ESTIMATING TIME SERIES MODELS FOR EMERGENCY ROOM EPISODES INVOLVING DRUGS

Charles K. Fairchild, CSR, Incorporated, Arthur L. Hughes, NIDA, and Richard E. Gruberg, CSR, Incorporated
Charles K. Fairchild, 1400 Eye St., N.W., Suite 300, Washington, DC, 20005

Key Words: Time Series, Hospital Emergency Rooms, Drug Abuse Data

1 Introduction

The Drug Abuse Warning Network (DAWN) is a reporting system designed to be an early warning indicator of the nature and extent of the drug abuse problem in the United States. Data on hospital emergency room (ER) episodes involving the abuse of licit and illicit drugs are processed monthly by the Substance Abuse and Mental Health Services Administration (SAMHSA). The secular trends and seasonal variations have long been important issues requiring attention when examining data collected from the DAWN ER sample. The purpose of this research is to apply time series analysis methods to the DAWN data to identify seasonal patterns, trends, or cycles in the data. The research reported in this paper led to the development of seasonal adjustment factors, examined the auto-correlation structure of the data, and developed a time series model.

2 Background of the DAWN Survey

The objectives of the DAWN survey are to identify drugs and substances currently being abused and to provide demographic and other data pertaining to drug abuse for national and local drug abuse policy planning. For this survey, "drug abuse is defined as the non-medical use of a drug or substance for psychic effect, for suicide attempt or gesture, or because of dependence" (National Institute on Drug Abuse [NIDA], 1992. p. 7). Participating hospitals submit data to SAMHSA from which weighted estimates of the total number of emergency room visits involving drug abuse (referred to as episodes) are produced, as well as the total number of ER visits involving the abuse of specific drugs (referred to as drug mentions). These estimates currently are produced for the nation as a whole as well as for 21 specific metropolitan areas. For most all common drugs, DAWN estimates also are broken down by demographic characteristics (i.e., age, race, and sex) as well as by drug use motive, reason for ER contact, source of procurement, form in which drug was acquired, and route of administration of the drug (NIDA, 1992).

2.1 Sample Design

DAWN data have been collected and reported from a sample of hospital ER's since 1972. The original sample was representative but the representativeness was not maintained as ER's left the sample. The sample was redesigned in 1982 as a probability sample representative of all ER's, and estimation procedures based on the new design began with the 1988 DAWN data. Hospitals eligible for selection into the DAWN survey must be non-Federal short-stay general surgical and medical hospitals in the coterminous United States having at least one 24-hour ER.

The new DAWN sample consists of a panel of approximately 500 hospitals located throughout the coterminous United States. The panel sample has virtually 100-percent overlap from month to month, with exceptions due primarily to sample attrition (non-response and ER closings) and the recruitment of newly eligible hospitals. The current sample has a 50-percent overlap with the hospital sample used prior to 1988, and preliminary weighted estimates have been produced for the 1978-87 period.

The new survey design incorporates a combined ratio estimator for all metropolitan area and national panel estimates. With this approach, each survey estimate is "benchmarked" by the ratio of the total number of hospital ER visits occurring during the period in the estimation region (obtained from the American Hospital Association) to the estimated number of ER visits obtained from the sample. NIDA's 1990 Annual Report was the first DAWN publication to be based completely on weighted estimates from the new sample.

2.2 Limitations of the DAWN Survey

Data from each hospital ER are collected by a DAWN reporter (usually a member of the hospital clerical staff) who reviews medical records for mentions of drug abuse. It should be noted that the DAWN data reflect mostly drug use that is self-reported by ER patients, and data may not be available for all patients entering the ER. In this regard, the DAWN survey has recently been criticized for underreporting drug abuse episodes associated with major trauma, such as motor vehicle accidents and violent assaults (Brookoff, Campbell, and Shaw, 1993). Such underreporting stems from the fact that trauma patients are often not in a position to give a history of drug abuse while being treated in the ER. Furthermore, medical records generated after the patient has left the ER (which may contain toxicology testing results) generally are not available to the DAWN reporter.

National estimates measuring ER drug abuse episodes from the DAWN survey are in large part driven by the data collected from the national panel. Since the majority of these participating hospitals are located in suburban areas, their response characteristics differ in significant ways from many of the hospitals located in the 21 metropolitan areas. Drug abuse episodes from the national panel are more likely to involve licit drugs, and less likely to involve such illicit drugs as crack-cocaine and heroin. As a result, the variability of the data representing illicit drug use from the national panel is often high, affecting the overall precision of the estimates.

Despite of these limitations, DAWN survey estimates are widely used by government and private research institutions alike and play a key role in the evaluation of national policy and goals related to drug abuse. The DAWN survey is generally considered a credible source to measure trends in consequences related to the abuse of common drugs.

3 Methodology

The application of time series methods to periodic surveys was developed by Scott and Smith (1974). It was extended by Scott, Smith, and Jones (SSJ) (1977) to cover complex survey designs, including partially overlapping surveys. The SSJ methodology has been applied in the United States to develop new methods for estimating monthly employment and unemployment for 39 States and the District of Columbia for the Current Population Survey (CPS) by Bell and Hillmer (1987) and Tiller (1992a, 1992b). The goal of the SSJ approach is to increase the precision of the estimates through the use of time series methods applied to repeated surveys, modeling seasonality and trends, and sampling error in a single model.

Consider that at time t a survey provides an estimate y_t of a population parameter θ_t based on data from the survey at time t only, with sampling error e_t. This can be expressed in the form

$$y_t=\theta_t+e_t \tag{1}$$

If y_t is an unbiased estimator, then for a given θ_t, $E(e_t) = 0$ and Var $(e_t) = s_t^2$. If θ_t is fixed over time, as in the classical model, then the current value of y_t will depend on the previous values y_{t-1}, y_{t-2} through the error terms e_t, e_{t-1}, e_{t-2},... which will be correlated to the extent that the surveys overlap. Scott and Smith (1974) showed, however, that θ_t may be treated as the realization of a stochastic process and that previous values of y_{t-1}, y_{t-2} can be used to estimate θ_t even in non-overlapping surveys, so long as the series is generated by a stationary process or one that can be reduced to a stationary process by differencing or some other simple transformation.

The basic model applicable to the time series analysis is an autoregressive (AR) model of the general form

$$\sum_{j=0} a_j y_{t-j}=\varepsilon_t \tag{2}$$

From such a model, the forecast of the population parameter may be obtained from

$$\hat{\theta}_t(l)=\hat{y}_t(l)-\hat{e}_t(l) \tag{3}$$

using standard time series methods (Box and Jenkins, 1976) to fit an autoregressive moving average (ARMA) model.

An important practical application of this work is in improving the precision of estimates available from repeated surveys. In single stage non-overlapping surveys, Scott and Smith show that the minimum mean square error estimate can be written as

$$\hat{\theta}_t=(1-\pi_t)y_t+\pi_t\hat{y}_t \tag{4}$$

with a mean square error

$$v_t^2=(1-\pi_t)s_t^2 \tag{5}$$

where $\hat{y}_t$ is the forecast of y_t from $Y_{t-1} = (y_{t-1}, y_{t-2}, \ldots)$ and

$$\pi_t=s_t^2/Var(y_t|Y_{t-1}) \tag{6}$$

which is the ratio of the survey variance for y_t to the variance of the forecast of y_t from previous surveys. The value of this approach in improving the precision of survey estimates depends on the predictive power of the methods used to make the forecast. In single-stage panel surveys, such as the DAWN survey, SSJ (1974) show that the same variance estimators apply, while surveys with partial overlaps or more complex designs have more complicated formulae.

In all cases, however, standard time series methods can be used to fit models to the data series and generate the conditional variance estimates from stationary data series. Two basic approaches to time series modeling have been widely used in the literature: autoregressive integrated moving average (ARIMA) models and state-space models. ARIMA models are based on the work of Box and Jenkins (1976) and require stationary linear series. While stationarity has a complex technical definition (Judge, et.al., 1985, p. 228), the practical definition of stationarity is that the data series have neither trend nor cyclical components. State-space models using Kalman filtering/smoothing may be computationally more efficient for long time series (Bell and Hillmer, 1987) and do not assume stationarity. For purposes of this exploratory analysis, ARIMA methods were used, as implemented in the personal computer-based MicroTSP (QMS, 1990).

4 Description of the Data over Time

The data used in this analysis are drug-related ER visits by month for the period January 1978 through September 1992 (1978.01 through 1992.09). Most discussions of drug abuse concentrate on the abuse of illegal drugs, but most of the drug abuse involved in ER visits does not fit this description. Overall, 15 percent of drug uses mentioned in ER reports were motivated by recreational use or psychic effects, 27 percent were motivated by dependence, and 47 percent were motivated by suicide (NIDA 1992, Table 2.13).

As shown in Chart 1, the 1991 annual data show wide variations by type of drug. By individual drug group, suicide motives were reported for 6.5 percent of the 102,727 cocaine mentions, 3.2 percent of the 36,576 heroin/morphine mentions, 39.2 percent of cases involving alcohol in combination with other drugs, and 62.6 percent of all other drug mentions (e.g., tranquilizers and analgesics). The primary non-suicide motive for drug use is dependence (78.6 percent of heroin/morphine mentions and 64.6 percent of cocaine mentions), with recreational and other psychic effects motivating 10 to 17 percent of episodes. Thus, the DAWN data represent events associated more often with chronic drug dependence than with recreational drug use.

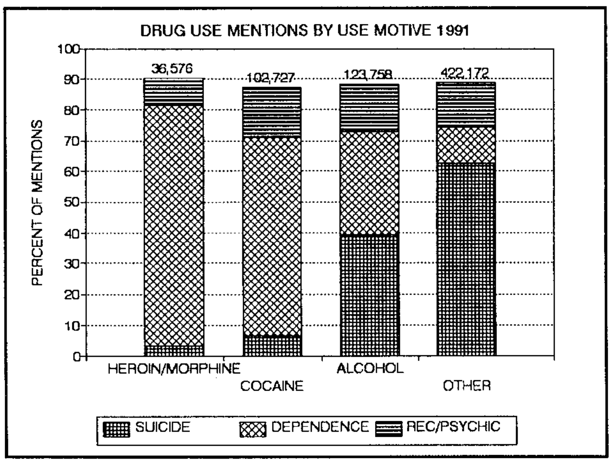

Chart 1

5 Results

The data used in the analysis are presented in Chart 2, which combines total episodes (solid line, left scale) and mentions of cocaine and heroin/morphine (right scale). Chart 2 shows that total monthly episodes were about 26,000 per month from 1978 through 1985. Between 1985 and 1989 there was a dramatic increase which culminated at 39,000 per month in 1989. After a sharp decline from 39,000 in 1989 to 28,000 in 1990, the monthly number of episodes increased in 1991 to near the 1989 levels. The growth in total episodes is driven largely by episodes in which cocaine was mentioned (dotted line, right scale). The number of cocaine mentions nearly tripled since 1986, while the number of heroin or morphine mentions (dashed line, right scale) remained fairly constant at approximately 3,000 mentions per month, until a rise to 5,000 mentions in 1992.

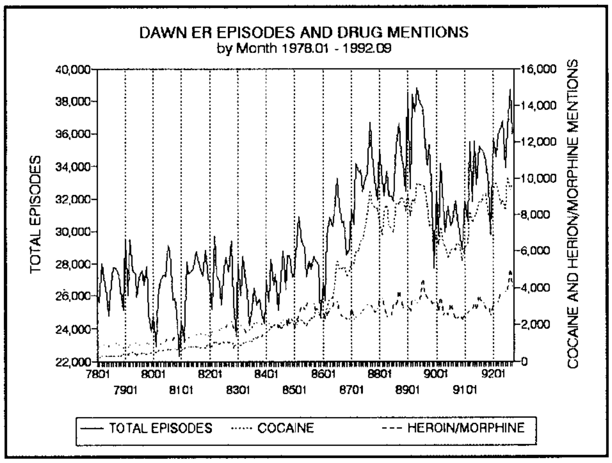

Chart 2

5.1 Seasonal Factors

A issue in any time series is the presence of seasonal factors that cause the reported estimates to vary systematically with the time of year. The presence of seasonal patterns was examined for each of the data series. To adjust for seasonal factors, a multiplicative method was used whereby the adjusted value is derived from the actual value by dividing the seasonal adjustment factor into the actual value. The seasonal adjustment factors were extracted using the MicroTSP SEAS function.

The seasonal adjustment factors for all five data series are shown in Chart 3, and the pattern is similar for each of the series. February (Month 2) generally has a lower number of episodes than January, followed by a rise in March, then a decline through May. July shows an increase in episodes (very large for heroin/morphine), followed by gradual declines through December. It should be noted that December estimates are among the lowest in the year, and January episodes represent a substantial increase from December. To test for sample weighting effects, the seasonal factors also were computed on unweighted data for 1978.1 through 1987.12, and the same pattern resulted.

Thus, there appear to be significant seasonal patterns for drug-related ER visits, and these patterns are similar for all five data series used in the analysis. The patterns are not markedly different between drug use

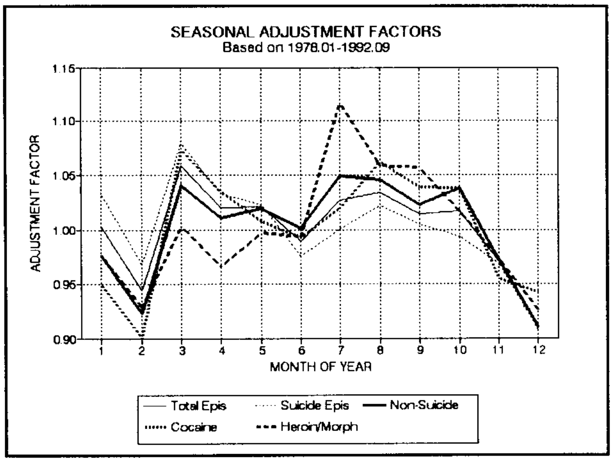

Chart 3

motivations (suicide or non-suicide) or drugs used (cocaine, heroin/morphine, or all other drugs, including alcohol in combination with other drugs). While the seasonal patterns may be significant, they do not completely explain the month-to-month variation. The range of the adjustments for total episodes is ± 0.05 (5 percent), while the relative variation in the data set is about 12 percent. When the seasonally adjusted data are plotted against the actual data, the adjusted and unadjusted series are similar. Thus, there is much more variability than can be explained by seasonal factors alone.

All five data series considered in this analysis (i.e., total episodes, suicide episodes, non-suicide episodes, cocaine mentions, and heroin/morphine mentions) have similar time patterns, as shown by correlations among the series. The lowest correlation among any of the series was 0.639 between suicide-related episodes and heroin/morphine mentions. Given these close correlations, there was no need for the exploratory analysis to examine all of them. In the interests of parsimony, the time series analysis presents results for total episodes only.

5.2 Time Series Models

One goal of any time series analysis is to identify any trends in the data, but as noted previously, time series methods require stationary series with trends removed. Two standard diagnostic approaches were applied to the data: (1) examination of the autocorrelation structure and (2) testing for unit roots. Autocorrelation analysis examines the autocorrelation coefficient (AC) and the partial autocorrelation coefficient (PAC). The AC measures the correlation of each data point with prior (lagged) data points in the series. For monthly data with seasonal patterns, the correlations should be examined over at least 1 year, so the AC's were calculated for lags from 1 to 18 months prior [$AC(k)$, k=1-18]. The simple correlations are calculated between each pair of lagged data points and then are averaged over all pairs for the given lag k. The PAC is the additional amount each lag k adds to explaining the total variance of the data series, after partialing out the effect of lags 1-(k-1).

In a stationary data series, the AC's taper off to 0 and PAC's disappear after a given lag. The original data were clearly non-stationary, because the AC's declined only very slowly. The PAC's decayed after Lag 2 and remained small until Lags 10, 12, and 13. This pattern is characteristic of data autocorrelated in degree 1 or 2 with a seasonal pattern. Chart 4 presents the AC and PAC analysis for the first difference of episodes. As can be seen in the chart, the AC's decay rapidly after Lag 1 but become larger again at Lags 11-15. The PAC's are large for Lags 1, 9, 11, and 12, suggesting a yearly pattern of seasonality. The autocorrelation analysis suggests that the total episode data are stationary in the first differences, and this was confirmed by additional analysis of unit roots (not shown here) using the Dickey-Fuller test (Dickey and Fuller, 1979).

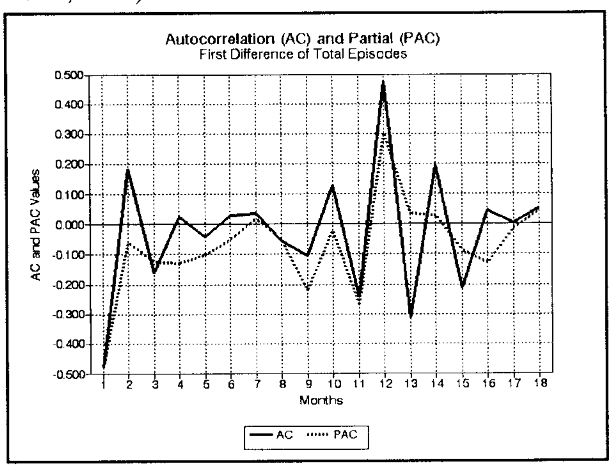

Chart 4

Time series model specifications were tested that used autoregressive (AR), moving average (MA), seasonal moving averages (SMA), and seasonal autoregressive (SAR) terms on both the differenced and undifferenced data. In theory, the models should be based on first differences, because the undifferenced data are not stationary. The search for a model began with a simple AR(1) model, then tested the addition of MA and higher order AR terms. The tests proceeded with an examination of the adjusted R^2, standard error of regression (SER), and F-ratio of each equation, plus a visual inspection of the residual AC and PAC plots.

The best first difference model found incorporates an autoregressive term {AR(1)}, a seasonal auto-regressive term {SAR(12)}, and a moving average {MA(12)}. The fitted model was as follows, using the notation of the TSP software:

$$Y_{t+1}=397.6-0.310*AR(1)+0.965*SAR(12)-0.920*MA(12)$$

where the constant term is insignificant and the other terms are significant at the 0.001 level or better. This equation has an adjusted R^2 of 0.55, an SER of 1,429.6, and an F-ratio of 63.3. The statistical significance of the 12-month moving average {MA(12)} is supported by the decrease in the F-ratio to 46.6 when it is omitted.

The ability of the model to fit the data is shown in Chart 5, which compares both fitted and forecast values with the actual data. Fitted values (dotted line) are single-period forecasts. That is, each period's fitted value begins with the last period's actual value. As shown in the chart, this close linkage does not allow the fitted values to depart far from the actual. The root mean square error (RMSE) of the fitted values is 1,472, or about 4 percent of the mean.

The forecast values (dashed line), on the other hand, are true multiperiod forecasts, with each period after the first month beginning with actual data up to the beginning of the forecast period, but continuing with forecast values after that date. In Chart 5, the forecast period began with 91.06 and continued through 92.09, the starting point of the best forecast (see further discussion below). The unlabeled lines parallel to the forecast line are the 95-percent confidence limits of the forecast, which increase over time taking into account both the variability of the data and the uncertainty of the coefficients. These confidence limits suggest that, while the forecasts track the data well, they should be used with caution.

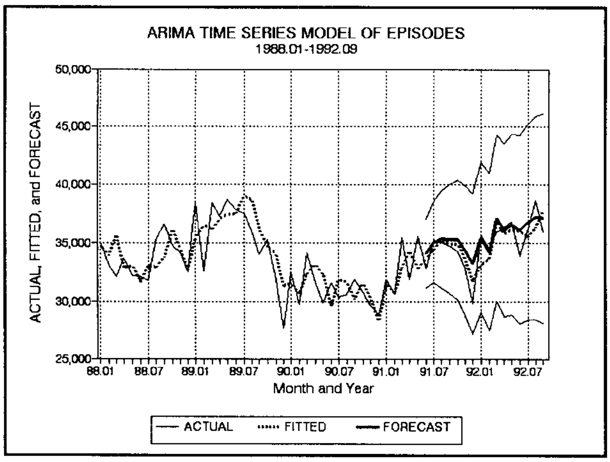

Chart 5

The forecast power of a model can be measured by a number of statistics. For the present paper, the forecast root mean squared error (FRMSE) (Kendall and Ord, 1990) was chosen for its utility in conjunction with the sampling error estimates. Most studies present a variety of forecast error measures, usually including the RMSE of the fitted values discussed above but calculated only once for each model. In the DAWN data, the magnitude of the forecast error depends heavily on the particular data chosen as the starting point. To illustrate this point, a series of 12 forecasts was made; each extended 6 months into the future, beginning in each of the 12 months of 1991 (the choice of year being arbitrary). The resulting FRMSE's are shown in Chart 6. The lowest root FRMSE was 922.8, for the forecast beginning in June, but those for forecasts beginning in January and March were more than 2,000, more than twice as high. These statistics show that, at least in this recent period, multi-period forecasts beginning in June have lower FRMSE's than single period forecasts beginning in other months of the year. This phenomenon merits further study to assess whether it may be attributable to some aspect of the modeling or to some peculiarity of the data set.

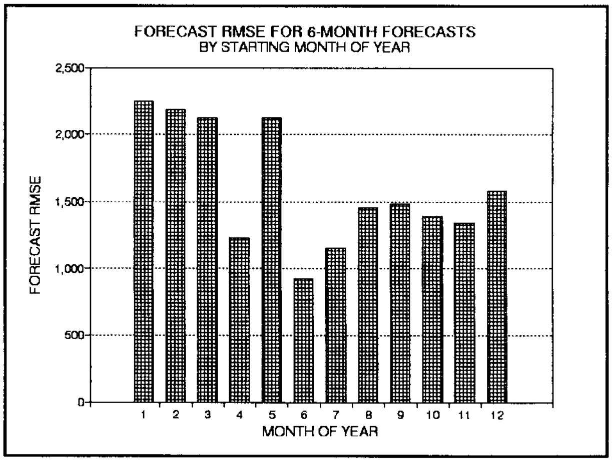

Chart 6

Finally, the study attempted to apply the SSJ methodology to improvements in the efficiency of sample estimates by incorporating the time series model results. Recall from Eq (4) that the gain in efficiency is proportional to $(1-\pi_t)$, defined as the ratio of the sampling variance s^2 to the MSE from the time series model. This implies that the sampling variance must always be smaller in magnitude than the MSE, so that $0 < \pi_t < 1$. In the literature, the magnitudes were in that range. However, nothing in the that restricts π_t, since the s^2 and the MSE are estimated completely independently. Indeed, when the calculations were performed using the DAWN data, s^2 was greater than the MSE for all calculations of the MSE from the fitted data as well as the forecast values. Therefore, the adjusted sampling variance, calculated as $(1-\pi_t)\ s^2$, was a negative number which is a logical impossibility.

The negative adjusted sampling result arises because the relative sampling errors (RSE's) for the DAWN monthly data exceed 5 percent of the monthly number of episodes, while the RMSE's calculated from the model results average approximately 4 percent of the monthly numbers. In the prior studies, the RSE's have been on the order of 1 to 2 percent so the issue never arose, although SSJ showed that high RSE's are possible in cluster survey data if the intra-cluster correlation coefficient is high. One solution to this problem is to model sampling error simultaneously with modeling the estimates (Tiller, 1992b), but it was not possible to do so under the time constraints for this study.

6 Discussion and Future Research

The DAWN data are important for tracking the impact of drug abuse on the health of drug abusers. While one might expect quite different trends and seasonal patterns for episodes motivated by suicide versus other motives, such as unexpected reactions or over-doses, the data suggest consistent patterns that also are reflected in separate mentions of cocaine and heroin/morphine. The data show seasonal patterns, but these are modest in relationship to the overall month-to-month variability of the data.

One potential goal of this research is to improve the precision of estimates from available data. At present, preliminary DAWN estimates are released quarterly, but the survey is designed to produce annual estimates. If the time series methods tested here prove fruitful, it might be possible to utilize monthly data to improve tracking and detection of trends in drug-related ER visits. Future research will attempt to model sampling error simultaneously with the time series model.

The work presented herein concentrates on total episodes, because there are strong correlations among the major indicators considered in this paper. However, there are also differences in drug use motivations, the types of drugs used, and the consequences of drug use among different demographic groups, between inner city and suburban areas, and across the urban areas represented in the sample. The application of time series methods might permit tracking ER episodes at a finer level of detail which would in turn result in more detailed data informing public officials and health professionals about particular new trends in drug use and their consequences.

REFERENCES

Bell, W.R., and Hillmer, S.C., (1987), "Time Series Methods for Survey Estimation," Bureau of the Census Statistical Research Division Report Series, #CENSUS/SRD/RR-87/20.

Box, G.E.P. and Jenkins, G.M. (1976) *Time Series Analysis: Forecasting and Control.* San Francisco: Holden Day.

Brookoff, D., Campbell, E.A., and Shaw, L.M. (1993), "The Underreporting of Cocaine-Related Trauma: Drug Abuse Warning Network Reports vs. Hospital Toxicology Tests," *American Journal of Public Health*, 83, 3, 369-371.

Dickey, E.B., and Fuller, W.A. (1979), "Distribution of the Estimators for Autoregressive Time Series with Unit Root," *Journal of the American Statistical Association*, 74, 427-431.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepol, H., and Lee, T.C.,1985, *The Theory and Practice of Econometrics.* New York, John Wiley and Sons.

Kendall, M. and Ord, J.K. *Time Series Third Edition.* New York: Oxford University Press.

National Institute on Drug Abuse (NIDA, 1992), *Annual Emergency Room Data 1991*, DHHS No. (ADM) 92-1955, Series 1, Number 11-A.

QMS (1990), *MicroTSP User's Manual.* Irvine, California: Quantitative Micro Software.

Scott, A.J. and Smith, T.M.F. (1974), "Analysis of Repeated Surveys Using Time Series Methods," *Journal of the American Statistical Association*, 69, 674-678.

Scott, A.J., Smith, T.M.F., and Jones, R.G. (1977), "The Application of Time Series Methods to the Analysis of Repeated Surveys," *International Statistical Review*, 45, 13-28.

Tiller, R. (1992a), "A Kalman Filter Approach to Labor Force Estimation Using Survey Data", *1992 Proceeding of the Business and Economic Statistics Section, American Statistical Association*, 16-25.

Tiller, R. (1992b), "A Time Series Approach to Small Area Estimation", unpublished paper presented to the American Statistical Association, September, 1992.