FORECASTING ANALOGOUS TIME SERIES

George Duncan
Wilpen Gorr*
Janusz Szczypula

H. John Heinz III School of Public Policy and Management,

Carnegie Mellon University, Pittsburgh, PA 15213-3890

*Contact:

E-mail: wg0g+@andrew.cmu.edu

Phone: (412) 268-8471

ABSTRACT

Organizations that use time series forecasting on a regular basis generally forecast many variables, such as demand for many products or services. Within the population of variables forecasted by an organization, we can expect that there will be groups of analogous time series that follow similar, time-based patterns. The co-variation of analogous time series is a largely untapped source of information that can improve forecast accuracy (and explainability).

This paper takes the Bayesian pooling approach to drawing information from analogous time series to model and forecast a given time series. Bayesian pooling uses data from analogous time series as multiple observations per time period in a group-level model. It then combines estimated parameters of the group model with conventional time series model parameters, using "shrinkage" weights estimated empirically from the data. Major benefits of this approach are that it 1) minimizes the number of parameters to be estimated (many other pooling approaches suffer from too many parameters to estimate), 2) builds on conventional time series models already familiar to forecasters, and 3) combines time series and cross-sectional perspectives in flexible and effective ways.

Provided are the necessary terms, concepts, and methods to understand Bayesian pooling and the conditions under which we can expect it to have comparative advantages over conventional time series methods. Useful for both practitioners and researchers are requirements stated on experimental data, treatments, and factors for comparative research on forecast accuracy of pooling methods. Lastly, the paper presents basic principles for applying pooling methods and supporting empirical results. The prospect for automatic pooling methods is good, although the best pooled forecasts at the current state of art will depend on expert judgment and manual interventions for time series that have frequent pattern changes.

Keywords: pooling, analogous time series, Bayesian methods, nonstationary time series

INTRODUCTION

Time series forecasting has the largest literature and number of applications of any approach to forecasting. Production planning, budgeting, inventory management, sales, marketing, and distribution all depend on accurate, short-term, time series forecasts. Many policy-level decisions in the energy, tourism, agricultural, and other areas depend on multivariate time series forecasts. While several avenues have been explored to improve time series forecasting, one of the most promising – pooling data from analogous time series – has had but scant attention (Fildes and Beard 1992).

Conventional time series methods – exponential smoothing, Kalman filters, Box Jenkins ARIMA methods, Census X11, Parzen's ARARMA method (Parzen 1982), multiple regression methods, etc. – all forecast single series in isolation. In contrast, nearly all organizations forecast multiple, analogous time series; e.g., similar products in the same geographic areas or the same products in different geographic areas. Often there are hundreds, thousands, and even tens of thousands of such time series that organizations forecast. We can expect analogous time series to co-vary and have similar time series patterns. Conventional series time methods do not use any of the information available from such analogous time series.

Some early attempts to incorporate data pooled from analogous time series into forecast models failed due to having too many parameters to estimate [Jones 1966] or capturing insufficient information from the analogous time series [Enns et al. 1982 as shown by Harvey 1986]. A class of pooling methods not well suited for time series forecasting are panel data methods (fixed and random effects models). These are models used by econometricians to control for nuisance cross-sectional variations while estimating multivariate causal models. Panel data models assume the coefficients of causal variables are constant across observational units. These models further assume that cross-sectional variations remaining after all causal model terms have been included in the model can be eliminated by adjustments only to the intercept term. In contrast, we expect coefficients for time trend, seasonality, or independent variables to vary from group to group within the population of time series to be forecasted.

An approach to using pooled data, which requires a large number of parameter estimates but nevertheless has been successful in forecasting, is Bayesian Vector Autoregressive models (BVAR). E.g., Lesage (1989) demonstrated the value of BVAR models for capturing leading indicator information from related geographic areas in forecasting. This paper does not pursue BVAR models, but instead examines Bayesian pooling models – also known as Bayesian shrinkage, Empirical Bayes, and Stein estimation. Bayesian pooling is applicable to univariate or multivariate time series models and has the advantages of requiring relatively few parameters for estimation, building directly on conventional time series models, being highly adaptive for nonstationary time series while also precise for stationary time series, and automatically switching from drawing on pooled data for precise estimation to rapid adjustment. BVAR models are more appropriate for macro-scale time series, but our interest has been in small-scale forecasting problems, where Bayesian pooling is more applicable.

The purposes of this paper are first to review concepts and methods for pooling time series and second to collect the limited (but promising) empirical results on pooling in terms of principles for forecasting. This paper focuses on univariate models, but also has many results applicable to multivariate models. Sections 2 and 3 provide background, terminology, and concepts for pooling time series. The fourth section describes Bayesian pooling and steps for its implementation. Section 5 provides principles for pooling and empirical support and Section 6 concludes the paper.

ANALOGOUS TIME SERIES

Often, groups of products (or services) are analogous in ways that make them follow similar time series patterns. For example, similar products may fall as a group within the same sphere of influence -- the same or similar consumer tastes, competition levels, local economic cycles, weather, regional trends, etc. -- therefore causing their time series to co-move (strongly correlate positively over time). We call such a collection an “equivalence group.” After standardizing each time series of an equivalence group to eliminate differences in magnitudes, etc., we can pool the time series by time period. The resulting pooled data have multiple data points per time period (one for each time series).

Spatial heterogeneity within the same geographic region can give rise to equivalence groups. For example, the time series trends of personal income in the 40 school districts of Allegheny County, Pennsylvania fall into three distinct groups according to diversity of local economies. One-mill towns are on one extreme and bedroom communities for white-collar workers are at the other extreme [Duncan et al 1993]. Economic cycles strongly affect the one-mill towns (causing turning points), but have only a slight impact on bedroom communities.

Spatial diffusion of innovations is a phenomenon that can yield equivalence groups with members widely distributed over space . For example, clothing fashions and disease epidemics start in major coastal cities in the U.S. (like New York City and Los Angeles), spread later to major inland cities (like Chicago and Washington DC), and later yet to tertiary cities (like Pittsburgh and Denver). Once in a city, cumulative growth at first is exponential, but later passes through an inflection point and eventually saturates. Such S-curve patterns may be dependent on population sizes and densities, with less intensity per capita as diffusion proceeds to lower classes of cities [Golub et al. 1993]. Sales of fashion clothing or the incidence level of infectious diseases can thus follow similar trends in disparate cities like Pittsburgh and Denver, and be members of the same equivalence group. (Furthermore, time series in Pittsburgh and Denver would have leading indicators series from Chicago and Washington DC.)

In summary, there are several phenomena and business practices that give rise to analogous products (and other analogous dependent variables) that organizations must forecast. The pooling methods discussed in this paper are intended, primarily, to draw on pooled time series data from analogous variables to improve forecast accuracy. Note that an additional benefit of pooling is forecast explanation, especially for univariate forecasts. Explaining a forecast can be aided by additional information from equivalence groups; e.g., that all or most members of analogous products are persisting in trends, have similar new trends, etc.

TIME SERIES VOLATILITY, PATTERN CHANGES, AND SCALE

This section defines terms and introduces concepts necessary for applying pooling methods. Bayesian pooling potentially has comparative advantages over conventional time series methods for forecasting time series that are volatile or nonstationary. Small-scale time series are most likely to suffer from volatility and nonstationarity.

Volatility refers to stationary time series with relatively large standard errors. For series

that do not have time trend or seasonality components, the coefficient of variation provides a useful index of volatility:

$$CV(x) = 100S / \bar{x}$$

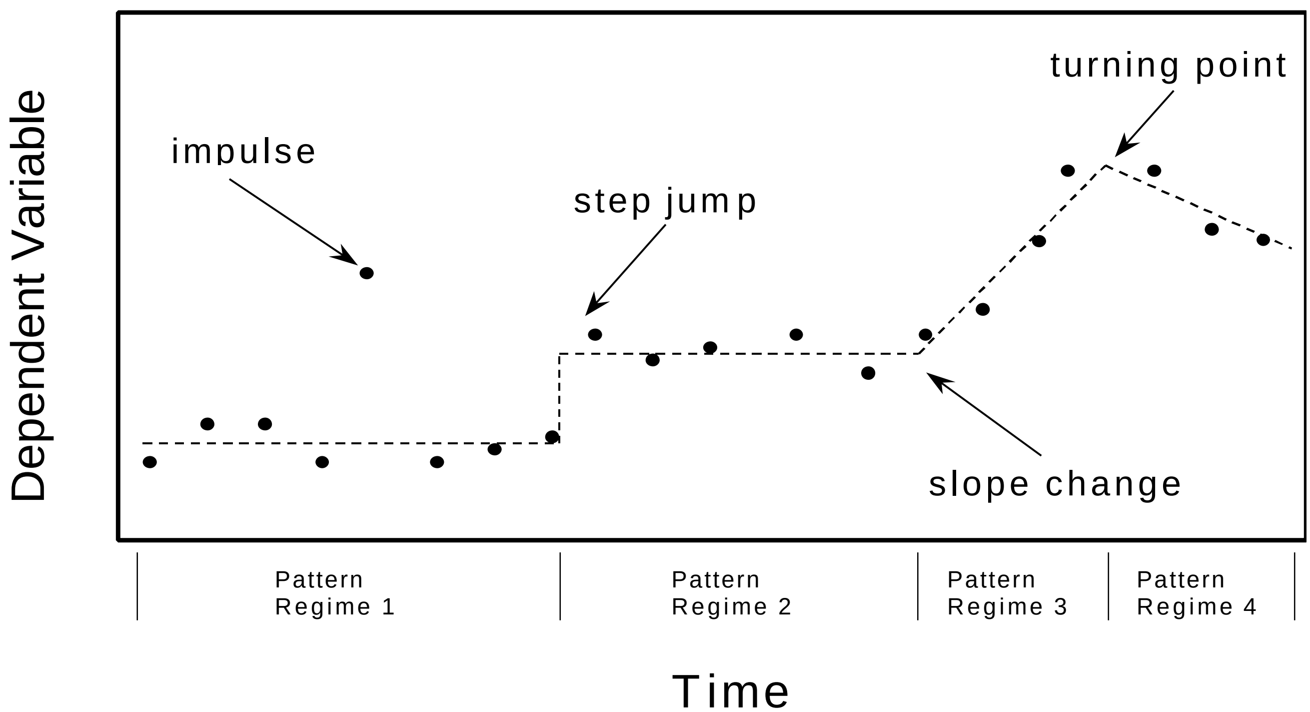
where x is the variable of interest, S is its sample standard deviation, and $\bar{x}$ is the sample mean.

For series with trend and/or seasonality, S can be replaced by the estimated standard error of an appropriate time series model. High values for this index indicate volatility and likely imprecise estimates from conventional time series models.

A time series pattern regime, or pattern regime for short, is an interval in which parameters of a time series model are constant. *A stationary time series* consists of a single pattern regime; whereas, *a nonstationary time series* has two or more pattern regimes. Exhibit 1 illustrates a nonstationary, nonseasonal, univariate time series with four pattern regimes defined by a step jump, time trend slope change, and a second time trend slope sign change (turning point). Also, more subtle patter change points, can be changes in error term parameters such as variance.

An *impulse* is an unusual, or outlier data point that occurs within a pattern regime. An impulse is not a pattern change, but merely an aberrant data value such as due a one-time shock to a system, a data collection error, or simply an extreme value occurring by chance. The estimation challenge presented by impulses is to ignore them (or screen them out) and not mistake them for pattern change points.

Exhibit 1
Illustration of Nonstationary Time Series with
Four Pattern Regimes



Note that while there is a large literature on detecting outliers in the quality control field [e.g., Fox 1972, Gardner 1983], Collopy and Armstrong (1992A) found few methods in the forecasting literature for handling outliers (and, to distinguish phenomena more finely, also for handling pattern regime changes). Of course, smoothing methods reduce the impact of outliers by damping adjustments to them ("smoothing"). Smoothing methods also allow parameter adjustments to accumulate ("drift") and catch up eventually with pattern regime changes. Lee (1990) provided a rule base for handling outliers and regime pattern changes based on quality control signals. Rule Based Forecasting [Collopy and Armstrong 1992B] has 11 rules for handling outliers out of a total of 99, based on expert judgment. The multi-state Kalman filter (MSKF) [Harrison and Stevens 1971, 1976], while having mixed results in forecasting competitions [Fildes 1983, Zellner 1986], has explicit and theoretically attractive mechanisms for modeling pattern regime changes. Duncan et al. (1993) suggested that the MSKF suffers from the large number of parameters needed for estimation. They therefore developed the cross-sectional MSKF using Bayesian pooling ("borrowing strength from neighbors") as a means to improve the precision of MSKF estimates.

Volatile and nonstationary time series can occur in any setting – small or macro scale – but are most characteristic of small-scale forecasting problems.

Small-scale time series have a relatively small number of individual transactions added up by time period (e.g., weeks, months), location (e.g., sales territory or municipality), and product/service category. Small-scale series in the private sector include stock keeping units of retail stores or warehouses, sales territory volumes, manufacturer's product and product family inventories, and firm-level sales. Small-scale time series in the public sector include administrative boundary totals (e.g., number of 911 calls by precincts of police departments), municipal totals, multi-municipality regions (like school districts and water districts), and counties.

Discrete (or special) events – price increases, competitors' promotional campaigns, opening of new shopping centers, etc. – play an important role in the small-scale time series. Impacts of discrete events that would average out in larger data aggregations (contributing to random noise in the model error term) instead can produce pattern regime changes in small-scale aggregations [Benjamin 1981, McLaughlin 1982, Lewandowski 1982, Benson 1983, Gorr 1986]. E.g., a shopping mall opening can cause a step jump increase in local income tax collections. A plant closing in a one-mill town can cause a downward turning point in local income tax collections.

Micro scale refers to aggregations so small that demand is intermittent with many periods having zero demand. Spare parts inventories and demand for certain big ticket items fall into this category. Simple exponential smoothing and Croton's smoothing are methods appropriate for the micro-scale setting [Willemain et al. 1994]. While pooling should provide increased precision for micro-scale time series models, we have not found any corresponding applications in the literature.

BAYESIAN POOLING

Bayesian pooling combines two models: a local model estimated for the target time series being forecasted and a group model estimated using the equivalence group's pooled data. Combination of the local and group models generally occurs at the parameter level (hence, the local and group models must have identical specifications) using "shrinkage" weights that have the effect of pulling (shrinking) local model parameter estimates to the central overall group estimates. The shrinkage weights are inversely proportional to the variance of parameter estimates they multiply, and they sum to one. Thus, if local model parameter estimates are more precise than corresponding group estimates, then more weight is placed on the local estimates, and visa versa. Bayesian pooling derives from maximum likelihood estimation of a hierarchical random effects model with distribution assumptions enabling Bayesian estimation [Duncan et al. 1993, Szczypula 1997].

For implementation, Bayesian pooling has the following steps: 1) *selection of an equivalence group* of analogous time series for the time series of interest (the *target series*), 2) *scaling each time series* to make pooled data homogeneous, 3) *construction of local and group models*: a conventional time series model for the target series and a pooled time series model for the group data, 4) *combination of local model and group model parameters* using Bayesian "shrinkage" weights, 5) *forecasting with the combined model*, and 6*) readjustment of target series forecasts* to raw data level. We proceed by explaining each step in detail.

1) Selection of an equivalence group – The first task is to identify analogous time series for pooling. The objective is to find time series that correlate highly over time (after synchronizing starting times if the series are not contemporaneous). There are three approaches: 1*) correlational co-movement grouping*: select time series that correlate highly with the target series; 2) *model-based clustering*: cluster time series using multivariate causal factors; and 3) *expert judgment*: have an expert use judgment for grouping. Zellner and Hong (1989), Greis and Gilstein (1991), Bunn and Vassilopoulos

(1993), Duncan et al. (1995b), and Szczypula (1997) have employed a variety of grouping approaches. Only Duncan et al. (1995b) and Szczypula (1997) have compared alternative grouping approaches experimentally.

Correlational co-movement grouping directly clusters series on the measure desired for pooling. Its danger is that series may correlate only by chance historically, or may react differently to changing environmental factors in the future. If grouped series diverge during forecast periods, then pooled forecast models will likely yield worse forecasts than conventional, single-series models.

Model-based clustering may use any one of several multivariate clustering methods commonly available in statistical packages (e.g., Johnson's hierarchical clustering). The variables used for clustering must yield equivalence groups with time series that co-move. For example, certain population density, age, income, and education ranges of populations may define sales territories with rapid versus low growth. The appeal of model-based clustering is twofold: it can use cross-sectional data not part of a time series, such as census data, and it can be based on theory and underlying causal relationships for co-movement. The danger of model-based clustering is that variables not used in clustering may also determine co-movement of time series.

Lastly, expert judgment may be the best approach in some settings. Experts may have the best insights into which series co-move [Szczypula 1997]. In practice, the most attractive approach to grouping may be a combination expert judgment or model-based clustering, followed by correlational co-movement check to remove some series from an equivalence group. Expert judgment and model-based clustering will identify potential equivalent groups that make sense theoretically and on other grounds, and are more likely to continue to co-move during forecast periods. Correlational co-movement will ensure that the theoretically grouped series actually have the same patterns.

Note that there is a fourth approach for grouping – simply pool all available time series. Total population pooling provides the strawman comparison for the first three approaches. Correlational co-movement grouping, model-based clustering, and expert judgment must yield overall more accurate forecasts than total population pooling to merit use in pooling.

2) Scaling each time series – Ideally, the pooled data from an equivalence group would form a random sample from a single stochastic process with independent and identically distributed, normal error terms. Theoretically, the conditionally independent hierarchical model [Kass and Steffey 1989] provides the needed formulation, including hyperparameters in a hierarchical random effects model.

For implementation, we can assist by homogenizing time series in various ways, to remove differences in magnitudes and variances. One alternative is to simply standardize each time series (subtract its sample mean and divide by its sample standard deviation) [Duncan, Gorr, and Szczypula (1993, 1994, 1995a, 1995b) and Szczypula (1997)]. Another (not yet applied, to our knowledge) would be to regress the target time series on each equivalence group time series, and use estimated target series values as standardized data. Note that standardized data need to be recalculated each time new data become available. For multivariate time series models, independent variables may be included that scale the dependent variable (e.g., population or total disposable income of sales territories. One other approach, appropriate for both univariate or multivariate applications is to use dimensionless dependent variables; e.g., Greis and Gilstein (1991) uses percentages of totals and Zellner and Hong (1989) use percentage growth rates.

3) Construction of local and group models – For multivariate time series models, this is a simple task. One uses the same model specification (e.g., linear in total population, per capita income, and marketing expenditures) for both the local model of the target series and the group model. The local model estimation uses only the target observation unit's time series. The group model uses the pooled data of the equivalence group.

Adaptive Bayesian pooling (ABP) [Duncan, Gorr, and Szczypula 1993, 1994, 1995a, 1995b; Szczypula 1997] uses the current level and time trend slope formulation of exponential smoothing models for its local univariate model. Instead of including an intercept term, this formulation recursively adjusts the time series level (current mean) so that, in effect, the intercept becomes the last historical time period. The local model for ABP is a univariate time series model that includes recursive updating of model parameters; e.g., exponential smoothing, Kalman filter, or multi-state Kalman filter. Estimated variances for local model parameters are also updated recursively using simple exponential smoothing. Lastly, ABP uses a short-memory group model. Duncan et al. (1993) use the sample mean of the pooled data's last historical period as an estimate of the level and the sample mean of the first differences of each local series as an estimate of the time trend slope. Similarly, the sample variances of these two estimates are used in step 4.

4) Combination of local model and group model parameters - At the heart of Bayesian pooling are "shrinkage" formulas that yield weights for combining local and group parameter estimates. These weights are inversely proportional to variances of estimated parameters. Below are empirical Bayes shrinkage calculations for the case of nonseasonal, univariate forecasts (shrinkage formulae for multivariate models are analogous to those of (1) and (2) [Zellner and Hong 1989]):

$$L'_{it} = u_1 L_{it} + u_2 \bar{x}_t \quad (1)$$

$$S'_{it} = w_1 S_{it} + w_2 \bar{\Delta}_t \quad (2)$$

where

i = target time series index

L = estimated level from the local model

S = estimated time trend slope from the univariate time series model

$\bar{x}$ = sample mean of group pooled data

$\bar{\Delta}$ = sample mean of group first differences

u_1, u_2 = shrinkage weights summing to 1.0 with u_1 inversely proportional to the estimated variance of L and u_2 inversely proportional to the estimated variance of $\bar{x}$

w_1, w_2 = shrinkage weights summing to 1.0 with w_1 inversely proportional to the estimated variance of S and w_2 inversely proportional to the estimated variance of $\bar{\Delta}$

L'_{it} = combined, final level estimate for target series i at forecast origin t

S'_{it} = combined, final slope estimate for target series i at forecast origin t.

Traditionally, Bayesian shrinkage has been used to improve the precision of estimates for volatile, but stationary time series. The same mechanism, implemented with shrinkage at each time period recursively and a short-memory group model, is also capable of rapid adjustment to time series pattern changes. ABP smoothes the estimated variances making up the shrinkage weights of (1) and (2). When a nonstationary period is entered, the parameters of the local model become imprecise, as reflected in large residuals and higher parameter variances. At the same time, the short-memory group model may suffer no or little increased variability, if all member time series of the equivalence group continue to co-move. The net effect is to increase the weights on the group components of estimates and decrease weights on the local components. The intended result is rapid and accurate adjustment to new pattern regimes.

It is easy to extend shrinkage formulas as in (1) and (2) to univariate time series with seasonality. The local model would be a conventional time series model including seasonal factors such as Winters. The group model, e.g., would average ratios of data points for each season to a group moving average to estimate group seasonal factors.

5) Forecasting with the combined model – The k-step-ahead forecast for the model in (1) and (2) is simply :

$$Y^{F}_{t+k} = L'_{it} + kS'_{it}$$

6*) Readjustment of the target series forecasts* – If the target series was transformed in step 2, the process needs to be reversed, as a final step, to produce forecasts at the raw data level.

EXPERIMENTAL DESIGNS

Designs for assessing pooled forecasting methods have additional requirements beyond those for conventional time series methods. We believe this to be the first attempt to specify such experimental designs. It may not be surprising then, that none of the available empirical studies on pooling methods include all data types, treatments, and factors discussed below as would be desired.

Forecasting comparisons for pooled forecasting methods require organization-based data, and not the random, unrelated time series used in past forecasting competitions (e.g., Makridakis et al. 1982). Time series are needed for analogous products and services that an organization forecasts. Production management and budgeting applications require contemporaneous time series reflecting effects of common environmental and controllable influences; e.g., regional economic cycles, marketing expenditure levels, competition levels etc. New product forecasting requires data banks of historical time series, with attributes on new products, environmental conditions, and management actions carefully recorded, for use in early phase planning and forecasting of new products. Note that while Mahajan and Wind (1988) discuss analogous products for use in forecasting new products, we have found no papers assessing forecast accuracy of such an approach.

Factors to include in experiments are: 1) level of volatility, 2) level of pattern change such as step jump, time trend slope change, or turning points at the ends of time series used for estimation, and 3) level of outlying with isolated outliers at the ends of time series. We expect comparative advantages for pooling with high levels of volatility, large pattern changes, and large outliers at the ends of times series.

Several treatments need to be included for empirical studies evaluating pooling methods. Treatments should include 1) the random walk as a straw man method; 2) local, group, and pooled versions of time series methods; and 3) comparison of total population, expert judgment, correlational co-movement, and model-based clustering grouping methods.

The random walk is often the best forecast method during early stages of a pattern regime change, because it is completely reactive. Other methods need to process signals from the data to discount historical data from before pattern changes, causing lags in responses and higher forecast errors. During steady time trends, however, the random walk will often be the worst forecast method, because it does not include a trend forecast component. An alternative to the random walk, is ABP's group model mean, which is the sample mean for the equivalence group from the last historical period.

Bayesian pooling combines parameter estimates of conventional time series methods used for local model estimation with parameter estimates for group models based on pooled data. Hence experiments have a natural basis upon which to assess the value added by pooling: compare forecast accuracy of the conventional time series methods used in estimating local models to forecast accuracy of the pooled methods. Of course, additional conventional time series methods may also be relevant for comparison purposes as well. Lastly, the group model, by itself without combination to the local model, may also be a competitive forecasting method [Greis and Gilstein 1991, Bunn and Vassilopoulos 1993].

Grouping the time series that an organization forecasts into analogous groups is a key step of pooling. Comparison of the major approaches need to be made: expert judgment, correlational co-movement, and model-based clustering. The strawman comparison for grouping methods is total population pooling, at least in organizations where the number of series to be forecasted is not very large. Pooling using groups needs to show improvements over pooling over the total population.

GUIDELINES

Many of the principles provided in this section seem obvious – after seeing them in print. Pooling as an area of forecasting has not had systematic development nor does it have a wide literature. Hence, after committing experimental design requirements to paper, in the previous section, we find that even our own research has missing components and considerations. We therefore begin at the beginning and put forward some basic principles: *volatility, outlier screening, clustering, simple pooling methods, and expert judgment for shrinkage*.

Volatility: *The higher the volatility of time series, the higher the improvement in forecast accuracy of pooling methods over conventional time series methods.* Borrowing strength from neighbors is the traditional purpose of Bayesian pooling. The pooled data model provides additional data, extending the sample size, thereby increasing the precision of model parameter estimates. Hence, pooling should have relatively higher payoffs for more noisy variables, as measured by the coefficient of variation . Few studies comparing pooled versus conventional time series forecasts have level of volatility reported. For some that do not, we can at least rank results by volatility or indicate that time series used should have been volatile.

First, Duncan, Gorr, and Szczypula (1995b) included volatility measures in their results. They forecasted annual total non-white live births, non-white infant deaths, and non-white infant mortality rate calculated as non-white infant deaths per 1000 non-white live births. These time series consist of annual data for 90 Pittsburgh neighborhoods for 1980 through 1992. The corresponding time series data are highly volatile with no steady trends and with coefficients of variation ranging from 60 to 308. Conventional univariate time series methods included the random walk, Holt, and time regression. The ABP method was the Cross-Sectional multi-state Kalman filter [Duncan et al. 1993] with total population pooling. One-year-ahead forecasts were calculated for each of the three variables using a rolling horizon design for 1987 through 1992. Expert judgment, correlational co-movement grouping, and model-based clustering were not able to improve forecast accuracy over total population pooling. Similarly, no method dominated within the set of conventional time series, hence their performance was averaged. Compared in Exhibit 2 are the percentage improvement in average mean absolute error (MAE) of the ABP method over conventional univariate methods. Results indicate an order of magnitude increase in forecast accuracy for the more volatile series.

Exhibit 2
Improvements of Adaptive Bayesian Pooling Over
Conventional Univariate Time Series for Vital Statistics with
Varying Levels of Volatility.

Variable	Coefficient of Variation	Percentage Improvement in MAE of ABP Method Over Conventional Time Series Methods
Births	60.0	2.1
Deaths	134.0	17.4
Mortality Rate	308.0	16.2

Greis and Gilstein (1991) compared pooled versus univariate forecasts for annual percentage of telephone circuit churn (percentage of circuits disconnected and then reconnected) for 939 wire centers of two telecommunications companies. The modal wire center had under 50 circuits, while the largest size category used was 1,000 or more circuits. Data consisted of five years of annual data by wire center. Even though many wire centers had increasing trends, the simple average of the first four years was used as the forecast of the fifth year. The group model was the four-year grand average by company or size range of wire center.

Exhibit 3 contains root mean squared forecast errors aggregated by company. Column 4 shows the percentage improvement in forecast accuracy due to using the company group model over individual wire center models: 17.2 percent for company 1 and 67.1 percent for company 2. Column 6 shows that Bayesian pooling of wire center and company models adds little more to accuracy improvements over the local models, which now stand at 20.0 percent and 67.6 percent for the two companies. Overall, the results demonstrate the benefits of using pooled data for local-level forecasting, even though Bayesian pooling does not make substantial improvements beyond the group model. A breakout of results by wire center size is not revealing.

Exhibit 3
Improvements in Forecast Accuracy due to Group and Pooled Models
for Percentage Churn Rates at Wire Centers.

(1) Data Source	(2) Local Wire Center Univariate RMSE	(3) Group Company Model RMSE	(4) Percent Improvement of Group Over Local Model	(5) Bayesian Pooling RMSE	(6) Percent Improvement of Bayesian Pooling Over Local Model
Company 1	11.7	9.7	17.2	9.3	20.0
Company 2	54.5	17.9	67.1	17.6	67.6

One promising area for borrowing strength is for estimating seasonal factors of time series models. Data for a particular seasonal factor is observed only once per complete cycle (e.g., once a year for monthly data); hence, seasonal factor estimates will often lack precision. Traditional pooling provides a means for increasing precision of seasonal factors. Bunn and Vassilopoulos (1993) compared group and local models (but not pooled models) for 12 groups of products consisting of 54 series from a UK chain of department stores. The series consisted of 4-week monthly sales volume, were highly seasonal, had slight increasing trends, covered the period 1/88 through 6/91, and were screened to eliminate nonstationary series. The forecast models consisted of two-parameter exponential smoothing on deseasonalized data, with seasonal factors calculated from individual series versus groups. The 12 groups were obtained by multivariate clustering of local seasonal factors. The overall improvement in mean squared error forecast accuracy due to grouped over local seasonal factors was modest, 6 percent.

Lastly, Duncan, Gorr, and Szczypula (1993) carried out forecast experiments using annual personal income tax collections from forty school districts in Allegheny County, Pennsylvania over a seventeen-year period, from 1972 to 1988. Expert judgment, based on diversity of local economies, led to three groups,: low, medium, and high diversity. No volatility statistics are available, however, visual inspection of time series plots indicate that the high economic diversity group has the least volatile time series and the low diversity group the highest. Forecast methods included the univariate multi-state Kalman filter and corresponding ABP approach, the cross-sectional multi-state Kalman filter. One, two, and three year ahead forecasts were made from 1978 using 4, 5, 6, and 7 historical data points. The entire period from 1972 through 1981 had stationary time series, thus any improvements in forecast accuracy must be from “borrowing strength” and increased precision of estimates. The percentage reduction in mean absolute percentage forecast error of the pooled model over the local model averaged 17.1 percent for the high diversity group, 29.3 percent for the medium diversity group, and 29.2 percent for the high diversity group, supporting the volatility principle. In summary, the limited evidence available supports the principle that pooling can improve forecast accuracy over conventional time series models, and more so with increasing time series volatility.

Outlier Screening: *Pooling models screen out (minimize) the distortions in time series model estimates caused by outlier data points.* Recursive time series methods, like exponential smoothing, react to outlier data points and erroneously adjust model estimates. Thus outliers can cause high forecast errors, especially when outliers are at the end of the historical or estimation time series. After a period of time, smoothing models can recover, forget the false signals sent by outliers, and return to the correct trend.

The traditional approach to outlier detection is to calculate a threshold value, based on smoothed estimates of a time series standard error [Fox 1972]. If a model residual exceeds the threshold, then an outlier is signaled (with certain type 1 and 2 error rates). Pooling provides a new approach to the literature on detecting outliers. With pooled data, a data point in the target series is a potential outlier if it is the only unusual data point at that time in the equivalence group. Bayesian pooling methods implicitly use this approach. If the sphere of influence of the event causing the impulse includes all or many members of the equivalence group, then adaptive pooling will make large adjustments in the direction of the impulse, in error.

A Monte Carlo study provide some clear evidence for the outlier principle. Exhibit 4 contains one-step-ahead forecast performance (mean absolute percentage errors) for Holt smoothing versus its ABP counterpart, cross-sectional Holt smoothing [Duncan et al. 1994], Each simulated time series has a positive time trend; 12 time series per equivalence group; single outlier data point as the last historical data point of the target time series, second to last historical data point, and up through the 4th to last data point; and 1,000 replications. Outliers varied from 3, to 5, and to 7 standard error deviations from the error term. Exhibit 4 has the results.

Exhibit 4
Monte Carlo Results on Outlier Screening by Holt Smoothing versus
Cross-Sectional Holt Smoothing (Mean Absolute Percentage Forecast Error).

Case	Forecast Method	Last Data Point is Outlier	2nd to Last Data Point is Outlier	3rd to Last Data Point is Outlier	4th to Last Data Point is Outlier
3 Sigma Outlier	Cross-Sectional Holt	3.4	2.8	2.3	1.9
	Holt	7.0	4.0	4.0	3.0
5 Sigma Outlier	Cross-Sectional Holt	3.0	3.5	1.8	2.1
	Holt	10.3	7.2	5.3	4.2
7 Sigma Outlier	Cross-Sectional Holt	3.1	2.4	2.2	1.9
	Holt	8.2	5.1	5.3	5.3

Clearly, Exhibit 4 shows that the pooling method successfully uses information from equivalence groups to screen the outliers. For example, when the target series has an outlier as the last historical data point, pooling has a forecast MAPE averaging 3.2 whereas Holt averages 8.5. By the time that the outlier is the 3rd to last data point, the pooling method forecast MAPE averages 2.1 and has "forgotten" the outlier completely, while Holt averages 4.9 and still "remembers" the outlier.

Note that these results on outliers have implications for estimating seasonal factors. They suggest that an aberrant value, that might heavily impact and distort a seasonal factor, can be effectively screened from a target time series. To accomplish screening, group models for seasonal factors as used by Bunn and Vassilopoulos (1993) would need to be extended to include pooling.

Clustering: *Pooling within clustered groups improves forecast accuracy over total sample pooling when there are large differences in time series patterns across groups and strong co-movement within groups.* This principle merely echoes properties of good clusters – large differences between clusters and large similarities within clusters. The school district revenue case discussed earlier has examples of both good and bad clusters. All clusters examined have strong co-movement with clusters, but for one forecast origin (1978) there are no between cluster differences and for a second origin (1983) there are major differences between clusters. During stable periods, all 40 time series of the case had similar growth. During pattern regime changes, there is also strong co-movement within groups, but each group has a different trend. After an economic downturn, the high economic diversity group had mildly decreased growth for a few years, the medium diversity group had growth flattened to a persistent no growth time series, and the low diversity group had sharp and dramatic downward turning points. For the first forecast origin, we thus expect total population pooling to be better than grouped pooling, but for the second origin we expect the opposite. These expectations are borne out in Exhibit 5 where we see that grouped pooling averages 5 percent worse than total population pooling when there is no difference between groups, but is from 34 to 51 percent better when there are large inter-group differences.

Exhibit 5
Percentage Improvement in MAPE Forecast Accuracy
of Grouped Over Total Population Pooling

Forecast Lead Time	Percent Improvement When No Differences in Groups: Stable Period (1972-1978 estimate, 1979-1981 forecast)	Percent Improvement When Large Difference in Groups: Pattern Change Period (1976-1983 estimate, 1984-1986 forecast)
One Year Ahead	-4	34
Two Years Ahead	-6	43
Three Years Ahead	-5	51

The infant mortality case, discussed earlier, has no consistent groups, whether designed by correlational co-movement, model-based clustering, or expert judgment (in the form of areas for programs designed to reduce infant mortality rates). While Pittsburgh has poverty and minorities in highly concentrated pockets (factors leading to high infant mortality), there is no consistent co-movement of neighborhoods' time series within clusters and forecast periods. Neighborhoods which had co-moved in equivalence groups during estimation periods frequently diverged in forecast periods. The impact on pooling is negative in such cases, because the cross-sectional means of group models are misleading. For example, if some series of an equivalence group increase but others decrease in forecast periods, the cross-sectional mean may have a value between the two divergent subsets of series and not be representative of any series. Then each time series is shrunk to an erroneous mean! Protecting pooling in such cases, however, is that the variance of the group model parameters is high. Hence pooling places more weight on the target series' univariate method – essentially zeroing out cross-sectional information. The end result is no gain from grouped pooling. Nevertheless, total sample pooling improved forecast accuracy over conventional time series methods as shown in Exhibit 2 above.

Simple Pooling Methods: *Use simple pooling methods and simple grouping methods. The limited evidence suggests that the best univariate pooling method is Cross-Sectional Holt and the best grouping methods are expert judgment or correlational co-movement.* The multi-state Kalman filter (MSKF) is a sophisticated univariate Bayesian method for modeling and forecasting nonstationary time series. The Cross-sectional MSKF, an adaptive pooling method, substantially improves the performance of the MSKF. Holt exponential smoothing is also an adaptive univariate time series method, but a very simple one. Likewise cross-sectional Holt is a simple pooling method. Lastly, model-based clustering is the more sophisticated than either expert judgment or correlational co-movement clustering.

Exhibit 6 provides evidence the simple methods are best for pooling. The cross-sectional Holt method dominates the cross-sectional MSKF. Furthermore, expert judgement and correlational co-movement clustering dominate model-based clustering.

Exhibit 6
Comparison of Sophisticated Versus Simple Methods for Pooling: Forecast MAPE for Rolling One-Year Ahead Forecasts for 1983 Through 1988

Type of Pooling	Cross-Sectional Holt MAPE	Cross-Sectional Multi-State Kalman Filter MAPE
Total Sample	6.1	11.4
Model-Based	6.3	10.7
Co-Movement	5.5	7.7
Expert Judgement	5.4	7.8

Expert Judgment for Shrinkage: *If the number of time series is not too large, we suggest monitoring times series and manually intervening to switch shrinkage weights for pattern changes.* Well-designed equivalence groups have the correct information for quickly identifying and accurately estimating new pattern regimes in univariate time series. The smoothed variances used in ABP methods as yet do not react quickly enough to switch weights from univariate time series models to the short-term group models. Hence, using either expert judgment or a rule base, at this time, will minimize estimation lags during pattern changes. Bretschneider and Gorr (1997) provide graphical methods, simple time series methods, and examples that illustrate judgmental adjustments.

Evidence supporting this last principle is from Monte Carlo experiments (which have the same overall design as those for the outlier principle). Results reported below are aggregated over low, medium, and high change cases for slope and step jump changes and various forecast origins. Forecast origins 1 and 2 are forecasts made 1 and 2 periods after a pattern change has occurred. Origins 3 and 4 are forecasts made 3 and 4 periods after a pattern change. The statistics reported are ratios of cross-sectional Holt forecast MAPE divided by random walk forecast MAPE. Ratios less than 1.0 favor the cross-sectional Holt method. Clearly, the random walk is best for origins 1 and 2 for both step jumps and slope changes, but the cross-sectional Holt is best for step and slope changes after origin 2. These results indicate that automatic pooling is not able to respond as quickly as the random walk.

Exhibit 7
Evidence that Automatic Adaptive Bayesian Pooling Does Not Adapt Quickly Enough.

Case	Cross-Sectional Holt MAPE ÷ Random Walk MAPE	
	Origins 1&2	*Origins 3&4*
Slope Change	1.25	0.71
Step Jump	4.07	0.92

CONCLUSION

Organizations need to start using the valuable information contained in the time series of analogous products, services, and other variables that they forecast. This paper has presented methods and principles for Bayesian pooling of analogous univariate time series. Empirical results provide guidelines on settings in which pooling has comparative advantages and best methods for pooling.

Future work has some clear needs. First, we are developing new forecasting software for univariate pooling. This software will allow further research on a number of topics including: 1) tuning smoothing factors for parameter variance estimates to increase responsiveness while maintaining reliability, 2) using rule bases to switch between time series and group models during pattern change periods, 3) heuristics for forecasting in the beginning of new pattern regimes when insufficient data exists to distinguish between step jumps and time trend slope changes, 4) implementation of pooling of seasonal factors, jointly estimated with time trends, and 5) reclustering groups at each forecast origin.

Second, future forecast competitions need to include analogous time series, and not just the isolated time series of past competitions. Competitions with multiple time series would most likely have to be case based, using data from a small number of organizations. Lastly, researchers need to carefully consider requirements for experimental designs, such as we have specified in this paper, for assessing pooled forecasting methods. Comparative research is needed on alternative grouping methods.

REFERENCES

Armstrong, J.S. (1985), *Long-Range Forecasting*, 2nd Edition, Wiley-Interscience, New York, 171.

Benjamin, R.I. (1981), Information technology in the 1990's: a long range planning scenario, *MIS Quarterly* 6, 11-31.

Bretschneider, S.I. and W. Gorr, Projecting Revenues, in Meyers, R.T, and A. Schick [ed.], *Handbook of Government Budgeting*, Jossey-Bass Inc., in press.

Bunn, D.W. and A.I. Vassilopoulos (1993), Using group seasonal indices in multi-item short-term forecasting, *International Journal of Forecasting*, 9, 517-526.

Collopy, F. and J.S. Armstrong (1992a) Expert opinions about extrapolation and the mystery of the overlooked discontinuities, *International Journal of Forecasting*, 8, 575-582.

Collopy, F. and J.S. Armstrong (1992b), Rule-based forecasting: development and validation of an expert systems approach to combining time series extrapolations, *Management Science*, 38.

Duncan, G., W. Gorr, and J. Szczypula (1993), Bayesian forecasting for seemingly unrelated time series: application to local government revenue forecasting, *Management Science* 39, 275-293.

Duncan, G., W. Gorr, and J. Szczypula (1994), comparative study of cross-sectional methods for time series with structural changes, Working Paper 94-23, Heinz School, Carnegie Mellon University.

Duncan, G., W. Gorr, and J. Szczypula (1995a), Bayesian hierarchical forecasts for dynamic systems: case study on backcasting school district income tax revenues, in Luc Anselin and Raymond Florax (eds.), *New Directions in Spatial Econometrics* (Springer), 322-358.

Duncan, G., W. Gorr, and J. Szczypula (1995b), Adaptive Bayesian Pooling Methods: Comparative Study on Forecasting Small Area Infant Mortality Rates, Working Paper 95-7, Heinz School, Carnegie Mellon University.

Enns, P.G., J.A. Machak, W.A. Spivey, and W.J. Wroblewski (1982), Forecasting applications of an adaptive multiple exponential smoothing model, *Management Science* 28, 1035-1044.

Fildes, R. (1983), An evaluation of bayesian forecasting, *Journal of Forecasting* 2, 137-150.

Fildes, R, and C. Beard (1992), Forecasting systems for production and inventory control, *International Journal of Operations and Production Management* 12, 4-27.

Fox. A.J. (1972) , Outliers in time series, *J. Royal Stat. Soc. Series B* 34, 350-363.

Gardner, E.S. (1983), Automatic monitoring of forecast errors, *Journal of Forecasting* 2, 1-21.

Golub, A., W. Gorr, and P. Gould (1993), Spatial diffusion of the hiv/aids epidemic: modeling implications and case study of aids incidence in Ohio, *Geographical Analysis* 25, 85-100.

Gorr, W. (1986), Special event data in shared databases, MIS *Quarterly*, 10, 239-250.

Greis, N.P. and Z. Gilstein (1991), Empirical Bayes methods for telecommunications forecasting, *International Journal of Forecasting*, 7, 183-197.

Harrison, P.J. and C.F. Stevens (1971), A Bayesian approach to short-term forecasting, *Operations Research Quarterly* 22, 341-362.

Harrison, P.J. and C.F. Stevens (1976), Bayesian Forecasting, *Journal of the Royal Statistical Society Series B* 38, 205-247.

Harvey, A.C. (1986), Analysis and generalization of a multivariate exponential smoothing model, *Management Science* 32, 374-380.

Jones, R.H. (1966), Exponential smoothing for multivariate time series, *Journal of the Royal Statistical Society Series B* 28, 241-251.

Kass, R.E. and D. Steffey (1989), Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models), *Journal of the American Statistical Association* 84, 717-726.

Lee, Y. (1990), *Conceptual design of a decision support system for local government revenue forecasting*, unpublished Ph.D. Dissertation, School of Urban and Public Affairs, Carnegie Mellon University.

Lesage, J.P. (1989), Incorporating regional wage relations in local forecasting models with a Bayesian prior, *International Journal of Forecasting* 5, 37-47.

Lewandowski, R. (1982), Sales forecasting by FORSYS, *Journal of Forecasting* 1, 205-214.

Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler (1982), The accuracy of extrapolation (time series) methods: results of forecasting competition, *Journal of Forecasting* 1, 111-153.

McLaughlin, R.L. (1982), A model of average recession and recovery, *Journal of Forecasting* 1, 55-65.

Parzen, E. (1982), ARARMA Models for time series analysis and forecasting, *Journal of Forecasting* 1, 67-82.

Szczypula, J. (1997), *Adaptive hierarchical pooling methods for univariate time series forecasting*, unpublished Ph.D. dissertation, Carnegie Mellon University.

Trigg, D. W. and D. H. Leach (1967), Exponential smoothing with an adaptive response rate, *Operations Research Quarterly* 18, 53-59.

Willemain, T.R, C.N. Smart, J.H. Shockor, and P.A. DeSautels (1994), Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method, International Journal of Forecasting 10, 529-538.

Zellner, A. (1986), A tale of forecasting 1001 series: The Bayesian knight strikes again, *International Journal of Forecasting* 2, 491-494.

Zellner, A. and C. Hong (1989), Forecasting international growth rates using Bayesian shrinkage and other procedures, *Journal of Econometrics* 40, 183-202.