MEMORANDUM | September 1, 2015

TO Craig O'Connor, NOAA
FROM Joseph Herriges
SUBJECT D4 - Precision Estimation

The purpose of this memo is to describe the approach used in computing estimates of the precision of the damages from the shoreline valuation model and the boating model. The structure of the models is the same, but the shoreline valuation covers two periods and two regions and the boating covers only one period and one region, the North Gulf. Hence the analysis for the shoreline model accounts for inter-period covariances and is more complex. The estimation of variances for the shoreline data is explained in detail. The boating model is a simplification of the shoreline model. These simplifications will be noted following the shoreline analysis.

VARIANCE ESTIMATES FOR SHORELINE DAMAGES

In this section, we describe, for two time periods of interest, the variances associated with (a) the estimated value per lost user day \( E_t \) in period \( t \) and (b) the estimated loss in period \( t \) \( L_t = V_t \cdot \Delta_t \), where \( \Delta_t \) denotes the estimated lost user days for period \( t \). The two time periods of interest are June 2010 to January 2011 (\( t = 1 \)) and February 2011 to November 2011 (\( t = 2 \)). We also compute the variance associated with the total loss, where the total loss is defined as

\[
L_T = V_1 \cdot \Delta_1 + V_2 \cdot \Delta_2.
\]

The final section of this memo provides the resulting estimates.

Three sources of uncertainty are taken into account in these calculations:

1. Uncertainty in the estimated lost user days for period \( t \) based upon the infield counts;
2. Uncertainty due to sampling variability in the valuation survey;
3. Uncertainty due to the imputation of income for those individuals in the valuation survey who did not provide household income or who provided only income bounds.\(^1\)

The first source of uncertainty is summarized by \( \sigma_{\Delta_t}^2 = Var[\Delta_t] \) for \( t = 1, 2 \) and \( \sigma_{\Delta_12} = Cov[\Delta_1, \Delta_2] \), estimates of which are constructed using a jackknife procedure.\(^2\)\(^3\) The other two sources of uncertainty both affect the estimated value per lost user day \( (V_t) \). The next section outlines how these sources of uncertainty are accounted

---

\(^1\) See Technical Memo E2 - Travel Cost Computation for a description of the income imputation process.
\(^3\) Throughout this memo, we use \( s^2 \) to denote variance terms. Sampling statisticians typically use \( s^2 \) rather than \( \sigma^2 \), emphasizing a difference in the finite population context.
for in our estimates of $\sigma^2_{V_t} = Var[V_t]$ for $t = 1, 2$ and $\sigma_{V12} = Cov[V_1, V_2]$. We then describe how these components are used to construct $\sigma^2_{L_t} = Var[L_t] (t = 1, 2)$. Finally, we outline the procedure for computing $\sigma^2_{L_T} = Var[L_T]$.

**ESTIMATING $\sigma^2_{V_t}$ AND $\sigma_{V12}$**

The overall variance $\sigma^2_{V_t}$ is constructed using a combination of a jackknife variance estimation procedure [2] and Rubin’s multiple imputation method with ignorable non-response ([1], [3]). The key inputs to this process are:

- A total of $R = 120$ sets of monthly replicate weights ($W_r, r = 1, \ldots, R$) and
- A total of $S = 5$ income imputations ($I_s, s = 1, \ldots, 5$).

Each set of monthly replicate weights is randomly paired with one of the five income imputations. Let $n_s$ denote the number of replicate weight sets assigned income imputations $s$, with $\sum_s n_s = R$, and $s_r$ denote the income imputation assigned to replicate weight set $r$. For each of these pairings $(r, s_r)$, the recreation demand model is re-estimated using the associated replicate monthly weights ($W_r$) in constructing trips and the associated incomes ($I_s$) in forming travel costs. The resulting estimated model is then used to form the estimated value per lost user day ($\hat{P}^{T,r}$) for pairing $(r, s_r)$. Let

$$\hat{V}_t^k = \frac{1}{n_k} \sum_{r=1}^{R} \delta_{s_r,k} \hat{V}_t^{r,s_r}$$

(2)

de note the mean value per lost user day across estimates using income imputation $k$, where $\delta_{s,k} \equiv 1(s = k)$.

The estimated variance within imputation $k$ for period $t$ is given by:

$$\Omega_{tt}^k = \frac{(R - 1)}{n_k} \sum_{r=1}^{R} \delta_{s_r,k} (\hat{V}_t^{r,s_r} - \hat{V}_t^k)^2.$$  

(3)

The estimated covariance between $V_1$ and $V_2$ within imputation $k$ is similarly given by:

$$\Omega_{12}^k = \frac{(R - 1)}{n_k} \sum_{r=1}^{R} \delta_{s_r,k} (\hat{V}_1^{r,s_r} - \hat{V}_1^k)(\hat{V}_2^{r,s_r} - \hat{V}_2^k).$$

(4)

The average within imputation variance for period $t$ is given by:

$$U_t = \frac{1}{S} \sum_{s=1}^{S} \Omega_{tt}^s.$$  

(5)

The between imputation variance for period $t$ is given by:

$$B_t = \frac{1}{S - 1} \sum_{s=1}^{S} (\hat{V}_t^s - V_t)^2,$$

(6)

\[\text{See Rusl (1985).}\]
where
\[ \bar{V}_t = \frac{1}{S} \sum_{s=1}^{S} \bar{V}_t^s. \] (7)

The overall variance in the value per lost user day is then given by:
\[ \sigma^2_{V_t} = U_t + \left(1 + \frac{1}{S}\right) B_t \quad t = 1, 2. \] (8)

**ESTIMATING \( \sigma^2_{U_t} \)**
Treating the uncertainty stemming from the counts as independent of the uncertainty stemming from the valuation survey (i.e., sources 2 and 3 in the introduction to this memo), we can compute the variance for the overall loss in period \( t \) as:
\[ \sigma^2_{L_t} = \sigma^2_{V_t} \cdot \sigma^2_{\Delta_t} + \sigma^2_{V_t} \cdot \Delta_t^2 + \sigma^2_{\Delta_t} \cdot V_t^2. \] (9)

The assumption that \( V_t \) and \( \Delta_t \) are independent ignores the fact that the counts are used to calibrate the model to reflect the percentage change in user days by zone as a result of the spill, which is in turn used to compute the value per lost user day. However, the assumption of independence is a reasonable approximation given that sensitivity analysis indicates that the value per lost user day is relatively insensitive to the percentage change in trips used in the calibration.\(^5\)

**ESTIMATING \( \sigma^2_{U_r} \)**
Computing the variance of the overall loss (\( L_T \)) is more complicated as there are two sources of correlation across the elements that make up \( L_T \). Specifically, our estimates of \( V_1 \) and \( V_2 \) are correlated since they are based on the same underlying parameter estimates. In addition, our estimates of \( \Delta_1 \) and \( \Delta_2 \) are correlated because they draw on some of the same infield counts in forming the baseline conditions in periods 1 and 2 for the North Gulf. Specifically, we have that:
\[ \Delta_1 = (y_{10}^N - y_{11}^N) + (y_{10}^F - y_{11}^F) \] (10)
\[ \Delta_2 = (y_{20}^N - y_{21}^N), \] (11)

where \( y_r^c \) denotes trips in region \( r \) (\( = N \) for North Gulf and \( = F \) for the Florida Peninsula), time period \( t \) (\( = 1, 2 \)), and case \( c \) (\( = 0 \) for baseline and \( = 1 \) for spill conditions). It is assumed that only \( y_{10}^N \) and \( y_{20}^N \) are correlated, so that:
\[ \sigma_{\Delta 12} = Cov(\Delta_1, \Delta_2) = Cov(y_{10}^N, y_{20}^N), \] (12)

an estimate of which is provided by the jacknife procedure described in Technical Memo B1 – Estimation Procedures for Count Data.

\(^5\) See Technical Memo E9--- Model Sensitivities.
Given these assumptions, the variance of $L_T$ in general takes the form:

$$\sigma^2_{L_T} = Var(L_T) = Var(L_1) + Var(L_2) + 2 \cdot Cov(L_1, L_2). \quad (13)$$

However, it is straightforward to show that:

$$Cov(L_1, L_2) = E(L_1 L_2) - E(L_1)E(L_2)$$
$$= E(V_1 \Delta_1 V_2 \Delta_2) - E(V_1)E(\Delta_1)E(V_2)E(\Delta_2)$$
$$= Cov(V_1, V_2) + E(V_1)E(V_2) \left[ Cov(\Delta_1, \Delta_2) + E(\Delta_1)E(\Delta_2) \right]$$
$$+ E(V_1)E(\Delta_1)E(V_2)E(\Delta_2)$$
$$= Cov(V_1, V_2)Cov(\Delta_1, \Delta_2) + Cov(V_1, V_2)E(\Delta_1)E(\Delta_2)$$
$$+ Cov(\Delta_1, \Delta_2)E(V_1)E(V_2) \quad (14)$$

The estimate of $\sigma^2_{L_T}$ proceeds as follows. Using equations (9), (13) and (14), the variance of the total loss using income imputation $s$ becomes

$$\Omega^s_{TT} = \left[ \Omega^s_{11} \cdot \sigma^2_{\Delta_1} + \Omega^s_{11} \cdot \Delta_1^2 + \sigma^2_{\Delta_1} \cdot \left( \bar{V}_1^s \right)^2 \right] + \left[ \Omega^s_{22} \cdot \sigma^2_{\Delta_2} + \Omega^s_{22} \cdot \Delta_2^2 + \sigma^2_{\Delta_2} \cdot \left( \bar{V}_2^s \right)^2 \right]$$
$$+ 2 \left[ \Omega^s_{12} \sigma_{\Delta_1 \Delta_2} + \Omega^s_{12} \Delta_1 \Delta_2 + \sigma_{\Delta_1 \Delta_2} \bar{V}_1^s \bar{V}_2^s \right] \quad (15)$$

The average within imputation variance for the total loss is given by:

$$U_T = \frac{1}{S} \sum_{s=1}^{S} \Omega^s_{TT}. \quad (16)$$

The between imputation variance for the total loss is given by:

$$B_T = \frac{1}{S-1} \sum_{s=1}^{S} (\bar{L}_T^s - \bar{L}_T)^2, \quad (17)$$

where

$$\bar{L}_T^k = \frac{1}{n_k} \sum_{r=1}^{R} \delta_{s,k} \bar{L}_{ir}^r$$

(18)

denote the mean total loss across estimates using income imputation $k$ and

---

1. The third equality follows from the assumption that the $V'$s are independent of the $\Delta'$s.
\[ L_T = \frac{1}{S} \sum_{s=1}^{S} \tilde{L}_s. \] 

\[ \sigma^2_{LT} = U_T + \left(1 + \frac{1}{S}\right) B_T. \] 

**SENSITIVITY ANALYSIS**

As a check on the importance of the independence assumption, we propose the following alternative bootstrap estimator for \( \sigma^2_{LT} (t = 1, 2) \) and \( \sigma^2_{LT} \) as a sensitivity. Three steps are involved in the procedure.

**STEP 1: BOOTSTRAP COUNT COMPONENTS**

The infield counts provide an estimate for a given time period \((\ell)\) of the user days by zone \((z = N\) for the North Gulf, \(z = F\) for the Florida Peninsula\) under baseline and spill conditions (denoted by \(y^B_{10}\) and \(y^F_{11}\), respectively). The estimated lost user days \( \Delta_t \) then corresponds to

\[ \Delta_t = \begin{cases} (y^N_{10} - y^N_{11}) + (y^F_{10} - y^F_{11}) & t = 1 \\ (y^N_{20} - y^N_{21}) & t = 2. \end{cases} \] 

\[ \rho^z = \frac{y^0_{10} - y^1_{11}}{y^0_{10}}. \] 

For \( b = 1, \ldots, B \), draw \( y^b \sim N (y, \Sigma_y) \), where \( y^b \equiv (y^{N^b}_{10}, y^{F^b}_{10}, y^{N^b}_{11}, y^{F^b}_{11})' \)

\[ y \equiv (y^N_{10}, y^F_{10}, y^N_{11}, y^F_{11})' \), and

\[ \Sigma_y = \begin{bmatrix} \sigma_{y_{11,0}} & \sigma_{y_{11,2}} & 0 & 0 & 0 & 0 \\ \sigma_{y_{12,0}} & \sigma_{y_{12,2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{y_{11,0}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{y_{11,1}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{y_{22,1}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{y_{11,1}} \end{bmatrix}. \]
where $\sigma^2_{y_{12,c}} = Var(y_{12,c}^t)$ and $\sigma^2_{y_{12,0}} = Cov(y_{12,0}^N, y_{12,0}^N)$. Given these draws, one can form

$$
\Delta_t^1 = \begin{cases} 
(y_{10}^{N} - y_{11}^{N}) + (y_{10}^{F} - y_{11}^{F}) & t = 1 \\
(y_{20}^{N} - y_{21}^{N}) & t = 2 
\end{cases}
$$

and

$$
\rho_t^z = \frac{y_t^z - y_{11}^z}{y_t^z - y_{10}^z} \quad (t, z) = (1, N), (2, N), (1, F).
$$

**STEP 2: BOOTSTRAP VALUATION MODEL PARAMETER ESTIMATES**

The variance-covariance matrix for the valuation model parameters ($\beta$) can be estimated using the income imputations and replicate weights, much as was done in constructing $\sigma^2_\theta$. Specifically, let $\hat{\beta}^{rs}$ denote the parameter estimates from the model using replicate weight $W_r$ and income imputations $I_s$, with

$$
\hat{\beta}^k = \frac{1}{n_k} \sum_{r=1}^{R} \delta_{s,k}\hat{\beta}^{rs_r}
$$

denote the mean parameter estimates using income imputation $k$. The variance within imputation $k$ is is given by:

$$
\Omega_{\beta_k} = (R - 1) \sum_{r=1}^{R} \delta_{s,k} (\hat{\beta}^{rs_r} - \hat{\beta}^k)(\hat{\beta}^{rs_r} - \hat{\beta}^k)',
$$

with the average within imputation variance given by:

$$
U_\beta = \frac{1}{S} \sum_{s=1}^{S} \Omega_{\beta_s}.
$$

The between imputation variance is given by:

$$
B_\beta = \frac{1}{S - 1} \sum_{s=1}^{S} (\hat{\beta}^s - \bar{\beta})(\hat{\beta}^s - \bar{\beta})',
$$

where

$$
\bar{\beta} = \frac{1}{S} \sum_{s=1}^{S} \hat{\beta}^s.
$$

The overall variance-covariance matrix for the parameters of the valuation model becomes:

$$
\hat{\Sigma}_\beta = U_\beta + \left(1 + \frac{1}{S}\right) B_\beta.
$$

For $b = 1, \ldots, B$, draw $\beta^b \sim N(\bar{\beta}, \hat{\Sigma}_\beta)$.
**STEP 3: CONSTRUCT BOOTSTRAPPED VALUES FOR** \( L_T \) **AND** \( V_T \) **(** \( T = 1, 2 \) **) AND** \( L_T \).**

For each \( b = 1, \ldots, B \), we can construct a bootstrapped value per lost user day \( V_T^b \) using the bootstrapped model parameters \( \beta^b \), with the corresponding bootstrapped percentage loss by zone (i.e., \( \rho^b_{z}, z = N, F \)) used in calibrating the model. The bootstrapped overall loss for period \( t \) then becomes \( L_T^b = V_T^b \cdot \Delta^b \), with the total loss \( L_T^b = L_1^b + L_2^b \). With these results, we can construct

\[
Var(V_T^b) = \frac{1}{B} \sum_{b=1}^{B} (V_T^b - V_T)^2; \tag{32}
\]

\[
Var(\Delta_T^b) = \frac{1}{B} \sum_{b=1}^{B} (\Delta_T^b - \Delta_T)^2; \tag{33}
\]

and

\[
Cov(V_T, \Delta_T^b) = \frac{1}{B} \sum_{b=1}^{B} (V_T^b - V_T)(\Delta_T^b - \Delta_T), \tag{34}
\]

from which \( \text{Corr}(V_T, \Delta_T) \) can be constructed. Finally, we can also compute

\[
Var(L_T) = \frac{1}{B} \sum_{b=1}^{B} (L_T^b - L_T)^2 \quad t = 1, 2; \tag{35}
\]

which is the bootstrapped value for \( L_T \).

**VARIANCE ESTIMATES FOR BOATING DAMAGES**

The variance for the boating valuation is much simpler because there is only one period and then only for the North Gulf.

The total losses are

\[
L_T = V_T \cdot \Delta_T
\]

so that all second period elements are identically zero and have no uncertainty. The variance of \( L_T \) is given by equation (9), where \( t = 1 \). The variance of this expression requires \( \sigma^2_{\text{lost trips}} \) and \( \sigma^2_{\text{lost value per lost trip}} \) in addition to the estimates of lost trips (\( \Delta_T \)) and lost value per lost trip (\( V_T \)). The variance estimate of lost trips (\( \sigma^2_{\text{lost trips}} \)) comes from the infield study and is explained in Technical Memo B1—Estimation Procedures for Count Data. The variance estimate for lost value per lost user day (\( \sigma^2_{\text{lost value}} \)) is explained in equations (2) through (9) in the shoreline section. These estimates provide the ingredients for the estimates for boating damages in Exhibit 1. The sensitivity analysis proceeds just as in the Shoreline case but with only one period is much simpler. The random draws of trips during the spill and baseline equation (23) use a diagonal variance-covariance matrix for trips during the baseline and during the spill period. The remainder of the analysis is analogous to Step 2 and Step 3 in the Sensitivity analysis for the Shoreline.
RESULTS

The following table summarizes the results from the precision calculation task outlined above. Each entry provides the relevant mean value, with the corresponding estimated standard error in parentheses.

**EXHIBIT 1. PRECISION RESULTS**

<table>
<thead>
<tr>
<th>TIMEFRAME</th>
<th>LOST USER DAYS Δt (MILLIONS)</th>
<th>VALUE PER LOST USER DAY Vt (Dollars)</th>
<th>TOTAL LOSS Lt (MILLIONS OF DOLLARS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shoreline</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period 1</td>
<td>10.30 (1.36)</td>
<td>34.81 (0.93)</td>
<td>358.38 (48.29)</td>
</tr>
<tr>
<td>Period 2</td>
<td>2.17 (1.08)</td>
<td>37.77 (0.97)</td>
<td>82.11 (40.96)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>440.50 (70.14)</td>
</tr>
<tr>
<td><strong>Shoreline - Sensitivity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period 1</td>
<td>10.34 (1.38)</td>
<td>35.15 (1.55)</td>
<td>362.76 (47.51)</td>
</tr>
<tr>
<td>Period 2</td>
<td>2.14 (1.09)</td>
<td>38.05 (1.05)</td>
<td>81.47 (41.31)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>444.23 (72.02)</td>
</tr>
<tr>
<td><strong>Boating</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period 1</td>
<td>0.215 (0.073)</td>
<td>15.75 (2.38)</td>
<td>3.39 (1.27)</td>
</tr>
<tr>
<td><strong>Boating - Sensitivity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period 1</td>
<td>0.216 (0.073)</td>
<td>16.36 (2.90)</td>
<td>3.54 (1.38)</td>
</tr>
</tbody>
</table>

Notes:
1. The estimates for the value per lost user day presented in this exhibit differ slightly from those presented in Technical Memo A - Summary of Recreation Assessment because the values presented here are based on the replicate weights described above, whereas the estimates in Technical Memo A are based on the use of standard weights.

REFERENCES
