INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

On the filtering of stochastically non-linear economic time series: An application to stock prices

Han, Sangwan, Ph.D.

City University of New York, 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A

ON THE FILTERING OF
STOCHASTICALLY NON-LINEAR ECONOMIC TIME SERIES

- An Application To Stock Prices -

by

SANGWAN HAN

A dissertation submitted to the Graduate Faculty in Economics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1991

This manuscript has been read and accepted for the Graduate Faculty in Economics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

March 23, 1991
Date

Salih N. Neftci
Chair of Examining Committee

March 23, 1991
Date

Michael Grossman
Executive Officer

Salih N. Neftci, Ph.D.

Michael Grossman, Ph,D.

Ronald Anderson, Ph.D.

Supervisory Committee

The City University of New York

Abstract

ON THE FILTERING OF STOCHASTICALLY NON-LINEAR ECONOMIC TIME SERIES

- An Application To Stock Prices -

by

Sangwan Han

Advisor: Professor Salih N. Neftci

The aim of this paper is to contribute to an understanding of the filtering theory of stochastically non-linear economic time series. Both the linear and non-linear filters are derived and applied to three different stock price indices (Standard & Poor's, Korean Composite Stock Price Index and Dow Jones Industrial Average) under the assumption that stock prices follow a two-state, first-order Markov process. The results will show that, under fairly general condition, the non-linear filter outperforms the linear filter. However, if cost effectiveness is a source of major concern, the linear filter is preferable.

Acknowledgements

I take this opportunity to express my deep appreciation to those who, throughout my years at the Graduate Center, were there when I needed them.

My special thanks to Professor Salih N. Neftci for the inspiration. He always made me see the positive in any situation and made me believe in my ability. Without him, I could not have come this far.

Professor Michael Grossman made my work at the Graduate Center a possibility. He helped me to get by every obstacle that I was faced with. I will be forever grateful to him.

Last, but not the least, I owe a deep debt of gratitude to my friends, Laura Bonanomi, Ahmet Enis Kocagil and Christopher Vaz, who pulled and pushed me all the way through my study. I am very fortunate to have such friends.

February 26, 1991

To My Parents And My Brother,
For All That I Am.

Table of Contents

List of Graphs

I. Introduction

1.1 Objective

In recent years, models for stochastically non-linear processes have gained increased importance in economic time series analysis. For example, Hamilton developed a first-order Markov model for U.S. real GNP and successfully estimated the business turning points for the U.S. economy. The reason for the increased importance is that linear models fail to successfully estimate the underlying parameters when the parameters are highly non-Gaussian. While linear models take information only from the first- and second-order moments of the time series, non-linear models require higher moments to be characterized. Hence, linear models are inadequate when the underlying economic time series exhibits highly non-Gaussian characteristics. For this reason, the development of Markov models is very useful for economic time series analysis.

The estimation of stochastically non-linear time series is optimally handled by both the non-linear and the linear filtering theory. The non-linear filtering theory provides a closed-form filter for the optimal estimation and the forecasting of parameters when the observed series follows a Markov Process. This filter removes the noise from the observed series and produces the trajectory of the unobservables, that is, the non-linear parameters. One advantage of non-linear filters is their sensitive reaction to the arrival of new information. Due to their high sensitivity, non-linear filters are capable of quickly detecting the onset of a new state that the parameter takes.

The estimation of stochastically non-linear time series can also be handled optimally by the linear filtering theory. Linear filters are less sensitive than non-linear filters since they take information only from the first- and the second-order moments of the process. Yet, due to their lower sensitivity, linear filters are able to avoid the problems that can arise from the noisiness and the inaccuracy of the data. Furthermore, since they take a less complex form, linear filters are less costly than non-linear filters. Therefore, it would be reasonable to consider a linear filter for its simplicity and lower sensitivity.

The objective of this paper is to compare the optimal non-linear filter with the optimal linear filter when the underlying economic time series follows a first-order Markov process, where the optimality is treated in a Mean Squared Error sense. [1] In doing so, this paper will present briefly the derivation of both the optimal non-linear filter and the optimal linear filter for Markov processes. The optimal non-linear filter is derived using Ito's formula while the optimal linear filter is obtained through the application of Kalman-Bucy method. The two filters are then applied to stock prices assuming that they follow a first-order Markov process. This assumption is not unreasonable since stock prices make sudden switches from one state to another, i.e. upward trends and downward trends.

Two assumptions are made throughout this paper. i) Investors are assumed to be rational in forming their expectation. ii) Markets are informationally efficient so that stock prices serve as an aggregator of the infromation available in the market. Under these two assumptions, investors can infer the state of the world from the observed stock prices. Therefore, it becomes possible that the non-linear and linear filters can optimally estimate the unobserved parameters from the underlying stock price.

---

1. The maximum a-posteriori criterion would yield the same result.

1.2 Data Sets

In the application of the non-linear and linear filters, three different financial series are carefully selected in order to reveal different aspects of the two filters. They are the Standard & Poor's 500 Index (S&P 500), the Korean Composite Stock Price Index (KCSPI), and the Dow Jones Industrial Average (DJIA).

A monthly series of the S&P 500 is incorporated in this study while the other two series are daily. The monthly series of S&P 500 is chosen because it is much less noisy compared to the daily series. The advantages and the disadvantages -with respect to the sensitivity issue- can be demonstrated by comparing the estimates calculated using series having different time intervals.

Although both the KCSPI and the DJIA data sets chosen in this study have daily intervals, they exhibit different characteristics. The KCSPI is much noisier than the DJIA. On the other hand, the DJIA exhibits many discrete jumps (discontinuities in the series such as the Black Monday of October, 1987) while the KCSPI hardly contains any. Comparison of the estimates from these two data sets will provide the different behaviors of the non-linear and the linear filter.

1.3 Outline

Section 2 of this paper discusses the optimal non-linear filter assuming that stock prices follow a first-order Markov process. The non-linear filter is derived in continuous time using Ito's formula and, then, is approximated for discrete time processes. The interpretation of the filter is presented, and the discussion on the stationarity of the filter follows the interpretation. In Section 3, the optimal linear filter for the Markov process is derived. In doing so, the Gaussian analog of the original process is derived, and, then, the Kalman-Bucy filtering theory is applied to the analog. The heuristic interpretation of the filter is included in this section. Also, the stationarity of the filter is discussed. In Section 4 through 6, the parameters of the Markov model are estimated by applying both the non-linear and the linear filter to three different time series: the S&P 500, the KCSPI, and the DJIA, respectively. The estimates are presented in the graphs attached at the end of each section. Section 7 compares the two filters on various issues: sensitivity, robustness, stationarity and cost efficiency. Section 8, finally, concludes this paper.

II. Non-Linear Filter

Non-Gaussian dynamics is the best characterization of non-linear stochastic processes. The parameter suddenly switches from one state to another, as is observed in non-linear business cycles, regime changes in monetary policy, and stock market crashes. The optimal non-linear filter for non-Gaussian dynamics is known. Its derivation will be briefly presented in this section. [2]

2.1 Model

2.1.1 Model

The model assumes that stock prices follow a process given by the following stochastic differential equation:

$$dX_t = A(X_t, \theta_t)dt + dw_t \qquad (1)$$

---

[2] The full proof of the theorem can be found in Lipster and Shiryayev, 1977.

where $\{X_t\}$ = the logarithmic first-order difference of stock prices

$\{\theta_t\}$ = a finite state Markov process

$\{w_t\}$ = a Wiener process with a variance σ_w^2 and is orthogonal to $\{\theta_t\}$ for all t.

$A(X_t,\theta_t)$ is a random function and is known with certainty given the information available at time t.

In this paper, $A(X_t,\theta_t)$ is assumed to be additive-separable in X_t and θ_t. Furthermore, the stock price is assumed to be in one of two states - either upward or downward trend. Let $A(X_t,\theta_t)$ be given as:

$$A(X_t,\theta_t) = \theta_t + \alpha X_t \qquad (2)$$

where $\theta_t = a_1$ if the stock price follows an upward trend

$= a_2$ if the stock price follows an downward trend,

a_i (for $i = 1,2$) are known constants.

α is also a known constant. The convergence of the model requires α to be negative.

Equations (1) and (2) describe a stock price path which switches between upward and downward trends depending on the value of θ_t.

Note that the process described by the system above is highly non-Gaussian and that it takes a sudden state change from a_1 to a_2.

2.1.2 Transition Probabilities

The Markov process $\{\theta_t\}$ has the following transition probabilities:

$$\begin{aligned} P(\theta_t=a_i|\theta_s=a_j,\{\theta_{s-u},u>0\}) &= P(\theta_t=a_i|\theta_s=a_j) \qquad i,j = 1,2, \quad t\geq s \\ &= P_{ij}(t-s). \end{aligned}$$

Let us assume that this transition probability is differentiable with respect to (t-s).

$$\begin{aligned} P_{ii}(t-s) &= 1 + f_{ii}(t-s) + o(s) \qquad t\geq s, \quad i=1,2 \\ P_{ij}(t-s) &= 0 + f_{ij}(t-s) + o(s) \qquad t\geq s, \quad i\neq j=1,2 \end{aligned} \tag{3}$$

with $P_{ii}(0)=1$, $P_{ij}(0)=0$, $f_{ii}<0$, $\delta f_{ii}/\delta t<0$, $f_{ij}>0$ and $\delta f_{ij}/\delta t>0$.

Notice from (3) that the system has a memory. By requiring that $f_{ii}<0$, $\delta f_{ii}/\delta t<0$, $f_{ij}>0$ and $\delta f_{ij}/\delta t>0$, as (t-s) increases, the probability that θ_t will remain in the same state decreases and the probability that it will switch to a new state increases, even in the absence of new information. In other words, the transition probabilities are duration dependent and, consequently, the duration of the stock price in a given state determines the magnitude of the transition probabilities.

2.2 Derivation

Now, the purpose of this paper is to infer, in some optimal fashion, the value of the unobserved parameter $\{\theta_t\}$ from the observed process $\{X_t\}$. The optimality is treated in a Mean Squared Error sense. Then, $\{\theta_t\}$ is estimated by:

$$\text{Min } E[(\theta_t-\hat{\theta}_t)^2|I_t] \qquad \text{w.r.t. } \hat{\theta}_t \qquad (4)$$

which yields:

$$\hat{\theta}_t = E[\theta_t|I_t] \qquad (5)$$

where I_t is the information set available at time t. Hence,

$$\hat{\theta}_t = a_1P(\theta_t=a_1|I_t) + a_2P(\theta_t=a_2|I_t)$$
$$= (a_1-a_2)P(\theta_t=a_1|I_t) + a_2$$

Let π_t be the best estimate of $P(\theta_t=a_1|I_t)$, i.e.,

$$\pi_t = P(\theta_t=a_1|I_t) \qquad (6)$$

Then, $P(\theta_t|I_t)$ determines the optimal estimates of the unobserved Markov process $\{\theta_t\}$. It is a non-linear function of I_t. The focus of this paper is to obtain the law of the motion for $\{\pi_t\}$.

2.2.1 Theorem

Let $\{X_t\}$ and $\{\theta_t\}$ be given by the system (1) - (3), then the optimal MSE estimate of θ_t will be given by $E[\theta_t|I_t]$:

$$E[\theta_t|I_t] = (a_1-a_2)\pi_t + a_2$$

where $\pi_t = P(\theta_t=a_1|I_t)$ satisfies the following Ito-Stochastic differential equation:

$$d\pi_t = [(\delta f_{11}/\delta t)\pi_t+(\delta f_{12}/\delta t)(1-\pi_t)]dt$$
$$+ [\pi_t(1-\pi_t)/\sigma^2](a_1-a_2)[dX_t-(\alpha X_t+a_1\pi_t+a_2(1-\pi_t))dt] \qquad (7)$$

where $\delta f_{ij}/\delta t$ are given by (3) and represent the time derivatives of the transition probabilities $P_{ij}(t-s)$.

We assume that π_0 is known.

2.2.2 Proof [3]

Define the innovation process $\{v_t\}$ by

$$\begin{aligned} dv_t &= (1/\sigma)(dX_t - E[dX_t | I_t]) \\ &= (1/\sigma)[(\theta_t - \hat{\theta}_t)dt + dw_t] \qquad (8) \end{aligned}$$

Lemma 1: $\{v_t\}$ is a Wiener process and it has no counterpart in discrete time since dv_t depends on a highly non-Gaussian variable θ_t.

Lemma 2: Given the model and Lemma 1,

$$\begin{aligned} E[\theta_t | \{X_u, u<t\}] &= E[\theta_t | \{v_u, u<t\}] \\ &= E[\theta_t | \{v_u, u\leq s\}] + E[\theta_t | \{v_u, s<u<t\}] \qquad (9) \end{aligned}$$

---

[3] See Frost & Kaliath, 1977 for the proof of lemmas 1 and 2.

It can be seen that $E[\theta_t]$ consists of two components: the past information, i.e. $E[\theta_t | \{v_u, u \le s\}]$, and the innovation, i.e. $E[\theta_t | \{v_u, s<u<t\}]$.

The first element on the RHS of (9) is:

$$E[\theta_t | \{v_u, u \le s\}] = (a_1 - a_2) P(\theta_t = a_1 | \{v_u, u \le s\}) + a_2 \tag{10}$$

$$= (a_1 - a_2) P(\theta_t = a_1 | \theta_s = a_1, \{v_u, u \le s\}) P(\theta_s = a_1 | \{v_u, u<s\})$$
$$+ (a_1 - a_2) P(\theta_t = a_1 | \theta_s = a_2, \{v_u, u \le s\}) P(\theta_s = a_2 | \{v_u, u<s\})$$
$$+ a_2$$

$$= (a_1 - a_2) P_{11}(t-s) \pi_s + (a_1 - a_2) P_{12}(t-s)(1-\pi_s) + a_2$$

Then, the RHS of (10) can be rewritten as:

$$E[\theta_t | \{v_u, u \le s\}] = (a_1 - a_2)[P_{11}(t-s)\pi_s + P_{12}(t-s)(1-\pi_s)] + a_2 \tag{11}$$

Taking the differential of equation (11),

$$dE[\theta_t | \{v_u, u \le s\}] = (a_1 - a_2)[(\delta f_{11}/\delta t)\pi_t + (\delta f_{12}/\delta t)(1-\pi_t)]dt \tag{12}$$

since $dP_{11}(t-s)/dt = \delta f_{11}/\delta t$

$dP_{12}(t-s)/dt = \delta f_{12}/\delta t$

The second element on the RHS of (9) measures the effect of the innovation component of the new information. From the definition of orthogonal projection,

$$\begin{aligned}\lim_{s\to t} dE[\theta_t|\{v_u, s<u<t\}] &= (E[\theta_t dv_t|I_t]/E[dv_t]^2)dv_t \\ &= E[\theta_t dv_t|I_t]dv_t \\ &= (1/\sigma)E[\theta_t((\theta_t-\hat{\theta}_t)dt+dw_t)|I_t]dv_t \\ &= (1/\sigma)E[a_1(a_1-\hat{\theta}_t)\pi_t+a_2(a_2-\hat{\theta}_t)(1-\pi_t)]dv_t\end{aligned}$$

Plugging $\hat{\theta}_t = (a_1-a_2)\pi_t + a_2$ into the equation, we obtain

$$\begin{aligned}\lim_{s\to t} dE[\theta_t|\{v_u, s<u<t\}] &= (1/\sigma)(a_1-a_2)^2\pi_t(1-\pi_t)dv_t \qquad (13) \\ &= (1/\sigma^2)(a_1-a_2)^2\pi_t(1-\pi_t)(dX_t-E[dX_t|I_t])\end{aligned}$$

Combining equations (12) and (13), we get the law of motion for $\hat{\theta}_t$.

$$dE[\theta_t|\{v_u,u<t\}] = (a_1-a_2)[(\delta f_{11}/\delta t)\pi_t+(\delta f_{12}/\delta t)(1-\pi_t)]dt + (1/\sigma^2)(a_1-a_2)^2\pi_t(1-\pi_t) * [dX_t-(\alpha X_t+a_1\pi_t+a_2(1-\pi_t))dt]$$

$$dE[\theta_t|\{v_u,u<t\}] = d[(a_1-a_2)\pi_t+a_2] = (a_1-a_2)d\pi_t$$

Thus, the law of motion for π_t is

$$d\pi_t = [(\delta f_{11}/\delta t)\pi_t+(\delta f_{12}/\delta t)(1-\pi_t)]dt + (1/\sigma^2)(a_1-a_2)\pi_t(1-\pi_t)[dX_t-(\alpha X_t+a_1\pi_t+a_2(1-\pi_t))dt] \qquad (14)$$

2.3 Discrete Time Approximation

The closed form representation discussed above applies to the continuous time case only. In other words, a closed form representation that gives the trajectory of π_t is not available for discrete time. Therefore, a discrete time approximation is required.

The non-linear filter given by equation (14) has the following structure:

$$d\pi_t = a(\pi_t)dt + b(\pi_t)dv_t \qquad (15)$$

where $a(\pi_t)$ measures the change in the estimate of π_t due to the memory of the system, and $b(\pi_t)dv_t$ measures the filter gain. Specifically, $b(\pi_t)$ is the weight while dv_t is the increment of a standard Wiener process (i.e. it is an innovation term). Thus, equation (15) is a stochastic differential equation with the diffusion term dependent on π_t. This equation can be discretized by using Ito-calculus which achieves the purpose by providing a correction factor.

Lemma 3: Given a stochastic differential equation

$$dX_t = a(t,X_t)dt + b(t,X_t)dw_t$$

we have the following first-order approximation with an error of order o(dt):

$$\begin{aligned} X_{t+1} - X_t = \; & a(t,X_t) \\ & + b(t,X_t)(w_{t+1}-w_t) \\ & + [db(t,X_t)/dX_t]b(t,X_t) \end{aligned}$$

Applying the Lemma to equation (14), we obtain the discrete time approximation:

$$\begin{aligned}\pi_{t+1} = {} & \pi_t + (\delta f_{11}/\delta t)\pi_t + (\delta f_{12}/\delta t)(1-\pi_t) \\ & + [\pi_t(1-\pi_t)/\sigma^2][a_1-a_2][(X_{t+1}-X_t)-(\alpha X_t+a_1\pi_t+a_2(1-\pi_t))] \\ & + [\pi_t(1-\pi_t)(1-2\pi_t)][(a_1-a_2)^2/\sigma^2]\end{aligned} \qquad (16)$$

2.4 Interpretation of The Filter

Equation (16) gives the closed form representation of the non-linear filter. Note that the RHS of the equation is composed of three blocks, the updating mechanism on the past information, the filter gain, and the correction factor.

The updating mechanism on the past information represents the change in π_t due to lapse of time. Given that no new information has arrived, the best estimate of π_{t+1} at time t+1 is π_t adjusted by the transition probability which, by assumption, is duration dependent.

$$\begin{aligned}E[\pi_{t+1}|I_{t+1}] & = E[\pi_{t+1}|I_t] \qquad \text{(since } I_{t+1} = I_t\text{)} \\ & = \pi_t + (\delta f_{11}/\delta t)\pi_t + (\delta f_{12}/\delta t)(1-\pi_t)\end{aligned}$$

Note that the system stores the past information in $\delta f_{11}/\delta t$ and $\delta f_{12}/\delta t$ and updates the estimate by $(\delta f_{11}/\delta t)\pi_t + (\delta f_{12}/\delta t)(1-\pi_t)$ as time lapses. At the same time, the transition probability adjusted component of the estimate makes the trajectory of π_t stationary, an issue which will be discussed in depth in section 2.5.

The second block, that is, the filter gain consists of two factors, which are the innovation term and the weight assigned to the innovation. The former is the unexpected portion of the new information that arrives at t+1, namely,

$$(X_{t+1}-X_t) - (\alpha X_t+a_1\pi_t+a_2(1-\pi_t)).$$

Note that when the innovation is zero, the filter gain also becomes zero, and, accordingly, the estimate will not be revised except for the effect of the system's memory.

The weight assigned to the innovation is comprised of two different factors: the covariance of the estimate with the innovation, $\pi_t(1-\pi_t)(a_1-a_2)$ and the variance of the innovation, σ^2. The stronger is the correlation between the signal and the estimate, the larger will be the weight. On the other hand, the larger is the variance of the innovation (which implies that the signal is noisy), the lower will be the weight.

The correction factor, $[\pi_t(1-\pi_t)(1-2\pi_t)][(a_1-a_2)^2/\sigma^2]$, offers an interesting interpretation. Provided that (a_1-a_2) does not vanish, $(a_1-a_2)^2/\sigma^2$ is always positive, and, therefore, it is of little concern. However, the sign of the term $\pi_t(1-\pi_t)(1-2\pi_t)$ varies depending on the values of π_t, as indicated by the following table.

$0 < \pi_t < 0.5$,	$\pi_t(1-\pi_t)(1-2\pi_t) > 0$
$\pi_t = 0.5$,	$\pi_t(1-\pi_t)(1-2\pi_t) = 0$
$0.5 < \pi_t < 1$,	$\pi_t(1-\pi_t)(1-2\pi_t) < 0$

It seems quite clear that the correction factor always pulls the estimate toward 0.5 thus yielding a more conservative estimate. This property is very similar to that of shrinkage estimation procedures (i.e. James-Stein estimator).

2.5 Stationarity

The non-linear filter is stationary in its continuous time formulation. As mentioned before, $(\delta f_{11}/\delta t)\pi_t + (\delta f_{12}/\delta t)(1-\pi_t)$ is the memory part of the system. It stores the past information in $\delta f_{11}/\delta t$ and $\delta f_{12}/\delta t$ and updates the estimate accordingly. In addition, it serves as a factor which makes the trajectory stationary. The table below illustrates this concept by showing the sign that each component takes when the probability estimate goes beyond the boundary limits.

	$(\delta f_{11}/\delta t)\pi_t$	$(\delta f_{12}/\delta t)(1-\pi_t)$	Sum of the Two
$\pi_t < 0$	+	+	+
$\pi_t = 0$	0	+	+
$\pi_t = 1$	-	0	-
$\pi_t > 1$	-	-	-

The table shows that, should a big shock arrive to the system and drive the estimate out of its boundary limits, the memory component will always work so as to force the estimate back into the boundary. In this sense, the continuous time formulation of the non-linear filter is stationary. Such conclusion does not apply to the discrete time approximation version of the model due to the presence of the correction factor.

In the discrete time formulation, the trajectory of π_t contains the correction factor as an additional term, i.e. $[\pi_t(1-\pi_t)(1-2\pi_t)][(a_1-a_2)^2/\sigma^2]$. As discussed previously, this correction factor pulls the estimate toward 0.5 whenever π_t ranges between 0 and 1. As long as no big shock arrives to the system, the model keeps generating acceptable new estimates. However, once π_t moves out of the boundary, the correction factor will drive the estimate even further out, as it can be seen by examining the following Table:

$\pi_t < 0$,	$\pi_t(1-\pi_t)(1-2\pi_t) < 0$
$\pi_t > 1$,	$\pi_t(1-\pi_t)(1-2\pi_t) > 0$

Under these circumstances, the model may fail to generate acceptable new estimates. In other words, the estimates may diverge to either $-\infty$ or $+\infty$. Whether the estimate will diverge or not depends on the sign of $\pi_{t+1} - \pi_t$, i.e. on the combined effects of the memory part of system, the filter gain and the correction factor. It turns out that, in most cases, the filter gain takes on the same sign as the correction factor; it is also known that, in absolute value, the magnitude of the memory component of the system is smaller than the correction factor. Then, by analysing the table below, it shoud be clear that divergence is very likely to occur.

	Memory	F. G.	C. F.
$\pi_t < 0$	+	random	-
$\pi_t > 1$	-	random	+

It is, in this sense, that we refer to the discrete time approximation as being non-stationary. In the rare event that the filter gain takes the same sign as the memory, it may occur that the combined effect of the two exceeds in magnitude that of the correction factor. When this is the case, convergence will be the result.

III. Linear Filter -- Kalman-Bucy Scheme

The non-Gaussian process can be approximated by the Gaussian analog of the original process, i.e. a process with the same first- and second-order moments. As a matter of fact, they are two different processes. Yet, the analog duplicates efficiently the original process since, through the first- and second-order moments, the analog contains all the information that the original process has, except for those contained in the higher moments.

The optimal linear filter for a non-Gaussian process is known to be the optimal linear filter for the Gaussian analog of the original process, that is a Gaussian process with the same expectation and correlation as the original. Unlike in the case of multiple states, it is an easy task to construct a Gaussian analog of a Markov process with two states only.

3.1 Model

Let us consider a filtering problem with an observable process $\{X_t\}$ in discrete time. We will consider the same model as in the non-linear case.

$$X_{t+1} - X_t = \theta_t + \alpha X_t + \delta_{t+1} \qquad (17)$$

where $\{X_t\}$ = the logarithmic first-order difference of stock prices.

$\{\delta_t\}$ = Gaussian random process independent of θ_t.

$E[\delta_t] = 0$

$E[\delta_t]^2 = \sigma^2$

$\{\theta_t\}$ = a Markov Process

= a_1 if the market is bullish

= a_2 if bearish.

a_i (for $i = 1,2$) are assumed to be given a priori.

α is a constant chosen a priori. The convergence of the model requires α to be negative.

P(t) is a matrix of transition probabilities of the Markov process $\{\theta_t\}$ with elements:

$$P_{ij}(t) = P(\theta_{t+1}=a_i \mid \theta_t=a_j) \qquad i,j = 1,2$$

3.2 Derivation of The Optimal Linear Filter [4]

Let us denote

$P_i(t) = P(\theta_t = a_i) \qquad i = 1,2$

Define $\pi(t) = (\pi_1(t), \pi_2(t))$ with $\pi_i(t) = I(\theta_t = a_i)$

The following results are needed for the derivation of the optimal linear filter.

Lemma 4: The sequences $\epsilon_j = (\epsilon_j(t))$, $j = 1,2$ with

$$\epsilon_j(t+1) = \pi_j(t+1) - \sum_{i=1}^{2} P_{ij}(t)\pi_i(t),$$

are the martingale differences with respect to the family of σ-algebras $\mathbb{F} = (F_t)$, $F_t = \sigma\{\pi(0), \ldots, \pi(t)\}$, i.e.,

---

4. Proofs and derivations can be found in Krichagina, Lipster, and Rubinovich, 1984.

$$E[\epsilon_j(t+1) \mid F_t] = 0 \qquad (P\text{-a.s.}).$$

Furthermore,

$$q_{jj}(t) = E[\epsilon_j(t+1)]^2 = p_j(t+1) - \sum_{i=1}^{2} p_{ij}(t)^2 p_i(t)$$

$$q_{ij}(t) = E[\epsilon_i(t+1)\epsilon_j(t+1)] = - \sum_{l=1}^{2} p_{li}(t) p_{lj}(t) p_l(t) \tag{18}$$

3.2.1 Gaussian Analog

Let us now construct the Gaussian analog of the process $\pi = (\pi(t))$.

Denote $\epsilon(t) = (\epsilon_1(t), \epsilon_2(t))$. From the definition of $\epsilon_j(t)$, it follows that

$$\pi(t+1) = P(t)^T \pi(t) + \epsilon(t+1) \qquad (19)^5$$

---

[5] Superscript T denotes matrix transposition.

Then, the Gaussian analog of the process $\{\pi(t)\}$ is the vector process $\{\tilde{\pi}(t)\}$ defined by

$$\tilde{\pi}(t+1) = P(t)^T\tilde{\pi}(t) + \tilde{\epsilon}(t+1) \qquad (20)$$

where $\tilde{\epsilon}(t) = (\tilde{\epsilon}_1(t),\tilde{\epsilon}_2(t))$, $t = 1,2,\ldots$

is the sequence of independent Gaussian vectors with

$E[\tilde{\epsilon}(t)] = 0$

$E[\tilde{\epsilon}(t)^T\tilde{\epsilon}(t)] = Q(t) = ||q_{ij}(t)||$ for $i,j = 1,2$

$\tilde{\pi}(0)$ is a Gaussian vector independent of $(\tilde{\epsilon}(t))$ with

$$E[\tilde{\pi}_j(0)] = P_j(0) \qquad (21)$$

$$\begin{aligned} Cov[\tilde{\pi}_i(0),\tilde{\pi}_j(0)] &= P_i(0)(1-P_i(0)) && i = j \\ &= -\,P_i(0)P_j(0) && i \neq j \qquad (22) \end{aligned}$$

Since $\pi_2(t) = 1 - \pi_1(t)$,

$\tilde{\pi}_2(t) = 1 - \tilde{\pi}_1(t)$

Therefore, from equation (20), it follows that

$$\tilde{\pi}_1(t+1) = [P_{11}(t)-P_{12}(t)]\tilde{\pi}_1(t) + P_{12}(t) + \tilde{\epsilon}_1(t+1) \qquad (23)$$

From the definition of $\pi_i(t)$, $i = 1,2$, we obtain

$$\theta_t = a_1\pi_1(t) + a_2\pi_2(t) = (a_1-a_2)\pi_1(t) + a_2$$

and the Gaussian analog for θ_t is

$$\tilde{\theta}_t = (a_1-a_2)\tilde{\pi}_1(t) + a_2. \qquad (24)$$

3.2.2 Kalman-Bucy Scheme

In order to obtain the optimal linear filter, it is sufficient to use a Kalman filter for the process $\{\tilde{\pi}_1(t)\}$ with the observable process $\{X_t\}$. Applying the Kalman-Bucy scheme, the optimal linear estimate m(t) for $\pi_1(t)$ is obtained as follows:

$$m(t+1) = E[\pi_1(t+1) | I_{t+1}]$$

where I_{t+1} = information available at t+1

Now, I_{t+1} can be decomposed into two parts, I_t and μ_{t+1}, where μ_{t+1} is the unexpected portion of the signal. Then, by Wald decomposition theorem,

$$m(t+1) = E[\pi_1(t+1)|I_{t+1}]$$
$$= E[\pi_1(t+1)|I_t] + E[\pi_1(t+1)|\mu_{t+1}]$$

The first element on the RHS of the equation is:

$$E[\pi_1(t+1)|I_t] = [P_{11}(t)-P_{12}(t)]m(t) + P_{12}(t) \tag{25}$$

The second element on the RHS of the equation is:

$$\begin{aligned} E[\pi_1(t+1)|\mu_{t+1}] &= (E[\pi_1(t+1)\mu_{t+1}]/E[\mu_{t+1}{}^2])\mu_{t+1} \\ &= [\sigma^2+(a_1-a_2)^2\Gamma(t)]^{-1}[P_{11}(t)-P_{12}(t)]\Gamma(t)[a_1-a_2] \\ &\quad * [X_{t+1}-(1+\alpha)X_t-(a_1-a_2)m(t)-a_2] \end{aligned} \tag{26}$$

Combining these two elements, we get the following optimal linear estimate:

$$\begin{aligned} m(t+1) = &[P_{11}(t)-P_{12}(t)]m(t) + P_{12}(t) \\ &+ [\sigma^2+(a_1-a_2)^2\Gamma(t)]^{-1}[P_{11}(t)-P_{12}(t)]\Gamma(t)[a_1-a_2] \\ &* [X_{t+1}-(1+\alpha)X_t-(a_1-a_2)m(t)-a_2] \end{aligned} \tag{27}$$

$$\begin{aligned} \Gamma(t+1) = &[P_{11}(t)-P_{12}(t)]^2\Gamma(t) + q_{11}(t) \\ &- [\sigma^2+(a_1-a_2)^2\Gamma(t)]^{-1}[P_{11}(t)-P_{12}(t)]^2[a_1-a_2]^2\Gamma(t)^2 \end{aligned} \tag{28}$$

where $q_{11}(t)$ is defined by equation (18)

$m(0)$ and $\Gamma(0)$ are obtained by equations (21) and (22).

Then, the optimal linear estimate of θ_t has the form

$$\hat{\theta}_t = (a_1-a_2)m(t) + a_2 \qquad (29)$$

3.3 Interpretation of The Filter

Note that the filter for $\pi_1(t)$ given in (27) is composed of three building blocks and is of the form:

$$m(t+1) = A[m(t)] + [K]\mu$$

The first block is the estimate of $m(t+1)$ given the past information, i.e. $A[m(t)]$ is given by $[P_{11}(t)-P_{12}(t)]m(t)+P_{12}(t)$. This is the best estimate of $m(t+1)$ when no new information arrives.

The second block represents the Kalman gain, $[K]\mu$. It is interesting to elaborate a little further on this Kalman filter gain. $[K]$ is the weight of the innovation and is of the form:

$$[K] = [\sigma^2+(a_1-a_2)^2\Gamma(t)]^{-1}[P_{11}(t)-P_{12}(t)]\Gamma(t)[a_1-a_2]$$

From this expression and from equation (28) which describes the trajectory of $\Gamma(t)$, it can be easily seen that the filter gain is exogenously determined. Unlike in the non-linear filter case, the Γ path will be set once and for all when the $\Gamma(0)$ is chosen a priori. The same is true, then, for [K].

This weight is, in turn, composed of two elements. The first one is the variance of the innovation term, $[\sigma^2+(a_1-a_2)^2\Gamma(t)]$. As the signal gets noisy, this term becomes large, thus assigning less weight to the new signal. This variance term itself can be decomposed into two parts: one being the variance of the Gaussian white noise δ and the other being the variance of one period ahead estimate errors. In other words, as the data becomes noisy and/or the magnitude of update of the estimates becomes large, the filter will give less weight to the new information arriving in each time period. The second element of the weight is the correlation between the unobserved process and the innovation fraction of the new information. The stronger the correlation between the signal and the parameter, the larger is this term and, therefore, the bigger is the weight assigned to the new signal. The Kalman filter, then, can be described as a mechanism which weighs the information according to its prediction power.

μ represents the innovation term. When there is no new information, $[K]\mu$ will vanish and the system will not update its estimate, except for the part that would be done by the transition probabilities. On the other hand, if a big shock is given to the system, μ becomes large and, as a result, the filter will update its estimate by a big amount.

3.4 Stationarity

It is very important to note that, unlike in the case of non-linear filter, the trajectory of m(t) is always stationary.

Given the estimate:

$$A[m(t)] = [P_{11}(t)-P_{12}(t)]m(t) + P_{12}(t),$$

m(t+1) will be forced back into the boundary when m(t) becomes less than 0 or greater than 1 since $[P_{11}(t)-P_{12}(t)]$ is smaller than 1.

The Kalman gain can be either negative or positive depending on the sign of the innovation and therefore, it will not affect the stationarity of the trajectory of π_t. In conclusion, then, the combined effect of the two components will make the trajectory stationary.

IV. Estimation -- Monthly Standard & Poor's 500

In this section, the probability π_t will be estimated using monthly data from Standard & Poor's 500 index.[6] The sample includes 490 observations, from Jan. 1947 to Oct. 1987. In order to eliminate the effect of inflation, the stock price index is deflated by the Consumer Price Index for the same time period.[7] The logarithmic first order difference of the observations was taken in order to get the monthly stock price returns. (See Graph 4-2.)

The estimation of the non-linear and linear models is not an easy task since it involves a great deal of prior knowledge. In the non-linear case, a_1, a_2, α, π_0, σ, δ_{11} and δ_{12} are the priors required. In the Kalman filter case, a_1, a_2, α, $m(0)$, $\Gamma(0)$, σ, and transition probabilities are the priors to be known by the analyst. In this application, empirical priors will be used. Section 4.1 describes the estimation procedures adopted to compute the empirical priors of both models.

---

[6] Data from CITIBASE.

[7] Data from CITIBASE.

In addition, the observations are divided into two groups: one including 389 observations and the other including the remaining 100. From the first group of data, the priors are estimated. These empirical priors are then applied in the filtering process of the remaining 100 observations. The results of the estimation of priors are shown in Table 1.

4.1 Estimation of Priors

4.1.1 Trend Dummy

In deciding whether the market is in an upward or downward trend, a long-term investment perspective is considered rather than a short-term speculative one. This is shown in Graph 4-1 where the duration of each trend is specified. A dummy variable with a value of 1 for the upward trend and 0 for the downward trend is assigned to the stock prices. This choice is plausible given that we are dealing with probabilities which should range between 0 and 1.

4.1.2 Estimation of a_i

a_1 and a_2 were estimated by calculating the average growth rate of the upward and downward trends, respectively. The 221 observations falling into the upward trend category provided an average growth rate of 0.99%. The remaining 168 observations falling into the downward trend yielded an average of -0.81%. In order to determine whether they were significantly different from each other, a t-statistics was applied to the null hypothesis.

H_0: $a_1 = a_2$.

$$t = \frac{\hat{a}_1 - \hat{a}_2}{S((1/N_1+1/N_2)^{\wedge}\tfrac{1}{2})}$$

where N_1 = the number of the observations falling into the upward trend

N_2 = the number of the observations falling into the downward trend

$$S^2 = \frac{(N_1-1)\mathrm{var}(a_1) + (N_2-1)\mathrm{var}(a_2)}{N_1 + N_2 - 2}$$

Given a t-statistics value of 8.08, the null hypothesis was rejected at a significance level over 99%.

4.1.3 Initial Point A-Priori

The estimate of $\pi_0 = m(0) = 0.57$ and was obtained by computing the ratio 221/389. $\Gamma(0)$ was estimated according to the formula given in (15) and turned out to be 0.25.

4.1.4 Estimation of α and σ

α and σ were estimated using the following equation with the $\hat{a}_i$'s estimated in 4.1.2.

$$X_{t+1} = \hat{a}_i + (1+\alpha)X_t + w_t$$

The value of $\hat{\alpha} = -0.91$ and was statistically significant. Since it is negative, it satisfied the condition for the convergence of the model. σ was estimated using the residuals from the above regression, i.e. $\hat{w}$'s.

4.1.5 δ_{ij}'s: Duration Dependence

For the non-linear case, it was assumed that the function $f_{ij}(t-s)$ could take one of two alternative forms: linear and log-linear.

Linear form:

$$f_{ij} = \delta_{ij}*(t-s), \qquad \delta f_{ij}/\delta t = \delta_{ij}$$

Log-linear form:

$$f_{ij} = \delta_{ij}*\ln(t-s+1), \qquad \delta f_{ij}/\delta t = \delta_{ij}/(t-s+1) = \delta_{ij}/2$$

Note that the continuous formula was discretized so that (t-s) became 1, and, therefore, (t-s+1) became 2.

In order to estimate δ_{11} and δ_{12}, $P_{11}(t-s)$ and $P_{12}(t-s)$ were calculated first from the trend dummy variable which was assigned in 4.1.1. In the linear formulation of $P_{ij}(t-s)$, (t-s) was allowed to vary from 1 to 15. Then, δ_{11} and δ_{12} were estimated as -0.0359 and 0.0505, respectively, by running the following regressions.

$$P_{11}(t-s) = 1 + \delta_{11}*(t-s) + e_1$$
$$P_{12}(t-s) = 0 + \delta_{12}*(t-s) + e_2$$

In the log-linear formulation of $P_{ij}(t-s)$, $(t-s)$ was allowed to vary from 1 to 50. δ_{11} and δ_{12} were estimated as -0.1077 and 0.1197, respectively, through the following regressions.

$P_{11}(t-s) = 1 + \delta_{11}*\ln(t-s+1) + e_1$

$P_{12}(t-s) = 0 + \delta_{12}*\ln(t-s+1) + e_2$

Then, $\delta_{11}/2$ and $\delta_{12}/2$ turned out to be -0.0539 and 0.0749, respectively.

The estimates from both the linear and the log-linear formulation satisfy the conditions given above by equation (3). The transition probabilities are shown in Graph 4-3.

4.1.6 Transition Probabilities

For the Kalman filter case, the transition probabilities, $P_{11}(t)$ and $P_{12}(t)$ were estimated as follows:

$$P_{ij}(t) = \frac{\sum_{l=1}^{100} d(\theta_{t-l}=a_i \mid \theta_{t-l-1}=a_j)}{\sum_{l=1}^{100} d(\theta_{t-l-1}=a_j)}$$

where $d(\cdot)$ is a dummy variable which takes the value of either 0 or 1.

Table 1. Empirical Priors

Number of Observations: 389

Variable	# of obs	Estimate	sd(•)	t-stat
a_1 (uptrend)	221	0.0099	0.0004	24.75
a_2 (downtrend)	168	-0.0081	0.0006	-13.50
H_0: $a_1 = a_2$				8.0761
α	389	-0.9051	0.0466	-19.4245
$\pi_0 = m(0)$	389	0.5681		
σ^2	388	0.0010		
Transition Probability:				
Linear to (t-s):				
δ_{11}	15	-0.0359		
δ_{12}	15	0.0505		
Log-linear to (t-s):				
$\delta_{11}/2$	100	-0.0539		
$\delta_{12}/2$	100	0.0749		

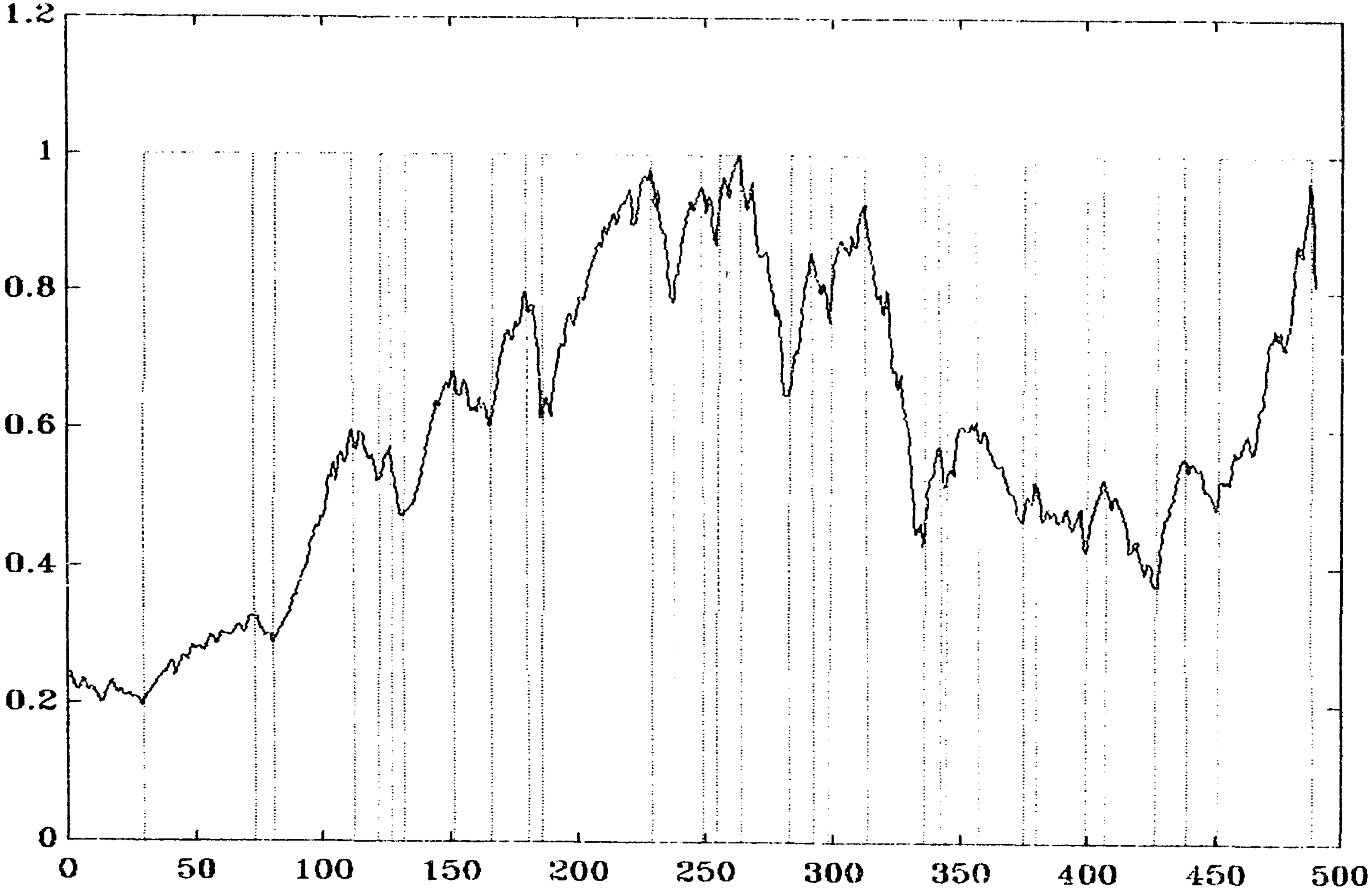
4.2 Estimation of The Parameters

Using the empirical priors, the non-linear and the linear filter were developed and the probabilities were estimated from the observed series $\{X_t\}$. These estimates are shown in the Graphs 4-4 through 4-8. Graph 4-4 shows the estimates of the probability using the non-linear filter with the transition probabilities being linear in (t-s). Graph 4-5 shows the estimates of the probability using still the non-linear filter but with the transition probabilities being log-linear in (t-s+1). Graph 4-6 compares the two non-linear filters. Graph 4-7 illustrates the estimates computed by using the linear filter - the optimal Kalman filter. Finally, Graph 4-8 compares the non-linear filter and the optimal Kalman filter, i.e. Graph 4-4 and 4-7.

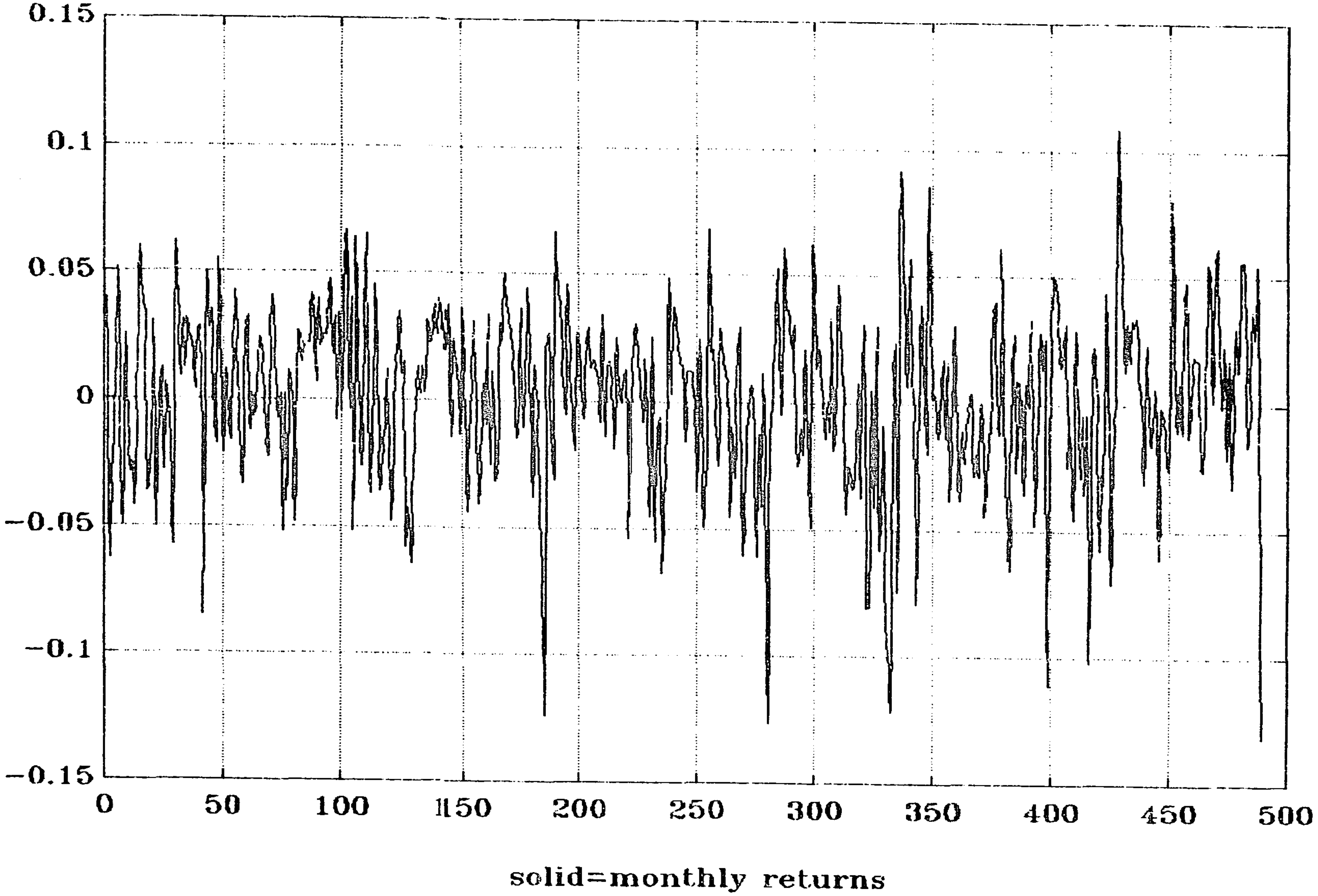
Note that some of the probability estimates are either less than 0 or greater than 1. These results are not acceptable as a probability measures. In the continuous time formulation this would not happen. However, due to the errors arising from the approximation, it happens in discrete time and there is nothing that can be done to prevent it.

It is also to be noted that both the filters turned out to be stationary. In other words, the non-linear filter did not diverge. This is attributable to the fact that the time series has monthly intervals and, therefore, it is less noisy and has less outliers.

GRAPH 4–1. S&P 500 INDEX, TREND

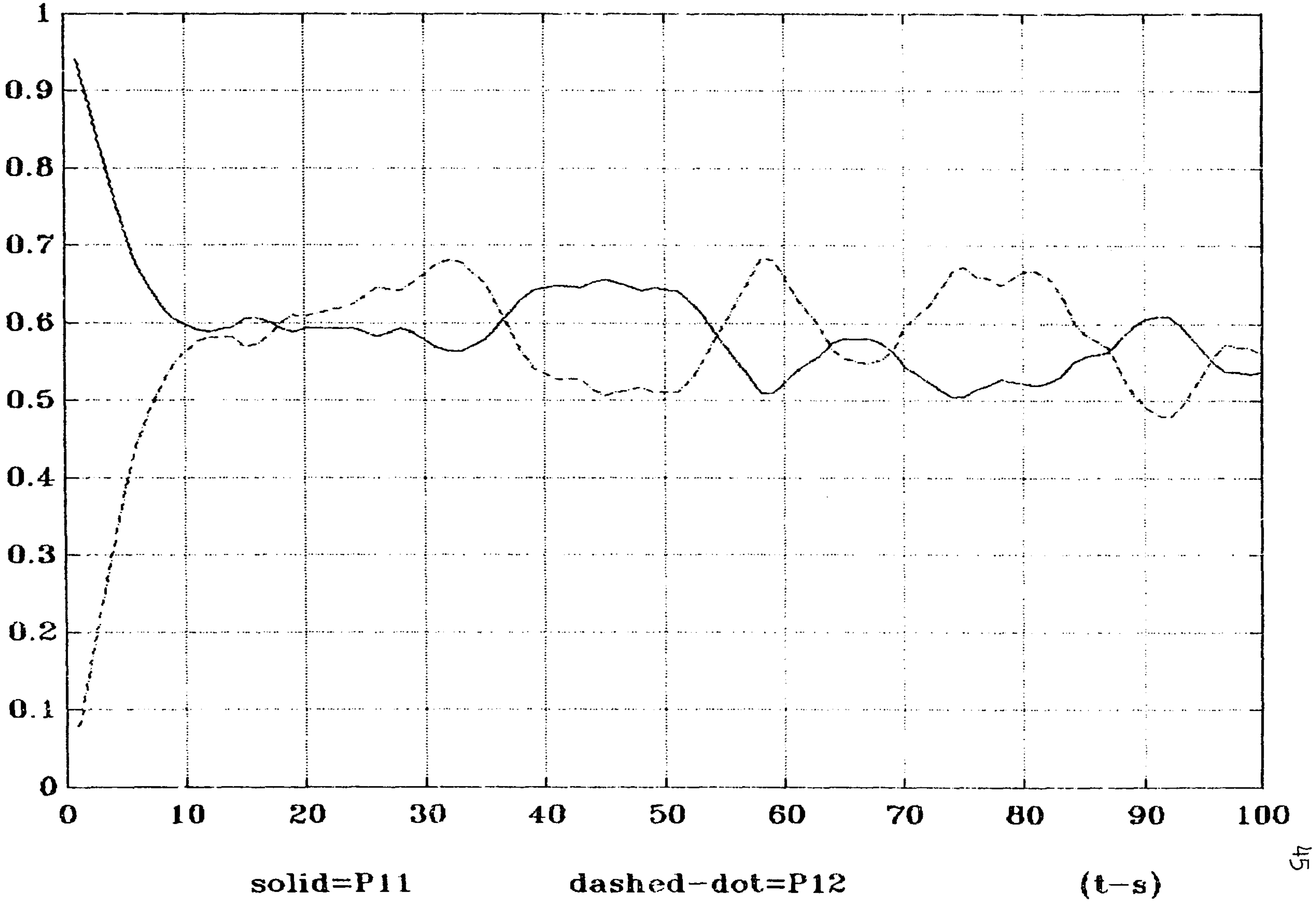


solid=S&P 500 dotted=trend Jan. 1947 – Oct. 1987

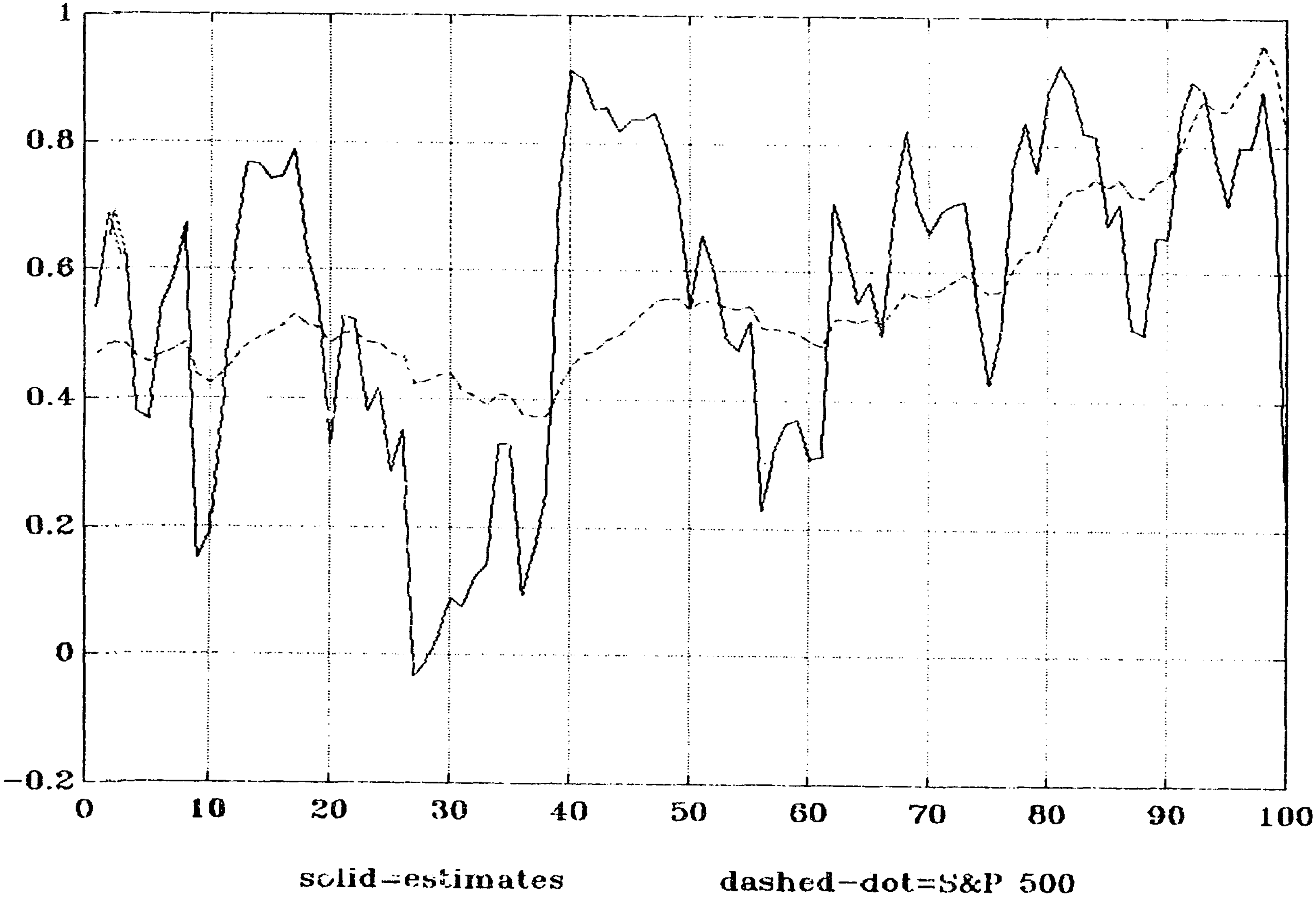


GRAPH 4-2. S&P 500: MONTHLY RETURNS

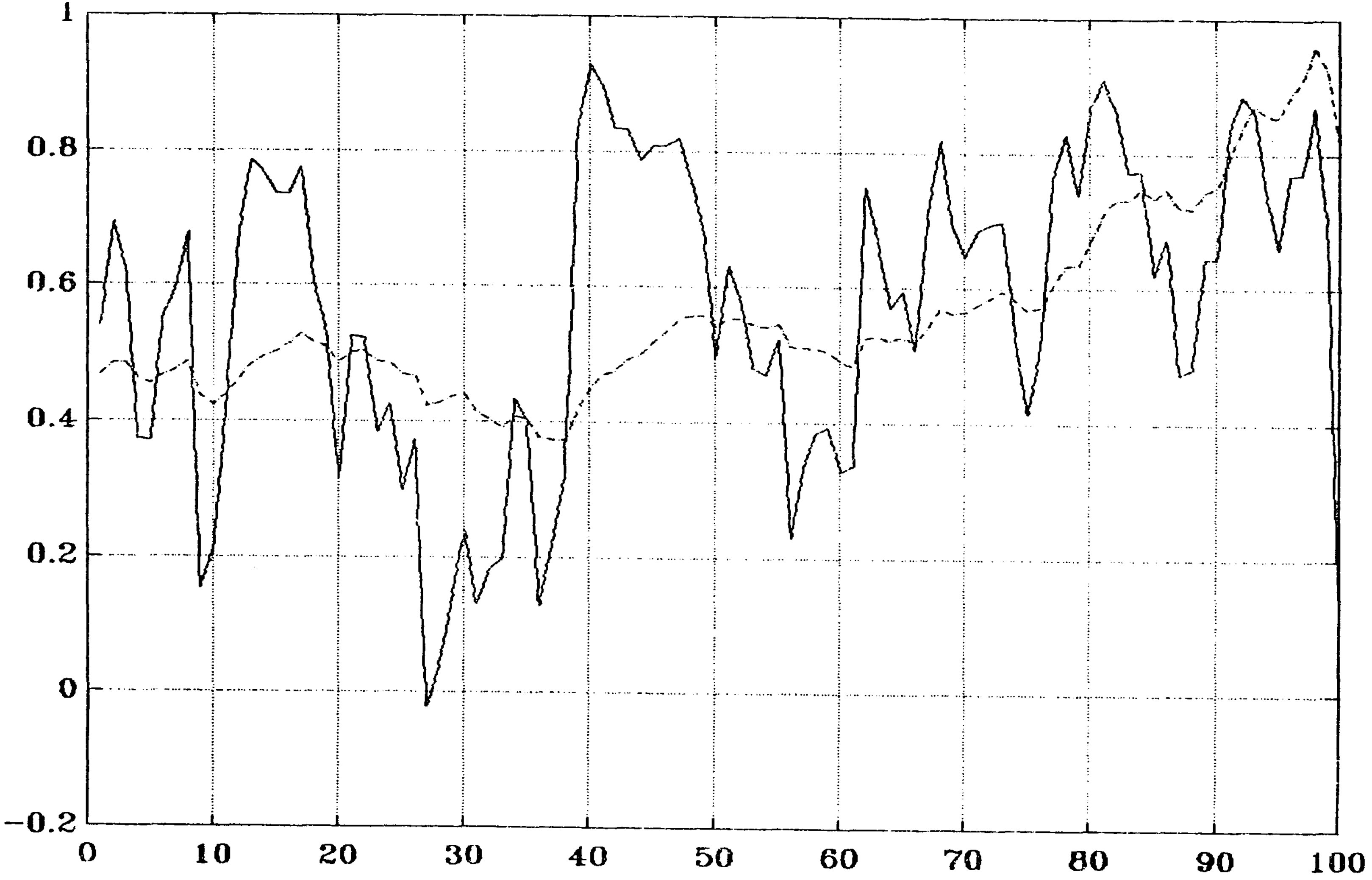
GRAPH 4-3. S&P 500: TRANSITION PROBABILITIES



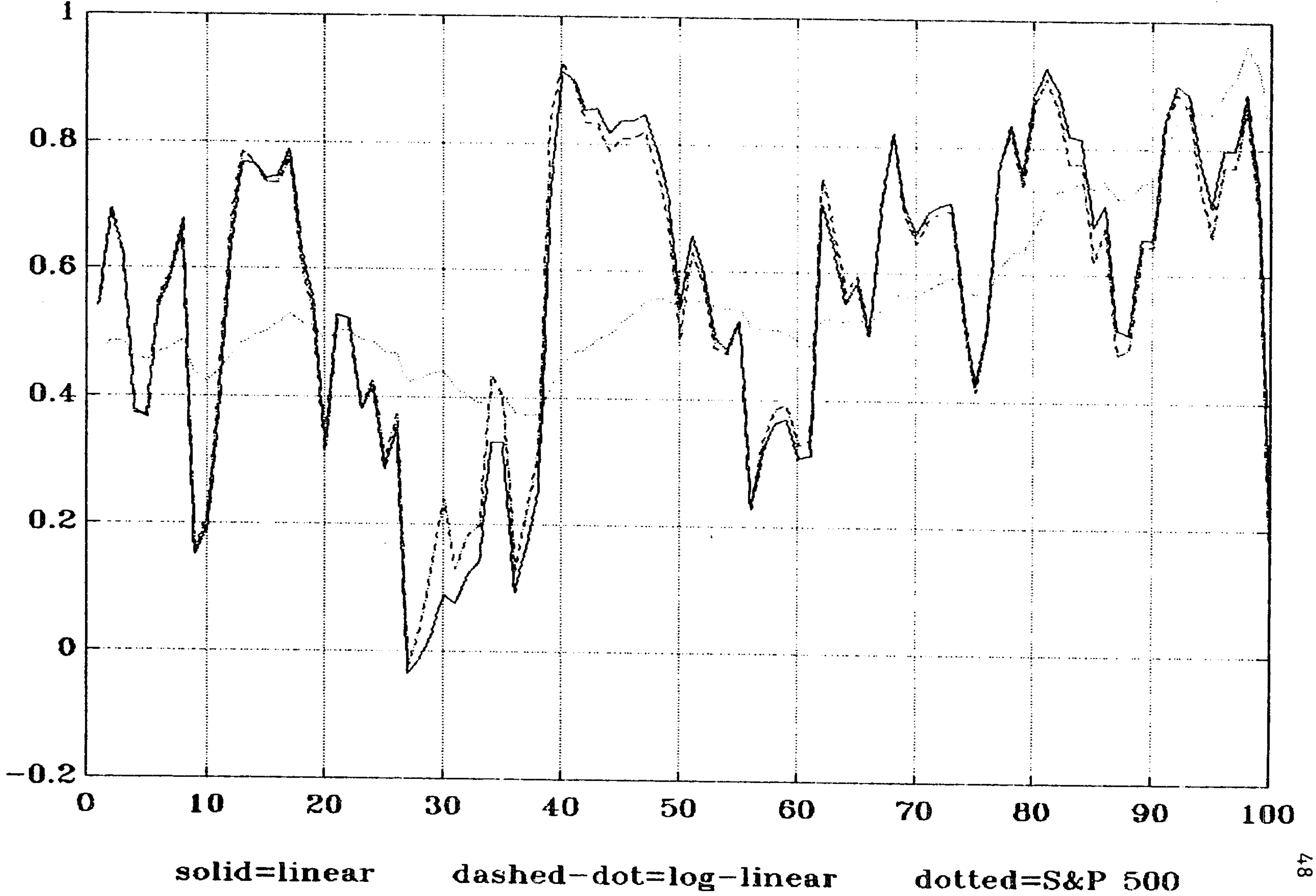
GRAPH 4-4. S&P 500: NON-LINEAR FILTER, LINEAR TRANS. PROB.



GRAPH 4-5. S&P 500: NON-LINEAR FILTER, LOG-LINEAR TRANS. PROB.

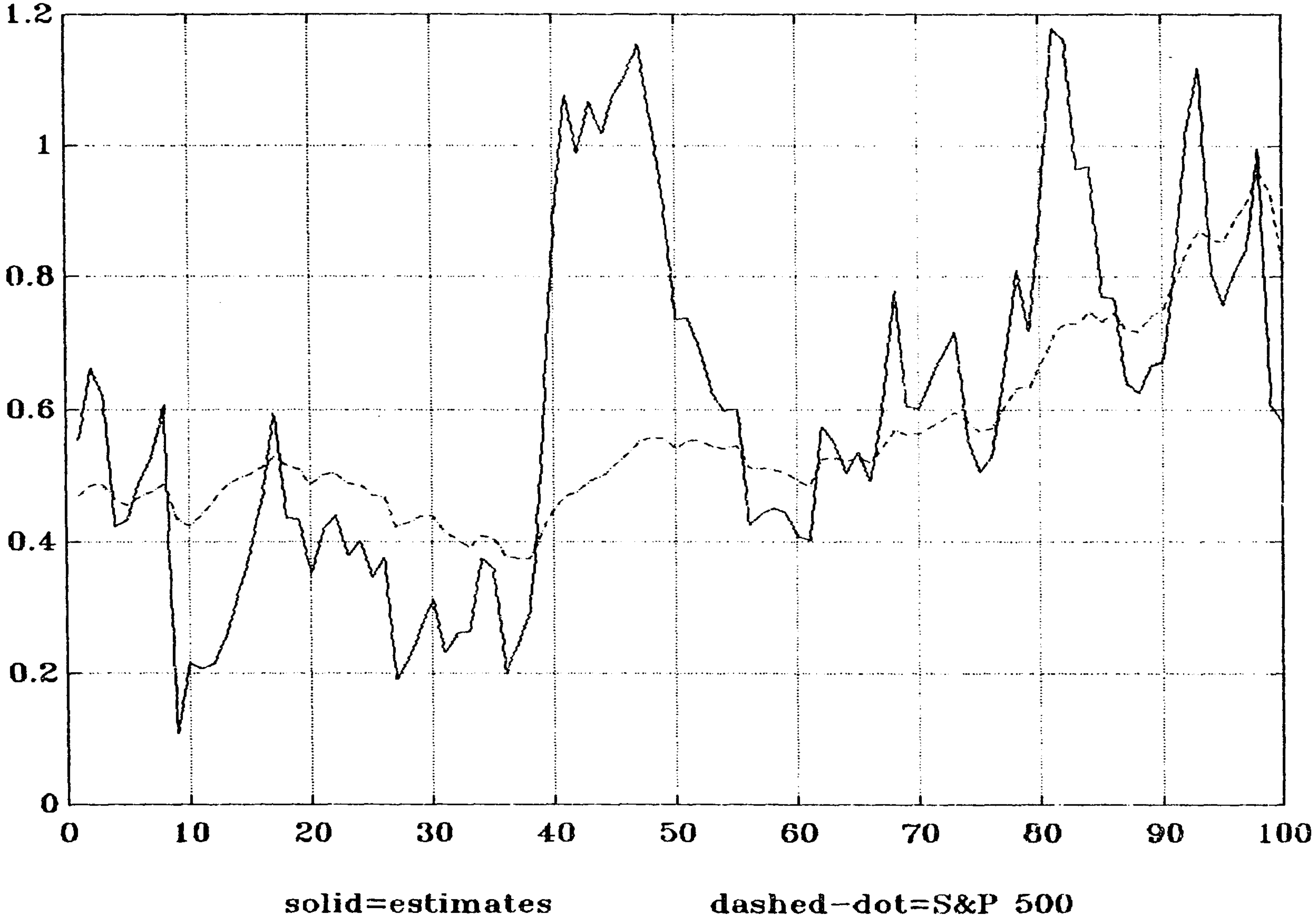


solid=estimates dashed-dot=S&P 500

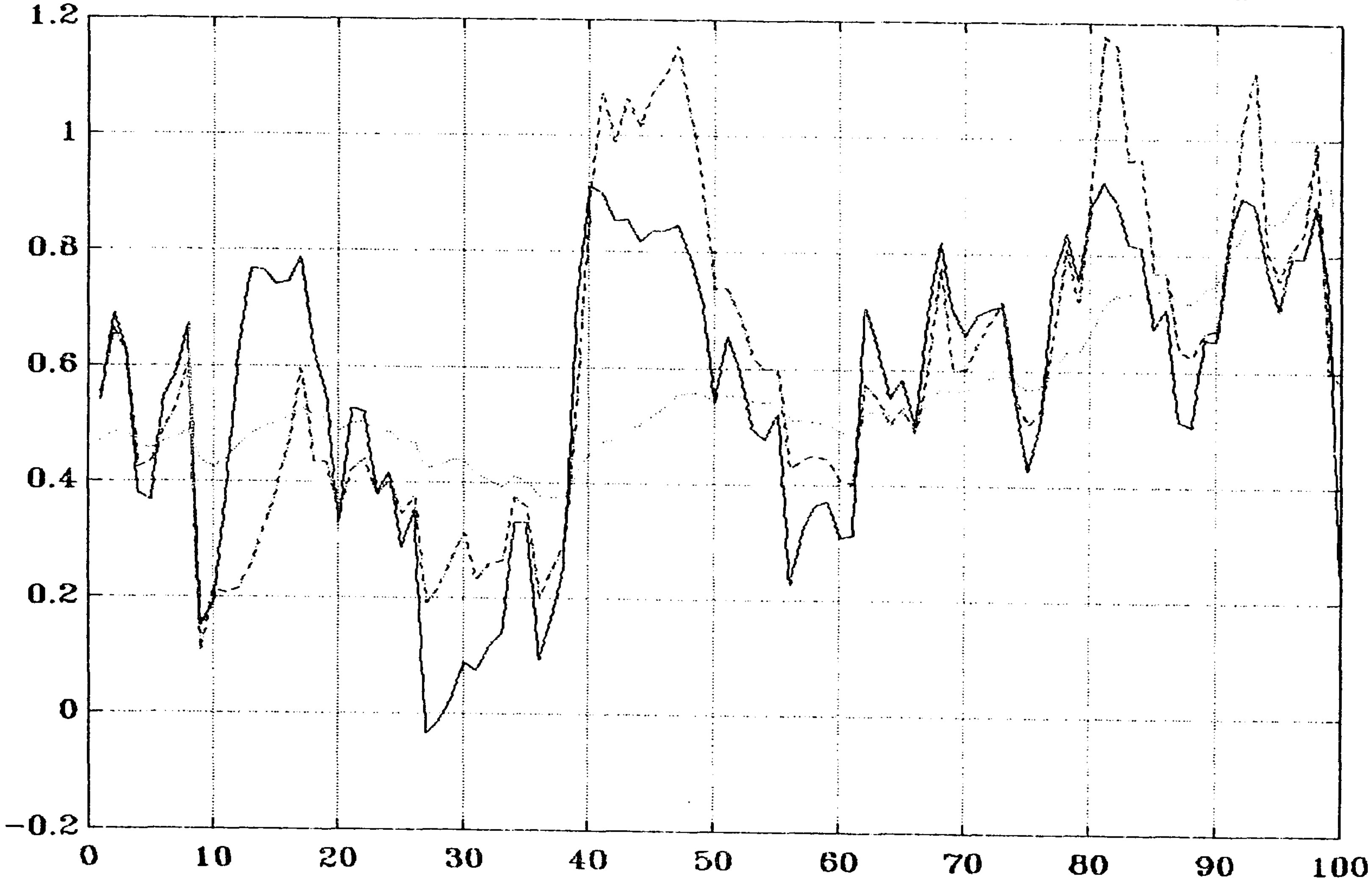


GRAPH 4-6. S&P 500: NON-LINEAR FILTER, LINEAR vs. LOG-LINEAR

GRAPH 4-7. S&P 500: KALMAN FILTER



GRAPH 4-8. S&P 500: NON-LINEAR FILTER vs. KALMAN FILTER



solid=non-linear dashed-dot=Kalman dotted=S&P 500

V. Daily Korean Composite Stock Price Index

This section estimates the parameters using the daily Korean Composite Stock Price Index. The data set consists of 1,353 observations from January 1986 to August 1990.[8] Since a daily deflator index is not available, no attempt has been made to eliminate the inflation effect from the stock price index. Once again, the logarithmic first order difference of the observations is taken in order to get the daily stock price returns. (See Graph 5-1 and 5-2 for the index and the daily returns, respectively.)

In this section, both empirical priors and a-priori knowledge were used in the estimation of the parameters. The procedure for the estimation of empirical priors is exactly the same as the one adopted for the S&P 500 case and, therefore, is not described here. Note that the 1,352 observations were divided into two equal size groups. The first one was used in the estimation of

---

[8] Data from "Korean Compsite Stock Price Index", Korea Stock Exchange.

the empirical priors and the second one in the estimation of the parameters.

$\pi_0 = m(0)$ were selected a priori to be 0.2. $\Gamma(0)$ was estimated according to the formula given in (22). For α, both empirical and a-priori selections were made and the results were then compared. In the estimation of δ_{11} and δ_{12}, only the linear formulation of P_{ij}'s was adopted.

5.1 Non-Linear Filter with Empirical α

When estimating the parameters using the empirical α, the same procedure as in the S&P case was followed. However, regardless of how the duration of the upward and downward trends was defined, the non-linear filter estimates always diverged to either $-\infty$ or $+\infty$. This outcome can be attributed to the presence of outliers in the daily stock price returns. In order to solve this problem, three alternative approaches were considered: (1) smooth the series by applying a moving average, (2) adopt a larger variance, σ^2, and (3) truncate the innovation.

5.1.1 Moving Average

First, the smoothing of the time series was attempted as a solution to the divergence problem. Weekly (6-day), bi-weekly (12-day), tri-weekly (18-day), and monthly (24-day) moving averages are calculated from the original series, with the understanding that the longer the time span is, the smoother the series will be. [9]

This type of solution has a deficiency in itself, however. Even though it might solve the problem of the outliers, it creates the problem of delays in timing. Specifically, the moving average will pick up the change in the original trend only after half of the time span has elapsed. For example, both peaks and troughs of a 6-day moving average will occur 3 days after those of the original series. Of course, the longer the time span is, the longer the delay will be. [10] In the financial market, time is of the essence and it might well be too late if the filter detects a new trend several days after its onset.

---

9. The number of working days in a week in Korea is 6 days.

10. This assumes, for ease of understanding, that the series is symmetric. However, it is generally believed that this is not the case.

When the non-linear filter was applied to these various moving averages, surprisingly enough, the filter estimates still diverged without an exception. Needless to say, this smoothing technique is not effective at all, and the reason becomes quite obvious after looking at the filter. The model filters the new information using the following filter gain:

$$\pi_t(1-\pi_t)(\text{innovation})/\sigma^2$$

As the series becomes smoother, the innovation also becomes smooth, and, therefore, σ becomes smaller proportionately. In other words, both the numerator and the denominator are scaled down in the same proportion and, therefore, their ratio $(\text{innovation})/\sigma^2$ remains unchanged. Consequently, this alternative is abandoned.

5.1.2 Adoption of A Larger Variance

In the second attempt, a larger variance, σ^2, became the candidate for a possible solution. If a large enough σ^2 is selected, their ratio $(\text{innovation})/\sigma^2$ will be scaled down and the estimate will not move out of the boundary. Then, the problem of divergence would be solved. However, soon after this candidate was considered, it became clear that the whole filter would become useless if such a scheme were deployed. In fact, if a very large variance is selected, the filter scales down not just the outliers but also the normal-size signals. Hence, the filter will take information only from the outliers and will ignore the normal-size signals, even though most of the useful information rests with the latter. This conclusion left only one alternative, namely the truncation of the innovation.

Yet, it should be noted that a larger σ^2 was sought for another reason as well. The Korean Composite Stock Price Index is very noisy, even after the noisiness of daily series is taken account. There are many reasons for such noisiness: the recent liberalization of the financial market, the political instability both within the country and with respect to North Korea, the dependence of the economy on international trade, and so on. In light of this situation, a larger variance was computed and incorporated as part of the third approach.

This larger variance was derived through the robustification of α estimate. The α estimate was robustized by $\pm 1.5\sigma$ until the 14th decimal digit convergence for both α and σ was achieved. Then, a new σ estimate was calculated using the original series and the robust α estimate. This yielded a new σ estimate nearly twice as large as the original.

5.1.3 Truncation of Innovations

The third approach to the outlier treatment was the truncation of innovations. The innovations, v_t, were metrically winsorized as follows:

$$\begin{aligned} v_t^* &= v_t && \text{if } |v_t| \le 2\sigma \\ &= -2\sigma && \text{if } v_t \le -2\sigma \\ &= 2\sigma && \text{if } v_t \ge 2\sigma \end{aligned}$$

The advantage of this type of treatment is that the filter is still sensitive to gradual changes while it is robust to outliers. Usually, in this model, outliers do not convey more information than normal signals do. In some instances, such as a discrete jump in stock prices, outliers convey no information at all. A perfect example would be a stock market crash which is not followed by an immediate rebound, such as the one in October 1987.

The problem which may arise from this type of mathematical treatment is that, by the recurrence equation of π_t, the path for the estimate of the parameter might be different from what it would be without the treatment. In order to test for this possibility, a large σ was chosen arbitrarily in order to prevent the estimate from diverging. Using this σ, two estimations were performed, one with v_t and the other with v_t^*. These estimates do not have any economic meaning, and are computed just for the purpose of comparison. From the comparison of the two groups of estimates, it became clear that this treatment would not make the path for the estimate be different from the one without the treatment. Suprisingly, the filter simply truncated the estimates of π_t whenever v_t was metrically winsorized. Therefore, the estimation with this outlier treatment was accepted.[11] The non-linear filter estimates are shown in Graph 5-3.

---

[11] This turned out to be the case also in the application of the Kalman filter to the DJIA, which will be discussed in the next section.

5.2 Non-Linear Filter with A-Priori α

This section estimated the parameters of the non-linear filter using a value of -2 for α. This number was selected a priori. By choosing this number, the innovation will have a memory in itself.

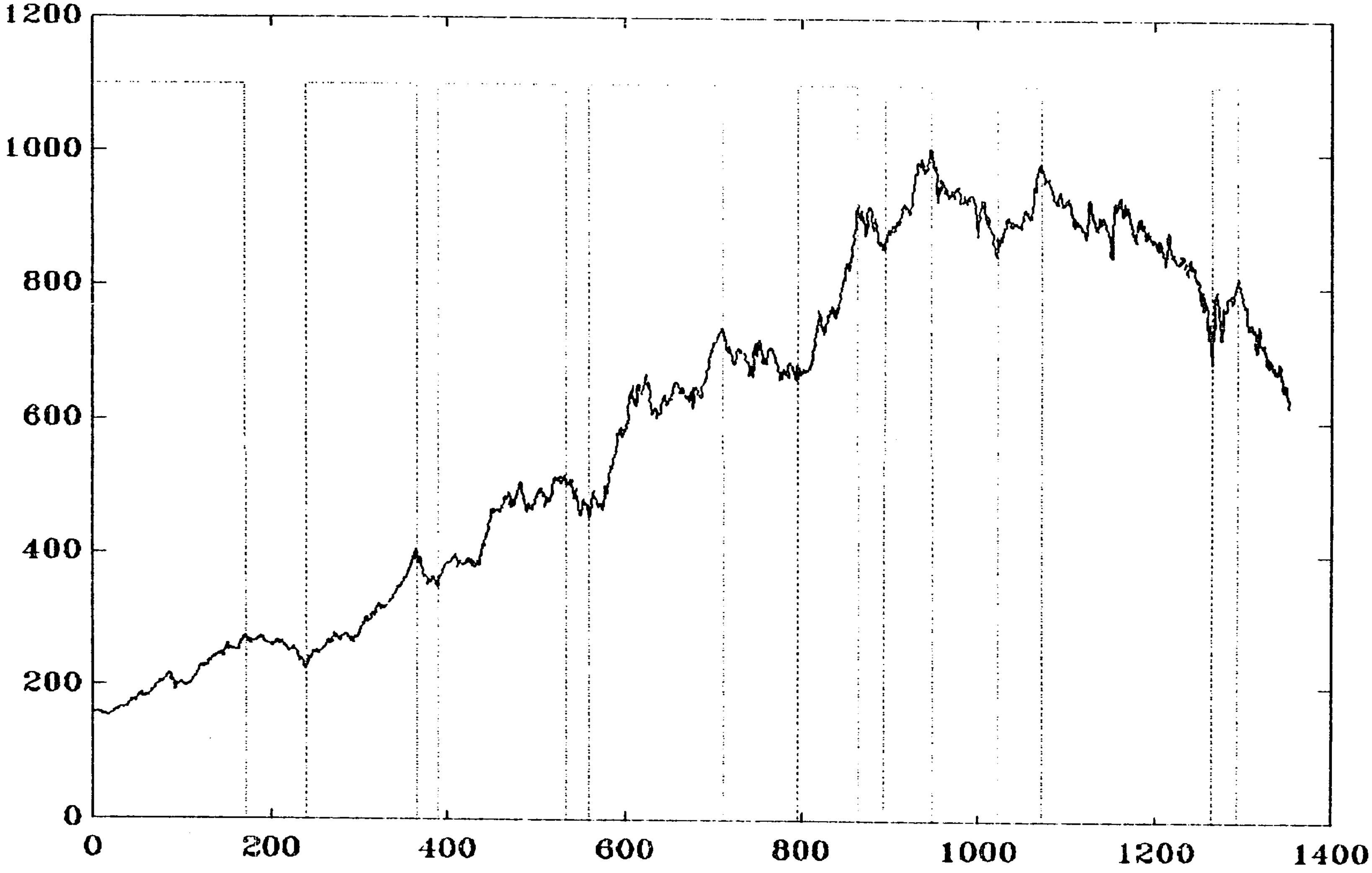
$$\begin{aligned} v_t &= X_{t+1} - (1+\alpha)X_t - \hat{a}_i \\ &= X_{t+1} + X_t - \hat{a}_i \end{aligned}$$

This configuration of innovation would give a larger weight to a trend and a smaller weight to stock price rallies. In order to see this clearly, assume that X_{t+1} and X_t have the same absolute magnitude. Then, v_t will be large in absolute value when the signs of X_{t+1} and X_t are the same and will be close to 0 when they are different. This configuration should perform better than the empirical α configuration in the sense that it will detect the onset of a new trend better and be resistant to the noise which takes its sign randomly. The estimates using an α value of -2 are shown in Graph 5-4.

5.3 Kalman Filter

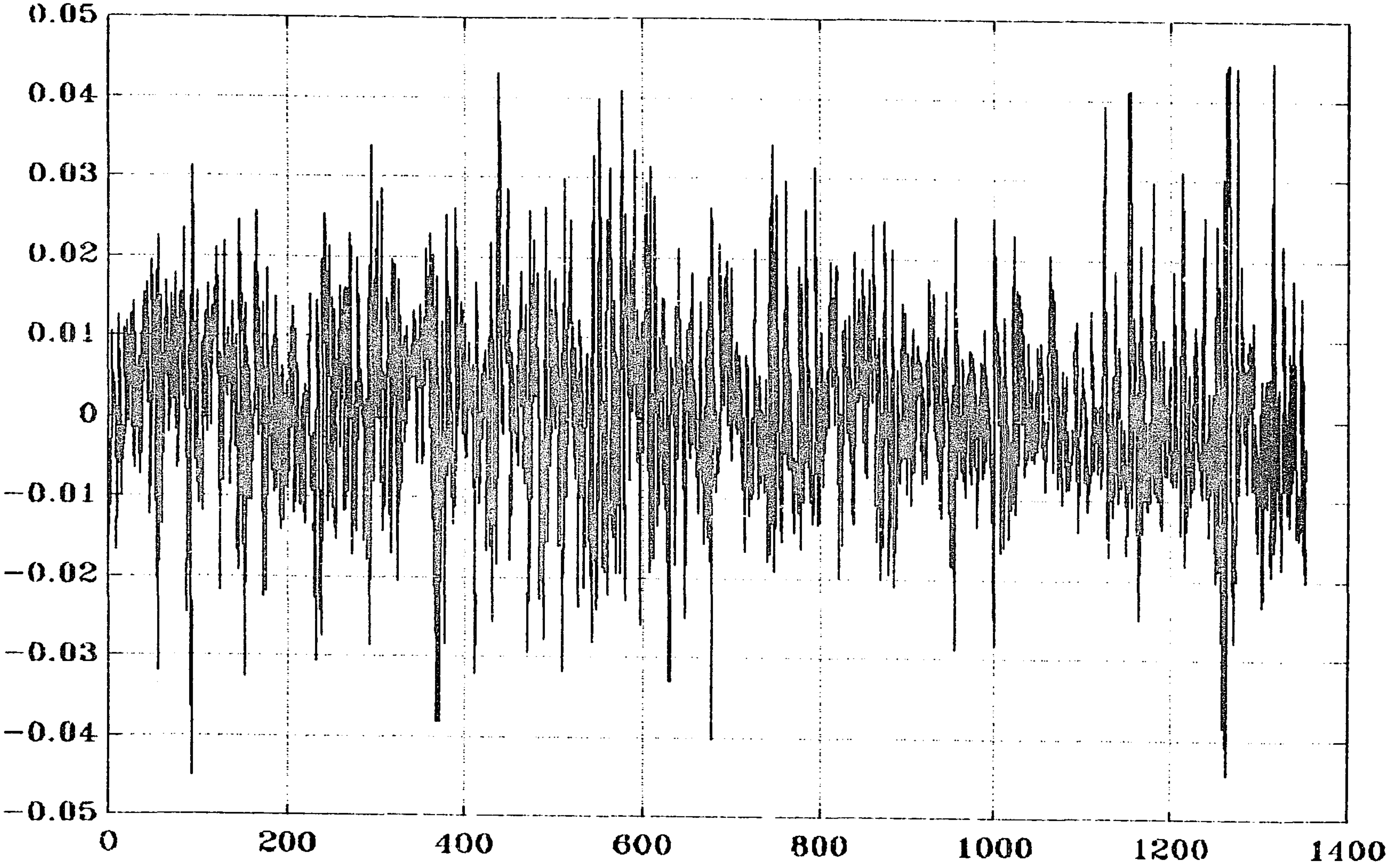
In the Kalman filter estimation, the outlier treatment was not necessary since the filter is less sensitive and is stationary. As expected, the optimal Kalman filter generated estimates which are generally acceptable, except in some instances where the estimate became greater than 1.0. The estimates are shown in Graph 5-5. In Graph 5-6, the Kalman filter estimates are compared with the non-linear filter estimates with the empirical α. Graph 5-7 illustrates the comparison between the Kalman filter and the non-linear filter with an a-priori α.

GRAPH 5-1. KCSPI INDEX, TREND



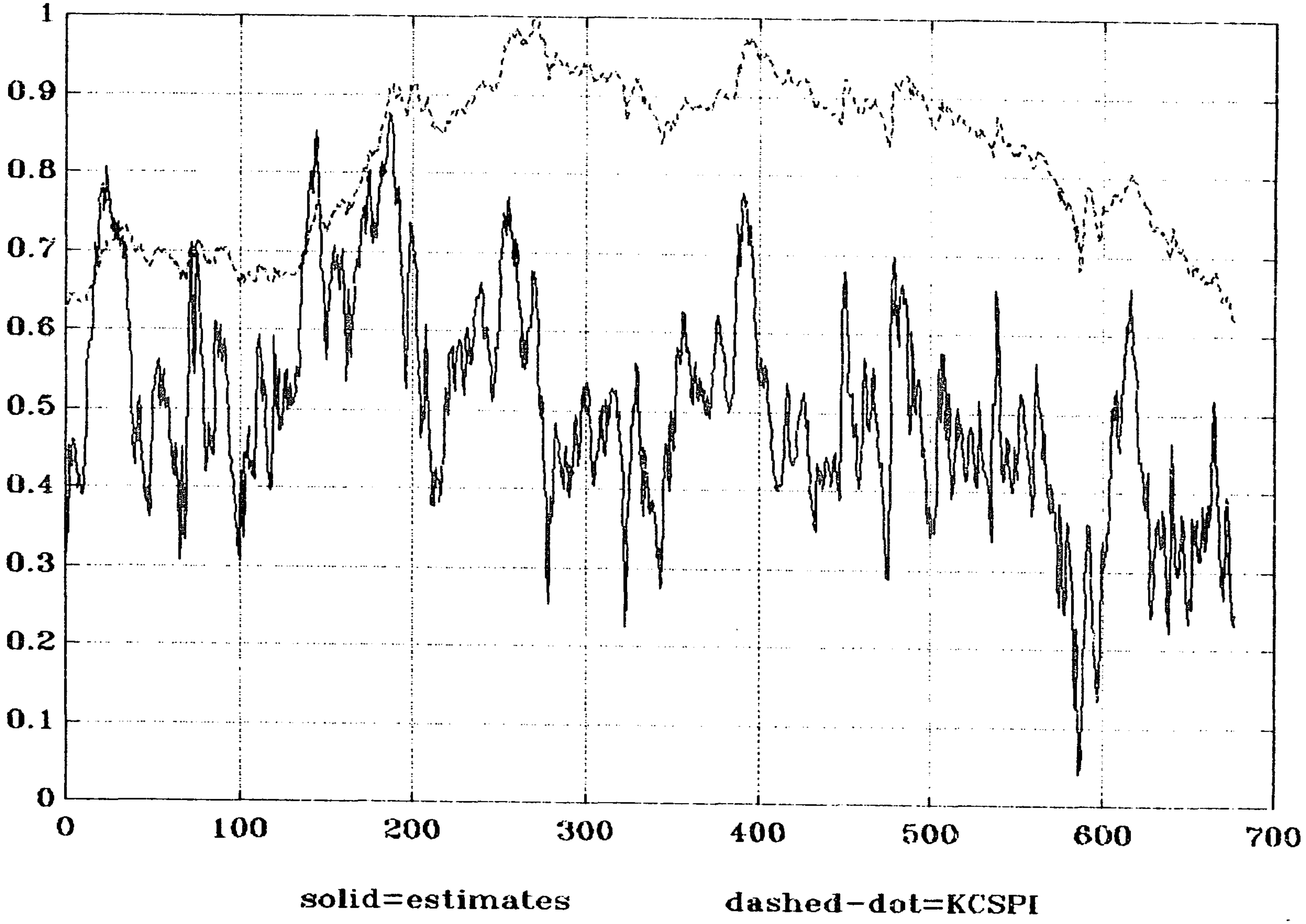
solid=KCSPI dotted=trend Jan. 1986 - Aug. 1990

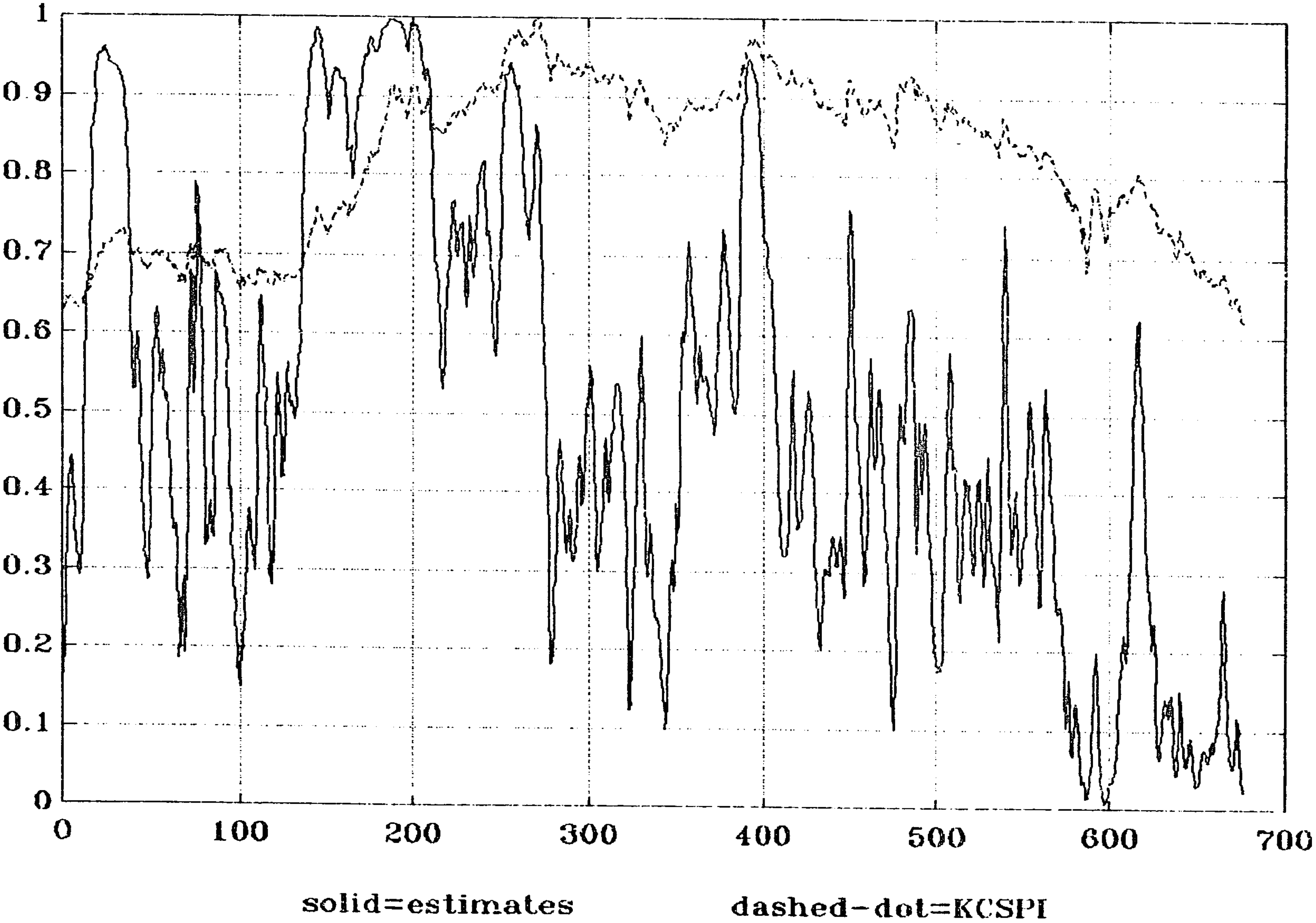
GRAPH 5-2. KCSPI: DAILY RETURNS



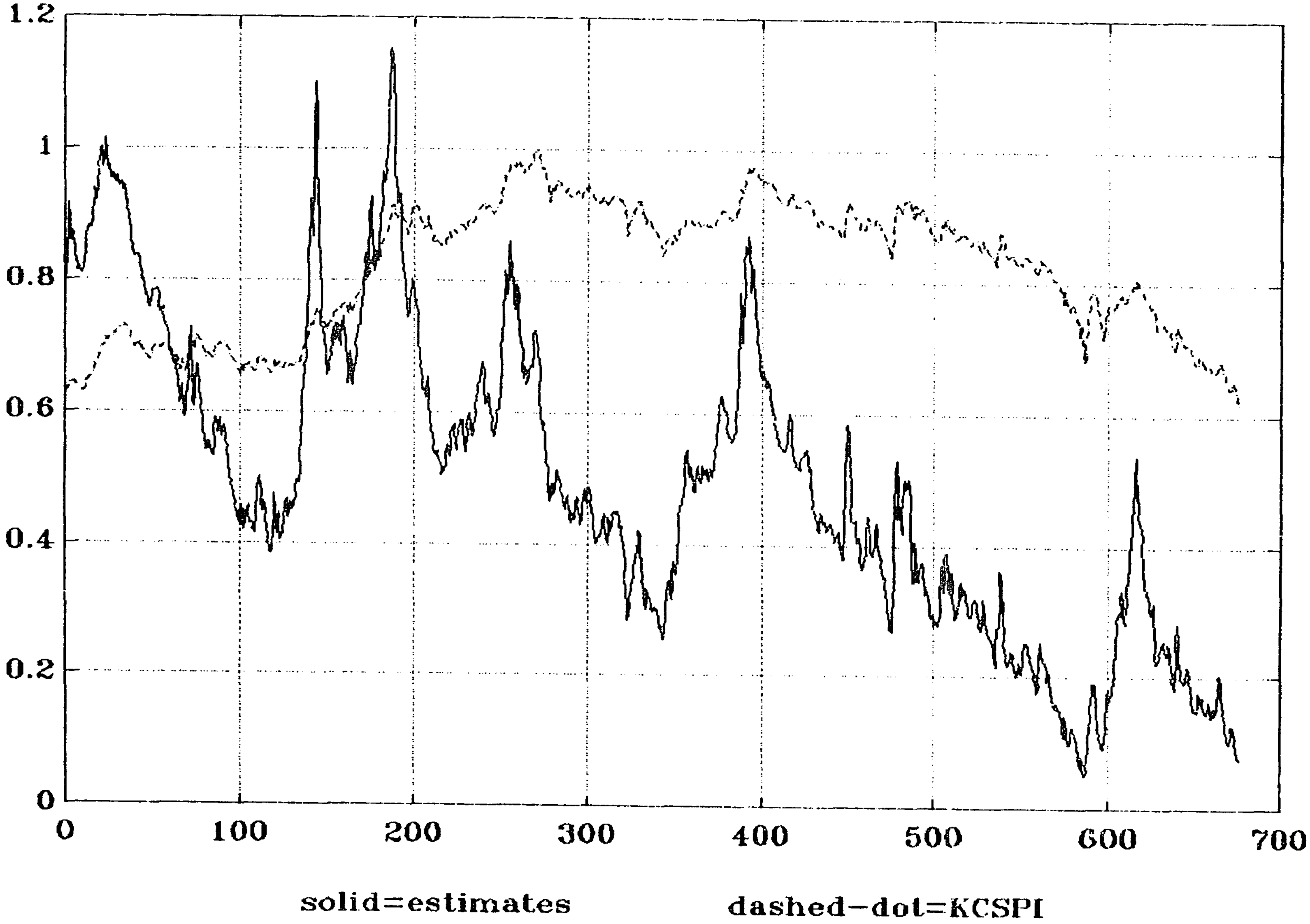
solid=daily returns

GRAPH 5-3. KCSPI: NON-LINEAR FILTER, EMPIRICAL ALPHA



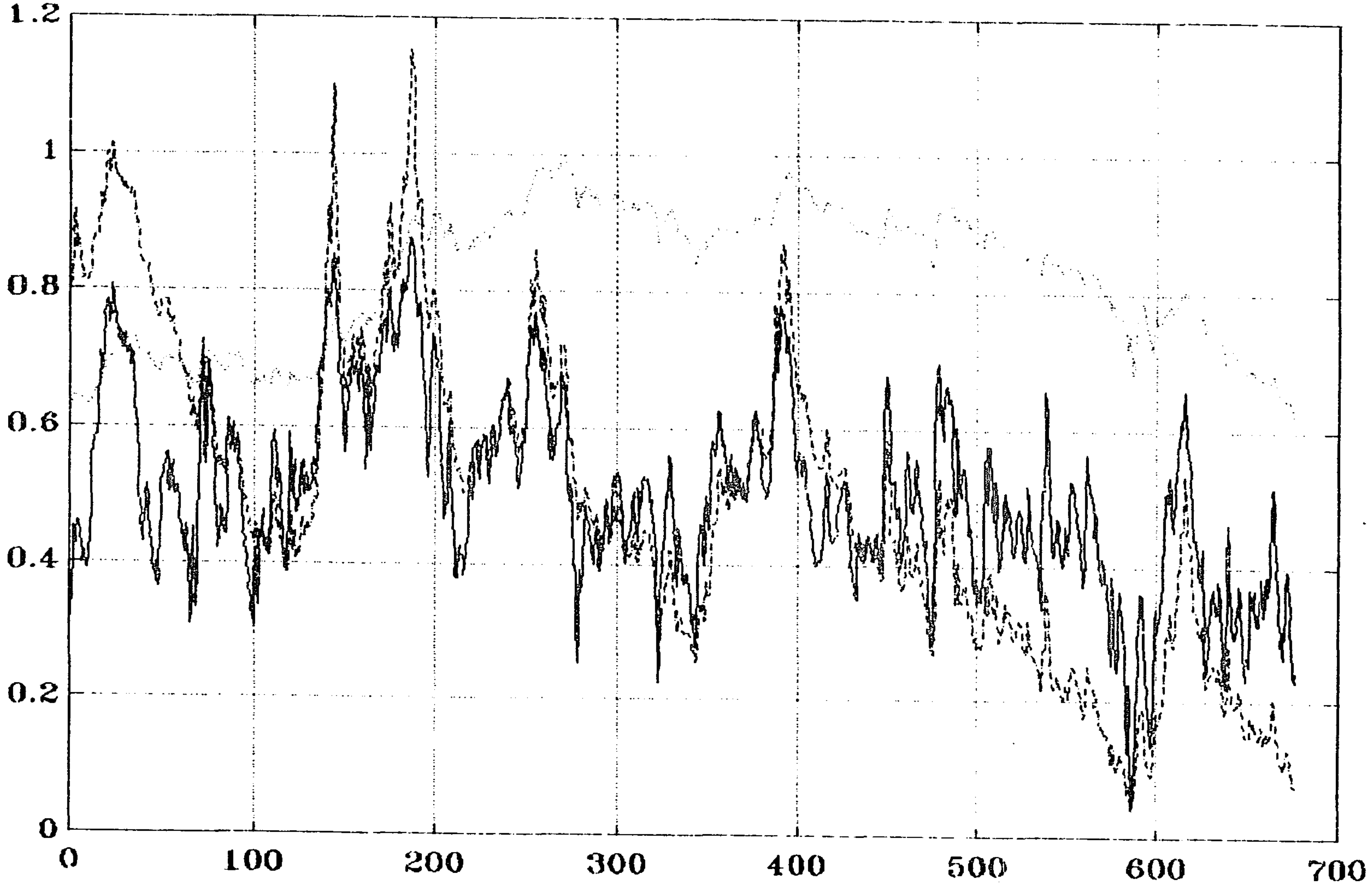


GRAPH 5-4. KCSPI: NON-LINEAR FILTER, A-PRIORI ALPHA



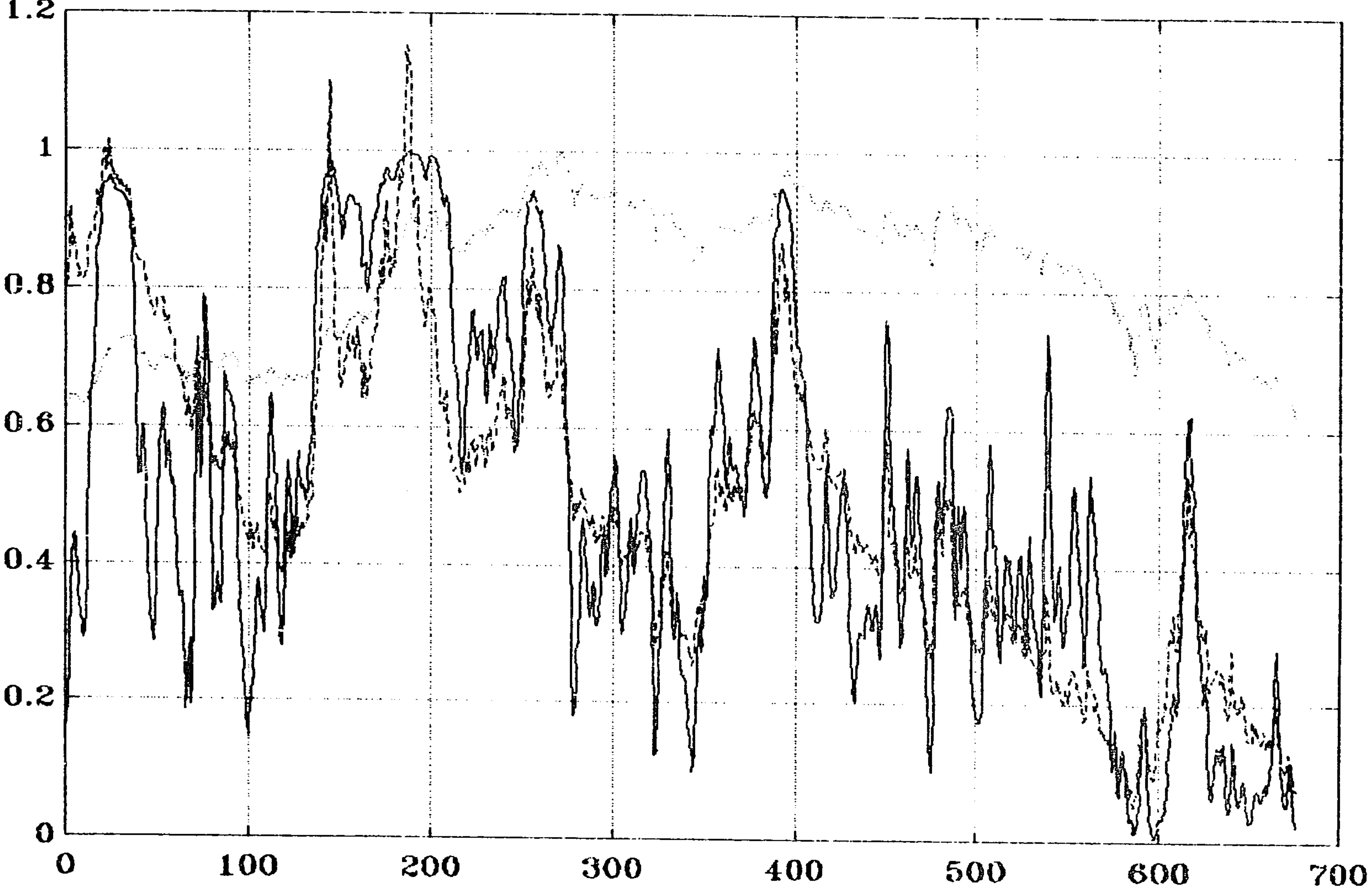
GRAPH 5-5. KCSPI: KALMAN FILTER

GRAPH 5-6. KCSPI: NON-LINEAR vs. KALMAN FILTER



solid=non-linear(empirical alpha) dashed-dot=Kalman

GRAPH 5-7. KCSPI: NON-LINEAR vs. KALMAN FILTER



solid=non-linear(a-priori alpha) dashed-dot=Kalman

VI. Daily Dow Jones Industrial Average

This section estimates the parameter using the daily Dow Jones Industrial Average. The data set consists of 1,939 observations from January 1983 to August 1990.[12] From this original series, daily returns $\{X_t\}$ were derived by taking the logarithmic first order difference. (Graph 6-1 and 6-2 show the index and the daily returns.) These observations were then grouped into two subsets, one with the first 938 observations and the other with the remaining 1000 observations. The first subset was used in the estimation of empirical priors. $\pi_0 = m(0)$ were chosen a priori to be 0.2. $\Gamma(0)$ is calculated according to the formula given by equation (15). a_1, a_2, α, and σ^2 were estimated with the procedure described in section 4 (the S&P 500 case). α and σ^2 were not robustized, the reason being that the series is not as noisy as the KCSPI.

---

[12] Data from Wall Street Journal Index.

6.1 Non-Linear Filter

It is assumed in this estimation that the transition probabilities are not duration dependent. Hence, δ_{11} and δ_{12} are equal to zero. This assumption is very important since it will adversely affect the stationarity of the model. As previously discussed, $\delta_{11}\pi_t + \delta_{12}(1-\pi_t)$ is that portion of the system which helps it to converge. Now, the trajectory of π_t no longer contains this term and, therefore, there is no guarantee that the model will generate an acceptable path for the estimate in both the continuous and discrete time versions of the model. Let us rewrite the recurrence equation:

$$\begin{aligned}\pi_{t+1} = {} & \pi_t \\ & + [\pi_t(1-\pi_t)/\sigma^2][a_1-a_2][(X_{t+1}-X_t)-(\alpha X_t+a_1\pi_t+a_2(1-\pi_t))] \\ & + [\pi_t(1-\pi_t)(1-2\pi_t)][(a_1-a_2)^2/\sigma^2]\end{aligned}$$

For now, let us ignore the correction factor. This leaves us with a continuous time version of the model. Then, the filter estimate is composed of two parts, π_t and the innovation portion. By definition, an innovation can take any value at random. Thus, the π_t estimate follows a random walk which is not acceptable due to the boundary limits of 0 and 1.

Now, the model depends only on the correction factor for its stationarity. As already discussed, the correction factor pulls the estimate toward 0.5 so long as the estimate ranges between 0 and 1. Therefore, while within the boundaries, this factor will serve as a stationary force in the model; if the estimate goes out of the boundary, however, the factor will serve as a divergent force.

For these reasons, this model is highly non-stationary. Yet, if, through the metrical winsorization of the innovations, the estimate is well bounded, the model can generate acceptable estimates.

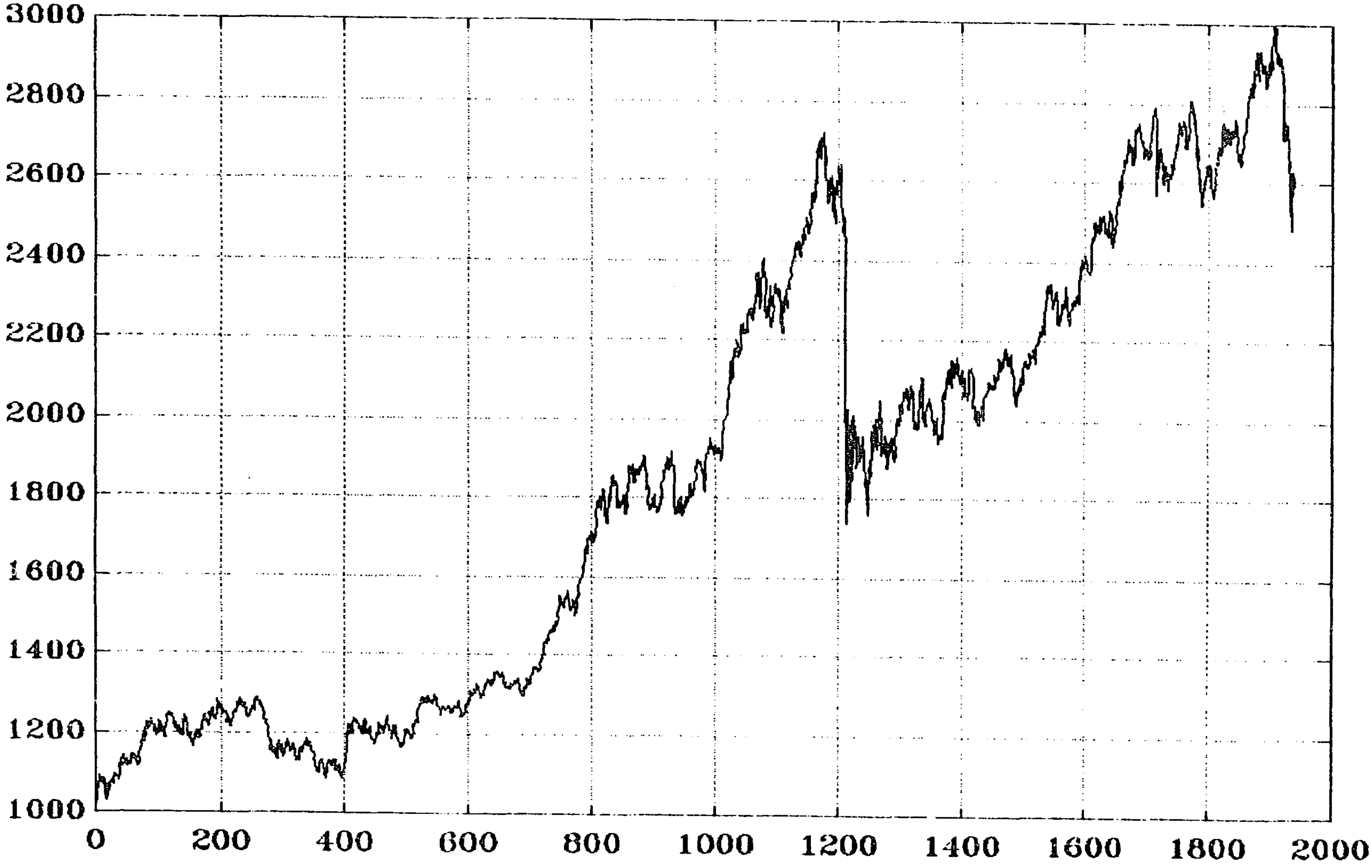
Outliers were again winsorized with the same procedure described in the previous section. This was necessary due to the fact that, unlike the monthly S&P 500 or the daily KCSPI, there are too many sudden discrete jumps in the DJIA. The best example of this is Black Monday, the stock market crash of October, 1987.

In general, these discontinuities do not convey much meaningful information to this type of analysis. Yet, upon the arrival of these outliers, the filter would revise the estimate by more than it would when a normal signal arrives. On the other hand, the outliers cannot be completely ignored since it cannot be determined ex ante whether a big shock represents a discontinuity, some kind of trend, a white noise, or some combination of the above. Hence, the innovations are truncated with the boundary values of ± 2σ. The results are shown in Graph 6-3 through 6-6, 6-9 and 6-10. Graph 6-3 illustrates the non-linear filter estimates calculated using empirical α while Graph 6-4 and 6-5 show the estimates calculated with a-priori α's (-2 and -1.5, respectively). Graph 6-6 compares the estimates with empirical α and the estimates with a-priori α (-2). The non-linear filter estimates using empirical α (i.e. Graph 6-3) are divided into two equal-size groups of 500 estimates each and shown in Graph 6-9 and 6-10.

6.2 Kalman Filter

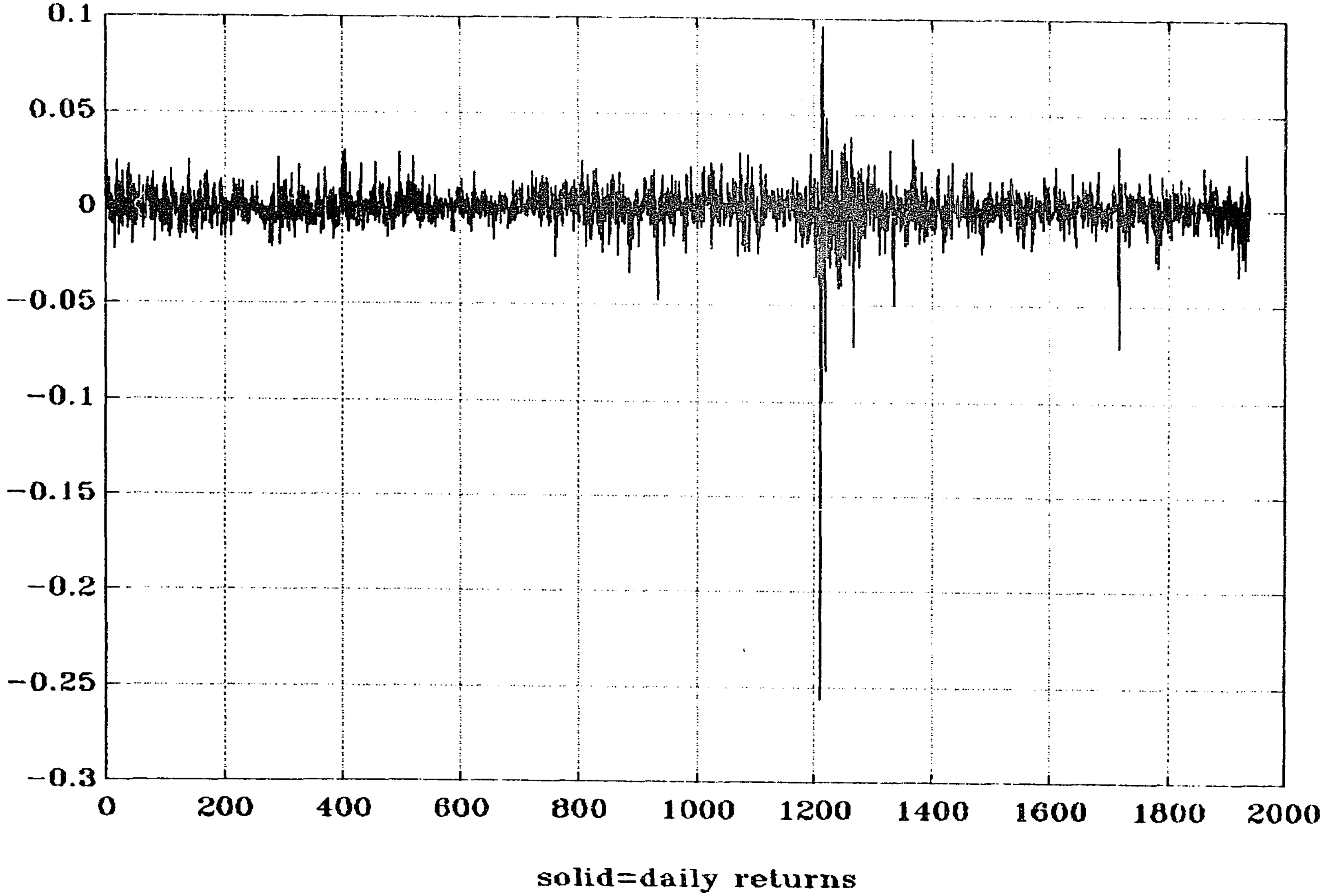
Since the optimal Kalman filter is robust and stationary, it was applied to the underlying series without the outlier treatment. The result was quite tragic. In one instance, the probability estimate soared over 40 and in another it plunged below -20. From Graph 6-7, it is clear that the model becomes useless in the presence of some anomalies. At the same time, it is also clear that, in spite of the anomalies, this model meets the convergence criterion. Owing to these outliers, the outlier treatment became necessary even in the Kalman filter case. The robustized estimates are reported in Graph 6-8. These 1,000 estimates are divided equally and shown in Graph 6-11 and 6-12.

GRAPH 6-1. DOW JONES INDUSTRIAL AVERAGE

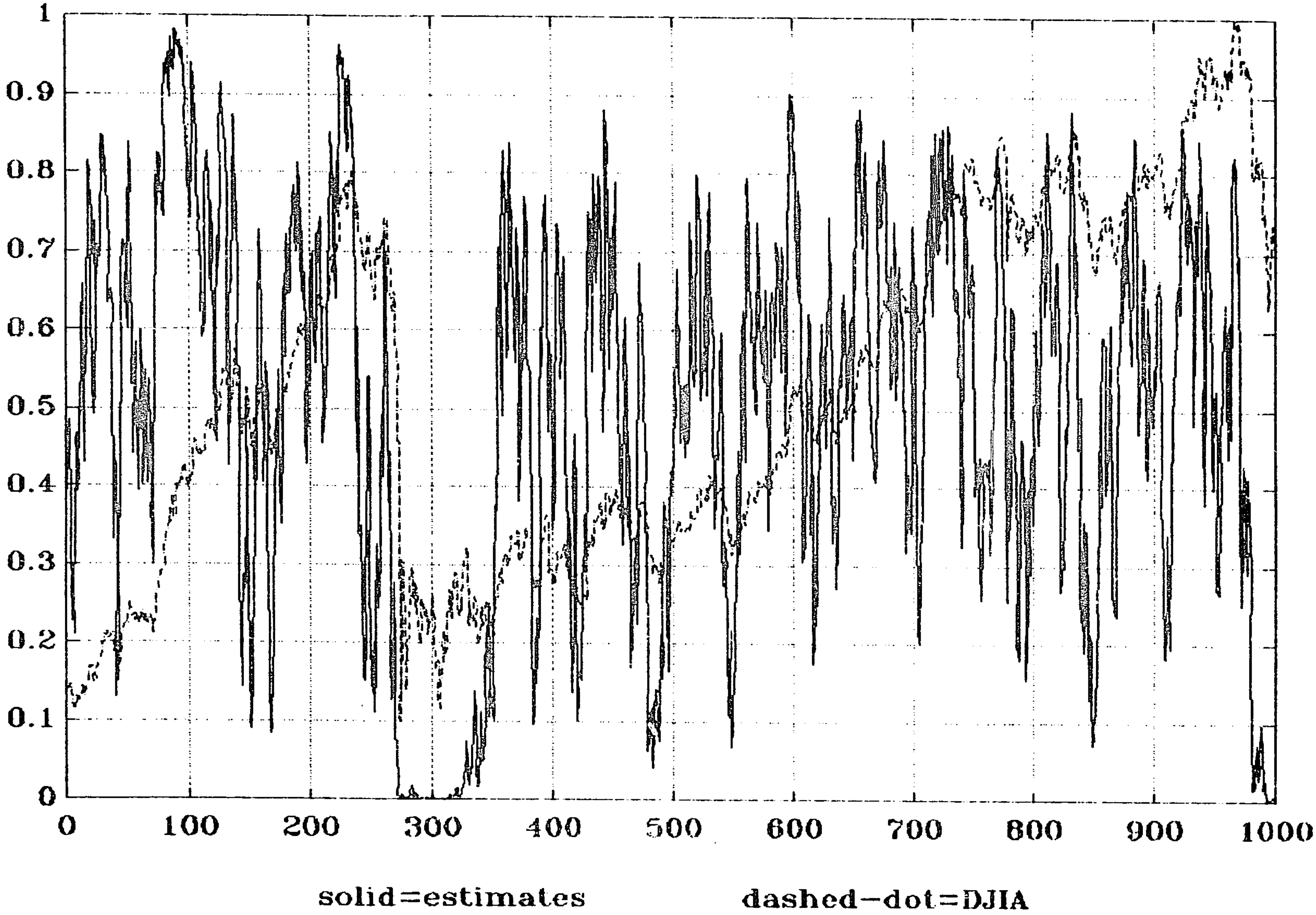


from Jan. 1983 to Aug. 1990

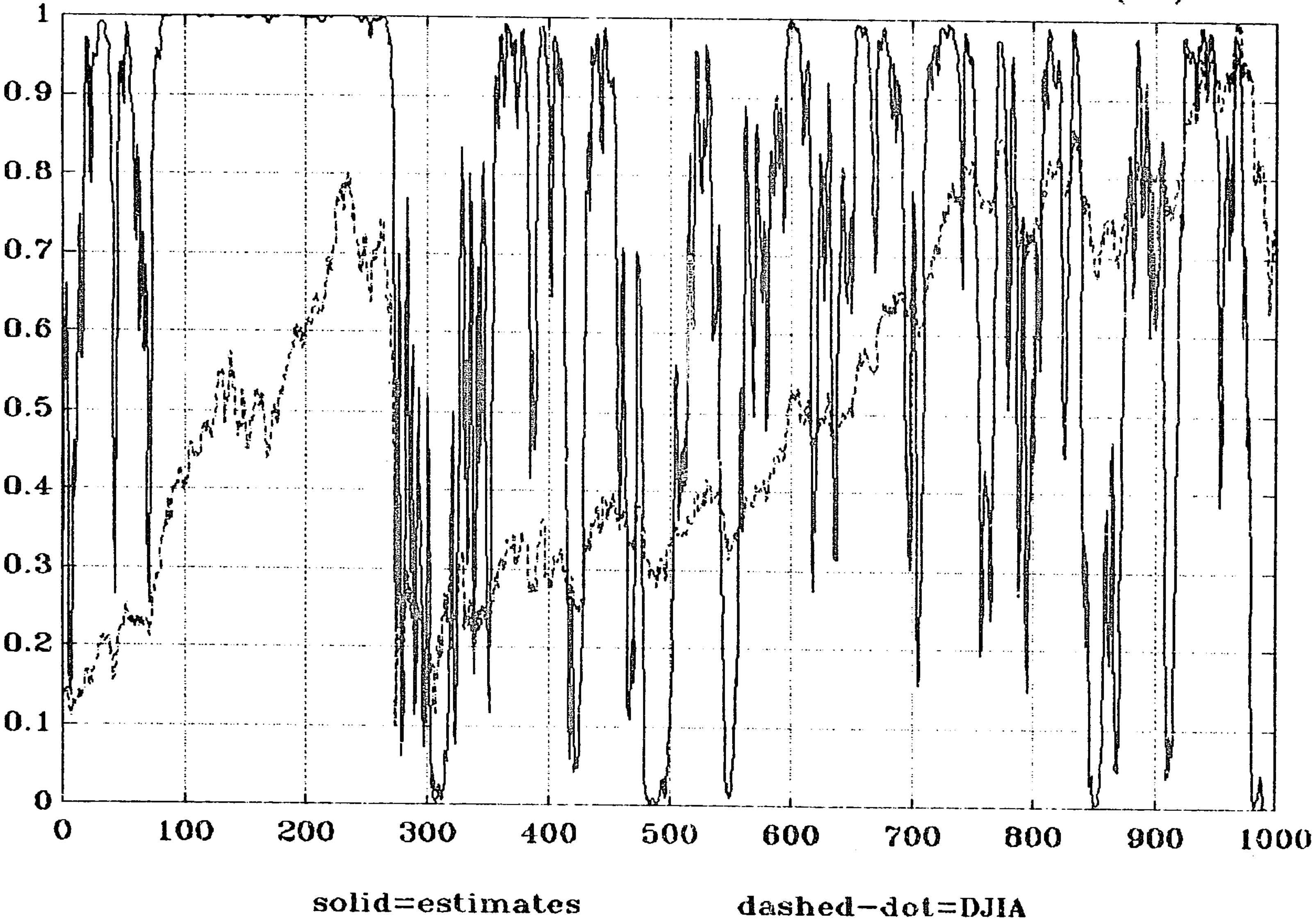
GRAPH 6-2. DJIA: DAILY RETURNS



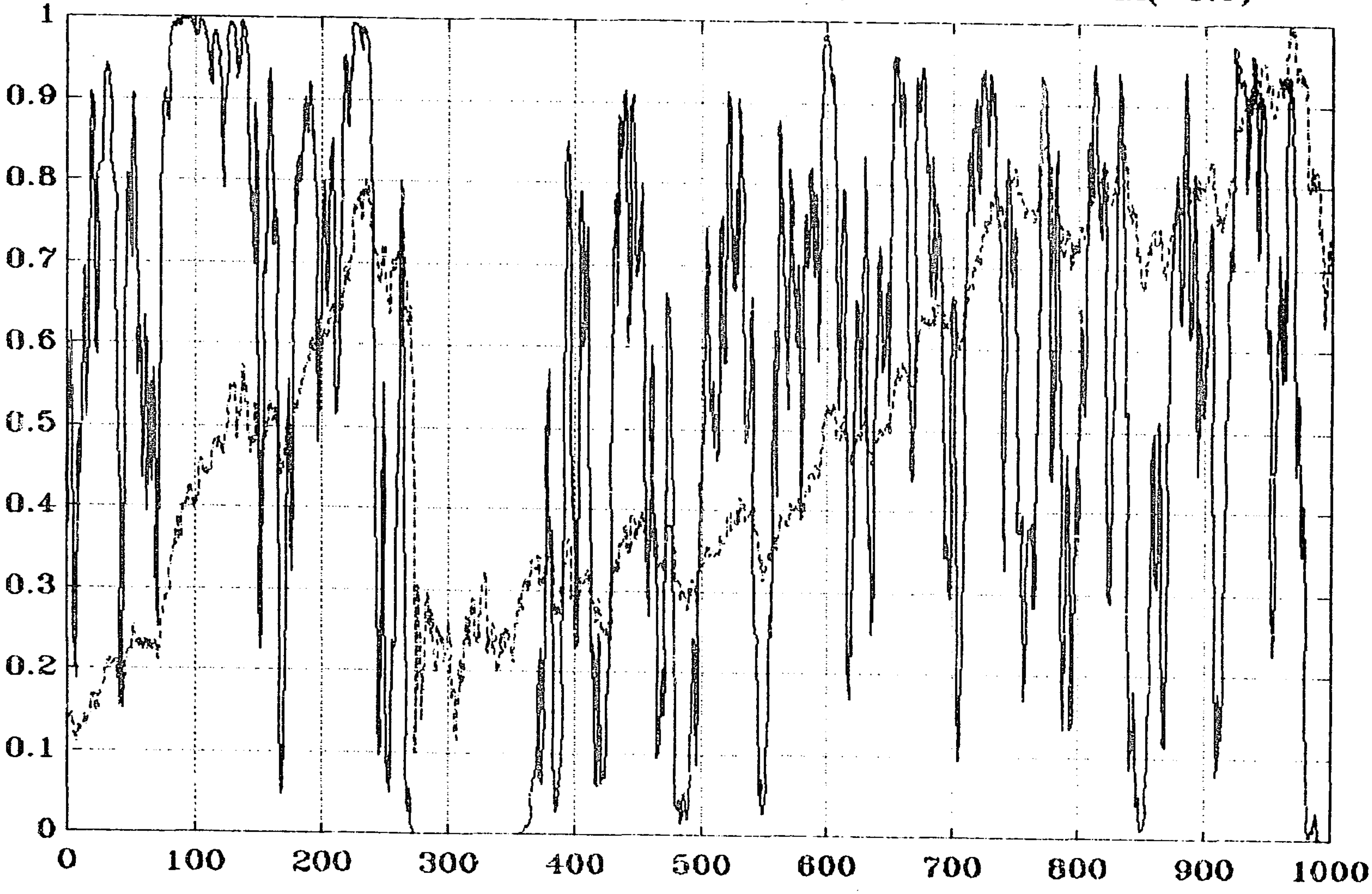
GRAPH 6-3. DJIA: NON-LINEAR FILTER, EMPIRICAL ALPHA



GRAPH 6-4. DJIA: NON-LINEAR FILTER, A-PRIORI ALPHA(-2)

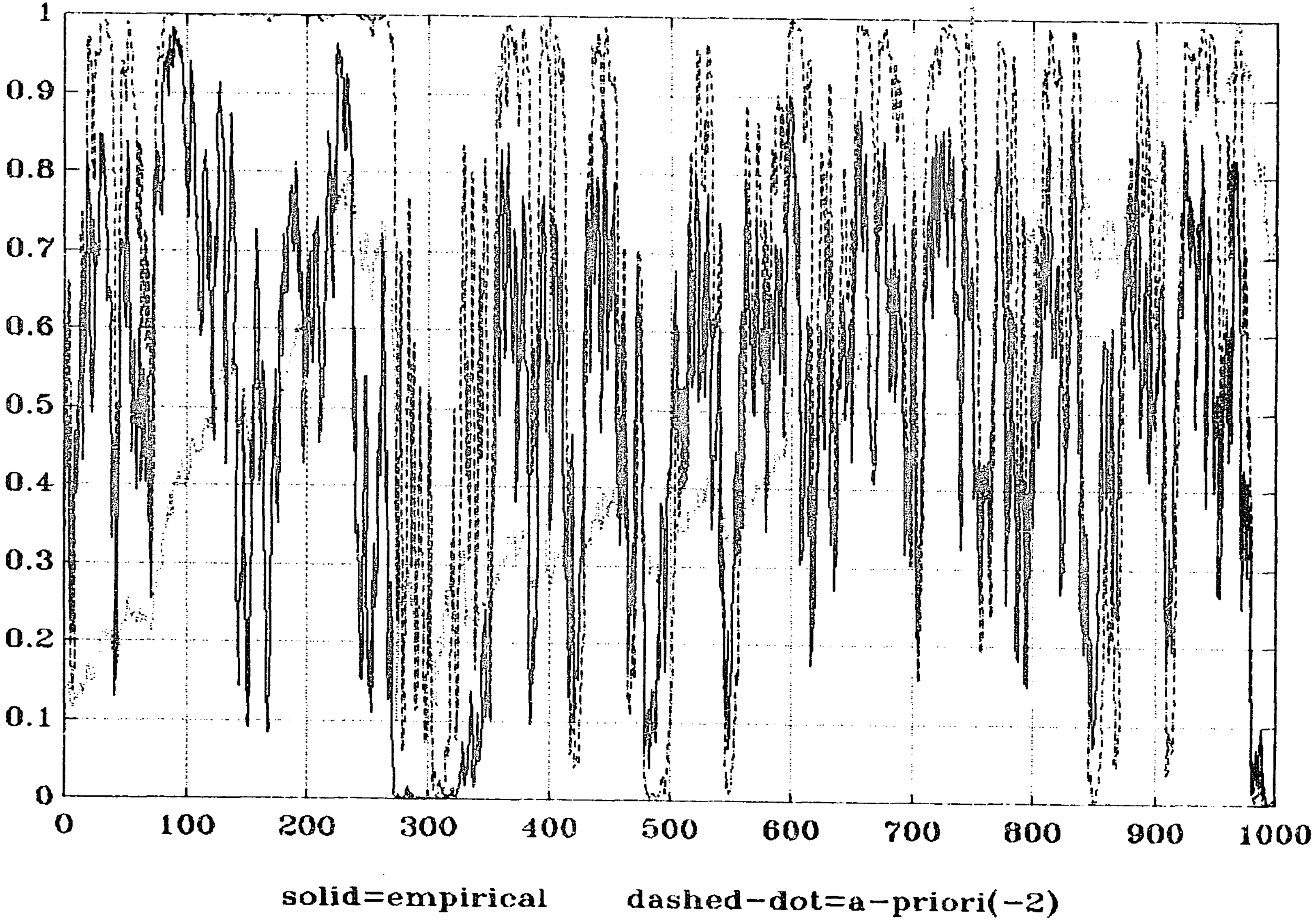


GRAPH 6-5. DJIA: NON-LINEAR FILTER, A-PRIORI ALPHA(-1.5)

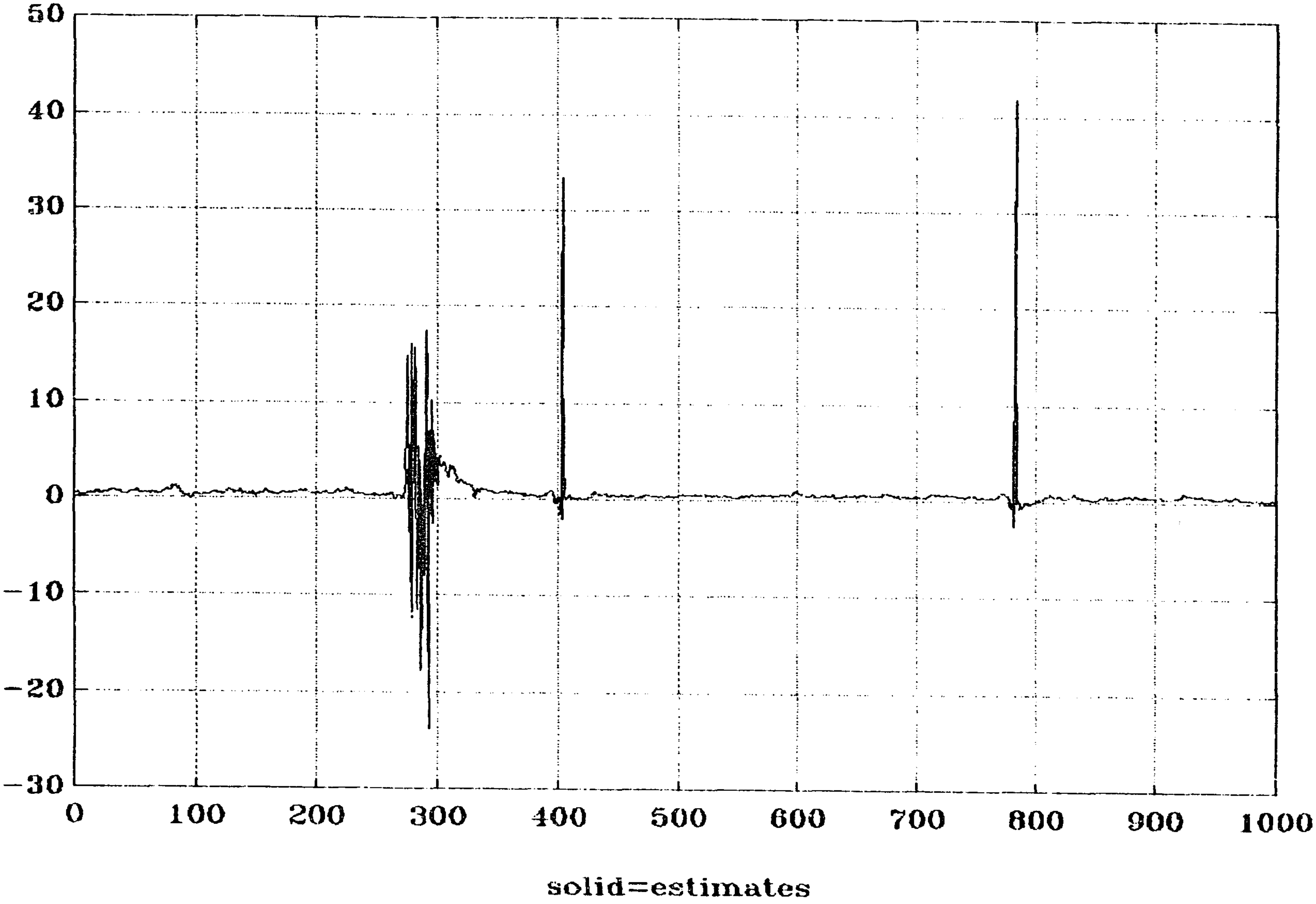


solid=estimates dashed-dot=DJIA

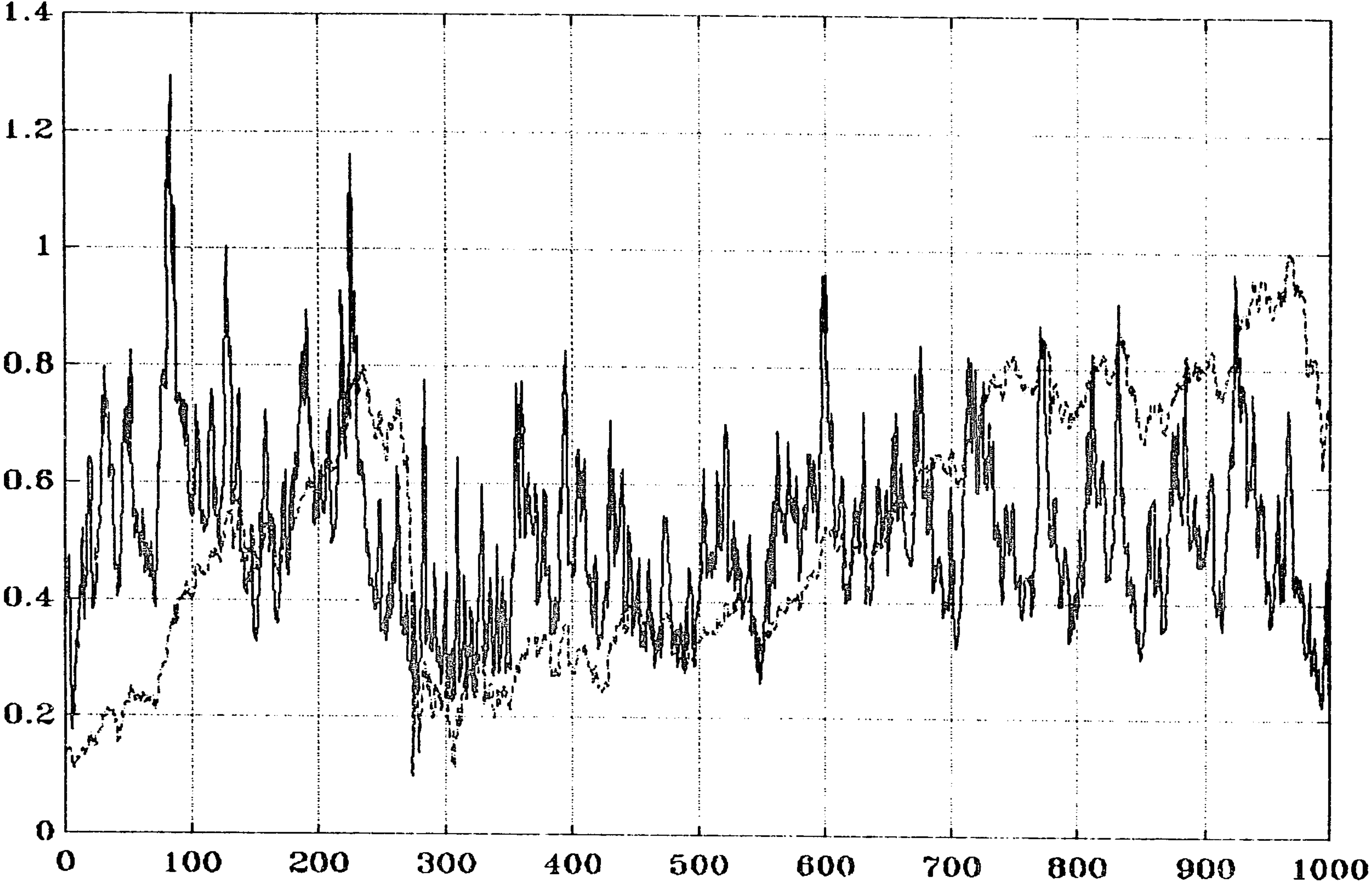
GRAPH 6–6. DJIA: NON–LINEAR FILTER, EMPIRICAL vs. A–PRIORI ALPHA



GRAPH 6–7. DJIA: KALMAN FILTER, NOT WINSORIZED

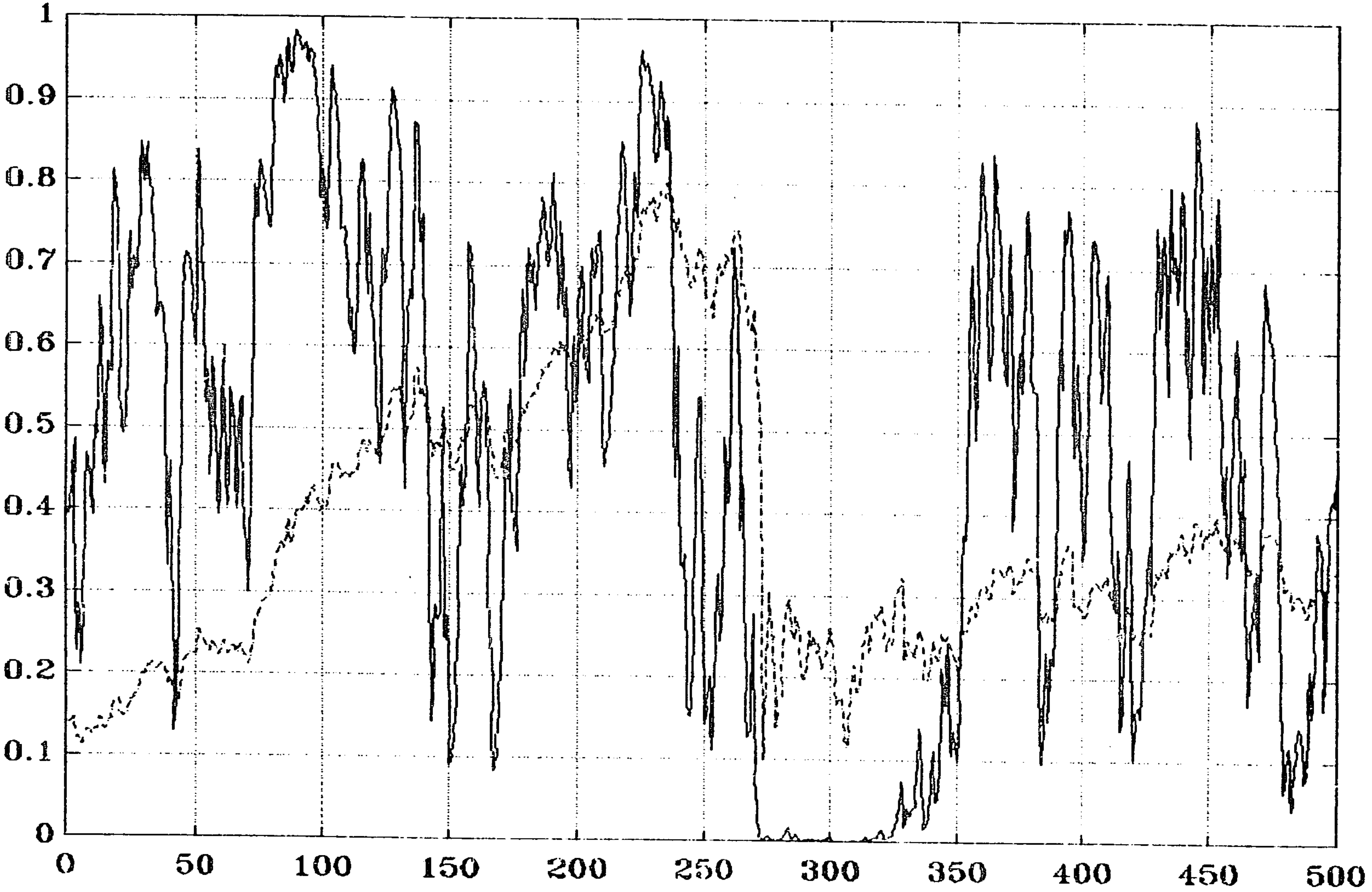


GRAPH 6-8. DJIA: KALMAN FILTER, WINSORIZED



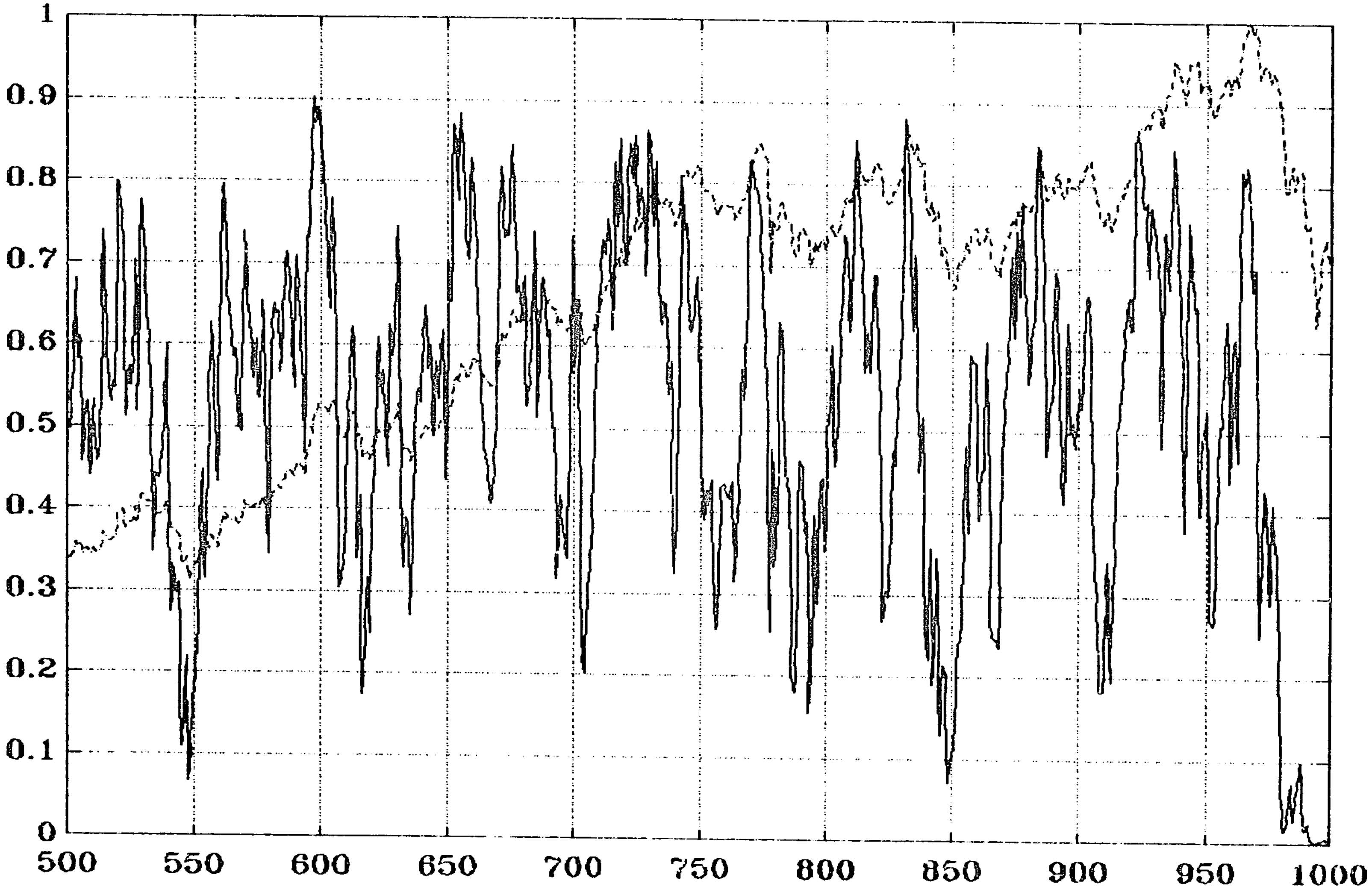
solid=estimates dashed-dot=DJIA

GRAPH 6-9. DJIA: NON-LINEAR FILTER, EMPIRICAL ALPHA



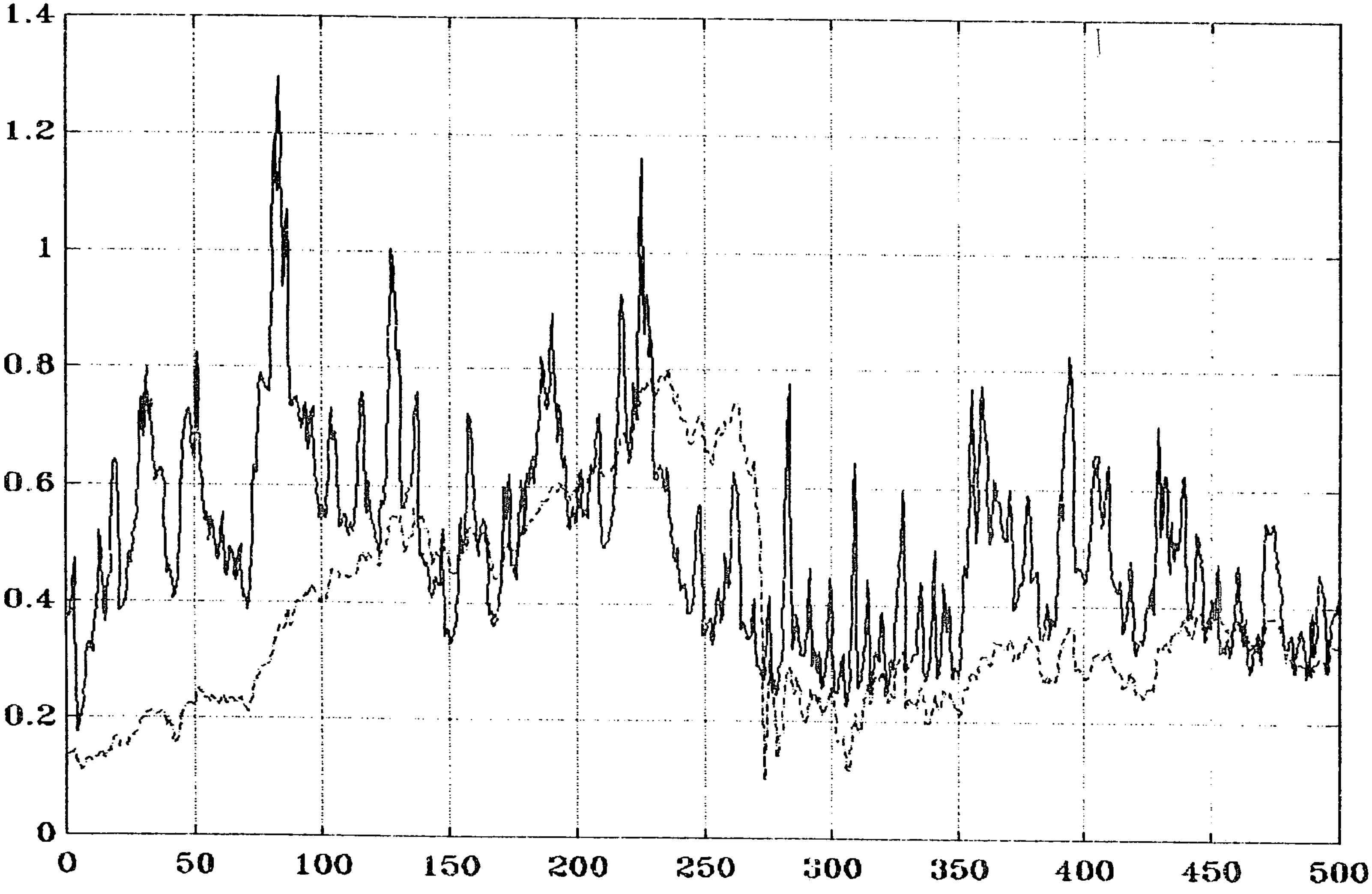
solid=estimates dashed-dot=DJIA first 500 obs.

GRAPH 6-10. DJIA: NON-LINEAR FILTER, EMPIRICAL ALPHA



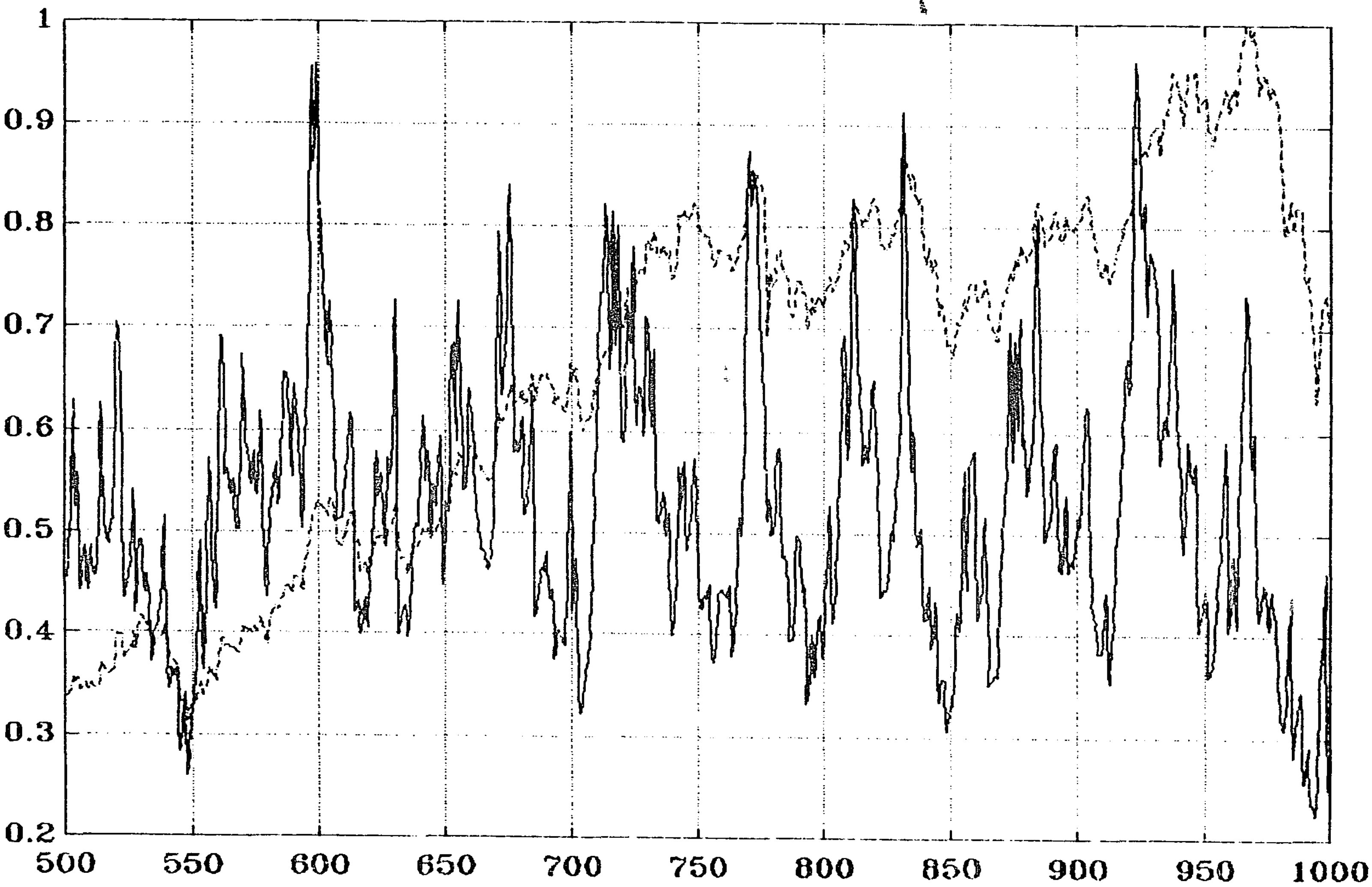
solid=estimates dashed-dot=DJIA second 500 obs.

GRAPH 6–11. DJIA: KALMAN FILTER



solid=estimates dashed–dot=DJIA first 500 obs.

GRAPH 6-12. DJIA: KALMAN FILTER



solid=estimates dashed-dot=DJIA second 500 obs.

VII. Comparison

There are four major issues to be discussed when comparing the two filters: sensitivity, robustness, stationarity, and cost. The following comparison will show that neither filter can claim an absolute advantage.

7.1 Sensitivity

Stated simply, the non-linear filter is much more sensitive than the Kalman filter. This does not necessarily imply that one is better than the other. Sensitivity can be an advantage or a disadvantage depending on the characteristics of the data set, primarily the noisiness.

7.1.1 S&P 500

Monthly data sets tend to be less noisy than daily data sets. For this reason, the non-linear filter generally performs better than the Kalman filter when monthly data are used. The Non-linear filter outperforms the Kalman filter when the change in the series is gradual, although, for big shocks, the optimal Kalman filter works as well as, but not necessarily better than, the non-linear filter. This can be easily understood by comparing the two filter gains:

Non-linear filter gain:

$[\pi_t(1-\pi_t)/\sigma^2](a_1-a_2)$*innovation

Kalman filter gain:

$[\sigma^2+(a_1-a_2)^2\Gamma(t)]^{-1}[P_{11}(t)-P_{12}(t)]\Gamma(t)[a_1-a_2]$*innovation

In both filters, the weights are multiplicative with respect to the innovation term. Therefore, the magnitude of the update of both filters is determined not only by their sensitivity but also by the size of the innovation. When the innovation term is large, the difference in sensitivity between the two filters is not so significant as to make the reaction of the two filters very different from each other. This explains why the Kalman filter performs as well as the non-linear filter in the presence of big shocks.

In the presence of small shocks, the non-linear filter works much better than the optimal Kalman filter. This is because the sensitivity of the filter becomes more significant when the innovation is small, that is when change is gradual, than it does in the presence of a large innovation. Therefore, the non-linear filter works better in detecting gradual changes from one state to another.

This point is well depicted in the estimation with the monthly S&P 500. Since monthly series is less noisy than daily series, the Non-linear filter reacts much more quickly to new signals than the optimal Kalman filter does. Specifically, when the change is gradual, the optimal Kalman filter is far slower in reaction. Only when the shock is comparatively large, does the optimal Kalman filter reacts as quickly as the non-linear filter. (See Graph 4-4 through 4-8.)

7.1.2 KCSPI

The sensitivity of the non-linear filter becomes a liability when the data set is very noisy. (See Graph 5-2 for the daily returns of KCSPI.) The noisiness of the data set is accounted for by the filter through the variance, σ^2. The filter will assign a bigger σ^2 to noisy data sets, and, therefore, will give less weight to the noisy signals. However, the filter still reacts so sensitively, even to a very small signal, that it generates estimates from which it is very hard to infer the state of world.

This can be easily seen from the Korean Composite Stock Price Index. Comparing Graph 5-3 (KCSPI: the non-linear filter estimates, empirical α) to Graph 5-5 (KCSPI: the Kalman filter estimates), it is clear that the Kalman filter is much less sensitive. (See Graph 5-7 for the comparison.) Yet, its lack of sensitivity makes it better able to estimate the trend of stock prices. Thus, the optimal Kalman filter performs better than the non-linear filter in estimating long-run trends. At the same time, however, the insensitivity of this filter works as a disadvantage in the sense that it delays the detection of trends.

With regard to the sensivity issue, the non-linear filter is better in detecting the onset of new trends while the Kalman filter is stronger in the estimation of long-run trends. The dilemma, then, is to choose the filter most suited to the type of data set.

Since both issues are important, it would be ideal to find a filter which possesses both characteristics. The solution was sought through the use of an a-priori α. As discussed in section

5.2 (KCSPI: the non-linear filter, a-priori α), the use of an a-priori α of -2 allows the non-linear filter not only to effectively detect the onset of a new trend, but also to estimate long-run trends. In fact, the filter revises the estimate a great deal at the onset of a new trend, and it does not update as much once the trend has begun. In other words, it is sensitive to the arrival of new information, yet it is less sensitive to the noise of the process. Specifically, it is defined such that it would give a large weight to the onset of a new trend and maintain the estimate at the revised level until the trend changes. This is shown in Graph 5-4 (KCSPI: the non-linear filter, a-priori α).

7.2 Robustness

When the economic time series exhibits a sudden big jump, both filters would generate estimates having values outside the boundary limits of 0 and 1. In general, both filters can handle the signals very well. The filters break down only when the innovation arrived is beyond 2 or 3 standard deviations from its mean. Due to the different sensitivity, the non-linear filter and the Kalman filter exhibit different reactions to the outliers. The non-linear filter is so sensitive that it gives bigger weight to the signals than does the Kalman filter. Consequently, the non-linear filter would break down more frequently and easily. In this sense, robustness is just a different face of the sensitivity issue.

7.2.1 S&P 500

Filtering monthly data did not create any problem in this respect. The monthly S&P 500 is smooth enough to generate estimates which are generally acceptable. Both the filter estimates ranged between -0.2 and 1.2. This is shown in Graph 4-8 (S&P 500: comparison between the non-linear & the Kalman filter). Note that the last observation in this series is October 1987. Both filters (especially the non-linear one) revised their estimates dramatically, yet they did not break down.

7.2.2 KCSPI

Filtering daily data, on the other hand, was not an easy task. The non-linear filter broke down in the KCSPI case, and refused to generate estimates at all. (This is also due to the non-stationarity of the filter.) For this reason, the filter was designed in such a way that it would metrically winsorize the innovation beyond the boundary of ± 2σ. Through the use of this procedure, the Non-linear filter was able to produce acceptable estimates.

On the other hand, the optimal Kalman filter was somewhat robust to outliers. In the Korean stock price analysis where there were hardly any discrete jumps, the filter did not have to be winsorized. In fact, it again generated acceptable estimates which ranged between 0 and 1.2.

7.2.3 DJIA

Suprisingly, not only the non-linear but also the Kalman filter broke down in the estimation with DJIA. This is due to the discontinuities that the American stock market experienced. Within a 7 year period, more than 5 discontinuities were discovered. The classical case is the Black Monday of October 1987. In one day, the stock price plunged by approximately 600 points or 25% of the closing price of the previous day. Due to these big sudden jumps, the Kalman filter generated estimates which are totally absurd, not to mention the non-linear filter's performance. The Kalman filter estimates ranged between -30 and 50 in some instances. This is shown in Graph 6-7 (DJIA: the

Kalman filter estimates, not winsorized). The only difference between the Kalman filter and the non-linear filter is that the former converged while the latter diverged. Hence, the winsorization was required for both filters in order to keep the estimates within the boundary limits. The windorized estimates are shown in Graph 6-3 through 6-5 and 6-8 (DJIA: the non-linear filter estimates & the Kalman filter estimates).

After winsorization, the non-linear filter still posed problems in the sense that it generated estimates that were not consistent with the market movements. It took sometime before the filter recovered completely from the shock. One explanation for this failure might be the increased after-crash volatility. (See Graph 6-2 for daily returns.) The market became chaotic after the crash and this was reflected in the volatility. It is possible that the real breakdown might be in the after-crash market, and not in the filter itself.

7.3 Stationarity

The stationary of the model naturally follows the robustness of the system. As discussed earlier, the Kalman filter is expected to be stationary, and this turned out to be the case. Even in the extreme case when the probability estimate sky-rocketed to 50, it converged back within the boundary limits. See Graph 6-7 for these estimates.

On the other hand, the non-linear filter is generally not stationary. Except in one instance, the estimates diverged to infinite values whenever they went out of the boundary limits. The exception was observed in the S&P 500 case, where the estimate dropped below 0 and did not diverge. The converging effect of the transition probability must have been greater than the diverging effect of the discrete time approximation's correction factor. In the discrete time analysis, the non-linear filter requires the transformation of the innovation both for the boundary limits and for the stationarity while the Kalman filter needs it for the boundary limits, only.

7.4 Cost

In light of the above discussion, it is quite clear that the non-linear filter requires much more effort in order to get it working. In both daily data sets, it required outlier treatment since, otherwise, it would have diverged to infinity. In general, the lon-linear filter requires much more attention from the researcher; also, it is more tedious and time consuming. On the contrary, Kalman filter did not at all require the innovation treatment in the KCSPI case. Even in the DJIA case, the treatment was not mandatory. Since the trajectory of π_t converges anyway, all that is required is to remove the abnormal estimates after filtering process is completed. This is the major advantage of the Kalman filter, along with its long-term trend prediction power.

VIII. Conclusion

In contrast to what is generally perceived, the non-linear filter is not always better than the linear filter. Both have advantages and disadvantages as well.

The non-linear filter is a very delicate tool in time series analysis. It is much more sensitive to shocks than the Kalman filter. It picks up the gradual changes in the parameter. It detects better the onset of a new trend. In general, it works better when the data set is less noisy.

On the other hand, the non-linear filter has its own disadvantages, particularly when the data set is noisy. It is too sensitive to estimate long-run trends. For the same reason, it is vulnerable to data inaccuracy. It cannot handle outliers because of its inherent sensitivity and non-stationarity. Furthermore, this delicate tool incurs a greater cost in terms of time and energy.

The advantages of the Kalman filter are its lower sensitivity, stationarity, and cost effectiveness. Since it is insensitive, it is less affected by the noise of the data. Therefore, it is better at estimating the long-term trends of a series. It is less affected by data inaccuracy and outliers. Also, it is stationary and, for all the reasons listed above, it is less costly.

However, it still has its disadvantages, too. It is slow in its reaction to new signals which could be fatal in financial markets. It also has problems in detecting the onset of new trends. It takes quite a while for this filter to detect a new trend, and, by that time, another trend may already have set in. To this end, its ability to detect long-run trends cannot be an absolute advantage over the non-linear filter since the non-linear filter can detect long-run trends at least as well as, maybe even better than, the Kalman filter when a value of α is carefully selected a priori.

It can be concluded from the discussion presented above that the non-linear filter, in general, is a better tool than the linear filter. However, the filter should be carefully chosen depending on the characteristics of the filters on which the emphasis is laid, e.g. ability to detect the onset of a new trend, long-run trend prediction power, etc. For example, the Kalman filter would be more recommendable when the marginal cost of calculation is greater than the marginal benefit from the improvement in the estimates of the parameters.

Bibliography

Aoki, Masanao, State Space Modeling of Time Series, New York: Springer-Verlag, 1987.

Barro, Robert J., Editor, Modern Business Cycle Theory, Cambridge: Harvard University Press, 1989.

Blanchard, Olivier Jean & Fischer, Stanley, Lectures on Macroeconomics, Cambridge, The MIT Press, 1989.

Brealey, Richard A. & Myers, Stewart C., Principles of Corporate Finance, 3rd Edition, New York: McGraw-Hill, 1988.

Breiman, Leo, Probability and Stochastic Processes, The Scientific Press, 1986.

Copeland, Thomas E. & Weston, J. Fred, Financial Theory and Corporate Policy, 3rd Edition, New York: Addison-Wesley Publishing Company, 1988.

Cox, John C. & Rubinstein, Mark, Options Markets, Englewood Cliffs, Prentice-Hall, 1985.

Easley, David & O'Hara, Maureen, "Price, Trade Size, and Information in Securities Markets", Journal of Financial Economics, Vol. 19, 1987, pp. 69-90.

Fama, Eugene F., "The Behavior of Stock Market Prices", Journal of Business, Vol. 38, 1965, pp. 34-105.

__________, "Efficient Capital Markets: A Review of Theory and Empirical Work", Journal of Finance, Vol. 25, 1970, pp. 383-417.

Frost, P. & Kaliath, T., "An Innovations Approach to Least-Squares Estimation, Part III", IEEE Trans. Automatic Control, AC-16, 1971, pp. 217-226.

Grossman, Sanford J., "On the Efficiency of Competitive Stock Markets Where Traders Have Diverse Information", Journal of Finance, Vol. 31, 1976, pp. 573-585.

__________, "The Existence of Future Markets, Noisy Rational Expectations and Informational Externalities", Review of Economic Studies, Oct. 1977, pp. 431-449.

Grossman, Sanford J.& Stiglitz, Joseph E., "On the Impossibility of Informationally Efficient Markets", Technical Report, No. 259, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1978.

__________, "Information and Competitive Price Systems", American Economic Review, Vol. 66, May 1976, pp. 246-253.

Hamilton, James D., "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", Econometrica, Vol. 57, Mar. 1989, pp. 357-384.

Hellwig, Martin F., "On the Aggregation of Information in Competitive Markets", Journal of Economic Theory, Vol. 22, 1980, pp. 477-498.

Huang, Chi-fu & Litzenberger, Robert H., Foundations for Financial Economics, New York: North-Holland, 1988.

Huber, Peter J., Robust Statistics, New York: John Wiley & Sons, 1981.

Johnston, J., Econometric Methods, New York: McGraw-Hill, 1984.

Judge, George G., et al., Introduction to the Theory and Practice of Econometrics, 2nd Edition, New York: John Wiley & Sons, 1988.

Karlin, Samuel & Taylor, Howard M., A First Course in Stochastic Processes, New York: Academic Press, 1975.

__________, A Second Course in Stochastic Processes, New York: Academic Press, 1981.

Krichagina, N.V., Lipster, R.S. & Rubinovich, E.Y., "Kalman Filter for Markov Processes", Steklov Seminar: Statistics and Control of Stochastic Processes, Edited by Krylov, N.V., Lipster, R.S. & Novikov, A.A., New York: Optimization Software, Inc., 1984.

Lipster, R.S. & Shiryayev, A.N., Statistics of Random Control I: General Theory, New York: Springer-Verlag, 1977.

________, Statistics of Random Control II: Applications, New York: Springer-Verlag, 1978.

Lucas, Robert E. Jr., "Asset Prices in an Exchange Economy", Econometrica, Vol. 46, Nov. 1978, pp. 1429-1445.

Maddala, G.S., Econometrics, New York: McGraw-Hill, 1977.

Neftci, Salih N., Optimal Prediction of Cyclical Downturns, Journal Economic Dynamics and Control, Vol. 4, Aug. 1982, pp. 225-241.

__________, "Are Economic Time Series Asymmetric over the Business Cycle?", Journal of Political Economy, Vol. 92, Apr. 1984, pp. 307-328.

Sargent, Thomas J., Dynamic Macroeconomic Theory, Cambridge: Harvard University Press, 1987.

Stiglitz, Joseph E., "Symposium on Bubbles", Journal of Economic Perspectives, Vol. 4, No. 12, Spring 1990, pp. 13-18.

Summers, Lawrence H., "Does the Stock Market Rationally Reflect Fundamental Values?", Journal of Finance, Jul. 1986, pp. 591-601.