Forecasting Crime[1]

By

Wilpen L. Gorr
Andreas M. Olligschlaeger
Janusz Szczypula
Yvonne Thompson

May 19, 1999

[1] Work on this paper was supported by Grant No. 98-IJ-CX-K005 awarded by the National Institute of Justice, Office of Justice programs, U.S. Department of Justice. Points of view in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice.

Introduction

Crime forecasting is a new field, just now being investigated by the National Institute of Justice through a cluster of four grants [National Institute of Justice 1998]. Indeed, we are aware of only one published paper on crime forecasting [Olligschlaeger 1998]. Hence, most of the material in this paper is conceptual, drawing on experience from other areas.

Organizations in the private sector must do strategic planning over long-term horizons to locate new facilities, plan new products, develop competitive advantages, and so forth. Consequently, long-term forecasts of demand, costs of raw materials, etc. are important in the private sector. There is no such strategic counterpart to police work; consequently, long-term forecasts are of little value to police. Police primarily need short-term forecasts; for example, crime levels one week or one month ahead. Currently, police mostly respond to new crime patterns as they occur. Client-server computing for real-time access to police records and computerized crime mapping have made it possible for police to keep abreast with crime. With short-term forecasting police may be able to get one step ahead of criminals by anticipating and preventing crime.

The organization of this paper proceeds first with a description of short-term forecasting models, to provide basic terms and concepts. Next is a discussion of unique features of crime space-time series data, and the need for data pooling to handle small-area model estimation problems. Lastly are a discussion of particular forecasting requirements of police and a summary.

Short-Term Forecasting Models

Short-term forecast models are of two primary kinds: 1) univariate, *extrapolative forecast models* and 2) multivariate, *leading indicator forecast models*. A major difference is that extrapolative models can only continue or extrapolate existing crime patterns into the future; whereas, leading indicator models can forecast new crime patterns not yet observed. Both types of models have value for police.

Extrapolative forecast models capture the historic time-based pattern in the variable of interest. The patterns most estimated and forecasted are linear time trend and seasonality. For example, robberies may be increasing by two per month on the average (linear time trend) with a peak 25% higher than normal in December and trough 20% lower in August (seasonality). Most often time trend estimates are made with “drift” estimators, such as exponential smoothing, that capture the most recent trend for extrapolation.

There are many time series forecasting models, ranging from simple to very complex. Fortunately, the literature has found that simple models are as accurate or more so than complex models. Simple, but good performing methods for extrapolation are classical decomposition for deseasonalizing data and Holt’s two parameter exponential smoothing for time trend estimation of deseasonalized data. Both additive and multiplicative seasonal coefficients may be estimated. See Makridakis (1982) and Armstrong (1999).

Leading indicator models require multivariate data: the dependent variable (e.g., number of robberies per month) and precursor leading indicator variables, generally lagged one or more time periods (e.g., number of gangs or drugs 911 calls from a month or more earlier). Lags may also be over space, such as total drug 911 calls at contiguous (nearby) block groups one month ago. Suppose that a model includes leading indicators lagged one or more months. Then a one-month-ahead forecast can be computed using the most recent month's and older data for indicators. Two or more month-ahead forecasts would require forecasts of some of the leading indicators themselves, thereby compounding the forecast problem. Hence leading indicator forecasts rarely extend beyond the shortest lag in the model (e.g., one month). If there is a pattern change in a significant leading indicator (e.g., a turning point or step jump) then the forecast will likely also have the same or related pattern change.

A successful method for estimating leading indicator models is the Bayesian Vector Autoregressive model (BVAR). Introduced by Litterman (1980, 1986), BVAR uses prior beliefs in the model to overcome collinearity and degrees of freedom problems that typically arise in applications of vector autoregressive (VAR) models. Doan, Litterman, and Sims (1984) introduced the so-called Minnesota priors for BVAR. LeSage and Pan (1995) introduced spatial contiguity to further specify the priors for regional studies. BVAR models have been successful time series analysis and forecasting models for regional data, especially when exploring for time and space lagged model specifications (LeSage 1989, 1990; LeSage and Pan 1995).

Crime Space-Time Series Data

A somewhat unusual feature of crime forecasting, relative to the general forecasting literature, is that multiple time series must be forecasted. The multiple series correspond to multiple areal units such as beats, block groups, or square grid cells. "Time series" refers to consecutive observations of a variable over time intervals, such as months or quarters. Most time series models were designed for forecasting a single time series in isolation (e.g., exponential smoothing, Box Jenkins, classical decomposition, etc.). "Space/time series" is the term that we use for crime data because, in addition to time series, the areal units form a continuous space series partitioning the police jurisdiction (e.g., city or county). Separate forecasts are needed for each areal unit, so that police can respond differently and appropriately in each unit.

A challenge of crime space/time series data is that they are likely volatile; that is, with relatively large residual error terms for models. They may also have abrupt changes in time series pattern. The cause of these problems is that the data are aggregates of relatively few crime incidents (e.g., number of robberies per month in areas averaging 1,000 population) so that special events like police interventions, release of a prison inmate to parole, etc. may have large impacts on crime levels. In larger data aggregates, such special events are numerous but relatively small parts of total variation, tend to cancel each other out, and merely contribute to a reasonably small error term. The consequence of volatile data is that parameter estimates of corresponding space/time series model are likely be unreliable with high standard errors. Unreliable parameter estimates lead ultimately to poor forecasts.

A remedy for the problems of crime space/time series is from the Bayesian pooling literature and is sometimes referred to as "borrowing strength from neighbors." [Duncan, Gorr, and Szczypula 1993, 1999] Pooling takes advantage of autocorrelation and other relationships between spatial units and in effect combines time series data from groups of areal units to increase sample sizes. The result is generally more reliable parameter estimates and therefore more accurate forecasts.

With pooling, the analyst has three alternative models to compare for each areal unit: 1) the *individual time series model* based on the unit's data, 2) a *group time series model* estimated from the pooled data of a selected group of related areal units, and 3) *a pooled model* that combines individual and group parameter estimates according to relative precision of estimates. The critical step in pooling is choosing the best areal units to make up the group for the areal unit being forecasted. Alternatives grouping approaches include using nearby areal units (contiguity) or areal units with correlating time series patterns (analogy). See Duncan, Gorr, and Szczypula (1999).

Forecasting Requirements of Police

The decisions most in need of forecasts by police are tactical deployment of existing resources such as directed patrol, focused enforcement, and hot spot enforcement. These are the kinds of decisions made in the popular COMPSTAT-style policing. Specific short-term forecasting needs include 1) counter-factual forecasts for evaluating police interventions, 2) time trend and seasonality forecasts for manpower allocations, and 3) hot spot forecasts for preventing serious crimes. We discuss each in turn.

Evaluation of Police Using Counter-Factual Forecasts:

A counter-factual provides the basis of comparison for judging whether or not there has been a change in crime. Two important kinds of changes are 1) *changes in crime levels* and 2) *changes in crime patterns*. To estimate changes in crime levels, while controlling for seasonality, police commonly calculate 12 month differences or percentage differences. If t is the most recent month for which we have crime data and A_t is the actual crime level for month t, then the percentage level change is $L_t = 100*(A_t - A_{t-12})/A_{t-12}$. An issue with this measure is that it ignores long-term trends. L_t includes a whole year's accumulation of level trend changes.

An alternative crime level change calculation that controls for seasonality is first differences of deseasonalized data: $L'_t = 100*(A'_t - A'_{t-1})/A'_{t-1}$ where $A'_t = A_t/S_t$ for multiplicative seasonality or $A'_t = A_t - S_t$ for additive seasonality and S_t is the appropriate seasonal estimate for the season of time t. The benefits of L'_t are that it focuses on the most recent change, while eliminating seasonal changes.

Changes in crime patterns address questions such as "Have we stopped the increases in robberies that occurred this past year?" The appropriate measure is $P_t = 100*(A_t - F_t)/F_t$, where F_t is the crime forecast made with data available up through t-1. By definition, F_t extrapolates or carries the status quo pattern forward in time, adjusting for known trends and seasonality.

For example, serious crime was on an increasing trend in Pittsburgh, Pennsylvania during the early 1990s due to the crack epidemic and rise of youth gangs. In Pittsburgh's Zone 2 (equivalent of a precinct), the January 1991 number of robberies of person was 20. In January 1992 the number had risen steadily by approximately one robbery per month to 32 robberies. Hence, there was no change of crime pattern (P_t = 0%), although the robberies of persons level had risen dramatically, L_t = 60%.

Manpower Scheduling Using Seasonality Forecasts:

Seasonality is especially useful for forecasting and scheduling manpower. For example, training and vacations might be scheduled in low crime months, while focused interventions may be targeted to high crime months. Crime maps are useful in targeting specific areas and times. Populations and land uses have widely varying characteristics across jurisdictions, hence we can expect that crime seasonality will also vary in patterns and magnitudes for different areas.

There are two forms of seasonality: additive or multiplicative adjustments to a time trend model. Additive adjustments, in units of number of crimes, reflect the scale of crime in different areas. Hence such seasonal factors are not readily transferable or comparable across areal units. Multiplicative adjustments have the advantage of being dimensionless and thus are well suited for application across areal units of differing crime scales.

For example, Gorr et al. [1998] used classical decomposition to estimate multiplicative seasonality of robberies of persons for two adjacent Pittsburgh police zones and all of Pittsburgh using five years of data. Zone 2 encompasses downtown Pittsburgh and nearby neighborhoods while Zone 6 includes a second urban center (Oakland).

Table 1 is the result. For Pittsburgh, the peak month is December with 1.2 seasonality (20% higher than normal), while the lowest months are August and November with 0.8 seasonality (20% lower than usual). There is spatial heterogeneity evident when breaking seasonality down to particular zones. For example, Zone 2, with its peak of 1.3 in September, follows a somewhat different seasonality pattern than either Zone 6 or Pittsburgh. In contrast, while Zone 2 and Pittsburgh have sharply increasing seasonality from August to September, Zone 6 has sharply decreasing seasonality.

The implications of seasonality for policing are fairly strong. They permit police to anticipate and target manpower. For example, Zone 2 should schedule vacations and training in August and focus manpower on robberies of persons in September. Zone 6 should do the opposite.

Table 1
Multiplicative Seasonality Estimates

Month	Zone 2	Zone 6	Pittsburgh
1	0.9	0.9	0.9
2	0.9	0.8	0.9
3	1.1	1.0	1.0
4	0.9	1.0	1.0
5	1.1	1.0	1.0
6	1.0	0.8	1.0
7	1.0	1.1	1.1
8	0.8	1.1	0.8
9	1.3	0.9	1.1
10	1.1	1.2	1.1
11	0.8	0.8	0.8
12	1.0	1.3	1.2

Hot Spot Forecasts:

“Broken windows” is a current hypothesis on crime leading indicators. It states that soft crimes (e.g, vandalism, disturbance), if left unabated, increase and lead to hard crimes (e.g., assault, robbery; Wilson and Kelling 1982, Kelling and Coles 1996). If this theory is correct, then police can detect, monitor, and suppress new soft crime hot spots to prevent later hard crimes. Gorr et al. [1998] will test this hypothesis with data from Pittsburgh and Rochester, New York using BVAR models and Granger causality tests and data from periods prior to hot spot enforcement. After hot spot enforcement, we expect the leading-indicator value of soft crimes for hard crimes (if it existed at all) will be weakened.

An emergent hot spot will have a vector of inputs including leading crime mix and numbers, proximity of other hot spots, land uses and population characteristics in the hot spot and surrounding cells, and arrests for incidents in adjacent cells (for displacement). Additional modeling methods to consider beyond BVAR include linear discriminant models, multi-layer perceptron neural network models to permit nonlinear boundaries of hotspot classes (Lapedes and Farber 1987; White 1988; Hornik, Stinchcombe, and White 1989, and Olligschlaeger 1998), and classification and regression trees (CART) for prediction.

Olligschlaeger (1998) found that leading indicator models forecasted emergence in drug dealing (from zero to one or more drug 911 calls) more accurately than non-leading indicator models. For example, multivariate forecast models were about 25% more accurate than the no change, random walk forecast. This was a very stringent test of leading indicators because emergent drug markets are initially detected with only a few 911 calls.

Summary

The forecasting literature has been dominated by the development and comparison of short-term, time series forecasting methods. This is fortunate for the emerging field of crime forecasting, because the major decisions in need of forecasts are tactical deployment of existing resources – in the short term.

A challenge, however, awaits crime forecasters. Crime space/time series data tend to be volatile and have abrupt pattern changes because of their small scale. Promising, for strengthening extrapolative forecast models, are pooling methods that combine data from nearby areas or from distance, analogous areas.

Also promising is the “broken windows hypothesis” which posits that “soft crimes” are leading indicators of “hard crimes.” If this is true then, police will generate their own leading indicator data – a real coup in the forecasting field. Practically, there are no other real-time, small scale leading indicators of crime, externally or internally generated. Soft crime leading indicators are the one shot that we have at forecasting pattern changes in serious crimes.

Research is underway to determine the best crime space/time forecasting methods, both extrapolative and leading indicator methods.

References

Armstrong, S. [ed.] (1999), *Forecasting Principles*, to appear, Kluwer Academic Publishers.

Doan, T. R. B. Litterman, and C. Sims (1984), “Forecasting and Conditional Projections Using Realistic Prior Distributions,” *Econometric Reviews*, 3, 1-100.

Dowd, Michael R. and James P. LeSage (1998) “Analysis of Spatial Contiguity Influences on State Price Level Formation,” *International Journal of Forecasting*.

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.L. Gorr, and J. Szczypula, “Forecasting Analogous Time Series,” S. Armstrong [ed.] (1999), *Forecasting Principles*, to appear in, Kluwer Academic Publishers.

Gorr, W.L., A.M. Olligschlaeger, and J. Szczypula (1998), “Crime Hot Spot Forecasting: Modeling and Comparative Evaluation,” Grant No. 98-IJ-CX-K005, National Institute of Justice, Office of Justice programs, U.S. Department of Justice.

Hornik, K., Stinchcombe, M. and White, H. (1989), "Multilayer Feedforward Network ate Universal Approximators", in *Neural Networks*, Vol. 2, Pergamon Press

Kelling, G. L. and C.M. Coles (1996), *Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities*, NY: Free Press.

Lapedes, A. and Farber, R. (1987), "Nonlinear Signal Processing Using Neural Networks: Prediction and System Modeling", Technical Report LA-UR-87-2662, Theoretical Division, Los Alamos National Laboratory, NM.

LeSage, J.P. (1989), “Incorporating Regional Wage Relations in Local Forecasting Models With a Bayesian Prior,” *International Journal of Forecasting* 5, 37-47.

LeSage, J.P. (1990), “Forecasting Turning Points in Metropolitan Employment Growth Rates Using Bayesian Techniques,” *Journal of Regional Science* 30, pp. 533-548.

LeSage, J.P. and Z. Pan (1995), “Using Spatial Contiguity as Bayesian Prior Information in Regional Forecasting Models',” *International Regional Science Review*, 18:1, pp. 33-53.

Litterman, R.B., *Techniques for Forecasting with Vector Autoregressions,* Ph.D. Dissertation, University of Minnesota 1980.

Litterman, R.B. (1986). “Forecasting with Bayesian Vector Autoregressions -- Five Years of Experience,” *Journal of Business & Economic Statistics*, 4:1, pp. 25-38.

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.

National Institute of Justice (1998), “Solicitation for Policing Research and Evaluation: Fiscal Year 1998;” Section VI, The Impact of Technology on Policing; part B, Developing Predictive Models.

Olligschlaeger, A. M. (1998), “Artificial Neural Networks and Crime Mapping”, in Weisburd, D. and McEwen, T. (eds.) *Crime Mapping, Crime Prevention*, Crime Prevention Studies 8, Criminal Justice Press, NY.

Wilson, J.Q. and G.L. Kelling (1982), “Broken Windows: The Police and Neighborhood Safety,” *Atlantic Monthly* 249: 29- 38.

White, H. (1988), "Economic Prediction Uring Neural Networks: The Case of IBM Daily Stick Returns", *Proceedings of the IEEE International Conference on Neural Networks,* San Diego.