PREDICTION THRESHOLD FILTERING

Robert M. KUNST

Forschungsbericht/
Research Memorandum Nr. 254
July 1989

CONTENTS

ABSTRACT

This paper is concerned with a modification of the robust filtering algorithm of Kleiner, Martin, and Thomson (1979). This improved version PTF (for prediction threshold filtering) is able to discriminate between outliers of the innovations and of the additive type (IO and AO). Since the effects of the different outliers species on parameter estimation and prediction are different, suggesting data correction for AO and no correction for IO, this discrimination could prove important in practice.

This paper provides some evidence on the qualities of PTF. First, the power of the filter is investigated using numerical formulae. This approach faces a limit when the assumption of known true parameters is dropped. Next, Monte Carlo simulation is used to reveal patterns based on the temporal dependence of series and decisions which are excluded from the numerical tables. Last, some theoretical lemmata explain some of the observed features and extend the results given in the first part of the paper.

1. Introduction [1]

About a decade ago, Kleiner, Martin, and Thomson (1979) presented an innovative method of robust time series estimation incorporating simultaneous data smoothing. Data are corrected iteratively by

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + c\sigma\Phi((x_t - \sum_{i=1}^{p} a_i y_{t-i})/c\sigma)$$

This filter formula, which is understood to be run consecutively in the direction of time, needs as input: the data x_t; preliminary estimates for the autoregressive coefficients a_i, the residual standard deviation σ, and the autoregressive order p; specification of the constant c and the function Φ. The generated y_t may be used to obtain new a_i and σ estimates, and filtering and estimating may be repeated until convergence.

Some of the problems involved with this procedure are merely technical, e.g. how to choose Φ and c. Martin (1979) uses redescending Hampel-like Φ, whereas Kunst (1987) favors the Huber (1964) function. Although small c seem to work well in Monte Carlo, as far as parameter estimation is concerned, values around 2 are recommended to avoid over-correction and deteriorating forecasting performance. Estimation of the a_i and maximal lag p could be done by OLS or Yule-Walker and AIC or robustified AIC, respectively. Even if standard AIC is used, the selected order rarely changes between iterations.

Another bulk of problems arises from the theoretical implications of the filter. It cannot be proved to converge, although it usually does unless c is reduced below one. However, it never converges to an efficient limit. If bounded Φ is used, residuals

---

1 A preliminary version of this paper was presented at the European Meeting of the Econometric Society at Copenhagen in September 1987.

will be bounded which is opposed to the assumption of normality. Any estimation procedure assuming infinite supports of the error distribution, like OLS, Yule-Walker, or Huber's M, necessarily is inefficient with bounded errors. There are three obvious solutions: first, the use of regression techniques designed for finite supports; second, the use of unbounded Φ; third, providing for an outlet to make the residuals unbounded retaining standard Φ and standard regression techniques. This paper is concerned with the third solution, since the former two entail extremely non-robust behavior inconsistent with the spirit of the overall procedure.

The most important problems obtain when using the filter for forecasting actual data. The parameter estimates are known to be quite reliable (Kunst 1987, 1989, Guzy 1986), but erroneously "corrected" data could be responsible for the mediocre forecasting performance (Tsay 1986, Kunst 1987). Section 2 outlines possible explanations for these spurious corrections and section 3 presents a method designed for reducing the possibility of their occurrence. We shall call it PTF (for prediction threshold filter).

Under the assumption of known parameters, Section 4 presents some theoretical results on the power of the PTF with respect to decisions about whether an observation at hand is an additive outlier or not. These results are extended to loose specifications approximating the original method by Kleiner et al. (1979) in Section 5. Section 6 gives a small illustrative example. To evaluate the usefulness of the theoretical formulae in the presence of estimated parameters, some Monte Carlo experiments are reported in Section 7. Again under the assumption of known parameters, the influence of filter decisions at time t on filter decisions at t+1 is illustrated in Section 8. Section 9 concludes the paper.

2. Innovation outliers

The filter by Kleiner, Martin, Thomson (1979, to be abbreviated KMT in what follows) was designed to cope with the effects of additive outliers (AO). An AO is a disturbance of the observed process which leaves the data-generating mechanism unaffected. A good example is the typing error. Subsequent data will not depend on the typed nonsense but on the actual realization. Fox (1972) already was conscious of the alternative phenomenon of innovations outliers (IO). Here, the disturbed observation feeds back into the data-generating process. This species of outliers is likely to be more common in economics where external shocks usually cause some kind of persistence in the response.

It is never known a priori whether data contain AO or IO. Those are impossible to distinguish at the time point of the aberrant observations, but the patterns of succeeding observations differ. After AO, the process logically continues the pre-outlier regime, leaving the outlier isolated in an eventual plot diagram. After IO, the process remains disturbed for some time or builds up a new regime starting with the outlier, depending on the memory of the process.

Tsay (1986) presented a statistic to discriminate between the two kinds of outliers. To calculate the statistic for an observation at time point t, all future observations t+1,t+2,...,T are needed. Within this paper, a simpler concept will be used relying on n-step forecasting. In the AO case, a two-step forecast at t-1 will predict the observation at t+1 quite well, whereas the one-step forecast at t, using the isolated aberrance, will fail. In the IO case, as in the case of undisturbed data, on average the two-step forecast will be less exact than the one-step forecast. Since IO and non-outliers ought to be treated alike (see below), this looks like a good basic idea. Of course, it will not be valid if data are allowed to incorporate "patches" of consecutive outliers.

Whereas, in the case of AO, what should be done ideally to gain access to process parameters is obvious - remove the added disturbance - this is not so clear in the IO case. Standard least-squares is known to perform very well with IO processes obeying regular laws (Hannan and Kanter, 1977). The same goes for LAD estimates (An and Chen, 1982) and the results should hold equally for M estimates if additional regularity assumptions are imposed. However, Monte Carlo (Denby and Martin, 1979) showed that OLS performance is satisfactory but not extremely good in finite samples. The Tsay procedure seems to suggest smoothing out the time series starting with the IO, which is out of the question in a forecasting framework. Treating the IO like AO as in the KMT filter algorithm can prove disastrous, especially if the erroneous corrections occur in the pre-period of the forecasting interval. Prediction then will rely on slightly wrong parameters and on harmfully wrong observations. In summary, IO should not be corrected, however, if possible, information about their occurrence used for re-designing the weighting function in estimation.

Of course, there are different possible reasons for a breakdown of the filter procedure. MA components should be a minor trouble [1]. Instationarities like trends or changes in coefficients - either slow changes or abrupt structural breaks - may impair performance. Trends or cycles, however, should be removed before filtering or, alternatively, included in the filter formula in a straightforward fashion. Changing coefficients have been treated successfully by Kalmanizing the filter (Martin 1979). As with all Kalmanized methods, it is unknown how much is gained in practice in return for abandoning the search for a constant structure.

1 The filter easily generalizes to MA, ARMA or similar models as it only requires specification of a forecast-generating mechanism and evaluation of forecasting errors.

3. The procedure

Throughout the paper Φ is assumed to be Huber's function (Huber, 1964) which is the identity function for all x with modulus less than one and equal to the sign function for larger absolute values. There is not much to be changed in argument if the latter condition is modified. However, a $\Phi(x)=x$ area in the center should be retained. It represents a window through which well predicted data pass unchanged.

If the forecasting error exceeds $c\sigma$, the KMT filter replaces the observation by its one-step prediction plus or minus $c\sigma$. The filter [2] presented here, keeping in mind the possibility of IO, only adjusts if the one-step error for x_{t+1} using the original x_t is greater than τ times the two-step error of a forecast in x_{t-1}, i.e.:

$$y_t = x_t \quad \text{if} \quad |x_t - E(x_t|y_{t-1},y_{t-2},\dots)| \le c\sigma$$

$$y_t = x_t \quad \text{if} \quad |x_t - E(x_t|y_{t-1},y_{t-2},\dots)| > c\sigma \text{ and}$$
$$|x_{t+1}-E(x_{t+1}|x_t,y_{t-1},\dots)| \le \tau|x_{t+1}-E(x_{t+1}|y_{t-1},\dots)|$$

$$y_t = E(x_t|y_{t-1},y_{t-2},\dots)$$
$$\text{if} \quad |x_t - E(x_t|y_{t-1},y_{t-2},\dots)| > c\sigma \text{ and}$$
$$|x_{t+1}-E(x_{t+1}|x_t,y_{t-1},\dots)| > \tau|x_{t+1}-E(x_{t+1}|y_{t-1},\dots)|$$

With normal data, one-step forecasts will tend to be better than two-step forecasts. However, especially as long as expectations have to be estimated (see e.g. Chong and Hendry 1986), the probability of a sign reversal in the inequality will be high, so the constant τ was inserted which is intended to be greater than one. The theoretical results in the next sections assume the

2 In the terminology of Martin and Yohai (1986) the formulae do not describe a filter since one future observation is needed for adjusting the present one. Thus, the PTF may be viewed as the first approximation of a "smoother" by a "filter".

expectation (i.e. the autoregressive parameters) as known. Unknown parameters will be treated via Monte Carlo in Section 7.

c and τ represent the tuning parameters set by the user. The following borderline cases should be noted:

$c=0$ All observations are submitted to the secondary decision. The original filter is not used.

$c=\infty$ No observations are changed. No decision is used.

$\tau=\infty$ No observations are changed. The original filter is used senselessly.

$\tau=1$ True absolute errors are compared. Many observations are changed.

$\tau=0$ All observations are changed in the second decision. Only the original filter is used.

In the following, this filter will be labeled PTF for prediction threshold filter because it uses decision thresholds put up by predictions.

4. Power of the filter

The PTF is designed to fulfil the following tasks:

- "good" data should not be corrected
- AO should be corrected (if possible, to the hidden original observation)
- IO should not be corrected

Whether the corrected values coincide with the original series, heavily depends on correct process parameters. If parameters are assumed to be known, this requirement can be set aside for the moment. Apart from knowledge of the parameters, it will be assumed that the process obeys a first-order autoregressive scheme with positive AR parameter a less than one. Extension of the results to higher-order AR will be straightforward but will complicate the calculations [3]. In this section, τ will be restricted to be greater than one. For some extensions to the case $0 < \tau < 1$, see section 5.

LEMMA 1: The probability for the type I error of correcting good data at t ($2 \le t \le T-1$) is

$$2 \int_{c\sigma}^{\infty} f(x) \int_{ax\tau/(\tau+1)}^{ax\tau/(\tau-1)} f(t)\,dt\,dx$$

where f(.) denotes the symmetric p.d.f. of the errors.

Proof: Type I errors occur for $|e_t| > c\sigma$ which determines the outer integral. Restrictions for e_{t+1} determine the inner integral.▪ [4]

If f is defined as the normal density, a borderline case may be calculated by an arithmetic expression:

---

3 Most results of this section hold for higher-order processes, if a denotes the first-order coefficient, and for negative a, if a is replaced by its modulus.

4 This symbol ▪ is used in the following to denote the end of a proof.

LEMMA 2: The probability for a type I error in t ($2 \le t \le T-1$) for c=0 is

$$\frac{2}{\pi}(\arctan(a\tau/(\tau-1))-\arctan(a\tau/(\tau+1)))$$

Proof: For small *a* the p.d.f. under the inner integral is replaced by its Taylor expansion. Integration gives the Taylor expansion of the arctan function. Since the function is holomorphic, its Taylor expansion is unique for all *a*. ■

For different specifications of f(.) and c, the integral of Lemma 1 has to be evaluated numerically [5]. Within the limits of this paper, this has only been done for the normal law and selected values of c, *a*, and τ to give an impression of the filter's performance. Results may be read from Table 1.

In summary, the size of the test strictly decreases with c or τ, but much faster with c. The size response to τ is almost linear. Since large τ put up problems with test power (see below), (c,τ)-combinations of (1.2,2.2) or (2,2) look like good suggestions. Response to *a* is not monotonous. Whereas for c less than 1.5, the procedure works best for loosely correlated data, high-memory processes prefer large c.

5 All numerical integrations have been performed with the Patterson algorithm of the NAG library software package.

TABLE 1: type I errors for undisturbed data

a=0.2

	0	0.5	1.0	1.5	2.0	2.5	3.0
1.0	.468	.281	.139	.057	.018	.005	.001
1.2	.244	.205	.122	.054	.018	.005	.001
1.5	.134	.116	.076	.037	.014	.004	.001
2.0	.079	.069	.046	.024	.009	.003	.001
3.0	.045	.040	.027	.014	.006	.002	.000
5.0	.025	.022	.015	.008	.003	.001	.000
10.0	.012	.011	.007	.004	.002	.001	.000

a=0.5

	0	0.5	1.0	1.5	2.0	2.5	3.0
1.0	.422	.240	.112	.042	.013	.003	.001
1.2	.313	.228	.108	.040	.012	.003	.001
1.5	.220	.178	.096	.037	.011	.003	.000
2.0	.148	.123	.072	.030	.009	.002	.000
3.0	.091	.076	.047	.020	.007	.002	.000
5.0	.052	.044	.027	.012	.004	.001	.000
10.0	.026	.022	.013	.006	.002	.000	.000

a=0.9

	0	0.5	1.0	1.5	2.0	2.5	3.0
1.0	.365	.191	.080	.026	.007	.001	.000
1.2	.297	.181	.074	.023	.006	.001	.000
1.5	.230	.161	.067	.020	.005	.001	.000
2.0	.167	.125	.057	.017	.004	.001	.000
3.0	.108	.084	.041	.013	.003	.000	.000
5.0	.064	.051	.025	.008	.002	.000	.000
10.0	.032	.025	.013	.004	.001	.000	.000

Specifications, which are equivalent with respect to type I errors, will differ with respect to the type II error of ignoring additive outliers. The probability of such an error depends on the size of the outlier. It would be possible to calculate the expectation of this event but this provides no help in practice, because neglecting large outliers is much more harmful to consequent results. The following lemma will again be the basis for numerical integration. Let F(.) denote the c.d.f. corresponding to the p.d.f. f(.).

LEMMA 3: The probability for the type II error of overlooking an outlier of size S at t ($2 \le t \le T-1$) is

$$F(-S+c\sigma)-F(-S-c\sigma)+$$

$$+\int_{-S+c\sigma}^{\infty} f(x)\int_{g_2(x)}^{\infty} f(t)\,dt\,dx+\int_{g_2(x)}^{-S-c\sigma} f(x)\int_{-\infty}^{g_2(x)} f(t)\,dt\,dx+\int_{-\infty}^{-S-c\sigma} f(x)\int_{-\infty}^{\infty} f(t)\,dt\,dx+\int_{g_1(x)}^{\infty} f(x)\int_{-S+c\sigma}^{g_1(x)} f(t)\,dt\,dx$$

with $g_1(x)=(-a\tau x-aS)/(\tau-1)$

$g_2(x)=(-a\tau x+aS)/(\tau+1)$

Proof: For $|e_t+S|<c\sigma$ the outlier is dropped by the first step decision. This gives the term $F(S+c\sigma)-F(S-c\sigma)$ if f is assumed symmetric.
For $|e_t+S|>c\sigma$ a type II error occurs if the outlier is overlooked by the second stage decision. The inequality restrictions for both events give the corners of the integration intervals like in Lemma 1. ∎

If the probability law is known, the distribution function F and the integrals may be evaluated numerically. Obviously, the F difference will be large only for small S. For large S, only the first integral matters, provided a standard unimodal density is assumed. Only in this integral the area around zero containing the principal mass of the density is covered twice by integration.

$S\to\infty$ takes the type II error to zero. Perhaps more interesting is what happens if S is finite but τ is taken to infinity. In this case, the probability increases, reaching an extremum for S=0 where everything sums up to one.

LEMMA 4: If $\tau \to \infty$ and the density is symmetric, the type II error probability of Lemma 3 converges to

$$F(S+c\sigma)-F(S-c\sigma)+2F(-c\sigma)$$

which is one for S=0

For c=0, this probability converges to one all the same but this case is of no interest as then no test at all is performed. In contrast, the case of Lemma 4 is important insofar as numerically infinite τ are reached rapidly. Again, numerical evaluations have been done for normal errors and several different values for S and τ while a was fixed at 0.5. The results are shown in Table 2. Only three different c designs (1.5; 2; 2.5) have been used but it should be easy for the interested reader to extend the table to his individual aims.

Contrary to AO cases, IO should not be corrected by the filter. As the original IO model is stochastic, information about the error distribution could and should be used for estimation during iterations. For the next result, this stochastic nature of the outliers is ignored, such as for the preceding lemmata.

LEMMA 5: The probability for a type I error with respect to spuriously detecting an IO of size S for an AO is

$$\int_{-\infty}^{-S-c\sigma} f(x) \int_{-h_2(x)}^{-h_1(x)} f(t)\,dt\,dx + \int_{-S+c\sigma}^{\infty} f(x) \int_{-h_1(x)}^{-h_2(x)} f(t)\,dt\,dx$$

which, for symmetric density f, is equivalent to

$$\int_{S+c\sigma}^{\infty} f(x) \int_{h_2(x)}^{h_1(x)} f(t)\,dt\,dx + \int_{-S+c\sigma}^{\infty} f(x) \int_{h_2(x)}^{h_1(x)} f(t)\,dt\,dx$$

where $h_1(x)=a(S+x)\tau/(\tau-1)$ and $h_2(x)=a(S+x)\tau/(\tau+1)$

Proof: The conditions for the corresponding error are $|e_t+S|>c\sigma$ and $|e_{t+1}|>\tau|e_{t+1}+ae_t+aS|$. After separating positive and negative cases, calculation of the integration bounds is straightforward. ■

TABLE 2: type II errors for AO outliers of size S

c=1.5

τ \ S =	1	2	3	4	5	6	7	8	9	10
1.0	.855	.555	.293	.172	.113	.073	.045	.026	.015	.008
1.2	.855	.570	.315	.196	.137	.094	.062	.040	.024	.014
1.5	.865	.601	.359	.237	.172	.126	.090	.063	.042	.028
2	.892	.682	.476	.343	.253	.188	.141	.105	.077	.057
3	.928	.789	.645	.532	.438	.356	.287	.229	.182	.144
4	.946	.843	.734	.645	.565	.491	.424	.362	.308	.259
5	.957	.875	.788	.715	.648	.585	.524	.467	.414	.365

c=2.0

τ \ S =	1	2	3	4	5	6	7	8	9	10
1.0	.930	.692	.371	.187	.114	.073	.045	.026	.015	.008
1.2	.932	.704	.392	.212	.138	.094	.062	.040	.024	.014
1.5	.937	.721	.424	.249	.173	.126	.090	.063	.042	.028
2	.946	.765	.516	.349	.253	.188	.141	.105	.078	.057
3	.962	.838	.666	.535	.438	.356	.287	.229	.182	.144
4	.971	.878	.749	.647	.565	.491	.424	.362	.308	.259
5	.977	.903	.799	.717	.648	.585	.524	.467	.414	.365

c=2.5

τ \ S =	1	2	3	4	5	6	7	8	9	10
1.0	.973	.820	.495	.228	.119	.073	.045	.026	.015	.008
1.2	.974	.828	.514	.252	.143	.094	.062	.040	.024	.014
1.5	.976	.839	.538	.285	.176	.126	.090	.063	.042	.028
2	.979	.859	.598	.371	.255	.188	.141	.105	.078	.057
3	.985	.898	.714	.546	.439	.356	.287	.229	.182	.144
4	.988	.922	.783	.655	.566	.491	.424	.362	.308	.259
5	.991	.938	.826	.723	.649	.585	.524	.467	.414	.365

Note that, for small S, the probability of Lemma 5 approaches that of Lemma 1. Even for moderate S, only the second integral matters. It is obvious that large τ entail narrower intervals for the inner integrals and help to diminish the IO errors. Table 3 gives numerical results for the Gaussian case. Note that the maximum of the error frequency relative to S is non-trivial. It depends on c and is usually attained at around three to four standard deviations. [6]

All three tables should be compared in order to understand the trade-off of the tuning constants τ and c. Suggestions for (τ,c) only make sense for specified costs of the three error types. For parameter estimation, low τ and c could be good solutions. For forecasting relying on corrected data, τ should be increased, even if many smaller AO are bypassed. Keeping c low could diminish the basic type II AO errors. In Section 7, we shall subject the procedure to Monte Carlo simulation and thus we shall see whether these observations still hold in the presence of estimated parameters.

6 It is seen easily that the probabilities do not depend on σ.

TABLE 3: type I errors for IO outliers of size S

c=1.5

τ \ S =	1	2	3	4	5	6	7	8	9	10
1.0	.093	.186	.207	.164	.113	.073	.045	.026	.015	.008
1.2	.088	.173	.189	.144	.094	.057	.033	.018	.009	.004
1.5	.081	.159	.169	.123	.075	.042	.022	.011	.005	.002
2	.067	.133	.141	.099	.057	.029	.013	.006	.002	.001
3	.047	.094	.102	.071	.038	.017	.007	.002	.001	.000
4	.035	.072	.078	.055	.029	.012	.005	.001	.000	.000
5	.028	.058	.064	.044	.023	.010	.003	.001	.000	.000

c=2.0

τ \ S =	1	2	3	4	5	6	7	8	9	10
1.0	.042	.122	.177	.158	.112	.073	.045	.026	.015	.008
1.2	.039	.113	.160	.139	.094	.057	.033	.018	.009	.004
1.5	.036	.102	.142	.118	.075	.042	.022	.011	.005	.002
2	.031	.087	.119	.095	.056	.029	.013	.006	.002	.001
3	.022	.064	.087	.068	.038	.017	.007	.002	.001	.000
4	.017	.049	.067	.053	.029	.012	.005	.001	.000	.000
5	.014	.040	.055	.043	.023	.010	.003	.001	.000	.000

c=2.5

τ \ S =	1	2	3	4	5	6	7	8	9	10
1.0	.016	.067	.134	.146	.111	.073	.045	.026	.015	.008
1.2	.014	.061	.120	.127	.092	.057	.033	.018	.009	.004
1.5	.013	.054	.105	.107	.074	.042	.022	.011	.005	.002
2	.011	.046	.087	.086	.055	.029	.013	.006	.002	.001
3	.008	.034	.064	.062	.037	.017	.007	.002	.001	.000
4	.006	.027	.050	.048	.028	.012	.005	.001	.000	.000
5	.005	.022	.041	.039	.023	.010	.003	.001	.000	.000

5. The case $\tau < 1$

If the parameter τ is decreased below one, the PTF approaches the original KMT filter. The following result shows that reaction somehow accelerates in this area.

LEMMA 6: The probability for a type I error with respect to spurious correction of undisturbed data for $\tau<1$ is

$$2\int_{c\sigma}^{\infty} f(x)\left[\int_{-\infty}^{-ax\tau/(\tau+1)} f(t)dt + \int_{ax\tau/(1-\tau)}^{\infty} f(t)dt\right]dx$$

Proof: completely analogous to Lemma 1. ∎

Note that the inner integrals are unbounded. This property is reflected by turning points of the fractile curves in a (c,τ) diagram at $(.,1.0)$. Further evidence evolves from the first derivative of the size function $Sf(c,\tau;a,\sigma,f)$ with respect to τ, i.e.:

$$\frac{d\, Sf(.,\tau;.)}{d\tau} = -2\frac{a}{(\tau-1)^2}\int_{c\sigma}^{\infty} xf(x)f(a\tau x/(\tau-1))dx -$$

$$-2\frac{a}{(\tau+1)^2}\int_{c\sigma}^{\infty} xf(x)f(-a\tau x/(\tau+1))dx$$

for any positive $\tau \neq 1$. For $\tau \to 1$, the first expression tends towards infinity. For $\tau<1$, the second derivative is negative and for $\tau>1$ it is positive.

Analogously, the counterparts to Lemmata 3 and 5 may be evaluated in the $\tau<1$ case:

LEMMA 7: The probability for a type II error with respect to overlooking an AO of size S for $\tau<1$ is

$$F(-S+c\sigma)-F(-S-c\sigma)+ + \int_{c\sigma-S}^{\infty} f(x) \int_{g_3(x)}^{g_4(x)} f(t)dtdx + \int_{-\infty}^{-c\sigma-S} f(x) \int_{g_4(x)}^{g_3(x)} f(t)dtdx$$

with $g_3(x)=a(S-\tau x)/(1+\tau)$

$g_4(x)=a(S+\tau x)/(1-\tau)$

The integrals of Lemma 7 converge to the first and second ones of Lemma 3 when τ reaches one while the remaining two terms of Lemma 3 disappear for the present case. Note that now all the inner integrals are bounded.

LEMMA 8: The probability for a type I error with respect to spurious correction of IO of size S for $\tau < 1$ is

$$\int_{-S+c\sigma}^{\infty} f(x) \int_{h_3(x)}^{\infty} f(t)dtdx + \int_{-S+c\sigma}^{\infty} f(x) \int_{-\infty}^{-h_2(x)} f(t)dtdx + \int_{-\infty}^{-S-c\sigma} f(x) \int_{-\infty}^{h_3(x)} f(t)dtdx + \int_{-\infty}^{-S-c\sigma} f(x) \int_{-h_2(x)}^{\infty} f(t)dtdx$$

with $h_2(x) = a(S+x)\tau/(1+\tau)$ and $h_3(x) = a(S+x)\tau/(1-\tau)$

Here, the second and fourth terms coincide with the integrals of Lemma 5 and the remaining terms converge to zero for $\tau\to1$. From Lemmata 7 and 8, it can be deduced that PTF performance changes qualitatively and not only quantitatively if $\tau<1$.

The overall conclusion seems to be that such a τ should not be used. With regard to IO errors, PTF power quickly deteriorates. On the other hand, if IO errors are considered not to be present at all, there is no reason to deviate from the KMT design with $\tau=0$.

6. An illustrative example

In order to enable a better understanding of what the PTF really does, the outcomes of an application to economic data are reported. The data are plotted in Figure 1. They represent Austrian GDP (gross domestic product) subjected to some stationarizing transformations during 1965.2-1985.4 (quarterly data).[7]

From visual inspection, several sharp peaks are noticed. These peaks mainly appear within the 1970s whereas the behaviour before and after looks calmer. These peaks are the obvious choices for naive outlier identification.

The PTF was applied to this data, starting with $(\tau,c)=(1,1)$ and gradually increasing both tuning constants in a grid with steps of 0.5 until no more outliers are found [8]. The 10 marked values are those for which the PTF with any of the specifications reports outlying behaviour and which are, consequently, corrected. (The case $(\tau,c)=(1,1)$ was omitted since it reported too many outliers.) Obviously, this set only contains three of the peaks and even ignores the sharpest aberrations. Instead, several observations are marked which nobody would have taken for outliers. It may be conceded that part of the visual difficulties stems from negative correlation.

The outliers show different persistence if the tuning constants are increased. The third from the left is the most persistent and even holds for $c=2.5$ and $\tau=2$. Other relatively persistent cases are numbers 1,4,9 (counted from the left). These are very likely to be AO aberrations. It may be interesting that the "classical" robust filter reproduces the judgment by visual inspection.

7 GDP raw data from the Austrian Institute for Economic Research (Wifo) were logged and first and fourth differences were taken.
8 The PTF was performed on the basis of Yule-Walker estimates of eighth-order AR processes.

Looking at the autoregressive parameters [9], the second-order coefficient is strongly influenced by PTF. It even declines considerably if only observation 3 (the OPEC shock) is modified and reaches -.2 for tightly tuned versions compared to the Yule-Walker estimate of -.04. Such a change in coefficients plays a role if forecasts should be generated.

TABLE 4: numbers of outliers corrected in the GDP example

τ \ c =	1	1.5	2	2.5	3
1	13	4	1	1	0
1.5	9	4	1	1	0
2	8	3	1	1	0
2.5	2	0	0	0	0
3	2	0	0	0	0
3.5	0	0	0	0	0

9 Parameter estimates are not reported in order to save space. All results are available on request from the author. The numbers of outliers in response to varying PTF design are shown in Table 4.

7. Some Monte Carlo evidence

Any conclusions about the filter's performance drawn at this stage from the formulae of sections 4 and 5 can at best be guesses since the results hold for known parameters only. It is not impossible to extend the results to the case where true and filtering parameters are different but the results are complicated and difficult to interpret, especially because they depend on y_{t-1} additionally to e_t and e_{t+1}. Therefore, this section rather concentrates on evidence from Monte Carlo simulation.

Before explaining the design of the simulation experiments and reporting the simulations results, it should be made clear why these will differ from the numerical tables 1 to 4. The possible sources are:

1. Estimation of the autoregressive parameter will be biased. Even for standard data, OLS or Yule-Walker estimation produces biased estimates in finite samples. Wrong estimates entail wrong decisions about the nature of possible outliers.

2. Even if estimates are unbiased, consequent decisions might be "biased" if non-linearities are present in the transformations.

3. Consecutive errors might play a role. This means that a wrong decision at t might be followed more frequently by another wrong decision at t+1 than a correct decision. Section 8 is devoted to some results concerning this phenomenon.

4. The scales estimate departs from its true value. Filtering could result in a downward bias of the scales due to a large amount of incorrectly modified observations. Such a bias might tighten the filter and cause, in turn, more spurious correction. As long as scales are, however, estimated via the median of the absolute errors, this effect only makes sense for

very small tuning constants c (less than .67 in the KMT suggestion) which should not be used anyway.

5. The PTF was implemented based on AR processes without an intercept. In the beginning and between iterations, the median is subtracted from the data. The deterministically simulated outliers may generate a shift in mean, especially in the cases of larger outliers.

Basically, decisions about the first few observations are impossible and any decision about the last one is arbitrary. Apart from this fact and possible lagging effects of point 3, PTF power will not depend on the position of the outlier within the sample. Thus, its position was set to t=60 rather arbitrarily for the whole of this simulation. In accordance with the practitioner's position, no specific probabilistic model will be used on the outliers.

The results of the Monte Carlo simulations are displayed in Tables 5 to 10. All of them were based on a sample size of 100 observations. It was not tried to investigate into asymptotics. First, the simulations are costly and more so if the sample size is increased. Second, the power of PTF relies on asymptotics only as far as parameter estimation is concerned whereas the AO/IO decision does not benefit from large samples. This means that asymptotics are less interesting.

The number of replications varies from 100 to 500. More than 500 random vectors would be necessary to smoothen curves like those in Figure 4 or to obtain statistics with prescribed accuracy. However, 100 is enough to illustrate certain patterns in the results.

All simulations are based on first-order autoregressive processes of the form

$$x_t = 0.5\, x_{t-1} + e_t$$

The parameter may be interpreted as a typical value or as the center of the interval [0,1] bounded by white noise and the random walk. The e_t have been generated as normal independent numbers by a NAGLIB routine. The mean of the e_t has been set to zero and the variance to 1. To get rid of starting effects, 120 numbers were generated and the first 20 numbers forgotten.

Table 5 reports the results of a 100 replications purely Gaussian experiment. c and τ were varied within the bivariate interval [1,2]x[1,2] over a grid of 0.2. The reported numbers are the respective sample means of: the preliminary Yule-Walker estimates of the autoregressive parameter; the corresponding PTF estimates; the numbers of iterations necessary to obtain convergence; the numbers of AO detected by PTF.

At first sight, tighter tuning (lower tuning constants) increases iterations as well as detected outliers. The errors frequency coincides quite well with the numbers given by Table 1 for c around 1.4 and less well for c=1.0 or c=2.0. This observation is closely related to the bias of the PTF coefficients which seems about zero along a curve connecting (c,τ) = (1.5,1.0), (1.4,1.4), and (1.3,2.0). The bias may be standardized by averaging over all the YW estimates (giving .4727) and adding the difference between the estimates. The dependence of this difference on the tuning constants is illustrated in Figure 2. The YW estimates suffer from a downward bias which is known as the Hurwicz bias whereas PTF causes an upward bias by smoothing. Along the "unbiasedness curve" in the (c,τ) plane, the bias effects cancel [10]. The position of this curve depends on the sample size and on the AR parameter. For a known design, it could be used as a guideline for selecting the tuning constants. It is possible that, in large samples, the demand for unbiasedness could lead to large tuning constants which are incompatible with the demands for AO robustness, as the Hurwicz bias asymptotically disappears. For typical economic

10 A similar effect was reported by Tuan (1984) with regard to a modified version of the original KMT filter.

TABLE 5: Type I errors in the basic specification[11]

τ \ c	1.0	1.2	1.4	1.6	1.8	2.0
1.0	.48196	.47550	.46123	.47870	.48461	.46158
	.55623	.52468	.49781	.50118	.49712	.46997
	4.51	4.25	4.17	3.74	3.45	3.30
	12.61	8.82	6.01	4.00	2.34	1.49
1.2	.47323	.47171	.48375	.47145	.47985	.45543
	.54877	.51556	.51342	.48991	.49211	.46254
	4.44	4.11	3.92	3.63	3.43	3.30
	12.02	8.08	5.26	3.21	1.93	1.38
1.4	.47995	.48283	.48118	.47163	.46798	.47571
	.54585	.52688	.50835	.48979	.47759	.48154
	4.38	4.18	3.86	3.63	3.37	3.19
	10.52	7.49	4.97	3.07	1.75	1.13
1.6	.48799	.47348	.47563	.46521	.47408	.47991
	.54731	.51205	.50046	.48606	.48506	.48789
	4.39	4.15	3.90	3.64	3.42	3.30
	9.68	6.65	4.33	3.17	1.93	1.29
1.8	.47426	.45746	.47413	.46876	.46535	.47477
	.52494	.49568	.49868	.48856	.47466	.48172
	4.26	4.03	3.81	3.59	3.35	3.28
	8.42	6.29	3.97	2.71	1.56	1.28
2.0	.47136	.46288	.48094	.45624	.46853	.46697
	.51496	.49785	.50204	.47102	.47937	.47444
	4.18	4.02	3.77	3.56	3.40	3.33
	7.35	5.80	3.98	2.42	1.68	1.26

line 1: PTF estimate
line 2: Yule-Walker estimate
line 3: number of iterations
line 4: number of outliers corrected by PTF
all lines: arithmetic means

11 100 replications using separate random numbers setups of sample size 100 for each of the 36 variants.

samples, this effect should not play a role.

Table 6 reports the results from a 100 replications experiment deviating from the Table 5 design by one artificial innovations outlier at t=60 of size 1.0. Three additional statistics are displayed: the frequencies of (spurious) data corrections at t=60,61,62. A rougher grid of 0.5 was used here.

Table 7 differs from Table 6 by increasing the size of the IO to two standard deviations. It may be contrasted with Table 6, remembering that the threshold for the residuals varies between one and two standard deviations. Here, 500 replications were used.

An interesting feature is illustrated in Figure 3. The frequency of corrections at t=60 is increased to about the numbers of Table 2. Hereafter, it falls below the average frequency of type I errors for uncontaminated data, approaching it again at t=62,63. In section 8, theoretical reasons for this phenomenon will be given.

Table 8 reports the effect of inserting an AO at t=60, using 200 replications. Again, the numbers are quite close to the calculated ones, but the unbiasedness curve seems to shift to the left, i.e the optimal c is lowered.

Tables 9 and 10 summarize the findings from an experiment where the size of the outliers was changed between one and ten standard deviations. The frequencies of detection at t=60 are analyzed graphically in Figure 4. (c,τ) were fixed to (1.5,1). Figure 4 clearly shows the maximum of wrong decisions for IO of two to three standard deviations. For larger outliers, discrimination almost reaches perfection.

TABLE 6: 1 innovations outlier sized 1 standard deviation at t=60 [12]

τ \ c	1.0	1.5	2.0
1.0	.47247	.46245	.47034
	.54577	.48977	.47933
	4.71	3.83	3.34
	12.09	4.81	1.48
	.15	.13	.06
	.12	.06	.02
	.15	.05	.01
1.5	.46785	.45009	.46301
	.52858	.47547	.46940
	4.39	3.90	3.26
	10.37	4.12	1.11
	.15	.09	.02
	.08	.03	.01
	.07	.04	.06
2.0	.46131	.46633	.47269
	.51197	.48582	.47758
	4.35	3.65	3.17
	8.06	2.98	.98
	.10	.06	.04
	.03	.02	.01
	.08	.01	.01

line 1: PTF estimate
line 2: Yule-Walker estimate
line 3: number of iterations
line 4: number of outliers corrected by PTF
line 5: probability of correcting observation # 60
line 6: probability of correcting observation # 61
line 7: probability of correcting observation # 62
all lines: arithmetic means

12 100 replications using separate random numbers setups of sample size 100 for each of the 9 variants.

TABLE 7: 1 innovations outlier sized 2 standard deviations at t=60 [13]

τ \ c	1.0	1.5	2.0
1.0	.46888	.46888	.46888
	.54505	.49768	.47819
	4.52	3.89	3.33
	12.41	4.81	1.61
	.27	.20	.13
	.07	.03	.02
	.12	.04	.01
1.5	.46888	.46888	.46888
	.53342	.49355	.47676
	4.43	3.83	3.29
	10.32	4.05	1.37
	.23	.18	.11
	.06	.03	.01
	.11	.03	.01
2.0	.46888	.46888	.46888
	.51775	.48886	.47559
	4.29	3.72	3.24
	7.66	3.21	1.16
	.18	.15	.10
	.05	.02	.01
	.08	.03	.01

line 1: PTF estimate
line 2: Yule-Walker estimate
line 3: number of iterations
line 4: number of outliers corrected by PTF
line 5: probability of correcting observation # 60
line 6: probability of correcting observation # 61
line 7: probability of correcting observation # 62
all lines: arithmetic means

---

12 500 replications using common random numbers setups of sample size 100 for all 9 variants.

TABLE 8: 1 additive outlier sized 2 standard deviations at t=60 [14]

τ \ c	1.0	1.5	2.0
1.0	.46529	.46529	.46529
	.54432	.49808	.47781
	4.52	3.87	3.42
	12.00	4.81	1.70
	.61	.47	.31
	.17	.07	.03
	.13	.06	.01
1.5	.46529	.46529	.46529
	.53135	.49319	.47621
	4.30	3.78	3.38
	9.93	4.08	1.43
	.51	.39	.26
	.14	.06	.03
	.08	.04	.02
2.0	.46529	.46529	.46385
	.51500	.48721	.47310
	4.25	3.68	3.29
	7.43	3.15	1.15
	.34	.26	.23
	.07	.05	.02
	.05	.03	.01

line 1: PTF estimate
line 2: Yule-Walker estimate
line 3: number of iterations
line 4: number of outliers corrected by PTF
line 5: probability of correcting observation # 60
line 6: probability of correcting observation # 61
line 7: probability of correcting observation # 62
all lines: arithmetic means

14 200 replications using common random numbers setups of sample size 100 for 8 of 9 variants. Last variant 200 extra replications after mainframe breakdown.

TABLE 9: 1 additive outlier of changing size at t=60 for $(\tau,c)=(1,1.5)$ [15]

S	YW est.	PTF est.	#it	#out	#60	#61	#62
1.0	.48211	.50785	3.86	4.55	.16	.04	.03
2.0	.46529	.49808	3.87	4.81	.47	.07	.06
3.0	.43801	.47856	4.09	5.04	.76	.10	.09
4.0	.41466	.47626	4.08	5.00	.81	.08	.02
5.0	.40532	.48798	4.16	5.40	.86	.04	.05
6.0	.36014	.47784	4.36	5.46	.92	.05	.02
7.0	.34621	.48096	4.20	5.12	.91	.05	.04
8.0	.30457	.46493	4.34	5.44	.93	.09	.04
9.0	.29568	.48435	4.36	4.93	.96	.02	.05
10.0	.26512	.48114	4.40	5.21	.92	.06	.06

TABLE 10: 1 INNOVATIONS OUTLIER of changing size at t=60 for $(\tau,c)=(1,1.5)$ [16]

S	YW est.	PTF est.	#it	#out	#60	#61	#62
1.0	.46245	.48977	3.83	4.81	.13	.06	.05
2.0	.46888	.49768	3.89	4.81	.20	.03	.04
3.0	.46481	.49637	3.87	4.85	.26	.04	.04
4.0	.48614	.51550	3.85	4.53	.23	.03	.03
5.0	.45606	.48310	3.77	4.60	.14	.02	.06
6.0	.47173	.49606	3.73	4.70	.04	.03	.07
7.0	.47140	.49346	3.70	4.77	.03	.04	.07
8.0	.48234	.50128	3.62	4.40	.03	.03	.05
9.0	.47214	.48825	3.63	4.56	.00	.04	.02
10.0	.48456	.50071	3.47	4.21	.02	.01	.07

15 100 replications using separate random numbers setups of sample size 100 for 9 of 10 variants. S=2 reuses the results of table 4.

16 100 replications using separate random numbers setups of sample size 100 for 8 of 10 variants. S=1 and S=2 reuse the results of tables 2 and 3, respectively.

8. Consecutive errors

An important property of PTF is that, following a data correction at time point t, succeeding decisions will be based on the corrected value whereas the original value will not be used any more. This could lead to the presumption that wrong decisions are very dangerous since they entail an increased propensity towards more wrong decisions for a phase whose length depends on the memory of the process. The simulation results, e.g. the graphic display of type I error frequencies around an IO in Figure 3, indicate that this sorrow is unfounded. Following the IO at t=60, the empirical type I error probability at t=61 falls below its average. Interestingly, at t=62 it slightly exceeds the average for some combinations of tuning constants.

Anyway, for the case of $\tau=0$, i.e. the KMT filter, a slight tendency towards increases of errors following a type I error may be proved. Here and in the following, all densities will be assumed as symmetric and unimodal. The lemmata are valid for first-order autoregressive processes with parameter *a* in (0,1).

LEMMA 9: In the case of KMT, the probability of a type I error at t+1 of correcting good data conditional on the appearance of a type I error at t is greater than the unconditional probability as long as $a>0$

Proof: The consecutive error probability CEP is given by

$$P(|x_{t+1}-ax_t^*|>c\sigma \text{ cond.on } |x_t-ax_{t-1}|>c\sigma) =$$

$$= P(|ae_t+e_{t+1}-ac\sigma sgn(e_t)|>c\sigma \text{ cond.on } |e_t|>c\sigma)$$

since $x_t^* = ax_{t-1}+c\sigma sgn(e_t)$

Let the average error propensity be denoted by AEP. Then the CEP may be expressed as

$$2\int_{c\sigma}^{\infty} f(x)\left(\int_{-\infty}^{c\sigma(a-1)-ax} f(t)dt + \int_{c\sigma(a+1)-ax}^{\infty} f(t)dt\right) dx \,/\, AEP$$

which is greater than AEP since the inner integral attains its minimum at $x=c\sigma$. ▪

If the correction at t is done on the basis of an actual AO, the lagging effects obey a slightly different formula :

LEMMA 10: If an outlier at t of size S is detected correctly, the conditional type I error probability at t+1 CEP(S) is given by :

$$CEP(S)=\left[\int_{c\sigma-S}^{\infty} f(x)\left(\int_{c\sigma(a+1)-ax}^{\infty} f(t)dt + \int_{-\infty}^{c\sigma(a-1)-ax} f(t)dt\right) dx +\right.$$

$$\left.+\int_{-\infty}^{-c\sigma-S} f(x)\left(\int_{-c\sigma(a-1)-ax}^{\infty} f(t)dt + \int_{-\infty}^{-c\sigma(a+1)-ax} f(t)dt\right) dx\right]/PAO(S)$$

where PAO(S) denotes the probability of detecting an AO of size S.

Proof: The error probability conditional on an outlier of size S, denoted CEP(S), is

$$P(|x_{t+1}-ax_t^*|>c\sigma \text{ cond.on } |e_t+S|>c\sigma)$$

Now, $x_t^* = ax_{t-1}+c\sigma sgn(e_t+S)$
Then the integrals give the common probability of both events. The denominator equals the conditioning event. ▪

Now the derivative of the numerator expression with respect to S is

$$(f(c\sigma-S)-f(-c\sigma-S))\left[\int_{c\sigma+aS}^{\infty} + \int_{-\infty}^{-c\sigma+aS}\right] > 0$$

with the validity of the inequality depending on the unimodality assumption. The conditioning denominator need not necessarily warrant the presumptions CEP(S)>AEP or CEP(S)>CEP(0). Numerical evaluation of the formula of Lemma 10 over a variety of design specifications, however, confirmed these properties except for outliers which are smaller than about one standard deviation. A direct implication of this is that any pairs - or even patches - of corrected observations are not to be interpreted as evidence on outliers patches but rather as additional evidence on the first outlier in the patch. These patchy patterns occur quite frequently in empirical applications of the KMT filter.

If the PTF with positive tuning constant τ is used, all results tend to get more intricate. The analogue to Lemma 9 in this case uses a triple integral:

LEMMA 11: The conditional error propensity in the case $\tau>1$ and S=0 is

$$CEP(\tau) = 2\Big[\int_{(\tau+1)c\sigma(1+a)/a}^{\infty} f(x) \int_{(1+a)c\sigma-ax}^{-a\tau x/(\tau+1)} f(y) \int_{(a^2c\sigma-a^2x-ay)\tau/(\tau-1)}^{(a^2c\sigma-a^2x-ay)\tau/(\tau+1)} f(z)dzdydx +$$

$$+ \int_{c\sigma}^{\infty} f(x) \int_{-ax\tau/(\tau-1)}^{(a-1)c\sigma-ax} f(y) \int_{(a^2c\sigma-a^2x-ay)\tau/(\tau+1)}^{(a^2c\sigma-a^2x-ay)\tau/(\tau-1)} f(z)dzdydx\Big] / AEP(\tau)$$

as long as $\tau<1/(1-a)$. If $\tau>1/(1-a)$, the lower integration bound of the second outmost integral changes.

Proof: $CEP(\tau) = P(|x_{t+1}-ax_t^*|>c\sigma \wedge \tau|x_{t+2}-a^2x_t^*|<|x_{t+2}-ax_{t+1}|$
cond. on $|x_t-ax_{t-1}|>c\sigma \wedge \tau|a^2x_{t-1}-x_{t+1}|<|x_{t+1}-ax_t|)$

The conditioning probability is AEP(τ). We have to calculate the common probability of the following inequality conditions to hold.

(1) $|x_{t+1}-ax_t^*| = |ae_t+e_{t+1}-ac\sigma sgn(e_t)|>c\sigma$
(2) $\tau|x_{t+2}-a^2x_t^*| = \tau|a^2e_t+ae_{t+1}+e_{t+2}-a^2c\sigma sgn(e_t)|<|e_{t+2}|$
(3) $|e_t|>c\sigma$
(4) $\tau|ae_t+e_{t+1}|<|e_{t+1}|$

Without loss of generality, we shall assume $e_t>0$ i.e. $e_t>c\sigma$ and double the resulting probabilities at the end. First we investigate into the direct conditions for e_{t+1} in (4) by separating four sign cases:

+ + $\tau ae_t+\tau e_{t+1} < e_{t+1}$ is contradictory
+ - $\tau ae_t+\tau e_{t+1} < -e_{t+1}$ implies $e_{t+1} < -a(\tau/(\tau+1))e_t$
- + $-\tau ae_t-\tau e_{t+1} < e_{t+1}$ is contradictory
- - $-\tau ae_t-\tau e_{t+1} < -e_{t+1}$ implies $e_{t+1} > -a(\tau/(\tau-1))e_t$

Further conditions are implied by (1)

+ $ae_t+e_{t+1}-ac\sigma > c\sigma$ or $e_{t+1} > (1+a)c\sigma-ae_t$
- $-ae_t-e_{t+1}+ac\sigma > c\sigma$ or $e_{t+1} < -(1-a)c\sigma-ae_t$

Note that the corresponding sign conditions $e_{t+1} >< ac\sigma-ae_t$ are fulfilled. The overall conditions on e_{t+1} must be the intersection of the areas imposed by (1) and (4). These areas are one finite interval and two half-finite intervals opening to $\pm\infty$. The intersection consists of two intervals which may be empty. We call them the lower and upper part, LP and UP. Now if

$-(1-a)c\sigma-ae_t < -a\tau e_t/(\tau-1)$ or $e_t< c\sigma(\tau-1)(1-a)/a$

LP is empty. On the other hand, if

$(1+a)c\sigma-ae_t < -a\tau e_t/(\tau+1)$ or $e_t< c\sigma(\tau+1)(1+a)/a$

UP is empty. Since $e_t > c\sigma$, LP is never empty for

$(\tau-1)(1-a)/a < 1$ or $\tau < 1/(1-a)$

Whereas this condition includes some conventional τ, its UP analogon is contradictory. This implies different restrictions on the admissible e_t sets for LP and UP:

$c\sigma < e_t < \infty$ for LP and $c\sigma(\tau+1)(1+a)/a < e_t < \infty$ for UP

Note that all admissible e_{t+1} are negative. We now evaluate the direct conditions on e_{t+2} by (2) by separating sign cases:

\+ \+ $\tau a^2 e_t + \tau a e_{t+1} - a^2 c\sigma\tau < (1-\tau)e_{t+2}$ or
$e_{t+2} < (a^2 c\sigma - a^2 e_t - a e_{t+1})\tau/(\tau-1)$ with
$e_{t+2} > 0$ and $e_{t+2} > a^2 c\sigma - a^2 e_t - a e_{t+1}$
The positivity condition is redundant.

\+ − $\tau a^2 e_t + \tau a e_{t+1} - a^2 c\sigma\tau < -(1+\tau)e_{t+2}$ or
$e_{t+2} < (a^2 c\sigma - a^2 e_t - a e_{t+1})\tau/(1+\tau)$ with
$e_{t+2} < 0$ and $e_{t+2} > a^2 c\sigma - a^2 e_t - a e_{t+1}$

− \+ $-\tau(a^2 e_t + a e_{t+1} + e_{t+2} - a^2 c\sigma) < e_{t+2}$ or
$e_{t+2} > (a^2 c\sigma - a^2 e_t - a e_{t+1})\tau/(1+\tau)$ with
$e_{t+2} > 0$ and $e_{t+2} < a^2 c\sigma - a^2 e_t - a e_{t+1}$

− − $-\tau(a^2 e_t + a e_{t+1} + e_{t+2} - a^2 c\sigma) < -e_{t+2}$ or
$e_{t+2} > (a^2 c\sigma - a^2 e_t - a e_{t+1})\tau/(\tau-1)$ with
$e_{t+2} < 0$ and $e_{t+2} < a^2 c\sigma - a^2 e_t - a e_{t+1}$

again with redundant sign constraints. These (2) conditions constitute the integrals

$$\int_{(a^2 c\sigma - a^2 x - ay)\tau/(\tau-1)}^{(a^2 c\sigma - a^2 x - ay)\tau/(\tau+1)} f(z)dz \quad \text{and} \quad \int_{(a^2 c\sigma - a^2 x - ay)\tau/(\tau+1)}^{(a^2 c\sigma - a^2 x - ay)\tau/(\tau-1)} f(z)dz$$

corresponding to negative and positive e_{t+2} and keeping the notation x, y, z for e_t, e_{t+1}, e_{t+2}. These areas impose further indirect conditions on x, y. If

$$a^2c\sigma-a^2x-ay < 0 \qquad \text{then} \qquad y > a(c\sigma-x)$$

which is contained in the restrictions for UP. Analogously, the positive z integral is compatible with LP. This proves the formula. ▪

The integrals may be evaluated numerically for given $f(.), a, c, \tau$. By setting $f(.)$ to the Gaussian density and a to 0.5, one may provide a counterexample to the presumption that always CEP > AEP as in the KMT case. Inside an area in the (c,τ) plane which is bounded by a non-trivial curve (see Figure 5), CEP is less than AEP. This curve looks different for the Cauchy density but the results are similar in the most important interval $1.0<c,\tau<2.0$. To obtain the curve in Figure 5, a result similar to Lemma 11 on the case $\tau<1$ was used.

This effect is not to be seen as outright benefactory. It logically implies that, after type I errors, actual outliers are prone to be "masked". To prove this statement formally, further conditional probabilities (type II AO conditional on type I) could be calculated. However, by the argument of continuity, this statement must hold for small outliers. On the other hand, pairs of detected outliers in the PTF should not be interpreted as additional evidence on the first one, as in KMT, but rather as definitely distinct outlying observations.

The PTF counterpart of Lemma 10 will close this section:

LEMMA 12: The conditional error propensity of a type I error following a correct detection of an AO of size S is

$$\mathrm{CEP}(\tau) = \left[\int_{\tau S+(\tau+1)c\sigma(1+a)/a}^{\infty} f(x) \int_{(1+a)c\sigma+aS-ax}^{a(S-\tau x)/(\tau+1)} f(y) \int_{(a^2(c\sigma+S-x)-ay)\tau/(\tau-1)}^{(a^2(c\sigma+S-x)-ay)\tau/(\tau+1)} f(z)\,dz\,dy\,dx \; + \right.$$

$$+ \int_{c\sigma-S}^{\infty} f(x) \int_{-a(S+\tau x)/(\tau-1)}^{(a-1)c\sigma+aS-ax} f(y) \int_{(a^2(c\sigma+S-x)-ay)\tau/(\tau+1)}^{(a^2(c\sigma+S-x)-ay)\tau/(\tau-1)} f(z)dzdydx +$$

$$+ \int_{-\infty}^{-c\sigma-S} f(x) \int_{a(S-\tau x)/(\tau+1)}^{-(1+a)c\sigma+aS-ax} f(y) \int_{(a^2(S-c\sigma-x)-ay)\tau/(\tau+1)}^{(a^2(S-c\sigma-x)-ay)\tau/(\tau-1)} f(z)dzdydx +$$

$$+ \int_{-\infty}^{-c\sigma-S} f(x) \int_{(1-a)c\sigma+aS-ax}^{-a(S+\tau x)/(\tau-1)} f(y) \int_{(a^2(S-c\sigma-x)-ay)\tau/(\tau-1)}^{(a^2(S-c\sigma-x)-ay)\tau/(\tau+1)} f(z)dzdydx] \; / \; PAO(S,\tau)$$

assuming $S>0$, $1<\tau$ and additional restrictions (see below). Here, $PAO(S,\tau)$ denotes the probability of correctly detecting an AO of size S.

Proof: The conditions (1)-(4) of the last lemma now look

(1) $|ae_t+e_{t+1}-ac\sigma sgn(e_t+S)-aS|>c\sigma$
(2) $\tau|a^2e_t+ae_{t+1}+e_{t+2}-a^2c\sigma sgn(e_t+S)-a^2S|<|e_{t+2}|$
(3) $|e_t+S|>c\sigma$
(4) $\tau|ae_t+e_{t+1}|<|e_{t+1}-aS|$

First, we shall treat the positive case of (3). Unlike in 8.3, the positive and negative cases are asymmetric and must be treated separately. The resulting sign cases of (4) are :

\+ + $\tau ae_t+\tau e_{t+1}-e_{t+1}<-aS$ is contradictory
\+ - $\tau ae_t+\tau e_{t+1}<aS-e_{t+1}$ or $e_{t+1}<a(S-\tau e_t)/(\tau+1)$
keeping $e_{t+1}>-ae_t$ but ignoring $e_{t+1}<aS$ as redundant
\- + $-a\tau e_t-\tau e_{t+1}<e_{t+1}-aS$ is contradictory
\- - $-a\tau e_t-\tau e_{t+1}<aS-e_{t+1}$ or $e_{t+1}>-a(S+\tau e_t)/(\tau-1)$

In summary, this implies (substituting x for e_t)
$-a(S+\tau x)/(\tau-1)<e_{t+1}<a(S-\tau x)/(\tau+1)$

The sign cases for (1) are

$+ \quad ae_t+e_{t+1}-ac\sigma-aS>c\sigma$ or $e_{t+1}>(a+1)c\sigma+aS-ae_t$

$- \quad -ae_t-e_{t+1}+ac\sigma+aS>c\sigma$ or $e_{t+1}<(a-1)c\sigma+aS-ae_t$

This again gives intersection intervals LP and UP. LP is empty if

$(a-1)c\sigma+aS-ae_t<-a(S+\tau e_t)/(\tau-1)$

or $e_t<-\tau S+(\tau-1)(1-a)c\sigma/a$

This is redundant (LP never empty) if

$c\sigma-S>-\tau S+(\tau-1)(1-a)c\sigma/a$

or $\tau(-aS+(1-a)c\sigma)<c\sigma-aS$

$aS<(1-a)c\sigma$ warrants a positive denominator implying $\tau<(c\sigma-aS)/((1-a)c\sigma-aS)$ with the r.h.s.>1

$aS>(1-a)c\sigma$ gives a negative denominator. As long as $aS<c\sigma$, the numerator is positive with any τ fulfilling the condition. Even for greater S, the conditioning expression can never reach 1. This gives

(A) lower bound $c\sigma-S$ for $S>(1-a)c\sigma/a$

(B) l.b. $c\sigma-S$ for $S<(1-a)c\sigma/a$ and $\tau<(c\sigma-aS)/((1-a)c\sigma-aS)$

(C) else l.b. $-\tau S+(\tau-1)(1-a)c\sigma/a$

UP is empty for

$a(S-\tau x)/(\tau+1)<(a+1)c\sigma+aS-ax$

or $x<\tau S+(a+1)(\tau+1)c\sigma/a$

This is greater than $c\sigma-S$ which gives the non-trivial bound. UP will be always empty in the limit $S\to\infty$. The "inner" conditions (2) give the integrals

$$\int_{(a^2(c\sigma+S-x)-ay)\tau/(\tau-1)}^{(a^2(c\sigma+S-x)-ay)\tau/(\tau+1)} f(z)dz \qquad \int_{(a^2(c\sigma+S-x)-ay)\tau/(\tau+1)}^{(a^2(c\sigma+S-x)-ay)\tau/(\tau-1)} f(z)dz$$

where the negative solution corresponds to the UP and the positive one to the LP of above. The detailed proofs of this part parallel those of 8.3 and are omitted. Up to now, we have the first two integrals of 8.4. The third and fourth one stem from the e_t of the negative (3) condition $e_t<-c\sigma-S$. In that case, the sign cases (+,-) and (-,-) imply contradictions whereas (+,+) and (-,+) give the condition

$$a(S-\tau x)/(\tau+1)<e_{t+1}<-a(S+\tau x)/(\tau-1)$$

The (1) inequalities are

+ $\quad ax+e_{t+1}+ac\sigma-aS>c\sigma \quad$ or $\quad e_{t+1}>(1-a)c\sigma+aS-ax$

- $\quad -ax-e_{t+1}-ac\sigma+aS>c\sigma \quad$ or $\quad e_{t+1}<-(1+a)c\sigma+aS-ax$

LP is empty for

$-(1+a)c\sigma+aS-ax<a(S-\tau x)/(\tau+1)$

or $\quad x>\tau S-(1+a)c\sigma(\tau+1)/a$

which makes the reverse condition non-redundant for

$\tau S-(1+a)c\sigma(\tau+1)/a<-c\sigma-S$

or $\tau(aS-c\sigma-ac\sigma)<c\sigma-aS$

or - if the denominator is negative -

$\tau>(c\sigma-aS)/(aS-c\sigma-ac\sigma)$

If $aS<c\sigma$ this is valid for all $\tau>0$. If $aS<(2+a)c\sigma/2$ it is still valid since τ must be greater 1 to be of interest, otherwise all the formulae would have to be reconsidered. If

$(2+a)c\sigma/2<aS<(1+a)c\sigma$

the above τ condition bounds the areas. For all greater S, any τ will imply the trivial upper bound $-c\sigma-S$.

UP is empty for

$(1-a)c\sigma+aS-ax>-a(S+\tau x)/(\tau-1)$

or $\quad \tau(c\sigma-ac\sigma+aS)>(1-a)c\sigma+ax$

or $\quad x<\tau S+(\tau-1)(1-a)c\sigma/a$

This gives a non-trivial upper bound for
$\tau S+(\tau-1)(1-a)c\sigma/a<-c\sigma-S$
or $\tau(aS+(1-a)c\sigma)<(1-a)c\sigma-ac\sigma-aS$

The l.h.s. expression is always positive but the implications are contradictory. The (2) conditions give the following z integrals ($z=e_{t+2}$)

$$\int_{(a^2(S-c\sigma-x)-ay)\tau/(\tau-1)}^{(a^2(S-c\sigma-x)-ay)\tau/(\tau+1)} f(z)dz \qquad \int_{(a^2(S-c\sigma-x)-ay)\tau/(\tau+1)}^{(a^2(S-c\sigma-x)-ay)\tau/(\tau-1)} f(z)dz$$

which completes the proof since these integrals again correspond to the UP and LP conditions, respectively. ∎

9. Tentative conclusions

More empirical research needs to be done until the power of the PTF may be evaluated fully. This research should rely on actual as well as on artificial data. It is to be expected that "optimal" (c,τ) pairs will change with varying samples. A guideline, however, should be fixed for practical applications.

The PTF, like robust filters in general, shows advantages for the user, in particular if he is interested in forecasting, over simple robust estimation without simultaneous data adjustment. Forecasting performance in medium-sized multivariate settings should be the acid test for the procedure.

The PTF presented may be extended in several ways. The weighting function for parameter estimation could be replaced by adaptive weighting (i.e. adaptive to residual distribution). Prediction thresholds may be raised to 3- or 4-step forecasts, thereby even overcoming the frequent problem of aberrant "patches". Of course, this extension path reaches an obvious limit where we are no more willing to identify a large amount of data as "outliers". Again, forecasting should be the acid test.

It may be seen from Tables 5 to 10, particularly if they are compared to Tables 1 to 3, that the assumption of known autoregressive parameters is never strong enough to render the numerical integrals useless. However, certain deviations show up that become larger if the tuning constants are moved away from the curve within the (c,τ) space where AR estimates are unbiased due to the right amount of smoothing to counteract the natural bias of least-squares estimation. On the other hand, the remaining reasons for deviations given in Section 7 - e.g. consecutive errors - only play a minor role.

This suggests the following procedure in practice to fix τ and c:

1. Find a rough estimate of the degree of data dependence to select the correct "unbiasedness curve" in the (c, τ) space.
2. Select a combination of tuning constants on the curve which results in the desired combination of test size and power according to the numerical integrals given in Section 3.
3. If necessary, repeat the procedure if parameters estimated by PTF deviate much from the rough estimates of step 1.

References:

AN, H. and Z. CHEN (1982): On Convergence of LAD Estimates in Autoregression with Infinite Variance; *Journal of Multivariate Analysis* 12,335-345.

CHONG,Y.Y. and D.F. HENDRY (1986): Econometric Evaluation of Linear Macro-Economic Models; *Review of Economic Studies* 53, 671-690.

DENBY, L. and R.D. MARTIN (1979): Robust Estimation of the First-Order Autoregressive Parameter; *Journal of the American Statistical Association* 74,140-146.

FOX, A.J. (1972) Outliers in Time Series. *Journal of the Royal Statistical Society, Series B* 34,350-363.

GUZY, C. (1986): Robust Estimation of Bivariate Autoregressive Processes, A Simulation Study; Mimeo, University of Vienna.

HANNAN, E.J. and M. KANTER (1977): Autoregressive Processes with Infinite Variance; *Journal of Applied Probability* 14,411-415.

HUBER, P.J. (1964): Robust Estimation of a Location Parameter; *Annals of Mathematical Statistics* 35,73-101.

KLEINER, B., R.D. MARTIN, D.J. THOMSON (1979) Robust Estimation of Power Spectra. *Journal of the Royal Statistical Society, Series B* 41,3,313-351.

KUNST, R.M. (1987): Multivariate Robust Filtering; mimeo, Institute for Advanced Studies, Vienna.

KUNST, R.M. (1989): The Performance of Robust Filtering: Some Monte Carlo Evidence; *Computational Statistics Quarterly*, forthcoming.

MARTIN, R.D. (1979): Approximate Conditional-Mean Smoothers and Interpolators; pp.117-143 in: Th.Gasser and M.Rosenblatt: Smoothing Techniques for Curve Estimation, Springer, New York.

MARTIN, R.D. and B. YOHAI (1986): Influence Functionals for Time Series; *Annals of Statistics* 14,3,781-818.

TSAY, R.S. (1986): Time Series Model Specification in the Presence of Outliers; *Journal of the American Statistical Association* 81,393,132-141.

TUAN, P.D. (1984) On Robust Estimation of Parameters for Autoregressive Moving Average Models. in: J.FRANKE, W.HÄRDLE, D.MARTIN (ed.) Robust and Nonlinear Time Series Analysis, Springer

APPENDIX

Figures 1 - 5

FIGURE 1: Outliers in Austrian GDP series

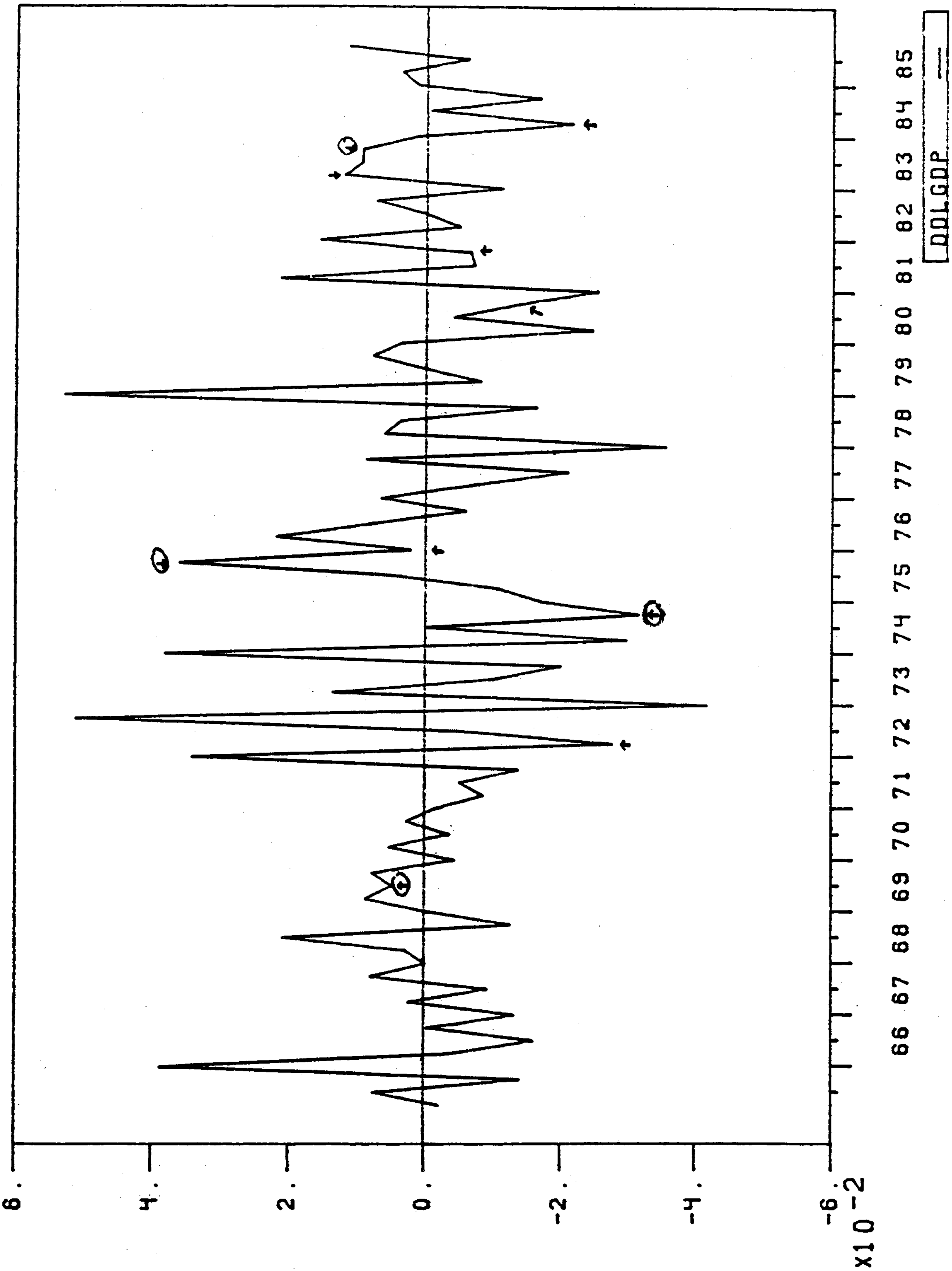
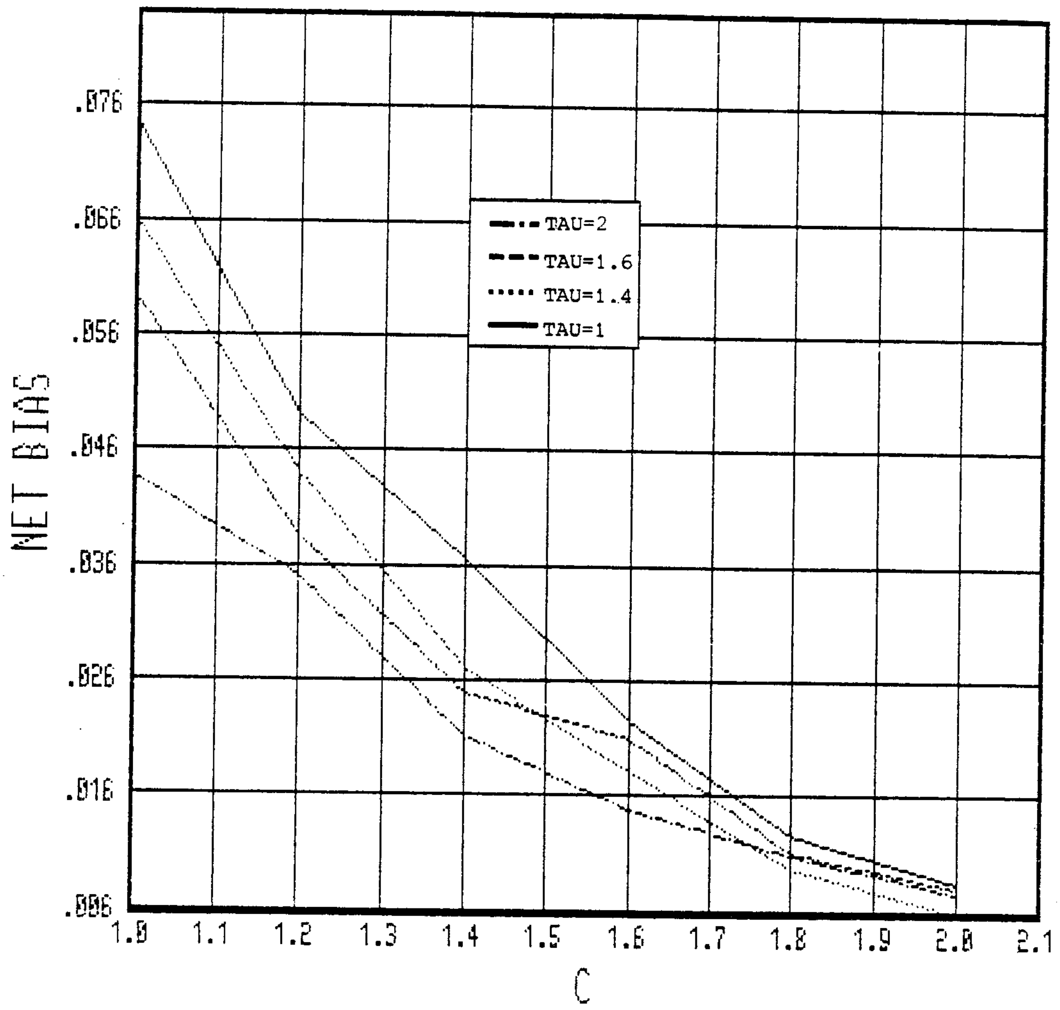
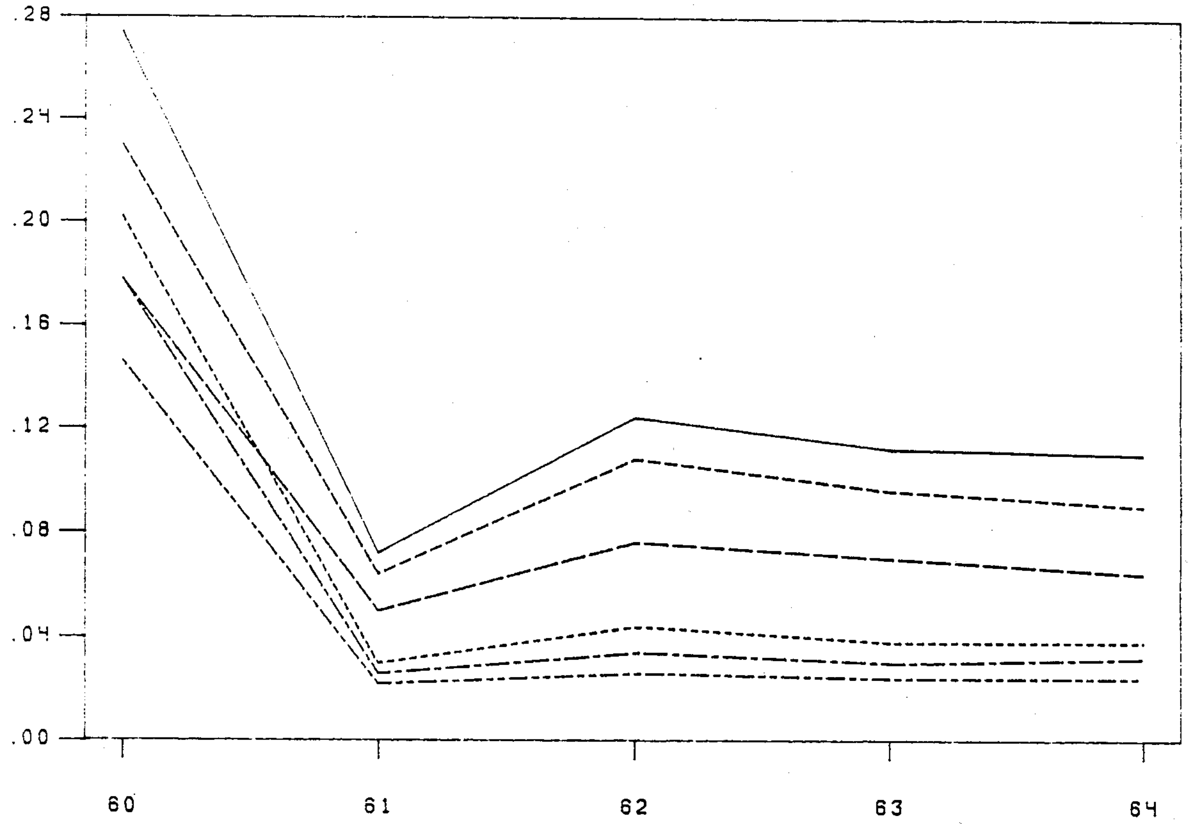


FIGURE 2: Net bias of PTF-2 estimates for a=0.5



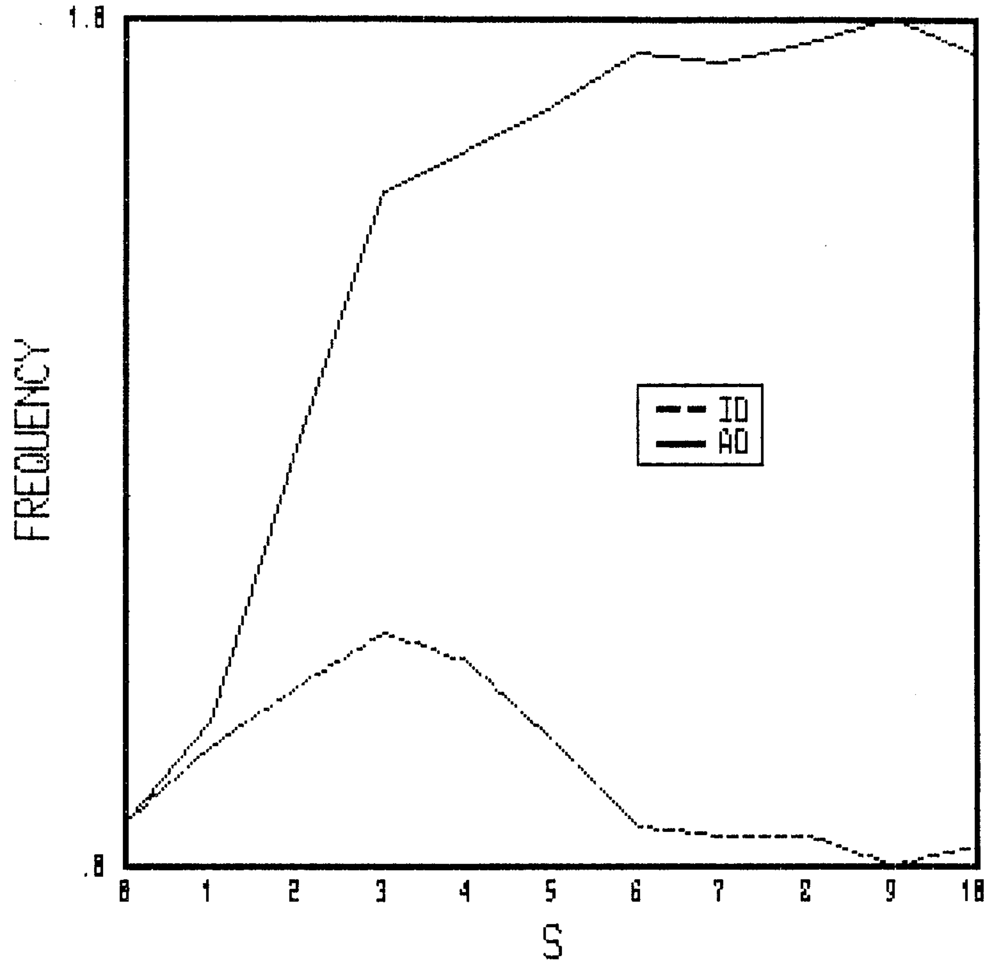
Note: Unbiased estimates for a net bias of 0.0273. C values on the X axis. The curves represent different TAU values.

FIGURE 3: Frequency of spurious corrections after IO at t=60



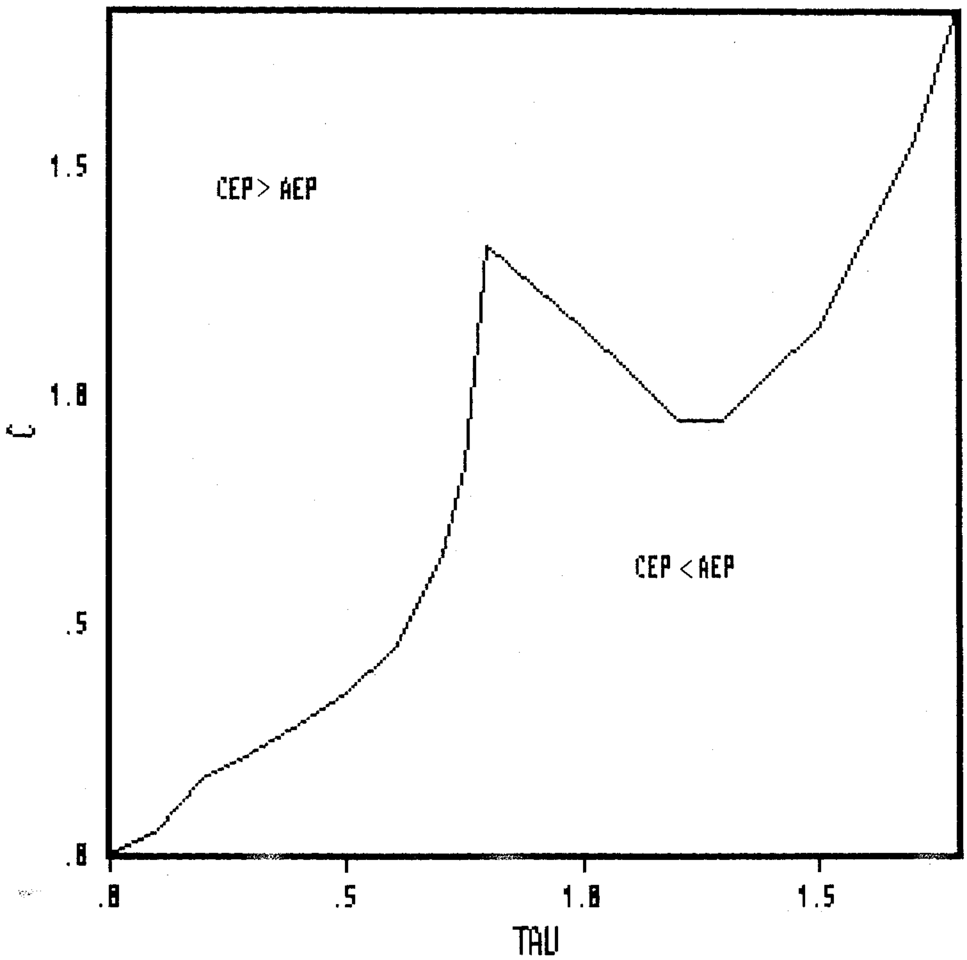
Note: IO sized two standard deviations. Design from above to below (right end):
(c, τ) = (1,1), (1,1.5), (1,2), (1.5,1), (1.5,1.5), (1.5,2)

FIGURE 4: Frequency of corrections for AO and IO



Note: Size of outliers relative to standard deviation of the error distribution at X axis.

FIGURE 5: Boundary curve AEP=CEP



Note: Theoretical curve for Gaussian errors and a=0.5