BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-90/11

MAKING DIFFICULT MODEL COMPARISONS

by

David F. Findley
U.S. Bureau of the Census
Statistical Research Division
Room 3000, FOB #4
Washington, DC 20233 U.S.A.

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Report completed:	June 29, 1990
Report issued:	June 29, 1990
Report revised:	November 19, 1990

MAKING DIFFICULT MODEL COMPARISONS

David F. Findley
Statistical Research Division
Bureau of the Census, Washington, D.C. 20233

SUMMARY

The process of statistical model selection often involves comparison of quite distinct competing models, none of which is correct in any strict sense In such situtations, the classical, nested–model comparison procedures are inapplicable. This paper describes an easily calculated and broadly applicable graphical diagnostic for model comparison and also some robust generalizations to the time series situation of a test statistic of Vuong (1989) for non–nested model comparisons of incorrect models. After presenting some supporting theory, we illustrate the application of these new procedures by comparing ARIMA models with "structural" component models for forty seasonal economic time series, continuing a study by Bell and Pugh (1989) who used AIC for this purpose. When the comparison procedures are decisive, ARIMA models are usually favored.

Keywords: ARIMA MODELS; GRAPHICAL DIAGNOSTICS; INCORRECT MODELS; MAXIMUM LIKELIHOOD; NON–NESTED COMPARISONS; ROBUST SPECTRUM ESTIMATION; STRUCTURAL MODELS; VUONG'S TEST.

1. INTRODUCTION

There are few situations where one could expect the properties and assumptions of a statistical model for observed data to perfectly describe the properties of the data, even if arbitrarily large sample sizes could be obtained with which to verify the convergence of the parameter estimates. Consequently, given the inherently approximate nature of statistical models, it is not surprising that modeling experts working independently of one another usually obtain different models for the same data. Thus there is a basic need for broadly applicable model comparison procedures. The very large role played in the theoretical statistical literature and in textbooks by the assumption that at least one of the model classes considered is correct has obscured the fact that such an assumption is not necessary for model comparisons, even of non–nested models, as we demonstrate in this paper. Our methods are based on the fact that if two models do not provide equally good fits to the data in large samples, then their log–likelihood–ratio will usually diverge linearly to an infinite value.

1.1. An Introductory Example

Consider the following simplistic example. Let y_n be a covariance stationary time series with mean zero, and with autocovariances $\gamma_k = Ey_ny_{n-k}$ and autocorrelations $\rho_k = \gamma_k/\gamma_0$, k=0, ± 1,... such that $\rho_2 \neq \rho_1^2 \neq 0$ and $|\rho_2|<1$. Suppose two autoregression models are considered for y_n, regression on y_{n-1} and regression on y_{n-2}. All of the usual estimates of the regression coefficients (least squares, m.l.e., etc.) converge to ρ_1 and ρ_2 respectively, with increasing sample size, yielding the competing asymptotic model equations

$$(1.1) \qquad y_n = \rho_1 y_{n-1} + e_n^{(1)}$$

and

$$(1.2) \qquad y_n = \rho_2 y_{n-2} + e_n^{(2)} .$$

Set $\sigma^{(j)^2} \equiv Ee_n^{(j)^2}$, j=1,2. The regression error processes $e_n^{(1)}$ and $e_n^{(2)}$ satisfy

$$(1.3) \qquad Ee_n^{(j)} y_{n-j} = 0; \;\; \sigma^{(j)^2} = \gamma_0(1-\rho_j^2) \quad (j=1,2),$$

and it can be shown that they are not white noise processes (uncorrelated series). Therefore (1.1) and (1.2) are incorrect models for y_n. Clearly the models are non–nested: neither is a restricted version of the other. The quantities $\sigma^{(1)^2}$ and $\sigma^{(2)^2}$ provide natural goodness–of–fit measures for these models. A first question to ask is whether one of these error variances is smaller than the other. This reduces, by (1.3), to a question about the magnitudes of ρ_1 and ρ_2, so the well–known limiting joint distribution of the sample autocorrelation estimates of ρ_1 and ρ_2 could be utilized to provide a classical test of hypothesis for choosing the asymptotically better fitting model if there is one, that is, if $|\rho_1| \neq |\rho_2|$. However, finding analogous hypothesis tests for more complex model pairs could be very difficult, so a different approach is needed in general. This paper describes two rather simple and very general procedures for detecting an asymptotically better fitting model according to a natural measure of fit. It will be apparent from the exposition that the range of applicability of these procedures is not limited to time series model comparisons. For reasons of space, only time series applications are presented.

1.2. General Introduction

Suppose one wishes to decide between two model families, not necessarily nested and not necessarily correct, for observed data $y_1,...,y_N$. Conceptually, there are two possible situations, illustrated by the example above: either the theoretically best–fitting models from the competing classes fit (i) equally well, or (ii) one model class is capable of providing a better fit than the other. In case (i), variability arising from parameter estimation can still cause one of the two classes to be preferred, and this is the playing field of the null hypothesis in classical statistical tests. However, general statistical procedures for identifying the preferred class in this case seem to require rather strong assumptions about the approximate correctness of the models. By contrast, as the results of this paper show, there are theoretically founded procedures with relatively weak requirements for making the more fundamental decisions, does (i) hold, or (ii)? — and, if (ii), which model class provides the better fit?

As section 2 explains, these two questions are answered by deciding if the log–likelihood difference $\hat{L}_N^{(1,2)}$ of the maximum likelihood values from the two families diverge to an infinite value as N ⟶ ∞ and, if so, whether to $+\infty$ of $-\infty$. The obvious diagnostic for such divergence is a graph of the log–likelihood differences from an appropriately selected increasing sequence of subsets of the observed data set. However, the calculation of the sequence of m.l.e.'s and the likelihood values required for such a graph can be quite demanding computationally and therefore poorly suited to interactive modeling. In section 3, we will describe a related graphical diagnostic which is suited to interactive modeling and is especially convenient because it requires only quantities which are usually available from the full–data–set likelihood maximizations. Also, a condition, (3.5), is given which guarantees that this diagnostic describes the relevant behavior of $\hat{L}_N^{(1,2)}$ for large enough sample sizes N. The proposition of section 4 shows that (3.5) can be verified under rather weak assumptions: for example, it is not required that the maximum likelihood parameter estimates converge uniquely.

The rest of the paper is devoted to time series models in preparation for the comparison study, in section 10, of ARIMA and structural component models for 40 economic time series. Sections 5 and 6 provide verifications of (3.5) for two classes of ARMA time series models. Section 7, which discusses the connection between log–likelihood differences for ARIMA models and those for the corresponding ARMA models, may be of independent interest.

The first general hypothesis testing procedure for detecting divergence to infinite values of the log–likelihood difference for non–nested, incorrect models seems to be that of the very insightful paper of Vuong (1989), which examines the case of independent and identically distributed data. A natural generalization of Vuong's statistics to the case of ARMA and ARIMA models is presented in section 9, based on the asymptotic distribution obtained for $N^{-1/2}\hat{L}_N^{(1,2)}$ in Proposition 8.4 of section 8. In section 10, two robustified versions of this statistic are applied to model comparisons for 40 economic time series and are shown to lead to the same model selection as the graphical diagnostic of section 3 in all situations in which both procedures have a preferred model . This is reassuring, also because our theoretical derivation of the asymptotic distribution of the test statistics in section 9 is incomplete.

2. WHEN AND HOW DO LOG–LIKELIHOOD–RATIOS DIVERGE TO ± ∞?

Let $L_N[\theta^{(1)}]$, $\theta^{(1)} \in \Theta^{(1)}$ and $L_N[\theta^{(2)}]$, $\theta^{(2)} \in \Theta^{(2)}$ be two parametric families of log–likelihoods defined by competing models for observed random variates $y_1,\ldots,y_N$. We do not assume that the competing log–likelihood functions have a similar form or are related in any way. A notation that more strongly emphasized possible differences of form would be $L^{(j)}[\theta^{(j)}]$, j=1,2, but we will avoid the duplicated superscript to reduce notational complexity, anticipating that this will not cause confusion for the reader. Usually each family of log–likelihood functions has an associated family of non–random "entropy" functions $\mathcal{E}_\infty[\theta^{(j)}]$, $\theta^{(j)} \in \Theta^{(j)}$ which are the limits (existing with probability one) of the sample–size–normalized log–likelihood functions,

(2.1) $$\mathcal{E}_\infty[\theta^{(j)}] = \lim_{N \longrightarrow \infty} N^{-1} L_N[\theta^{(j)}] \quad (j = 1,2) \quad (\text{w.p.1}).$$

If maximum likelihood estimates $\hat{\theta}_N^{(j)}$ exist and if we consider the entropy supremum

(2.2) $$\mathcal{E}_\infty^{(j)} \equiv \sup_{\theta^{(j)} \in \Theta^{(j)}} \mathcal{E}_\infty[\theta^{(j)}],$$

then, ordinarily (with p–lim denoting convergence in probability)

$$\text{(2.3)} \qquad \text{p–lim}_{N \longrightarrow \infty} N^{-1} L_N[\hat{\theta}_N^{(j)}] = \mathcal{E}_{\infty}^{(j)}$$

will hold for j = 1,2: see White (1990) and the proof of Theorem 7.4.10 of Hannan and Deistler (1988) where (2.3) is established without the assumption that the model classes under consideration contain the true model. Recently Pötscher (1990) has established (2.3) for incorrect *noninvertible* ARMA models. We will use the abbreviations $\hat{L}_N^{(j)} \equiv L_N[\hat{\theta}_N^{(j)}]$, j = 1,2, and, for the log–likelihood difference (= log likelihood–ratio),

$$\hat{L}_N^{(1,2)} \equiv \hat{L}_N^{(1)} - \hat{L}_N^{(2)} .$$

Then, from (2.3), we obtain

$$\text{(2.4)} \qquad \text{p–lim}_{N \longrightarrow \infty} N^{-1} \hat{L}_N^{(1,2)} = \mathcal{E}_{\infty}^{(1)} - \mathcal{E}_{\infty}^{(2)}.$$

Consequently,

(2.5) *If* $\mathcal{E}_{\infty}^{(1)} \neq \mathcal{E}_{\infty}^{(2)}$, *then*

$$\text{p–lim}_{N \longrightarrow \infty} \hat{L}_N^{(1,2)} = \pm \infty,$$

where the sign of the limit and its asymptotic slope are the sign and the slope of $(\mathcal{E}_{\infty}^{(1)} - \mathcal{E}_{\infty}^{(2)})N$. *Thus,* $\hat{L}_N^{(1,2)} = O_p(N)$.

This result has a particularly straightforward interpretation when Gaussian likelihood functions are used to model the mean and covariance structure (without assuming the data are Gaussian), because then, usually,

$$\mathcal{E}_{\infty}^{(j)} = -\tfrac{1}{2}\log 2\pi e \sigma_{\infty}^{(j)^2}, \tag{2.6}$$

where $\sigma_{\infty}^{(j)^2}$ is the variance of the asymptotic "residuals" process associated with the best fitting model(s) in the class being considered: for example, we shall show in sections 5 and 7 that for competing Gaussian ARMA (or ARIMA) time series models, $\sigma_{\infty}^{(j)^2}$ is the variance of the one–step–ahead forecast error process of a model determined by a limiting value $\theta_{\infty}^{(j)}$ of $\hat{\theta}_{N}^{(j)}$, j=1,2. Under (2.6), the sign of $\mathcal{E}_{\infty}^{(1)} - \mathcal{E}_{\infty}^{(2)}$ is that of $\sigma_{\infty}^{(2)^2} - \sigma_{\infty}^{(1)^2}$. This means that the model class with better one–step–ahead forecasting properties for the observed time series is the model class whose log–likelihood function will dominate the log–likelihood difference as $N \rightarrow \infty$.

3. GRAPHICAL DIAGNOSTICS FOR MODEL COMPARISON: DETECTING DIVERGENCE PROPERTIES OF THE LOG–LIKELIHOOD RATIO.

The result (2.5) immediately suggests a graphical method for detecting whether or not one of two log–likelihoods will, in large samples, strongly dominate the other and thereby identify itself as the model which is to be preferred. The method is: reestimate the model over an increasing sequence of subsets of the available observations and plot the resulting sequence of likelihood ratios as a function of sample size, looking for a persistently sloping trend movement in the later part of the graph. In a situation where the data index has a natural ordering, as with time series, this procedure would suggest calculating $\hat{\theta}_{M}^{(1)}$ and $\hat{\theta}_{M}^{(2)}$ from $y_1,\ldots,y_M$ for, say, $N/2 < M \leq N$, and plotting the log–likelihood differences

$$\hat{L}_{M}^{(1,2)} \quad , N/2 < M \leq N \tag{3.1}$$

as a function of increasing M.

However, the calculation of the quantities in (3.1) could be time consuming and also quite inconvenient with some software packages. We shall argue in this and the next sections that plotting

$$L_M[\hat{\theta}_N^{(1)}] - L_M[\hat{\theta}_N^{(2)}], \; N/2 < M \le N \tag{3.2}$$

versus M is much less burdensome computationally, yet also, when N is large enough, equally informative about the divergence properties of $\hat{L}_N^{(1,2)}$. Calculating (3.2) rather than (3.1) has the clear advantage that only a single likelihood maximization is necessary for each model class to obtain $\hat{\theta}_N^{(1)}$ and $\hat{\theta}_N^{(2)}$. Less obvious is the fact that the quantities in (3.2) are often immediately available as a byproduct of the calculation of $\hat{\theta}_N^{(j)}$, j = 1,2. This is easy to see when the data are modeled as though they were independent and identically distributed: in this case the log–likelihoods have the form

$$L_N[\theta^{(j)}] = \sum_{n=1}^{N} \log g[\theta^{(j)}](y_n) ,$$

and, for any $M \le N$,

$$L_M[\hat{\theta}_N^{(j)}] = \sum_{n=1}^{M} \log g[\hat{\theta}_N^{(j)}](y_n), \tag{3.3}$$

for j = 1,2.

In the case of dependent time series data modeled as a Gaussian ARMA or ARIMA model, there is an analogue of (3.3) which arises from the conditional decomposition

$$L_N[\theta^{(j)}] = \log g[\theta^{(j)}](y_1) + \sum_{n=2}^{N} \log g[\theta^{(j)}](y_n | y_{n-1},\ldots,y_1)$$

and which has the form

$$(3.4) \qquad L_M[\hat{\theta}_N^{(j)}] = -\frac{1}{2} \sum_{n=1}^{M} \left\{ \log 2\pi\sigma^2_{n|n-1}[\hat{\theta}_N^{(j)}] + \left[\frac{y_n - y_{n|n-1}[\hat{\theta}_N^{(j)}]}{\sigma_{n|n-1}[\hat{\theta}_N^{(j)}]}\right]^2 \right\} .$$

In this expression, $y_{n|n-1}[\theta]$ denotes the linear function of $y_{n-1},\ldots,y_1$ which would provide the best predictor of y_n if the data were Gaussian with the mean and covariance structure specified by θ (when n=1, $y_{n|n-1}[\theta]$ is the mean specified for y_1 by θ); $\sigma^2_{n|n-1}[\theta]$ is the function of θ equal to the mean square of $y_n - y_{n|n-1}[\theta]$ calculated with respect to the joint density $\exp(L_N[\theta])$. All of these quantities can be calculated from one pass over the data with the Kalman filter algorithm, given a state space representation of the time series model and a suitable initialization, see Jones (1980) or Bell and Hillmer (1990), for example. If the Kalman filter has been used to evaluate the likelihood function in the maximization routine, all of the quantities required for (3.4) and (3.2) will be available after the last maximization step.

Graphs of (3.1) and (3.2) for competing models (described in section 10) for eight economic time series are presented in Figs. 1–8. The subfigures (a) are graphs of (3.1), and the subfigures (b) are graphs of (3.2). There is further discussion of these Figures in section 10.

In subsequent sections, we shall demonstrate the large–sample equivalence of the sets of statistics (3.1) and (3.2) for model comparison by verifying the condition

$$(3.5) \qquad \lim_{M\to\infty} \sup_{N\geq M} |M^{-1}L_M[\hat{\theta}_N] - \mathcal{E}_\infty| = 0 \quad \text{(w.p.1)}$$

for the classes of models under consideration. This condition shows that either set of statistics can be used to determine the sign of $\mathcal{E}_{\infty}^{(1)} - \mathcal{E}_{\infty}^{(2)}$ if this quantity is non–zero.

Remark 3.1. When (3.5) holds, then, in theory, the lower bound, N/2, of M in (3.1) and (3.2) could be replaced by N/r for any fixed r > 1. In practice, if too large a value of r is used, early values of $\hat{L}_M^{(1,2)}$ can have a graph quite different from $L_M[\hat{\theta}_N^{(1)}] - [L_M[\hat{\theta}^{(2)}]$.

4. VERIFICATION OF (3.5): A GENERAL RESULT

One attractive feature of the result of this section is that it accommodates the situation (observed by Kabaila (1983) to occur sometimes with an incorrect first order moving average time series model) wherein the m.l.e.'s $\hat{\theta}_N$ do not converge to a unique value, because $\mathcal{E}_{\infty}[\theta]$ has a set of maximizing values,

$$\Theta_0 \equiv \{\theta: \mathcal{E}_{\infty}[\theta] = \mathcal{E}_{\infty}\} . \tag{4.1}$$

For a set F containing Θ_0, we shall write

$$\hat{\theta}_N \longrightarrow \Theta_0 \text{ (in F, w.p.1)} \tag{4.2}$$

if, excluding realizations $\{y_n(\omega)\}_{1\leq n<\infty}$ of $\{y_n\}_{1\leq n<\infty}$ which form an event of probability 0, every subsequence of $\{\hat{\theta}_N(\omega)\}_{1\leq N<\infty}$ contains a subsequence which is in F and which converges to a point in Θ_0. This is equivalent to saying that, given any neighborhood V of Θ_0 in F, the probability is 1 that only finitely many of the events $\hat{\theta}_N \notin V$, N = 1,2,..., occur. The result we are after is the following.

Proposition 4.1. Suppose there is a set F *containing the* $\mathcal{E}_\infty[\theta]$*–maximizing set* Θ_0 *defined in* (4.1) *above such that* (i) *with probability one, the log–likelihood functions* $L_N[\theta]$, $N \geq N_0$ *are continuous on* F; (ii) $N^{-1}L_N[\theta]$ *converges uniformly to* $\mathcal{E}_\infty[\theta]$ *on* F (w.p.1); *and* (iii) *the condition* (4.2) *is satisfied. Then* (3.5) *holds.*

Proof. It follows from (i) and (ii) that $\mathcal{E}_\infty[\theta]$ is continuous on F. Given $\delta > 0$, the set V = $\{\theta \in F: |\mathcal{E}_\infty[\theta] - \mathcal{E}_\infty| < \delta/2\}$ is thus an open set in F containing Θ_0. Therefore, by (4.2), the probability is one that only finitely many of the events $\hat{\theta}_N \notin V$ occur and also, by (ii), that only finitely many of the events

$$\sup_{\theta \in V}|M^{-1}L_M[\theta] - \mathcal{E}_\infty[\theta]| \geq \delta/2 \quad (M = 1,2,...)$$

occur. Since

$$|M^{-1}L_M[\theta] - \mathcal{E}_\infty| \leq |M^{-1}L_M[\theta] - \mathcal{E}_\infty[\theta]| + |\mathcal{E}_\infty[\theta] - \mathcal{E}_\infty| ,$$

it follows that, with probability one, at most finitely many of the events

$$\sup_{N \geq M}|M^{-1}L_M[\hat{\theta}_N] - \mathcal{E}_\infty| \geq \delta \quad (M = 1,2,...)$$

occur. This establishes the condition (3.5).

In many situations, the condition (ii) of the Proposition is a uniform law of large numbers, see Pötscher and Prucha (1989) and their references.

5. VERIFICATION OF (3.5): ARMA MODELS WITH POLES AND ROOTS BOUNDED OUTSIDE THE UNIT CIRCLE

We will now verify the conditions of Proposition 4.1 for some important classes of time series models. For each pair of non–negative integers p, q and each pair of non–negative real numbers α,β define $K^{p,q}_{\alpha,\beta}$ to be the set of rational functions k(z) of the form b(z)/a(z), where a(z) and b(z) are polynomials of degrees at most p, respectively q, satisfying a(0) = b(0) = 1, whose zeros belong to $\{|z|\geq 1+\alpha\}$ and $\{|z|\geq 1+\beta\}$, respectively. We shall show in Appendix A that $K^{p,q}_{\alpha,\beta}$ is compact, meaning that every sequence $k_n(z)$ has a subsequence $k_m(z)$ which converges to some $k(z)\in K^{p,q}_{\alpha,\beta}$ in the sense that the coefficients k_{jm} of the power series expansions $k_m(z) = \Sigma^{\infty}_{j=0} k_{jm}z^j$ ($|z|<1$) converge to k_j, where $k(z) = \Sigma^{\infty}_{j=0} k_jz^j$. We now assume that p,q,α,β are fixed, that $\alpha>0$ if $p>0$, and $\beta>0$ if $q>0$. (A separate discussion will be given in section 6 for unrestricted autoregressive models.) Define $\Theta = (0,\infty)\times K^{p,q}_{\alpha,\beta}$. With each $\theta = (\sigma^2,k)\in\Theta$ we associate a stationary, invertible ARMA model with spectral density function

$$f[\theta](\lambda) \equiv \frac{\sigma^2}{2\pi}|k(e^{i\lambda})|^2 .$$

Also, if we define $G_N[k]$ to be the covariance matrix whose (r,s)–entry is given by

$$(5.1)\qquad g_{r-s}[k] \equiv \frac{1}{2\pi}\int_{-\pi}^{\pi}|k(e^{i\lambda})|^2\cos(r-s)\lambda d\lambda ,$$

we can associate each θ with a Gaussian log–likelihood function $L_N[\theta]$ for a vector of N observations $Y_N = [y_1 \dots y_N]'$ by means of

$$(5.2)\qquad -2L_N[\theta] \equiv N\log 2\pi\sigma^2 + \log\det G_N[k] + \frac{1}{\sigma^2} Y_N' G_N^{-1}[k] Y_N .$$

It is known that $\det G_N[k] \geq 1$, see Lemma 3.5 of Dunsmuir and Hannan (1976).

Let us assume that the zero mean process y_n has stationary second moments to which the sample estimates converge almost surely. We also assume that the spectral distribution $F_y(\lambda)$ has a density $f_y(\lambda) \equiv dF_y/d\lambda$ which is not almost everywhere 0. Then a variety of relevant properties of $L_N[\theta]$ and of the related quantity

$$\sigma_N^2[k] \equiv \frac{1}{N} Y_N' G_N^{-1}[k] Y_N$$

hold with probability 1 when N is sufficiently large. Most of those listed in (5.3) below are given explicitly, or follow easily, from arguments in Deistler and Pötscher (1984), and Pötscher (1987), hereafter referred to as DP and P. They will help us show that $\hat{\theta}_N$ exists and belongs to a set F as required in Proposition 4.1.

(5.3) $L_N[\theta]$ <u>is continuous on</u> Θ, <u>and</u> $|G_N[k]|$ <u>and</u> $\sigma_N^2[k]$ <u>are continuous on</u> $K_{\alpha,\beta}^{p,q}$. (<u>Theorem</u> 3.3 <u>of</u> DP). $\sigma_N^2[k]$ <u>converges uniformly on</u> $K_{\alpha,\beta}^{p,q}$ <u>to</u>

$$\sigma_\infty^2[k] \equiv \int_{-\pi}^{\pi} |k(e^{i\lambda})|^{-2} dF_y(\lambda)$$

<u>with probability</u> 1. (established in the proof of Lemma 3.7 of P.)

It follows that $\sigma_\infty^2[k]$ is non-zero and continuous on $K_{\alpha,\beta}^{p,q}$ and therefore has a non-zero minimum $\sigma_{\min}^2$ and a finite maximum $\sigma_{\max}^2$ on this compact set. If we choose $\delta = \sigma_{\min}^2/2$, then since

$$\sup_{k \in K_{\alpha,\beta}^{p,q}} |\sigma_N^2[k] - \sigma_\infty^2[k]| < \delta$$

holds for all but finitely many N, with probability one, the same is true of

$$0 < \tfrac{1}{2}\,\sigma^2_{\min} \le \min_{k \in K^{p,q}_{\alpha,\beta}} \sigma^2_N[k] \le \max_{k \in K^{p,q}_{\alpha,\beta}} \sigma^2_N[k] \le \tfrac{3}{2}\,\sigma^2_{\max} \,.$$

Hence, with probability one, for sufficiently large N, the concentrated log–likelihood

$$L^c_N[k] = -\tfrac{N}{2}\log 2\pi e \sigma^2_N[k] - \tfrac{1}{2}\log|G_N[k]|$$

is finite and continuous on $K^{p,q}_{\alpha,\beta}$, and so has a maximizing value $\hat{k}_N$ in this compact set. Then $\hat{\theta}_N \equiv (\sigma^2_N[\hat{k}_N], \hat{k}_N)$ maximizes $L_N[\theta]$ over Θ. We conclude that, with probability 1, $\hat{\theta}_N$ exists and is an element of the compact set $F \equiv [\sigma^2_{\min}/2, 3\sigma^2_{\max}/2] \times K^{p,q}_{\alpha,\beta}$. By Lemma 3.2 and the proof of Lemma 3.7 of P, $N^{-1}L_N[\theta]$ converges on Θ and uniformly on F to

$$\mathcal{E}_\infty[\theta] = -\tfrac{1}{2}\log 2\pi\sigma^2 - \tfrac{1}{2}\,\sigma^2_\infty[k]/\sigma^2,$$

whose maximum value over Θ is given by

$$\mathcal{E}_\infty = -\tfrac{1}{2}\log 2\pi e \sigma^2_{\min} \,.$$

Consequently,

$$\lim_{N\longrightarrow\infty} N^{-1}L_N[\hat{\theta}_N] = \mathcal{E}_\infty \quad (\text{w.p. } 1),$$

verifying (2.3). Under the assumptions of Proposition 5.1 below, Theorem 7.4.10 of Hannan and Deistler (1988) establishes that

$$(5.4) \qquad \Theta_0 = \{(\sigma^2[k], k) \in \Theta;\ \sigma^2[k] = \sigma^2_{\min}\}$$

has the property (4.2). Thus Proposition 4.1 applies and we have

Proposition 5.1. Suppose that y_n is a purely non–deterministic covariance stationary time series whose one–step–ahead forecast error (innovations) process e^y_n, which determines the Wold representation $y_n = \Sigma^{\infty}_{j=0} k^y_j e^y_{n-j}$, has the property that its sample moments converge w.p.1 *to the process moments,*

$$(5.5) \qquad (N-j)^{-1} \sum_{n=j+1}^{N} e^y_n e^y_{n-j} \longrightarrow E e^y_n e^y_{n-j} \quad (\text{w.p.1}) \quad (j = 0,1,\ldots).$$

Then the property (3.5) *holds for the* ARMA(p,q) *Gaussian model log–likelihood family* $L_N[\theta]$, $\theta \in \Theta$, *with* $\Theta = (0, \infty) \times K^{p,q}_{\alpha,\beta}$, *assuming* $\alpha > 0$ *if* $p > 0$ *and* $\beta > 0$ *if* $q > 0$.

The somewhat abstract parametrization used above for ARMA models is needed when p,q>0 because of the possibilities that, in the large–sample limit, the highest order estimated coefficients will become zero, or that common roots will occur in the AR and MA polynomials. In these situations, the resulting $k_{\infty}(z)$ will have a multiplicty of representations of the form $k_{\infty}(z) = b(z)/a(z)$ with $\deg\{a(z)\} \leq p$ and $\deg\{b(z)\} \leq q$.

A more familiar parametrization can be used for the stationary models associated with the structural component models utilized in section 10. For these, the spectral densities have the form

$$(5.6) \qquad f[\sigma^2,\mu,\nu,\eta](\lambda) \equiv \sigma^2\{\mu|1-e^{i\lambda}|^4 + \nu|1-\eta e^{i\lambda}|^2|u(e^{i\lambda})|^2 + |1-e^{i\lambda}|^4|u(e^{i\lambda})|^2\},$$

where $u(z) = 1 + z + \dots + z^{11}$ and

$$(\sigma^2,\mu,\nu,\eta) \in \Theta \equiv (0,\infty)\times[0,\infty)\times[0,\infty)\times[0,1].$$

It is easy to see that $f[\sigma^2,\mu,\nu,\eta](\lambda)$ is non–zero everywhere in $[-\pi,\pi]$ (that is, the model is invertible) if and only if (σ^2,μ,ν,η) belongs to the subset

$$\Theta^* \equiv (0,\infty)\times(0,\infty)\times(0,\infty)\times[0,1),$$

and to verify that each compact set F in Θ^* corresponds to a set of (variance, transfer function)–pairs (σ^2,k) contained in a compact set of the form $[\sigma_0^2,\sigma_1^2]\times K_{0,\beta}^{0,13}$ with $0 < \sigma_0^2 < \sigma_1^2 < \infty$ and $\beta > 0$. If the true spectral density $f(\lambda)$ is bounded away from zero in a neighborhood of the zeros of $1-e^{i\lambda}$ and $u(e^{i\lambda})$, then the set Θ_0 defined in (5.3) will belong to Θ^*, and the reasoning leading to Proposition 5.1 shows that (3.5) holds.

On the other hand, if $f(\lambda)$ shares zeros with $1-e^{i\lambda}$ or $u(e^{i\lambda})$ (an indication of "overdifferencing") and if some parameter vector with $\mu = 0$ or $\nu = 0$ or $\eta = 1$ belongs to Θ_0, then all we can say at present is that, under (5.5), the basic convergence result (2.3) holds, now with $\mathcal{E}_\infty$ defined as the supremum of the $\mathcal{E}_\infty[\theta]$. This motivates the graph of (3.1) as a diagnostic. ((2.3) follows from (7.4.47) of Hannan and Deistler (1988, p. 347).) In our analyses, the graph of (3.2) in such situations has almost always resembled the graph of (3.1), see Figs. 6–8, which suggests that further research may yield a justification for the use of (3.2) with an overdifferenced model. (The one or two exceptions we observed may turn out to be explained by likelihood maximization difficulties.)

6. A STRENGTHENED FORM OF (3.5): AUTOREGRESSIONS WITH GAPS

There is a restricted class of ARMA models for which we have been able to establish somewhat deeper results, including an upper bound on the rate of convergence to 0 in (3.5) which shows that sample paths of $L_M[\hat{\theta}_N]$ will not differ greatly from $M\xi_\infty$. This is the class of autoregressions, some of whose coefficients might be constrained to be zero, whose parameters are estimated via sample moments (Yule–Walker estimates). We assume in this section that y_n is a mean zero, stationary time series whose lag j autocovariance $\gamma_j \equiv Ey_ny_{n-j}$ is estimated by

$$\hat{\gamma}_j(N) \equiv \frac{1}{N} \sum_{n=|j|+1}^{N} y_n y_{n-|j|} \quad (j=0,\pm1,\dots,\pm(N-1)) .$$

We assume that all autocovariance matrices $\Gamma_m = [\gamma_{i-j}]_{0 \leq i,j \leq m-1}$ are non–singular. Set

$$\hat{f}_N(\lambda) \equiv (2\pi)^{-1} \sum_{j=-N+1}^{N-1} \hat{\gamma}_j(N)\cos j\lambda \ (-\pi \leq \lambda \leq \pi).$$

Given a vector of lags, $\ell \equiv (\ell_1,\dots,\ell_m)$, with $1\leq\ell_1<\dots<\ell_m$, we define the coefficient vector $\phi^\ell = [\phi_1 \dots \phi_m]'$ to be the solution of

$$(6.1) \qquad [\gamma_{\ell_i-\ell_j}]_{1\leq i,j\leq m}\, \phi^\ell = [\gamma_{\ell_1} \dots \gamma_{\ell_m}]' .$$

The coefficient matrix in (6.1) is a submatrix of Γ_{ℓ_m} and so is nonsingular. We define the error process e_n^ℓ by mean of

$$(6.2) \qquad y_n = \phi_1 y_{n-\ell_1} + \dots + \phi_m y_{n-\ell_m} + e_m^\ell .$$

Therefore e_n^ℓ is a mean zero, stationary process which, from (6.1), is uncorrelated with $y_{n-\ell_1}, \dots, y_{n-\ell_m}$, so its variance is given by

(6.3) $$\sigma_{\ell}^{2} \equiv \gamma_0 - \phi_1\gamma_{\ell_1} - \cdots - \phi_m\gamma_{\ell_m} .$$

Given data $y_1,\ldots,y_N$, we can fit an autoregression with gaps of the form (6.2) by replacing the autocovariances in (6.1) with their sample estimates. If $\hat{\phi}^{\ell}(N) \equiv [\hat{\phi}_1(N) \ldots \hat{\phi}_m(N)]$ is the resulting coefficient vector and if we define

(6.4) $$\hat{\sigma}_{\ell}^{2}(N) \equiv \hat{\gamma}_0(N) - \hat{\phi}_1(N)\hat{\gamma}_{\ell_1}(N) - \cdots - \hat{\phi}_m(N)\hat{\gamma}_{\ell_m}(N) ,$$

then the pair $\hat{\sigma}_{\ell}^{2}(N)$, $\hat{\phi}^{\ell}(N)$ is the unique maximizer of

(6.5) $$L_N[\sigma^2, \phi^{\ell}] = -\frac{N}{2}\{\log 2\pi\sigma^2 + \frac{1}{\sigma^2}\int_{-\pi}^{\pi} |\phi^{\ell}(e^{i\lambda})|^2 \hat{f}_N(\lambda)d\lambda\} ,$$

where $\phi^{\ell}(e^{i\lambda}) \equiv 1 - \phi_1 e^{i\ell_1\lambda} - \phi_m e^{i\ell_m\lambda}$. The maximum value is

$$\hat{L}_N^{(\ell)} = -\frac{N}{2}\log 2\pi e\hat{\sigma}_{\ell}^{2}(N)$$

Under conditions weaker than the more easily stated assumptions of Proposition 6.1 below, Theorem 7.4.3 of Hannan and Deistler (1988) establishes that

(6.6) $$\sup_{1\leq j<\infty} |\gamma_j - \hat{\gamma}_j(N)| = O\{(\log N/N)^{1/2}\} \quad (\text{w.p.1}) .$$

In (6.6), $\hat{\gamma}_j(N) \equiv 0$ if $j \geq N$. A straightforward argument based on (6.6), which we omit, leads to (6.7) and (6.8) below.

Proposition 6.1. Suppose that y_n *has a linear representation*

$$y_n = \sum_{j=0}^{\infty} k_j e_{n-j}$$

in which $e_n \sim \text{I.I.D.}(0,\sigma^2)$ *and* $Ee_n^4 < \infty$. *Suppose, too, that* $k(z) = \Sigma_{j=0}^{\infty} k_j z^j$ *is non–zero on* $\{|z| \leq 1\}$ *and that* (i) and (ii) hold:

(i) $$\sum_{j=1}^{\infty} j^{1/2}|k_j| < \infty .$$

(ii) $$\sup_j j|k_j| < \infty .$$

Then $\hat{\phi}_1(N),\dots,\hat{\phi}_m(N)$ *and* $\hat{\sigma}_\ell^2(N)$ *satisfy*

(6.7) $$\sup_{1 \leq j \leq m} |\phi_j - \hat{\phi}_j(N)| = O\{(\log N/N)^{1/2}\} \quad (\text{w.p.1})$$

and

(6.8) $$|\sigma_\ell - \hat{\sigma}_\ell^2(N)| = O\{(\log N/N)^{1/2}\} \quad (\text{w.p.1}).$$

Set $\hat{\theta}_N^\ell \equiv [\hat{\sigma}^2(\ell)\ \hat{\phi}_1(N) \dots \hat{\phi}_m(N)]'$ and $\mathcal{E}_\infty^\ell \equiv (-1/2)\log 2\pi\sigma_\ell^2$. Note that

$$M^{-1}L_M[\hat{\theta}_N] - \mathcal{E}_\infty^\ell = -\tfrac{1}{2}\log\left(1 + \left(\frac{\hat{\sigma}_\ell^2(N) - \sigma_\ell^2}{\sigma_\ell^2}\right)\right).$$

Since $|\log(1 + x) - x| \leq 2x^2$ when $|x| \leq 1/2$, and since $\log N/N$ is a decreasing function for $N \geq 3$, a strengthened form of (3.5) follows from (6.8):

Corollary 6.1. Under the assumptions of Proposition 4.1, the bound (6.9) holds:

$$(6.9) \qquad \sup_{N\geq M} |M^{-1}L_M[\hat{\theta}_N^{\ell}] - \mathcal{E}_{\infty}^{\ell}| = O\{(\log M/M)^{1/2}\} \quad (\text{w.p.1}) .$$

It follows that if two such autoregressive models, with lag vectors ℓ and $\tilde{\ell}$ respectively, are being considered for y_n, and if we define $\mathcal{E}_{\infty}^{\ell,\tilde{\ell}} = (-1/2)\log\sigma_{\ell}^2/\sigma_{\tilde{\ell}}^2$, then $M^{-1}\{L_M[\hat{\theta}_N^{\ell}] - L_M[\hat{\theta}_N^{\tilde{\ell}}]\}$ converges to $\mathcal{E}_{\infty}^{\ell,\tilde{\ell}}$ uniformly over $N/2 < M \leq N$ as $N \longrightarrow \infty$ (w.p.1). The two autoregressive models will be non–nested if each lag vector has an entry not shared with the other. Autoregressions with gaps have been proposed as alternatives to ARMA models, see Kenny and Durbin (1982).

7. LOG–LIKELIHOODS FOR ARIMA MODELS AND (3.5)

The models we wish to compare in section 10 are nonstationary ARIMA models. The nonstationarity introduces some subtle complications which we address in this section. Consider the situation in which a stationarizing backshift–operator polynomial $\delta(B)$ of degree d with $\delta(0) = 1$ is applied to the observed series $y_1,\ldots,y_N$ to obtain the data $w_n \equiv \delta(B)y_n$, $n = d+1,\ldots,N$ which are actually modeled, from a log–likelihood family $L_{N,d}[\theta] = L[\theta](w_{d+1},\ldots,w_N)$. If we let $L_d(y_1,\ldots,y_d)$ denote the (unknown) true log–density of $y_1,\ldots,y_d$, then

$$(7.1) \qquad L_N[\theta] \equiv L_{N,d}[\theta] + L_d$$

is a log–likelihood for $y_1,\ldots,y_N$, in the sense that

$$\int_{\mathbb{R}^N} \exp(L_N[\theta])dy_1 \cdots dy_N = 1 \quad ,$$

if the integral of $\exp(L_{N,d}[\theta])$ over $\mathbb{R}^{N-d}$ is 1: indeed, since the Jacobian of the transformation $(y_1,\ldots,y_N) \longrightarrow (y_1,\ldots,y_d,w_{d+1},\ldots,w_N)$ is 1, we have

$$\int_{\mathbb{R}^N} \exp(L_N[\theta])dy_1 \cdots dy_N =$$

$$\int_{\mathbb{R}^d} \exp(L_d)dy_1 \cdots dy_d \int_{\mathbb{R}^{N-d}} \exp(L_{N,d}[\theta])dw_{d+1} \cdots dw_N$$

$$= \int_{\mathbb{R}^{N-d}} \exp(L_{N,d}[\theta])dw_{d+1} \cdots dw_N \ .$$

For a more general discussion, see Findley (1990a). Note that $L_N[\theta^{(0)}]$ will be the correct log density for $y_1,\ldots,y_N$ if $L_{N,d}[\theta^{(0)}]$ is the correct log density of $w_{d+1},\ldots,w_N$, and if, at the same time, $y_1,\ldots,y_d$ are <u>independent</u> of the w_n process. This independence is usually assumed, in order to insure that the one–step–ahead forecast–error (innovations) process of y_n coincides with that of w_n, see Bell (1984).

Remark 7.1. In the case of "overdifferencing", when some proper divisor $\delta^{(0)}(B)$ of degree $d^{(0)} < d$ transforms y_t into a stationary series $w_n^{(0)} = \delta^{(0)}(B)y_t$, there is an issue concerning $L_N[\theta]$ as defined in (7.1) which deserves mention. In this case, although $L_N[\theta]$ can properly be called a model log–likelihood function for $y_1,\ldots,y_N$, it cannot, in general, describe the correct log–likelihood of $y_1,\ldots,y_N$, because $L_{N,d}[\theta](w_{d+1},\ldots,w_N)$ will not be the conditional log–likelihood of $w_{d+1},\ldots,w_N$ given $y_1,\ldots,y_d$. The problem is that the independence of the $w_n^{(0)}$ from $y_1,\ldots,y_d(0)$ can preclude the independence of $w_n = \delta(B)y_t$ from $y_{d(0)+1},\ldots,y_d$, as simple calculations with $\delta^{(0)}(B) = 1-B$ and $\delta(B) = 1-B^2$ reveal. Perhaps the error in $L_N[\theta]$ in this situation will be unimportant when N is large, a possibility that deserves further investigation. Another technical problem that arises when $\delta(B)$ is too large is the presence of zeros in the spectral density of w_t, which places $L_{N,d}[\theta]$ beyond the reach of the verifications of (3.5) given in this paper. With overdifferencing, it will still be possible, usually, to verify (2.3), so the graph of (3.1) can be used, see below.

Suppose we have candidate transformations $\delta^{(j)}(B)$ of degree $d^{(j)}$ and candidate m.l.e. models $\hat{L}^{(j)}_{N,d}(j) \equiv L_{N,d}(j)[\hat{\theta}^{(j)}_{N,d}(j)]$ for the data $w^{(j)}_n = \delta^{(j)}(B)y_n$, $n = d^{(j)}+1,\ldots,N$, $j = 1,2$. Then the log–likelihood difference $\hat{L}^{(1,2)}_N \equiv L_N[\hat{\theta}^{(1)}_{N,d}(1)] - L_N[\hat{\theta}^{(2)}_{N,d}(2)]$ satisfies

$$\hat{L}^{(1,2)}_N = \hat{L}^{(1)}_{N,d}(1) - \hat{L}^{(2)}_{N,d}(2) + \{L_d(1) - L_d(2)\}. \tag{7.2}$$

So $\hat{L}^{(1,2)}_N$ *is known to within a summand which is a function of* $y_1,\ldots,y_{\max\{d^{(1)},\, d^{(2)}\}}$. Of course, when $d^{(1)} = d^{(2)}$, then $\hat{L}^{(1,2)}_N$ is known,

$$\hat{L}^{(1,2)}_N = \hat{L}^{(1)}_{N,d}(1) - \hat{L}^{(2)}_{N,d}(2) \qquad (d^{(1)} = d^{(2)}), \tag{7.3}$$

and the graphical procedures of section 2 can be applied. In fact, since $L_d(1) - L_d(2)$ does not change with N, it follows from (7.2) that, whether or not $d^{(1)} = d^{(2)}$,

$$\hat{L}^{(1,2)}_N \longrightarrow \pm\infty \ \text{if and only if}\ \hat{L}^{(1)}_{N,d}(1) - \hat{L}^{(2)}_{N,d}(2) \longrightarrow \pm\infty, \tag{7.4}$$

so the graph of $\hat{L}^{(1)}_{M,d}(1) - \hat{L}^{(2)}_{M,d}(2)$, $N/2 < M \leq N$ *can be examined to see if an ultimate direction for* $\hat{L}^{(1,2)}_N$ *is suggested*.

When $d^{(1)} \neq d^{(2)}$, we shall refer to the quantities in (7.5) and (7.6) below as *pseudo–log–likelihood–ratios* (pseudo–LLR's):

$$\hat{L}^{(1)}_{M,d}(1) - \hat{L}^{(2)}_{M,d}(2), \quad N/2 < M \leq N\,. \tag{7.5}$$

$$(7.6) \qquad L_{M,d^{(1)}}[\hat{\theta}_N^{(1)}] - L_{M,d^{(2)}}[\hat{\theta}_N^{(2)}], \ \ N/2 < M \leq N.$$

In section 10, graphs of (7.5) and (7.6) are given for two series with models having $d^{(1)} \neq d^{(2)}$, see Figs. 3(c), (d) and 7(c), (d).

Remark 7.2. To make the theoretical situation concerning the case $d^{(1)} \neq d^{(2)}$ clearer, we point out that if both $w_n^{(1)}$ and $w_n^{(2)}$ are (covariance) stationary (and have continuous spectral distribution functions), then it follows from the result (3.1) of Findley (1985a) that there is a common divisor $\delta^{(0)}(B)$ of $\delta^{(1)}(B)$ and $\delta^{(2)}(B)$ such that $w_n^{(0)} = \delta^{(0)}(B)y_n$ is stationary. Thus Remark 7.1 applies. On the other hand, if, say, $\delta^{(1)}(B)$ is a divisor of $\delta^{(2)}(B)$ and $w_n^{(2)}$ is stationary but $w_n^{(1)}$ is not ("underdifferencing"), then the movement of $\hat{L}_N^{(1,2)}$ toward $-\infty$ can be at a rate proportional to N^r with $r > 1$. For autoregressive models, this follows from results of Tiao and Tsay (1983) or Chan and Wei (1988).

Remark 7.3. If the approach described here to defining $\hat{L}_N^{(1,2)}$ for ARIMA models is used, there are some implications concerning the applicability of Akaike's AIC criterion. Consider the difference of AIC values,

$$(7.7) \qquad \Delta AIC_N^{(1,2)} \equiv -2\hat{L}_N^{(1,2)} + 2(\dim\theta^{(1)} - \dim\theta^{(2)}),$$

where $\dim\theta^{(j)}$ denotes the number of estimated parameters in the j–th model family, j=1,2. It is clear from (7.2) than when $d^{(1)} \neq d^{(2)}$, since $L_d^{(1)} - L_d^{(2)}$ has non–zero mean, the calculable analogue of (7.7),

$$(7.8) \qquad -2\{\hat{L}_{N,d}^{(1)}(1) - \hat{L}_{N,d}^{(2)}(2)\} + 2(\dim\theta^{(1)} - \dim\theta^{(2)}),$$

will not have the same asymptotic mean as the uncalculable quantity $\Delta \mathrm{AIC}_N^{(1,2)}$ when the means of the sequence $\hat{L}_N^{(1,2)}$ converge. As a consequence, in this case the bias calculations motivating the use of AIC (see Findley (1985b) and Findley and Wei (1989)) do not support the use of the sign of (7.8) for model selection. Of course, if $\hat{L}_N^{(1,2)} \rightarrow_p \pm \infty$, the finite bias correction term $2(\dim\theta^{(1)} - \dim\theta^{(2)})$ is inconsequential.

There is another technical problem in the situation of "overdifferencing", regardless of whether $d^{(1)} \neq d^{(2)}$ or $d^{(1)} = d^{(2)}$: here the spectral density function of at least one of the series, $w_n^{(j)}$, j=1,2 may have a zero, and in this situation, it does not appear to be known if the Fisher information matrix plays the role in the limiting distribution of $N^{1/2}(\hat{\theta}_N^{(j)} - \theta_\infty^{(j)})$ needed for the derivation of AIC, see Findley (1985b).

8. WEAKLY EQUIVALENT MODELS AND THE ASYMPTOTIC DISTRIBUTION OF $N^{-1/2}\hat{L}_N^{(1,2)}$.

The discussion so far is completely general in the sense that it applies both to nested and non–nested model comparisons of model classes which might or might not contain the correct model. In sections 9 and 10, we will discuss a hypothesis testing procedure for reaching the same conclusions about the limiting behavior of $\hat{L}_N^{(1,2)}$ which takes advantage of a component of the log–likelihood–ratio which ordinarily only occurs when non–nested and incorrect models are fit, a phenomenon we shall explain precisely in this section with two easy Propositions and a somewhat deeper result. The properties of interest concern the models defined by the large–sample limits of the m.l.e. models.

In this section, y_n denotes a mean zero, purely non–deterministic stationary time series with innovations representation

$$(8.1) \qquad y_n = \sum_{j=0}^{\infty} k^y_j \, e^y_{n-j} \quad (k^y_0 = 1),$$

$$= k^y(B)e^y_n$$

and spectral density $f_y(\lambda)$. In sections 5 and 6, the assumption that the candidate model (innovations) transfer function $k(z) = \Sigma^{\infty}_{j=0} k_j z^j$ $(k_0 = 1)$ was a rational function simplified the presentation. That is not the case for this section, so no special form will be assumed. We suppose that $k(z) \neq 0$ if $|z| < 1$. If $k(z)$ is such that

$$(8.2)^{*} \qquad \sigma^2[k] \equiv \int_{-\pi}^{\pi} |k(e^{i\lambda})|^{-2} f_y(\lambda) d\lambda$$

is finite, then $\sigma^2[k]$ is the variance of the covariance stationary series $e_n[k] \equiv k(B)^{-1} y_n$, and there is a moving average representation for y_n with transfer function $k(z)$,

$$(8.3) \qquad y_n = \sum_{j=0}^{\infty} k_j e_{n-j}[k] \, .$$

The candidate spectral density function for y_n defined by

$$(8.4) \qquad f[k](\lambda) \equiv \frac{\sigma^2[k]}{2\pi} |k(e^{i\lambda})|^2$$

coincides with the true spectral density $f_y(\lambda)$ if and only if $e_n[k]$ is a white noise process (an uncorrelated series), in which case the representations (8.1) and (8.3) are identical, $k_j = k^y_j$ and $e_{n-j}[k] = e^y_{n-j}$, j=0,1,... . We shall explain shortly why it is appropriate, in general, to regard $e_n[k]$ as a one-step-ahead forecast-error process and $\sigma^2[k]$ as the mean square forecast error. Thus $\sigma^2[k]$ defined by (8.2) is a theoretical goodness–of–fit measure for transfer function models k(z).

We shall say that two transfer function models $k^{(1)}(z)$ and $k^{(2)}(z)$ are weakly equivalent if $\sigma^2[k^{(1)}] = \sigma^2[k^{(2)}]$. There are two theoretically important modeling situations in which weak equivalence implies that the models coincide, $k^{(1)}(z) = k^{(2)}(z)$.

Proposition 8.1. Suppose the transfer function model k(z) for y_n is weakly equivalent to the true innovations transfer function $k^y(z) = \Sigma_{j=0}^{\infty} k_j^y z^j$; that is, suppose $\sigma^2[k]$ coincides with the innovations variance $\sigma_y^2 = E(e_n^y)^2$. Then k(z) is correct, $k(z) = k^y(z)$. On the other hand, if k(z) and $k^y(z)$ are not weakly equivalent, then $\sigma^2[k] > \sigma_y^2$.

Proof. Note that, since $k^y(0) = k(0) = 1$,

$$e_n[k] - e_n^y = \{1/k^y(B) - 1/k(B)\}y_n$$

is a linear function of y_{n-j}, $j \geq 1$, and so is uncorrelated with e_n^y. Because $e_n[k] = e_n^y + \{e_n[k] - e_n^y\}$, we therefore have $\sigma^2[k] = \sigma_y^2 + E\{e_n[k] - e_n^y\}^2$ from which the asserted results follow.

Our null hypothesis in section 9 will hypothesize weakly equivalent models with distinct covariance structures. Proposition 8.1 above shows that the situation where one of the models is correct is excluded by this hypothesis. We observe next that nested models are also excluded when the m.l.e.'s from the larger class of models have a unique limit. The result is obvious but fundamental.

Proposition 8.2. Let $k[\theta^{(j)}](z)$, $\theta^{(j)} \in \Theta^{(j)}$, i=1,2 be two families of innovations transfer function models for y_n such that (i) – (iii) hold:

(i) $\Theta^{(1)} \subseteq \Theta^{(2)}$.

(ii) *There exist* $\theta_\infty^{(j)} \in \Theta^{(j)}$ *such that*

$$\sigma^2[\theta_\infty^{(j)}] = \int |k[\theta_\infty^{(j)}](e^{i\lambda})|^{-2} f_y(\lambda)d\lambda$$

is finite and equal to $\inf_{\theta^{(j)} \in \Theta^{(j)}} \sigma^2[k[\theta^{(j)}]]$ $(j = 1,2)$.

(iii) *There is only one such minimizer* $\theta_\infty^{(2)}$ *in* $\Theta^{(2)}$.

Then, if $k[\theta_\infty^{(1)}]$ and $k[\theta_\infty^{(2)}]$ *are weakly equivalent models, that is, if* $\sigma^2[\theta_\infty^{(1)}] = \sigma^2[\theta_\infty^{(2)}]$, *then* $k[\theta_\infty^{(1)}](z) = k[\theta_\infty^{(2)}](z)$.

We turn now to the derivation of the limiting distribution of $N^{-1/2}\hat{L}_N^{(1,2)}$ in the situation in which the best transfer function models from the two competing families are weakly equivalent but not coincident. We will utilize the conditional decomposition of the Gaussian log–likelihood function for the covariance structure for y_n specified by $f[k](\lambda)$ defined in (8.4). Let $y_{n|n-1}[k]$ denote the linear function of $y_{n-1},\dots,y_1$ defining the best linear predictor of y_n (the Gaussian conditional mean) under the model $f[k](\lambda)$, and let $\sigma^2_{n|n-1}[k]$ denote the mean square of

$$e_{n|n-1}[k] = y_n - y_{n|n-1}[k]$$

(the Gaussian conditional variance) calculated under $f[k](\lambda)$. The interpretation of $e_n[k]$ and $\sigma^2[k]$ as prediction error quantities arises naturally from the fact that the differences $e_n[k] - e_{n|n-1}[k]$ and $\sigma^2[k] - \sigma^2_{n|n-1}[k]$ tend to zero as $n \longrightarrow \infty$. We need a specific rate of convergence. Baxter (1962) provides rates of convergence for the case in which $f[k](\lambda)$ is the true spectral density of y_n, but minor and straightforward modifications of his proofs show that the same rates apply whatever the autocovariance structure of y_n, see also Findley (1990b). We state the variant of his Proposition 3.1 that we need as

Proposition 8.3. Suppose that $k(z) = \Sigma_{j=0}^{\infty} k_j z^j$ satisfies (i) and (ii):

(i) $k(z) \neq 0$ for all $|z| \leq 1$.

(ii) $\sum_{j=1}^{\infty} j^{\alpha} |k_j| < \infty$, for some $\alpha \geq 0$.

Then (8.5) and (8.6) hold:

$$\lim_{n \to \infty} n^{2\alpha}\{\sigma^2[k] - \sigma^2_{n|n-1}[k]\} = 0. \tag{8.5}$$

$$\lim_{n \to \infty} n^{2\alpha} E\{e_n[k] - e_{n|n-1}[k]\}^2 = 0\ . \tag{8.6}$$

These convergence results lead to a useful approximation to the Gaussian log–likelihood determined by $f[k](\lambda)$, which is obtained from (5.2) by setting $\theta = (\sigma^2[k], k)$ and which we will denote by $L_N[k]$. This log–likelihood has a decompositon like (3.4),

$$L_N[k] = -\frac{1}{2} \sum_{n=1}^{N} \left\{ \log 2\pi\sigma^2_{n|n-1}[k] + \sum_{n=1}^{N} \frac{e^2_{n|n-1}[k]}{\sigma^2_{n|n-1}[k]} \right\} . \tag{8.7}$$

The following approximation result is proved in Appendix B.

Corollary 8.1. Suppose that the innovations transfer function model k *for* y_n *satisfies the hypotheses of Proposition 8.3 with* $\alpha = 1/2$. *Then* $N^{-1/2}L_N[k]$ *can be approximated in mean absolute deviation as indicated in* (8.8):

$$\text{(8.8)}\qquad \lim_{N \to \infty} N^{-1/2} E|L_N[k] + \tfrac{1}{2}\{N\log 2\pi\sigma^2[k] + \sum_{n=1}^{N} e_n^2[k]/\sigma^2[k]\}| = 0 .$$

If two such transfer functions $k^{(1)}$ and $k^{(2)}$ are given, consider the mean zero series

$$\text{(8.9)}\qquad \zeta_n^{(1,2)} = \frac{e_n^2[k^{(1)}]}{\sigma^2[k^{(1)}]} - \frac{e_n^2[k^{(2)}]}{\sigma^2[k^{(2)}]} ,$$

and suppose that the series y_n is fourth–order stationary. Then $\zeta_n^{(1,2)}$ is a covariance stationary series. Let $f^{(1,2)}(\lambda)$ denote its spectral density function. If y_n is strictly stationary and satisfies Assumption 2.6.1 of Brillinger (1975), requiring the absolute convergence of the cumulants of each order, then so does $\zeta_n^{(1,2)}$, and Brillinger's Theorem 4.4.1 applies, resulting in

$$\text{(8.10)}\qquad N^{-1/2} \sum_{n=1}^{N} \zeta_n^{(1,2)} \to_{\text{dist.}} \mathcal{N}(0,2\pi f^{(1,2)}(0)) .$$

If we set $L_N^{(1,2)} \equiv L_N[k^{(1)}] - L_N[k^{(2)}]$, it follows from (8.8) and (8.10) that

$$\text{(8.11)}\qquad 2N^{-1/2}L_N^{(1,2)} + N^{1/2}\log\frac{\sigma^2[k^{(1)}]}{\sigma^2[k^{(2)}]} \to_{\text{dist.}} \mathcal{N}(0,2\pi f^{(1,2)}(0)) .$$

The next result, concerning the asymptotic distribution of $2N^{-1/2}\hat{L}_N^{(1,2)}$ for log–likelihood differences $\hat{L}_N^{(1,2)}$ of *estimated* models, now follows immediately via the decompositions

$$(8.12) \quad N^{-1/2}\hat{L}_N^{(j)} = N^{-1/2}\{L_N[\hat{\theta}_N^{(j)}] - L_N[\theta_\infty^{(j)}]\} + N^{-1/2}\, L_N[\theta_\infty^{(j)}] \quad (j=1,2),$$

because the first expression on the right tends to zero in the usual situations, as we explain below.

Proposition 8.4. Let y_n be a strictly stationary, zero–mean time series satisfying Assumption 2.6.1 of Brillinger (1975, p. 26). *Suppose that $L_N[\theta^{(j)}]$, $\theta^{(j)} \in \Theta^{(j)}$, j=1,2 are Gaussian log–likelihood functions families for $y_1,\dots,y_N$ having maximizing values $\hat{\theta}_N^{(1)}$ and $\hat{\theta}_N^{(2)}$ which converge in probability as $N \longrightarrow \infty$ to limiting values $\theta_\infty^{(1)}$ and $\theta_\infty^{(2)}$. Suppose these define innovations filters $k^{(1)} = k[\theta_\infty^{(1)}]$ and $k^{(2)} = k[\theta_\infty^{(2)}]$, respectively, which satisfy the assumptions of Corollary 8.1. Let $\sigma^2[k^{(1)}]$ and $\sigma^2[k^{(2)}]$ denote the associated innovations variances, defined as in* (8.2), *and let $f^{(1,2)}(\lambda)$ denote the spectral density of the series $\zeta_n^{(1,2)}$ defined by* (8.9).

Then if the log–likelihood difference sequences $L_N[\hat{\theta}_N^{(1)}] - L_N[\theta_\infty^{(1)}]$ and $L_N[\hat{\theta}_N^{(2)}] - L_N[\theta_\infty^{(2)}]$, N=1,2,... are bounded in probability, it follows that the log–likelihood difference sequence $\hat{L}_N^{(1,2)} \equiv L_N[\hat{\theta}_N^{(1)}] - L_N[\hat{\theta}_N^{(2)}]$ satisfies (8.13):

$$(8.13) \qquad 2N^{-1/2}\hat{L}_N^{(1,2)} + N^{1/2}\log \frac{\sigma^2[k^{(1)}]}{\sigma^2[k^{(2)}]} \longrightarrow_{\text{dist.}} \mathcal{N}(0,2\pi f^{(1,2)}(0)) \; .$$

Concerning the boundedness–in–probability assumption, we note that if $L_N[\theta^{(j)}]$ is a differentiable function of a vector parameter $\theta^{(j)}$ and if $\hat{\theta}_N^{(j)}$ is in the interior of the parameter set, so that the gradient $\dot{L}_N[\hat{\theta}_N^{(j)}] \equiv \partial L_N[\hat{\theta}_N^{(j)}]/\partial\theta^{(j)}$ is zero, it follows via Taylor's formula that there is a $\bar{\theta}_N^{(j)}$ on the line segment between $\hat{\theta}_N^{(j)}$ and $\theta_\infty^{(j)}$ such that

$$2\{L_N[\theta_\infty^{(j)}] - L_N[\hat{\theta}_N^{(j)}]\} =$$

$$\{N^{1/2}(\hat{\theta}_N^{(j)} - \theta_\infty^{(j)})\}'\{N^{-1}\ddot{L}_N[\bar{\theta}_N^{(j)}]\}\{N^{1/2}(\hat{\theta}_N^{(j)} - \theta_\infty^{(j)})\},$$

for $j = 1,2$. Hence, the boundedness in probability of these sequences will follow from convergence in distribution of $N^{1/2}(\hat{\theta}_N^{(j)} - \theta_\infty^{(j)})$ and convergence in probability of the sample–size normalized Hessian matrices $N^{-1}\ddot{L}_N[\bar{\theta}_N^{(j)}]$.

This proposition shows that if $f^{(1,2)}(0) \neq 0$ and if a consistent estimator $\hat{\sigma}_N^{(1,2)^2}$ of $2\pi f^{(1,2)}(0)$ can be found, then when $\sigma^2[k^{(1)}] = \sigma^2[k^{(2)}]$,

(8.14) $$Z^{(1,2)}(N) \equiv 2N^{-1/2}\hat{L}_N^{(1,2)}/\hat{\sigma}_N^{(1,2)} \longrightarrow_{\text{dist.}} \mathcal{N}(0,1)\ .$$

The test statistic $Z^{(1,2)}(N)$ can be used to test the hypothesis $\sigma^2[k^{(1)}] = \sigma^2[k^{(2)}]$ against the two–sided alternative $\sigma^2[k^{(1)}] \neq \sigma^2[k^{(2)}]$ and the associated one–sided alternatives.

This test is a time series analogue of the test presented in Vuong (1989) for the case of i.i.d. observations. Vuong shows that, in the i.i.d. case, under rather general assumptions, if density functions $g^{(1)}[\theta^{(1)}](y)$ and $g^{(2)}[\theta^{(2)}](y)$ are being fit to $y_1,\ldots,y_N$, then

$$\hat{\sigma}_N^{(1,2)^2} = N^{-1}\sum_{n=1}^{N}\{\log g^{(1)}[\hat{\theta}_N^{(1)}](y_n)/g^{(2)}[\hat{\theta}_N^{(2)}](y_n)\}^2$$

is a consistent estimator of the asymptotic variance of $N^{-1/2}\hat{L}_N^{(1,2)}$, and that $\hat{V}_N^{(1,2)} \equiv N^{-1/2}\hat{L}_N^{(1,2)}/\hat{\sigma}_N^{(1,2)}$ has a $\mathcal{N}(0,1)$ limiting distribution when $\mathcal{E}_\infty^{(1)} \equiv E\{\log g^{(1)}[\theta_\infty^{(1)}]\}$ is equal to $\mathcal{E}_{\infty}^{(2)} \equiv E\{\log g^{(2)}[\theta_{\infty}^{(2)}]\}$. (Our notation differs from Vuong's.) In the discussion of the hypothesis test associated with $\hat{V}_N^{(1,2)}$ following the statement of Theorem 5.1 of Vuong (1989), it is assumed that the sign of $\hat{L}_N^{(1,2)}$ indicates the direction of linear movement of $\hat{L}_N^{(1,2)}$ toward ± ∞ if the null hypothesis of no such movement is rejected. While this must be correct for large enough N, it need not be true for all N, and the analysis of the series *blqrrs* in section 10 suggests that the sign of $\hat{L}_N^{(1,2)}$ could be misleading in situations encountered in practice. It therefore seems prudent to use Vuong's test in conjunction with a graphical diagnostic like those discussed in section 3.

Remark 8.1. The distributional result (8.13) concerning $\hat{L}_N^{(1,2)}$ also holds if the competing models are invertible ARIMA models with the same stationarizing polynomials $\delta^{(1)}(B) = \delta^{(2)}(B)$, providing $\hat{L}^{(1,2)}$ is defined as in section 7. In this case, the series $e_n^{(1)}$ and $e_n^{(2)}$ defining $\zeta_n^{(1,2)}$ are obtained from the stationary series $\delta^{(1)}(B)y_n = \delta^{(2)}(B)y_n$. On the other hand, if, say, $\delta^{(1)}(B)$ is a proper divisor of $\delta^{(2)}(B)$ and $\delta^{(1)}(B)y_n$ is stationary, so that $\delta^{(2)}(B)$ represents "overdifferencing", then the result (8.10) holds as before but the status of (8.8) for $\delta^{(2)}(B)y_n$ and of (8.13) is uncertain , because $k^{(2)}(\lambda)$ may or may not have zeros, see Corollary 3.1 of Pötscher (1990). In practice, we have found that the use of $\hat{L}_{N,d}^{(1)}(1) - \hat{L}_{N,d}^{(2)}(2)$ in place of $\hat{L}_N^{(1,2)}$ in $Z^{(1,2)}(N)$ when $d^{(1)} \neq d^{(2)}$ often leads to incorrect conclusions: the difference in the numbers of observations utilized for the likelihood calculation does not seem to be negligible in its effect at the usual sample sizes, even though this effect must vanish as $N \longrightarrow \infty$ because of the divisor $N^{1/2}$ in $Z^{(1,2)}(N)$. We do not recommend the use of hypothesis tests based on (8.14) when $\delta^{(1)}(B) \neq \delta^{(2)}(B)$.

Our efforts to verify that natural estimators of $2\pi f^{(1,2)}(0)$ are consistent, and in this way obtain a theoretically complete time series analogue of Vuong's test, have only been partially successful, as we shall explain in the next section.

9. TIME SERIES ANALOGUES OF VUONG'S STATISTIC

To be able to use the distributional result (8.14) for testing the asymptotic weak equivalence of two maximum likelihood time series models, we need an estimator $\hat{\sigma}^{(1,2)^2}(N)$ of $2\pi f^{(1,2)}(0)$, where $f^{(1,2)}(\lambda)$ is the spectral density of the unobserved process $\zeta_n^{(1,2)}$ defined in (8.8). One approach is to use an autoregressive spectrum estimator, with order determined by Akaike's minimum AIC criterion, considering orders up to $N^{1/2}$, see Berk (1974) and Shibata (1981). This autoregression is fit to the process

$$(9.1) \qquad \hat{\zeta}_n^{(1,2)} = \frac{e^2_{n|n-1}[\hat{\theta}_N^{(1)}]}{\sigma^2_{n|n-1}[\hat{\theta}_N^{(1)}]} - \frac{e^2_{n|n-1}[\hat{\theta}_N^{(2)}]}{\sigma^2_{n|n-1}[\hat{\theta}_N^{(2)}]},$$

where $e_{n|n-1}[\hat{\theta}_N^{(i)}] \equiv y_n - y_{n|n-1}[\hat{\theta}_N^{(i)}]$, and $y_{n|n-1}[\hat{\theta}_N^{(i)}]$ and $\sigma^2_{n|n-1}[\hat{\theta}_N^{(i)}]$ are defined as in section 3. Thus $\hat{\zeta}_n^{(1,2)}$, n=1,...,N can be obtained from the Kalman filter output for the two maximum likelihood models. If the autoregressive model fit to this data has order p, estimated coefficients $\hat{a}_1(N),...,\hat{a}_p(N)$, and estimated innovations variance $\hat{\sigma}^2(N)$, then an estimate of $2\pi f^{(1,2)}(0)$ is given by

$$(9.2) \qquad \hat{\sigma}^{(1,2)^2}(N) \equiv \hat{\sigma}^2(N)(1 - \hat{a}_2(N) - \cdots - \hat{a}_p(N))^{-2} .$$

There are several possible ways of estimating the coefficients and innovations variance used in (9.2). The empirical study presented in the next section makes clear that it is essential to use robust estimates. We will use the notation $\hat{\sigma}_{YW}^{(1,2)^2}$ when (9.2) is calculated from sample moment estimates discussed in section 6, and we will use $\hat{\sigma}_{GM}^{(1,2)^2}$ when the robust scale and GM–estimates of coefficients from the routine ar.gm of S–PLUS are used, see Martin (1981) for definitions and supporting theory. We define the corresponding test statistics

$$Z^{(1,2)}_{YW} \equiv 2N^{-1}\hat{L}^{(1,2)}_N / \hat{\sigma}^{(1,2)}_{YW}(N) \tag{9.3}$$

and

$$Z^{(1,2)}_{GM} \equiv 2N^{-1/2}\hat{L}^{(1,2)}_N / \hat{\sigma}^{(1,2)}_{GM}(N)\ . \tag{9.4}$$

It is to be expected that $\hat{\sigma}^{(1,2)^2}_{YW}(N)$ and $\hat{\sigma}^{(1,2)^2}_{GM}(N)$ will converge in probability to $2\pi f^{(1,2)}(0)$, see Berk (1974) and Shibata (1984), but we have not been able to prove this. Our best result to date applies only to the comparison of autoregressions with gaps for y_n and shows, based on slight generalizations of (6.6) – (6.8), that, for <u>fixed</u> p, the Yule–Walker estimates from $\hat{\zeta}^{(1,2)}_n$ converge to the same limiting values as a p–th order autoregression fit to the unobservable process $\zeta^{(1,2)}_n$. This result will not be proved here, because it is inadequate for the examples of the next section, which involve models with moving average terms. It can be seen there that results compatible with the graphical diagnostics of section 3 are usually obtained if $Z^{(1,2)}_{GM}$ is referred to a standard normal distribution, but that $Z^{(1,2)}_{YW}$ is often misleadingly small. Consistency results for GM–estimates exist but are difficult, see Boente, Fraiman and Yohai (1987). A second autoregression–based robust estimator will also be used in section 10. When using robust estimates of $f^{(1,2)}(0)$, we do not know how to verify that $f^{(1,2)}(0) \neq 0$.

<u>Remark</u> <u>9.1</u>. It is not surprising that robust methods are needed, because the two series $e_{n|n-1}[\hat{\theta}^{(j)}_N]$, j=1,2 which define $\hat{\zeta}^{(1,2)}_N$ are forecast errors from imperfect models. Also, even in the situation where the series $e_n[k^{(1)}]$ and $e_n[k^{(2)}]$ defining $\zeta^{(1,2)}_n$ are Gaussian, the distribution of the series $\zeta^{(1,2)}_n$ will be heavier–tailed: for example, the marginal

distributions will be linear combinations of independent chi–square variates with one degree of freedom.

Remark 9.2. One would like model selection procedures to have the property of transitivity: if model 1 is preferred over model 2, and model 2 over model 3, then model 1 should be preferred over model 3. For the graphical diagnostics of section 3, transitivity follows simply from the additivity of the log–likelihood ratios: for example, since

$$\hat{L}_M^{(1,3)} = \hat{L}_M^{(1,2)} + \hat{L}_M^{(2,3)} , \tag{9.5}$$

it follows that if, for positive numbers α, β and for all M satisfying $N_0 < M \leq N$, the inequalities $\hat{L}_M^{(1,2)} - \hat{L}_{M-1}^{(1,2)} \geq \alpha$ and $\hat{L}_M^{(2,3)} - \hat{L}_{M-1}^{(2,3)} \geq \beta$ hold, then $\hat{L}_M^{(1,3)} - \hat{L}_{M-1}^{(1,3)} \geq \alpha+\beta$ holds for these values of M. In other words, the ultimate slope of the graph of $\hat{L}_M^{(1,3)}$ will exceed the ultimate slopes of the graphs of $\hat{L}_M^{(1,2)}$ and $\hat{L}_M^{(2,3)}$. For our generalization of Vuong's test, it is not clear that transitivity will hold for any fixed sample size, but it can be obtained asymptotically from (9.5) and the strict inequality

$$(f^{(1,3)}(0))^{1/2} < (f^{(1,2)}(0))^{1/2} + (f^{(2,3)}(0))^{1/2} ,$$

which holds if the joint spectral density matrix of $\zeta_n^{(1,2)}$ and $\zeta_n^{(2,3)}$ is non–singular at $\lambda=0$.

10. COMPARISONS OF MODEL PAIRS FOR SOME ECONOMIC TIME SERIES

In this section, we elaborate the study of Bell and Pugh (1989). They used AIC to compare the basic structural component model (BSM) of Harvey and Todd (1983) with ARIMA models fit individually to a large set of log–transformed economic time series, from the Business, Industry and Construction Statistics Divisions of the U.S. Census Bureau. The BSM can be written

$$(10.1) \qquad y_n = S_n + T_n + I_n$$

where S_n, T_n and I_n are independent series presumed to satisfy

$$(1+B + \dots B^{11})S_n = e_{1n},\ e_{1n} \sim \text{i.i.d. } \mathcal{N}(0,\mu\sigma^2)$$

$$(1-B)^2T_n = (1-\eta B)e_{2n} \quad (\eta \geq 0),\ e_{2n} \sim \text{i.i.d. } \mathcal{N}(0,\nu\sigma^2)$$

$$I_n \sim \text{i.i.d. } \mathcal{N}(0,\sigma^2)\ .$$

In our study, if the estimated value of η exceeded 0.9, we often used a different model for the "trend" component T_n,

$$(1-B)T_n = C + e_{2n}\ ,\ \ e_{2n} \sim \text{i.i.d. } \mathcal{N}(0,\nu\sigma^2),$$

with C a constant term, in order to avoid the technical problems caused by non–invertibility which were discussed in section 5. We refer to this model as a *modified component model*.

In addition to the three components of (10.1), the models considered for most of the series have a *mean component* consisting of a sum of indicator variables for highly significant additive outliers or level shifts together with linear regression expressions modeling calendar effects, see Bell and Hillmer (1983). Table 3 indicates, for each series, which effects were included in the mean function. The theoretical discussion in the preceding sections concerned mean zero time series, so we need to say something about the additional assumptions and developments required to cover the situation of estimated mean functions. The estimation of the coefficients of the indicator variables for additive outliers and limited–duration level shifts has an asymptotically negligible effect on the likelihood function. So, for theoretical purposes, we assume that such coefficients are fixed and not reestimated as $N \longrightarrow \infty$. The calendar effect variables can be regarded as periodic with long periods, and they satisfy Grenander's conditions as discussed in Hannan (1973) as does a constant mean variable. Our method of simultaneously estimating regression and ARMA coefficients is described in Findley, Monsell, Otto, Bell and Pugh (1988). We shall assume that with properly chosen coefficients (perhaps zero) these regression variables completely describe the mean function of y_n, even though the remainder of the model might not completely describe the covariance structure of the series. With this assumption, the methods used for the proof of Theorem 4 of Hannan (1973) can be utilized to obtain a generalization of Proposition 5.1 which covers models with such mean functions. The analysis of section 8, which concerned the asymptotic models' covariance structures, carries over without change if one replaces y_n with $y_n - Ey_n$.

Bell and Pugh used Akaike's AIC as the basic comparison statistic. Hence they used the sign of (7.7) to indicate the preferred model, the first model being preferred over the second if $\Delta \mathrm{AIC}_N^{(1,2)} < 0$. In our comparisons, the ARIMA model family is designated the first family (j=1) and the component or modified component model family is always the second family (j=2).

10.1 Comparison Results

The model comparison results have been divided into two categories, determined by the extent of conformity to assumptions utilized above. Table 1 presents the results for the ten series for which both the ARIMA and component models obtained were invertible and utilized the same transformation to stationarity. This is the category of comparisons best supported by the theoretical results of the earlier sections concerning (3.2) and the $Z^{(1,2)}(N)$–statistics. For the other comparisons, given in Table 2, one of the models had to be overdifferenced to achieve $\delta^{(1)}(B) = \delta^{(2)}(B)$, a condition which seemed to be necessary in order to obtain reliable $Z^{(1,2)}(N)$–statistics with series of the length used in the study. The results now to be discussed have led us to conclude that the $Z_N^{(1,2)}$–statistics are useful complements to the graphical diagnostics if two limitations are taken into account. First, their values are sometimes distorted by nonstationarities in the $\hat{\zeta}_n^{(1,2)}$ so it is worthwhile to graph this series. Second, being summary statistics, the $Z^{(1,2)}(N)$ are inherently less informative than (3.1) and (3.2) about changes in the statistical properties of the most recent observations.

The designations of the modeled series are explained in Table 3. The values of the statistics ΔAIC, Z_{YW} and Z_{GM} and the interpretation of the graph of (3.2) for each series are given in Tables 1 and 2. This graph was interpreted as inconclusive (I) unless two requirements were satisfied: first, the general trend of the later portion of the graph must move toward an infinite value, linearly or faster. Second, the subinterval of (N/2, N] in which this later movement occurs must be longer than any earlier subinterval in which the general movement is in the opposite direction. As the discussion of the examples below shows, these requirements may be a bit too restrictive. For instance, if the trend of the graph is linearly upward over (N/2, 7N/8] but level over (7N/8, N], the graph would be interpreted as inconclusive in the study, but it would usually be reasonable to select the model favored by the upward trend, since it was never dominated by the other model. However, the leveling out at the end indicates a change in the nature of the time series being modeled, so one should investigate the adequacy of the fits of both models over (7N/8, N] before completing the selection.

The ambiguity about whether the hypothesis tests based on $Z_{GM}^{(1,2)}$ should be intepreted as one–sided tests (based on the sign of $\hat{L}_N^{(1,2)}$) as in Vuong (1989), or as two–sided tests, leads to ambiguity in the definition of the significance levels and critical regions of the test. If we choose a five–percent significance level and assume that the asymptotic distribution is appropriate for the test statistic, then $(1.645, \infty)$ and $(-\infty, -1.645)$ are the critical regions for the one–sided tests, and $(-\infty, -1.96) \cup (1.96, \infty)$ is the critical region for the two–sided test. There are two series, *icmeti* ($Z_{GM} = -1.94$) and *bdptrs* ($Z_{GM} = 1.88$) where this ambiguity affects the conclusion of the test. For *icmeti*, the graphical diagnostics (Figs. 5a, 5b) offer support for the rejection of the null hypothesis in accord with the one–sided test. For *bdptrs*, the graphs (Figs. 1a, 1b) were classified as inconclusive. To help resolve this ambiguity, a more elaborate "cleaned residuals" estimate of $f^{(1,2)}(0)$ was calculated utilizing the fitted robust AR models as prefilters, as described in section XIII of Martin (1981), and using a 10% data taper and the c(3, 5) periodogram smoothing window of S–PLUS. The values of the test statistic $Z^{(1,2)}(N)$ obtained from these estimates are presented as Z_{SP}–values in Tables 1 and 2. For *icmeti* and *bdptrs*, these values are 9.44 and 1.60 respectively, which are unambigously in accord with the interpretation of the graphical diagnostic. Thus, in Table 1, there is complete agreement between the interpretation of (3.2) and the one–sided test based on Z_{SP}. By constrast, in about a third of the comparisons in Table 2, one of the new comparison methods is decisive (usually the test based on Z_{SP}) and the other (the graph of (3.2)) is not. Some of the causes of such disagreements are discussed in the next subsection.

10.2 <u>Examples of Graphs of</u> (3.1), (3.2), (7.5), (7.6) <u>and</u> (9.1)

Figs. 1–5 present graphs associated with series in Table 1. Fig. 6–10 are associated with Table 2. In the subfigures labelled (a),(b), (c) and (d), the horizontal axis is the time = sample size axis for the stationary series obtained by applying 1–B and $1+B+ \dots + B^{11}$

as needed. The vertical axis gives the values of the log–likelihood differences in the (a) and (b) figures and the values of the pseudo–log–likelihood–ratios in (c) and (d) of Figs. 3 and 7.

A feature of some of these graphs facilitates the use of AIC. For the model comparisons associated with these figures, the parameter dimension difference, $\dim\theta^{(1)} - \dim\theta^{(2)}$, took on the values –1, 0, and 1. If

$$\hat{L}_N^{(1,2)} > \dim\theta^{(1)} - \dim\theta^{(2)} , \tag{10.2}$$

then model (1) is preferred by the minimum AIC procedure. If the opposite inequality holds, model (2) is preferred. When both models are invertible and have the same stationarizing transformation, so that $\hat{L}_N^{(1,2)}$ is known, a horizontal line is drawn in the figures at the level $\dim\theta^{(1)} - \dim\theta^{(2)}$ (provided that this ordinate value is within the range of the graph) in order to make it possible to see how persistently the graph stays above or below this level. This provides information about the stability of AIC's choice. The stability of such statistics is a matter of special concern in non–nested comparisons, because if $f^{(1,2)}(0) \neq 0$, the variance of $\Delta\mathrm{AIC}_N^{(1,2)}$, like that of $\hat{L}_N^{(1,2)}$, has order N, asymptotically, in the situation of weakly equivalent models. This is precisely the situation in which the means of these quantities usually have a finite limit, and, consequently, the graphical diagnostics are expected to be inconclusive and the Z–statistic small. It is often the natural context in which to use the minimum AIC criterion, see Findley and Wei (1989).

The subfigures (a) and (b) of Figs. 1–8 present graphs of (3.1) and (3.2) for situations in which the ARIMA model and the component model have the same stationarizing polynomial, $\delta^{(1)}(B) = \delta^{(2)}(B)$, at the expense, in Figs. 6–8, of using the basic structural component model even when it was non–invertible. In all cases, the graphs of (3.1) and (3.2) are quite similar.

Figs. 1–2 are graphs which were classified as inconclusive. The movements of the graphs relative to the dashed line at level $\dim\theta^{(1)} - \dim\theta^{(2)}$ suggest that AIC's preference for the ARIMA model is rather stable in the case of *bdptrs*, but tentative in the case of *bfrnrs*, because of the downward movement at the end in Fig. 2. For the other series in Table 1 for which (3.2) was inconclusive ,*bgasrs*, this graph (not given) stayed above the level $\dim\theta^{(1)} - \dim\theta^{(2)}$ in a way that suggested that AIC's preference for the ARIMA model was stable.

The graphs in Figs. 3–5 favor the ARIMA model as does the Z_{SP}–statistic. The graphs of Figs. 6–8 are the only ones in the study which could be interpreted as favoring the component model, although in none of these graphs does the downward trend seem to have a constant slope. At least one of the models in each of these last three comparisons is non–invertible, so there is uncertainty about the applicability of the $Z^{(1,2)}(N)$–statistics.

The subfigures (c) and (d) of Fig. 3 and Fig. 7 present the pseudo–log–likelihood–ratio graphs (7.5) and (7.6) for situations in which the stationarizing polynomials $\delta^{(1)}(B)$ and $\delta^{(2)}(B)$ differ by a factor 1–B. In the case of Figs. 3(c) and 3(d), a non–invertible component model was used in place of the preferred modified component model to obtain a comparison in which $\delta^{(2)}(B) = (1-B)\delta^{(1)}(B)$. In the case of Figs. 7(c) and 7(d), the use of the modified component model to avoid noninvertibility resulted in $\delta^{(1)}(B) = (1-B)\delta^{(2)}(B)$. In both cases, the shapes of the graphs closely resemble those of the subfigures (a) and (b).

Figs. 3(e) and 7(e) present plots of the series $\hat{\zeta}_n^{(1,2)}$ for two series for which the non–robust statistic Z_{YW} is strongly biased toward 0 by "outliers". Tables 1 and Table 2 contain further examples of too–small values of this statistic.

10.2.1 *Series with an early change of regime.*

Fig. 7(e) reveals that the first six years of the series $\hat{\zeta}_n^{(1,2)}$ from *ifatvs* have a different character than the rest. Since the graph of the original series (Fig. 9(i)) also suggests a change of regime after the first six years, these early observations were deleted, the models were reestimated from the remaining data, and the diagnostics were recalculated. The results are presented in parenthesis in Table 2. The new graph of (3.2) (Fig. 9(ii)) is erratic and is interpreted as inconclusive (in agreement with the new, insignificant value of $Z_{SP} = -.47$): the downward movement is restricted to the middle of the graph and is stepped rather than linear.

In a similar way, the graphs of (9.1) (not given) for *ihapvs* and *bhwws* called our attention to the fact that these series, too, appear to have undergone an early change of regime, after the first three, respectively, six years. The results associated with the shortened series are likewise given parenthetically in Table 2. The graph of *ihapvs* and its new (3.2) are given in Fig. 10. The values of (3.2) tend strongly upward for the first two thirds of the graph, in line with the significant new $Z_{SP} = 4.71$, but the final third has a level trend, so this diagnostic was inconclusive. Had it become downsloping in the final years, rather than level, we would have first considered the component model if short term forecasting of *ihapvs* was the goal. Since the graph remained level, however, it is reasonable to prefer the ARIMA model for this purpose. For *bhdwws*, there is a rapid downward movement in the first third of the graph of (3.2) (not given) from the shortened series (in accord with the value $Z_{SP} = -1.72$) followed by a large oscillation to a lower value, then, in the final third, a generally upward movement becoming rather level at the end near the value −1.8. Thus the diagnostic graph is not conclusive, which is not surprising, because the graph of the original series (not given) is quite erratic in the final years.

10.2.2 *Disagreements between graphs and tests.*

For approximately one–third of the comparisons of Table 2 a situation similar to that just described for shortened *ihapvs* occurred, in which Z_{SP} was significant ($|Z_{SP}| > 1.65$) but the graphical diagnostic was inconclusive, because of a leveling out or a change in the sign of the slope in the last years of the graph. There is one series, *blqrrs*, where the graph of (3.2) (Fig. 6(b)) was interpreted as favoring the component model, but the value $Z_{SP} = 1.14$ is insignificant. There is a change of direction early in the graph and its subsequent decreasing movement is not very linear and never carries the graph below the positive value 2.0. Thus the insignificance of Z_{SP} reinforces the lingering ambiguity in the graph.

In summary, we find the diagnostic graphs (3.1) and (3.2) more informative than the $Z^{(1,2)}(N)$–statistics, but we do find these statistics to be useful adjuncts to the graphs, either to confirm, or to provoke closer scrutiny of, their interpretation. When these diagnostics and tests are conclusive, the ARIMA model is usually favored over the structural component model.

11. OTHER TESTS AND GENERALIZATIONS

Vuong (1989) has already discussed the fundamental differences between the hypothesis testing procedure for non–nested models of sections 8 and 9 and the tests of Cox (1961, 1962), which were generalized to ARMA models by Walker (1967) and have stimulated the development of a vast theoretical literature, see White (1990). To Vuong's discussion we wish to add one remark concerning the comparison of ARMA models: not only are the null and alternative hypothesis of the Cox tests different from the simple and intuitive $\sigma^2[k^{(1)}] = \sigma^2[k^{(2)}]$ versus $\sigma^2[k^{(1)}] \neq \sigma^2[k^{(2)}]$ hypotheses of section 9, also the two quantities needed for the Cox test statistic, the null–hypothesis mean of $\hat{L}_N^{(1,2)}$ and a consistent estimator of the variance of the limiting distribution under the null hypothesis, are so complex that Cox tests seem not to have been implemented for models with moving average terms. By contrast, producing the data (3.2) for a graph or the series $\hat{\zeta}_n^{(1,2)}$ of (9.1) for the test statistic $Z^{(1,2)}(N)$ is unproblematic. The rather interesting problem posed by the series $\hat{\zeta}_n^{(1,2)}$ is that it seems inherently to require robust methods of analysis. Fortunately, there is a widely available statistical software package that provides such procedures.

Finally, we mention that there are some generalizations of the diagnostics of this paper that we plan to implement in software and apply to the series and models discussed in section 10. These generalizations concern the selection of time series models for m–step–ahead prediction, where m>1. (Recall from the discussion preceding Proposition 8.3 that the time series model selection procedures of this paper were seen to focus upon the selection of one–step–ahead forecasting models because of (2.6).) If neither of the two competing time series model classes is capable of describing the correct innovations transfer function of the observed series, then it can happen that the class with the better forecast function for m=1 will have the worse forecast function for some m>1: for the models of subsection 1.1, if $|\rho_1|>|\rho_2|$, it follows from $\rho_2 \neq \rho_1^2$ that the m–step–ahead forecast function $\rho_1^m y_N$ of (1.1) has mean square error which is smaller than that of the forecast function of model (1.2) when m=1, but larger when m=2. The proof of Proposition 8.1 easily generalizes to show that a correct model has strictly smaller mean square forecast error for all values of m than an incorrect model. A generalization of Proposition 8.3 for multistep forecasting is given in Findley (1990b).

ACKNOWLEDGEMENTS

The substantial amount of computing and special purpose programming required for the analysis of section 10 was done by Larry Bobbitt with initial help from Mark Otto. The author is very grateful for their careful, insightful computing support. He wishes to also thank William Bell for stimulating conversations on several aspects of this paper and Larry Bobbitt for helpful comments on an earlier draft.

The maximum likelihood calculations required for estimating the models, graphical diagnostics and test statistics presented in section 10 were accomplished with the program

REGCMPNT designed by William Bell and Steven Hillmer, and developed by them and by Mark Otto, Marian Pugh, Larry Bobbitt and James Bozik. The ARIMA model selection for the full series in Tables 1 and 2 was done by Peter Burman and Mark Otto for a study of the effects of outliers on forecasting, Burman and Otto (1988).

DISCLAIMER

This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.

REFERENCES

Baxter, G. (1962). An asymptotic result for the finite predictor. Math. Scand. 10, 137–144.

Bell, W. R. (1984). Signal extraction for non–stationary time series. Ann. Statist. 12, 646–664.

Bell, W. R. and Pugh, M. G. (1989). Alternative approaches to the analysis of time series components. Proceedings of the Statistics Canada Symposium on Analysis of Data in Time. Ottawa: Statistics Canada (to appear).

Bell, W. R. and Hillmer, S. C. (1983). Modeling time series with calendar variation. J. Amer. Statist. Ass. 78, 526–534.

Bell, W. R. and Hillmer, S. C. (1990). Initializing the Kalman Filter for Non–Stationary Time Series Models. J. Time Ser. Anal. 11 (to appear).

Berk, K. N. (1974). Consistent autoregressive spectral estimates. Ann. Statist. 2, 489–502.

Boente, G., Fraiman R. and Yohai, V. J. (1987). Qualitative robustness for stochastic processes. Ann. Statist. 15, 1293–1312.

Brillinger, D. (1975). Time Series Analysis. New York: Holt, Rinehart and Winston.

Burman, J. P. and Otto, M. C. (1988). Outliers in Time Series. Research Report Number 88/14, Statistical Research Division, Bureau of the Census.

Chan, N. H. and Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* 16, 367–401.

Cox, D. R. (1961). Tests of separate families of hypotheses. *Proc. 4th Berkeley Symp.* 1, 105–123.

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *J. R. Statist. Soc.* B24, 406–424.

Deistler, M. and Pötscher, B. M. (1984). The behavior of the likelihood function for ARMA models. *Adv. Appl. Prob.* 16, 843–866.

Dunsmuir, W. and Hannan, E. J. (1976). Vector linear time series models. *Adv. Appl. Prob.* 8, 339–364.

Findley, D. F. (1985a). Backshift–operator polynomial transformations to stationarity for nonstationary time series and their aggregates. *Commun. Statist.* A13(1), 49–61.

Findley, D. F. (1985b). On the unbiasedness property of AIC for exact or approximating linear time series models. *J. Time Ser. Anal.* 6, 229–252.

Findley, D. F., Monsell, B. C., Otto, M. C., Bell, W. R. and Pugh, M. G. (1988). Toward X–12–ARIMA. *Proceedings of the Fourth Annual Research Conference*, 591–622. Washington, D.C.: Bureau of the Census.

Findley, D. F. (1990a). Conditional densities as densities, and likelihoods defined via singular transformations. (manuscript in preparation)

Findley, D. F. (1990b). Convergence rate of the finite multi–step–ahead predictors. *Note di Matematica* (to appear).

Findley, D. F. and Monsell, B. C. (1989). REG–ARIMA model–based preprocessing for seasonal adjustment. *Proceedings of the Statistics Canada Syposium on Analysis of Data in Time*. Ottawa: Statistics Canada (to appear).

Findley, D. F. and Wei, C. Z. (1989). Beyond chi–square: likelihood ratio procedures for comparing non–nested, possibly incorrect regressors. Statistical Research Division Research Report No. RR–89/08. Washington, D.C.: U.S. Bureau of the Census.

Hannan, E. J. (1973). The asymptotic theory of linear time series models. *J. Appl. Probab.* 10, 130–145.

Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. New York: Wiley.

Harvey, A. C. and Todd, P.H.J. (1983). Forecasting economic time series with structural and Box–Jenkins models: a case study (with discussion). *J. Bus. Econ. Stat.* 1, 299–315.

Jones, R. H. (1980). Maximum likelihood estimation of ARMA models for time series with missing observations. *Technometrics* 22, 389–395.

Kabaila, P. (1983). Parameter values of ARMA models minimizing the one–step–ahead prediction error when the true system is not in the model set. J. Appl. Prob. 20, 405–408.

Kenny, P. D. and Durbin, J. (1982). Local trend estimation and seasonal adjustment of economic and social time series (with discussion). J. R. Statist. Soc. A 145, 1–41.

Kitagawa, G. (1987). Non–gaussian state–space modeling of non–stationary time series: Rejoinder, J. Amer. Statist. Ass. 82, 160–163.

Loeve, M. (1977). Probability Theory, 4th edn. Vol. I. New York: Springer–Verlag.

Martin, R. D. (1981). Robust methods for time series. In Applied Time Series Analysis II (ed. D. F. Findley), pp. 683–760. New York: Academic Press.

Pötscher, B. M. (1987). Convergence results for maximum likelihood type estimators in multivariable ARMA models. J. Mult. Anal. 21, 29–52.

Pötscher, B. M. (1990). Noninvertibility and quasi–maximum likelihood estimation of misspecified ARMA models. Technical Report, Department of Economics, University of Maryland (College Park).

Pötscher, B. M. and Prucha, I. R. (1989). A uniform law of large numbers for dependent and heterogenous data processes. Econometrica 57, 675–683.

Shibata, R. (1981). An optimal autoregressive spectral estimate. Ann. Statist. 9, 300–306.

S–PLUS. Statistical Sciences, Inc. Seattle.

Tiao, G. C. and Tsay, R. S. (1983). Consistency properties of least squares estimates of autoregressive parameters in ARMA models. Ann. Statist. 11, 856–871.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non–nested hypotheses. Econometrica 57, 307–333.

Walker, A. M. (1967). Some tests of separate families of hypotheses in time series analysis. Biometrika 54, 39–68.

White, H. (1990). Estimation, Inference and Specification Analysis. New York: Cambridge University Press.

APPENDIX A: COMPACTNESS OF $K_{\alpha,\beta}^{p,q}$.

We need to show that every sequence $k_n(z)(=\Sigma_{j=0}^{\infty} k_{jn}z^j)$, n=1,2,... in $K_{\alpha,\beta}^{p,q}$ has a subsequence $k_m(z)$ whose coefficients k_{jm} converge, $k_{jm} \xrightarrow{m} k_j$, in such a way that $k(z) =$

$\Sigma_{j=0}^{\infty} k_j z^j$ belongs to $K_{\alpha,\beta}^{p,q}$.

By definition, $k_n(z) = b_n(z)/a_n(z)$, where $a_n(z)$ and $b_n(z)$ are polynomials of degree not larger than p, respectively q, of the form $\Pi_j(1-\zeta_{jn}^{-1}z)$, with $|\zeta_{jn}|\leq 1$ (more precisely, $|\zeta_{nj}| \leq (1+\alpha)^{-1}$ for $a_n(z)$, and $|\zeta_{nj}| \leq (1+\beta)^{-1}$ for b(z)). Therefore, the absolute values of the coefficients of a_{jn} of $a_n(z)$ and b_{jn} of $b_n(z)$ are uniformly bounded above, by the binomial coefficients ${}_pC_{[p/2]}$ and ${}_qC_{[q/2]}$, respectively. Hence we can find convergent subsequences of the coefficients, and there is a subsequence $k_m(z)$ for which the coefficients of the corresponding polynomials $a_m(z)$ and $b_m(z)$ converge to the coefficients of polynomials a(z) and b(z) of degrees at most p and q, respectively, whose roots also belong to $\{|z|\geq 1+\alpha\}$ and $\{|z|\geq 1+\beta\}$, respectively, because they are limits of roots in these regions. From $a_m(z)k_m(z) = b_m(z)$, we obtain recursion formulas for the coefficients k_{jm} of $k_m(z)$: $k_{0m} = 1$, and

$$k_{jm} = -\sum_{i=1}^{\min(j,p)} a_{im}k_{j-i,m} + b_{jm},$$

setting $b_{jm} \equiv 0$ if $j > \deg b_m(z)$. It follows by induction that $k_{jm} \longrightarrow k_j$, with $k(z) \equiv \Sigma_{j=0}^{\infty} k_j z^j$ satisfying $a(z)k(z) = b(z)$. Hence $k(z) \in K_{\alpha,\beta}^{p,q}$. This completes the proof.

APPENDIX B: PROOF OF COROLLARY 8.1

The proof is based on a special case of Toeplitz's Lemma (Loeve 1977, p. 250): <u>If the sequence</u> Δ_n <u>satisfies</u> $\lim_{n\to\infty} n^{1/2}\Delta_n = 0$, <u>then</u> $\lim_{N\to\infty} N^{-1/2}\Sigma_{n=1}^{N}\Delta_n = 0$.

It is clear from (8.7) that the quantity whose mean absolute convergence is at issue in (8.8) can be written as one–half the sum of

$$N^{-1/2}\sum_{n=1}^{N}\log\frac{\sigma^2_{n|n\text{-}1}[k]}{\sigma^2[k]}$$

and

$$N^{-1/2}\sum_{n=1}^{N}\left\{\frac{e^2_{n|n-1}[k]}{\sigma^2_{n|n-1}[k]}-\frac{e^2_n[k]}{\sigma^2[k]}\right\}.$$

Thus (8.1) will follow from Toeplitz's Lemma if we verify (B.1) and (B.2),

(B.1) $$\lim_{n\longrightarrow\infty} n^{1/2}\log\frac{\sigma^2_{n|n\text{-}1}[k]}{\sigma^2[k]}=0 .$$

(B.2) $$\lim_{n\longrightarrow\infty} n^{1/2}E\left|\frac{e^2_{n|n-1}[k]}{\sigma^2_{n|n-1}[k]}-\frac{e^2_n[k]}{\sigma^2[k]}\right|=0 .$$

Concerning (B.1), it follows from (8.5) with $\alpha = 1/2$ that $n(\sigma^2_{n|n-1}[k]-\sigma^2[k])\longrightarrow 0$. Since $|\log(\sigma^2_{n|n-1}[k]/\sigma^2[k])| \leq 2|\sigma^2[k]-\sigma^2_{n|n-1}[k]|/\sigma^2[k]$ when $\sigma^2_{n|n-1}[k]/\sigma^2[k]\geq 1/2$, we conclude that a stronger result holds under the assumptions of the Corollary, namely $n(\log\sigma^2_{n|n-1}[k]/\sigma^2[k])\longrightarrow 0$.

For (B.2), let us use the abbreviations $\bar{e}_{n|n-1}\equiv e_{n|n-1}[k]/\sigma_{n|n-1}[k]$ and $\bar{e}_n = e_n[k]/\sigma[k]$. Noting that $\bar{e}^2_{n|n-1}-\bar{e}^2_n = (\bar{e}_{n|n-1}-\bar{e}_n)(\bar{e}_{n|n-1}+\bar{e}_n)$, it follows from the Cauchy–Schwarz inequality and the triangle inequality for $\|\cdot\|_2\equiv(E\{\cdot\}^2)^{1/2}$ that

(B.3) $$E|\bar{e}^2_{n|n-1}-\bar{e}^2_n| \leq (E\{\bar{e}_{n|n-1}-\bar{e}_n\}^2)^{1/2}\cdot(\|\bar{e}_{n|n-1}\|_2+\|\bar{e}_n\|_2) .$$

holds. Since $\|\bar{e}_n\|_2 = 1$ and since, by (B.5) below, $\|\bar{e}_{n|n-1}\|_2 \longrightarrow 1$, (B.2) follows from (B.3) and (8.6). Concerning $\|\bar{e}_{n|n-1}\|_2$, note that since $\|e_n[k]\|_2 = \sigma[k]$, we have

$$\text{(B.4)} \qquad \big|\,\|e_{n|n-1}[k]\|_2 - \sigma[k]\big| \le \|e_{n|n-1}[k] - e_n[k]\|_2 \,.$$

From (8.5) and (8.6) with $\alpha = 1/2$, and (B.4) we obtain the last result needed,

$$\text{(B.5)} \qquad \lim_{n \longrightarrow \infty} n^{1/2}\big|\,\|\bar{e}_{n|n-1}\|_2 - 1\big| = 0 \,.$$

Table 1. ARIMA vs. Component Model (Both Invertible)

	ΔAIC[1]	Z_{YW}[2]	Z_{GM}[2]	Z_{SP}[2]	Graph[3] of (3.2)
bgasrs	-2.7	0.03	0.08	0.16	I
icmeti	-7.2	1.09	1.94	9.44	A
ifmeti	-29.0	2.14	9.92	5.49	A
itvrti	-20.6	1.59	2.62	2.28	A
bdptrs[4]	-5.7	.54	1.88	1.60	I
bfrnrs[4]	-.3	.27	1.07	1.27	I
bmncrs[4]	-22.1	2.80	6.82	5.73	A
cnctbp[4]	-21.2	1.03	5.94	2.26	A
cncths[4]	-27.7	1.19	6.84	4.61	A
cneths[4]	-21.6	1.40	2.91	2.84	A

[1] Negative values favor the ARIMA model

[2] If Z > 1.65 (Z <- 1.65), the ARIMA model (the component model) is favored.

[3] I = inconclusive; A = ARIMA model favored; C = component model favored.

[4] For these series, the modified component model was used.

Graph Summary: 7 A's, 0 C's, 3 I's

Table 2. ARIMA vs. Component Model (One Non-Invertible)

	ΔAIC[1]	Z_{YW}[2]	Z_{GM}[2]	Z_{SP}[2]	Graph of (3.2)[3]
bapprs	-7.3	.18	0.71	0.61	I
bautrs	-14.5	.56	4.89	4.95	I
belgws	-3.7	.12	1.10	0.92	I
bfrnws	-18.3	.43	5.47	2.88	A
bgrcrs	.7	0.08	-.11	-.08	I
bgrcws	-6.2	1.42	3.57	4.61	A
bhdwws	1.9(.8)[4]	-.32(-.62)[4]	-2.65(-1.3)[4]	-2.15(-1.72)[4]	I(I)[4]
blqrrs	-4.3	.83	1.38	1.14	C
bshors	-.3	-.49	-0.83	-0.51	I
bvarrs	-1.0	0.05	0.32	0.21	I
bwaprs	-13.9	1.08	2.74	1.72	I
c1ftbp	-26.0	1.19	6.95	5.62	A
c24tbp	-5.1	1.18	2.09	1.31	I
c5ptbp	-9.3	.62	3.36	3.22	I
caopvp	-71.2	4.33	10.93	11.77	A
cnetbp	-8.9	.46	3.23	2.56	I
cwsths	-2.7	1.09	1.65	1.97	I
iapevs	1.6	-.92	-2.03	-0.97	I
ibevti	-49.6	2.79	10.03	6.07	A
ibevvs	-20.3	1.41	6.04	9.37	A
icmevs	-9.7	.96	3.42	2.74	A
ifatvs	2.7(-.02)[4]	-.23(-.18)[4]	-10.14(-.62)[4]	-4.39(-.47)[4]	I(I)[4]
ifrtvs	-23.2	1.17	2.50	2.11	A
iglcvs	-29.6	1.84	7.24	11.42	I
ihapti	-36.7	2.65	8.94	18.46	I
ihapvs	.7(-16.2)[4]	-1.00(1.88)[4]	-2.24(3.25)[4]	-1.93(4.71)[4]	C(I)[4]
inewuo	-47.7	3.23	7.77	6.72	I
irrevs	.1	-.19	-1.25	-1.68	I
itobvs	-15.8	1.70	6.77	5.47	A
itvrvs	-17.0	1.53	5.36	8.50	A

(1) Negative values favor the ARIMA model, if it can be shown that AIC is applicable to comparisons involving non-invertible models.
(2) If Z > 1.65 (Z < - 1.65), the ARIMA model (the component model) is favored if it can be shown that the same limiting distribution applies when one or both of the models being compared is non-invertible.
(3) I = inconclusive; A = ARIMA model favored; C = component model favored.
(4) The values in parentheses were obtained from shortened series as explained in section 10.

Graph Summary: 10 A's, 2(1) C's, 18(19) I's

Table 3: Series and ARIMA Models (with Regression Variables) Used in the Study

Series	Years	Selected ARIMA Model	Outliers	Series Description
bapprs	67-83	(010)(011)12+TD	6/72	Retail sales of household appliance stores
bautrs	67-82	(110)(011)12+TD	3/75	Total retail sales of automotive dealers (total)
belgws	67-83	(011)(011)12+TD	-	Wholesale sales of electrical goods
bfrnws	67-82	(011)(011)12+TD	12/77,12/78, 2/79,1/80	Wholesale sales of furniture and home furnishings
bgasrs	67-82	(011)(011)12+TD	-	Retail sales of gasoline stations
bgrcrs	67-82	(013)(011)12+TD+E	12/70,1/72, 4/75,12/77	Retail sales of grocery stores
bgrcws	67-83	(013)(011)12+TD	-	Wholesale sales of groceries and related products
bhdwws	67-83	(011)(011)12+TD	-	Wholesale sales of hardware, plumbing, heating equipment, and supplies
bdhwws	73-83	(011)(011)12+TD	Level shift: 4/80	
blqrrs	67-83	(012)(011)12+TD	-	Retail sales of liquor stores
bshors	67-83	(011)(011)12+TD+E	12/69,1/70	Retail sales of shoe stores
bvarrs	67-83	(013)(011)12+TD+E	4/67,4/76	Retail sales of variety stores
bwaprs	67-83	(012)(011)12+TD+E	8/73	Retail sales of women's clothing stores
clftbp	64-83	(011)(011)12TD	2/66,1/70, 12/70,12/78, 3/79	Total 1 family dwelling building per permits
c24tbp	64-83	(011)(011)12+TD	3/75,8/75, 6/78,4/80	Total 2 to 4 unit building permits
c5ptbp	64-83	(013)(011)12	12/74	Total 5+ unit building permits
caopvp	64-83	(310)(011)12	4/69,8/70, 7/77	Value put in place, all other private residences
cnetbp	64-83	(011)(011)12+TD	1/67,12/74, 3/77, 1/82	Total Northeast building permits
cwsths	64-83	(013)(011)12	-	Total West housing starts

iapevs	68-83	(011)(011)12	12/71,3/73, 1/78	Value shipped of aircraft parts and equipment
ibevti	62-81	(012)(011)12	-	Total inventories of beverages
ibevvs	62-81	(014)(011)12+TD	-	Value shipped of beverages
icmeti	68-84	(310)(011)12	1/69,9/69, 11/82	Total inventories of communications equipment
icmevs	68-83	(210)(011)12	8/74,12/75, 12/76,8/83, 12/83	Value shipped of communications equipment
ifatvs	62-81	(011)(011)12	10/73,8/74	Value shipped of fats and oils
ifatvs	68-81	(011)(011)12	10/73,8/74	
ifmeti	62-81	(210)(011)12	10/71,10/75, 11/75,10/76, 11/76,2/77, 11/77,2/78, 12/78,11/81	Total inventories of farm machinery and equipment
ifrtvs	62-81	(011)(011)12	4/78,5/78, 1/79,4/79,5/79	Value of fertilizer shipped
iglcvs	62-81	(012)(011)12	4/65,2/68, 3/68,12/70, 12/74,7/75, 3/76,3/77,8/77	Value of glass containers shipped
ihapti	62-81	(012)(011)12	8/66,1/72, 4/80	Total inventories of household appliances
ihapvs	62-81	(011)(011)12	-	Value of household appliances shipped
ihapvs	65-81	(011)(01 1&6)12	1/80,12/66,8/70 level shifts: 6/77,4/80	
inewuo	64-83	(011)(011)12	9/68,3/82	Unfilled newspaper, periodical, and magazine orders
irrevs	62-81	(011)(011)12	-	Value of railroad equipment shipped
itobvs	64-81	(013)(011)12	11/75,10/77, 6/79,10/79	Value of tobacco shipped
itvrti	64-83	(011)(011)12	1/69,1/76	Total television and radio inventories
itvrvs	62-81	(012)(011)12	4/67	Value of televisions and radios shipped
bdptrs	67-83	(101)(011)12+TD+E	-	Retail sales of department stores

FIGURE CAPTIONS

Figure 1. (*bdptrs*). In the plots of (3.1) and (3.2) ((a) and (b)), there is no persistent suggestion of linear movement toward ± ∞ and therefore no suggestion that the asymptotic fit of one model is better. However, the way in which the graphs mainly stay above the dashed line indicating the value of $\dim\theta^{(1)} - \dim\theta^{(2)}$ shows that AIC's preference for the ARIMA model is rather stable, so this would be our preferred model.

Figure 2. (*bfrnrs*). The plots of (3.1) and (3.2) ((a) and (b)) do not suggest linear movement toward ± ∞. Also, their final, downward movement to the dashed line at level $\dim\theta^{(1)} - \dim\theta^{(2)}$ suggests that AIC's preference for the ARIMA model is fragile, so these procedures, by themselves, do not lead us to a preferred model for this series.

Figure 3. (*cnetbp*). A linearly increasing trend, favoring the ARIMA model, is suggested both by the log–likelihood–ratio plots (3.1) and (3.2) ((a) and (b)), for the comparison with the modified component model, and also by the pseudo log–likelihood–ratio plots (7.5) and (7.6) ((c) and (d)), for the comparison with the standard component model. The similarity between (7.5) and (7.6) suggests that the asymptotic behavior is insensitive to identical overdifferencing in both models. The obvious "outliers" in the graph (e) of (9.1) provide an explanation for the substantial differences between the robust statistics Z_{GN} and Z_{SP} and the non–robust statistic Z_{YW} in Table 1.

Figure 4. (*cneths*). An example in which the linear trend movement toward + ∞ in the graphs (3.1) and (3.2) ((a) and (b)), favoring the ARIMA model, is especially clear.

Figure 5. (*icmeti*). An example in which the graphs of (3.1) and (3.2) were interpreted as tending linearly toward + ∞, even though there are substantial fluctuations. There is no ambiguity about AIC's preference for the ARIMA model, since the plots stay above the value 1 of $\dim\theta^{(1)} - \dim\theta^{(2)}$, which is below the range of the graph.

Figure 6. (*blqrrs*). Overall, the plots of (3.1) and (3.2) ((a) and (b)) appears to be moving toward − ∞, although the stepped movement at the end doesn't suggest the linear trend expected. The graph was interpreted as favoring the non–invertible component model.

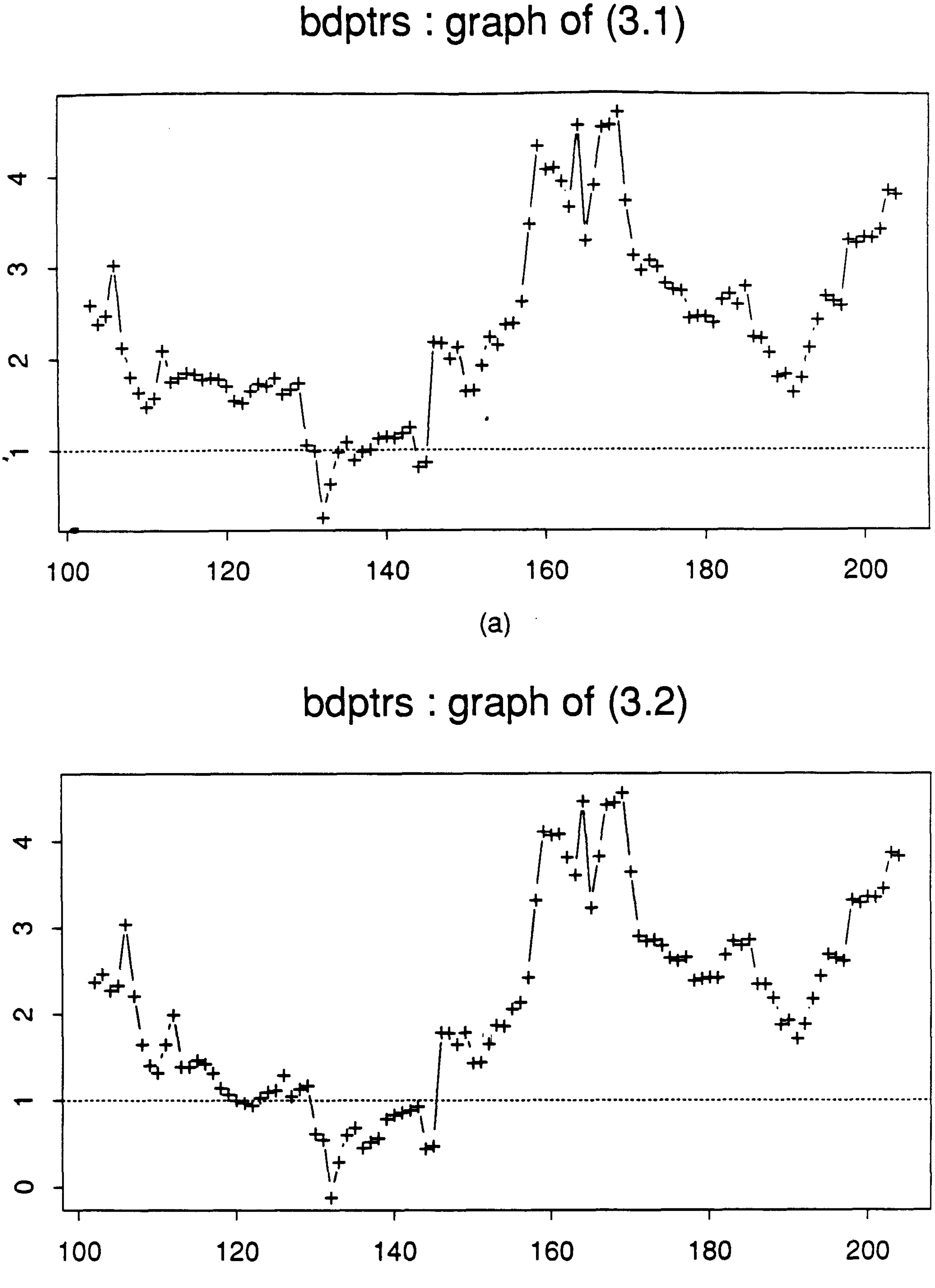
Figure 7. (*ifatvs*). Movement toward − ∞ is a possible interpretation both of the log–likelihood–ratio plots (3.1) and (3.2), for the comparison with an overdifferenced non–invertible component model, and of the similarly shaped pseudo–log–likelihood–ratio plots (7.5) and (7.6) ((c) and (d)), for the comparison with the modified component model. This preference for the component model is also supported by the values of the robust statistics, $Z_{GM} = -10.14$ and $Z_{SP} = -4.39$ in Table 2. However, the graph (e) of (9.1) not only shows the need for robust estimation, it also reveals non–stationarity of the $\zeta_n^{(1,2)}$, whose earliest six years are less variable than the later years. See Fig. 9 for graphs associated with models fit to the shortened series with these years deleted..

Fig. 9. Graph of *ifatvs* and the graph of (3.2) for models fit to the shortened series beginning in 1968. The middle part of the diagnostic graph has some indications of downward movement, but levels out at the end.

bfrnrs	67-82	(101)(011)12+TD	-	Retail sales of furniture stores
bmncrs	67-83	(101)(011)12+TD+E	-	Retail sales of men’s and boys’ clothing stores
cnctbp	64-83	(100)(011)12+TD	12/64,1/65, 1/79	Total North Central building permits
cncths	64-83	(101)(011)12	2/64,1/73, 1/75,1/77, 1/79,2/79	Total North Central housing starts
cneths	64-83	(101)(011)12	1/65,2/75, 2/78,2/80	Total Northeast housing starts

Fig. 10. Graph of *ihapvs* and the diagnostic graph (3.2) from models fit to the shortened series beginning in 1965. The strong initial upward movement levels out later, so the graphical diagnostic is not conclusive with the clarity suggested by the value of Z_{SP} (= 4.71) given in Table 2.

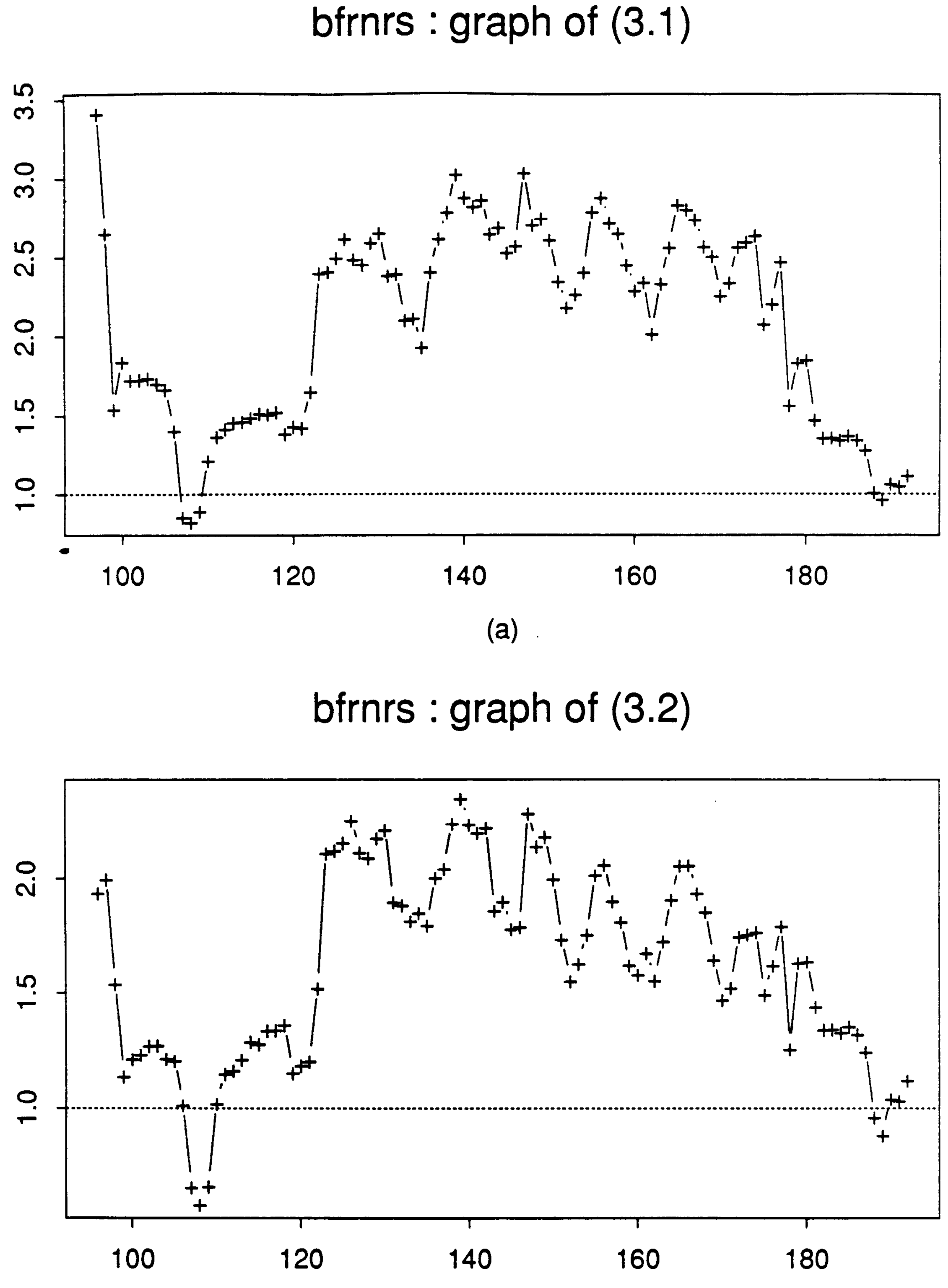
Figure 1.


bdptrs : graph of (3.1)
1
2
3
4
100
120
140
160
180
200
bdptrs : graph of (3.2)
0
1
2
3
4
100
120
140
160
180
200

(a)

(b)

Figure 2.



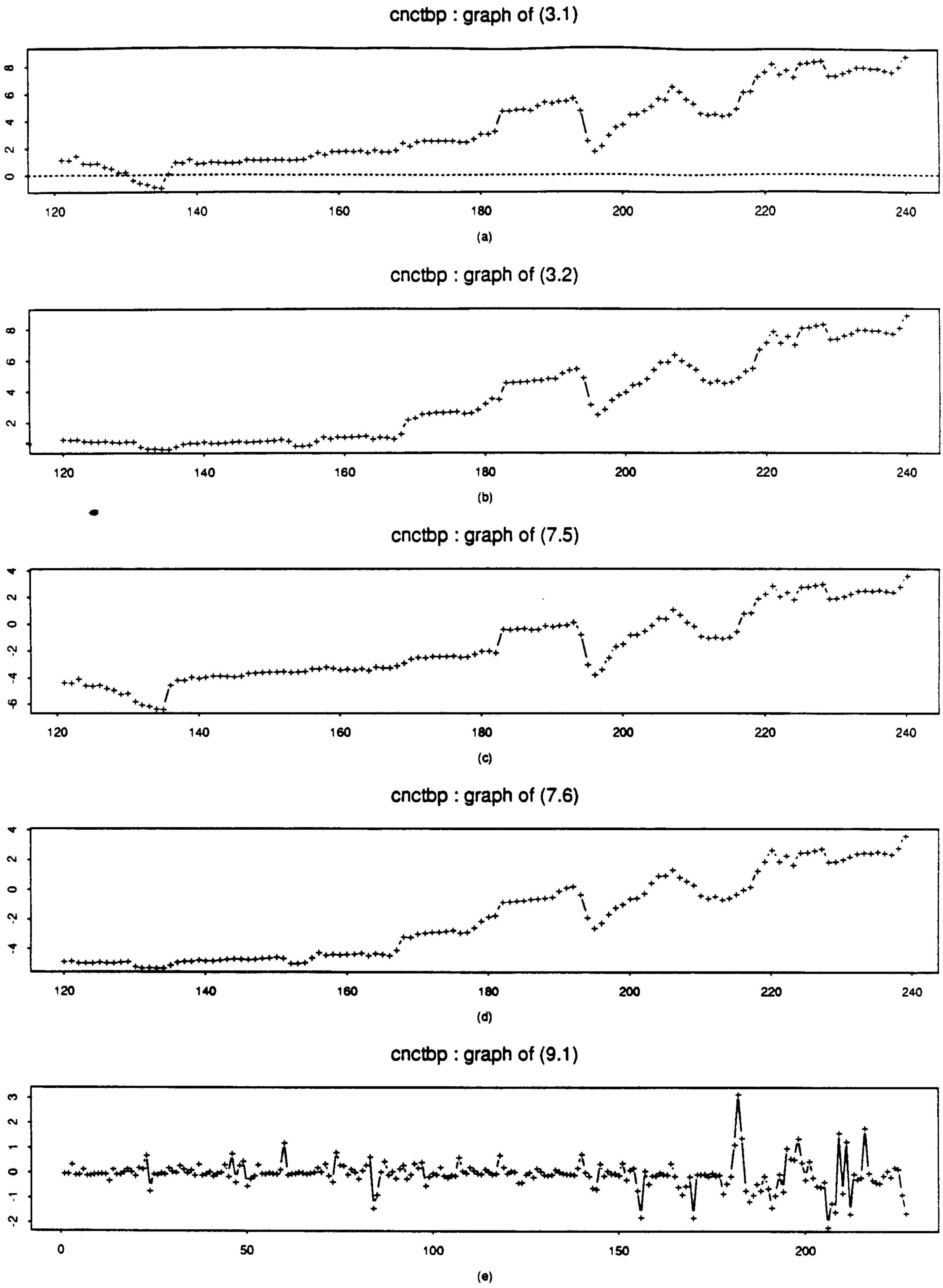
bfrnrs : graph of (3.1)

(a)

bfrnrs : graph of (3.2)

(b)

Figure 3.



(a)

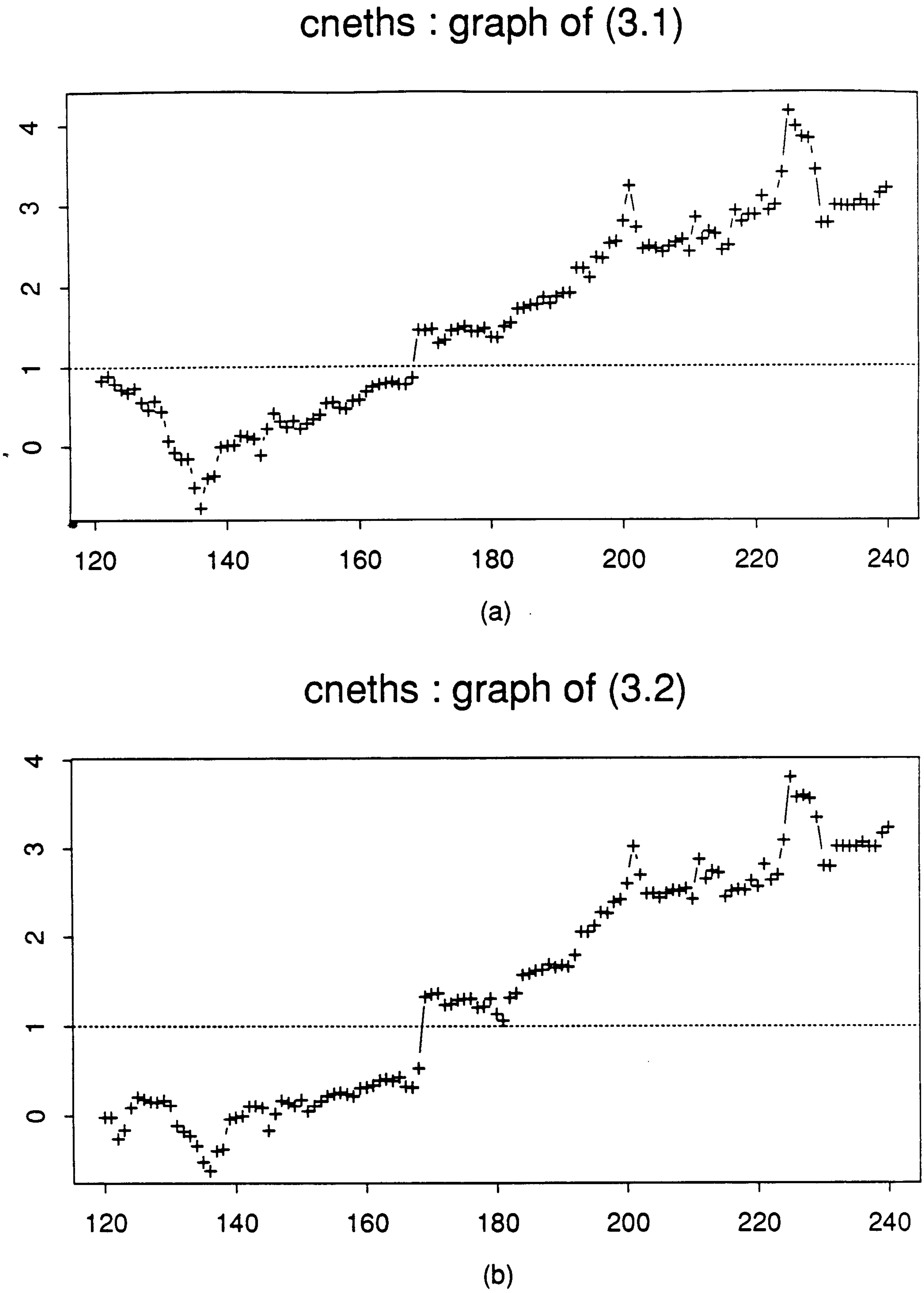
(b)

(c)

(d)

(e)

Figure 4.



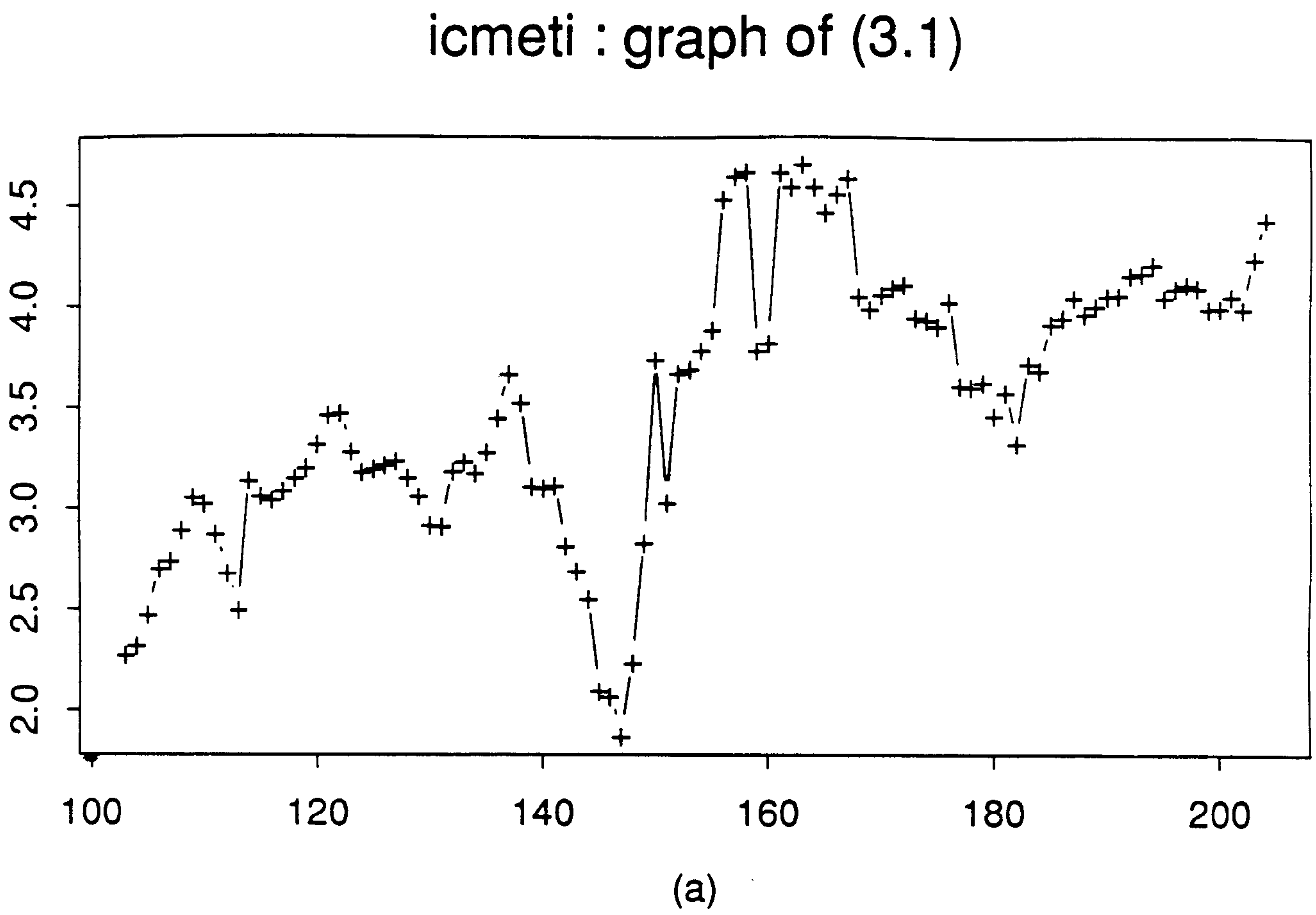
cneths : graph of (3.1)

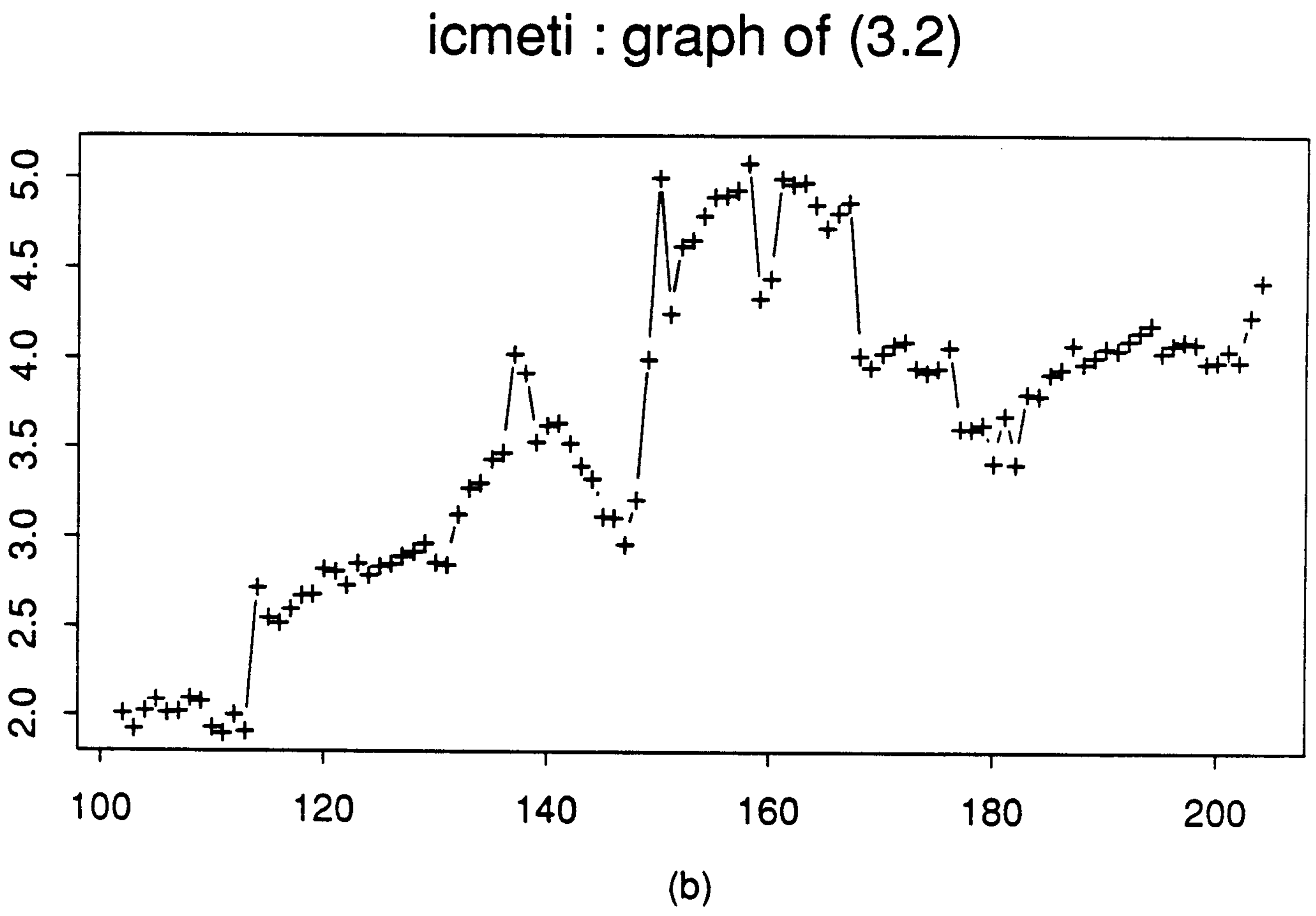
(a)

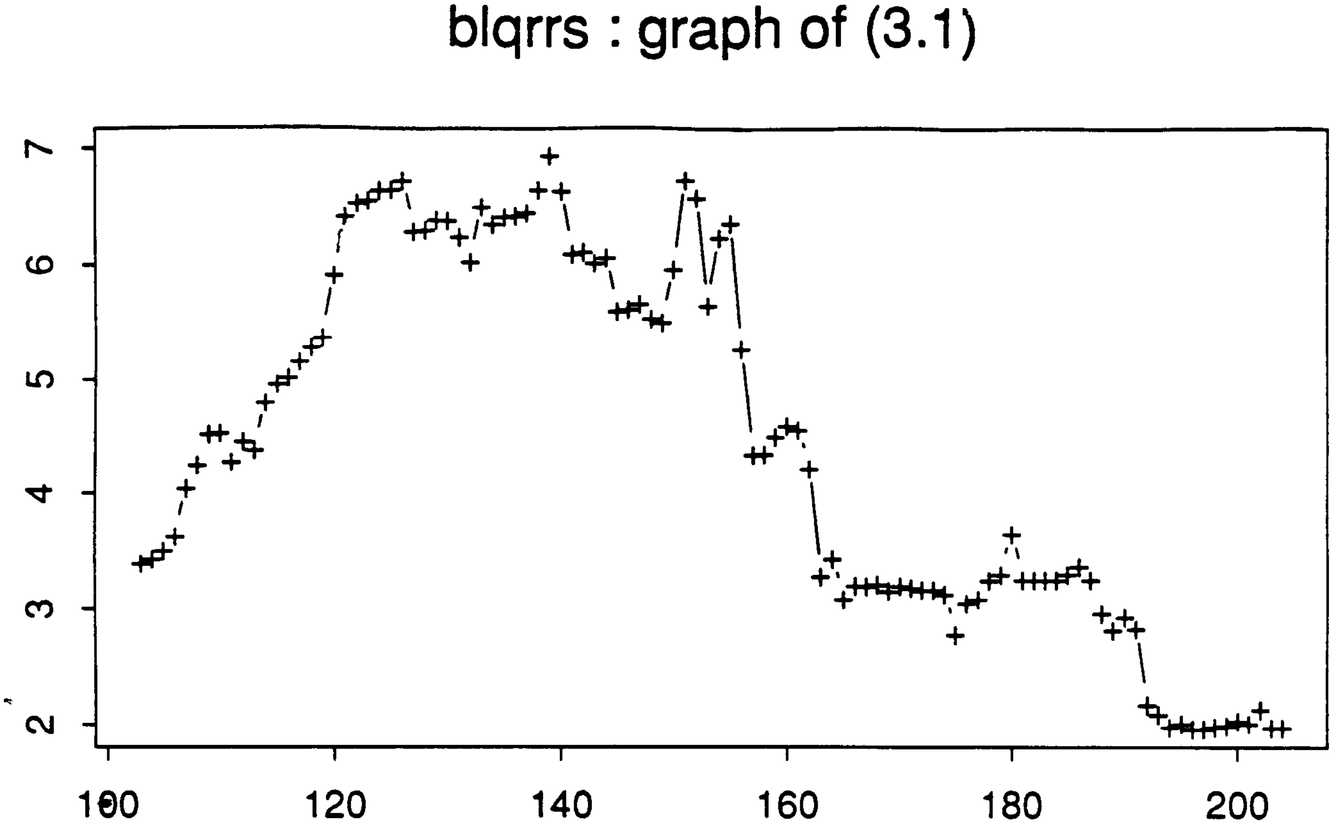
cneths : graph of (3.2)

(b)

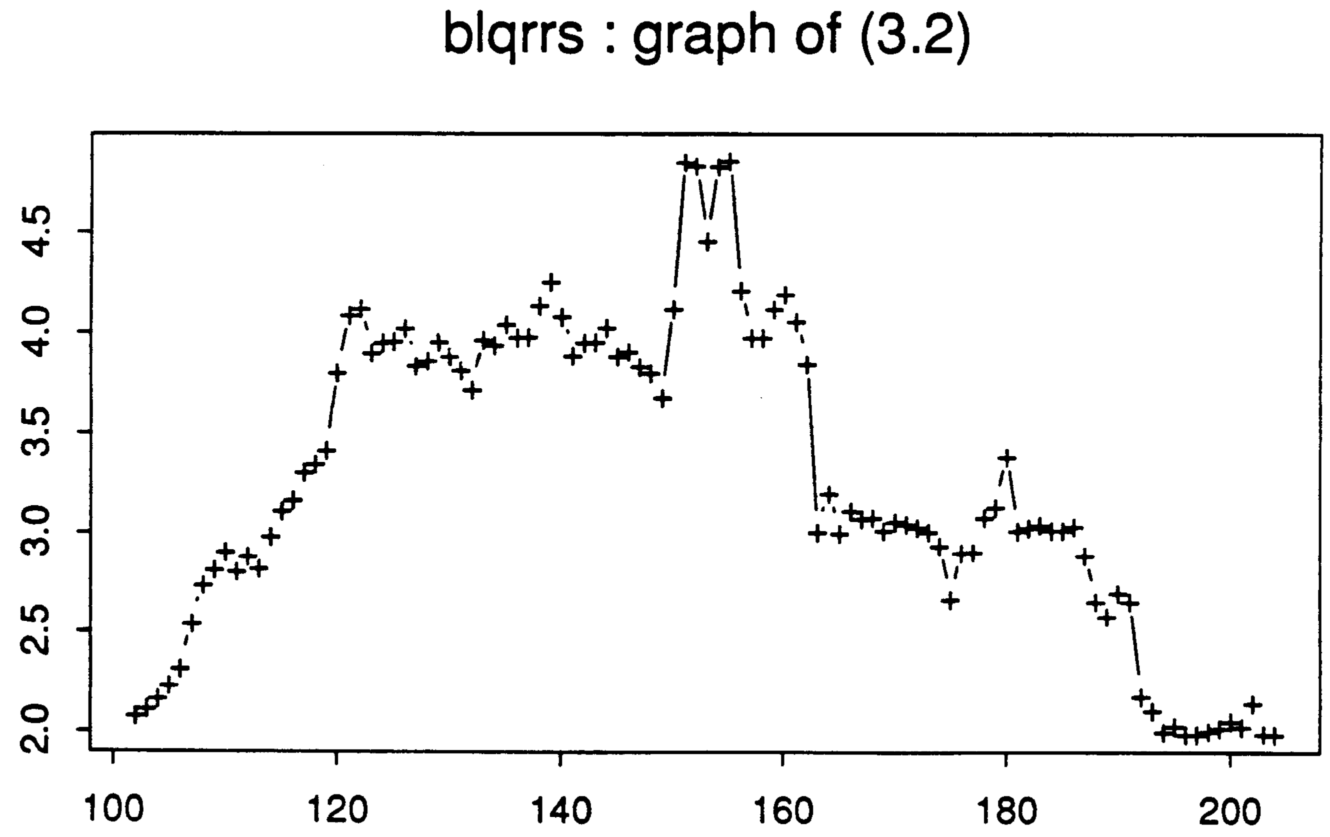
Figure 5.


icmeti : graph of (3.1)
4.5
4.0
3.5
3.0
2.5
2.0
100
120
140
160
180
200
(a)


icmeti : graph of (3.2)
5.0
4.5
4.0
3.5
3.0
2.5
2.0
100
120
140
160
180
200
(b)

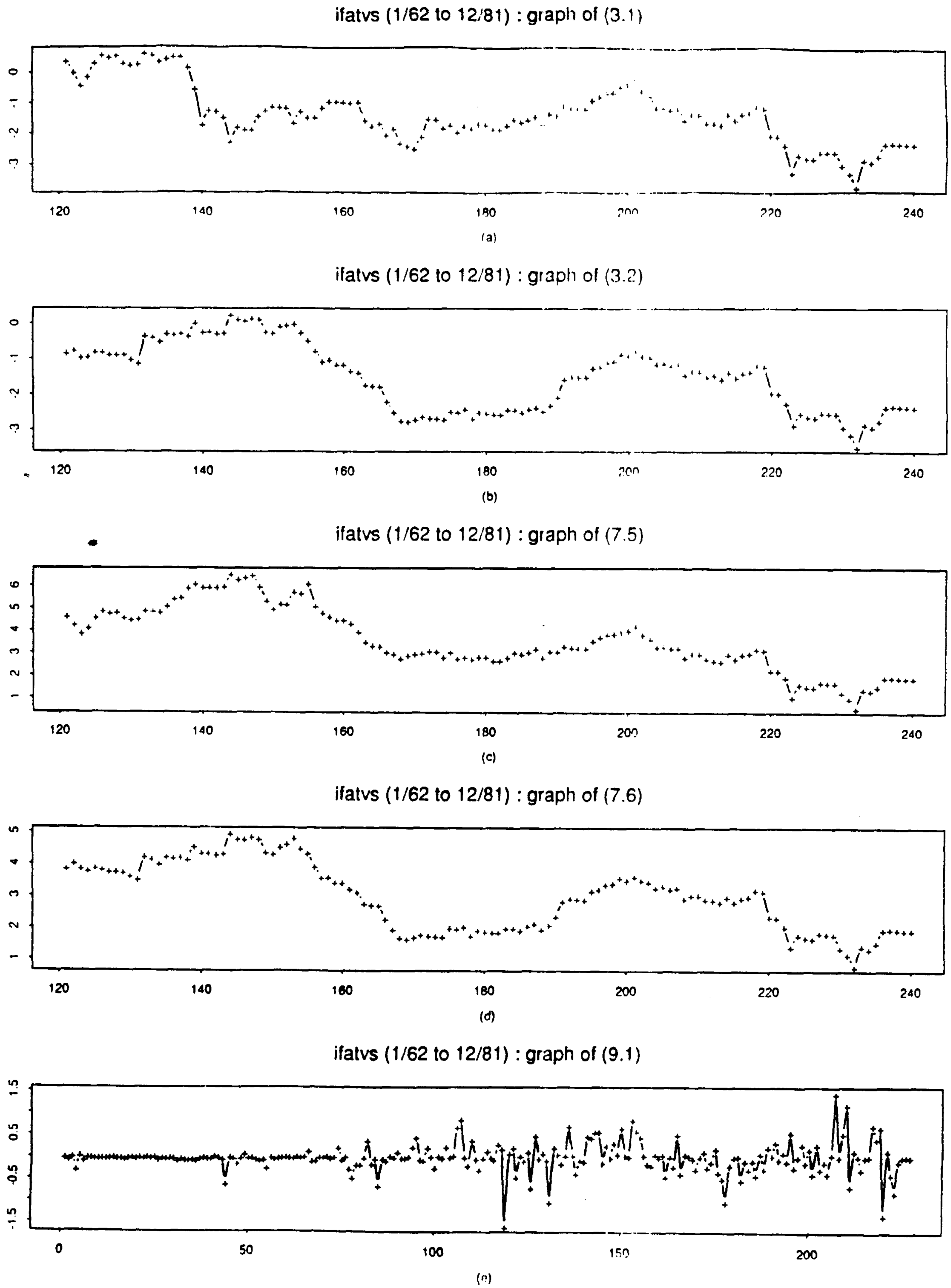

blqrrs : graph of (3.1)

(a)


blqrrs : graph of (3.2)

(b)

Figure 7



ifatvs (1/62 to 12/81) : graph of (3.1)

(a)

ifatvs (1/62 to 12/81) : graph of (3.2)

(b)

ifatvs (1/62 to 12/81) : graph of (7.5)

(c)

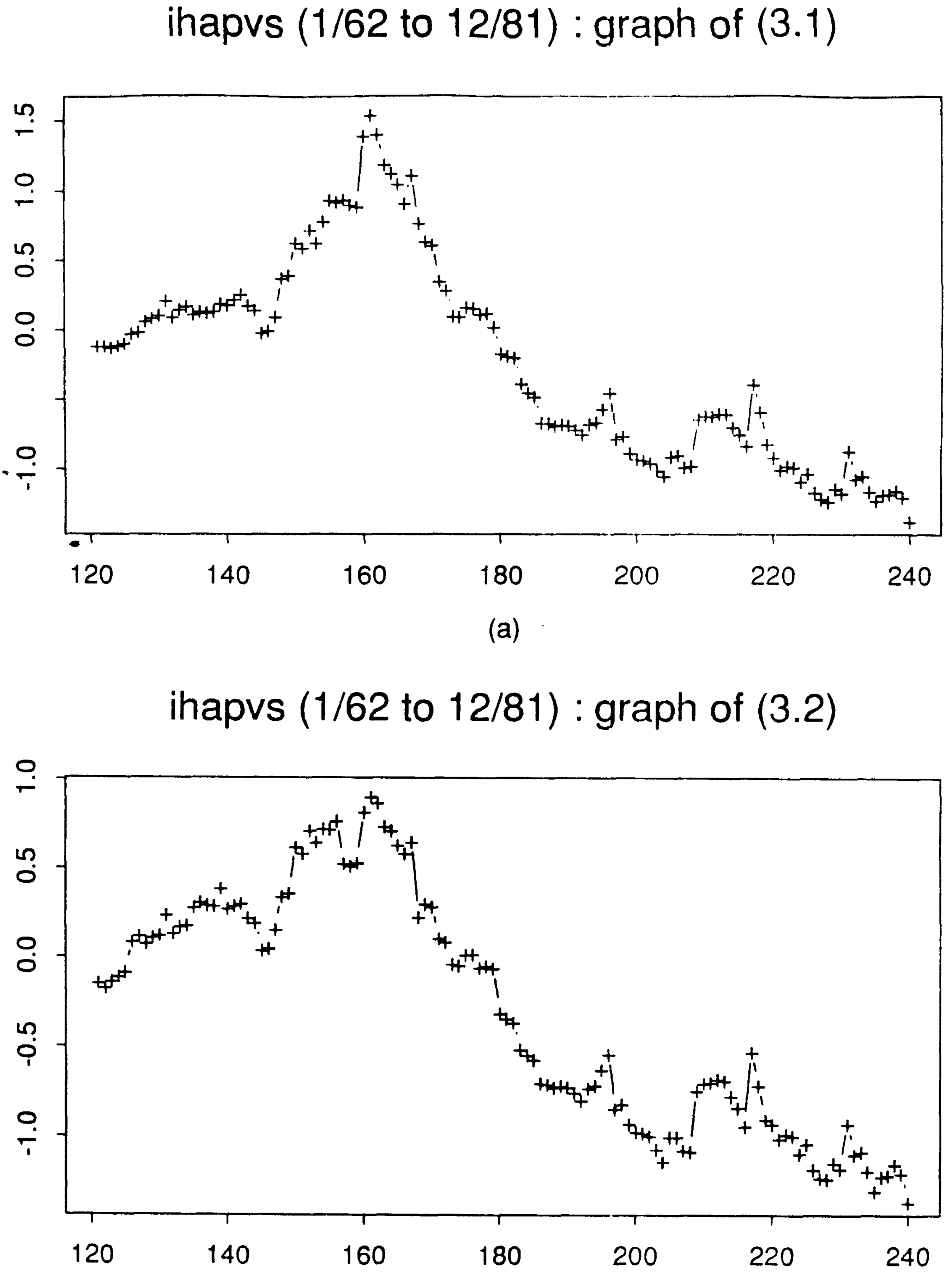
ifatvs (1/62 to 12/81) : graph of (7.6)

(d)

ifatvs (1/62 to 12/81) : graph of (9.1)

(e)

Figure 8



ihapvs (1/62 to 12/81) : graph of (3.1)

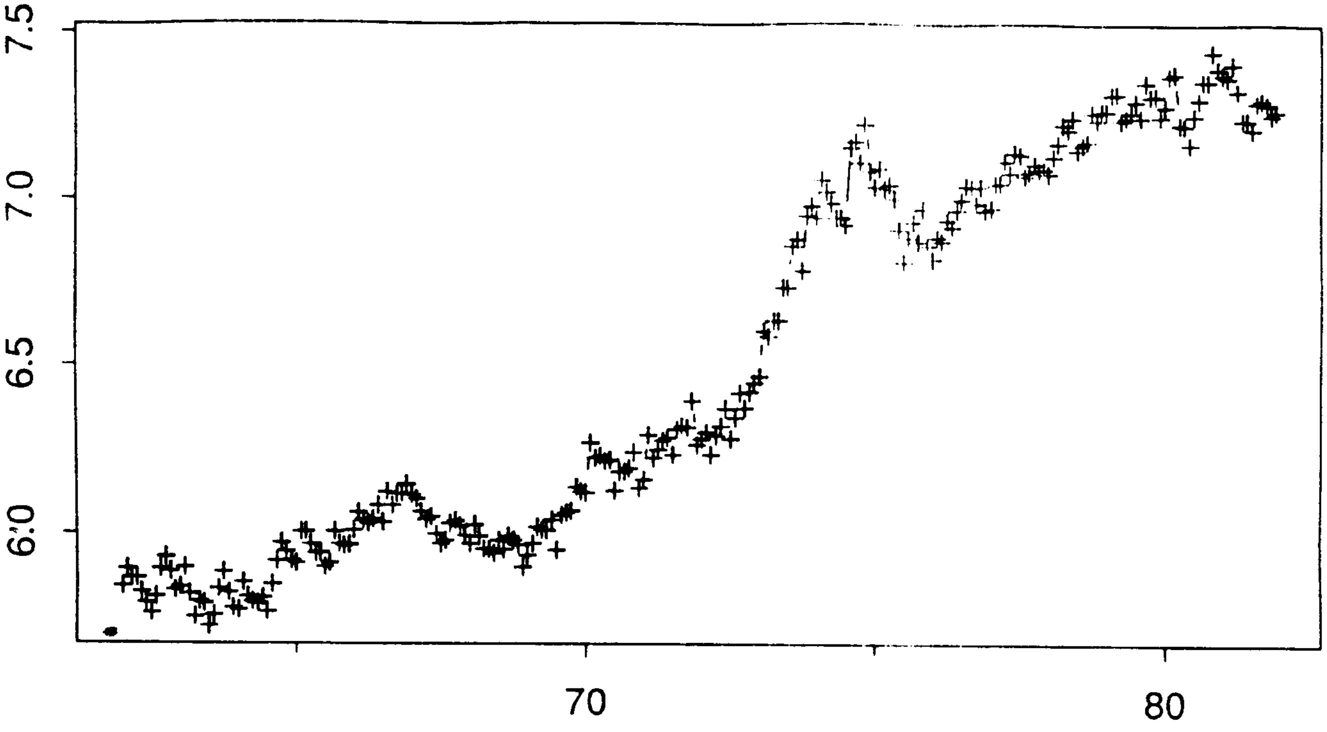
(a)

ihapvs (1/62 to 12/81) : graph of (3.2)

(b)

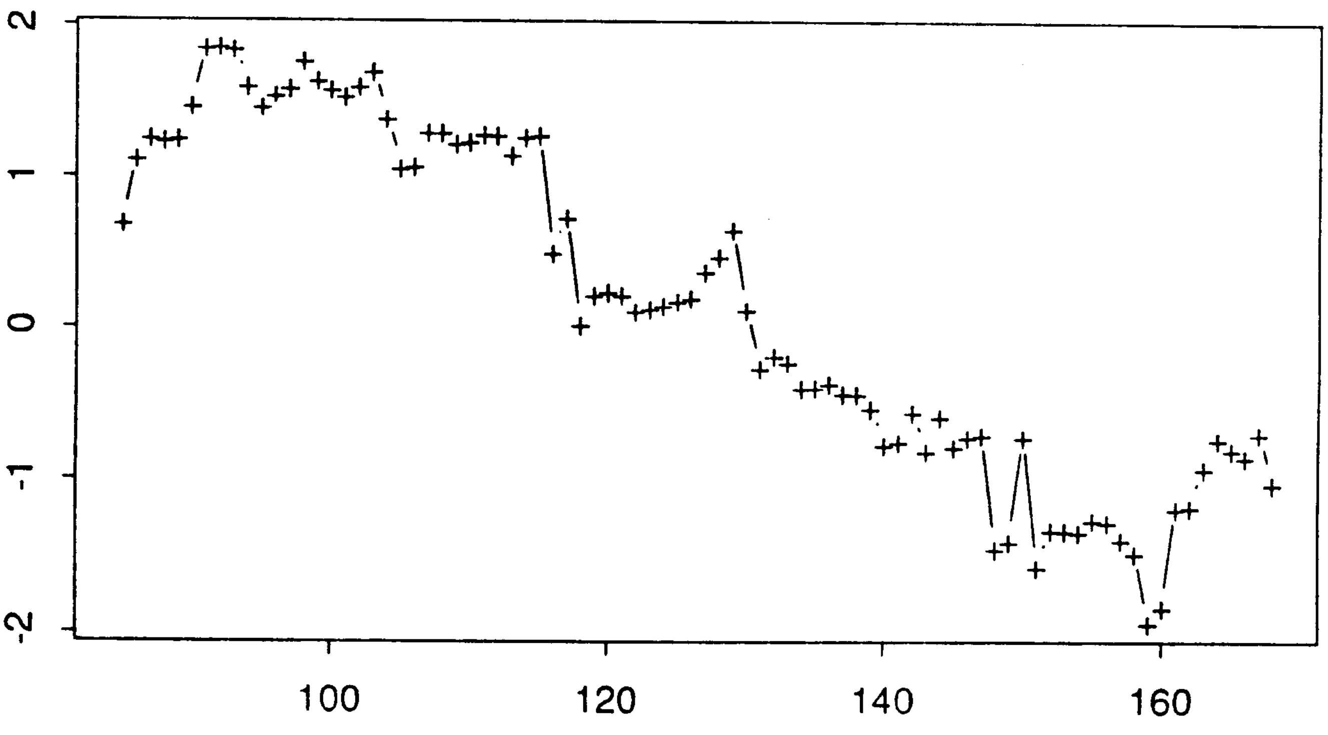
Figure 9

Ln(ifatvs) (1/62 to 1/81)



(i)

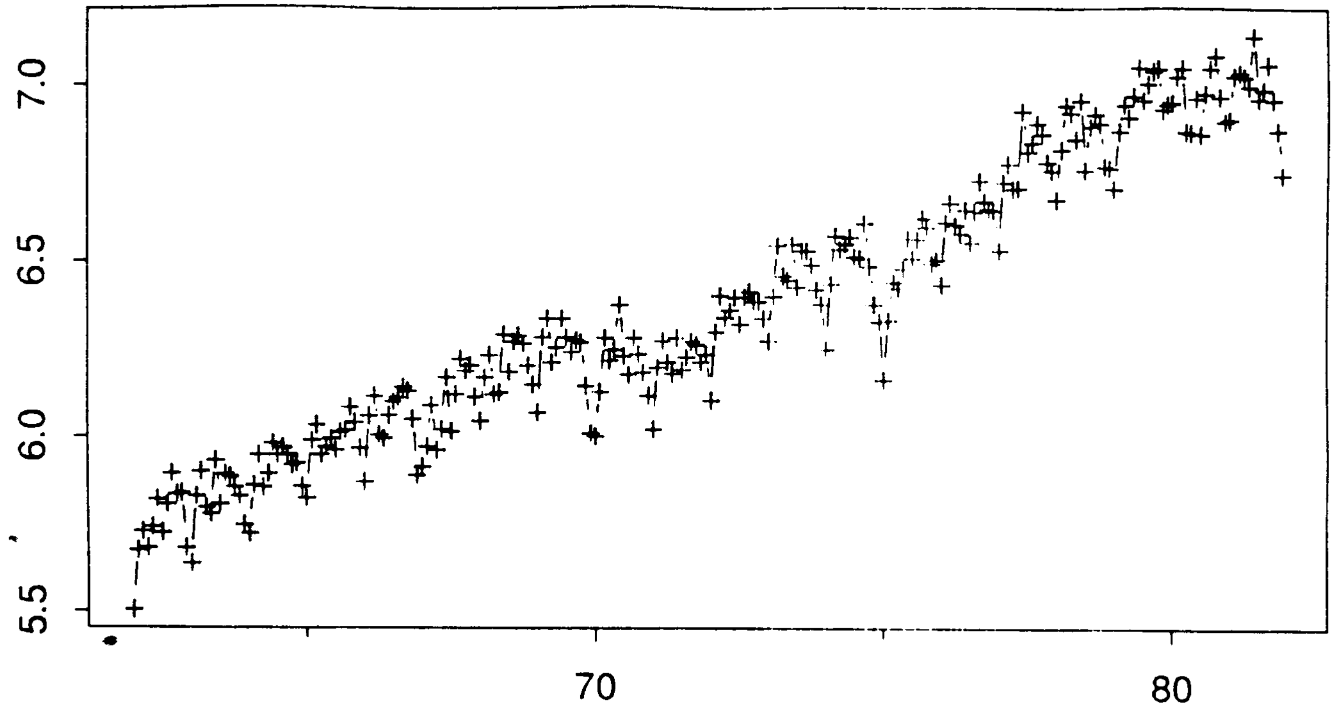
ifatvs (1/68 to 1/81) : graph of (3.2)



(ii)

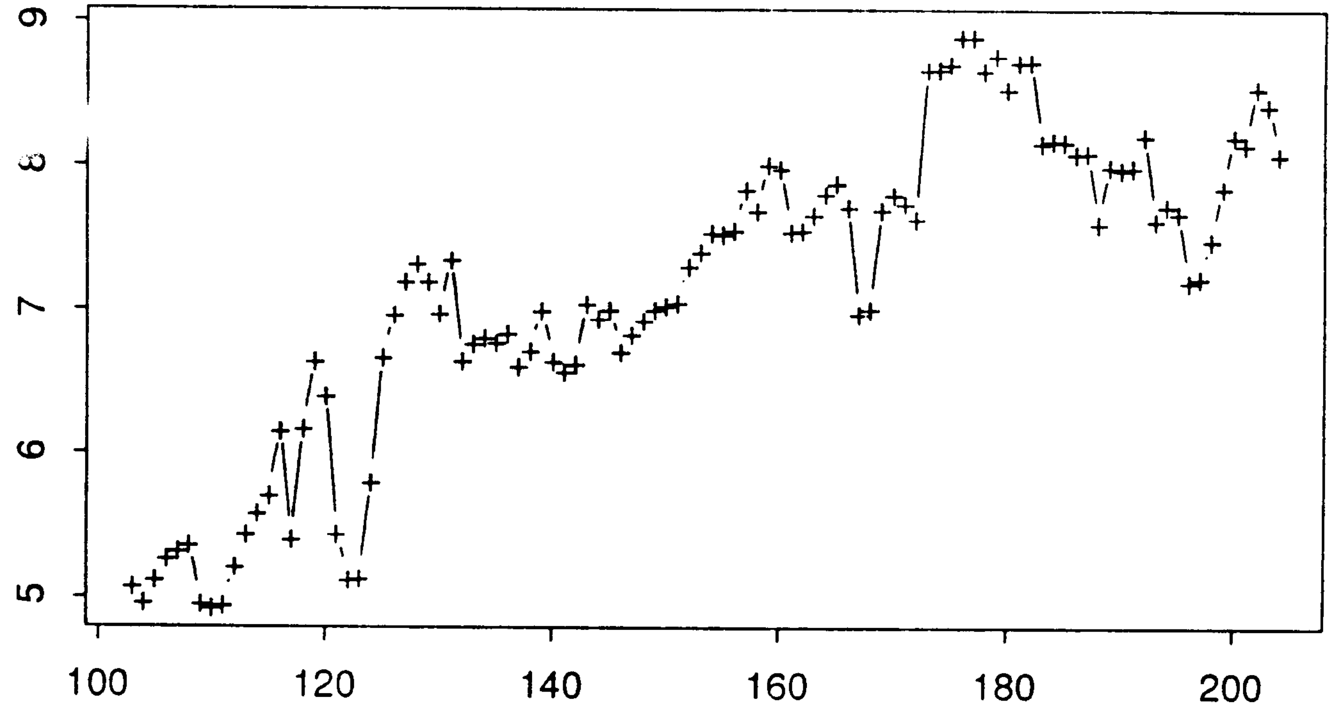
Figure 10.

Ln(ihapvs) (1/62 to 1/81)



(i)

ihapvs (1/65 to 1/81) : graph of (3.2)



(ii)