PRELIMINARY ESTIMATES FOR THE NATIONAL CRIME SURVEY USING REGRESSION AND TIME SERIES

Paul G. Wakim, Bureau of the Census

1. Introduction

In the National Crime Survey (NCS) conducted by the Bureau of the Census for the Bureau of Justice Statistics, a sampled household is interviewed every six months for three years. The first of the seven interviews, the bounding interview, is used only to set a time frame in order to avoid duplicating reported crimes on subsequent visits. The estimated crime levels and rates that are computed from the NCS are based on the last six interviews only. At the interview, the victimizations that occurred during the past six months are reported. As a result, all the reports on victimizations that occurred during the year of interest are not collected until June of the following year.

The Bureau of Justice Statistics is interested in producing preliminary annual estimates as early as possible. That is, the goal is to predict the final estimate of the annual crime level obtained when all the needed interview reports are collected. In that sense, the "true" value is not the population crime level, but its final estimate. In this paper, the population crime level is not mentioned at all, and the final estimate of the crime level is sometimes referred to as just the crime level. Moreover, by considering interviews up to January of year t+1, the suggested methods might seem to predict the number of victimizations that occurred in the past, i.e., year t; however, the "true" value will not be known until June of year t+1; hence, the word "prediction" rather than "estimation" is often used.

The method that has been used for the 1983 preliminary estimates (BJS Bulletin, June 1984) considers the collection year, that is, includes in the estimation procedure all the crimes that were reported in interviews conducted in the year of interest regardless of whether they occurred during that year. Wakim (1984) describes this method in more detail and compares it to the regression approach. The results showed that the simple linear regression model tends to lead to smaller relative prediction errors on the average. In this paper, three methods within the regression approach are described (section 2). Section 3 proposes several methods for obtaining preliminary annual estimates by combining predictions from regression and time series models. In Section 4, these methods are applied to two types of crime and compared to other methods based on regression alone or time series alone. The methods are not restricted to this particular problem; they can easily be applied in any situation where the dependent variable is different for each time unit.

The chart (at the end of the report) shows the interviewing pattern. Each X represents all victimizations that took place during the specified month of occurrence and were reported during the specified month of interview. The chart also illustrates the fact that it takes six months of interview (e.g., May through October) to obtain complete data for a single month (the April victimizations). Similarly, all of the December victimizations are not available until June (of the following year). Moreover, if we were to collect all reports only through the January (of the following year) interviews, we would only have a small part (about one sixth) of the December occurrences, about two sixths of the November occurrences, and so on up to about five sixths of the August occurrences. On the other hand, the reports on victimizations that occurred during the months of January through July (of the year of interest) would all be available.

For this analysis, the first step is to set the last month of interview through which all reports will be collected. Considering the interviews only up to December is not recommended since absolutely no information on the December occurrences would be known. Throughout this paper, the interviews up to January are considered. The suggested methods still apply if interviews up to February or later are considered.

2. The Regression Approach

Let W denote the annual crime level of the year of interest when all the needed reports are available. This represents 72 X's in the chart or equivalently the sum of the 12 monthly crime levels (January - December) where each month of occurrence is represented by 6 X's (read vertically). Let Z_1 denote the crime level for the months with complete data, i.e., January through July. Z_1 is represented by the sum of 42 X's (7 months of occurrence x 6 X's read vertically). On the other hand, let Z_2 denote the crime level for the months with incomplete data, i.e., August through December ($W = Z_1 + Z_2$). Z_2 is represented by the sum of 30 X's (5 months of occurrence x 6 X's read vertically).

Finally, we let z_2 denote the number of crimes that occurred during the months with incomplete data (August - December) and that were reported in interviews conducted up to January (of the following year); z_2 is represented by the sum of 15 X's (5+4+3+2+1).

The regression approach basically tries to predict W given Z_1 and z_2.

The data used to fit the models and obtain estimates of the parameters consist of the monthly levels from January, 1973 to December, 1982, broken down by month of interview.

2.1 Predicting the annual crime level directly

A direct way to predict the annual crime level W using a regression approach is by considering $(Z_1 + z_2)$ as the independent variable and writing

$$W = \hat{c}_A\,(Z_1 + z_2) + \hat{b}_A + e_A$$
$$= \hat{W}_A + e_A$$

where c_A and b_A are the parameters of the regression line; their estimates $\hat{c}_A$ and $\hat{b}_A$ are obtained by fitting the line through the 10 data points; $\hat{W}_A$ is the annual crime level prediction; and $E[e_A] = 0$.

2.2 Predicting the crime level for the period with incomplete data

Instead of predicting the annual crime level directly, one can predict Z_2 which is unknown and add this prediction to Z_1 which is known. We consider two methods:

(1) Sum of the monthly levels:
The regression model for predicting Z_2, from z_2 can be written as

$$Z_2 = \hat{c}_B z_2 + \hat{b}_B + e_B$$
$$= \hat{Z}_2 + e_B$$

where $\hat{c}_B$ and $\hat{b}_B$ are the parameter estimates, $\hat{Z}_2$ the prediction of the crime level for the months with incomplete data and e_B the predicting error with $E[e_B]=0$.
Now, we can write the annual crime level as

$$W = Z_1 + \hat{Z}_2 + e_B$$
$$= \hat{W}_B + e_B$$

where $\hat{W}_B$ is the annual crime level prediction.

(2) The monthly levels separately:
The idea is to use separate regression lines to predict the levels of the five months with incomplete data (i.e., Aug - Dec). Let Y_{t+i} denote the crime level for month t+i; this corresponds to the 6 X's in the chart for the specified month of occurrence. Let x_{t+i} denote the partial crime level for month t+i; that is, for t = 7, 19, 31, 43, ..., x_{t+1} corresponds to the sum of the first 5 X's for August, x_{t+2} to the sum of the first 4 X's for September and so on up to x_{t+5} corresponding to the first X for December. Note that after the reports from the January (month t+6) interviews are collected, $x_{t+1}, \ldots, x_{t+5}$ are known; moreover, for that particular year of interest,

$$z_2 = \sum_{i=1}^{5} x_{t+i}.$$

The regression model for each of the 5 months can be written as

$$Y_{t+i} = \hat{c}_i x_{t+i} + \hat{b}_i + e_{t+i} \qquad i=1,\ldots,5$$
$$= \hat{X}_{t+i} + e_{t+i}$$

where $\hat{c}_i$ and $\hat{b}_i$ are estimates of the parameters, $\hat{X}_{t+i}$ the predicted monthly crime level and e_i the predicting error, i=1,...,5.

We can therefore predict Z_2 by $\hat{Z}_2 = \sum_{i=1}^{5} \hat{X}_{t+i}$

and the annual crime level by

$$\hat{W}_C = Z_1 + \hat{Z}_2$$

where $W = \hat{W}_C + e_C$

and $e_C = \sum_{i=1}^{5} e_{t+i}$.

3. The Time Series/Regression Approach

3.1 Time series model

The regression approach tries to predict the final annual crime level from the known part of the data. When fitting the regression lines, the pattern that monthly crime levels might follow is completely ignored. A Bureau of Justice Statistics report (1980) showed that several types of crime do in fact follow seasonal patterns. Their occurrences can, therefore, be described quite appropriately by a time series model. Including such information may lead to more accurate predictions and smaller prediction variances. For such a type of crime, a time series model can be written as

$$\sum_{j=0}^{\infty} a_j Y_{t-j} = \varepsilon_t \qquad a_0 = 1$$

where $\varepsilon_t, \varepsilon_{t-1}, \ldots$ are uncorrelated random variables with mean zero and common variance σ^2.

For the moment, let Y_t denote the monthly crime level for July. Then, with interviews up to January, $Y_t, Y_{t-1}, \ldots$ are known and the five predictions for August through December can be written as

$$Y_{t+i} = \hat{Y}_t(i) + e_t(i) \qquad i=1,\ldots,5$$

where $\hat{Y}_t(i)$ is the prediction for month t+i and $e_t(i)$ the prediction error associated with $\hat{Y}_t(i)$. The crime level for the period with incomplete data can be written now as,

$$Z_2 = \hat{Z}_2 + \sum_{i=1}^{5} e_t(i)$$

where $\hat{Z}_2 = \sum_{i=1}^{5} \hat{Y}_t(i)$, and the annual crime level as

$$W = (Z_1 + \hat{Z}_2) + \sum_{i=1}^{5} e_t(i).$$

There are two main disadvantages associated with using the time series model alone: the first one is that the forecasting variance increases very rapidly with the lead time; as a result, the variance of $\sum_{i=1}^{5} e_t(i)$ is expected to be large. The second is that it ignores the part of the data that is known, namely z_2. One solution is a method that combines the time series and regression models. But first, a few assumptions about the models' prediction error terms need to be made.

3.2 Assumptions about the correlation between the error terms

This paper makes the following assumptions:

(1) e_A and $\sum_{i=1}^{5} e_t(i)$ are correlated.

(2) e_B and $\sum_{i=1}^{5} e_t(i)$ are correlated.

(3) The error terms from regression line i (for the monthly levels separately) are serially uncorrelated for every i=1,...,5.

(4) e_{t+i} and e_{t+j} are correlated, for i,j=1, ...,5.

(5) e_{t+i} and ε_{t+j} are correlated for i=j and uncorrelated for i≠j, i, j=1,...,5.

Note: The fifth assumption can be relaxed, namely, assume that e_{t+i} and ε_{t+j} are correlated for i≠j; the methods described below would still be applicable; only the computations would be more complex.

3.3 Combination at the annual level

Using the input variable x_t, the typical time series/regression model can be written as (see for example, Box and Jenkins, 1976):

$$Y_t = \beta x_t + \frac{\theta(B)}{\phi(B)\ \delta(B)} a_t$$

where Y_t and x_t are defined as in the previous sections, B is the backshift operator, θ (B) and $\phi\cdot$ (B) are polynomial functions in B, δ (B) is a differencing operator, and a_t is white noise. However, for the NCS preliminary estimates problem, the input variable is not the same for each of the five months with incomplete data (August through December). Therefore, the above model does not apply and other methods need to be investigated. The basic idea of the models suggested in this paper is to linearly combine the regression and time series predictions in an optimal way, in the sense of minimizing the variance of the final error term. These models basically differ in terms of the level at which the combination is made. The first method combines the prediction of the annual crime level from the regression model, namely $\hat{W}_A$ (section 2.1) with the one from the time series model,

namely $Z_1 + \sum_{i=1}^{5} \hat{Y}_t(i)$. In other words, we express the new annual crime level prediction as

$$\hat{W}_A' = K\hat{W}_A + (1-K)\ (Z_1 + \sum_{i=1}^{5} \hat{Y}_t(i)).$$

The new error term becomes,

$$e_A' = W - \hat{W}_A' = Ke_A + (1-K)\ (\sum_{i=1}^{5} e_t(i)).$$

In this case, the optimal value of K is

$$K = \frac{var(\Sigma e_t(i)) - cov(e_A, \Sigma e_t(i))}{var(e_A) + var(\Sigma e_t(i)) - 2cov(e_A, \Sigma e_t(i))}$$

and the corresponding optimal value of the variance of the new error term is

$$var(e_A') = \frac{var(e_A).var(\Sigma e_t(i)) - cov^2(e_A, \Sigma e_t(i))}{var(e_A) + var(\Sigma e_t(i)) - 2\ cov(e_A, \Sigma e_t(i))}$$

which is smaller than each of the variances $var(e_A)$ and var $(\Sigma e_t(i))$ (for a proof, see for example Bates and Granger, 1969).

3.4 Combination of the crime levels for the period with incomplete data.

For each of the two linear regression models described in Section 2.2, we consider a corresponding combination model.

(1) Sum of the monthly levels:
This method combines the prediction of the crime level for the period with incomplete data from the regression line, namely $\hat{Z}_2$ (from section 2.2, method (1)) with the one from the time series model, namely $\sum_{i=1}^{5} \hat{Y}_t(i)$ and then adds the combination to Z_1 in order to obtain the new annual crime level prediction $\hat{W}_B'$.
The new error term becomes

$$e_B' = W - \hat{W}_B' = Ke_B + (1-K)(\sum_{i=1}^{5} e_t(i)).$$

The optimal value of K and the corresponding optimal value of the variance of the new error term can be similarly (as in the previous section) obtained.

(2) The monthly levels separately:
This method combines the sum of the monthly predictions from the five regression lines, namely $\sum_{i=1}^{5} \hat{X}_{t+i}$ (see section 2.2, method (2)) with the corresponding sum from the time series model, namely $\sum_{i=1}^{5} \hat{Y}_t(i)$; the new prediction of Z_2 is added to Z_1 in order to obtain the final annual crime level prediction $\hat{W}_C$. The new error term becomes

$$e_C' = W - \hat{W}_C' = Ke_C + (1-K)\ (\sum_{i=1}^{5} e_t(i)).$$

The optimal value of K and the corresponding optimal value of the variance of the new error term can be similarly obtained.
Note: In this case, the error terms are correlated as a result of the correlation between e_{t+i} and ε_{t+i} (section 3.2).

3.5 Combination at the monthly level

The third level at which a combination can be made is at each of the five months with incomplete data. In this case, we need the predictions from the five separate regression lines of section 2.2 (2nd method) and the five forecasts from the time series model. We propose three ways of combining these predictions.

(1) Simple combination:
For each month with incomplete data, its new prediction $\hat{Y}_t'(i)$ is a linear combination of the monthly prediction from the regression $(\hat{X}_{t+i})$ and time series $(\hat{Y}_t(i))$ models. The final annual crime level prediction is written as $Z_1 + \sum_{i=1}^{5} \hat{Y}_t'(i)$.
The new error term is equal to

$$W - (Z_1 + \sum_{i=1}^{5} \hat{Y}_t'(i)) = \sum_{i=1}^{5} e_t'(i)$$

where $e_t'(i) = K_i e_{t+i} + (1-K_i)e_t(i)$.
For each i, the optimal value of K_i and the corresponding optimal value of the variance of $e_t(i)$ can be obtained; the variance of the error term associated with the final annual crime level predictor is equal to the variance of

the sum, $\sum_{i=1}^{5} e_t'(i)$.

(2) Intertwined combination:
The first of the five monthly crime levels, the August level, is predicted as in the simple combination. The new August prediction, $\hat{Y}_t(1)$, is used in the time series model to obtain the two-step-ahead forecast, $\hat{Y}_t(2)$. This forecast is then combined with the September prediction, $\hat{X}_{t+2}$, from the corresponding regression model. The new September prediction, $\hat{Y}_t(2)$ is used in the time series model to obtain the three-step-ahead forecast, $\hat{Y}_t(3)$, and so on up to obtaining the new

December prediction, $\hat{Y}_t'(5)$.

The final annual crime level prediction is then equal to $Z_1 + \sum_{i=1}^{5} \hat{Y}_t'(i)$, and the error term associated with it is $\sum_{i=1}^{5} e_t'(i)$.

At each step, the optimal K_i is chosen so as to minimize the variance of $e_t(i)$.

(3) Minimizing the variance of the sum:
An extension to each of the previous two methods is to express the variance of the error terms $e_t(i)$ in terms of all the K_i's and the estimated variances and covariances, and then find the value of the K_i's that minimize the variance of the sum. This extension for the case of the intertwined combination leads to a rather complicated minimization problem and is not studied in this paper. In the case of the simple combination, the minimization problem is easily reduced to solving a system of five linear equations with the five unknowns $K_1,\ldots,K_5$.

3.6 Which combination to use

The answer is the one with the smallest prediction error and since all the error terms have their expected value equal to zero, that translates, for our purposes, into the model with the smallest error variance. In this sense, the time series model, if it exists, is bound to improve the prediction error when its prediction is combined with the one from the regression model. So for the three regression models, namely at the annual level, the sum of the monthly levels and the monthly levels separately, their corresponding time series/regression model will lead to a smaller error variance, respectively.

Now the question becomes: which combination to use from among the time series/ regression models? The simple combination at the monthly level will lead to a smaller variance than the one from the monthly levels separately for the period with incomplete data simply because more coefficients are considered with the same set of error terms. Moreover, the method that minimizes the variance of the sum will lead to a smaller variance than the one from the simple combination at the monthly level, by definition.

4. Applications

Under the National Crime Survey program, the Bureau of Justice Statistics publishes an annual report providing information on criminal victimization in the United States. In this paper, we consider two of the types of crime that are tabulated in the reports.

4.1 Personal larceny without contact and total household crimes

For each type of crime, the ten described methods of obtaining preliminary estimates were applied and compared (see Table 4.3 for a complete list of the different methods).

(1) Personal larceny without contact:
This type of crime is described as "theft or attempted theft, without direct contact between victim and offender, of property or cash from any place other than the victim's home or its immediate vicinity. Examples of personal larceny without contact include the theft of a briefcase or umbrella from a restaurant, a portable radio from the beach, clothing from an automobile parked in a shopping center, etc." (Criminal Victimization in the United States, 1981). Figure 4.1 is a plot of the monthly levels from January 1973 to December 1982. It is clear that the series is seasonal with peaks in the fall of the year and low points in the summer months. The mean of the series is about 1,300,000 victimizations per month and its standard deviation about 115,600 victimizations. The estimated standard deviation of the white noise term of the time series model is about 50,800 victimizations. Table 4.1 shows the estimated variance of the forecast errors at each lead time using tne five methods that lead to separate monthly predictions. For the regression model alone (I) and the time series model alone (II), the variance of the prediction error increases as the lead time increases. However, in the case of the three time series/regression models, the variance peaks at the 4th step-ahead forecast and decreases for the 5th step-ahead forecast. As expected, the time series model alone led to the highest variances for each step ahead forecast. On the other hand, the simple time series/regression combination (III) consistently led to the smallest variances. The method that minimizes the variance of the sum (V) led to relatively high error variances for each prediction since it does not necessarily minimize the variance at each step. Table 4.1 also shows the coefficients K_i's of the regression predictions in the time series/regression models. The simple and intertwined combinations (III and IV) have similar coefficients because they both try to achieve the same goal, namely to minimize the variance at each lead time.

Table 4.3 shows the estimated variance of the error term associated with the final annual crime level predictor which is equivalently the estimated variance of the sum of the error terms for each lead time. As expected, the variance from the regression models alone is higher than the one from the corresponding time series/ regression models. For the reasons explained in Section 3.6, the variance from the time series/ regression model from the monthly levels separately ($11,920 \times 10^6$) is larger than the one from the simple combination ($11,219 \times 10^6$) which is, in turn, larger than the one from the method that minimizes the variance of the sum ($10,537 \times 10^6$). By far, the largest variance was from the time series model alone ($48,506 \times 10^6$). Among the regression models alone, the lowest variance was from the monthly levels separately ($12,956 \times 10^6$). Among the time series/regression models, the lowest variance was from the one that minimizes the variance of the sum ($10,537 \times 10^6$). Therefore, if we consider the "best" two models in their respective category, the reduction in variance is about 18.7% confirming the advantage of incorporating a time series model.

(2) Total household crimes:
The Bureau of Justice Statistics annual report describes this type of crime as "burglary or larceny of a residence, or motor vehicle theft, crimes that do not involve personal confrontation" (Criminal Victimization in the United

States, 1981). Household larceny includes theft or attempted theft by someone with a right to be there, such as a maid, a delivery person, or a guest. Figure 4.2 is a plot of the monthly levels for the same period of time. It is clear that this series is seasonal too, with peaks in July and August when families leave their house for vacation and low points in January and February when people typically stay home. The mean of the series is about 1,500,000 incidents per month and its standard deviation is about 205,300 incidents. The estimated standard deviation of the white noise term of the time series model is about 50,100 incidents. Table 4.2 shows the estimated variance of the forecast errors at each lead time. The figures differ from Table 4.1 in a few points. The first one is that in the time series/regression model that minimizes the variance of the sum (V), the variances do not follow the same pattern as in the other two combination models (III and IV). Moreover, among the time series/regression models, none consistently led to either the smallest or the largest variances. As for the coefficients of the regression predictions, the three time series/ regression models show similar values for each lead time, except for the one-step-ahead forecast. Table 4.3 shows another interesting difference between the two types of crime, namely that among the regression models alone, the model at the annual level led to the smallest variance ($13,508 \times 10^6$); the model for the monthly levels separately led to a considerably largest variance ($26,801 \times 10^6$) which is due mainly to the large covariance between the error terms of the regression lines (Section 3.2, 4th assumption). On the other hand, among the time series/regression models, the model that combines the sum of the monthly levels led to the smallest variance ($12,161 \times 10^6$). Again, if we consider the "best" two models in their respective category, the reduction in variance is about 10%.

4.2 Conclusion

It is important to realize that the variance of the "best" regression model is theoretically larger than the variance of the "best" time series/regression model; in other words, what the previous tables have shown is not a special case resulting from the particular data used. On the other hand, we do need to know the variance of the variance estimates before making any general statements about the error term variance of the following four time series/ regression models:

- At the annual level
- For the period with incomplete data: (1) sum of the monthly levels
- At the monthly level: (2) intertwined combination
- At the monthly level: (3) minimizing the variance of the sum

The comparison of the different methods was based only on the estimated variance of the error terms. A more thorough comparison should also involve the number of parameters estimated (including K or the K_i's) and the number of observations that were used to estimate the parameters. Criteria such as Akaike's AIC need to be computed. For the time series/ regression models this task is rather complex and is not carried out in this paper; however, for the sake of completeness, the number of linear parameters and the number of observations were included in Table 4.3.

Another important point to stress is that the time series/regression models described in this paper are applicable not only to the NCS problem, but to any time series/regression situation where the independent variable is different for each lead time and hence where the regular time series/regression model described in Section 3.3 would not apply.

One final note is that with more data being available, the estimates of the regression and time series model parameters will be more accurate and so will the estimates of the variances and covariance of the error terms and the estimate of the optimal K. As a result, the final combined prediction will improve too. However, with the collection approach (mentioned in the Introduction), the predictions will not necessarily improve with time since, except for the previous year, all earlier observations are ignored.

5. References

- Bates, J.M. and Granger, C.W.J. (1969), "The Combination of Forecasts", Operational Research Quarterly, 20, 451-468.
- Box, G.E.P., and Jenkins, G.M. (1976), Time Series Analysis: Forecasting and Control, San Francisco: Holden Day.
- Bureau of Justice Statistics, Crime and Seasonality, U.S. Department of Justice, May 1980.
- Bureau of Justice Statistics, Criminal Victimization in the United States, 1981, U.S. Department of Justice, November 1983.
- Bureau of Justice Statistics (Bulletin), Criminal Victimization 1983, U.S. Department of Justice, June 1984.
- Draper, N.R., and Smith H. (1981), Applied Regression Analysis, 2nd ed., Wiley, New York.
- Wakim, P.G. (1984), "NCS Preliminary Estimates-A Preliminary Analysis", unpublished report.

Chart: NCS-Month of Interview by Month of Occurrence

Mo. of Interview	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	
J	X	X	X	X	X	X																		J
F		X	X	X	X	X	X																	F
M			X	X	X	X	X	X																M
A				X	X	X	X	X	X															A
M					X	X	X	X	X	X														M
J						X	X	X	X	X	X													J
J							X	X	X	X	X	X												J
A								X	X	X	X	X	X											A
S									X	X	X	X	X	X										S
O										X	X	X	X	X	X									O
N											X	X	X	X	X	X								N
D												X	X	X	X	X	X							D
J													X	X	X	X	X	X						J
F														X	X	X	X	X	X					F
M															X	X	X	X	X	X				M
A																X	X	X	X	X	X			A
M																	X	X	X	X	X	X		M
J																		X	X	X	X	X	X	J
J																								J

<---Year of Interest-->

Figure 4.1
Personal Larceny without Contact

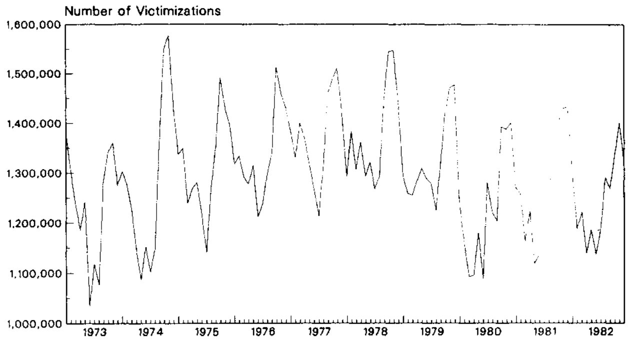

Figure 4.2
Total Household Crimes

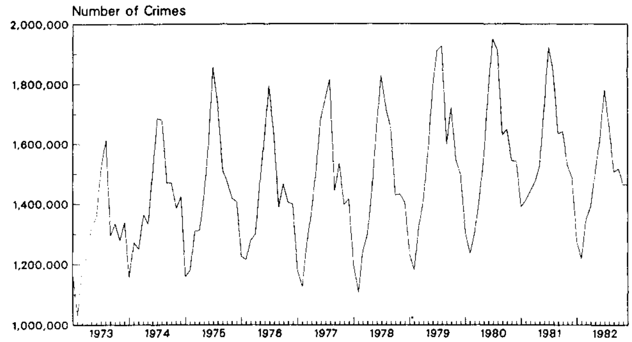

TABLE 4.1 Personal Larceny Without Contact - Estimated Variance of the Forecast Errors Using Different Methods (x10^6) (and the coefficients for the time series/regression models)

Lead Time	I	II	III	IV	V
			Method Used*		
1	515	2582	499 (K_1=1.097)	499 (K_1=1.097)	842 (K_1=1.543)
2	682	3190	676 (K_2=0.955)	681 (K_2=1.018)	1230 (K_2=0.507)
3	768	3748	750 (K_3=0.928)	762 (K_3=0.949)	1670 (K_3=1.442)
4	1995	4042	1790 (K_4=0.768)	1845 (K_4=0.724)	2127 (K_4=0.471)
5	2245	4231	1500 (K_5=0.657)	1657 (K_5=0.616)	1530 (K_5=0.726)

TABLE 4.2 Total Household Crimes - Estimated Variance of the Forecast Errors Using Different Methods (x10^6) (and the coefficients for the time series/regression models)

Lead Time	I	II	III	IV	V
			Method Used*		
1	253	2513	187 (K_1=0.856)	187 (K_1=0.856)	1546 (K_1=1.510)
2	1503	3049	589 (K_2=0.621)	530 (K_2=0.590)	593 (K_2=0.645)
3	2399	4062	1324 (K_3=0.615)	1308 (K_3=0.530)	1643 (K_3=0.405)
4	2811	4206	1776 (K_4=0.605)	1863 (K_4=0.528)	1802 (K_4=0.667)
5	3048	4565	1669 (K_5=0.592)	1806 (K_5=0.523)	1712 (K_5=0.519)

* I-Regression approach:the monthly levels separately.
II-Time series model alone.
III-Time series/regression:simple combination.
IV-Time series/regression:intertwined combination.
V-Time series/regression:minimizing the variance of the sum.

TABLE 4.3 Estimated Variance of the Error Term Associated with the Final Annual Crime Level Predictor (x10^6)

METHOD USED	TYPE OF CRIME: Personal Larceny Without Contact	TYPE OF CRIME: Total Household Crimes	# of Parameters*	# of Observations
Regression Alone:				
- At the annual level:	15,682	13,508	2	10
- Period with incomplete data:				
(1) Sum of the monthly levels	14,087	14,627	2	10
(2) Monthly levels separately	12,956	26,801	10	50
Time Series Alone:	48,506	51,147	6	120
Time Series/Regression Combination:				
- At the annual level:	15,314	13,109	9	120
- Period with incomplete data:				
(1) Sum of the monthly levels	13,600	12,161	9	120
(2) Monthly levels separately	11,920	15,968	17	120
- At the monthly level:				
(1) Simple combination	11,219	15,022	21	120
(2) Intertwined combination	11,550	15,301	21	120
(3) Minimizing the variance of the sum	10,537	13,874	21	120

* The number of linear parameters estimated from fitting the models.