Journal of Business & Economic Statistics, October 1989, Vol. 7, No. 4

Time Series Models for Count or Qualitative Observations

A. C. Harvey and C. Fernandes
Department of Statistical and Mathematical Sciences, London School of Economics, London WC2A 2AL, United Kingdom

Time series sometimes consist of count data in which the number of events occurring in a given time interval is recorded. Such data are necessarily nonnegative integers, and an assumption of a Poisson or negative binomial distribution is often appropriate. This article sets ups a model in which the level of the process generating the observations changes over time. A recursion analogous to the Kalman filter is used to construct the likelihood function and to make predictions of future observations. Qualitative variables, based on a binomial or multinomial distribution, may be handled in a similar way. The model for count data may be extended to include explanatory variables. This enables nonstochastic slope and seasonal components to be included in the model, as well as permitting intervention analysis. The techniques are illustrated with a number of applications, and an attempt is made to develop a model-selection strategy along the lines of that used for Gaussian structural time series models. The applications include an analysis of the results of international football matches played between England and Scotland and an assessment of the effect of the British seat-belt law on the drivers of light-goods vehicles.

1. INTRODUCTION

It is not unusual to find time series consisting of *count data*. Such series record the number of events occurring in a given interval and are necessarily nonnegative integers. An example would be the number of accidents occurring in a given period. The number of goals scored by England against Scotland in international football matches also has the characteristics of count data. Count-data models are usually based on distributions such as the Poisson or negative binomial. If the means of these distributions are constant or can be modeled in terms of observable variables, then estimation is relatively easy; see, for example, the book on generalized linear models (GLIM) by McCullagh and Nelder (1983). The essence of time series models, however, is that the mean of a series cannot be modeled in terms of observable variables but depends on some stochastic mechanism.

The nature of count data makes autoregressive integrated moving average (ARIMA) models inappropriate unless the values of the observations are large enough to justify the assumption of normally distributed disturbances as a reasonable approximation. The structural time series models used, for example, by Harvey and Durbin (1986) or Kitagawa and Gersch (1984) offer a way out of these difficulties because they are set up in terms of permanent and transitory components, which are driven by separate disturbance terms. The simplest structural model, the local level plus noise, takes the form

$$y_t = \mu_t + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{1.1}$$

and

$$\mu_t = \mu_{t-1} + \eta_t, \tag{1.2}$$

where μ_t is a permanent or level component that can move up or down because of the disturbance term η_t and ε_t is a transitory disturbance term. If both η_t and ε_t are normally distributed, with zero means and variances σ_η^2 and σ_ε^2, respectively, then the model is equivalent to an ARIMA(0, 1, 1) process but is inappropriate for count data. The structural framework, however, allows one to specify ε_t in such a way that the distribution of y_t conditional on μ_t is Poisson or negative binomial. The stochastic process governing the evolution of μ_t should then be such that it is always positive.

The structural framework can also be adapted to handle qualitative variables. The simplest such case is binary data, in which the two possible outcomes are dependent on a binomial distribution. The setup may be generalized to several outcomes, dependent on a multinomial distribution.

The essence of the problem is to formulate a model that allows the distribution of y_t, given past observations, to be obtained. If this can be done, the likelihood function can be formed and used as the basis for estimating unknown parameters in the model. Predictions of future observations may then be made. The solution to the problem rests on the use of natural-conjugate distributions of the type used in Bayesian statistics. Our approach, however, is still essentially a classical one, and we attempt to develop a model-fitting procedure based on the kind of methodology used by Harvey and Durbin (1986).

We focus our attention on formulating models for count and qualitative data that are analogous to (1.1) and (1.2) in that they allow the underlying mean of the process to change over time. By introducing a hyperparameter, ω, into these local-level models, past ob-

servations are discounted in making forecasts of future observations. Indeed it transpires that in all cases the predictions can be constructed by an *exponentially weighted moving average* (EWMA) procedure. This is exactly what happens in (1.1) and (1.2) under the normality assumption (see Muth 1960).

Explanatory variables can be introduced into our local-level models via the kind of link functions that appear in GLIM models. Time trends and seasonal effects can be introduced as special cases. The framework does not extend to allowing these effects to be stochastic, as is typically the case in linear structural models. We believe that this is not a serious restriction. Even with data on continuous variables, it is not unusual to find that the slope and seasonal effects are close to being deterministic (e.g., see Harvey and Durbin 1986; Harvey and Todd 1983). With count and qualitative data, it seems even less likely that the observations will provide enough information to pick up changes in the slope and seasonal effects over time.

The use of natural-conjugate distributions to formulate local-level models for count and qualitative observations was suggested by Smith (1979, 1981). He observed that such procedures gave rise to EWMA predictions. His models were set up within a Bayesian framework, however, and he did not advocate the estimation of the hyperparameter, ω, by maximum likelihood (ML). The same is true of the work of West, Harrison, and Migon (1985) and West and Harrison (1986), in which various approximations are used to tackle a more general problem in which components other than the level are allowed to change over time. The approach adopted in this article is more like that employed by Smith and Miller (1986) in their study on predicting records in athletics. Their article is not concerned with count or qualitative observations, however, but with observations that, when transformed, are exponentially distributed.

The plan of the article is as follows. Section 2 sets out the basis of our approach and applies it to modeling Poisson data. Section 3 deals with data from a binomial distribution, and Section 4 makes the extension to the multinomial distributions. The negative binomial is examined in Section 5. Explanatory variables are introduced into the models in Section 6, and Section 7 deals with model selection and applications.

2. OBSERVATIONS FROM A POISSON DISTRIBUTION

Suppose that the observation at time t is drawn from a Poisson distribution,

$$p(y_t \mid \mu_t) = \mu_t^{y_t} e^{-\mu_t}/y_t!. \tag{2.1}$$

This corresponds to the measurement equation of (1.1).

The conjugate prior for a Poisson distribution is the gamma distribution. Let $p(\mu_{t-1} \mid Y_{t-1})$ denote the pdf of μ_{t-1} conditional on the information at time $t-1$. Suppose that this distribution is gamma; that is, it is given by

$$p(\mu; a, b) = \frac{e^{-b\mu}\,\mu^{a-1}}{\Gamma(a)b^{-a}}, \quad a, b > 0, \tag{2.2}$$

with $\mu = \mu_{t-1}$, $a = a_{t-1}$, and $b = b_{t-1}$, where a_{t-1} and b_{t-1} are computed from the first $t-1$ observations, Y_{t-1}. In Models (1.1) and (1.2) with normally distributed observations, $\mu_{t-1} \sim N(m_{t-1}, p_{t-1})$ at time $t-1$ implies that $\mu_t \sim N(m_{t-1}, p_{t-1} + \sigma_\eta^2)$ at time $t-1$. In other words, the mean of $\mu_t \mid Y_{t-1}$ is the same as that of $\mu_{t-1} \mid Y_{t-1}$, but the variance increases. The same effect can be induced in the gamma distribution by multiplying a and b by a factor less than 1. We therefore suppose that $p(\mu_t \mid Y_{t-1})$ follows a gamma distribution with parameters $a_{t|t-1}$ and $b_{t|t-1}$ such that

$$a_{t|t-1} = \omega a_{t-1}, \tag{2.3}$$

$$b_{t|t-1} = \omega b_{t-1}, \tag{2.4}$$

and $0 < \omega \leqslant 1$. Then

$$E(\mu_t \mid Y_{t-1}) = a_{t|t-1}/b_{t|t-1} = a_{t-1}/b_{t-1} = E(\mu_{t-1} \mid Y_{t-1}),$$

whereas

$$\begin{aligned}\operatorname{var}(\mu_t \mid Y_{t-1}) &= a_{t|t-1}/b_{t|t-1}^2 \\ &= \omega^{-1}\operatorname{var}(\mu_{t-1} \mid Y_{t-1}).\end{aligned}$$

The stochastic mechanism governing the transition of μ_{t-1} to μ_t is, therefore, defined implicitly rather than explicitly. It is possible, however, to show that it is formally equivalent to a multiplicative transition equation of the form $\mu_t = \omega^{-1}\mu_{t-1}\eta_t$, where η_t has a beta distribution, of the form (3.2), with parameters ωa_{t-1} and $(1-\omega)a_{t-1}$ (see Smith and Miller 1986).

Once the observation y_t becomes available, the posterior distribution, $p(\mu_t \mid Y_t)$, is given by a gamma distribution with parameters

$$a_t = a_{t|t-1} + y_t \tag{2.5}$$

and

$$b_t = b_{t|t-1} + 1. \tag{2.6}$$

The initial prior gamma distribution—that is, the distribution of μ_t at time $t = 0$—tends to become diffuse, or noninformative, as $a, b \to 0$, although it is actually degenerate at $a = b = 0$ with $\Pr(\mu = 0) = 1$. None of this prevents the recursions [(2.3)–(2.6)] being initialized at $t = 0$ with $a_0 = b_0 = 0$, however. A proper distribution for μ_t is then obtained at time $t = \tau$, where τ is the index of the first nonzero observation. It follows that, conditional on Y_τ, the joint density of the observations $y_{\tau+1}, \ldots, y_T$ is

$$p(y_{\tau+1}, \ldots, y_T; \omega) = \prod_{t=\tau+1}^{T} p(y_t \mid Y_{t-1}). \tag{2.7}$$

The predictive pdf's are given by

$$p(y_t \mid Y_{t-1}) = \int_0^\infty p(y_t \mid \mu_t)\, p(\mu_t \mid Y_{t-1})\, d\mu_t, \tag{2.8}$$

and, for Poisson observations and a gamma prior, this

operation yields a negative binomial distribution

$$p(y_t \mid Y_{t-1}) = \binom{a + y_t - 1}{y_t} b^a(1 + b)^{-(a+y_t)}, \tag{2.9}$$

where $a = a_{t|t-1}$ and $b = b_{t|t-1}$ and

$$\binom{a + y_t - 1}{y_t} = \frac{\Gamma(a + y)}{\Gamma(y + 1)\,\Gamma(a)},$$

although, since y is an integer, $\Gamma(y + 1) = y!$ Hence the log-likelihood function for the unknown hyperparameter ω is

$$\log L(\omega) = \sum_{t=\tau+1}^{T} \{\log \Gamma(a_{t|t-1} + y_t) - \log y_t! - \log \Gamma(a_{t|t-1}) + a_{t|t-1}\log b_{t|t-1} - (a_{t|t-1} + y_t)\log(1 + b_{t|t-1})\}. \tag{2.10}$$

It follows from the properties of the negative binomial that the mean and variance of the predictive distribution of y_{T+1}, given Y_T, are, respectively,

$$\tilde{y}_{T+1|T} = E(y_{T+1} \mid Y_T) = a_{T+1|T}/b_{T+1|T} = a_T/b_T \tag{2.11}$$

and

$$\text{var}(y_{T+1} \mid Y_T) = a_{T+1|T}\,(1 + b_{T+1|T}) \mid b^2_{T+1|T} = \omega^{-1}\text{var}(\mu_T \mid Y_T) + E(\mu_T \mid Y_T). \tag{2.12}$$

Repeated substitution from (2.3)–(2.6) shows that the forecast function is

$$\tilde{y}_{T+1|T} = a_T/b_T = \sum_{j=0}^{T-1} \omega^j y_{T-j} \Big/ \sum_{j=0}^{T-1} \omega^j. \tag{2.13}$$

This is a weighted mean in which the weights decline exponentially. It has exactly the same form as the discounted least squares estimate of a mean. In large samples, the denominator of (2.13) is approximately equal to $1/(1 - \omega)$ when $\omega < 1$ and the forecasts can be obtained recursively by the EWMA scheme

$$\tilde{y}_{t+1|t} = (1 - \lambda)\,\tilde{y}_{t|t-1} + \lambda y_t, \qquad t = 1, \ldots, T, \tag{2.14}$$

where $y_{1|0} = 0$ and $\lambda = 1 - \omega$ is the smoothing constant. When $\omega = 1$, the right side of (2.13) is equal to the sample mean. Regarding this as an estimate of μ, the choice of zeros as initial values for a and b in the filter is seen to be justifed insofar as it yields the classical solution. It is also worth noting that, unlike the Gaussian case, no approximations are involved in the use of a diffuse prior in this model.

Now consider multistep prediction. The l-step-ahead predictive distribution at time T is given by

$$p(y_{T+l} \mid Y_T) = \int_0^\infty p(y_{T+l} \mid \mu_{T+l})\, p(\mu_{T+l} \mid Y_T)\, d\mu_{T+l}. \tag{2.15}$$

It could be argued that the assumption embodied in (2.3) and (2.4) suggests that $p(\mu_{T+l} \mid Y_T)$ has a gamma distribution with parameters

$$a_{T+l|T} = \omega^l a_T \tag{2.16}$$

and

$$b_{T+l|T} = \omega^l b_T. \tag{2.17}$$

This would mean that the predictive distribution for y_{T+l} was negative binomial (2.9), with a and b given by $a_{T+l|T}$ and $b_{T+l|T}$ in (2.16) and (2.17). Unfortunately, the evolution that this implies for μ_t is not consistent with that which would occur if observations were made at times $T + 1, T + 2, \ldots, T + l - 1$. In the latter case, the distribution of y_{T+l} at time T is

$$p(y_{T+l} \mid Y_T) = \sum_{y_{T+l-1}} \cdots \sum_{y_{T+1}} \prod_{j=1}^{l} p(y_{T+j} \mid Y_{T+j-1}). \tag{2.18}$$

It is difficult to derive a close-form expression for $p(y_{T+l|T})$ from (2.18) for $l > 1$, but it can, in principle, be evaluated numerically.

Although finding a closed-form expression for $p(y_{T+l} \mid Y_T)$ is difficult, it is possible to show that

$$E(y_{T+l} \mid Y_T) = a_T/b_T \tag{2.19}$$

for all lead times. To see this result, first note that taking the conditional expectation of y_{T+l} at time $T + l - 1$ gives, from (2.11),

$$\underset{T+l-1}{E}(y_{T+l}) = a_{T+l-1}/b_{T+l-1}.$$

Using (2.3)–(2.6) and taking conditional expectations at time $T + l - 2$ gives

$$\underset{T+l-2}{E}\,\underset{T+l-1}{E}(y_{T+l}) = \underset{T+l-2}{E}\left[\frac{\omega a_{T+l-2} + y_{T+l-1}}{\omega b_{T+l-2} + 1}\right] = \frac{a_{T+l-2}}{b_{T+l-2}}, \qquad l \geq 2.$$

Repeating this procedure by taking conditional expectations at time $T + l - 3$ and so on gives (2.19).

3. BINOMIAL DISTRIBUTION

If the observations at time t are generated from a binomial distribution then

$$p(y_t \mid \pi_t) = \binom{n_t}{y_t} \pi_t^{y_t}(1 - \pi_t)^{n_t - y_t}, \qquad y_t = 0, \ldots, n_t, \tag{3.1}$$

where π is the probability that y_t is unity when n_t is 1. The value of n_t is assumed to be fixed and known. Thus observations from the binomial can be regarded as a special case of count data in which there is a fixed number of opportunities for the event in question to occur. When n_t is 1, the data are *binary* or *dichotomous*.

The conjugate prior for the binomial distribution is

the beta distribution

$$p(\pi; a, b) = [B(a, b)]^{-1}\pi^{a-1}(1 - \pi)^{b-1}, \quad (3.2)$$

where the beta function is

$$B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b).$$

Let $p(\pi_{t-1} \mid Y_{t-1})$ have a beta distribution with parameters a_{t-1} and b_{t-1}. Assume that $p(\pi_t \mid Y_{t-1})$ is also beta with parameters given by equations exactly the same as those in (2.3) and (2.4). This again ensures that the mean of $\pi_t \mid Y_{t-1}$ is the same as that of $\pi_{t-1} \mid Y_{t-1}$, but the variance increases. Specifically

$$E(\pi_t \mid Y_{t-1}) = a_{t|t-1}/(a_{t|t-1} + b_{t|t-1}) = a_{t-1}/(a_{t-1} + b_{t-1})$$

and

$$\operatorname{var}(\pi_t \mid Y_{t-1}) = \frac{a_{t|t-1}b_{t|t-1}}{(a_{t|t-1} + b_{t|t-1})^2(a_{t|t-1} + b_{t|t-1} + 1)} = \frac{a_{t-1}b_{t-1}}{(a_{t-1} + b_{t-1})^2(\omega a_{t-1} + \omega b_{t-1} + 1)}.$$

Once the tth observation becomes available, the distribution of $\pi_t \mid Y_t$ is beta with parameters

$$a_t = a_{t|t-1} + y_t \quad (3.3)$$

and

$$b_t = b_{t|t-1} + n_t - y_t. \quad (3.4)$$

The predictive distribution, $p(y_t \mid Y_{t-1})$, is beta-binomial

$$p(y_t \mid Y_{t-1}) = \frac{1}{n_t + 1}\frac{B(a + y_t, b + n_t - y_t)}{B(y_t + 1, n_t - y_t + 1)B(a, b)}, \quad (3.5)$$

where $a = a_{t|t-1}$ and $b = b_{t|t-1}$. The likelihood function is again (2.7) with τ defined as the first time period for which

$$0 < \sum_{t=1}^{\tau} y_t < \sum_{t=1}^{\tau} n_t.$$

This condition ensures that a_τ and b_τ are strictly positive, although again there is nothing to prevent us starting the recursions (2.3)–(2.4) and (3.3)–(3.4) at $t = 1$ with $a_0 = b_0 = 0$ (see Lehmann 1983, p. 243). [This does not correspond to the use of a uniform prior. Since the range of a beta distribution is between 0 and 1, a uniform prior is a proper prior and the summation in the likelihood runs from $t = 1$ to T. A uniform distribution is obtained by setting the beta parameters, a and b, equal to unity; alternatively, Jeffreys (1961) argued that it is more appropriate to have $a = b = \frac{1}{2}$, corresponding to arc $\sin(\pi\frac{1}{2})$ having a uniform distribution.]

From the properties of the beta-binomial distribution, the mean and variance of y_{T+1} conditional on the information at time T are

$$\tilde{y}_{T+1|T} = E(y_{T+1} \mid Y_T) = n_{T+1}a_T/(a_T + b_T) \quad (3.6)$$

and

$$\operatorname{var}(y_{T+1} \mid Y_T) = \frac{n_{T+1}a_Tb_T(a_T + b_T + \omega^{-1}n_{T+1})}{(a_t + b_T)^2(a_T + b_T + \omega^{-1})}. \quad (3.7)$$

By substituting repeatedly from the recursive equations (3.3) and (3.4), it can be seen that, for n_t constant, $\tilde{y}_{T+1|T}$ is effectively an EWMA.

For binary data, $n_t = 1$, the beta-binomial distribution in (3.5) reduces to a binomial distribution with

$$\Pr(y_t = 1 \mid Y_{t-1}) = a_{t|t-1}/(a_{t|t-1} + b_{t|t-1}).$$

As regards forecasts, it can be shown, by evaluating (2.18), that the distribution of $y_{T+l} \mid Y_T$ is binomial with expected value $a_T/(a_T + b_T)$ for all $l = 1, 2, 3, \ldots$. Hence its variance does not increase with the lead time.

The model for binary data can easily be extended to handle Markov chains in which there are two parameters evolving over time, but the one which pertains at time t depends on the observation at $t - 1$. Thus if y_{t-1}, then $\Pr(y_t = 1) = \pi_{t|t-1}$, whereas if $y_{t-1} = 0$, then $\Pr(y_t = 1) = \pi^*_{t|t-1}$. When $y_{t-1} = 1$, the parameters $a_{t|t-1}$ and $b_{t|t-1}$, associated with $\pi_{t|t-1}$, are used to form the predictive distribution for y_t and are updated via (3.3) and (3.4) when y_t becomes available. From the point of view of π^*_t, y_t is treated as though it were missing. When $y_{t-1} = 0$, the situation is reversed.

4. MULTINOMIAL DISTRIBUTON

When there are more than two categories, the observations are said to be *polytomous* and the multinomial distribution is appropriate. Let there be $m + 1$ possible categories, and suppose that the probability that at time t an object belongs to the ith category is π_{it}. If there are n_t trials and the number of objects in the ith category is y_{it}, then

$$p(y_{ot}, \ldots, y_{mt}) = \binom{n_t}{y_{1t}, \ldots, y_{mt}} \prod_{i=0}^{m} \pi_{it}^{y_{it}}, \quad (4.1)$$

with $\sum_{i=0}^{m} y_{it} = n_t$ and $\sum_{i=0}^{m} \pi_{it} = 1$.

The conjugate prior for the multinomial distribution is the multivariate beta or *Dirichlet* distribution

$$p(\pi_0, \ldots, \pi_m; a_0, \ldots, a_m) = \frac{\Gamma(\sum a_i)}{\prod \Gamma(a_i)} \prod_{i=0}^{m} \pi_i^{a_i - 1}, \quad (4.2)$$

where the summations are from $i = 0$ to m. (When $m = 1$, this collapses to the beta distribution with $a_0 = a$ and $a_1 = b$.) Proceeding as in the previous section, it is not difficult to show that the recursive equations corresponding to (2.3) and (3.3) become

$$a_{i,t|t-1} = \omega a_{i,t-1} \quad (4.3)$$

and

$$a_{i,t} = a_{i,t|t-1} + y_{it}, \qquad i = 0, \ldots, m. \tag{4.4}$$

The likelihood for ω is as in (2.7), with τ the first value of t, which yields $a_{i,t} > 0$ for all $i = 0, \ldots, m$. The predictive distribution in this case is known as the Dirichlet multinomial. The forecasts can again be expressed in terms of EWMA's.

5. NEGATIVE BINOMIAL

The negative binomial distribution is

$$p(y_t \mid \pi_t) = \binom{v + y_t - 1}{y_t} \pi_t^v (1 - \pi_t)^{y_t}, \qquad y_t = 0, 1, 2, \ldots, \tag{5.1}$$

where $0 < \pi_t < 1$ and $v > 0$. This is known as the Pascal distribution if v is an integer and $v = 1$ corresponds to the geometric distribution. The mean and variance are

$$E(y_t \mid \pi_t) = v(1 - \pi)/\pi \tag{5.2}$$

and

$$\operatorname{var}(y_t \mid \pi_t) = E(y_t \mid \pi_t)[1 + v^{-1} E(y_t \mid \pi_t)]. \tag{5.3}$$

The distribution, therefore, exhibits overdispersion compared with the Poisson distribution; that is, the variance exceeds the mean. If the mean is kept constant, however, the negative binomial tends toward the Poisson distribution as $v \to \infty$.

The conjugate prior distribution for the negative binomial is the beta distribution, (3.2). At first sight, it might appear that the recursions in (2.3) and (2.4) are again appropriate. In view of (5.2), however, it is the expected value of $(1 - \pi)/\pi$, rather than π, that needs to be kept constant while the variance increases. For a beta distribution (3.2),

$$E((1 - \pi)/\pi) = \int_0^1 \pi^{a-2}(1 - \pi)^b \, d\pi = \frac{B(a - 1, b + 1)}{B(a, b)} = \frac{b}{a - 1}, \tag{5.4}$$

provided that a > 1. Hence we require that

$$\frac{b_{t|t-1}}{a_{t|t-1} - 1} = \frac{b_{t-1}}{a_{t-1} - 1}. \tag{5.5}$$

This can be achieved by multiplying the numerator and denominator in the expression on the right side of (5.5) by ω. The prediction equation, (2.3), is, therefore, modified to

$$a_{t|t-1} = \omega a_{t-1} + (1 - \omega), \qquad 0 < \omega \leq 1, \tag{5.6}$$

but (2.4) remains unchanged. The updating equations have the more standard form

$$a_t = a_{t|t-1} + v \tag{5.7}$$

and

$$b_t = b_{t|t-1} + y_t. \tag{5.8}$$

The predictive distribution is the beta-Pascal

$$p(y_t \mid Y_{t-1}) = \frac{1}{v + y_t} \frac{B(v + a_{t|t-1}, y_t + b_{t|-1})}{B(v, y_t + 1)B(a_{t|t-1}, b_{t|t-1})}, \tag{5.9}$$

and the likelihood function is as in (2.7) with τ the first value of t for which y_t is nonzero.

The expected value of the one-step-ahead predictive distribution at time T is

$$\tilde{y}_{T+1|T} = E(y_{T+1} \mid Y_T) = \int_0^1 E(y_{T+1} \mid \pi_{T+1}) p(\pi_{T+1} \mid Y_T) \, d\pi_{T+1}$$

$$\frac{vb_{T+1|T}}{a_{T+1|T} - 1} = \frac{vb_T}{a_T - 1} \tag{5.10}$$

in view of (5.4) and the way in which prediction operates via (5.5). As in the previous models, b_T can be written as an exponentially weighted average of past weights, and repeatedly substituting from (5.6) and (5.7) gives

$$a_T = v \sum_{j=0}^{T-1} \omega^j + (1 - \omega) \sum_{j=0}^{T-1} \omega^j. \tag{5.11}$$

As $T \to \infty$, $(a_T - 1) \to v/(1 - \omega)$, so $\tilde{y}_{T+1|T}$ again has the EWMA form of (2.13). Furthermore, using an argument similar to that employed to show (2.19), it is possible to verify that the forecasts l steps ahead are also given by (5.10).

The parameter v can be estimated by ML along with ω. Alternatively, it may be preset (see Cameron and Trivedi 1986).

6. EXPLANATORY VARIABLES

Explanatory variables may be brought into the random-walk-plus-noise model [(1.1)–(1.2)] with normal disturbances by extending (1.1), so it becomes

$$y_t = \mu_t + x_t' \delta + \varepsilon_t, \tag{6.1}$$

where x_t is a $k \times 1$ vector of exogenous explanatory variables and δ is a $k \times 1$ vector of parameters. If δ were known, the optimal prediction of y_t given Y_{t-1} would be $E(\mu_t \mid Y_{t-1}) + x_t'\delta$, and the updating of the estimate of the state would be carried out in the same way as for (1.1) and (1.2) but with y_t replaced by $y_t - x_t'\delta$ in the Kalman filter. When δ is unknown, the linearity of the model may be exploited to concentrate δ out of the likelihood function. Thus nonlinear optimization only needs to be carried out with respect to the signal–noise ratio, $q = \sigma_\eta^2/\sigma_\varepsilon^2$.

In a model with Poisson, binomial, or negative binomial observations but no dynamic structure, explanatory variables are introduced via a *link* function; see the discussion of the GLIM framework by McCullagh and Nelder (1983). For a Poisson model, the exponen-

tial link function

$$\mu_t = \exp(x_t'\delta) \tag{6.2}$$

ensures that μ_t remains positive. For a binomial model, the logit link function

$$\pi_t = \exp(x_t'\delta)/\{1 + \exp(x_t'\delta)\}$$

ensures that π_t remains between 0 and 1. Recent applications of GLIM models to count data include those of Hausman, Hall, and Griliches (1984), who studied the number of patents applied for by firms, and Cameron and Trivedi (1986), who studied the number of consultations with a doctor. Note that it may be unsatisfactory to include a lagged dependent variable as an explanatory variable when the observations are small and discrete.

Explanatory variables can be introduced into models for count data as follows. Consider the Poisson model of Section 2. As in (6.1), the level μ_t may be thought of as a component that has a separate effect from that of the explanatory variables in x_t, none of which is a constant. Suppose that $\mu_{t-1} \sim \Gamma(a_{t-1}, b_{t-1})$ and that, conditional on Y_{t-1}, $\mu_t \sim \Gamma(\omega a_{t-1}, \omega b_{t-1})$. This level component may be combined multiplicatively with an exponential link function for the explanatory variables so that the distribution of y_t conditional on μ_t is Poisson with mean

$$\mu_t^\dagger = \mu_t \exp(x_t'\delta). \tag{6.3}$$

It follows from the properties of the gamma distribution that, conditional on Y_{t-1}, $\mu_t^\dagger \sim \Gamma(a_{t|t-1}^\dagger, b_{t|t-1}^\dagger)$, where

$$a_{t|t-1}^\dagger = \omega a_{t-1}$$

and

$$b_{t|t-1}^\dagger = \omega b_{t-1}\exp(-x_t'\delta), \tag{6.4}$$

respectively. The log-likelihood of the observations is, therefore, as in (2.10) with $a_{t|t-1}$ and $b_{t|t-1}$ replaced by $a_{t|t-1}^\dagger$ and $b_{t|t-1}^\dagger$. This must be maximized with respect to ω and δ.

As regards updating, $\mu_t^\dagger \sim \Gamma(a_t^\dagger, b_t^\dagger)$, where $a_t^\dagger$ and $b_t^\dagger$ are obtained from $a_{t|t-1}^\dagger$ and $b_{t|t-1}^\dagger$ via updating equations of the form (2.5)–(2.6). Therefore, the posterior distribution of μ_t is $\Gamma(a_t, b_t)$, where a_t and b_t are given by

$$a_t = \omega a_{t-1} + y_t \tag{6.5}$$

and

$$b_t = \omega b_{t-1} + \exp(x_t'\delta), \qquad t = \tau + 1, \ldots, T. \tag{6.6}$$

Thus the only amendment as compared with the recursions in Section 2 is the replacement of unity by $\exp(x_t'\delta)$ in the equation for b_t.

For a given value of δ, we can proceed as in (2.19) to show that the mean of the predictive distribution of y_{T+l} is

$$\underset{T}{E}(y_{T+l}) = \exp(x_{T+l}'\delta) \sum \omega^j y_{T-j} \Big/ \sum \omega^j \exp(x_{T-j}'\delta)$$
$$= \exp(x_{T+l}'\delta)\text{EWMA}(y)/\text{EWMA}[\exp(x'\delta)], \tag{6.7}$$

where EWMA(y) is given by (2.13) and EWMA [$\exp(x'\delta)$] is defined similarly.

It is interesting to compare (6.7) with the result obtained from the Gaussian model (6.1) for a given discount factor, ω. Since the level and explanatory variables are combined multiplicatively in (6.3), it seems sensible to make the comparison with a Gaussian model in which logarithms have been taken. The optimal estimator of μ_t is obtained by applying the EWMA operation to $\log y_t - x_t'\delta$. The optimal estimate of $\log y_{T+l}$ can then be expressed as

$$\underset{T}{E}(\log y_{T+l})$$
$$= x_{T+l}'\delta\text{EWMA}(\log y) - \text{EWMA}(x'\delta). \tag{6.8}$$

The other point of comparison with the Gaussian model is in the maximization of the respective likelihood functions. In the Gaussian case, the computational burden is eased considerably by the fact that δ may be concentrated out of the likelihood function by estimating it by generalized least squares (see Kohn and Ansley 1985). This suggests that it may be possible to use estimates from the Gaussian model as starting values; the difficulty lies in how to handle zero observations when logarithms are being taken.

Explanatory variables can also be introduced into a Gaussian model via the level. Thus (6.1) and (1.2) become

$$y_t = \mu_t^* + \varepsilon_t \tag{6.9}$$

and

$$\mu_t^* = \mu_{t-1}^* + (\Delta x_t)'\delta + \eta_t. \tag{6.10}$$

This is equivalent to (6.1) and (1.2), the only difference being that μ_t^* includes the effect of the exogenous variables, whereas μ_t does not; in fact, $\mu_t^* = \mu_t + x_t'\delta$. Following Smith and Miller (1986), explanatory variables can be brought into a Poisson model in an analogous way. If $\mu_{t-1}^* \sim \Gamma(a_{t-1}^*, b_{t-1}^*)$, then, conditional on Y_{t-1}, μ_t^* is gamma with

$$a_{t|t-1}^* = \omega a_{t-1}^*, \qquad b_{t|t-1}^* = \omega b_{t-1}^*\exp(-\Delta x_t'\delta). \tag{6.11}$$

The posterior distribution of μ_t^* is then $\Gamma(a_t^*, b_t^*)$, where a_t^* and b_t^* are obtained from $a_{t|t-1}^*$ and $b_{t|t-1}^*$ via equations of the form (2.5)–(2.6). Repeated substitutions show that

$$b_T^* = 1 + \sum_{j=1}^{T-1} \omega^j \exp(-\Delta_j x_T'\delta),$$

where $\Delta_j = 1 - L^j$, and further rearrangement yields a forecast function that is exactly the same as (6.7).

The appropriate way of proceeding with the negative binomial is to introduce the explanatory variables directly into the distribution of $y_t \mid \pi_t$ via an exponential link function. This may be done by replacing v with $v_t = v \exp(x_t' \delta)$. Such a negative binomial distribution has, for a constant, π, a constant variance–mean ratio (see Cameron and Trivedi 1986, p. 33). Proceeding in this way leads to the updating equations (5.7) being modified to

$$a_t = a_{t|t-1} + v \exp(x_t' \delta), \tag{6.12}$$

but (5.8) remains unchanged. Combining the prediction and updating equation for a_t gives

$$a_t = \omega a_{t-1} + (1 - \omega) + v \exp(x_t' \delta).$$

The mean of the predictive distribution is

$$\begin{aligned} \underset{T}{E}(y_{T+l}) &= v_{T+l}\, b_{T+l|T}/(a_{T+l|T} - 1) \\ &= v \exp(x_{T+l}' \delta) b_T/(a_T - 1), \end{aligned}$$

and it is not difficult to deduce that it can be expressed in terms of an equation identical to (6.7).

It was noted earlier that bringing stochastic slope and seasonals components into count-data models is not easy. They can, however, enter in a deterministic fashion, and this is done by treating them as explanatory variables. A slope is introduced by setting one of the elements of x_t equal to time, t. The seasonals are modeled by $s - 1$ explanatory variables constructed so that the seasonal effects sum to 0 over a period of one year. The form of (6.3) means that the trend and seasonals combine multiplicatively, just as in a logarithmic Gaussian model. As in such a model, the coefficient of the slope is to be interpreted as a growth rate, whereas the seasonal coefficients are multiplicative seasonal factors.

The extension of count-data models to include explanatory variables opens up the possibility of carrying out *intervention analysis*. The explanatory variables x_t are replaced or augmented by a variable w_t that is designed to pick up the effect of some event or policy change.

7. MODEL SELECTION AND APPLICATIONS FOR COUNT DATA

Many of the issues that arise in the selection of GLIM models are also relevant here. There is the additional problem of testing for serial correlation; however, this is currently under investigation.

The standardized (Pearson) residuals are defined by

$$v_t = \frac{y_t - E(y_t \mid Y_{t-1})}{SD(y_t \mid Y_{t-1})}. \tag{7.1}$$

If the parameters in the model are known, it follows from the decomposition of the likelihood in (2.7) that these residuals are independently distributed with mean 0 and unit variance. They are not, however, in general, identically distributed.

The following diagnostic checks are suggested:

1. An examination of the plot of the residuals against time and against an estimate of the level.
2. An examination of a plot of the cumulative sum (CUSUM) of the residuals against time. A test for systematic underprediction or overprediction may be effected by comparing the mean of the residuals with a standard normal distribution. This test is analogous to the recursive t test suggested for Gaussian regression models by Harvey and Collier (1977).
3. A check on whether the sample variance of the residuals is close to 1. A value greater than 1 indicates overdispersion relative to the model that is being fitted. (Note that since the mean of the residuals is not necessarily 0, the sample variance and raw second moment will not usually be the same.)
4. As in the standard GLIM models, a goodness-of-fit test may be carried out by comparing the sum of the squared residuals with a χ^2 distribution in which the number of degrees of freedom is equal to $T - \tau$ minus the number of fitted parameters. Whether or not such a distribution is a reasonable approximation under the null hypothesis that the model is correctly specified is, however, a matter for investigation. The same would apply to the goodness-of-fit statistic referred to in the GLIM literature as the deviance. This statistic is computed by subtracting the likelihood from the likelihood obtained by fitting the observation model exactly by using T parameters and then multiplying by 2. In the case of a Poisson observation equation, the likelihood given by fitting each observation exactly is

$$\log L^* = \sum_{t=1}^{T} (y_t \log y_t - y_t - \log y_t!).$$

An alternative measure of deviance for dynamic models would be to construct an L^* based on the predictive distribution with ω set equal to the ML estimate obtained for the actual model. The degrees of freedom would then be $T - \tau - k$. In a univariate Poisson model, the test statistic would then be as in expression (7.2) with the summation running from $t = \tau + 1$ to T.

5. An examination of the CUSUM-of-squares plot.

Post-sample predictive testing may also be carried out. For the model with Poisson observations, the post-sample predictive test statistic is

$$\begin{aligned} \xi(l) = {} & 2 \sum_{t=T+1}^{T+l} a_{t|t-1} \log(a_{t|t-1}/y_t b_{t|t-1}) \\ & - 2 \sum_{t=T+1}^{T+l} (a_{t|t-1} + y_t) \log(y_t + a_{t|t-1}/(1 + b_{t|t-1}) y_t), \end{aligned} \tag{7.2}$$

where $a_{t|t-1}$ and $b_{t|t-1}$ are computed from the recursions (2.5)–(2.6). In the special case in which y_t is 0, the term in $\xi(l)$ at time t is

$$-2\, a_{t|t-1}\log((1 + b_{t|t-1})/b_{t|t-1}).$$

Under the null hypothesis that the model is correctly specified, $\xi(l)$ is asymptotically χ^2 with l df. The test is analogous to the test developed by Chow (1960) for a Gaussian regression model. The derivation in the Appendix is based on the introduction of a dummy variable into the model for each of the observations in the post-sample period.

7.1 Goals Scored by England Against Scotland

Figure 1 shows the number of goals scored by England in international football matches played against Scotland at Hampden Park in Glasgow. Except during the war years, these matches were played in Glasgow every other year. (The year 1985 is also an exception; the match should have been played at Wembley, but Margaret Thatcher decreed that it be played in Scotland to save Londoners from the ravages of marauding Scottish football supporters.) With the observations treated as though they were evenly spaced, estimation of the Poisson–gamma model gave $\tilde{\omega} = .844$. The variance of the standardized residuals is 1.269, and a plot of them shows no indication of misspecification. A post-sample predictive test carried out over the last five observations gave no hint of model breakdown with $\xi(5) = 4.54$.

The forecasted value for the mean of future observations is .82. This corresponds to the forecast that would have been obtained from the Gaussian random-walk-plus-noise model (1.1)–(1.2) by setting $q = .029$. The general formula is

$$q = (1 + \omega^2 - 2\omega)/\omega. \qquad (7.3)$$

The full one-step-ahead predictive distribution is shown in Table 1. The smooth line in Figure 2 shows the estimate of the level of the process obtained by applying the fixed-interval smoothing algorithm to a random-walk-plus-noise model with $q = .029$. Since there is no firm theoretical foundation for constructing a smoother in this way, we will refer to it as the *quasi smoother*.

Table 1. Predictive Probability Distribution of Goals in Next Match

0 goal	1 goal	2 goals	3 goals	4 goals	>4 goals
.471	.326	.138	.046	.013	.005

Fitting the negative binomial-beta model using a crude grid search for v yields an estimate of $v = 5$ with the corresponding estimate of ω being .916. Thus the introduction of an adjustable scale parameter has resulted in less movement in the level. The variance of the standardized residuals is 1.077 and the prediction is 1.187. The likelihood function is relatively insensitive with respect to changes in v. Furthermore, its value at the maximum is only marginally greater than the maximized likelihood for the Poisson–gamma model. If an allowance is made for the extra parameter via the Akaike information criterion, the Poisson–gamma model gives a better fit.

We now consider the full set of results of England–Scotland matches, by including the matches played at Wembley in England. Playing at home tends to confer an advantage, so we extend the model by introducing a dummy explanatory variable, W_E, that takes a value of unity when England is at home, and is 0 when the team is away. The ML estimates of the parameters are $\tilde{\omega} = .903$ and $\tilde{\delta} = .498$. As expected, the estimate of δ is positive. The likelihood-ratio test statistic is 10.57; this statistic is asymptotically χ^2_1 under the null hypothesis that δ is 0 and so is clearly highly significant. Since exp(.498) is equal to 1.64, the results can be interpreted as saying that the expected number of goals scored by England rises by 64% when they are playing at home.

7.2 Purse Snatching in Chicago

In their textbook, McCleary and Hay (1980) listed a time series of reported purse snatchings in the Hyde Park neighborhood of Chicago. The observations are 28 days apart, starting in January 1968. McCleary and

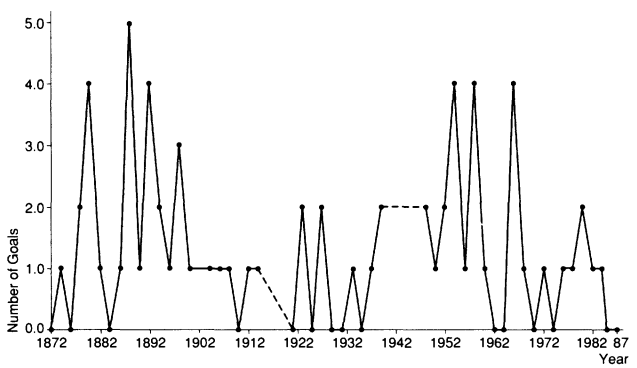

Figure 1. Goals Scored by England Against Scotland at Hampden Park, Glasgow.

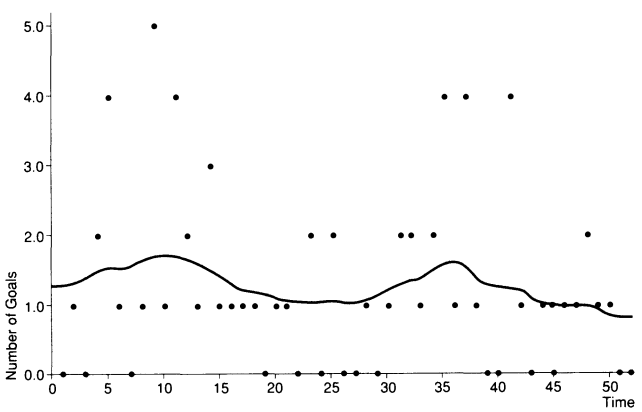

Figure 2. Goals Scored by England Against Scotland at Hampden Park, With Underlying Trend.

Hay decided that the series was stationary, and on the basis of the correlogram and the sample partial autocorrelation function they fitted an (autoregressive) AR(2) model.

The assumption of stationarity for this series implies that the level of purse snatchings remained constant throughout the period in question and that the variations observed were simply fluctuations around this constant level. This in turn implies that purse snatching is in some kind of equilibrium. Although this may be true, a more plausible working hypothesis is that the level of this crime is gradually changing over time. This suggests a random-walk-plus-noise model [(1.1)–(1.2)]. Estimating such a model under the normality assumption gives a signal–noise ratio of $q = .208$. The residuals give no indication of serial correlation. For example, $Q(8)$ is equal to 7.88, and this should be tested against a chi-square distribution with 7 df. The prediction-error variance is estimated to be 38.94, and this is only slightly above the figure reported by McCleary and Hay for their AR(2) model, which, of course, contains one more parameter. The practical implications for forecasting become apparent when it is noted that the forecasted level of future observations is 7.39 but the mean of the series is 13.92.

In summary, basic a priori considerations give rise to a structural time series model which not only has a clearer interpretation than the ARIMA model fitted by McCleary and Hay but is more parsimonious as well. The model is not, however, strictly speaking, data admissible. The forecast function is horizontal and cannot be negative, but a prediction interval of one root mean squared error on either side rapidly strays into the region of negative values of y. A logarithmic formulation, on the other hand, is not satisfactory because it fails the Bowman–Shenton normality test. A much better model is obtained by carrying out a square-root transformation before fitting the model. This yields an estimated signal–noise ratio of $q = .132$ while squaring the forecasted values gives predictions of 7.432 and a much narrower prediction interval.

Of course, the purse snatchings are an example of count data, but since the numbers are not too small, fitting various Gaussian models is a useful preliminary exercise. (For example, extending the model to include a stochastic slope indicates that such a component is unnecessary.) In particular it should be noted that the square-root transformation is the variance-stabilizing transformation for a Poisson distribution (see McCullagh and Nelder 1983, pp. 129–130).

When the data are treated explicitly as count data, a negative binomial-beta model with $v = 20$ seems to give the best fit. The estimate of ω is $\tilde{\omega} = .697$. The predicted level is 7.573, corresponding to predictions from the Gaussian model with $q = .131$. A plot of the residuals shows no evidence of heteroscedasticity, and the sample variance of the standardized residuals is 1.008.

7.3 Effect of the Seat-Belt Law on Van Drivers in Great Britain

The effect of the seat-belt law of January 1983 on various classes of road users in Great Britain was analyzed by Harvey and Durbin (1986). For certain categories, the numbers involved were relatively small, with the result that a Gaussian model could not be regarded as a reasonable approximation. One such series is the monthly totals of drivers of light-goods vehicles killed. Here the numbers, from January 1969 to December 1984, range from 2 to 17.

Since the series contains no zero observations, a Gaussian model can be fitted to the logarithms of the observations. This gives preliminary estimates for the seasonal and intervention effects that can be used as starting values in the iterative procedure used to calculate the ML estimators in a count-data model. It is clear from doing this, however, that a Gaussian model is not at all satisfactory in these circumstances, and the results are very different for different specifications. In particular, fitting a model with fixed seasonals and no slope gives an estimate of the intervention effect that implies a 45% fall in fatalities as a result of the seat-belt law. This is quite out of line with estimates obtained for other series and indeed with the results obtained when a slope is included.

For the Poisson model, it is reassuring to note that the conclusions regarding the effect of the seat-belt law are affected very little by the inclusion or otherwise of a slope term. In fact, the preferred specification does not have a slope. The explanatory variables are, therefore, an intervention and seasonals, and fitting the model gives the estimates of ω and the intervention effect as $\tilde{\omega} = .934$ and $\tilde{\lambda} = -.276$. The estimate of λ implies a 24.1% reduction in fatalities, which is quite close to the figures reported earlier for car drivers by Harvey and Durbin (1986). The likelihood-ratio test statistic for the inclusion of the intervention variable is 25.96, and this is clearly significant when set against a χ^2_1 distribution. Finally, the estimated seasonal factors, given by exponentiating the estimated seasonal coefficients, are very reasonable and not dissimilar to the seasonal factors reported by Harvey and Durbin (1986) for car drivers killed and seriously injured. See Table 2.

ACKNOWLEDGMENTS

An earlier version of this article was presented at the European Meeting of the Econometric Society, Bologna, September 1988, at the Economic and Social Research Council (ESRC) Econometric Study Group Meeting in Bristol in June 1988, and at the National Bureau of Economic Research Time Series Conference in Chicago in October 1988. We thank Jim Durbin and Richard L. Smith for helpful comments. The research was supported by the ESRC Grant B00-23-2155 (Harvey) and the Conselho Nacional de Desenvolvimento Scientifico e Technologico (Fernandes).

Table 2. Estimated Seasonal Factors for LGV Drivers Killed

Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1.16	.79	.94	.89	.91	1.06	.97	.92	.92	1.16	1.19	1.19

APPENDIX: THE POST-SAMPLE PREDICTIVE TEST FOR THE POISSON MODEL

The post-sample predictive test statistic for the Poisson observation model is obtained by introducing l dummy variables into the model at times $T + 1$ to $T + l$. The statistic $\xi(l)$ is obtained by subtracting the likelihood obtained without these variables from the likelihood with these variables and multiplying by a factor of two.

To find the likelihood function for the model with dummy variables in the post-sample period, first consider the case of $l = 1$. The likelihood function is of the form (2.10) with T replaced by $T + 1$. The dummy variable parameter δ, however, only enters the likelihood via $b^{\dagger}_{t|t-1}$, which from (6.4) is

$$b^{\dagger}_{T+1|T} = \omega b_T e^{-\delta}. \tag{A.1}$$

Thus the log-likelihood can be written as

$$\log L_{T+1} = \log L^{*}_{T+1} + a_{T+1|T}\log b^{\dagger}_{T+1|T} - (a_{T+1|T} + y_{T+1})\log(1 + b^{\dagger}_{T+1|T}), \tag{A.2}$$

where L^{*}_{T+1} does not depend on δ. Differentiating (A.2) with respect to δ yields

$$\exp(-\tilde{\delta}) = a_{T+1|T}/y_{T+1}b^{\dagger}_{T+1|T} = a_T/y_{T+1}b_T \tag{A.3}$$

and so, from (A.1),

$$\tilde{b}_{T+1|T} = a_{T+1|T}/y_{T+1}. \tag{A.4}$$

Substituting into (A.2) gives the likelihood concentrated with respect to δ. Note that, in the special case when $y_{T+1} = 0$, the last two terms on the right side of (A.2) are 0 when taken together, so $\log L_{T+1} = \log L^{*}_{T+1}$.

Now consider $l > 1$. The log-likelihood function with l dummy variables, $\delta_1, \ldots, \delta_l$, in the post-sample period is

$$\log L_{T+l} = \log L^{*}_{T+l} + \sum_{t=T+1}^{T+l} a_{t|t-1}\log b^{\dagger}_{t|t-1} - \sum_{t=T+1}^{T+l} (a_{t|t-1} + y_t)\log(1 + b^{\dagger}_{t|t-1}), \tag{A.5}$$

where $b^{\dagger}_{t|t-1}$ obeys the recursion (6.4) and (6.6). This implies that $b^{\dagger}_{T+j|T+j-1}$ depends on $\delta_1, \ldots, \delta_{j-1}$ for $j = 2, \ldots, l$, thereby making differentiation of $\log L_{T+l}$ with respect to $\delta_1, \ldots, \delta_l$ rather tedious. If we differentiate with respect to δ_l first, however, we obtain a result analogous to (A.4), namely,

$$\tilde{b}_{T+l|T+l-1} = a_{T+l|T+l-1}/y_{T+l}. \tag{A.6}$$

This is independent of previous values of $b^{\dagger}_{t|t-1}$ and hence of $\delta_1, \ldots, \delta_{l-1}$.

Concentrating the likelihood with respect to δ_l and proceeding to treat $\delta_1, \ldots, \delta_{l-1}$ in the same way gives the following concentrated log-likelihood function:

$$\log L_{T+l} = \log L^{*}_{T+l} + \sum_{t=T+1}^{T+l} a_{t|t-1}\log(a_{t|t-1}/y_t) - \sum_{t=T+1}^{T+l} (a_{t|t-1} + y_t)\log(1 + a_{t|t-1}/y_t). \tag{A.7}$$

The likelihood function under the null hypothesis—that is without dummy variables—is

$$\log L_{T+l} = \log L^{*}_{T+l|T} + \sum_{t=T+1}^{T+l} a_{t|t-1}\log b_{t|t-1} - \sum_{t=T+1}^{T+l} (a_{t|t-1} + y_t)\log(1 + b_{t|t-1}), \tag{A.8}$$

where $b_{t|t-1}$ is computed via the recursion in (2.4) and (2.6). Subtracting (A.8) from (A.7) and multiplying by 2 gives the likelihood ratio test statistic (7.2).

[*Received February 1989.*]

REFERENCES

Cameron, C. C., and Trivedi, P. K. (1986), "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests," *Journal of Applied Econometrics*, 1, 29–53.

Chow, G. C. (1960), "Tests of Equality Between Sets of Coefficients on Two Linear Regressions," *Econometrica*, 28, 591–605.

Harvey, A. C., and Collier, P. (1977), "Testing for Functional Misspecification in Regression Analysis," *Journal of Econometrics*, 6, 103–119.

Harvey, A. C., and Durbin, J. (1986), "The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling," *Journal of the Royal Statistical Society*, Ser. A, 149, 187–227.

Harvey, A. C., and Todd, P. H. J. (1983), "Forecasting Economic Time Series With Structural and Box–Jenkins Models" (with discussion), *Journal of Business and Economic Statistics*, 1, 229–315.

Hausman, J. A., Hall, B. H., and Griliches, Z. (1984), "Econometric Models for Count Data With an Application to the Patents–R & D Relationship," *Econometrica*, 52, 909–938.

Jeffreys, H. (1961), *The Theory of Probability*, Oxford, U.K.: Oxford University Press.

Key, P., and Godolphin, E. J. (1981), "On the Bayesian Steady Forecasting Model," *Journal of the Royal Statistical Society*, Ser. B, 43, 92–96.

Kitagawa, G., and Gersch, W. (1984), "A Smoothness Priors–State Space Modeling of Time Series With Trend and Seasonality," *Journal of the American Statistical Association*, 79, 378–389.

Kohn, R., and Ansley, C. F. (1985), "Efficient Estimation and Prediction in Time Series Regression Models," *Biometrika*, 72, 694–697.

Lehmann, E. L. (1983), *Theory of Point Estimation,* New York: John Wiley.

McCleary, R., and Hay, R. A., Jr. (1980), *Applied Time Series Analysis for the Social Sciences,* Beverly Hills, CA: Sage Publications.

McCullagh, P., and Nelder, J. A. (1983), *Generalised Linear Models,* London: Chapman & Hall.

Muth, J. F. (1960), "Optimal Properties of Exponentially Weighted Forecasts," *Journal of the American Statistical Association,* 55, 299–305.

Smith, J. Q. (1979), "A Generalization of the Bayesian Steady Forecasting Model," *Journal of the Royal Statistical Society,* Ser. B, 41, 375–387.

——— (1981), "The Multiparameter Steady Model," *Journal of the Royal Statistical Society,* Ser. B, 43, 256–260.

Smith, R. L., and Miller, J. E. (1986), "A Non-Gaussian State Space Model and Application to Prediction of Records," *Journal of the Royal Statistical Society,* Ser. B, 48, 79–88.

West, M., and Harrison, P. J. (1986), "Monitoring and Adaptation in Bayesian Forecasting Models," *Journal of the American Statistical Association,* 81, 741–750.

West, M., Harrison, P. J., and Migon, H. S. (1985), "Dynamic Generalized Linear Models and Bayesian Forecasting" (with discussion), *Journal of the American Statistical Association,* 80, 73–97.