SERIE RESEARCH MEMORANDA

ADAPTIVE FORECASTING WITH HYPERFILTERS

A.H.Q.M. Merkies

Researchmemorandum 1988-18

June 1988

VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
EN ECONOMETRIE
AMSTERDAM

ADAPTIVE FORECASTING WITH HYPERFILTERS

Arnold H.Q.M. Merkies
Faculty of Economics and Econometrics, Vrije Universiteit, Amsterdam, The Netherlands

Ivo J. Steyn
Faculty of Economics and Econometrics, University of Amsterdam, Amsterdam, The Netherlands

ABSTRACT

Adaptive forecasting with exogenous variables has developed from recursive regression (RR) to Kalman filters (KF). The estimation of variances within the K.F. is time consuming and often imprecise. An alternative is the use of optimally updated variance estimates. Ljung & Söderström's "Recursive Prediction Error" method is amended somewhat and used to predict housing starts in the Netherlands.

KEY WORDS: Adaptive estimation, Recursive Residuals, Kalman Filter, State Space Models.

1. INTRODUCTION

Adaptive forecasting with exogenous information has developed from recursive regression (RR) to Kalman filters (KF). The adaptation is restricted to means, estimates of variances are rarely updated. Their estimation within the KF is time consuming and often imprecise. Proper application of KF therefore requires prior knowledge about variances or a sample period sufficiently large to detect stable optima. Moreover, in continued forecasting the question remains when reestimation is due. A possible solution to these problems is to replace efficient estimation by optimal adaptation. If variance estimates are optimally updated their original values lose significance as time passes by. Optimal updating formulae for the variances are provided by Ljung and Söderström (1983). In this paper their recursive prediction error method is applied with some amendments to predict housing starts in the Netherlands from permits issued. This approach does not introduce a different model e.g. with time varying variances. Yet it may be considered a generalization of the KF and a fortiori of RR as it involves time varying estimates. Because of its generalization aspect we will call it a hyperfilter (HF). If updated variance estimates converge through time to some constant the HF also offers an alternative to numerically obtained estimates in a KF.

In the sections 2, 3 and 4 we will subsequently discuss the RR, the KF and the hyperfilter. In section 5 we reveal some of the merits of the HF in an empirical setting. Section 6 summarizes.

2. RECURSIVE REGRESSION

In a regression model the relation between a dependent variable y_t and K explanatory variables is given by

$$(1) \qquad y_t = x_t'\alpha + u_t$$

where x_t' is a row vector with K explanatory variables at time t, α a

column vector with K coefficients and u_t an error term, with

(2) $\quad E\ u_t = 0$

and

(3) $\quad E\ u_t u_r = \sigma^2 \quad \text{if} \quad t = r$
$\qquad\qquad = 0 \quad \text{if} \quad t \neq r.$

If time series of x and y are available α can be estimated with least squares:

(4) $\quad \hat{\alpha} = (X'X)^{-1}\ X'y$

with T-1 observations X and y are a (T-1) * K matrix and a (T-1) vector respectively. A forecast of y for period T with (1) needs at least one additional observation of each of the explanatory variables to give the row vector x_T'. The forecast is then computed as

(5) $\quad \hat{y}_T = x_T'\ \hat{\alpha}.$

With one additional value of the dependent variable this forecast can be evaluated from the forecasting error

(6) $\quad e_T = y_T - \hat{y}_T$

A sharper but also more restrictive evaluation of e_T is possible if an assumption on the statistical distribution of u_t is added e.g.

(7) $\quad u_t \sim N(0,\sigma^2)$

for all t, which leads to

(8) $\quad e_T \sim N(0,\sigma^2 s_T)$

with

(9) $\quad s_T = 1 + x_T'(X'X)^{-1}x_T .$

Before the forecasting procedure is repeated for period T+1 the estimate $\hat{\alpha}$ can be updated with (4) using for X a T*K matrix and for y a vector of length T. To indicate that this new estimate refers to period T+1 whereas the previous one refers to period T we will indicate the two estimates by $\hat{\alpha}_{T+1}$ and $\hat{\alpha}_T$ respectively and write (9) as

(10) $\quad s_T = 1 + x_T' P_T x_T$

with

(11) $\quad P_T = (X'X)^{-1}$

where X is as before (T-1)*K.
Alternatively $\hat{\alpha}_{T+1}$ can be computed with Recursive least squares (RR):

(12) $\quad \hat{\alpha}_{T+1} = \hat{\alpha}_T + M_T e_T$

where

(13) $\quad M_T = P_T \, x_T / s_T .$

The charm of RR is that all new information is contained in $M_T e_T$; the information on the exogenous variables in M_T and that of the dependent variable in e_T. Substitution of (13),(10) and (11) into (12) gives a formula equivalent to (4). The K * K updating matrix P_T can also be computed recursively by

(14) $\quad P_T = P_{T-1} - M_{T-1} \, M_{T-1}' \, s_{T-1}$

with s_t from (10) and M_t from (13), see Harvey (1981, par.7.1). The forecasting procedure then consists of the following steps for each subsequent period:

1) Given all variables for t = T-1,
we update : P_T from (14)

2) Given also x_t' for t = T,
we update : s_T from (10)
M_T from (13)

and predict : $\hat{y}_T$ from (5) with $\hat{\alpha} = \hat{\alpha}_T$

3) Given also y_t for t = T,
we evaluate : e_T from (6)

and update : $\hat{\alpha}_{T+1}$ from (12)

We may start the forecasting procedure from T = 1 with initial condition $P_0 = \mu I$ where μ is a large scalar or we may start at T = K with P_T from (11). For T ≥ K the two procedures coincide.

We have concentrated upon updating $\hat{\alpha}_t$. The remaining parameter of the model i.e. σ^2, sometimes called hyperparameter, has drawn less attention, which is not surprising as this parameter is not needed in the forecasting procedure. In this case σ^2 can be estimated by maximum likelihood. The log likelihood function expressed in forecasting errors multiplied by -2 is, see (8):

$$(15) \qquad -2 \log L = T \log 2\pi\sigma^2 + \Sigma \log s_t + \sigma^2 \Sigma e_t^2/s_t .$$

This function of σ^2 attains its maximum at

$$(16) \qquad \hat{\sigma}_T^2 = 1/T \Sigma e_t^2/s_t$$

which is the ML estimator of σ^2 with T observations, given the estimate $\hat{\alpha}_T$. Clearly

$$(17) \qquad \hat{\sigma}_T^2 = (1 - 1/T)\hat{\sigma}_{T-1}^2 + 1/T \, e_T^2/s_T$$

may be used as an updating formula if so desired. It converges to some constant $\hat{\sigma}^2$ for $T \to \infty$.
Note that substitution of $y_t - x_t'\alpha_t = e_t/\sqrt{s_t}$ transforms (15) into the likelihood function of α_{T+1}. As $\hat{\alpha}_T$ is a ML estimator based on T-1 observations, e_T in (17) is already optimized with respect to $\hat{\alpha}_T$. Note also that the standard error of $\hat{\alpha}_T$ can be estimated by $\hat{\sigma}_T^2 P_T$ and computed recursively.

3. KALMAN FILTERS

In estimating $\hat{\alpha}_{T+1}$ from (12) we assumed the underlying model (1) did not change from T-1 to T. Such is in general not realistic in an economic context. As time passes by preferences and reactions change. That is why the state space model was introduced. In the present context this means the following pattern through time of the unknown parameter α:

(18) $$\alpha_t = D\,\alpha_{t-1} + \varepsilon_t$$

with

(19) $$E\varepsilon_t = 0$$

and

(20) $$E\varepsilon_t\varepsilon_t' = \Sigma = \sigma^2 Q.$$

Note that ε_t is a K-vector like α_t and hence Σ and Q are of order K*K. Again a more restrictive model is obtained if we add a distribution:

(21) $$\varepsilon_t \sim N(0, \sigma^2 Q)$$

where Q is introduced to relate the variance-covariance matrix of ε_t to σ^2. The two terms in the righthand side of (18) indicate deterministic and stochastic changes in α respectively. Note that, if $\varepsilon_t = 0$ for all t, (18) can be rewritten as $\alpha_t = D^t\alpha_0$; so if we transform x_t in

(1) by $z_t = x_t D^t$ and adopt the initial condition $\alpha_0 = \alpha$ we have cast the deterministic component of the state space model in a recursive regression setting. Similarly the stochastic component can be annexed to the recursive regression model by taking the variance-covariance matrix of u as Ω with typical element $\omega_{ts} = \sigma^2 x_t' Q_{ts} x_s + Eu_t u_s$ instead of $\sigma^2 I$. Hence the state space model is only a simple redraft of a rather cumbersome recursive regression model. Its advantage lies in its superior presentation of the underlying dynamic forces of the model. As we now have a regression problem with a non-scalar variance-covariance matrix Ω simple least squares as in (4) is no longer efficient and should be replaced by generalized least squares. Consequently the updating parameter estimation formulae (12),(13) and (14) change too and we get the Kalman filter, introduced by Kalman (1960).

Instead of (12),(13) and (14) we have

$$\hat{\alpha}_{T+1} = D\,\hat{\alpha}_T + M_T e_T \tag{22}$$

$$M_T = DP_T\, x_T / s_T \tag{23}$$

$$P_T = DP_{T-1}D' - M_{T-1}M_{T-1}' s_T + Q. \tag{24}$$

Forecasting and forecast evaluation, however, is still performed by (5) and (6). For given D and Q we may start as before from $T = 1$ with $P_0 = \mu I$; the alternative for $T \geq K$ would now require $P_T = (X'\Omega X)^{-1}$ with Ω as the special structured matrix referred to above and is therefore less suitable in this context. Note that $\hat{\alpha}_{T+1}$ may differ from $\hat{\alpha}_T$ even if the forecasting error e_T is zero due to the deterministic component of the Kalman filter. And if the forecasting error e_T is non zero, the adaptation of $\hat{\alpha}_T$ differs from that of the recursive regression setting as disturbances ε_t should, but disturbances u_t should not lead to a different α.

The problem remains what value we should adopt for D and Q. Any value will do, for instance with $D = I$ and $Q = 0$ we are back with formulae (12),(13) and (14), but this is not what we had in mind, when we

introduced the Kalman filter. We are interested whether there are superior values of D and Q. In this paper we will concentrate on Q only and use a given matrix D. As no specific alternative for Q offers itself we may let the data decide.
To estimate Q in an optimal sense we need a criterion. A suitable optimizing criterion is the likelihood function (15), where the solution (16) is now explicitly taken as a function of Q, which in itself may be a function of a limited number of parameters $\theta_1, \theta_2, \ldots$ Minimization of $\hat{\sigma}_T^2$ with respect to Q or to $\theta_1, \theta_2, \ldots$ does not lead to an explicit analytical formula, so one resorts to numerical optimization. This implies that $\hat{Q}$ is approximated by $\hat{Q}_{nT}$ in which the suffix n indicates what numerical procedure is used and T refers to the number of observations. All procedures such as DFP (Davidon, Fletcher and Powell), BFSG (Broyden, Fletcher, Shanno and Goldfarb), Powell, HJ (Hooke and Jeeves) and BHHH (Berndt, Hall, Hall and Hausman), see e.g. Fletcher (1980), start with some initial value of Q and form successive approximations to the maximum value of the likelihoodfunction.
The use of numerical solutions has three important implications:

a) The results become less precise.
b) The computational costs rise.
c) It evokes a tendency to restrict the problem.

a) If all procedures would end up with the same $\hat{Q}$ the suffix n would be of relevance only if one is interested in speed of convergence. If they may end up with different approximations of the same $\hat{Q}$ the suffix is more crucial. Such situations occur frequently in time series analysis when the surface of (15) is rather flat or contains multiple maxima. In the present context the problem of the best approximating estimation procedure has not been solved yet. In such circumstances the results lack precision.

b) The speed of convergence also presents a problem. Each of the alternatives is rather costly: for three unknown parameters - θ_1, θ_2 and θ_3 - in Q an average of 80 likelihood evaluations is not uncommon and each evaluation takes appreciable time: for each choice D=I and

$Q = \hat{Q}$ we need the procedure to calculate the $\hat{\alpha}_1, \ldots, \hat{\alpha}_T$ consecutively with (22),(23) and (24) from the initial choice $\hat{\alpha}_0 = 0$ and the starting conditions $T = 1$ and $P_0 = \mu I$.

c) Due to the inaccuracy of the results repeating the estimation of Q for each t=1,2,...,T may generate erratic time paths of $\hat{Q}_{nt}$. Moreover such a procedure evokes a huge computational burden. Hence the common practice is to select a particular T for which the optimal $\hat{Q}_{nT}$ is computed and this value is maintained for all t beyond T.

4. HYPERFILTERS

In the recursive regression model updating of the variance was not needed but could be done quite easily. In the Kalman filter model, where updating should take part in the forecasting procedure it is too laborious.
A solution to the problem was given by Ljung and Söderström (1983). These authors reverse the sequence in which the problem is presented. In the K.F. first a satisfactory $\hat{Q}_{nT}$ is sought and the question of updating is usually not even posed. In the Recursive Prediction Error method of Ljung and Söderström updating comes first. LS start with an arbitrary initial estimate $\hat{Q}_t$ of Q at t=0 and update $\hat{Q}_t$ at each t. This is an obvious solution as it can be shown that if an adequate updating formula is used $\hat{Q}_t$ converges to the ML estimate. If at some T the value of $\hat{Q}_t$ has converged the numerical problems of K.F. have been circumvented and Q may still be updated to capture model changes, which is a second improvement of H.F. over K.F.
The forecasting procedure in the two filters is similar. In H.F. we only insert an updating formula for Q and use the most recent estimate of Q in the remaining formulae.
The practical implications of the hyperfilter are considerable:
- We no longer need a separate numerical optimization routine.
- The data are run through the hyperfilter only once.
- A new observation requires only one iteration to process.

The crucial question is how to adapt our estimates optimally if new

information enters. Striving for maximum likelihood estimators we can use (17); the introduction of state space (18) changed the model without violating the likelihood in terms of forecasting errors. New information contained in y_T and x_T' cannot reduce the first term of the right hand side of (17). If we assume that $\hat{\sigma}^2_{T-1}$ is already optimal for T-1 observations the updating question is reduced to finding that change in $\hat{Q}$ which minimizes e_T^2/s_T given the observations y_T and x_T'. Assume that Q depends upon a (column)vector θ of parameters. This vector contains at maximum K(K+1)/2 elements e.g. $\theta=(q_{11},\ldots,q_{KK})$. The function f: $\theta_T \rightarrow e_T^2/s_T$ is rather complex; to minimize this function with respect to θ_T we approximate it by a Taylor expansion up to a quadratic term in our previous estimate $\hat{\theta}_{T-1}$:

$$(25) \qquad f(\theta_T) \simeq f(\hat{\theta}_{T-1}) + \beta_T'(\theta_T-\hat{\theta}_{T-1}) + (\theta_T-\hat{\theta}_{T-1})'B_T(\theta_T-\hat{\theta}_{T-1})$$

where

$$\beta_T = \delta f/\delta\theta \text{ evaluated at } \hat{\theta}_{T-1} \text{ (gradient)}$$

and

$$B_T = \delta^2 f/\delta\theta\delta\theta' \text{ evaluated at } \hat{\theta}_{T-1} \text{ (Hessian)}.$$

The optimum of (25) with respect to θ is obtained if

$$(26) \qquad \theta_T - \hat{\theta}_{T-1} = -B_T^{-1}\,\beta_T.$$

Hence if we add some reluctancy to deviate from our previous estimate the optimal value of θ_T may be taken as

$$(27) \qquad \hat{\theta}_T = \hat{\theta}_{T-1} - \gamma B_T^{-1}\beta_T$$

with $\gamma = 1/T$.

Add the simplification $B \simeq \beta\beta' + \tau I$ and we are left with the single but still impressive problem of finding β_T. This is a technical matter

which is dealt with in the Appendix.
With β_T known we can compute $\hat{\theta}_T$ from (27) and rearrange its elements into $\hat{Q}_T$. If the time series is long enough and the model remains valid $\hat{Q}_T$ will converge to some $\hat{Q}$. It is interesting to compare this $\hat{Q}$ with 0 to evalualate the recursive regression model and with $\hat{Q}_{nT}$ to evaluate the Kalman filter estimate with T observations and numerical procedure n. If $\hat{Q}_T$ is not too different from $\hat{Q}_{nT}$ it can be considered to be the limit value and be used in a standard Kalman setting from T onwards. In this sense the hyper filter procedure embodies the Kalman filter. It does not enclose the KF in the sense that all estimates $\hat{Q}_{nT}$ for any or some n are nested in the hyperfilter. For small T the values of $\hat{Q}_T$ and $\hat{Q}_{nT}$ usually differ and the former provides another numerical procedure to estimate Q. Of course it remains wise to use $\hat{Q}_T$ even if it coincides with $\hat{Q}_{nT}$ to be prepared for changes that may occur in the model as time passes by.

5. HOUSING STARTS IN THE NETHERLANDS

The hyperfilter model was applied to forecast housing starts in the Netherlands. Construction in the Netherlands requires a permit, which expires after three months. So with adequate administrations and law abiding constructors construction in quarter t only starts if a permit is issued in that or the previous quarter. This suggests the relation

$$(28) \qquad B_t = \alpha_{0t} V_t + \alpha_{1t} V_{t-1} + u_t$$

where B_t is the amount of construction started (in millions of 1969 guilders) and V_t is the amount of permits issued (in millions of 1969 guilders).
If permits that are not used or that relate to construction which is discontinued are removed from the observations we should have $\alpha_{1t} = 1 - \alpha_{0t-1}$. Errors may occur if administrators or constructors deviate from the rules. Some care must be observed with respect to these errors in this context as argued by Hendry and Richard (1982). The issue may be clarified: if V_t equals a constant c from period t=0 onwards ideally construction starts B_t should be equal to that constant too from t=1

onwards as both V_t and B_t refer to the same amount of construction. Hence one would expect some restriction of the kind $\Sigma_t u_t = 0$. We, however, adopt the view that (small) errors may occur in the equality between $\Sigma_t V_t$ and $\Sigma_t B_t$ due to statistical errors and the like. This justifies our assumption

(29) $\quad u_t \sim N(0,\sigma^2)$

for all t.

The coefficient α_{it} is the fraction of the total amount licensed in quarter t-i which results in housing starts in quarter t. This fraction depends upon the distribution of issuing dates of permits over the period and the reactions of constructors. If both remain constant over time α_{it} is a constant too. We will take this as our initial assumption.

The Data

Time series for V_t and B_t for the period 1971.1-1985.4 are portrayed in Figure 1.

Figure 1. *Permit Issued and Construction Starts Dwellings, The Netherlands 1971.1-1985.4*

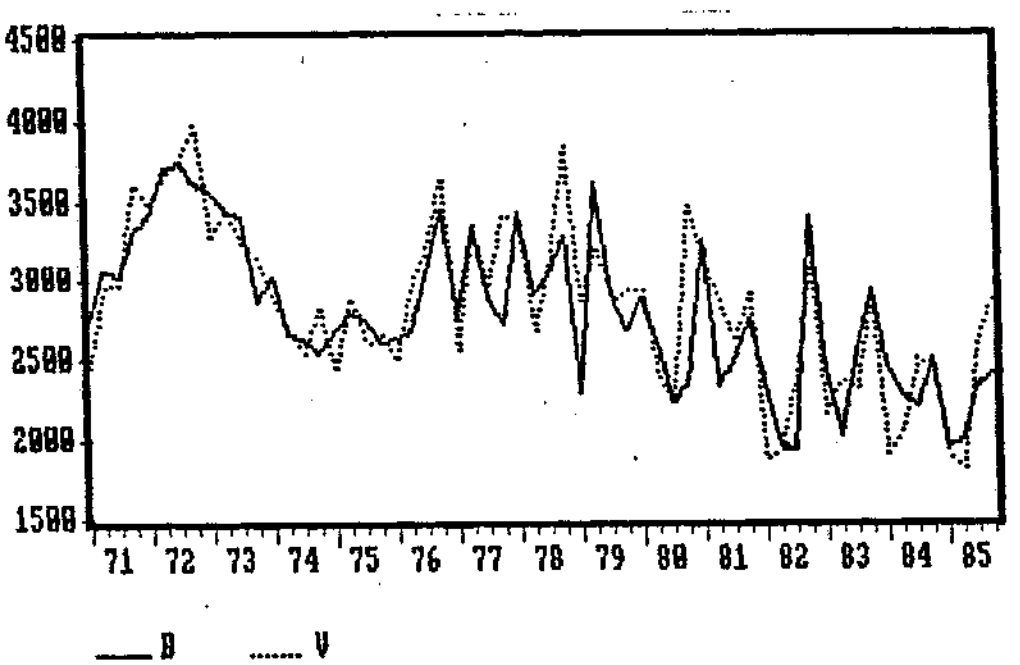


Figure 1 shows the relative smoothness of the series B_t in the first five years compared with V_t and also compared with B_t in later years. These later years reflect a slightly downward trend for both series. Remarkable is the peak in housing starts in 1979.2 and the peaks in permits issued in 1978.4 and 1980.4, not followed by a similar peak in construction starts. Remarkable also are the diverging developments since 1982.4. Explanations are partly available. Political pressure put V_t up at the end of 1978 (but also in 1977.4) to attain plan levels irrespective of their effect on B_t, and 1979.1 was apart from 1963.1 the most severe winter period of this century. The Central Bureau of Statistics (C.B.S.) changed its measurement system of V_t at the end of 1984 and claimed to have thus recovered a number of permits never announced. The reasons for the odd behaviour of recent data have not been clarified yet. Changes in measurement methods may be partly responsible.
A plot of the two series V_t and B_t for dwellings shows 1980.4 as a clear outlier.

Preliminary calculations

As indicated above we first assume both α_{0t} and α_{1t} to be constant through time but do not require α_{1t} to be equal to 1 - α_{0t}. This leads to the regression results of Table 1.

Table 1. The relation between housing starts (B_t) and permits issued (V_t), 1971.2 - 1985.4

V_t	V_{t-1}	D_1	D_2	D_3	D_4	R^2	$\hat{\sigma}$	Lik.
.670 (.07)	.314 (.07)	-	-	-	-	.65	.2530	-409.2
.75 (.09)	.24 (.09)	58.05 (74)	80.28 (55)	5.34 (55)	-142.66 (65)	.76	.2468	-406.1

Standard errors are within brackets

On the average 70% of the total amount licensed starts in the same quarter. The remainder follows the next quarter. The standard error (S.E.) of the regression $\hat{\sigma}$ will be used as a criterion below. As the mean of the series for 'housing starts' is 2.81 the regression equations of Table 1 generate forecasting errors of about 9% (measured by the coefficient of variation). The errors in the first equation show plan effects (in 77.4 stronger than in 78.4), a winter effect in both 79.1 and 79.2, the C.B.S. effect in 80.4 but also in 81.2, and recent uncertainties resulting in outliers in 82.4 (+) and 85.4 (-). The second equation with additive seasonal dummies shows that seasonal effects in V_t and B_t are positively correlated so that the immediate effect α_t is correspondingly higher. It is tempting to interpret the negative dummy D_4 as the plan effect but in dynamic regressions such conlusions are too quickly made. The plan effect is too cursory to be grasped which is reflected in the residuals of the second equation. The negative dummy is merely an artefact arising from the failure to incorporate the lag pattern in the seasonals. Below this artefact is avoided by incorporating the seasonal effect in the coefficient α_t.

Filter models

In our filter models we must specify measurement equation (1) and state space equation (18), in the Appendix indicated as equations (A2) and (A1). In the preliminary calculations above the measurement equation was specified by (28). We specify the state space equation (18) now as

$$(30) \qquad \alpha_t' = [\alpha_t \ \alpha_{t-1} \ \alpha_{t-2} \ \alpha_{t-3}]$$

and distinguish between a non seasonal model with

$$(31) \qquad D = D_1^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and a seasonal model with

$$(32) \qquad D = D_1^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Seasonality in α_t expresses the idea that housing starts reacts restrained to excessive issuing of permits e.g. at the end of the year. The variance-covariance matrix (20) is now specified as

$$(33) \qquad Q = \text{diag}\ [\theta,0,0,0]$$

with some scalar value θ. Thus we have the following modelspecifications:

	Code	D_1	θ
Deterministic α			
1) R.R. with constant α	RRC	$D_1^{(1)}$	0
2) R.R. with seasonal α	RRS	$D_1^{(2)}$	0
Stochastic α			
3) K.F. with seasonal α_t	KF	$D_1^{(2)}$	$\hat{\theta}$
4) H.F. with seasonal α_t	HF	$D_1^{(2)}$	$\hat{\theta}_t$

The models with $\theta = 0$ lead to deterministic α, which are estimated by R.R. Models with stochastic α can be handled with either KF or HF. With K.F. there is only one estimate for the whole sample period; with

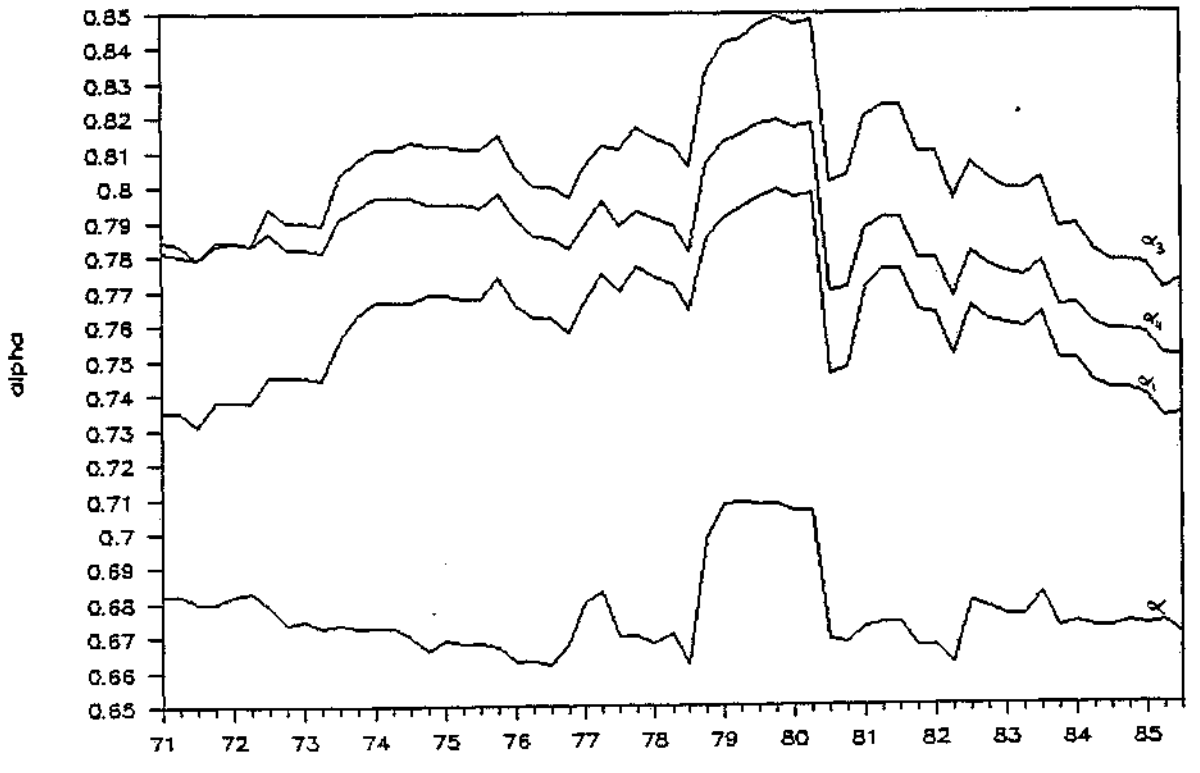
H.F. there is an estimate of θ for each quarter of the sample period. Note that in (25) θ is a vector containing more elements. We have studied cases with more elements but report only our scalar results here.

Results

RR

Figure 2 shows how α_1 gradually obtains its value of Table 1 if the sample is extended. Also the coefficients for the first, third and fourth quarter of the RRS model are drawn; the second quarter alpha shows roughly the same pattern. These α's circulate around the average value of .75 from Table 1. Note the jump in all α-series from 1979.1 to 1980.3. Note also the absence of any negative effect in the fourth quarter as in D_4 above. The reaction pattern does not differ very much among the seasons, although B_t seems to react a little bit slower in the first and a little bit quicker in the third quarter.

Figure 2. *Recursive estimates of alpha*



On the whole RRS performs somewhat better than RRC. Table 2 shows that the standard deviation of its prediction error S.E. (u_t) in the final quarter 1985.4 is .2481 versus .2530 for RRC and the likelihood values are -408.69 and -409.20 respectively. In view of the two additional degrees of freedom used up in the seasonal model the difference is not significant.

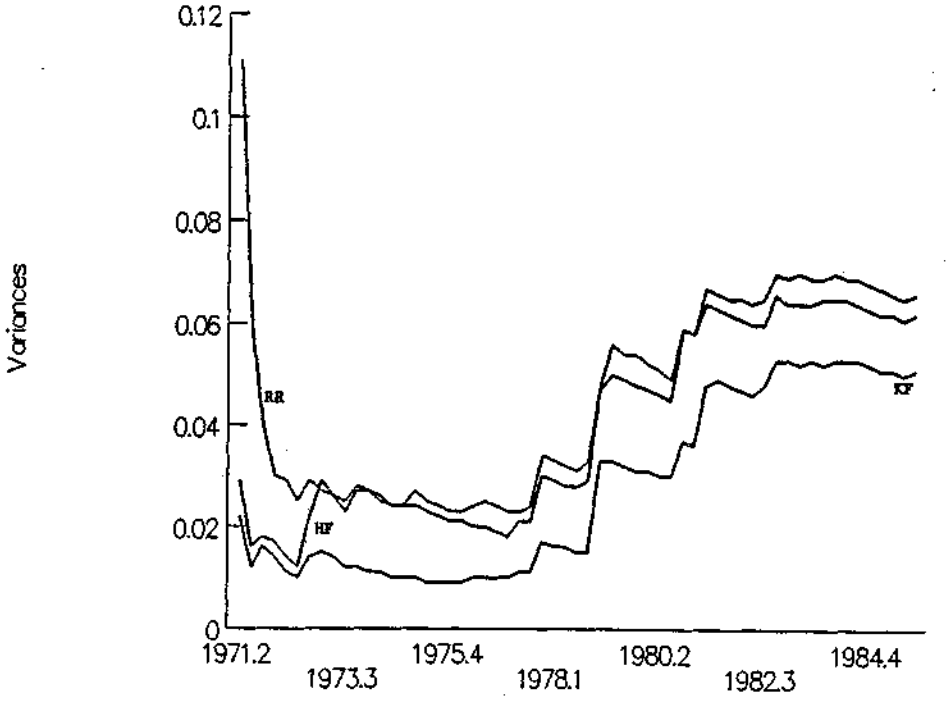
Table 2. Housing starts 1971.1 - 1985.4

Model	RRC	RRS	KF	HF
S.E.($u_{1985.4}$)	.2530	.2481	.2337	.2530
$\hat{\theta}_{1985.4}$	0	0	.011	.000
Loglik	-409.20	-408.69	-407.42	*)

*) not defined

This dispute is of little relevance as the assumption of a constant variance σ^2 over time that underlies both specifications is clearly violated. This is shown in figure 3, where the updated variance according to (16) or (17) is pictured.

Figure 3. *Measurement Variances*



The K.F is more flexible; it captures part of the increase in the residual variance in adapting the modelcoefficients α_t, see figure 4. Allowing these coefficients to vary with variance measured at .011 rather than zero spreads the uncertainty over two sources and reduces the S.E of the measurement error to .2337, less than 8.5% of the average level of the series. Accordingly the fluctuations of the KF α's in Figure 4 are wilder than those of RR. The shift in 1979.1 both in these coefficients and in the residual variance are alike in both models. Again we conclude that KF is a little superior but its likelihood is not significant to justify θ to differ from zero. There are more drawbacks for the KF. Its optimal value of .011 can only be calculated afterwards and it is not clear how many observations we must select to obtain an appropriate value.

Figure 4. *Recursive estimates of alphas*

Fig. 4a : Recursive estimates of alpha1

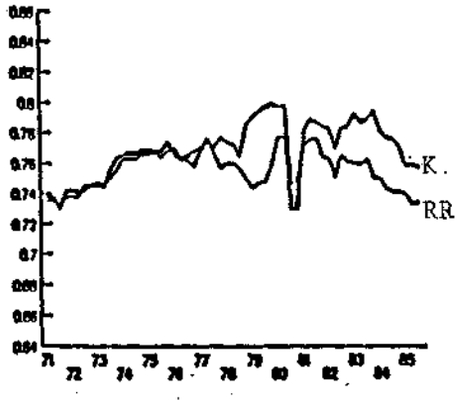


Fig. 4b : Recursive estimates of alpha2

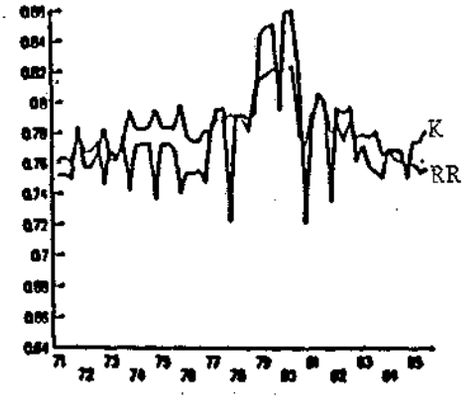


Fig. 4c : Recursive estimates of alpha3

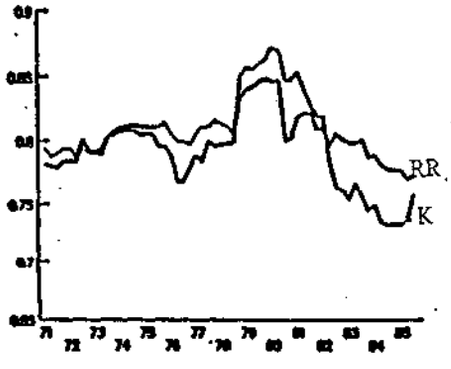
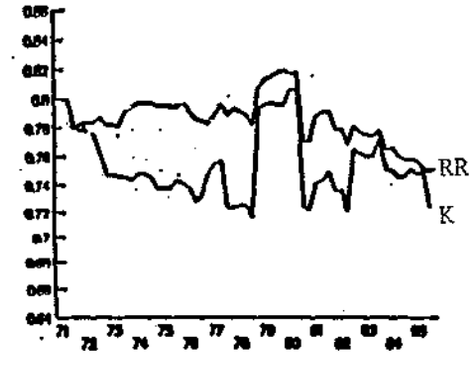


Fig. 4d : Recursive estimates of alpha4



How does the HF perform in this little race? First we observe that it cannot be compared with the other models by likelihood values as HF does not use the same likelihood function. It is too complicated to be of value here and moreover it is not clear how both likelihood functions must be compared. As both KF and HF only differ with respect to it is appropriate to cast a comparison in terms of θ.
The $\hat{\theta}_t$ for the HF are given in Table 3.

Table 3B

Estimated ratio's of variance of model parameters to measurement variance (Values of $\hat{\theta}_t$). 1971.1 - 1985.4

	Quarter			
year	I	II	III	IV
1971		.010	.010	.010
1972	.010	.010	.000	.008
1973	.000	.000	.000	.000
1974	.015	.011	.000	.008
1975	.020	.000	.000	.004
1976	.000	.002	.000	.000
1977	.000	.000	.000	.000
1978	.000	.007	.011	.009
1979	.000	.000	.000	.000
1980	.000	.000	.000	.000
1981	.000	.000	.000	.000
1982	.000	.000	.006	.005
1983	.005	.000	.000	.000
1984	.000	.000	.000	.000
1985	.000	.000	.000	.000

It is remarkable that $\hat{\theta}_t$ hardly deviates from zero, which brings us back to the RR model. We do not need to mention either the time path of the α's or that of $\hat{\sigma}^2$ of the HF model. Apart from its initialization which takes about 8 quarters the results are those of RR, see e.g. fig.3. The value of $\hat{\sigma}_t^2$ in the final quarter can even hardly be distinguished from the minimal level .2530 of RRC, see table 2. Does this repudiate HF? We have seen that all differences are insignificant and θ in KF can only be computed afterwards. There is another

advantage in using HF. The $\hat{\theta}_t$ that differ from 0 (the end of 1978 and the end of 1982) provide another source of information. Reactions of B_t to V_t in those periods were different which was not caused by measurement errors. In spite of the increased inaccuracy with which the reactions are measured by the CBS according to fig.3 these changes in the reaction patterns could be detected with the hyperfilter. This information could not be provided with the KF.

6. Conclusions

Our hyperfilter - a variant of Ljung and Söderström's recursive prediction error method - appears to be a satisfactory alternative to a Kalman Filter estimation with numerically determined variances. It avoids the use of information which is not available at time t, it is computationally more efficient and the fluctuations in the estimates provide information about the constancy of the variances. In our example of forecasting starts from permits issued in the Netherlands, RR appeared to be a sufficiently flexible approach; non-zero values of $\hat{\theta}_t$ during the sample period were of a transient nature. Hence assuming non-zero θ is clearly unnecessary in this application.

References

Fletcher, R. "Practical Methods of Optimization", John Wiley & Sons, New York, 1980.

Goldfeld, S.M., R.E. Quandt & H.F.Trotter "Maximization by Quadratic Hill-Climbing", Econometrica, vol.34, 1966, pp. 541-551.

Harvey, A.C., "Time Series Models", Philip Allan, Oxford, 1981.

Hendry, D.F. & J.-F. Richard, "On the formulation of empirical models in dynamic econometrics", Journal of Econometrics, vol.20, 1982, pp.3-33.

Kalman R.E., "A new approach to linear filtering and prediction problems", Transactions ASME Journal of Basic Engineering, vol.82, 1960, pp.35-45.

Ljung, L. & T. Söderström, "Theory and practice of recursive identification", MIT Press, Massachusetts, 1983.

Appendix

I. The model and the updating procedure

Our model is, see (18),(21),(1) and (7) :

(A1) $\alpha_t = D\,\alpha_{t-1} + \epsilon_t \qquad \epsilon_t \sim N(0,\sigma^2 Q)$

(A2) $y_t = x_t'\,\alpha_t + u_t \qquad u_t \sim N(0,\sigma^2)$

The forecasting procedure of the Kalman filter consists of updating,predicting and evaluating as follows:

(A3) $P_t = D\,P_{t-1}\,D' - M_{t-1}\,M'_{t-1}\,s_t + \hat{Q}_{t-1}$

(A4) $s_t = 1 + x'_t\,P_t\,x_t$

(A5) $M_t = D\,P_t\,x_t/s_t$

(A6) $\hat{y}_t = x'_t\,\hat{\alpha}_t$

(A7) $e_t = y_t - \hat{y}_t$

(A8) $\hat{\alpha}_{t+1} = D\alpha_t + M_t\,e_t$

Equations (A3) - (A8) coincide with respectively (24), (10), (23), (5), (6) and (22). With all information known in period t-1 equation (A3) can be elaborated. Equations (A4)-(A6) require knowledge of the exogenous values for period t, i.c. x_t. If the dependent variable is also known for period t equations (A7) and (A8) can be handled. P_t is called the standardized variance matrix of α_t ($\sigma^2 = 1$) and M_t the Kalman gain.

II. The optimal gradient β_t.

The optimal prediction occurs if $f_t() = e_t^2/s_t$ is minimized. This is obtained, see (27) , if

(A9) $\quad \hat{\theta}_t = \hat{\theta}_{t-1} - \gamma B_t^{-1}\beta_t$

with $\quad \gamma = 1/t$ and

(A10) $\quad \beta_t = \delta f_t/\delta\theta = e_t/s_t \; [2\,(\delta e_t/\delta\theta) - e_t/s_t\,(\delta s_t/\delta\theta)]$

B_t in (A9) is approximated as

(A11) $\quad B_t = \beta_t'\beta_t + \tau\, I$

where τ is chosen such that B_t is positive definite, see Goldfeld, Quandt and Trotter (1966). Formula (A10) is computed recursively: e_t and s_t follow from (A7) and (A4) respectively and the gradients $\delta e_t/\delta\theta$ and $\delta s_t/\delta\theta$ are derived below.

A. We first compute $\delta s_t/\delta\theta$. This is facilitated by the following definitions:

(A12) $\quad \pi(i)_t = \delta P_t/\delta\theta_i$

(A13) $\quad \sigma(i)_t = \delta s_t/\delta\theta_i$

(A14) $\quad \mu(i)_t = \delta M_t/\delta\theta_i$

where θ_i is the i^{th} component of θ and $\pi(i)_t$, $\sigma(i)_t$ and $\mu(i)_t$ are resp. a K * K matrix, a scalar and a K * 1 vector. Differentiating (A3),(A4) and (A5) gives in terms of these new definitions

(A15) $\quad \pi(i)_t = D\,\pi(i)_{t-1}\,D' + \delta\hat{Q}_{t-1}/\delta\theta_i - \sigma(i)_{t-1}M_{t-1}M'_{t-1} -$

$$- s_{t-1}\mu(i)_{t-1}M'_{t-1} - s_{t-1}M_{t-1}\mu(i)'_t$$

(A16) $\sigma(i)_t = x'_t \; \pi(i)_t \; x_t$

(A17) $\mu(i)_t = D \; \pi(i)_t \; x_t/s_t - M_t \; \sigma(i)_t/s_t$

Given all variables for period t-1 the value for $\pi(i)_t$ follows from (A15) and that of $\sigma(i)_t$ from (A16).

B. Now we turn to $\delta e_t/\delta\theta$.

From A(7) and (A6) we have

(A18) $\delta e_t/\delta\theta = - x'_t \; (\delta\hat{\alpha}_t/\delta\theta)$

From (A8) we have

(A19) $\delta\hat{\alpha}_{t+1}/\delta\theta = D \; (\delta\hat{\alpha}_t/\delta\theta) + M_t \; (\delta e_t/\delta\theta) + (\delta M_t/\delta\theta) \; e_t$

which after substitution of (A18) simplifies into

(A20) $\delta\hat{\alpha}_{t+1}/\delta\theta = [D - M_t \; x'_t] \; (\delta\hat{\alpha}_t/\delta\theta) + (\delta M_t/\delta\theta) \; e_t$

Given all variables for period t-1 the value of $\delta\hat{\alpha}_t/\delta\theta$ follows from (A20) and $\delta e_t/\delta\theta$ from (A18). Our procedure differs from that of Ljung and Soederstroem in two respects. First LS use $\delta s_t/\delta\theta$ in elaborating $\pi(i)_t$ like we did in (A15) and (A17), but they ignore this term in (A10). Secondly they use instead of (A11) $B_t = 1/t \; [\beta_t'\beta_t + (t-1) \; B_{t-1}]$. Our procedure (A11) defeated that of LS in speed of convergence at least in our application.