tag{7}$$

where e_t is an independent random error with mean 0 and variance σ_t^2 / n_t. Consider the following three transition equations

$$ {}_t - {}_{t-1} = v_t \tag{8a}$$

$$ {}_t - 2\ {}_{t-1} + {}_{t-2} = v_t \tag{8b}$$

$$ {}_t - 3\ {}_{t-1} + 3\ {}_{t-2} - {}_{t-3} = v_t \tag{8c}$$

where v_t is an independent random error with variance $p\sigma_t^2/n_t$ and p is a parameter that is estimated from the data. As p gets larger, the influence of the transition equation diminishes and the mixed estimator approaches the mean for each panel. Each transition equation leads to a mixed estimator with somewhat different characteristics. Likewise, each equation represents a different prior assumption about how ${}_t$ is related to ${}_{t-1}$. The transition equations (8a-c) constrain the first, second, and third differences of the s and lead to progressively smoother estimates of trend. The transition equations also state that past values give an indication of current values. This seems eminently plausible, since the forest won't change much from 1 year to the next, barring catastrophe. Transition equation (8b) represents an intermediate smoothness constraint and would make a reasonable choice for FIA purposes.

The estimation process is best described using matrix notation. It follows that there is no particular reason to use only the most recent 5 years of data. The equations stay the same regardless of how much data are used, and the estimates will usually improve with more data. The matrix estimation equations for years 1 through T are

$$\hat{} = \left[\Sigma^{-1} + \frac{1}{p}R'\ {}^{-1}R\right]^{-1}\Sigma^{-1}\bar{Y} \tag{9a),}$$

and

$$V(\hat{}) = \left[\Sigma^{-1} + \frac{1}{p}R'\ {}^{-1}R\right]^{-1} \tag{9b}$$

where $\hat{} = \left[{}_{1,}\ {}_{2},\ldots,{}_{T}\right]$, $\Sigma^{-1} = DIAG\left({}^2_1/n_{1,}\ {}^2_2/n_2,\ldots,{}^2_T/n_T\right)$, R depends on the selected transition equation, contains transition variances on the diagonal, and $\bar{Y} = [\bar{y}_{1,}\bar{y}_2,\ldots,\bar{y}_T]$. More details can be found in Van Deusen (1999).

The important point here is that relatively simple equations are available to estimate current status and its variance. Trend estimates are also available, since equation (9a) provides estimates of ${}_1$ through ${}_T$. To estimate the the change from time t–k to t, simply use $\hat{}_t - \hat{}_{t-k}$. A complete covariance-matrix is available from (9b), which makes it possible to estimate the variance of change. Thus, mixed estimators provide very general capabilities for estimating status and trend.

SUMMARY

FIA is replacing the periodic inventory with an annual inventory. Even though data will be acquired annually, FIA plans to produce official estimates for each state every 5 years using a 5-year moving average. The 5-year moving average has been selected as the default estimator for the annual survey, in part because it seems easy to understand and compute. Users want FIA procedures that are statistically valid, not unnecessarily complicated, and that meet their needs. The MA is statistically valid and easy to implement, but it doesn't fully meet user needs. In particular, the MA does not estimate current status at time t. Regardless, it is similar to what was done under the old periodic design and might be an adequate approximation of current status.

Users also want estimates of trend between any 2 years, say t and t–k. In particular, they should be able to obtain change estimates between the current year and the previous year. It was shown in equation (6) that the difference between consecutive 5-year moving averages gives an estimate of the average annual growth over the last 5 years using only 20 percent of the plots. Therefore, this cannot be the best trend estimator available.

Mixed estimators were discussed that do provide estimates of current status and trend between any 2 years. These estimators are more complex than the MA, but will give users a wider array of estimates. For a few years following annual inventory implementation, the moving average may be sufficient. In the long run, FIA should give serious thought to finding alternatives to the MA.

REFERENCES

American Forest Council. 1992. The report of the first blue ribbon panel on forest inventory and analysis. Washington, DC.

American Forest and Paper Association. 1998. The report of the second blue ribbon panel on forest inventory and analysis. Washington, DC.

Reams, G.A.; Van Deusen, P.C. 1999. The southern annual forest inventory system. Journal of Agricultural Biological & Environmental Statistics. 4(4): 346–360.

Roesch, F.A.; Reams, G.A. 1999. Analytical alternatives for an annual inventory system. Journal of Forestry. 12: 33–37.

Rubin, D.B. 1987. Multiple imputation for nonresponse in surveys. Wiley.

USDA Forest Service. 1999. A strategic plan for forest inventory and monitoring. Washington, DC.

Van Deusen, P.C. 1997. Annual forest inventory statistical concepts with emphasis on multiple imputation. Canadian Journal of Forest Research. 27: 379–384.

Van Deusen, P.C. 1997. Alternative sampling designs and estimators for annual surveys. In: Hansen, M.; Fairweather, S., eds. Proceedings of international conference on the inventory and monitoring of forested ecosystems. Boise, ID; 1998 August; USDA Forest Service General Technical Report.