ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

A COMPARISON OF METHODS FOR FORECASTING DEMAND FOR SLOW MOVING CAR PARTS

RALPH D. SNYDER AND ADRIAN BEAUMONT

December 2007

Working Paper 15/07

A COMPARISON OF METHODS FOR FORECASTING DEMAND FOR SLOW MOVING CAR PARTS

RALPH D. SNYDER AND ADRIAN BEAUMONT

Abstract. This paper has a focus on *non-stationary* time series formed from small non-negative integer values which may contain many zeros and may be over-dispersed. It describes a study undertaken to compare various suitable adaptations of the simple exponential smoothing method of forecasting on a database of demand series for slow moving car parts. The methods considered include simple exponential smoothing with Poisson measurements, a finite sample version of simple exponential smoothing with negative binomial measurements, and the Croston method of forecasting. In the case of the Croston method, a maximum likelihood approach to estimating key quantities, such as the smoothing parameter, is proposed for the first time. The results from the study indicate that the Croston method does not forecast, on average, as well as the other two methods. It is also confirmed that a common fixed smoothing constant across all the car parts works better than maximum likelihood approaches.

1. Introduction

Simple exponential smoothing, in its original form (Brown, 1963), was designed for forecasting demand of fast moving inventories. Given that businesses often also have slow moving inventories, the traditional recommendation was to forgo the use of exponential smoothing in such cases and use a Poisson distribution with a mean that is estimated with a simple sample average. The Poisson distribution, however, only really applies to the case of equi-dispersed demand series where the variance and mean are equal. Demand series for slow moving items are typically over-dispersed in the sense that the variance is greater than the mean.

Over-dispersion can be achieved with a Poisson distribution by allowing its mean, designated by λ, to be a random variable. The Poisson distribution then represents the distribution of demand *conditional* on a particular value of λ. When λ has a gamma distribution, the negative binomial distribution (Greenwood and Yule, 1920) is obtained as the corresponding marginal (unconditional) distribution of demand. This suggests that the Poisson distribution should be supplanted by the negative-binomial distribution for representing the demand of slow moving inventories (Taylor, 1961). Alternatively, it means that the Poisson distribution can be retained in models of demand provided it is treated as a conditional distribution supplemented with a mixing distribution like the gamma distribution.

The negative binomial approach, and indeed the earlier Poisson approach, were based on the assumption that means and variances are constant over time. The

Date: November, 2007.
Key words and phrases. count time series, forecasting, exponential smoothing, Poisson distribution, negative binomial distribution, Croston method.
Research for this paper has been funded by an Australian Research Council Discovery grant.

implied stationarity of the demand series, however, is inconsistent with the fact that most such series are non-stationary. Moreover, these approaches rely on an assumption of inter-temporal independence, something that is at odds with the autocorrelation that is commonly present in demand series. Adaptations are needed if the non-stationarity and autocorrelation are to be accommodated.

Autocorrelation and over-dispersion have been introduced with Poisson measurements by allowing the mean λ to change randomly over time according to some sort of autoregressive process (Chan and Ledolter, 1995; Heinen, 2003; Jung, Kuluk and Leisenfield, 2006; Snyder, Martin, Gould and Feigen, 2007). An alternative has been to allow the mean to follow a kind of moving-average process (Davis, Dunsmuir and Streett, 2003). The stationarity assumption that accompanies these approaches is at odds with the non-stationary character of most demand series. In this paper, therefore, approaches which allow for autocorrelation and over-dispersion and which are based on unit root processes, are considered. The first approach is an adaptation of simple exponential smoothing (Chatfield, Koehler, Ord and Snyder, 2001; Ord, Koehler and Snyder, 1997) where the traditional Gaussian distribution is replaced by a Poisson distribution. The second is a discounted least squares approach (Brown, 1963; Gilchrist, 1967) with negative binomial measurements (Harvey and Fernandes, 1989). The third is the traditional Croston (1972) method for forecasting slow moving inventories where, for the first time, a logically sound model underlying this method is proposed. These three approaches are introduced in Section 2, and their forecasting performance is compared in an empirical study on car parts data in Section 3.

2. Models for Non-Stationary Count Time Series

2.1. **Local Poisson Model.** The Gaussian innovations local level model (Chatfield, 2001; Ord et. al. 1997) underpins simple exponential smoothing. In this section we introduce a variation of this model that is better suited to low count data. The usual Gaussian distribution of demand y_t in typical period t is replaced by the Poisson mass function

$$\mathrm{p}(y_t|\lambda_{t-1}) = \frac{\lambda_{t-1}^{y_t}}{y_t!}\exp(-\lambda_{t-1}). \tag{2.1}$$

The mean λ_{t-1} of this distribution, called the local level, changes over time according to the smoothing equation

$$\lambda_t = \delta\lambda_{t-1} + \alpha y_t. \tag{2.2}$$

The parameters δ and α are constrained to be non-negative and to satisfy the constraint $\delta + \alpha = 1$. The only source of randomness is the Poisson distribution itself: the model has only one source of randomness. Nevertheless, the associated time series is over-dispersed because the local level λ_t is now a random rather than a fixed quantity.

Simple exponential smoothing is used to calculate the local levels, so λ_t corresponds to the exponentially weighted moving average

$$\lambda_t = \delta^t\lambda_0 + \alpha\sum_{j=0}^{t}\delta^j y_{t-j}. \tag{2.3}$$

Since λ_{t-1} is governed by the lagged version of 2.3, it must be a fixed quantity when $y_1, y_2, \ldots, y_{t-1}, \lambda_0, \alpha$ are fixed. This implies that $y_t|y_1, y_2, \ldots, y_{t-1}, \lambda_0, \alpha$ is governed by the Poisson distribution 2.1. The likelihood function is based on the joint distribution $\mathrm{p}(y_1, y_2, \ldots, y_n|\lambda_0, \alpha)$. It is found from the products of the one-step ahead prediction distributions $p(y_t|y_1, y_2, \ldots, y_{t-1}, \lambda_0, \alpha)$ for $t = 1, 2, \ldots, n$ and is given by

$$L(\lambda_0, \alpha|y_1, y_2, \ldots, y_n) = \prod_{t=1}^{n} \frac{\lambda_{t-1}^{y_t}}{y_t!} \exp(-\lambda_{t-1}) \tag{2.4}$$

The seed level λ_0 and the smoothing parameter α are selected to maximize this Poisson likelihood. Point predictions are obtained by extrapolating the final level λ_n. Prediction distributions must be simulated because analytical expressions for them are not currently known.

2.2. **Local Negative Binomial Distribution.** The Harvey and Fernandes(1989) approach, when reduced to its bare essentials, uses a discounted moving average (Brown, 1963; Gilchrist, 1967) instead of an exponentially weighted average and a negative binomial distribution instead of a Poisson distribution. A discounted average is a measure of central tendency that minimizes the discounted sum of squared errors. It derives from two quantities a_t and b_t which are calculated with the recurrence relationships

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

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

seeded with $a_0 = 0$ and $b_0 = 0$. The parameter δ is included to provide a discounting effect. The one-step ahead prediction is given by

$$\hat{y}_{t|t-1} = a_{t-1}/b_{t-1}. \tag{2.7}$$

It may be established that $\hat{y}_{t|t-1} = \frac{\sum_{j=1}^{t-1} \delta^{j-1} y_{t-j}}{\sum_{j=1}^{t-1} \delta^{j-1}}$. This indicates that the one-step ahead prediction is an average. Moreover, older observations are given less weight in its calculation than more recent observations. It converges in large samples to an exponentially weighted average, so the approach may be considered a slight variation of simple exponential smoothing.

A one-step ahead prediction distribution always has a mean corresponding to the one-step ahead prediction. In the case of the negative binomial distribution

$$\mathrm{p}(y_t|y_1, \ldots, y_{t-1}, \delta) = \frac{\Gamma(\delta a_{t-1} + y_t)}{\Gamma(\delta a_{t-1}) y_t!} \left(\frac{\delta b_{t-1}}{1 + \delta b_{t-1}}\right)^{\delta a_{t-1}} \left(\frac{1}{1 + \delta b_{t-1}}\right)^{y_t}. \tag{2.8}$$

the mean is given by 2.7. The likelihood function is formed from the product of these one-step ahead prediction distributions. However, there is one qualification. The initial values of a_t and b_t are zero until the period following that period in which the first non-zero value of y_t is observed. The one-step ahead prediction distributions are all degenerate, in the sense that they equal zero, over this run-in period. So the product is formed from the prediction distributions for the periods following the period containing the first non-zero observation.

2.3. **Croston Method.** A popular method for forecasting the demand of slow moving time series was developed by Croston (1972). It accounts for the time gap between periods with positive demands as well as the positive demands themselves. The time gap for period t is designated by τ_t. If there is a non-zero demand in period t, then τ_t is the number of periods that have elapsed since the period with the previous non-zero demand; otherwise it is arbitrary.

Simple exponential smoothing is applied to both the positive demands and the time gaps. More specifically, there are two exponentially weighted averages designated by $\bar{q}_t$ for the non-zero demands and $\bar{\tau}_t$ for the time gaps in typical period t. Both averages effectively rely on the same smoothing parameter α. There is, however, a slight twist. We use a time dependent smoothing parameter α_t which equals α in those periods with a non-zero demand and equals zero otherwise. A corresponding time dependent discount factor $\delta_t = 1 - \alpha_t$ applies.

It is assumed that seed values for the averages, designated by $\bar{q}_0$ and $\bar{\tau}_0$, have been specified. The method, as applied in typical period t, can be stated as:

$$\begin{aligned} \bar{q}_t &= \delta_t \bar{q}_{t-1} + \alpha_t y_t \\ \bar{\tau}_t &= \delta_t \bar{\tau}_{t-1} + \alpha_t \tau_t \end{aligned}$$

An exception occurs in the period with the first non-zero demand, because here the first time gap is still unknown. In this particular period, the second equation is replaced by the simple equation $\bar{\tau}_t = \bar{\tau}_{t-1}$. The point predictions of future demands all equal the ratio $\bar{q}_n/\bar{\tau}_n$.

On the basis of experience, Croston recommends that the smoothing parameter α should take a value between 0.1 and 0.2. Little is said about the choice of seed values for the exponentially weighted averages. In an attempt to eliminate the ambiguity surrounding this choice of seed values and the smoothing parameter, we now set out to explore the stochastic foundations of his approach in a quest for maximum likelihood estimates.

Croston attempted to identify the stochastic foundations, but it turned out to be flawed (Snyder, 2002). He suggested in his Appendix B, that the probability π of a positive demand in a period is constant over time. The consequent stationarity of the time gaps is, however, inconsistent with the use of simple exponential exponential smoothing.

In order to identify the correct statistical foundations, we assert that the empirical distributions of the positive demands are typically right skewed. To allow for this possibility, we assume that the positive quantities are governed locally by a Poisson distribution, but as they cannot take the value 0, its domain is shifted by one to the right. The mass function of this distribution in typical period t is represented by $p_{tj} = \frac{(\bar{q}_{t-1}-1)^{j-1}}{(j-1)!} \exp(-\bar{q}_{t-1}+1)$ where j is the value of the positive demand.

Local Bernoulli distributions govern whether there are positive or zero demands. The Bernoulli probability in typical period t is designated by π_t: it represents the probability of a positive demand. It is related to the exponentially weighted average of the time gaps by the simple formula $\pi_t = 1/\bar{\tau}_{t-1}$.

The probability of demand in period t is given by

$$\Pr\{y_t = j\} = \begin{cases} (1-\pi_t) & \text{if} j = 0 \\ \pi_t p_{tj} & \text{if} j > 0 \end{cases} \tag{2.9}$$

The likelihood function is the product of these mass functions for periods $t_1, t_1 + 1, \ldots, n$. Maximum likelihood estimates of the seed averages $\bar{q}_{t_1-1}$, $\bar{\tau}_{t_1-1}$ and the

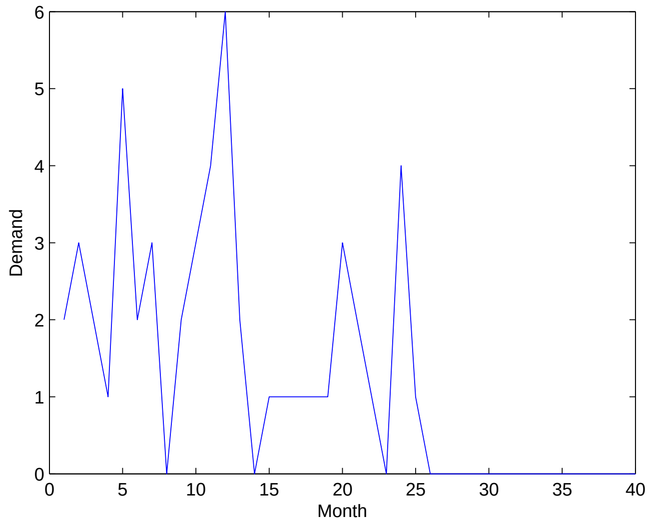

FIGURE 1. Time series generated from Poisson local level model: $\ell_0 = 2$; $\alpha = 0.5$.

smoothing parameter α are sought. For numerical stability this is done by maximizing the log-likelihood.

2.4. **Convergence problem.** The local level model in Section 2.1 has a fixed point of $\lambda_t = 0$. This fixed point is an attractor: there is a finite probability that λ_t drops to zero. The problem with this particular fixed point is that all subsequent series values are forced to be zero. Figure 1 illustrates this phenomena with some data generated by a local Poisson model from Section 2.1 where $\lambda_0 = 2$ and $\alpha = 0.5$. One must be aware of this problem when using the model to simulate prediction distributions. A detailed explanation of this phenomenon is provided by Grunwald, Hamza and Hyndman (1997).

The other methods also suffer from this problem. In the Croston method, it takes a slightly different form: the simulated gaps and positive quantities both eventually converge to one rather than zero. The important message here is that care must be taken when simulating prediction distributions. Longer-run prediction distributions may be problematic.

3. Empirical Study

Forecasts from the three methods were compared on 2674 demand series for parts supplied by a US auto manufacturer. The series, representing the monthly sales for slow moving parts, cover a period of 51 months from January 1998 to March 2002. The 2509 series without missing values have an average gap between positive demands of 2.9 months and an average positive demand of 2. Eighty-nine percent of the series were over-dispersed. The dispersion ratio, averaged across all series, was 2.3.

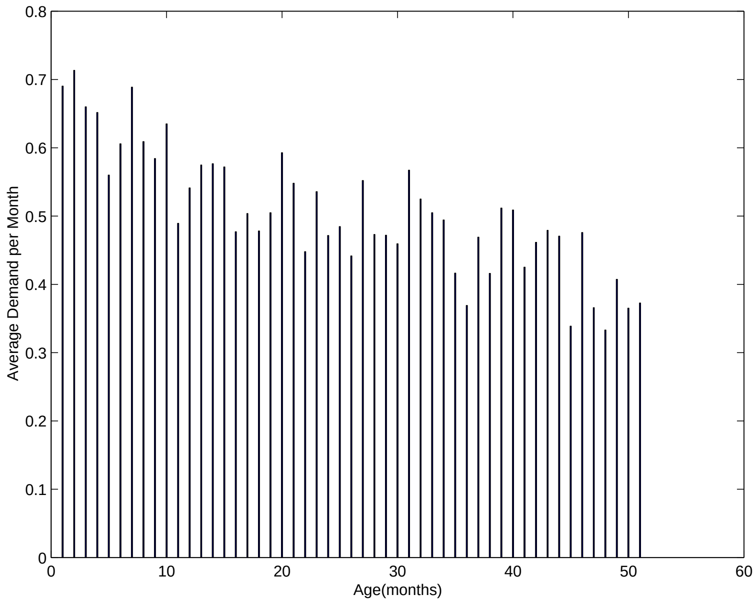

FIGURE 2. Time profile of demands averaged across series: CARPARTS database.

The time profile of aggregate demand for all the car parts is shown in Figure 2. It indicates that there is a tendency for demands to decline as the age of a part increases. Demands appear to be non-stationary.

Although a downward trend is discernable in the aggregate data, such trends will clearly not always operate at the individual parts level. It is important to allow for other possible trajectory shapes which may only be observed at the individual part level. One possibility, ignoring the zero demands, is a gradual increase to some peak and then a slow decline. Because such patterns are not known in advance, there is a need for an approach that adapts to whatever pattern that emerges as a part ages. Given the uncertainty over the trajectory, it is best to treat the underlying level as a random variable and assume that its evolution over time is governed by a stochastic process.

To minimize computational problems that arise with series with a small number of positive values, the database was further culled to eliminate those series which:

- possessed less than 10 positive monthly demands;
- had no positive demand in the first 15 and final 15 months.

There were 1046 series left after this additional cull.

Six approaches to forecasting were compared in the study. The simplest, designated by ZERO, was to set all the predictions to the value zero, on the grounds that the empirical distributions of many series have a mode of zero. The second was based on a *global* Poisson distribution (PSNG) where it is optimal to use a simple average of observed demands. This assumes that there is no structural change in the market for a product. The others were the methods described in the previous section that allow for random changes in the underlying level: the *local*

Poisson distribution (PSNL); the *local* negative binomial distribution (NEGB); and the Croston method (CROST).

The maximum likelihood (ML) versions of these approaches were considered. The folk law of exponential smoothing (Brown, 1959); Croston, 1972) suggests that especially with small samples, it is best to use fixed values of α or, equivalently, fixed values of the discount factor δ, across an entire range of products, so this possibility was also considered. In those cases of a fixed parameter which rely on seed levels, the latter continued to be estimated by maximizing the likelihood function. In other words, a *partial* maximum likelihood approach was used in these cases.

Estimation was undertaken with the first 45 observations of each series. Forecasting performances were compared over periods 46-51 using the mean absolute scalar error statistic (Hyndman and Koehler, 2006), defined as

$$\text{MASE} = \frac{1}{h}\sum_{j=1}^{h}\frac{|y_{n+j} - \hat{y}_{n+j|n}|}{\text{MAE}} \tag{3.1}$$

where

$$\text{MAE} = \frac{1}{n-1}\sum_{t=2}^{n}|y_t - y_{t-1}|. \tag{3.2}$$

where n is the fitting sample size, h is the forecast horizon, and $\hat{y}_{n+j|n}$ is the prediction of y_{n+j} made at the end of period n. The means, medians and standard deviations of the MASEs calculated across the 1046 series are given in Table 1. The approaches are ordered by the median.

Model	**Parameter**	**Mean**	**Median**	**Stdev**
ZERO		0.42	0.30	0.47
PSNL	0.300	0.63	0.55	0.42
PSNL	0.200	0.64	0.56	0.38
NEGB	0.800	0.64	0.56	0.38
NEGB	ML	0.65	0.59	0.40
CROST	0.300	0.65	0.60	0.40
CROST	0.200	0.65	0.61	0.39
PSNL	ML	0.68	0.64	0.36
CROST	0.100	0.68	0.65	0.38
CROST	ML	0.68	0.67	0.40
CROST	0.000	0.70	0.70	0.39
PSNG		0.82	0.75	0.31

Table 1. MASE for each method applied to car parts time series.

The zero method had the best performance in terms of the median MASE. However, a very high standard deviation of the MASE indicates that its performance lacked consistency. Moreover, such a method is unlikely to work as well on time series with fewer zeros.

Intriguingly, the traditional *global* Poisson distribution had the worst performance. The associated simple average, which places an equal weight on all the observations (including the zeros), has little value for the type of count data represented by car parts demands. It does not allow the mean to change over time; nor does it allow for autocorrelation.

Of the maximum likelihood methods (ML), the local negative binomial distribution was best. This was followed by the local Poisson distribution. Surprisingly, the Croston method had the worst performance.

The folk law about the use of fixed parameter values was also confirmed. Global fixed value approaches did better than their maximum likelihood counterparts. The results of the local Poisson and the local negative binomial distributions were reversed, but the Croston method continued to have the worst performance.

A deeper analysis of the results indicates that the local Poisson model may or may not be better than the local negative binomial model at the individual series level. The NEGB 0.8 method was better than PSNL 0.2 method for 56 percent of the series and they tied for 28 percent of the series. However, the maximum difference in the MASE between these two models was only 0.15, so both approaches are very similar with regard to point predictions.

In a further attempt to separate the PSNL and NEGB methods, 90 percent prediction intervals were simulated for periods 46-51. In both cases about 95 percent of withheld series values were found to lie within these prediction intervals. This result was obtained whether or not the smoothing parameter (discount factor) was optimized. In these circumstances, the NEGB approach appears to have little advantage over PSNL approach on this particular data set. The prediction intervals are a little too wide, a result of having integer counts. Curiously, it was found that 91 percent of all the future observations across all time series lie in the closed interval [0,1].

4. Conclusions

Three approaches to predicting demands for slow moving inventories have been compared in this paper. In essence, they consisted of various adaptations of simple exponential smoothing to accommodate count data with many zero observations. Emphasis was placed on providing stochastic versions of these methods to enable estimation of pertinent parameters using maximum likelihood methods and to enable the simulation of prediction distributions.

The methods were applied to a database of demand series for parts from an auto manufacturer. It was found that predictions based on simple averages did not work well and that considerable benefits accompanied a move to methods based in some way on exponential smoothing. Little separated the exponential smoothing methods from Poisson and negative binomial measurements. However, the study did raise serious questions about the practical value of the Croston method. It also indicated that there may be little reward from the use of maximum likelihood methods, something that confirms the folk-law of practitioners of business forecasting.

In general it was observed in the study that reductions in the median MASE corresponded to increases in its standard deviation. This suggests that a multi-model approach might work better than any single model approach by selecting the model that best fits the individual structures of the time series. However, it is not clear how this can be done. Sample quantities, in practice, are often too small to withhold data for a prediction validation approach. And the models are based on different probability distributions, something that precludes the use of an information criterion approach. We have here an issue that warrants further investigation.

References

[1] Brown, R.G.(1959) *Statistical Forecasting for Inventory Control*, McGraw-Hill

[2] Brown, R.G.(1963) *Smoothing, Forecasting and Prediction of Discrete Time Series*, Prentice Hall.

[3] Chan, K. S. and Ledolter, J. (1995) Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, 90, 242-252.

[4] Chatfield, C, A. B. Koehler, Ord, J.K., Snyder, R.D. (2001) A new look at models for exponential smoothing, *The Statistician*, 50, 147-159.

[5] Croston, J. D. (1972) Forecasting and stock control for intermittent demands, *Operational Research Quarterly*, 23, 289-303.

[6] Davis, R.A. Dunsmuir, W.T.M. and Streett, S.B. (2003) Observation-driven models for poisson counts, *Biometrika*, 90, 777-790.

[7] Gilchrist, W. G. (1967) Methods of estimation involving discounting, *Journal of the Royal Statistical Society: Series B*, 355-369.

[8] Greenwood, M. and Yule, G.U. (1920) An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents, *Journal of the Royal Statistical Society*, 83, 255-279.

[9] Grunwald, G.K., Hamza, K. and Hyndman, R.J. (1997) Some properties and generalizations of non-negative Bayesian time series models. *Journal of the Royal Statistical Society: Series B*, 59, 615-626.

[10] Harvey, A. C. and C. Fernandes (1989) Time series models for count and qualitative observations, *Journal of Business and Economic Statistics*, 407-417.

[11] Heinen, A. (2003) Modeling Time Series Count Data: an Autoregressive Conditional Poisson Model, *Draft Paper, University of California, San Diego.*

[12] Hyndman, R.J. and A.B. Koehler (2006) Another look at measures of forecast accuracy International Journal of Forecasting, 22, 679-688

[13] Jung, R.C., Kukuk, M. and Leisenfeld, R. (2006) Time Series of Count Data: Modeling, Estimation and Diagnostics, *Computational Statistics and Data Analysis*, 51, 2350-2364.

[14] Ord, Koehler and Snyder, (1997) Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models, Journal of The American Statistical Association, 92, 1621-1629.

[15] Snyder, R.D. (2002) Forecasting sales of slow and fast moving inventories, *European Journal of Operational Research*, 140, 684-699.

[16] Snyder, R.D. Martin, G., Gould, P., Feigen, P. (2007), An assessment of alternative state space models for count time series, working paper,*Department of Econometrics and Business Statistics*

[17] Taylor, C.J. (1961) The application of the negative binomial distribution to stock control problems, *Operational Research Quarterly*, 12, 81-88.

Department of Econometrics and Business Statistics, P.O. Box 11E, Monash University, Victoria 3800, Australia
E-mail address: ralph.snyder@buseco.monash.edu.au

Department of Econometrics and Business Statistics, P.O. Box 11E, Monash University, Victoria 3800, Australia
E-mail address: adrian.beaumont@buseco.monash.edu.au