Conference on Seasonality, Seasonal Adjustment and their implications for Short-Term Analysis and Forecasting

10-12 May 2006

Finite Sample Revision Variances for ARIMA Model-Based Signal Extraction

Tucker McElroy and Richard Gagnon
Finite Sample Revision Variances for ARIMA Model-Based Signal Extraction

Tucker McElroy and Richard Gagnon
U.S. Census Bureau
Introduction

- **Revision Measures**: as more data becomes available, signal extraction estimates get updated. How much do the estimates change? How can we quantify this?

- **Need for Revision Measures**: a concurrent signal extraction estimate depends on past and present data. Official agencies revise their published estimates as more data becomes available.

- **SEATS approach**: historically (Pierce, 1980) one computes the variance of the update, or revision. Exact calculation is possible in a model-based framework.

- **Finite vs. semi-Infinite Sample**: previous approaches assume data span extends to the infinite past. We assume a finite sample.
Notation

• Observed data $Y_1, \cdots, Y_n$. Additional data $Y_{n+1}, \cdots, Y_{n+h}$ for a revision lead $h > 0$.

• Additive decomposition: $Y_t = S_t + N_t$ into signal plus noise. Suppose (ARIMA) models are known for $S_t$ and $N_t$.

• Optimal signal extraction estimate $\hat{S}_{t|n}$ for $S_t$ given data in span from 1 to $n$.

• Revision = New - Old = $\hat{S}_{t|1}^{n+h} - \hat{S}_{t|1}^n$. Denoted its variance by $R_t(h)$ ($n$ is suppressed).
Revision Variances

Consider the following orthogonal decomposition:

$$\hat{S}_t|_1^n - S_t = (\hat{S}_t|_1^n - \hat{S}_t|_{n+h}) + (\hat{S}_t|_{n+h} - S_t)$$

The terms on the right are orthogonal. Hence the revision variance is

$$R_t(h) = V_{t|_1^n} - V_{t|_{n+h}}$$

where $V_{t|_1^n}$ is the signal extraction MSE for time $t$ based on a sample from 1 to $n$. Note that $h = \infty$ is allowed.
Finite Sample Implementation

• Assuming a finite sample, the covariance matrix for the signal extraction error process can be easily computed using formulas from McElroy (2005). Denote this by $M^{(n)}$, where $n$ denotes the dimension.

• Then the revision variance is

$$R_t(h) = M^{(n)}_{tt} - M^{(n+h)}_{tt}$$

• Holds for $h < \infty$. $R_t(\infty)$ is computed in another way (see below).
Properties

- $R_t(h)$ increases in $h$, maximum of $R_t(\infty)$.

- Depends on $t$ (position in sample) and $n$.

- SEATS uses revision measure $1 - \sqrt{1 - R_t(h)/R_t(\infty)}$. The quantity

$$\frac{R_t(\infty) - R_t(h)}{R_t(\infty)} = \frac{V_{t|_1}^{n+h} - V_{t|_1}^\infty}{V_{t|_1}^n - V_{t|_1}^\infty}$$  \hspace{1cm} (1)$$

gives proportion of “total revision variance” that remains, unaccounted for by revising at $h$ revision lead.
Obtaining $R_t(\infty)$

- Need to know $V_t|_{1}\infty$; semi-infinite filter goes back $t - 1$ data points (from current position at time $t$) and forward infinitely far.

- Adapt Bell and Martin (2004), which is concerned with infinite past-finite future filters. Formulas are similar; obtain autocovariance generating function for the error process.

- The procedure involves computing certain partial fraction decompositions, which depend on $m = t - 1$. 
Implementation/Partial Fraction Decomposition

• We obtain partial fraction decompositions by solving linear systems.

• Two decompositions used, depending on whether $m$ is large or small. Since $m$ determines the degree of a certain polynomial, numerical instabilities can result from polynomial division and multiplication if $m$ is large. The large $m$ decomposition essentially ameliorates this problem.

• Recursion in $m$; obtains $m + 1$ case from $m$ case. This is useful to compute $V_t|_1$ for various values of $t$. 
Empirical Illustrations

• Compare SEATS revision variances to the exact values (our method). Consider concurrent (so \( t = n \)).

• SEATS’ calculation in our notation:

\[
\tilde{R}(h) = V_{n\mid -\infty}^{n} - V_{n\mid -\infty}^{n+h}
\]

Note this quantity does not depend on \( n \). But \( R_{n}(h) \) does.

• So \( \tilde{R}(\infty) = V_{n\mid -\infty}^{n} - V_{n\mid -\infty}^{\infty} \) is the SEATS maximum. These approximate revision variances are calculated by a different method (Pierce, 1980 and Maravall, 1986).
Empirical Illustrations

• Compare $R_n(h)$ to $\tilde{R}(h)$ for various $n$ and $h$ and various models. In each case, we compute the revision measure (1).

• Consider Airline Models with $\theta = .6$ and $\Theta = .6, .7, .8, .9$ for monthly data. So $(1 - B)(1 - B^{12})Y_t = (1 - \theta B)(1 - \Theta B^{12})\epsilon_t$.

• Take $n = 60$ to 132 (5 to 11 years), and $h = 12$ to 60 (1 to 5 years).

• Results presented in Tables 1 through 4.
Table 1. Revision Measure for (.9, .6) Airline Model.

<table>
<thead>
<tr>
<th>Lead</th>
<th>60</th>
<th>72</th>
<th>84</th>
<th>96</th>
<th>108</th>
<th>120</th>
<th>132</th>
<th>SEATS</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>.4015</td>
<td>.4006</td>
<td>.4001</td>
<td>.3999</td>
<td>.3999</td>
<td>.3999</td>
<td>.3999</td>
<td>.3999</td>
</tr>
<tr>
<td>24</td>
<td>.6412</td>
<td>.6404</td>
<td>.6401</td>
<td>.6399</td>
<td>.6399</td>
<td>.6399</td>
<td>.6399</td>
<td>.6399</td>
</tr>
<tr>
<td>36</td>
<td>.7848</td>
<td>.7842</td>
<td>.7840</td>
<td>.7840</td>
<td>.7839</td>
<td>.7839</td>
<td>.7839</td>
<td>.7839</td>
</tr>
<tr>
<td>48</td>
<td>.8709</td>
<td>.8705</td>
<td>.8704</td>
<td>.8703</td>
<td>.8703</td>
<td>.8703</td>
<td>.8703</td>
<td>.8703</td>
</tr>
<tr>
<td>60</td>
<td>.9225</td>
<td>.9223</td>
<td>.9223</td>
<td>.9222</td>
<td>.9222</td>
<td>.9222</td>
<td>.9222</td>
<td>.9222</td>
</tr>
</tbody>
</table>
Table 2. Revision Measure for (.9, .7) Airline Model.

<table>
<thead>
<tr>
<th>Lead</th>
<th>60</th>
<th>72</th>
<th>84</th>
<th>96</th>
<th>108</th>
<th>120</th>
<th>132</th>
<th>SEATS</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.3059</td>
<td>0.3028</td>
<td>0.3013</td>
<td>0.3006</td>
<td>0.3003</td>
<td>0.3001</td>
<td>0.3000</td>
<td>0.2999</td>
</tr>
<tr>
<td>24</td>
<td>0.5162</td>
<td>0.5129</td>
<td>0.5114</td>
<td>0.5107</td>
<td>0.5103</td>
<td>0.5101</td>
<td>0.5100</td>
<td>0.5099</td>
</tr>
<tr>
<td>36</td>
<td>0.6620</td>
<td>0.6594</td>
<td>0.6581</td>
<td>0.6575</td>
<td>0.6572</td>
<td>0.6571</td>
<td>0.6570</td>
<td>0.6570</td>
</tr>
<tr>
<td>48</td>
<td>0.7636</td>
<td>0.7617</td>
<td>0.7608</td>
<td>0.7603</td>
<td>0.7601</td>
<td>0.7600</td>
<td>0.7600</td>
<td>0.7599</td>
</tr>
<tr>
<td>60</td>
<td>0.8346</td>
<td>0.8332</td>
<td>0.8325</td>
<td>0.8322</td>
<td>0.8321</td>
<td>0.8320</td>
<td>0.8320</td>
<td>0.8319</td>
</tr>
</tbody>
</table>
Table 3. Revision Measure for (.9, .8) Airline Model.

<table>
<thead>
<tr>
<th>Lead</th>
<th>60</th>
<th>72</th>
<th>84</th>
<th>96</th>
<th>108</th>
<th>120</th>
<th>132</th>
<th>SEATS</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>.3831</td>
<td>.3744</td>
<td>.3690</td>
<td>.3657</td>
<td>.3636</td>
<td>.3623</td>
<td>.3615</td>
<td>.3600</td>
</tr>
<tr>
<td>36</td>
<td>.5108</td>
<td>.5022</td>
<td>.4970</td>
<td>.4937</td>
<td>.4916</td>
<td>.4903</td>
<td>.4895</td>
<td>.4880</td>
</tr>
<tr>
<td>48</td>
<td>.6108</td>
<td>.6032</td>
<td>.5985</td>
<td>.5955</td>
<td>.5937</td>
<td>.5925</td>
<td>.5917</td>
<td>.5904</td>
</tr>
<tr>
<td>60</td>
<td>.6897</td>
<td>.6832</td>
<td>.6792</td>
<td>.6767</td>
<td>.6751</td>
<td>.6741</td>
<td>.6735</td>
<td>.6723</td>
</tr>
</tbody>
</table>
Table 4. Revision Measure for (.9, .9) Airline Model.

<table>
<thead>
<tr>
<th>Lead</th>
<th>60</th>
<th>72</th>
<th>84</th>
<th>96</th>
<th>108</th>
<th>120</th>
<th>132</th>
<th>SEATS</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>.1441</td>
<td>.1328</td>
<td>.1250</td>
<td>.1193</td>
<td>.1150</td>
<td>.1118</td>
<td>.1094</td>
<td>.1000</td>
</tr>
<tr>
<td>24</td>
<td>.2578</td>
<td>.2412</td>
<td>.2293</td>
<td>.2206</td>
<td>.2140</td>
<td>.2090</td>
<td>.2051</td>
<td>.1900</td>
</tr>
<tr>
<td>36</td>
<td>.3506</td>
<td>.3317</td>
<td>.3180</td>
<td>.3078</td>
<td>.3000</td>
<td>.2940</td>
<td>.2893</td>
<td>.2710</td>
</tr>
<tr>
<td>48</td>
<td>.4280</td>
<td>.4086</td>
<td>.3943</td>
<td>.3835</td>
<td>.3752</td>
<td>.3688</td>
<td>.3638</td>
<td>.3439</td>
</tr>
<tr>
<td>60</td>
<td>.4938</td>
<td>.4748</td>
<td>.4605</td>
<td>.4497</td>
<td>.4414</td>
<td>.4349</td>
<td>.4298</td>
<td>.4095</td>
</tr>
</tbody>
</table>
Summary of Tables

- Across the rows: values decrease in $n$ and get fairly close to the SEATS value. Tighter approximation for higher revision leads when $\Theta = .6$ and $.7$.

- Down the columns: as expected most of the revisions have occurred by the fourth or fifth year. But for larger values of $\Theta$ slower convergence.

- For $\Theta = .9$ all the values are under 50 percent.

- The largest discrepancies between $SEATS$ and the finite-sample approach occur for large $\Theta$ and small sample size.
Conclusion

• We correct SEATS’ revision variance for finite sample; SEATS assumes an infinite past of data, but our method does not. Our method is implemented in $X - 13A - S$.

• In practice, the discrepancy depends on the model parameters, sample size, and revision lead. Our method takes more time on a computer.

• There are extensions to calculating revision variances for growth rates.

• Acknowledgements: thanks to David Findley, Bill Bell, and Agustín Maravall.
References


