Journal of Forecasting Vol. 2, 331–350 (1983)

The Filtering of Transitory Noise in Earnings Numbers

Z. LIEBER
Tel Aviv University, Tel Aviv, Israel

E. I. MELNICK
J. RONEN
New York University, New York, U.S.A

ABSTRACT

This paper applies the Kalman filtering procedure to estimate persistent and transitory noise components of accounting earnings. Designating the transitory noise component separately (under a label such as extraordinary items) in financial reports should help users predict future earnings. If a firm has no foreknowledge of future earnings, managers can apply a filter to a firm's accounting earnings more efficiently than an interested user. If management has foreknowledge of earnings, application of a filtering algorithm can result in smoothed variables that convey information otherwise not available to users. Application of a filtering algorithm to a sample of firms revealed that a substantial number of firms exhibited a significant transitory noise component of earnings. Also, for those firms whose earnings exhibited a significant departure from the random walk process, the paper shows that filtering can be fruitfully applied to improve predictive ability.

KEY WORDS Kalman filter Annual earnings Predictive ability Persistent and transitory components

The subject of reporting extraordinary items has drawn considerable attention (Ronen *et al.*, 1977, provide a partial representative list of references). Barnea *et al.* (1976) proposed that managers use their judgement to determine which income items are persistent and which are transitory, and classify the first as ordinary and the second as extraordinary. The claim that such a classification is useful to investors is based on the following propositions:

1. The persistent component of income contributes more than the transitory component to the value (and changes therein) of securities. This proposition is consistent with extant security valuation models (such as that of Miller and Modigliani, 1961; see also Ball and Watts, 1972). Some empirical support for this hypothesis is provided by Ronen and Sadan (1981, pp. 111–113).
2. Prediction of future earnings is facilitated by presenting the two components separately. This proposition is also supported by empirical evidence given by Ronen and Sadan (1981, pp. 106–110).

0277-6693/83/040331-20$02.00

3. The security valuation mechanism uses (among other variables) expectations of future earnings. This proposition is assumed by all extant valuation models, and is implicit in the Financial Accounting Standards Board's (FASB's) assertion that prediction of future earnings is an important intermediate step in assessing future cash flows (FASB, 1978).

These propositions are interrelated. Future earnings predictions, which figure in the determination of security prices, are based on the persistent component of income; hence, the importance of the persistent component in valuing securities. By contrast, the transitory component is not useful in forming expectations of earnings, so its role in valuation is likely to be minor.

Opinion 30 of the Accounting Principles Board (AICPA, 1973), currently in effect, prescribes that items be classified as extraordinary only if they are both unusual and non-recurring. These criteria disqualify all but extremely rare events (such as earthquakes) from being labelled extraordinary. Two questions related to the criteria and to the reporting of extraordinary items in general arise. First, will the separate reporting of the transitory component as an extraordinary item facilitate improved forecasting of earnings? Secondly, what statistical method should be used to estimate the unobservable persistent and transitory components of income?

On the first question, the answer depends upon the time-series process underlying annual accounting earnings. Ball and Watts (1972), Albrecht *et al.* (1977) and Watts and Leftwich (1977) seem to suggest that annual accounting earnings behave as a random walk. One implication is that earnings classification is meaningless since no transitory component exists, and thus no improvement in predictive ability is expected. However, Beaver *et al.* (1980) presented evidence consistent with earnings being perceived as a compound process consisting of both a persistent and transitory component, when viewed from the perspective of other information reflected in security prices. Attempting to reconcile these contradictory pieces of evidence, Beaver *et al.* suggested that the near zero serial correlation in earnings differences reported in past research resulted from opposite-sign serial correlations associated with the persistent and transitory components.

This paper assumes that earnings have persistent and transitory components. It is argued that, for a specific class of earnings processes studied by Barnea *et al.* (1976), the application of the Kalman filter procedure (Kalman, 1960) improves the estimation of the persistent and transitory components.[1] Although both the Box–Jenkins technique and the Kalman filtering technique produce the same estimation results for the assumed earnings process when the sample is very large, the filtering procedure is optimal in the case of a finite sample, and thus advantageous.[2]

Beaver *et al.* (1980) describe an interesting alternative approach to estimating transitory and persistent earnings components, and predicting future earnings. They assume an earnings process that differs from the one used here only in that theirs allows for the possibility that the two earnings components might be correlated. Using a simple market valuation model, they infer the parameters of the hypothesized process contingent on their assumptions, and offer preliminary evidence that price-based forecasts of earnings perform better than those on a model of random walk with a drift. Underlying their approach is the premise that security prices act as if the transitory component and the persistent component are separately known to market participants.

The present study complements the Beaver *et al.* approach by addressing *how* information on the two components is likely to be acquired by the market. Specifically, it is assumed that management possesses greater knowledge than others regarding the transitory and persistent components. The paper then investigates whether the Kalman filtering procedure offers a means

---

[1] Appendix 1 describes Kalman's state-space representation theory, and Appendix 2 presents the Kalman filter algorithm.

[2] Muth (1960) showed that exponential smoothing produces asymptotically the same forecast as the filter. The latter is used in this study because of the small sample sizes.

by which the persistent component can be extracted efficiently from past earnings series by market participants (a descriptive question). Further, in the normative domain, examination of the potential of the Kalman procedure provides a means by which management can convey to the market varying degrees of knowledge regarding the persistent component. That is, management's superior information can be used in the Kalman procedure so that 'good' estimates of the persistent components can be disclosed in the income statement.

The following section describes the assumed state space representation for earnings, and its implications for separating the transitory from the persistent component of income. The second section summarizes the Kalman filter estimation technique and defines parameters that measure improvement in prediction from using the filter. In the third section, the filtering procedure is tested on a sample of firms. Its predictive ability is compared with that of a random walk model within the parameter estimation period. Predictions derived from Box–Jenkins procedures are also compared. The fourth section presents evidence on predictions outside the parameter estimation period. Concluding remarks are provided in the last section.

1. THE EARNINGS PROCESS

The earnings process studied is the linear stochastic process suggested by Barnea *et al.* (1975). The following formulation states the process from the perspective of the beginning of a period t_0. The model is:

$$\tilde{E}_t = \tilde{X}_t + \tilde{u}_t \tag{1}$$

and

$$\tilde{X}_t = X_{t-1} + \tilde{w}_t, \qquad t = t_0,\ t_0+1, \ldots, T$$

where $\tilde{E}_t$ is the net earnings for period t, X_t is the persistent component of $\tilde{E}_t$ and $\tilde{u}_t$ is a transitory component of $\tilde{E}_t$, such that:

$$E(\tilde{u}_t) = 0, \qquad \mathrm{Var}(\tilde{u}_t) = \sigma_u^2$$

$$\mathrm{Cov}(\tilde{u}_t, \tilde{u}_{t+\tau}) = 0 \qquad \text{for all } \tau \neq 0$$

and $\tilde{w}_t$ is a noise component satisfying:

$$E(\tilde{w}_t) = \beta, \qquad \mathrm{Var}(\tilde{w}_t) = \sigma_w^2$$

$$\mathrm{Cov}(\tilde{w}_t, \tilde{w}_{t+\tau}) = 0 \qquad \text{for all } \tau \neq 0$$

where

$$\mathrm{Cov}(\tilde{w}_t, \tilde{u}_{t+\tau}) = 0 \qquad \text{for all } \tau$$

The ~ (tilde) denotes a random variable. In this model, only the E_t series is observable.

It follows from (1) that:

$$\Delta\tilde{E}_t = \tilde{E}_t - E_{t-1} = \tilde{w}_t + \tilde{u}_t - u_{t-1} \tag{2}$$

$$E(\Delta\tilde{E}_t) = \beta \tag{3}$$

and

$$E_t = \sum_{i=1}^{t} w_i + X_0 + u_t \tag{4}$$

where E_t, w_t and X_0 reflect realized magnitudes. In the estimation procedure and prediction ability tests described in Section 4, β is assumed to be zero for simplification.

This earnings process is a special case of the one assumed by Beaver *et al.* (1980). They assumed that the transitory component consisted of (1) a noise element representing garbling induced exclusively by an accounting system and (2) a 'real flow' element proportional to $\tilde{w}_t$. The present formulation also assumes that the transitory component consists of these two elements. However, it is assumed that the 'real flow' element is uncorrelated with $\tilde{w}_t$. The plausibility of this zero correlation assumption is difficult to assess empirically since the 'observable' extraordinary items presented in income statements do not necessarily equal the true transitory component. This is due to the restrictive criteria of APB 30 described earlier. The validity of the zero correlation assumption thus remains an open issue. Nonetheless, the analysis proceeds with this caveat.

Let $p = \sigma_w^2/\sigma_u^2$. If $p = 0$, the model in (1) reduces to a mean reverting process. On the other hand, the process becomes a random walk with drift as p approaches infinity. In general, the lower p, the more similar the process is to a mean reverting process. Conversely, the larger p is, the closer the process resembles a random walk with a drift. Muth (1960) showed that the above generating process can be asymptotically described by an exponential smoothing model. Since the latter can be characterized by its weighting factor, $\alpha(0 \leq \alpha \leq 1)$, (1) can be either characterized by the parameter p or by α, where $p = \alpha^2/(1-\alpha)$ (Fisher, 1971).

Denote the serial correlation coefficient of the first differences of earnings of (1) by R. Then, using (2):

$$R = \frac{\mathrm{Cov}(\Delta E_t, \Delta E_{t+1})}{\mathrm{Var}(\Delta E_t)} = \frac{-\sigma_u^2}{2\sigma_u^2 + \sigma_w^2} = \frac{-1}{2+p} \tag{5}$$

Thus, the smaller p is, the higher is the serial correlation of ΔE_t. For example, for $p = 0$, i.e. mean reverting process, the serial correlation coefficient is $-1/2$. For $p = \infty$, i.e. a random walk with drift, the serial correlation coefficient is 0.

The ability to identify the persistent and transitory components of earnings assumes special importance when p is low. Correctly classifying the two components enhances the ability to predict future earnings to a degree that is monotone increasing with the size of the transitory component. For example, assume that X_t, the persistent component, and u_t, the transitory component, are known, and thus correctly classified. Assume further that net earnings at period $t+1$, $\tilde{E}_{t+1} = E_t + \beta$. However, if $X_t + \beta$ is used as the estimate, the prediction error will be $\tilde{E}_{t+1} - X_t - \beta$. The reliability (variance of prediction error) of X_t as a predictor of $\tilde{E}_{t+1}$ is contrasted with the reliability of E_t as predictor below:

$$\mathrm{Var}(\tilde{E}_{t+1} - E_t - \beta) = 2\sigma_u^2 + \sigma_w^2 = 2\sigma_u^2 + p\sigma_u^2 = \sigma_u^2(2+p) \tag{6}$$

$$\mathrm{Var}(\tilde{E}_{t+1} - X_t - \beta) = \sigma_u^2 + \sigma_w^2 = \sigma_u^2 + p\sigma_u^2 = \sigma_u^2(1+p) \tag{7}$$

The difference is σ_u^2.

The relative improvement in the predictive reliability as a result of correct classification (i.e knowledge of X_t) is:

$$\Gamma = \frac{\sigma_u^2(2+p) - \sigma_u^2(1+p)}{\sigma_u^2(2+p)} = \frac{1}{2+p} = |R| \tag{8}$$

Thus, the absolute value of the serial correlation measures the relative improvement in predictive reliability. When $p = 0$, i.e. the process is mean reverting, Γ achieves its maximum at 1/2. That is, the maximum relative improvement in reliability resulting from correct classification is 50 per cent. Since X_t is unobservable, this 50 per cent improvement cannot be attained in practice. X_t has to be estimated using the observed series, E_t. Assuming (1) to be the earnings process, X_t can be estimated either subjectively by management or statistically (or both).

Contingent on (1), it is shown in the next section that the use of an X_t^F, estimated from E_t using a filtering procedure, can result in improved predictions in relation to use of E_t as the predictor.

2. THE FILTER PROCEDURES

The problem of estimating the unobservable, X_t, based on the observed time series $E_1, E_2, \ldots, E_t$ is called a filtering problem (Kalman, 1960; Sunder, 1975). Assuming that w and u in (1) follow a Gaussian process independent over time, the algorithm provided by Kalman (1960) yields an estimate for X_t that is optimal in the sense of minimizing the expected value of any monotonically non-decreasing loss function. Alternatively, if w and u do not follow a Gaussian process, then the Kalman filter minimizes any quadratic loss function.

The Kalman estimation procedure is conditional on the value of the unknown p. X_t is estimated (using the Kalman algorithm) by first estimating p using the maximum likelihood function due to Schweppe (1965).[3] The expression for the likelihood function is given in Appendix 2. The resulting estimate of X_t is denoted X_t^F.

When prediction is based on the filtered estimate, X^F, rather than on the actual X, the relative improvement in predictive reliability will be less than Γ, defined in equation (8). In particular, the asymptotic variance of the prediction error using X_t^F has been shown by Fisher (1971) to be:

$$\operatorname{Var}(\tilde{E}_{t+1} - X_t^F - \beta) = \sigma_u^2(1 + \alpha + p) \tag{9}$$

Denote the improvement in prediction reliability relative to the situation where E_t is used as predictor as Φ. Then:

$$\Phi = \frac{\sigma_u^2(2+p) - \sigma_u^2(1+\alpha+p)}{\sigma_u^2(2+p)} = \frac{1-\alpha}{2+p} \tag{10}$$

or equivalently:

$$\Phi = |R|(1-\alpha) \tag{11}$$

where α is the exponential smoothing weighting factor, $0 \leq \alpha \leq 1$. Since $\Gamma = |R|$, the relative improvement in predictive reliability is decreased, as a result of the necessity to estimate X, by a factor of $(1-\alpha)$. Denoting the improvement in predictive reliability when X is known relative to the situation where the estimate X^F is used as θ:

$$\theta = \frac{\Gamma - \Phi}{\Phi} = \frac{|R| - |R|(1-\alpha)}{|R|(1-\alpha)} = \frac{\alpha}{1-\alpha} \tag{12}$$

Thus, when $\alpha = 0$, i.e. $p = 0$ and the process is mean reverting, X and X^F produce the same improvement in predictive reliability relative to the naïve prediction rule which uses only E_t. In this case, X^F is the mean of all available observations and the contribution of the last observation is asymptotically zero. On the other hand, when $\alpha = 1$, i.e. $p = \infty$, the process is a random walk with drift, and knowledge of X will infinitely improve predictive reliability relative to using X^F.

If exact foreknowledge of $\tilde{E}_t$ is assumed, i.e. if a firm can formulate exact expectations of future earnings, managers can use this information to provide a better estimate of the persistent

[3] Considered also by Mehra (1970) and Garbade (1975).

component X (the smoothed estimate) relative to the filtered estimate. Such an extreme case is presented as a benchmark representing perfect foreknowledge that could be exploited to convey improved estimates of X. Whereas, in this unlikely circumstance, managers could directly present the future values of E_t, such forecast disclosures are unlikely to arise. They would be outside the domain of audited financial statements and of incentive contracts. They might also trigger adverse litigation. The more interesting and likely scenario is when a manager has limited foreknowledge in terms of estimating a p that could be used to extract X_t^F from past earnings.

Results presented in the foregoing section reflect the application of three different degrees of foreknowledge in the Kalman procedure: (1) zero foreknowledge, where p is estimated on the basis of past realizations only and then applied within the Kalman procedure to those same past realizations in extracting X_t^F; (2) an intermediate degree of foreknowledge as would result in using a superior estimate of p in relation to (1) above; and (3) complete foreknowledge of E_t up to a given future period T, resulting in a smoothed estimate X_t^S superior to X_t^F generated in (2) above, but not as good as perfect knowledge of X_t.

Kalman (1960) defines the smoothing estimation problem as that of estimating X_t based on the time series $E_1, E_2, \ldots, E_t, \ldots, E_T$. The foreknowledge regarding $t+1, \ldots, T$ used in producing X_t^S, the smoothed variable, improves predictive reliability beyond X_t^F. Being based upon future data, it is possible to construct the filter producing X_t^S as a two-sided low bandpass filter. However, the asymptotic variance of the prediction error using X_t^S, $\text{Var}(\tilde{E}_{t+1} - X_t^S - \beta)$, in terms of σ_u^2, α and p is unavailable for this study. The smoothing algorithm used to compute X is provided in Appendix 2.

3. SAMPLE AND INITIAL RESULTS

This section describes the results of predictive tests that provide evidence on the performance of the filtering model relative to a random walk model and the Box–Jenkins procedure. Comparative performance is evaluated for alternative degrees of foreknowledge. For the purpose of applying the filtering procedure and the prediction tests, a zero growth rate was assumed. This assumption avoids error associated with estimating β, the growth rate, and facilitates estimating the p.

Based on the procedures in Appendix 2, p was estimated for 114 firms using their annual net earnings data for the period 1953–1972.[4] The sample consisted of firms (1) from four-digit SIC industries having a minimum of ten companies each, (2) that had a December fiscal year end, and (3) whose securities were traded on the NYSE, ASE or OTC stock exchanges. Further, firms without consecutive earnings data for 1953–1972 were eliminated. These criteria were imposed principally for the purpose of future analysis. In particular, the years were chosen to precede 1973 when APB Opinion 30 (AICPA, 1973) became effective.

Exhibit 1 shows the distribution of the p-ratios for the 114 firms.[5] The lower the p-ratio, the smaller the α. From equations (5) and (12), the filtering procedure is expected to become more efficient in predicting future earnings relative to simply using the last earnings number for predictive purposes. As can be derived from Exhibit 1, 37 firms had a p-ratio less than or equal to 2. Exhibit 2 lists those 37 firms. For those firms, $|R|$ and Γ are expected to be greater than 0.25.

---

[4] The earnings data of the 114 firms and the computer programs to compute p are available from the authors upon request.
[5] When a trend was allowed for, as expected a larger number of firms had low p values. Trend was initially estimated using $(E_{20} - E_1)/20$, and then the filtering procedure was applied to estimate the maximum likelihood estimates for p. Using the trend procedure, 54 firms out of 114 had $p \leq 1$, whereas 64 firms out of 114 had $p \leq 2$. The results in the text are based on models which do not allow for a trend.

p	Number of firms
0.0	7
0.2	5
0.4	3
0.6	4
0.8	2
1.0	4
1.2	2
1.4	0
1.6	1
1.8	4
2.0	5
3.0	5
4.0	3
5.0	6
7.0	1
8.0	0
9.0	2
10.0	0
11.0	0
≥12.0	60
	114

Exhibit 1. The distribution of the p-ratio for the sample of 114 firms

Further, Φ should be greater than 0.07, from equations (8) and (11). Since it was reasonable to assume that for those companies the filtering procedure would be helpful in improving predictive ability relative to a strict random walk assumption, predictive tests were applied to those firms.

For the 37 firms, the filtered estimates, X_t^F, were computed as estimates of X_t for the years 1953–1972. All estimates were based on all E_t observations, up to and including period t. In addition, the smoothed values X_t^S were computed as estimates of X_t for the years 1953–1972 based on all the observations up to the period $t = 1972$. The Box–Jenkins estimates of X_t, denoted X_t^B, were also computed using the observations of the period 1953–1972 to identify and estimate the model. The Box–Jenkins procedure is described by Nelson (1973). Finally, the predictor for E_{t+1} based on the assumption the E_t follows a random walk was computed as $X_t^M = E_t$.

Let d_{jt}^M, d_{jt}^F, d_{jt}^S and d_{jt}^B be the prediction errors for firm j at period t resulting from using the random walk model, filtered, smoothed, and Box–Jenkins predicted values, respectively, defined as follows:

$$d_{jt}^M = E_{j,t+1} - E_{jt}$$

$$d_{jt}^F = E_{j,t+1} - X_{jt}^F$$

$$d_{jt}^S = E_{j,t+1} - X_{jt}^S$$

$$d_{jt}^B = E_{j,t+1} - X_{jt}^B$$

where $t = 1, \ldots, 19$ (the first year being 1953).

Exhibit 3 provides data on the predictive ability of the filtered estimate, X^F, the smoothed estimate, X^S, and the Box–Jenkins estimate, X^B, relative to the random walk estimate, X^M, based

Firm number	SIC code	Ticker	Firm name
1	10000274	AMX	American Metal Climax Inc.
2	10004600	N	International Nickel Company of Canada
3	10006087	MLY	Molybdenum of America
4	26004083	HML	Hammermill Paper Co.
5	26005828	MEA	Mead Corp.
6	26007935	SRT	St. Regis Paper Co.
7	26009055	QUCC	Union Camp Corp.
8	28010253	ACY	American Cyanamid Co.
9	28014270	HPC	Hercules Inc.
10	28030102	AXO	Akzona
11	28032271	CNK	Crompton and Knowles Corp.
12	28035006	KOP	Koppers Co.
13	28035542	MAF	MacAndrews and Forbes
14	28359153	UPJ	Upjohn Co.
15	29126267	MHR	Murphy Oil Corp.
16	30002657	DLP	Dunlap Holding Ltd.
17	30005644	MSF	Mansfield Tire and Rubber Co.
18	30006083	MWK	Mohawk Rubber Co.
19	30009091	R	Uniroyal Inc.
20	32410239	AAC	Amcord Inc.
21	32414515	IDL	Ideal Basic Industries Inc.
22	32415248	LPT	Lehigh Portland Cement Co.
23	33112176	COS	Copperweld Corp.
24	33114587	IK	Interlake Inc.
25	33115498	LUC	Luken Steel Co.
26	37140997	BOR	Borg-Warner Corp.
27	37141188	BF	Budd Co.
28	37143135	FMO	Federal-Mogul Corp.
29	37145300	LOF	Libbey-Owen-Ford Co.
30	37145664	MAR	Maremont Corp.
31	37148873	TKR	Timken Co.
32	37215398	LK	Lockheed Aircraft Corp.
33	37215801	MD	McDonnell-Douglas Corp.
34	37218841	THI	Theokol Chemical Corp
35	45110172	ALA	Allegheny Airlines Inc.
36	45111054	BNF	Braniff Airways Inc.
37	45119575	WAL	Western Airlines Inc.

Exhibit 2. List of firms with $p \leq 2.0$

	L_j^F Number of years (out of 19) for which $\lvert d_{jt}^F \rvert < \lvert d_{jt}^M \rvert$	L_j^S Number of years (out of 19) for which $\lvert d_{jt}^S \rvert < \lvert d_{jt}^M \rvert$	L_j^B Number of years (out of 19) for which $\lvert d_{jt}^B \rvert < \lvert d_{jt}^M \rvert$	AA_j^F $\frac{\sum_{t=m}^{19} \lvert d_{jt}^F \rvert}{\sum_{t=m}^{19} \lvert d_{jt}^M \rvert}$	AA_j^S $\frac{\sum_{t=m}^{19} \lvert d_t^S \rvert}{\sum_{t=m}^{19} \lvert d_t^M \rvert}$	AA_j^B $\frac{\sum_{t=m}^{19} \lvert d_t^B \rvert}{\sum_{t=m}^{19} \lvert d_t^M \rvert}$
1	2	3	4	5	6	7
Mean for 37 firms	10.11	12.84	10.16	0.940	0.752	0.915
Sampling standard error	—	—	—	0.014	0.014	0.038
$t = \left(\frac{1\text{-mean}}{\text{std. error}}\right)$	—	—	—	4.29	17.70	2.22
Panel A. Based on all predictions						
Mean for 37 firms	3.78	4.32	3.92	0.856	0.706	0.850
Sampling standard error	—	—	—	0.033	0.017	0.048
$t = \left(\frac{1\text{-mean}}{\text{std. error}}\right)$	—	—	—	4.36	17.29	3.10
Panel B. Based on last six years' predictions						

Exhibit 3. Absolute prediction errors by companies ($m = 1$ in Panel A and $m = 14$ in Panel B)

	QA_j^F	QA_j^S	QA_j^B
	$\frac{\sum_{t=m}^{19} (d_{jt}^F)^2}{\sum_{t=m}^{19} (d_{jt}^M)^2}$	$\frac{\sum_{t=m}^{19} (d_{jt}^S)^2}{\sum_{t=m}^{19} (d_{jt}^M)^2}$	$\frac{\sum_{t=m}^{19} (d_{jt}^B)^2}{\sum_{t=m}^{19} (d_{jt}^M)^2}$
1	2	3	4
Mean for 37 firms	0.838	0.548	0.870
Sampling standard error	0.022	0.013	0.074
$t = \left(\frac{1\text{-mean}}{\text{std. error}}\right)$	7.36	34.77	1.76
Panel A. Based on all predictions			
Mean for 37 firms	0.841 (0.742)	0.535	0.818
Sampling standard error	0.108 (0.29)	0.038	0.092
$t = \left(\frac{1\text{-mean}}{\text{std. error}}\right)$	1.47 (5.48)	12.24	1.982

Panel B. Based on last six years' predictions (the numbers in parentheses show the statistics for QA^F when an outlier firm with a QA^F of 4.583 was excluded)

Exhibit 4. Squared prediction errors by companies ($m = 1$ in Panel A and $m = 14$ in Panel B)

on absolute prediction errors.[6] Corresponding data, based on squared prediction errors, are provided in Exhibit 4.

Accuracy measures

Exhibit 3 presents the sample means of the following statistics:

L_j^F: the number of years for which $|d_{jt}^F| < |d_{jt}^M|$

L_j^S: the number of years for which $|d_{jt}^S| < |d_{jt}^M|$

L_j^B: the number of years for which $|d_{jt}^B| < |d_{jt}^M|$

$$AA_j^F = \frac{\sum_{t=m}^{19} |d_{jt}^F|}{\sum_{t=m}^{19} |d_{jt}^M|}; \quad AA_j^S = \frac{\sum_{t=m}^{19} |d_{jt}^S|}{\sum_{t=m}^{19} |d_{jt}^M|} \quad \text{and} \quad AA_j^B = \frac{\sum_{t=m}^{19} |d_{jt}^B|}{\sum_{t=m}^{19} |d_{jt}^M|}$$

The ratios AA_j^F, AA_j^S and AA_j^B are measures of relative absolute accuracy.

[6] The statistical tests and their results reported in the exhibits and in the discussion below implicitly assume no cross-sectional dependence among the transitory earnings components. Cross-sectional dependence would imply the existence of a transitory 'market factor' that is significantly correlated with the transitory earning component of individual firms. To test for such dependence, the transitory earnings 20 year sequences of 588 firms were regressed on the transitory 'market factor', defined as the sum over the 588 firms of their transitory components. For about 70 per cent of the firms, the R^2 statistics did not significantly differ from zero at the 0.05 level. This finding lends support to the assumption of cross-sectional independence. Additional support for the assumption is provided in the Spearman rank correlation coefficients between the prediction error series (19 years), computed for all distinct pairs among the sample of 37 firms. Only 28.2 per cent of 666 pairwise correlation coefficients (37 × 36/2) under the martingale model differed significantly from zero at the 0.05 level. Only 30 per cent (out of 666) and 48 per cent differed significantly from zero (at the 0.05 level) under the filtering and the smoothing models, respectively.

In Exhibit 3, Panel A, all 19 prediction errors are available (i.e. $m = 1$). In Panel B, however, only the last 6 years' prediction errors are available (i.e. $m = 14$), because it is assumed that the filtering procedure becomes significantly more efficient once it has been applied to a large number of years.[7]

For a majority of firms, the sum of the absolute deviations of the filtered, smoothed, and Box–Jenkins predicted earnings is less than that of non-filtered earnings. This is indicated by the fact that out of the 37 firms, 30 show $AA_j^F < 1$, 29 show $AA_j^B < 1$, and all 37 show $AA_j^S < 1$. Moreover, L_j^F, L_j^S, and L_j^B exceed 9.5 for 25 firms, 25 firms and 33 firms, respectively. Under the null hypothesis of equal predictive ability between the filtered, smoothed, and Box–Jenkins predicted variables on the one hand, and the random walk on the other, the expected value of L_j is 9.5. Therefore, if L_j exceeds 9.5 for a sufficiently large number of firms, the null hypothesis can be rejected in favour of the alternative hypothesis of superior predictive ability for X^F, X^S, and X^B, relative to X^M. It can be easily verified that the null hypothesis—that $L_j > 9.5$ for half the total number of firms—can be rejected in favour of the alternative hypothesis at the 0.01 level of significance (Z is 3.04 for L_j^F and L_j^B and 4.77 for L_j^S, where Z is the normal approximation to the binomial distribution). In addition, the averages[8] $\bar{L}^F$ (10.11), $\bar{L}^B$ (10.16) and $\bar{L}^S$ (12.84) significantly exceeded 9.5, the mean under the null hypothesis (Z was 2.44, 2.64, and 13.36, respectively).

The measures of relative accuracy, the averages $\overline{AA}^F$ (0.490), $\overline{AA}^B$ (0.915), and $\overline{AA}^S$ (0.752), were significantly smaller than 1, the expected mean under the null hypothesis (the t values were 4.36, 2.22 and 17.29, respectively). It is interesting to note that although the means of AA_j^F and AA_j^B were extremely close, the standard error for AA_j^B was almost three times larger than the standard error for AA_j^F. Note that when the earnings process is as described by equation (1), the filtering procedure minimizes the loss function. For earnings processes not described by equation (1), an ARIMA model chosen by the Box–Jenkins procedure might generate more accurate forecasts compared to the filtering procedure. Hence, the above result that the filtered estimates had lesser forecast variability than the Box–Jenkins estimates implies that the earnings data of our sample of firms are described by a process consistent with the process assumed in equation (1).

As expected, improved predictive ability via filtering was significantly stronger when only the last six observations were used.[9] (See Panel B of Exhibit 3, where, $\overline{AA}^F = 0.856$ and $\overline{AA}^B = 0.850$, as compared with means 0.940 and 0.915 in Panel A.) In addition, the value of Z for $\bar{L}^F$ in Panel B was 3.9, as compared with 2.44 in Panel A. Similar to Panel B, all the tests discussed in the context of Panel A were significant at the 0.01 level.

It can be seen from both panels of Exhibit 3 that the smoothed earnings variable, X^S, which assumes complete foreknowledge, produced considerably higher improvement in predictive ability relative to the random walk than did the filtered variable, X^F. Overall, the variables X^F, X^B, and X^S were significantly better predictors than X^M, the random walk.

The above results are based on absolute prediction errors and are indicators of predictive ability when losses contingent on prediction errors are linearly related to their absolute magnitudes. Exhibit 4 presents results based on squared prediction errors. This error metric is appropriate for individuals with quadratic loss functions. The measures of relative quadratic accuracy, QA^F, QA^S

---

[7] This initialization problem is well known in the filtering literature. See. for example, the discussion by Garbade (1975).

[8] $\bar{L} = \frac{1}{37}\sum L_j$, and $\overline{AA} = \frac{1}{37}\sum A_j$.

[9] Notice that no such improvement is expected to occur with respect to the smoothed variable. In both cases, X^S uses all 20 year observations. Comparison of the results in the two parts of Exhibit 3 shows that they are consistent with the expectation of no improvement.

and QA^B were computed as:

$$QA_j^F = \frac{\sum_{t=m}^{19} (d_{jt}^F)^2}{\sum_{t=m}^{19} (d_{jt}^M)^2}; \qquad QA_j^S = \frac{\sum_{t=m}^{19} (d_{jt}^S)^2}{\sum_{t=m}^{19} (d_{jt}^M)^2}; \qquad QA_j^B = \frac{\sum_{t=m}^{19} (d_{jt}^B)^2}{\sum_{t=m}^{19} (d_{jt}^M)^2}$$

Better results regarding the relative predictive ability of the filtering, smoothing, and Box–Jenkins procedures are expected in this case, because these procedures were derived by minimizing a quadratic loss function. Consistent with this contention, the accuracy measures presented in Exhibit 4, based on squared prediction errors (QA^F, QA^S and QA^B), provided stronger evidence of the comparative superior predictive ability of X^F, X^S and X^B.

One of the companies in Exhibit 4, Panel B, had a QA^F of 4.583, and appeared to be an outlier. Removing this outlier resulted in a significantly lower QA^F and a higher t-value. Both sets of results (including and excluding the outlier) are presented in Exhibit 4.

Reliability measures

As is well known, accuracy based on squared prediction errors includes both reliability and bias. The following relationship holds for each company:

$$\frac{1}{20-m+1}\sum_{t=m}^{19} d_{jt}^2 = \frac{1}{20-m+1}\sum_{t=m}^{19} (d_{jt} - \bar{d}_j)^2 + \bar{d}_j^{\,2}$$

where

$$\bar{d}_j = \frac{\sum_{t=m}^{19} d_j}{20-m+1}$$

The element on the left hand side is the accuracy measure, the first element on the right hand side is the reliability measure, and the second element is the bias.

Exhibit 5 provides the measure $\bar{d}_j$ (the square root of the bias) and the relative reliability measures defined as follows:

$$R_j^F = \frac{\sum_{t=m}^{19} (d_{jt}^F - \bar{d}_j^F)^2}{\sum_{t=m}^{19} (d_{jt}^M - \bar{d}_j^M)^2}; \qquad R_j^S = \frac{\sum_{t=m}^{19} (d_{jt}^S - \bar{d}_j^S)^2}{\sum_{t=m}^{19} (d_{jt}^M - \bar{d}_j^M)^2}; \qquad R_j^B = \frac{\sum_{t=m}^{19} (d_{jt}^B - \bar{d}_j^B)^2}{\sum_{t=m}^{19} (d_{jt}^M - \bar{d}_j^M)^2}$$

R^F is the empirical counterpart of $1-\Phi$ (equation (11)). The mean values of $1-\Phi_j$ are shown in Exhibit 5. There is a large degree of correspondence between the $1-\Phi$ measure and the observed statistic R_j^F. The average $1-\bar{\Phi}$ over the firms in the sample equalled 0.780. This means that the predictive reliability of X^F relative to X^M can be expected to improve by 22 per cent. In fact, $\bar{R}^F$ is 0.805 and 0.748 for the 19 years and 6 years, respectively. The close correspondence between $1-\bar{\Phi}$ and its empirical counterpart provides evidence regarding the efficiency of the maximum likelihood procedure employed to estimate p for each firm using only 20 observations. Had the procedure resulted in a poor estimate of the ps, it is unlikely that such a close correspondence would have been obtained.

The results presented in this section demonstrate that, conditional on the process assumed in (1), a manager's superior knowledge could be exploited with a Kalman filter procedure to provide

	$\bar{d}^F$	$\bar{d}^S$	$\bar{d}^M$	$1-\Phi$	R_j^F $\frac{\sum_{t=m}^{19}(d_{jt}^F-\bar{d}^F)^2}{\sum_{t=m}^{19}(d_{jt}^M-\bar{d}^M)^2}$	R_j^S $\frac{\sum_{t=m}^{19}(d_{jt}^S-\bar{d}^S)^2}{\sum_{t=m}^{19}(d_{jt}^M-\bar{d}^M)^2}$	R_j^B $\frac{\sum_{t=m}^{19}(d_{jt}^B-\bar{d}^B)^2}{\sum_{t=m}^{19}(d_{jt}^M-\bar{d}^M)^2}$
Mean for 37 firms	1.6799	0.8641	1.1187	0.780	0.805	0.543	0.824
Sampling standard error	—	—	—	—	0.022	0.033	0.074
$t = \left(\frac{1\text{-mean}}{\text{std. error}}\right)$	—	—	—	—	8.86	35.15	2.38
Panel A: Based on all predictions							
Mean	0.3036	0.1924	0.0728	0.780	0.748	0.502	0.770
Sampling standard error	—	—	—	—	0.045	0.016	0.087
$t = \left(\frac{1\text{-mean}}{\text{std. error}}\right)$	—	—	—	—	5.6	31.13	2.63
Panel B. Based on last six years' predictions							

Exhibit 5. Bias and reliability measures ($m = 1$ in Panel A and $m = 14$ in Panel B)

financial statement users with estimates of X that would allow them to obtain better forecasts of future earnings than those based solely upon past realized earnings (e.g. a random walk model).

Two forms of superior knowledge were assumed in the prediction tests. First, an extreme case of complete foreknowledge of the future earnings series up a future period T was used in computing a smoothed estimate X_t^S. This was then used as benchmark against which the effect of lesser knowledge could be evaluated. Secondly, a more limited form of superior knowledge was assumed—one that allowed a manager to apply, within the Kalman procedure, an estimate of p superior to that which could be obtained from the observation of only past realizations of E_t. Specifically, it was assumed that management could use 'informed' expectations regarding possible future changes in the structure of earnings (relation between the persistent and the transitory) in updating, and thus improving, its estimate of the ratio of the variance of the two earnings components, p, in relation to an estimate extracted only from past observations. The surrogate used for such improved estimates was the one extracted from all 20 years of sample data. Moreover, the results indicated that the earnings data of the sample were consistent with the process assumed in (1), and that Kalman-based estimates exhibited less variability than the Box–Jenkins estimates that required subjective model selection based on 20-year sample data.

4. PREDICTIVE ABILITY TESTS USING NO SUPERIOR KNOWLEDGE

This section evaluates predictions using the Kalman procedure. Such predictions assume *no* superior knowledge on the part of managers (or, alternatively, whatever superior knowledge is possessed would not be revealed). Only past realizations of E_t are assumed available for these computations. Similarly, it is assumed that users of financial statements cannot exploit inside information. In this phase of the analysis, all predictions are made outside the period in which the p-ratio was estimated. The results could provide some joint but indirect and limited evidence on: (1) the process assumed in equation (1), and (2) the efficiency of the filtering procedure. The limitation of the evidence stems from the small finite samples used to estimate p. For each firm, the following computation was made (notations have the same meaning as in the preceding section).

For each period i between 15 and 19, the p-ratio was estimated on the basis of the first i observations in the 19 year subsample. On the basis of this estimate, the prediction errors, $d_{j,i}^F$, were computed. Thus, for example, for $i = 15$, $d_{j,15}^F = E_{j,16} - X_{j,15}^F$ is the filtered estimate based on the p-ratio estimated from the first 15 years' observations. These prediction errors were compared with d_{ji}^M, random walk-based prediction errors, as shown in Exhibit 6. So, for each firm, only 5 filtered values, X_t^F, were estimated for the last 5 of the 19 year subsample. Exhibit 6 includes only firms whose estimated p-ratios based on the first 15 observations were equal to, or less than 2.0 (note that p was updated in each consecutive prediction year). Exhibit 6 shows the mean values L_j^F and R_j^F based on the five prediction years.

On a global examination of Exhibit 6, the filter did not produce better predictions than the random walk prediction. This is evidenced by the t-statistic of -1.02, computed on the QA_j^F data. But p-ratios estimated on very small samples are likely to be biased downward. For example, in the special case of only two observations, the process will always be estimated as a mean reverting process, and the p-ratio will always be estimated as zero. The downward bias in the estimate of p in a small sample could perhaps be at least partially responsible for the relatively poor predictive performance. Asymptotically, the estimate of p is best asymptotically normal since it is a maximum likelihood estimator.

On the other hand, assuming that $\beta = 0$ when it is actually different from zero introduces an upward bias in the estimate of p. Since the ultimate aim is to test the extent to which filtering and

	L_j^F Number of years (out of 5) for which $\lvert d_{jt}^F \rvert \le \lvert d_{jt}^M \rvert$	QA_j^F $\dfrac{\sum_{t=15}^{19} (d_{jt}^F)^2}{\sum_{t=15}^{19} (d_{jt}^M)^2}$
1	2	3
Mean for 44 firms	2.59	1.14 (1.02)
Sampling standard error	—	0.14 (0.06)
$t = \left(\dfrac{\text{1-mean}}{\text{std. error}}\right)$	—	−1.02 (−0.28)

Exhibit 6. Squared prediction errors by companies, based on last five years' predictions. The p-estimate is updated each period. The numbers in parentheses show the statistics for QA^F when an outlier firm with a QA^F of 6.53 was excluded

smoothing improve predictive ability conditional on the model, the potential for such improvement is greater for firms with low ps. The upward bias results in reducing the apparent efficiency of filtering and smoothing in determining predictive ability. Thus, the true potential of filtering and smoothing might be higher than is apparent in the results.

In any event, for the cases where $p < 0.04$, the predictions based on the filtered values are superior. For example, out of the 30 cases, QA_j^F exceeded 1.0 only in 10 cases. Similarly, $L_j^F \le 2$ only in 10 cases.

5. CONCLUDING REMARKS

The results of the preceding section—describing predictive tests using no knowledge beyond observed past realizations—were hampered by small sample sizes. Nonetheless, when the empirically estimated p-ratio was less than or equal to 0.04, predictions appeared to be superior to random walk based predictions. Thus, for the set of firms exhibiting $p \le 0.04$, the evidence was consistent with the process assumed in (1), and the filtering procedure was efficient in relation to a random walk model. The implication is that users of financial statements could apply the Kalman procedure with advantage in those cases where the estimate of p, based on no more than information available in income statements, is extremely small (zero in our sample). Thus, an effective device, p is provided for determining when the persistent component can be separated from the transitory component with advantage.

The results reported in Section 3 were more definitive. When superior knowledge exists, it can be exploited with considerable advantage to improve predictive accuracy and reliability in relation to a random walk based model which has been shown to be quite robust (Ball and Watts, 1972; Albrecht *et al.*, 1977; Watts and Leftwich, 1977, among others). Moreover, given the uncertainty inherent in selecting a proper model under the Box–Jenkins procedure, the Kalman procedure can provide less variable estimates in relation to the former when the underlying earnings process is correctly assessed. The analysis herein yielded results consistent with the procedure assumed in (1). The results seemed also to be consistent with Beaver *et al.*'s (1980) findings that security price behaviour implies the perceived earnings process to be a compound mechanism, which includes a transitory component.

Two general scenarios were examined: (1) managers have no foreknowledge, i.e. no comparative advantage relative to outsiders with respect to knowledge of the future earnings series or p, and (2) managers have some foreknowledge. First, a manager's estimated p could be viewed as incorporating expected changes in the earnings structure. Secondly, and at the extreme, managers could be viewed as knowledgeable of all future earnings. In Scenario (1), either managers or outside users can filter the earnings series using only available observations and thus improve the ability to predict future earnings when the presence of a transitory component is empirically indicated. It could be argued that since both outsiders and managers could filter available published earnings numbers, one cannot derive useful implications from such a test regarding accounting policy. However, if the total social cost of filtering the earnings number is reduced when firms' managers do it rather than individual users, it is reasonable to conclude that firms should communicate their filtered earnings to users assuming that incentives for 'cheating' or 'falsifying' can be minimized by audits and other monitoring procedures. A convenient vehicle for doing this within generally accepted accounting principles (GAAP) is to separate out the transitory noise ($E_t - X_t^F$) as an extraordinary item. Thus, an FASB standard based on such a procedure may be appropriate.

In the second scenario, where managers are assumed to have total or some foreknowledge, the implication for accounting policy is much clearer. Managers should convey superior knowledge of future earnings or of p to market participants so that the ordinary income series X_T^S will convey useful information to the market. This could be provided within GAAP by applying the Kalman procedure (or perhaps some equivalent thereof) and segregating the transitory component ($E_t - X_t^S$) as an extraordinary item.

Some limitations of this study should be pointed out. The shortness of the series examined limits the ability to draw useful implications for policy matters. In this regard, the overall quality of the Compustat data used should also be considered. But, although the examination of a larger series should have increased confidence in the implications drawn, it seems reasonable, on the basis of the analysis presented in this paper, to conclude tentatively that the rules governing the characterization of extraordinary items as stated in Accounting Principles Board No. 30 are too restrictive.

It should also be noted that earnings data reported after 1972 were generated under a different set of GAAP than before 1972. In fact, this was one reason our sample included only data up to and including 1972. Although there was also a change in GAAP related to extraordinary items in 1966 (APB Opinion 9, AICPA, 1966), this was not as drastic as the one introduced by APB Opinion 30. In future studies, comparison of the transitory component estimated by the Kalman procedure and extraordinary items reported by management separately in the periods before and after Opinion 30, should yield additional useful insights with respect to the advisability of the classification criteria included in the Opinion. Finally, future research can compare earnings predictions based on Kalman filters and other models with security price based predictions so some further insight can be gained regarding the prediction method implied in information contained within security prices.

ACKNOWLEDGEMENTS

We are indebted to Edwin Elton, Larry Fisher and an anonymous referee for helpful comments.

APPENDIX 1

Description of Kalman theory

There are essentially two ways of filtering linear stochastic systems that result in minimizing mean square error. Box and Jenkins (1970) approaches the problem from a statistical viewpoint by fitting autoregressive and moving average (ARIMA) models to a set of data. The second approach uses Kalman's general theory of state-space representation for linear systems. This algorithm is useful for estimation problems when prior information on the internal structure of the dynamic system exists. The two approaches are equivalent since they are both based on linear models that describe observed data as the output of a system subjected to a white noise input. Since Kalman's state-space representation theory is not as well known as the Box and Jenkins approach, this Appendix presents a brief description of Kalman's state-space representation theory.

Consider a single input/output system in which W_t denotes the input, X_t the true (unobservable) output, U_t an additive noise disturbance and E_t the observed output so that:

$$E_t = X_t + U_t \tag{13}$$

and, for all t, the relationship between the true input and output can be expressed as a finite order linear difference equation of the form:

$$\begin{aligned} X_t &= \sum_{j=0}^{n} \theta_j W_{t-j} \\ &= \Phi_n(B) W_t \end{aligned} \tag{14}$$

where B is a lag operator and $\Phi_n(B)$ is a polynomial in B. Assume, further, that $\Phi_n(B)$ is a rational function of the form $\Phi_n(B) = \Psi_n^{-1}(B)\lambda_n(B)$, where

$$\Psi_n(B) = 1 - \Psi_1 B - \cdots - \Psi_n B^n, \qquad \text{and} \qquad \lambda_n(B) = \lambda_1 + \lambda_2 B + \cdots + \mathrm{j}_n B^{n-1}$$

It will be assumed that the zeros of $\Psi_n(B)$ and $\lambda_n(B)$ lie outside the unit circle. Thus, the relationship between E_t and W_t is:

$$E_t = \Phi_n(B) W_t + U_t \tag{15}$$

and equation (14) can be rewritten as:

$$\Psi_n(B) X_t = \lambda_n(B) W_t \tag{16}$$

so that equations (13) and (14) can be put in the form:

$$E_t = H_m(B) X_t + U_t \tag{17}$$

$$X_{t+1} = \Psi_n'(B) X_t + \lambda_n(B) W_{t+1} \tag{18}$$

Equations (17) and (18) are known as the state-space representation. A linear unbiased minimum variance estimator, $\hat{X}_{t+1}$, of X_{t+1}, given the observed sequence $E_1 \ldots E_t$, is obtained from the Kalman filter algorithm. This algorithm is based on the assumptions that: (1) $X_t = \Phi_n(B) W_t$ is an ARMA model with Gaussian residuals, (2) U_t is an independent zero mean Gaussian random variable with known covariance and (3) $\Psi_n'(B)$, $\lambda_n(B)$ and $H_m(B)$ are known linear filters. Output from the algorithm contain the following information: (1) The estimate of X_t at time t, $X_{t+1|t} = \Psi_n'(B) X_{t|t}$, (2) The updated predicted variance $V_{t+1|t} = E(X_{t+1} - X_{t+1|t})^2$ and (3) The filter gain function F_{t+1}, which is a function of $H_m(B)$ and $V_{t+1|t}$. After observing E_{t+1}, the smoothed estimate of X_{t+1} is estimated by:

$$X^s_{t+1|t+1} = X_{t+1|t} + F_{t+1}(E_{t+1} - H_m(B) X_{t+1|t}) \tag{19}$$

Let $Z_{t+1} = E_{t+1} - H_m(B) X_{t+1|t}$ in equation (19). It can be shown that $E(Z_t Z_{t'}) = 0$ if $t \neq t'$. The transformation of E_t to Z_t is similar to a Gram–Schmidt orthogonalization procedure which transforms E_t into a vector of independent random variables. The estimate of X_{t+1} given the observations $E_1, \ldots E_{t+1}$ is equal to the prediction of X_{t+1} using observations $E_1, \ldots, E_t$ plus an adaptive term, representing additional information. This is a generalization of the exponentially weighted moving average model.

The state space representation approach differs from the Box–Jenkins approach in the amount of data required for model fitting. ARMA models represent a class of black-box time-series models fitted to an observable set of data for the purpose of predicting future events. Usually many models describe the input–output relationship of the historical data equally well, but generate widely different forecasts. State space models, on the other hand, use *a priori* information of the physical nature of the system and thus provide a mechanistic internally descriptive model of the process that is usually more informative to the model builder. The requirement of specifying the parameters allow the application of the Kalman filter being applied to forecasting situations in which little data is available (such as for the problem in this paper). The construction of ARMA models requires a great deal of data in order to choose reasonable models and to estimate the parameters. In the absence of *a priori* information, a modelling procedure begins by fitting all reasonable ARMA models to the data. Each model is then transformed into its equivalent state-space form and then the investigator chooses the model which most closely describes his knowledge of the dynamics of the system under study.

Finally, ARMA models require that the underlying process be weakly stationary whereas the state-space representation only requires that X_t follow a wide-sense Markov process.

APPENDIX 2

Kalman filter algorithm

The assumed state-space representation of the earnings-generating process in this paper is:

$$E_t = X_t + U_t$$

$$X_t = X_{t-1} + W_t$$

Let X_t^F be the filtered estimate of X_t, the persistent component. Let $X_1^F = E_1$. For $t > 1$, X_t^F is obtained from the following recursive equation (assuming a given value of p):

$$X_t^F = X_{t-1}^F + \frac{R_t}{1 + R_t}(E_t - X_{t-1}^F)$$

R_t and S_t are defined by the following recursive equations:

$$R_t = S_{t-1} + p$$

and

$$S_t = R_t - \frac{R_t^2}{1 + R_t}$$

(The problem in this paper is concerned with estimating the state of a linear, discrete, stochastic system. A similar problem is estimating a set of parameters in a variable parameter regression model where the parameters vary according to an *a priori* known parameter variation law. In the second problem, R_t is proportional to the one-step ahead predicted coefficient estimates, and S_t is proportional to the updated coefficient estimate).

The maximum likelihood estimators of p and σ_u

The log likelihood function of net earnings, $E_1, \ldots, E_T$, is:

$$L(E_1, \ldots E_T) = -\frac{1}{2} T \ln(2\pi) - \frac{1}{2} \sum_{t=2}^{T} \ln[\sigma_u^2(1+R_t)] - \frac{1}{2} \sum_{t=2}^{T} [(E_t - X_{t-1}^{\mathrm{F}})^2 / \sigma_u^2(1+R_t)]$$

Maximizing $L(.)$ with respect to σ_u^2 yields, for a given p, the following maximum likelihood estimator for σ_u^2:

$$\hat{\sigma}_u^2 = \frac{1}{T-1} \sum_{t=2}^{T} [(E_t - X_{t-1}^{\mathrm{F}})^2 / (1+R_t)]$$

Notice that $\hat{\sigma}_u^2$ is a function of p. Different values of p will yield different filtered series X_t^{F} and thus different $\hat{\sigma}_u^2$.

Substituting back into the maximum likelihood function, L can be expressed as a function of p only:

$$L(p) = C - (T-1) \ln(\hat{\sigma}_u(p)) - \frac{1}{2} \sum_{t=2}^{t} \ln(1+R_t)$$

where C is a constant independent of p.

There is no analytical solution for the maximizing value of p. The search for the optimal p was performed using a grid search.

The Kalman smoothing algorithm

Let X_t^{S} be the smoothed estimate of X_t, the persistent component. Set $X_T^{\mathrm{S}} = X_T^{\mathrm{F}}$. For $t < T$, X_t^{S} is obtained from the following backward recursive equation (assuming a given value of p):

$$X_t^{\mathrm{S}} = X_t^{\mathrm{F}} + G_t(X_{t+1}^{\mathrm{S}} - X_t^{\mathrm{F}})$$

where G_t is defined by:

$$G_t = S_t / (S_t + p)$$

where X_t^{F} and S_t are the values derived by the filter algorithm described above.

Initialization

Let $X_1^{\mathrm{F}} = E_1$, $S_1 = 1$ and $p = 0$. For each value of p, compute σ_u^2, X_t^{F} and X_t^{S}. The value of p is updated in steps of 0.2 units, and the algorithm terminates when the likelihood function, $L(p)$, reaches its maximum. This algorithm assumes that β, the growth element (see discussion in the text), equals zero, which is also the assumption made in the empirical tests.

REFERENCES

Albrecht, W. S., Lookabill, L. L. and McKeown, J. C., 'The time series properties of annual earnings', *Journal of Accounting Research*, **15** (1977), 226–244.

American Institute of Certified Public Accountants, Accounting Principles Board, Accounting Principles Board Opinion No. 9, *Reporting the Results of Operations*, New York: AICPA, December 1966.

American Institute of Certified Public Accountants, Accounting Principles Board Opinion No. 30, *Reporting the Results of Operations, Reporting the Effects of Disposal of a Segment of a Business, and Extraordinary, Unusual and Infrequently Occurring Events and Transactions*, New York: AICPA, June 1973.

Ball, R. and Watts, R. L., 'Some time series properties of accounting income', *The Journal of Finance*, **27** (1972), 663–681.

Barnea, A., Ronen, J. and Sadan, S., 'The implementation of accounting objectives: an application to extraordinary items', *The Accounting Review*, **50** (1975), 58–68.
Barnea, A., Ronen, J. and Sadan, S., 'Classificatory smoothing of income with extraordinary items', *The Accounting Review*, **51** (1976), 110–122.
Beaver, W., Lambert, R. and Morse, D., 'The information content of security prices', *Journal of Accounting and Economics*, **2** (1980), 3–28.
Box, G. E. P. and Jenkins, G. M., *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1970.
Financial Accounting Standards Board, Statement of Financial Accounting Concepts No. 1, *Objectives of Financial Reporting by Business Enterprises*, Stamford, Conn.: FASB, 1978.
Fisher, L., 'A replacement of exponential smoothing', *Business and Economics Statistics Section Proceedings of the American Statistical Association* (1971).
Garbade, K., 'The initialization problem in variable parameter regression', Research paper, New York: Graduate School of Business, New York University, 1975.
Kalman, R. E., 'A new approach to linear filtering and prediction problems', *Journal of Basic Engineering*, **82** (1960), 35–45.
Mehra, R. K., 'On the identification of variances and adaptive Kalman filtering, *IEEE Transactions on Automatic Control*, **15** (1970), 175–184.
Miller, A. M. and Modigliani, F., 'Dividend policy, growth and the valuation of shares', *The Journal of Business*, **34** (1961), 411–433.
Muth, J. F., 'Optimal properties of exponentially weighted forecasts', *Journal of the American Statistical Association*, 60, (1960), 299–306.
Nelson, C., *Applied Time Series Analysis for Managerial Forecasting*, San Francisco: Holden-Day, Inc., 1973.
Ronen, J. and Sadan, S., *Smoothing Income Numbers: Objectives, Means, and Implications*, Reading, Mass: Addison-Wesley Publishing Company, 1981.
Ronen, J. S., Sadan, S. and Snow, C., 'Income smoothing: a review', *The Accounting Journal*, **1** (1977), 11–26.
Schweppe, F. C., 'Evaluating of the Likelihood functions for Gaussian signals', *IEEE Transactions of information Theory*, **1** (1965), 61–70.
Sunder, S., 'Accounting changes in inventory valuation', *The Accounting Review*, **50** (1975), 305–315.
Watts, R. L. and Leftwich, R., 'The time series of annual accounting earnings', *Journal of Accounting Research*, **15** (1977), 253–271.

Authors' biographies:
Joshua Ronen is at present Professor of Accounting at New York University's Graduate School of Business Administration. His main research interest lies in the interface between capital markets and corporate management in its role as producer of accounting information with discretion over the use of alternative accounting techniques.

Edward I. Melnick is Professor of Statistics at New York University. Dr. Melnick's research interests focus on time-series analysis and information theory.

Zvi Lieber is Senior Lecturer in Accounting, Tel Aviv University, Israel.

Authors' addresses:
Joshua Ronen and Edward I. Melnick, New York University Graduate School of Business Administration, New York University, New York, NY 10006, U.S.A.

Zvi Lieber, Faculty of Management, Tel Aviv University, Tel Aviv, Israel.