Comparative Study of Cross Sectional Methods for Time Series with Structural Changes

George Duncan
Wilpen Gorr
Janusz Szczypta

June 1994

H. John Heinz III School of Public Policy and Management
Carnegie Mellon University
Pittsburgh, PA 15213-3890
Email: js1m@andrew.cmu.edu

We thank Roman Kapuscinski for his helpful comments and suggestions.
1. Introduction

Hierarchical models for pooling and borrowing strength allow combining data from several parallel observational units (e.g., census tracts, experimental studies, products, etc.). The hierarchical models both describe the observational model for data and the structural model for the unobserved parameters. Further, they provide for inferences about any particular parameter through “borrowing strength” from the entire set of observational units. A variety of statistical ideas and terms are used to described borrowing-strength concepts, including (in some contexts) shrinkage and Stein estimation (James and Stein 1961); empirical Bayes (e.g., Efron and Morris 1975); and random effects models (Harville 1977). Observational units can be assigned to “equivalence sets” based on a formal relation that partition the units into sets, most importantly as defined by a common influence of exogenous factors, for example food and environmental conditions for fish in a lake, social and economic conditions for teenagers in a neighborhood, etc.

Duncan, Gorr, and Szczyupa (DGS: 1993, 1994) were the first to develop a new application, called C-MSKF, of recursive “borrowing strength” to dynamic settings. Their goal was to determine if borrowing strength in the C-MSKF provided significantly improved forecasting accuracy over the univariate Multi-State Kalman Filter (MSKF). Shrinkage formulae derived by DGS rapidly adjusted univariate time series models in response to cyclical or structural changes (e.g., step changes, time trend slope changes, turning points, etc.) using recursively estimated, time-varying- parameter models. For example, suppose that most or all members of an equivalence set of univariate time series experienced a sizable step-change increase at the last historical data point. Individual time series estimates would have small (damped) increases but greatly increased variances. The overall mean for the equivalence set’s last data point would be at the new level caused by the step jump and its variance would be relatively low (depending of course on the cohesiveness of the equivalence set). Weighting, proportional to inverse variances, might very well throw weight appropriately on the equivalence set's mean.
DGS integrated the Multi-State Kalman Filter (MSKF: Harrison and Stevens 1971) with a Conditionally Independent Hierarchy Model (CIHM) (e.g., Kass and Steffey 1989) to produce the cross-Sectional Multi-State Kalman Filter (C-MSKF). The MSKF is perhaps the most sophisticated and complicated univariate method for dynamic settings. Maximum likelihood estimates of parameters from the resultant Bayesian, hierarchical random effects model yielded shrinkage formulae appropriate for dynamic settings. Out-of-sample comparisons of school district revenue forecasts made by the MSKF versus C-MSKF provided evidence that the "borrowing strength" is effective in increasing forecast accuracy.

In summary, the DGS's contributions were that: 1) the C-MSKF was the first method to recursively use the "borrowing strength" for dynamic model estimates, 2) DGS rigorously applied the CIHM yielding shrinkage formulae, and 3) the number of C-MSKF parameters increased only linearly with the size of the equivalence set. On the latter point, earlier cross-sectional methods either had 1) too much structure and therefore the number of parameters increasing exponentially with equivalence set size (e.g., Jones 1966) or 2) too little structure and therefore drew little information from cross-sections (e.g., Enns et al. 1982).

Fildes (1993) pointed out that simple forecast methods are often as accurate or more accurate than sophisticated methods, and suggested that benchmark univariate time series methods be included in empirical comparisons of the C-MSKF. This paper follows that suggestion, plus it implements the "borrowing strength" approach to derive a new cross-sectional counterpart to a relatively simple univariate method. While the C-MSKF only has its number of parameters increasing linearly with equivalence set size, its number of parameters is large because the MSKF has a large number of parameters. In contrast, there are only four hyperparameters per time series model parameter in DGS's formulation (e.g., eight cross-sectional parameters for a linear time trend). Thus the question arises whether a simpler univariate time series model, with fewer parameters to estimate, might lead to an overall better cross-sectionally adjusted method. The new method presented in this paper is thus based on Holt's exponential smoothing model and is a
heuristic application of DGS's shrinkage formulae, that yields the Cross-Sectional Holt method (CSH).

Section 2 describes the CSH. Section 3 describes the comparative empirical study which uses school district income tax revenue and simulated Monte Carlo data. Section 4 presents the findings from the comparison. Section 5 concludes the paper.

2. Cross-Sectional Time Series Methods

Similar to the C-MSKF of DGS we develop here a cross-sectional counterpart, CSH, to Holt's two parameter linear exponential smoothing. Holt's method (see Makridakis et al. 1989) is given by the following recursive equations:

\begin{align*}
L_t &= \alpha X_t + (1 - \alpha)(L_{t-1} + S_{t-1}), \\
S_t &= \beta (L_t - L_{t-1}) + (1 - \beta)S_{t-1}, \text{ and} \\
F_{t+m} &= L_t + mS_t
\end{align*}

where

- $L_t$ is the current smoothed level of the time series,
- $X_t$ is the current datum for the series,
- $S_t$ is the current smoothed slope of the time series unit,
- $F_{t+m}$ is the $m$ step-ahead forecast, and
- $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$ are smoothing parameters.

In our implementation the values of the smoothing parameters are estimated by minimizing the sum of squared errors over a grid with values ranging from 0 to 1 and a step of 0.1. The first observation and the difference of second and first observation are taken as the initial values of smoothed level and slope, respectively.
The CSH method is defined by the following equations:

\[ L_{it} = \alpha X_{it} + (1 - \alpha) (L^*_{i,t-1} + S^*_{i,t-1}), \]  
(4)

\[ S_{it} = \beta (L^*_{it} - L^*_{i,t-1}) + (1 - \beta) S^*_{i,t-1}, \]
and

\[ L^*_{it} = \frac{(\sigma^2_{L_{it}} X_{it} + \tau^2_t L_{it})}{(\sigma^2_{L_{it}} + \tau^2_t)}, \]
(6)

\[ S^*_{it} = \frac{(\sigma^2_{S_{it}} \Delta_t + \lambda^2_t S_{it})}{(\sigma^2_{S_{it}} + \lambda^2_t)}, \]
(7)

\[ F_{i,t+m} = L^*_{it} + m S^*_{it}, \]
(8)

where

- \( i \) is the index of the time series in the cross-sectional group of size \( N \) (i=1, 2, ..., N),
- \( L_{it} \) is the current smoothed level at time \( t \) for series \( i \),
- \( X_{it} \) is the current datum at time \( t \) for series \( i \),
- \( S_{it} \) is the current smoothed slope at time \( t \) for series \( i \),
- \( L^*_{it} \) is the cross-sectionally adjusted level at time \( t \) for series \( i \),
- \( S^*_{it} \) is the cross-sectionally adjusted slope at time \( t \) for series \( i \),
- \( X_\cdot t \) is the sample mean of time series values in the group at time \( t \),
- \( \Delta_{it} = X_{it} - X_{i,t-1} \),
- \( \Delta_\cdot t \) is the sample mean of the first differences of time series in the group at time \( t \),

\[ \tau^2_t = \frac{1}{(N-1)} \sum_{i=1}^{N} (X_{it} - X_\cdot t)^2, \]
(9)

\[ \lambda^2_t = \frac{1}{(N-1)} \sum_{i=1}^{N} (\Delta_{it} - \Delta_\cdot t)^2, \]
(10)

\[ \sigma^2_{L_{it}} = \gamma (X_{it} - L_{it})^2 + (1 - \gamma) \sigma^2_{L_{i,t-1}}, \]
(11)

\[ \sigma^2_{S_{it}} = \rho (\Delta_{it} - S_{it})^2 + (1 - \rho) \sigma^2_{S_{i,t-1}}, \]
(12)

- \( F_{i,t+m} \) is the \( m \) step-ahead forecast at time \( t \) for series \( i \), and
- \( 0 \leq \alpha \leq 1, \ 0 \leq \beta \leq 1, \ 0 \leq \gamma \leq 1, \) and \( 0 \leq \rho \leq 1 \) are parameters to be estimated.

The CSH method updates the estimated level and slope twice at each time period \( t \). The first step
of updating (according to equations (4) and (5)) is performed similarly to the standard smoothing formulas (1) and (2) above. The second step of updating (according to equations (6) and (7)) is based on a “shrinkage formula” analogous to the one used in DGS. The new level (slope) for individual series is a weighted sum of the current individual level (slope) of the series and the grand average level (slope) over the cross section. The weights are the reciprocals of estimated variances of the two terms being weighted. The variance of the cross section at a given time period is calculated as a sample variance of the appropriate cross-sectional terms at that time. The variance of the individual smoothed level and slope are derived from a smoothing process defined by equations (11) and (12). The equations (11) and (12) the updating equation by Bretschneider (1986). The initial values for the level and slope of each individual time series are respectively the first data point and the difference of the second and first data point respectively for the level and slope of each unit. We estimate the four smoothing parameters by minimizing the mean squared error of one step-ahead forecast errors through grid search.

3. Comparative study

The comparative study of this paper has two parts. First is an empirical validation based on income tax revenue data from 40 school districts in Allegheny County, Pennsylvania as in DGS. Second is a Monte Carlo experiment that is realistic (similar to the school district case study) but limited in scope. It demonstrates feasibility of a more extensive Monte Carlo experimentation, provides more control, and therefore sharper results.

3.1 School District Case Study

The data used in this paper are annual income tax collections from 1972 to 1988, two additional years (1987-1988) beyond those used in DGS, for 40 school districts in Allegheny
County, obtained from the Pennsylvania Department of Education. The data contain significant pattern changes due to recessions, starting in 1974/75 (with a relatively short impact) and 1981/82 (longer lasting and much stronger impact).

DGS had an expert divide school districts into three similar categories according to their sensitivity to economic cycles. We call them: Category I (24 school districts with low sensitivity to external influences), Category II (9 school districts with medium sensitivity to external influences), and Category III (7 school districts with high sensitivity to external influences). Scale differences across school districts’ time series within each category were removed by calculating standardized values. Lastly, for the convenience of dealing with only positive numbers, we added the constant value of “5” to each standardized data point. Figure 1 presents the standardized time series by category. Categories II and III have quite large structural changes in the individual series at the time of economic downturns in 1981/82.

[ Figure 1 goes about here.]

In this work, the time series data of each individual unit are divided into two periods: 1) the historical period (e.g. 1972 - 1982), taken to be available for estimation of the parameters, and 2) the forecast period (e.g. 1983 - 1985). We extend the design of DGS by further varying the location of forecasting origin relative to the structural change. The forecasting origin is located 1, 2, 3, and 4 periods/years after the 1981/82 structural change (creating respectively Groups 1 to 4). This allows us to increase the number of cases for the comparison and vary an additional factor important for testing forecasting and structural change.

We vary the estimation sample size by systematically dropping the earliest historical data points one at a time. Thus, the data sets created for estimation varies from a maximum of eleven to a minimum of only four data points. Each series created for the comparison is then classified by category and group, and then used for forecasting. We forecast each series one, two, and three steps (years) ahead.
3.2 Monte Carlo Simulation

The Monte Carlo experiment matches the impact of 1981/82 recession on school district data and the design of section 3.1 by creating three categories of time series patterns, four groups, and several data sets of different length (by dropping earlier observations one at a time). Figure 2 presents the structural change characteristic (the changing slope) for generating the three categories. Category I is a category where all series have slopes averaging 1.0 before the change and 0.8 after the change. Category II groups series that have slopes of about 1.0 before and 0.0 after the change. Category III represents a turning point design where the slopes change from about 1.0 before the change to about -1.0 after the change. Each category has 12 time series units.

[Figure 2 goes about here.]

Figure 3 depicts the true means of Category III and variation in groups 1 through 4. Group 1 is comprised of series that have only one historical observation after the structural change. Series in Group 2 have two historical observations after the change, etc. Forecasting series in Group 1 is the most challenging due to "long" irrelevant history (indicating increasing trend) and only one relevant observation at the forecasting origin (indicating decreasing trend).

[Figure 3 goes about here.]

Figure 4 presents the data sets generated for each of the three Categories. The observations in each of the 12 time series in every Category with a randomly generated noise given by the following equations:

\[ Y_{it} = L_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, \sigma^2) \]  

(13)
\[ L_{it} = L_{i,t-1} + (S_t + \nu_{it}), \quad \nu_{it} \sim N(0, 0.01) \] (14)

where

- \( i \) is the index of time series unit \( i = 1, \ldots, 12 \),
- \( \varepsilon_{it}, \nu_{it} \) are normally distributed disturbance terms,
- \( L_{it} \) is current level, and
- \( S_t \) is a deterministic series with first \( m = 10 \) values equal to 1.0 (same for all three categories) and the remaining values equal to 0.8 for Category I, 0.0 for Category II, and -1.0 for Category III.

Since the variance of \( \nu_{it} \) is relatively small, the series conform to the overall patterns (the slope changes) of the three categories. To match the patterns of the empirical data more closely the variance of \( \varepsilon_{it} \) is smaller for Category I \( (\sigma^2 = 1) \) than for Category II and III \( (\sigma^2 = 4) \). Further, to avoid any transformation of data, initial values \( L_{i0} \) were set equal to 10 for all \( i \).

[ Figure 4 goes about here.]

### 3.3 Treatments

We applied the Random Walk (RW), Holt’s exponential smoothing, and the MSKF separately to each individual series. Then, we used the CSH and C-MSKF method for one category at a time. We anticipated that the C-MSKF as well as the CSH method would outperform the RW, Holt, and MSKF method in most of the cases, thereby showing the value of the “borrowing strength from neighbors” approach.

Forecast error measures include mean absolute percentage error (MAPE) for precision and mean percentage error (MPE) for bias computed by forecast horizon (one, two, and three years ahead) using the following formulas:
\[
\text{MAPE}_{T+k} = \left( \frac{100}{N} \right) \sum_{i=1}^{N} \frac{|Y_{i,T+k} - Y_{i,T+k}^f|}{Y_{i,T+k}} \quad \text{and} \\
\text{MPE}_{T+k} = \left( \frac{100}{N} \right) \sum_{i=1}^{N} \left( \frac{Y_{i,T+k} - Y_{i,T+k}^f}{Y_{i,T+k}} \right)
\]

where

\( T \) is the forecast origin,

\( k \) is the forecast horizon (1, 2, or 3),

\( Y_{i,T+k} \) is the actual value for the origin \( T \), horizon \( k \), and unit \( i \),

\( Y_{i,T+k}^f \) is the forecast for origin \( T \), horizon \( k \), and unit \( i \), and

\( N \) is the number of observational units in the given Category/Group.

4. Results

We expected that the cross-sectional methods would adjust much more quickly to structural changes that the univariate methods but with increasing number of observations after the structural change the forecast precision of all methods would improve. We also expected that increasing length of historical time series would have a negative influence on forecasting precision given the structural changes near the ends of the historical data sets. Category II is an exceptional case where the series become flat after the structural change (slopes average zero). In this category we expected that RW would dominate other methods.

Tables 1 and 2 present mean absolute percent error (MAPE) comparisons of the cross-sectional methods to benchmark time series methods (RW and Holt) for the school district data by Category and Group. To make the presentation of results manageable, we aggregated the forecast errors over the horizon (average of one, two, and three step-ahead forecast errors). Furthermore, we aggregated over sample series to two types of series: short (4 to 6 historical observations) and
longer series (7 to 11 historical observations). The reported results are the same, qualitatively, as for the disaggregated data statistics.

The C-MSKF is significantly better than RW only for Category I (low change). For the other two categories, the C-MSKF is worse than RW but it closes the gap with increasing number of observations after the structural change. The C-MSKF is worse than Holt for most cases in each category. RW is significantly better than CSH in Category II and III (see Table 2). CSH fully dominates Holt in almost all cases. Holt is better than CSH only in a few cases.

[Tables 1 and 2 go about here.]

Tables 3 and 4 present mean absolute percent error (MAPE) comparisons of the cross-sectional methods with benchmark methods (RW and Holt) for the Monte Carlo experiment by Category, Group, and for short and longer series. The C-MSKF (Table 3) is significantly better than RW in Category I. In Categories II and III, the C-MSKF is better for Groups 3 and 4. Comparing the C-MSKF with the Holt method, the C-MSKF is relatively worse only for the longer series in Category I. Otherwise, C-MSKF dominates Holt in Categories II and III. CSH is significantly better than RW in all of the cases in Category I. RW is better than CSH in Category II but the differences are not significant. Comparing CSH with Holt, CSH is significantly better than Holt for every Group in each Category. The following is a more detailed comparison of all methods for both school district and Monte Carlo studies.

[Tables 3 and 4 go about here.]

School District case study:

Figure 5 presents the mean absolute percent error (MAPE) results by Category, Group and time series length. The MAPE performance generally improved with increasing number of
observations after the change as well as with the increasing time series length. The cross-sectional methods were the most precise across nearly all categories and groups especially for short series. In Category II, RW dominated other methods in Group 1 and 2. The inappropriateness of history in Category II was represented in almost the same MAPE values for each method for different time series length (no improvement for longer series). Category III has a turning point near the end of the series making the predictions more challenging. The C-MSKF forgets the irrelevant history much faster than MSKF which is visible in the increasing gap between the two methods. Similarly the gap between CSH and Holt method increases for longer time series. The exponential smoothing methods (both CSH and Holt) are more accurate than the multiple data Kalman filter methods (C-MSKF and MSKF) in Category III. RW is the most accurate method in most cases of Category III since it is not influenced by contradicting information.

[ Figure 5 goes about here.]

Figure 6 presents the mean percent error (MPE) or biasedness for the school district study. In general, the results confirm the expectations as stated above. Biasedness diminishes with increased time series length, with increased history after the change, and with decreasing size of the structural change. Both cross-sectional methods are reliable as indicated by mostly monotonic (smoothly changing) curves of the errors with increasing length of the series. Holt and MSKF have a lot of fluctuation — the errors are not improving smoothly with increasing time series length. They are often less biased than their cross-sectional equivalents (CSH and C-MSKF methods) but it is often noticeable that the increasing history is increasing their bias significantly. Overall, there is no winner in terms of MPE. CSH, however, performs quite well, having relatively low MPE compared with other methods. CSH is even better than RW for most cases in Category II. In Category III, RW is the least biased method for the most challenging case (Group 1 – relatively large structural changes and only one observation after the change).
Monte Carlo experiment

Figure 7 presents the MAPE results for the Monte Carlo experiment by Category, Group, and time series length as in the school district case. Precision increases with increasing number of observations after the change, with increasing time series length, and with decreasing size of the structural change. In general, one can notice that both cross-sectional methods are better than the other methods for most of the cases, especially when three or more observations after the change are available (Group 3 and Group 4). When there are only one or two observations after the change (Group 1 and Group 2) the CSH method is overall the best. The CSH is even better than RW for most cases in Category II where RW was supposed to dominate. History is not influencing the cross-sectional methods as much as the individual time series methods for which it either helps in getting more precise forecasts (Category I) or decreases the precision (Category II and III). This is particularly noticeable in Category III where increasing the time series length causes the errors to become larger. The longer the history for Category III, the stronger the indication of opposite to current trend causing the non cross-sectional methods to perform poorly due to using the contradicting information about the slope too much.

Figure 8 presents MPE results for the Monte Carlo experiments by Category, Group, and time series length. The role of history is more notable here than for the MAPE case above. Relevant history (Category I) lowers the MPE error but the improvements are not smooth. The MPE performance fluctuates and one can see dramatic changes for different time series lengths (e.g. for MSKF results). In Category II, RW dominates other methods, as expected, having very small MPE. Overall in Category II, both cross-sectional methods perform quite well and seem to
be much more reliable than the other methods. Holt is less biased than the C-MSKF but it is also less reliable having bigger error differences for series of different length. In Category III the best method is the CSH. The C-MSKF is also very dependable in Groups 3 and 4. Performance of MSKF and Holt fluctuate and deteriorate significantly for longer series.

[Figure 8 goes about here.]

5. Conclusions

We evaluated the concept of “borrowing strength from neighbors” in comparison with univariate time series methods used for forecasting time series with structural changes. We compare two methods that use the borrowing strength mechanism — the C-MSKF as in DGS and a newly developed method, cross-sectional Holt exponential smoothing (CSH) — with three univariate time series methods: two benchmark methods (Random Walk and Holt’s exponential smoothing) and the MSKF. The comparative validation of this paper uses income tax revenue data from 40 school districts in Allegheny County, Pennsylvania and Monte Carlo experiment. The experimental design of the validation uses three different categories of data sets, four different groups within each category and different length of time series. Each category combines sets of series that present similar slope changes at the same time period. The higher the category the stronger the slope changes. Within each category we controlled for the location of the forecasting origin creating four distinct groups. Group 1 has forecasting origin one period after the structural change. Group 4 has the forecasting origin four periods after the change.

All methods considered in this paper (except Random Walk) confirmed the expectations on school district data as well as on the Monte Carlo data. Their accuracy and unbiasedness increased with increasing number of historical observations after the change, with increasing time series length (for Category I), and with decreasing size of the structural change. There are clear
differences between Groups 1, 2, and 3. The differences in performance between Groups 3 and 4 are less significant suggesting that for a method to increase its accuracy it needs at least three observations after a structural change in the data, but after that the gains in precision may be minor.

Increasing length of historical series influences the forecasting performance in two ways. When history is relevant (confirming the recent trend - as in Category I), it helps lower forecasting errors. When it is irrelevant (contradicting the recent trend - as in Category II or III), it usually lowers the accuracy of a method. This is particularly the case for the non cross-sectional methods in Category III where the longer the history before the change (stronger indication of opposite trend) the poorer the performance. The cross-sectional methods usually are much less influenced by the irrelevant history of an individual time series. Thus, the cross-sectional methods recognize and forget irrelevant history, put more emphasis on the cross-sectional data, and gain forecasting accuracy.

Both cross-sectional methods are better than the other methods for most cases, especially when three or more observations after the change are available for the parameters estimation. They are reliable and accurate because they utilize the cross-sectional information from similar categories (“borrowing strength”) and respond well to structural changes. The good precision of CSH (especially in Category II) and its low biasedness in most Categories suggest that this method “borrows the strength” in a more effective way than the C-MSKF method.

The poor performance of the Holt method in MAPE and its good performance in MPE indicates that the Holt method is unreliable for the particular forecasting environment that we considered here. MAPE fluctuates with increasing time series length also indicate the unreliability of the method since the past history influences its performance too much. “Borrowing strength” applied to Holt (i.e., through the CSH method) can significantly improve the method, making it more accurate and reliable.
References


### Table 1. Mean Absolute Percentage Error Comparison of C-MSKF with Random Walk and Holt by Category, Group, and Time Series length for School District Case Study. §

<table>
<thead>
<tr>
<th>Category I</th>
<th>Random Walk Groups</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Holt Groups</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Short TS</strong></td>
<td>CMSKF</td>
<td>10.3</td>
<td>7.5 ***</td>
<td>7.0 ***</td>
<td>8.9 ***</td>
<td>CMSKF</td>
<td>10.3</td>
<td>7.5</td>
<td>7.0 ***</td>
<td>8.9</td>
</tr>
<tr>
<td></td>
<td>CW</td>
<td>7.5 ***</td>
<td>9.4</td>
<td>9.9</td>
<td>11.2</td>
<td>Holt</td>
<td>9.9</td>
<td>6.2 ***</td>
<td>7.8</td>
<td>7.8 ***</td>
</tr>
<tr>
<td><strong>Longer TS</strong></td>
<td>CMSKF</td>
<td>7.5</td>
<td>6.6 ***</td>
<td>6.9 ***</td>
<td>8.5 ***</td>
<td>CMSKF</td>
<td>7.5</td>
<td>6.6</td>
<td>6.9</td>
<td>8.5</td>
</tr>
<tr>
<td></td>
<td>CW</td>
<td>7.5</td>
<td>9.4</td>
<td>9.9</td>
<td>11.2</td>
<td>Holt</td>
<td>7.0</td>
<td>4.8 ***</td>
<td>7.0</td>
<td>6.8 ***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category II</th>
<th>Random Walk Groups</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Holt Groups</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Short TS</strong></td>
<td>CMSKF</td>
<td>17.4</td>
<td>16.5</td>
<td>10.5</td>
<td>9.5</td>
<td>CMSKF</td>
<td>17.4</td>
<td>16.5</td>
<td>10.5</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td>CW</td>
<td>5.7 ***</td>
<td>7.1 ***</td>
<td>7.0 ***</td>
<td>9.0</td>
<td>Holt</td>
<td>15.5</td>
<td>10.5 **</td>
<td>8.7</td>
<td>9.3</td>
</tr>
<tr>
<td><strong>Longer TS</strong></td>
<td>CMSKF</td>
<td>14.7</td>
<td>15.7</td>
<td>12.5</td>
<td>10.7</td>
<td>CMSKF</td>
<td>14.7</td>
<td>15.7</td>
<td>12.5</td>
<td>10.7</td>
</tr>
<tr>
<td></td>
<td>CW</td>
<td>5.7 ***</td>
<td>7.1 ***</td>
<td>7.0 ***</td>
<td>9.0 **</td>
<td>Holt</td>
<td>13.9</td>
<td>12.7 ***</td>
<td>8.6 **</td>
<td>8.7 **</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category III</th>
<th>Random Walk Groups</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Holt Groups</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Short TS</strong></td>
<td>CMSKF</td>
<td>28.5</td>
<td>19.4</td>
<td>19.3</td>
<td>17.2</td>
<td>CMSKF</td>
<td>28.5</td>
<td>19.4</td>
<td>19.3</td>
<td>17.2</td>
</tr>
<tr>
<td></td>
<td>CW</td>
<td>7.7 ***</td>
<td>12.8 ***</td>
<td>8.6 ***</td>
<td>6.4 ***</td>
<td>Holt</td>
<td>24.3</td>
<td>12.5 ***</td>
<td>12.9 ***</td>
<td>11.9 ***</td>
</tr>
<tr>
<td><strong>Longer TS</strong></td>
<td>CMSKF</td>
<td>25.3</td>
<td>20.0</td>
<td>14.0</td>
<td>14.8</td>
<td>CMSKF</td>
<td>25.3</td>
<td>20.0</td>
<td>14.0</td>
<td>14.8</td>
</tr>
<tr>
<td></td>
<td>CW</td>
<td>7.7 ***</td>
<td>12.8 ***</td>
<td>8.6 ***</td>
<td>6.4 ***</td>
<td>Holt</td>
<td>22.0 *</td>
<td>15.6 **</td>
<td>15.3</td>
<td>10.9 ***</td>
</tr>
</tbody>
</table>

§ Significant differences in mean accuracy are based on the Wilcoxon Matched Pairs Signed Rank two-tailed Test

(Pfaffenberger and Patterson 1977, p. 675):

* = 0.10 significance level

** = 0.05 significance level

*** = 0.01 significance level
Table 2. Mean Absolute Percentage Error Comparison of CSH with Random Walk and Holt by Category, Group, and Time Series length for School District Case Study.

<table>
<thead>
<tr>
<th>Category</th>
<th>Short TS</th>
<th>Longer TS</th>
<th>Holt</th>
<th>Groups</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>CSH</th>
<th>5.7 ***</th>
<th>5.6</th>
<th>6.1 ***</th>
<th>6.9 ***</th>
<th>9.9</th>
<th>6.2</th>
<th>7.8</th>
<th>7.8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CSH</td>
<td>5.8 ***</td>
<td>5.7 ***</td>
<td>6.1 ***</td>
<td>6.9 ***</td>
<td>11.2</td>
<td></td>
<td>5.8 ***</td>
<td>5.7</td>
<td>6.1</td>
<td>6.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RW</td>
<td>7.5</td>
<td>9.4</td>
<td>9.9</td>
<td>11.2</td>
<td></td>
<td></td>
<td>7.0</td>
<td>4.8 ***</td>
<td>7.0</td>
<td>6.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CSH</td>
<td>9.3</td>
<td>7.2</td>
<td>9.0</td>
<td></td>
<td>6.9 **</td>
<td>9.1</td>
<td>7.2</td>
<td>6.9 **</td>
<td>15.5</td>
<td>10.5</td>
<td>8.7</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RW</td>
<td>5.7 ***</td>
<td>7.1 **</td>
<td>7.0</td>
<td>9.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Longer TS</td>
<td>CSH</td>
<td>11.1</td>
<td>7.6</td>
<td>7.2 ***</td>
<td>7.6</td>
<td>7.2 **</td>
<td>13.9</td>
<td>12.7</td>
<td>8.6</td>
<td>8.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RW</td>
<td>5.7 ***</td>
<td>7.1 ***</td>
<td>7.0</td>
<td>9.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category III</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CSH</td>
<td>21.7</td>
<td>10.5</td>
<td>9.1</td>
<td></td>
<td>10.5</td>
<td>12.6</td>
<td>10.5</td>
<td>9.1</td>
<td>24.3</td>
<td>12.5</td>
<td>12.9</td>
<td>11.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RW</td>
<td>7.7 ***</td>
<td>8.6 *</td>
<td>6.4 ***</td>
<td>8.6 *</td>
<td>10.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Longer TS</td>
<td>CSH</td>
<td>21.9</td>
<td>11.1</td>
<td>10.3</td>
<td></td>
<td>11.1 ***</td>
<td>11.1 ***</td>
<td>10.3</td>
<td>22.0</td>
<td>15.6</td>
<td>15.3</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RW</td>
<td>7.7 ***</td>
<td>12.8</td>
<td>8.6 *</td>
<td>6.4 ***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

§ Significant differences in mean accuracy are based on the Wilcoxon Matched Pairs Signed Rank two-tailed Test
(Pfaffenberger and Patterson 1977, p. 675): * = 0.10 significance level
** = 0.05 significance level
*** = 0.01 significance level
<table>
<thead>
<tr>
<th>Category I</th>
<th>Short TS</th>
<th>Longer TS</th>
<th>Holt</th>
<th>All Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CMSKF</td>
<td>RW</td>
<td>CMSKF</td>
<td>Holt</td>
</tr>
<tr>
<td>1</td>
<td>4.9 ***</td>
<td>6.8</td>
<td>4.9</td>
<td>3.9 ***</td>
</tr>
<tr>
<td>2</td>
<td>3.9 ***</td>
<td>6.3</td>
<td>4.6</td>
<td>3.7 ***</td>
</tr>
<tr>
<td>3</td>
<td>3.8 ***</td>
<td>6.2</td>
<td>4.2</td>
<td>3.7 ***</td>
</tr>
<tr>
<td>4</td>
<td>3.7 ***</td>
<td>6.2</td>
<td>4.3</td>
<td>3.7 ***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category II</th>
<th>Short TS</th>
<th>Longer TS</th>
<th>Holt</th>
<th>All Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CMSKF</td>
<td>RW</td>
<td>CMSKF</td>
<td>Holt</td>
</tr>
<tr>
<td>1</td>
<td>16.2</td>
<td>12.9 ***</td>
<td>16.2</td>
<td>9.4 ***</td>
</tr>
<tr>
<td>2</td>
<td>12.8</td>
<td>11.4</td>
<td>23.2</td>
<td>18.5</td>
</tr>
<tr>
<td>3</td>
<td>9.4 ***</td>
<td>12.2</td>
<td>18.5</td>
<td>27.3</td>
</tr>
<tr>
<td>4</td>
<td>8.8</td>
<td>11.4</td>
<td>20.2</td>
<td>27.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category III</th>
<th>Short TS</th>
<th>Longer TS</th>
<th>Holt</th>
<th>All Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CMSKF</td>
<td>RW</td>
<td>CMSKF</td>
<td>Holt</td>
</tr>
<tr>
<td>1</td>
<td>30.7</td>
<td>20.9 ***</td>
<td>30.7</td>
<td>13.4 **</td>
</tr>
<tr>
<td>2</td>
<td>26.2</td>
<td>21.6 **</td>
<td>26.2</td>
<td>12.2</td>
</tr>
<tr>
<td>3</td>
<td>16.7</td>
<td>18.5</td>
<td>25.6</td>
<td>18.5</td>
</tr>
<tr>
<td>4</td>
<td>13.4 **</td>
<td>18.5</td>
<td>34.6</td>
<td>37.8</td>
</tr>
</tbody>
</table>

$\text{Significant differences in mean accuracy are based on the Wilcoxon Matched Pairs Signed Rank two-tailed Test} (\text{Pflüger and Patterson 1977, p. 675}):$

* $= 0.10$ significance level
** $= 0.05$ significance level
*** $= 0.01$ significance level
Table 4. Mean Absolute Percentage Error Comparison of CSH with Random Walk and Holt by Category, Group, and Time Series length for the Monte Carlo Case Study.

<table>
<thead>
<tr>
<th>Category</th>
<th>Random Walk</th>
<th>Holt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Groups</td>
<td>Groups</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Category I</td>
<td>CSH</td>
<td>3.1 **</td>
</tr>
<tr>
<td></td>
<td>RW</td>
<td>6.8</td>
</tr>
<tr>
<td>Short TS</td>
<td>CSH</td>
<td>3.3 ***</td>
</tr>
<tr>
<td></td>
<td>RW</td>
<td>6.8</td>
</tr>
<tr>
<td>Longer TS</td>
<td>CSH</td>
<td>15.7</td>
</tr>
<tr>
<td></td>
<td>RW</td>
<td>12.9</td>
</tr>
<tr>
<td>Short TS</td>
<td>CSH</td>
<td>15.4</td>
</tr>
<tr>
<td></td>
<td>RW</td>
<td>12.9 **</td>
</tr>
<tr>
<td>Longer TS</td>
<td>CSH</td>
<td>23.4</td>
</tr>
<tr>
<td></td>
<td>RW</td>
<td>20.9</td>
</tr>
<tr>
<td>Category III</td>
<td>CSH</td>
<td>22.0</td>
</tr>
<tr>
<td></td>
<td>RW</td>
<td>20.9</td>
</tr>
</tbody>
</table>

$\$ Significant differences in mean accuracy are based on the Wilcoxon Matched Pairs Signed Rank two-tailed Test (Pfaffenberger and Patterson 1977, p. 675):

* = 0.10 significance level
** = 0.05 significance level
*** = 0.01 significance level
Fig. 1 Standardized Income Tax Revenue by Category (1972-1988).
Fig. 2  Types of structural change by Category
Fig. 3  Location of Structural Change in Category III by Group
Fig. 4 Monte Carlo Data by Category
Fig. 6 MPE Forecasting Accuracy for School District Case Study by Category, Group, and Time Series Length
Fig. 7  MAPE Forecasting Accuracy for Monte Carlo Design by Category, Group, and Time Series Length

Key:  
- - - - MSKF  
--- - - C-MSKF  
--- --- --- Holt’s  
--- --- - - CSH  
--- --- --- RW
Fig. 8  MPE Forecasting Accuracy for Monte Carlo Design by Category, Group, and Time Series Length