

Smit, E.

Article

Combining quantitative and qualitative data in business cycle research

South African Journal of Business Management

Provided in Cooperation with:

University of Stellenbosch Business School (USB), Bellville, South Africa

Suggested Citation: Smit, E. (1986) : Combining quantitative and qualitative data in business cycle research, South African Journal of Business Management, ISSN 2078-5976, African Online Scientific Information Systems (AOSIS), Cape Town, Vol. 17, Iss. 3, pp. 143-148, <https://doi.org/10.4102/sajbm.v17i3.1048>

This Version is available at:

<https://hdl.handle.net/10419/217935>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Combining quantitative and qualitative data in business cycle research

E. Smit

Department of Statistics, University of the Western Cape, Bellville

In this article a leading indicator of the South African business cycle is proposed which combines the traditional quantitative data inputs with qualitative data. The integration is achieved via the Kalman filter technique. It is shown that this model surpasses the traditional approaches in accuracy.

S. Afr. J. Bus. Mgmt. 1986, 17: 143–148

In hierdie artikel word 'n leidende indikator van die Suid-Afrikaanse konjunktuur voorgestel wat op sowel die tradisionele kwantitatiewe data-insette as op kwalitatiewe data-insette steun. Die integrasie geskied via die Kalman-filteertegniek. Daar word aangetoon dat hierdie model die tradisionele benaderings in akkuraatheid oortref.

S.-Afr. Tydskr. Bedryfsl. 1986, 17: 143–148

Paper presented at the 1985 Conference in Management and Economic Sciences, Cape Town, 29–31 October 1985

Introduction

It may be too much of a generalization to state that a schism exists between the protagonists and users of quantitative and qualitative data respectively. However, the main thrust of business cycle research in the United States has been directed towards the construction of quantitative indicators, whereas the accent in Europe has been on the analysis of qualitative survey data. Efforts directed towards a synthesis were mainly due to econometricians trying to improve the forecasting accuracy of models based on quantitative data by introducing qualitative variables into their equations. Our own efforts have shown that quantitative variables tend to suppress qualitative variables when used in combination and that qualitative variables tend to suppress each other due to high levels of multicollinearity (Smit, 1983). On balance, qualitative data have the advantages of early availability, of capturing expectations data, and of providing information on a broader spectrum of variables than traditionally published in quantitative form, whereas quantitative data excel providing 'richer' or 'fuller' information due to measurement on a higher measurement scale. The problem is how to optimally integrate and use both types of data in economic analyses and forecasting. It is believed that the Kalman filter technique opens new horizons in this regard.

In 1975 Aiginger used the Kalman filter to illustrate how quantitative investment anticipations could be used to correct constants in the investment equation (Aiginger, 1975). This paper reports on research aimed at incorporating qualitative data into a quantitative model via the Kalman filter. The problem is the establishment of a leading indicator for the South African business cycle. Two approaches are described: a quantitative technique used by Barr (1983) and a qualitative technique after Smit & Kotze (1984). The Kalman filter is then used to integrate these approaches.

Measuring the business cycle: van Coller's coinciding indicator

Since 1977 the Bureau for Economic Research at the University of Stellenbosch has been publishing composite leading and coinciding indicators of the South African business cycle on a monthly basis. These indices were developed by van Coller (1980) using the methods of the NBER. The five series constituting the coinciding indicator are as follows: registered unemployment, index of physical volume of manufacturing production, wholesale sales, retail sales, and imports.

This composite coinciding indicator has proved to be of

E. Smit

Graduate School of Business, University of Stellenbosch, P.O. Box 610, Bellville, 7530 Republic of South Africa

Accepted April 1986

great value in the South African context and many firms use this variable as an input in their own forecasting models, justifying the effort exerted in forecasting the coinciding indicator. The leading indicator developed by van Coller, speaking on the average, has sufficient lead time to be a truly *ex-ante* indicator, but it has the disadvantage of inconsistent lead times at peaks and troughs ranging between 3 and 21 months. This has led to efforts to establish an alternative leading indicator showing more consistent timing at peaks and troughs.

An alternative based on quantitative data: Barr's proposal

Barr (1983) applied a stepwise regression method for selecting, in an optimal way, from 23 variables considered, those variables which represent the best set of predictors for van Coller's coinciding indicator. This method, he argued, obviates the need to make *ex-ante* predictions about the relevance of the variables as leading indicators or to analyse consistency of turning-point relations because the variable with the highest correlation will necessarily provide the best average correlation with the business cycle and thus be the most consistent measure of turning points. Furthermore, this method provides weights which are automatically selected in an optimal way and hence obviates the selection of subjective weightings when averaging.

The model is fitted over the period 1967 M1 – 1981 M10 and the following explanatory variables contribute towards a R^2 value of 0,95: new motor cars licenced, imports, commercial bank loans and discounts, and commercial bank demand deposits. Due to a lag of three months between the dependent and all of the independent variables it is possible to calculate a predicted value for the coinciding indicator three months ahead of its realization. However, due to delays in the publication of quantitative data, data on the variables at time t may only become available after another four months have elapsed, destroying the true *ex-ante* character of the leading indicator.

An alternative based on qualitative data: Smit and Kotze's proposal

Smit and Kotze (1984) stressed the importance of the psychological disposition of businessmen and argued that the expectations and plans of economic agents provide the link between future and present economic activity. They showed that psychological links do not only occur in the so-called psychological business cycle theories, but that they also feature within the so-called real theories, although not in the same exalted position. The underlying idea is that if changes in moods and expectations should lead to changes in real activity, then in theory it should be possible to predict the development of the business cycle when measurements on psychological variables are available.

They tested this hypothesis using the subjective qualitative expectations and plan data published by the Bureau for Economic Research at the University of Stellenbosch as part of the Opinion Survey and Building Survey results. All *ex-ante* variables related to expectations (statements about future events outside the direct sphere of influence of the firm) and plans (statements about the firm's own future course of action) are included in the analysis provided that sufficient observations are available.

These qualitative series, when compared to the quantitative series usually used in indicator research, have the disadvantages of being available only on a quarterly instead of a monthly basis and of covering only a few sectors in the

economy. Furthermore about 50% of the series only date back to 1974:4. However, these series have the major advantage of being available three months ahead of the event they relate to. The following variables have qualified for selection:

MSALES	: manufacturing; expected volume of sales;
MPROD	: manufacturing; expected volume of production;
MUNOR	: manufacturing; expected ratio of unfilled orders to demand;
MRAWMAT	: manufacturing; stocks of raw materials relative to planned production;
MFINPRO	: manufacturing; stocks of finished goods relative to expected demand;
MBUSCLI	: manufacturing; expected business climate;
WSALES	: wholesale; expected volume of sales;
WSTOCK	: wholesale; stocks relative to expected demand;
WBUSCLI	: wholesale; expected business climate;
RSALES	: retail; expected volume of sales;
RSTOCK	: retail; stocks relative to expected demand;
RBUSCLI	: retail; expected business climate;
CSTOCK (N)	
(U) (S)	: motor trade; stocks of new cars/used cars/spares relative to expected demand;
CBUSCLI	: motor trade; expected business climate;
BPCAP	: manufacturers of building materials; expected production capacity;
BBUSCLI	: builders; expected business climate.

The variables are measured relative to the same quarter in the previous year (higher, the same, lower or better, the same, worse) with the exception of the variables related to stocks which are reported as too high, just sufficient or too low. The qualitative measurements are transformed to balance series on the interval [0, 100].

It has been argued that it is unnecessary to make *ex-ante* evaluations about the economic significance of the variables as leading indicators or to analyse the consistency of turning-point relations when a stepwise regression approach is used. Nor is it necessary to select subjective weightings when averaging because these will also be selected optimally. Similar arguments apply in the case where the weighting is done by means of factor- and regression analysis. The latter should also take care of the consistency of timing at peaks and troughs. As far as economic significance is concerned, important sectors of the economy are not covered by the survey (amongst others agriculture, mining and banking) but because of the high degree of economic interdependence between the manufacturing sector, the wholesale, retail and motor trade and the sectors which are not surveyed, it may be argued that the subjective opinions expressed in the survey are representative of a much wider field of economic activity.

The high intercorrelations between the different balance series necessitate a reduction in the dimensions of the explanatory data base before regression relationships between the balance series and a quantitative business cycle indicator can be explored. The original data base is reduced to three factors by means of principal components analysis followed by a Varimax factor rotation in order to provide a representation of the factor loadings matrix which facilitates interpretation. This rotation simply transforms the axes retained from the principal components solution into a new orthogonal space of the same dimensions. The procedure is aimed at improving the ease of interpretation of the reference

Table 1 Rotated factor matrix

Variable	Factor loadings		
	Factor 1	Factor 2	Factor 3
MSALES	0,80	0,47	-0,21
MPROD	0,77	0,50	-0,21
MUNOR	0,89	0,34	-0,19
MRAWMAT	-0,76	-0,21	0,14
MFINPRO	-0,64	-0,43	0,38
MBUSCLI	0,88	0,44	-0,11
WSALES	0,62	0,36	-0,13
WSTOCK	-0,78	-0,04	0,15
WBUSCLI	0,80	0,33	-0,08
RSALES	0,29	0,76	-0,25
RSTOCK	-0,22	-0,40	0,13
RBUSCLI	0,64	0,51	-0,09
CSTOCK (N)	-0,68	-0,13	0,59
CSTOCK (U)	-0,46	-0,36	0,59
CSTOCK (S)	0,02	-0,17	0,80
CBUSCLI	0,73	0,43	-0,31
BPCAP	0,16	0,10	-0,73
BBUSCLI	0,77	0,41	-0,25

dimensions. The rotated factor matrix is presented in Table 1. It is based on observations over the period 1974:4–1984:4.

The total variance accounted for is 77,6%. Unfortunately the object of obtaining a relatively high loading on a single factor only is not achieved which complicates the identification of the underlying factors. The inverse relationship between the stock variables and the rest of the pro-cyclical variables is emphasized by the above factor loadings.

In the factor analysis standardized values of the survey variables are used, therefore the original balance series are standardized over the period 1974:4–1984:4 before the three factors, F1, F2 and F3, are calculated as linear combinations of the survey variables. CI, van Coller's coinciding indicator averaged on a quarterly basis and transformed to percentual yearly changes on a quarterly basis.

$$CIC_t = (CI_t - CI_{t-4})/CI_{t-4} \times 100$$

is used as the dependent variable in the regression analysis. The rationale for using percentual changes in the case of variables measured relative to the same quarter in the previous year is to be found in the work of Anderson (1952), whilst Smit (1983) demonstrated that the stock variables (measured as high, just sufficient or too low) correlate significantly with yearly differences in certain components of final demand.

After the application of the Durbin two-stage procedure for the correction of autocorrelation, the following estimated regression relationship is obtained:

$$\widehat{CIC}_t = 4,213 F1_{t-1} + 2,447 F2_{t-1} - 1,884 F3_{t-1} + 0,006$$

(6,01) (4,85) (3,54)

$$R^2=0,56; \quad DW=1,28; \quad PF: 75:2-84:4$$

The respective beta-coefficients are 43,2%, 32,9% and 23,9% with the *t* values in brackets.

The model is completed by the addition of a second equation which transforms the percentual differences into levels of CI:

$$\widehat{CI}_t = CI_{t-4} + \widehat{CIC}_t \times CI_{t-4}/100$$

The model is simulated over the sample period (1975:2–1984:4) by means of forecasts one period ahead. This reflects the practical forecasting situation for which the model is designed. It is possible to forecast a value for CI almost three months early and the lead time from forecast to the publication of the data is more than six months.

The Kalman filter

Meinhold & Singpurwalla (1983) 'demystified' the Kalman filter by putting it in the context of Bayesian inference. Their exposition was as follows.

Given a data set Y_t, Y_{t-1}, \dots, Y_1 where Y_t is dependent on a non-observable quantity θ_t known as the state of nature and the object is inference about θ_t , the *observation equation* describes the relationship between Y_t and θ_t as

$$Y_t = F_t \theta_t + v_t.$$

θ_t is a scalar or vector independent of the dimension of Y_t , and v_t is the observational error which is assumed to be normally distributed with mean 0 and variance V_t .

The state of nature changes over time according to the *systems equation*:

$$\theta_t = G_t \theta_{t-1} + w_t.$$

The systems error is w_t which is also assumed to be normally distributed with mean 0 and variance W_t . Furthermore it is assumed that v_t and w_t are independent.

The Kalman filter is a recursive method of inference about θ_t based on a Bayesian procedure.

At time $t-1$ the knowledge about θ_{t-1} conditional on the information Y_{t-1} is embodied in the following probability statement

$$(\theta_{t-1}/Y_{t-1}) \sim N(\widehat{\theta}_{t-1}, \Sigma_{t-1})$$

where $\widehat{\theta}_0$ and Σ_0 are best guesses for the mean and variance of (θ_{t-1}/Y_{t-1}) at time 0.

Looking ahead at t in two stages, the best choice of θ_t before Y_t is observed is given by the systems equation, therefore

$$(\theta_t/Y_{t-1}) \sim N(G_t \widehat{\theta}_{t-1}, R_t = G_t \Sigma_{t-1} G_t' + W_t).$$

This is the prior distribution of (θ_t/Y_{t-1}) . Define e_t as the error made in the prediction of Y_t from time $t-1$, i.e.

$$\begin{aligned} e_t &= Y_t - \widehat{Y}_t \\ &= Y_t - F_t G_t \widehat{\theta}_{t-1}. \end{aligned}$$

Observing Y_t is thus equivalent to the observation of e_t and after Y_t has been observed

$$P(\theta_t/Y_t, Y_{t-1}) = P(\theta_t/e_t, Y_{t-1}).$$

It can now be shown that the posterior distribution of θ_t is given by

$$\begin{aligned} (\theta_t/e_t, Y_{t-1}) &\sim N(G_t \widehat{\theta}_{t-1} + R_t F_t' (V_t + F_t R_t F_t')^{-1} e_t, \\ &R_t - R_t F_t' (V_t + F_t R_t F_t')^{-1} F_t R_t). \end{aligned}$$

The procedure may be summarized in the following way:

- (i) After time $t - 1$ the posterior distribution of θ_{t-1} has mean $\widehat{\theta}_{t-1}$ and variance Σ_{t-1}
- (ii) At time t , form a prior distribution with mean $G_t \widehat{\theta}_{t-1}$ and variance $R_t = G_t \Sigma_{t-1} G_t' + W_t$ and
- (iii) evaluate a likelihood given $e_t = Y_t - F_t G_t \widehat{\theta}_{t-1}$.
- (iv) This leads to a posterior distribution of θ_t with mean $\widehat{\theta}_t = G_t \widehat{\theta}_{t-1} + R_t F_t' (V_t + F_t R_t F_t')^{-1} e_t$ and variance $\Sigma_t = R_t - R_t F_t' (V_t + F_t R_t F_t')^{-1} F_t R_t$.
- (v) Continue with the recursive procedure.
- (vi) $K_t = R_t F_t' (V_t + F_t R_t F_t')^{-1}$ is known as the Kalman filter.
- (vii) The mean of the posterior distribution is the regression of θ_t on e_t .
- (viii) The mean of the posterior distribution is therefore equal to the sum of the mean of the prior distribution and the regression coefficient of θ_t on e_t , multiplied by the error committed in predicting the next observation.

The aforementioned principles can now be applied to a dynamic linear system consisting of p observation equations and q systems equations

$$Y_t = H_t S_t + U_t^y + v_t$$

$p \times 1 \quad p \times q \quad q \times 1 \quad p \times 1 \quad p \times 1$

$$S_t = \Phi_t S_{t-1} + U_t^s + w_t$$

$q \times 1 \quad q \times q \quad q \times 1 \quad q \times 1 \quad q \times 1$

The usual assumptions apply. Given Y_1, Y_2, \dots, Y_{t-1} , the problem is to estimate S_k . If $k = t - 1$, the problem is referred to as Kalman filtering, if $k > t - 1$ it is referred to as Kalman forecasting, and if $k < t - 1$ it is a problem in Kalman smoothing.

It can be shown that the minimum mean squared estimate of S_t based on Y_1, Y_2, \dots, Y_{t-1} can be recursively expressed as

$$\widehat{S}_{t/t-1} = U_{t-1}^s + \Phi_t \widehat{S}_{t-1/t-2} + K_{t/t-1} (Y_{t-1} - U_{t-1}^y - H_{t-1} \widehat{S}_{t-1/t-2}).$$

$K_{t/t-1}$ is called the Kalman gain for one stage prediction and is given by

$$K_{t/t-1} = \Phi_t \Sigma_{t-1/t-2}^s H_{t-1}' (H_{t-1} \Sigma_{t-1/t-2}^s H_{t-1}' + \Sigma_{t-1}^v)^{-1}$$

where by definition

$$\widetilde{S}_{t-1/t-2} = S_{t-1} - S_{t-1/t-2}$$

and

$$\Sigma_{t-1/t-2}^s = E(\widetilde{S}_{t-1/t-2} \widetilde{S}_{t-1/t-2}')$$

$\Sigma_{t-1/t-2}^s$ is the covariance matrix of the error in predicting S_{t-1} based on observations up to time $t - 2$. The covariance matrix can be recursively updated by

$$\Sigma_{t/t-1}^s = (\Phi_t - K_{t/t-1} H_{t-1}) \Sigma_{t-1/t-2}^s + \Sigma_t^w$$

The estimate $\widehat{S}_{t/t-1}$ is obtained by predicting forward the estimate of the immediately preceding period and then correcting with a weighted combination of the components of the innovation vector $Y_{t-1} - U_{t-1}^y - H_{t-1} \widehat{S}_{t-1/t-2}$.

The filtered estimate of S_t based on Y_1, Y_2, \dots, Y_t can be

recursively expressed as

$$\widehat{S}_{t/t} = U_t^s + \Phi_t \widehat{S}_{t-1/t-1} + K_t [Y_t - U_t^y - H_t (U_t^s + \Phi_t \widehat{S}_{t-1/t-1})]$$

where K_t is the Kalman filter.

$$K_t = \Phi_t^{-1} K_{t+1/t} \quad \text{and}$$

$$\Sigma_{t/t}^s = (I - K_t H_t) \Sigma_{t/t-1}^s$$

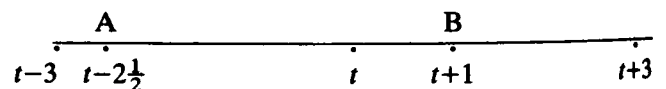
where $\Sigma_{t/t-1}^s$ is the prior and $\Sigma_{t/t}^s$ is the posterior variance.

In order to apply the filtering technique to the data described in earlier sections, the values of the coinciding indicator, CI_t , are associated with the systems variable S_t and the expectations about CI_t , namely CI_t^e , as generated by Smit and Kotze, are associated with the observed variable Y_t . Equations generating CI_t and CI_t^e , respectively, are introduced and the information is combined via the Kalman filter.

It must be kept in mind that the formulae were derived under the assumption that no past or present observations on the state variable are available. In the problem considered here, in addition to the observations of Y_t , it is also possible to observe S_{t-1} . This being the case, it is possible to replace $\widehat{S}_{t-1/t-1}$ by S_{t-1} in all the equations and because this modified estimate is obtained by using more information, it is to be expected that the resultant estimates will be more efficient than those containing $\widehat{S}_{t-1/t-1}$.

The results

In the analysis of the results presented here, it should be kept in mind that not all the forecasts are made at the same point in time. The following diagram presents the earliest stage at which it is possible to apply a particular technique (time is measured in months and the aim is to forecast the value of CI_t for the quarter ranging from $t - 3$ to t):



At point A, two weeks into the quarter, expectations data become available and the Smit/Kotze technique can be applied and an expectation for CI_t , namely CI_t^e can be calculated. At the same time the Kalman filter estimate for an unknown S_{t-1} can be calculated. At point B, one month into the next quarter, quantitative information about the lagged variables in the Barr model becomes available and the quantitative model can be applied. At the same time the Kalman filter estimate for a known S_{t-1} can be calculated.

Model 1¹

In this model a weak generator of CI_t is tested. The system equation contains van Coller's leading indicator as an explanatory variable, whilst it is well known that the lead time of this variable tends to be highly erratic. Expectations are generated using an adaptive expectations mechanism.

$$CI_t = 6,61846 + 0,06218 LI_{t-1} + 0,86465 CI_{t-1}$$

$(1,20) \quad (2,68) \quad (14,86)$

$$R^2=0,901; \quad SEE=2,03654; \quad PF:1975:2-1984:3$$

$$CI_t^e = 7,87296 + 0,60393 CI_{t-1}^e + 0,31869 CI_t$$

$(1,11) \quad (4,79) \quad (2,73)$

$$R^2=0,840; \quad SEE=2,38950; \quad PF:1975:2-1984:3$$

$$\Sigma_t^v = (2,38950)^2 = 5,7097$$

$$\Sigma_t^w = (2,03654)^2 = 4,1475$$

$$\Sigma_{1/0}^s = 1,000$$

The variables CI_t and CI_t^e have been defined before and LI_t represents van Coller's leading indicator. All error measures reported in Tables 2, 3 and 4 are over the full sample period.

In both cases (S_{t-1} unknown and S_{t-1} known) the filtered estimates perform better than the estimates based on the systems equation only. Both of these models, however, are inferior to the Smit/Kotze expectations generating mechanism. This is due to the poor specification of the systems equation.

Table 2 Root mean square errors: Model 1

Smit/Kotze	S_{t-1} unknown	S_{t-1} known
3,2270	5,0190 (Systems equation) 4,2011 (Filtered)	2,002 (Systems equation) 1,872 (Filtered)

Model 2

In this model a better specification of the systems equation is used, using the variables proposed by Barr in a Koyck lag scheme. The same expectations generating mechanism as used in Model 1 applies.

$$CI_t = 32,38065 + 0,2105 MV_{t-1} + 0,00142 DEMPOS_{t-1} + 1,41889 IMPORT_{t-1} - 0,0094 LOAND_{t-1} + 0,53058 CI_{t-1}$$

(4,98) (2,48) (4,63)
(1,51) (4,80)
(4,65)

$$R^2=0,966; \quad SEE=1,29266; \quad PF:1975:2-1984:3$$

$$CI_t^e = 7,87296 + 0,60393 CI_{t-1}^e + 0,31869 CI_t$$

(1,11) (4,79) (2,73)

$$R^2=0,840; \quad SEE=2,38950; \quad PF:1975:2-1984:3$$

$$\Sigma_t^w = (1,29266)^2 = 1,6710$$

$$\Sigma_t^v = (2,28950)^2 = 5,7097$$

$$\Sigma_{1/0}^s = 1,000.$$

The new variables are MV_t , new motor vehicles registered, $DEMPOS_t$, commercial banks demand deposits, $IMPORT_t$, real imports and $LOAND_t$, commercial banks loans and discounts.

Table 3 Root mean square errors: Model 2

Smit/Kotze	S_{t-1} unknown	S_{t-1} known
2,557	1,466 (Systems equation) 1,393 (Filtered)	1,194 (Systems equation) 1,136 (Filtered)

Once more, the filtered estimates outperform the estimates based on the systems equation only. Table 3 also illustrates the improvement in results due to a better specification of the systems equation. In this case, the improvement due to filtering is considerably less than in the case of Model 1.

Model 3

In the period 1983:3-1984:3 the South African economy experienced a mini-boom in the midst of a downswing phase due to excessive government expenditure. This caused most leading indicators to break down² and the indicator based on qualitative data, for one, never registered this phenomenon. At the same time a quick growth in the money supply also invalidated the quantitative indicator of Barr. Both models were re-estimated and the following equations materialized:

$$CI_t = 32,28295 + 0,02327 MV_{t-1} + 2,61147 IMPORT_{t-1} + 0,44390 CI_{t-1}$$

(4,80) (4,41) (4,05)

$$R^2=0,933; \quad SEE=1,67676; \quad PF:1975:2-1984:4$$

$$CI_t^e = 7,90723 + 0,21530 CI_t + 0,70795 CI_{t-1}^e$$

(1,25) (2,74) (7,27)

$$R^2=0,87; \quad SEE=1,85962; \quad PF:1975:2-1984:4$$

$$\Sigma_t^w = (1,67676)^2 = 2,8115$$

$$\Sigma_t^v = (1,85962)^2 = 3,4582$$

$$\Sigma_{1/0}^s = 1,000.$$

All the variables have been defined previously.

Table 4 Root mean square errors: Model 3

Smit/Kotze	S_{t-1} unknown	S_{t-1} known
3,5043	2,1572 (Systems equation) 2,1433 (Filtered)	1,6208 (Systems equation) 1,6110 (Filtered)

Once more it is evident that the Kalman filter improves the accuracy of the estimate of CI_t .

Conclusion

This article has explored an approach aimed at using qualitative expectations data as well as the lead time in the publication of this data. Three approaches have been compared: an approach using quantitative data, one based on qualitative data and one integrating the two data sets via the Kalman filter. For three different models the Kalman filter has led to improved forecasts and further research into the application of the Kalman filter to qualitative data seems a worthwhile pursuit.

Notes

1. Models 1 and 2 do not use the same factor loadings as presented in Table 1. This is due to the fact that the principal components analysis has been done twice over the past ten quarters. The factors loadings have remained very stable over time and the addition of new information.
2. This breakdown was already evident in Model 2 where the $LOAND$ variable switched signs in the systems equation.

References

- Aiginger, K. 1975. Trade Cycle Survey and Investment Anticipation Data in Econometric Models for Austria. 12th CIRET-Conference, Stockholm.
- Anderson, O. 1952. The Business Test of the IFO-Institute for Economic Research and its Theoretical Model. *Rev. Int. Stat. Inst.*, vol 20.
- Barr, G.D.I. 1983. A Leading Indicator for the South African Economy. *J. Stud. Econ. Econometrics*, Nr 15.
- Mariano, R.S. & Schleicher, S. 1972. On the Use of Kalman Filters in Economic Forecasting. Discussion Paper 247, Wharton School of Finance and Commerce.
- Meinhold, R.J. & Singpurwalla, N.D. 1983. Understanding the Kalman Filter. *Am. Stat.*, vol. 37, Nr 2.
- Smit, E. vd M. 1983. 'n Ekonometriese Bestedingsmodel van die Suid-Afrikaanse Ekonomie gebaseer op Kwalitatiewe Data Insette, *J. Stud. Econ. Econometrics*, Nr 15.
- Smit, E. vd M. & Kotze, D. 1984. A Psychological Profile of the South African Business Cycle based on Qualitative Data. *J. Stud. Econ. Econometrics*, Nr 20.
- Van Coller, J.E.M. 1980. A System of Composite Leading and Coinciding Indicators for the South African Business Cycle, *J. Stud. Econ. Econometrics*, Nr 8.