Chapter 4

Low Breakdown Robust Unit Root Tests

Chapters 4, 5, and 6 form the second part of this thesis. They all deal with the combination of robust estimation and unit root testing for univariate autoregressive time series. The aim of this chapter is to classify the effects of outlying observations in the data on both classical and Bayesian unit root inference.[1] Moreover, a relatively simple procedure is proposed to alleviate the outlier problem in both the classical and Bayesian setting.

The chapter is set up as follows. First, the importance of unit roots in economic time series is briefly discussed in Section 4.1. In Section 4.2, some of the relevant robustness concepts for time series models are introduced. These include the concepts of additive outliers, innovative outliers, and the influence function. In Section 4.3, the Student t based pseudo maximum likelihood estimator is proposed for testing the unit root hypothesis. It is shown that this estimator has a bounded influence function in the time series context, which contrasts with the i.i.d. regression setting. In Section 4.4, some brief comments can be found on the outlier nonrobustness of one of the common Bayesian approaches to testing for unit roots. This section also presents a simple suggestion for alleviating the outlier problem for unit root inference in the Bayesian setting. Section 4.5 applies the developed outlier robust procedures to empirical data. Apart from the extended Nelson-Plosser data, which are also used in Chapter 5, the Finland-U.S. real exchange rate and a series from marketing are considered. Section 4.6 gathers the main conclusions from this chapter. The appendix contains the proof of Proposition 4.1 in Section 4.3.

4.1 Introduction

In this introduction, some intuition is provided for the unit root model. This is done in Subsection 4.1.1. In Subsection 4.1.2, some comments on the rel-

[1] Most of the material in this chapter (except Section 4.1) is taken from Hoek, Lucas, and van Dijk (1995). For style compatibility reasons, I replaced the occurrences of 'we' etc. in the original text by 'I' etc.

evance of unit roots for economics and econometrics are presented. Finally, in Subsection 4.1.3, some standard (nonrobust) univariate unit root testing procedures are briefly discussed.

4.1.1 A Simple Autoregressive Unit Root Model

A large number of economic time series reveals a trending behavior. Let $\{y_t\}_{t=1}^T$ be an observed time series and consider the model

$$\phi(L)(y_t - \alpha - \beta t) = \varepsilon_t, \tag{4.1}$$

with $\{\varepsilon_t\}_{t=0}^\infty$ an i.i.d. process with zero mean, $\phi(z)$ a pth order polynomial in the complex variable z, $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$, L the lag operator, $L^p y_t = y_{t-p}$, and $\alpha, \beta, \phi_1, \ldots, \phi_p$ the set of unknown parameters. Model (4.1) can describe the two types of trending behavior that are mostly studied in econometrics. First, (4.1) can be seen as a deterministic trend model. In that case, the roots of the polynomial $\phi(z)$ must lie outside the unit circle, such that the fluctuations around the deterministic time trend, $\alpha + \beta t$, are stationary. Second, (4.1) can describe a stochastic trend or unit root model, in which case it must hold that $\phi(1) = 0$. The simplest model in this class of unit root models is the one where $\phi(1) = 0$ and the remaining roots of $\phi(z)$ lie strictly outside the unit circle. Note that in that case $\phi(z)$ can be rewritten as $\phi(z) = \phi^*(z)(1 - z)$, with $\phi^*(z)$ a polynomial of degree $p - 1$ with all roots outside the unit circle. For $p = 1$, the deterministic and stochastic trend model reduce to (4.2) and (4.3), respectively:

$$\begin{aligned} y_t &= \phi y_{t-1} + (1-\phi)\alpha + \phi\beta + (1-\phi)\beta t + \varepsilon_t, && (4.2)\\ \Delta y_t &= \beta + \varepsilon_t, && (4.3) \end{aligned}$$

with Δ the first difference operator, $\Delta y_t = (1 - L)y_t = y_t - y_{t-1}$, and $|\phi| < 1$.

The difference between (4.2) and (4.3) most clearly emerges if one rewrites (4.2) and (4.3) as

$$\begin{aligned} y_t &= \phi^t y_0 + (1-\phi^t)\alpha + \beta t + \sum_{i=0}^{t-1} \phi^i \varepsilon_{t-i}, && (4.2')\\ y_t &= y_0 + \beta t + \sum_{i=0}^{t-1} \varepsilon_{t-i}. && (4.3') \end{aligned}$$

Both models contain a deterministic linear trend with a slope coefficient β. The influence of the first observation (y_0), however, differs markedly between the two models. In (4.2′) the influence of y_0 on y_t becomes negligible if t increases. In contrast, in (4.3′) the influence of y_0 on all future values of the time series remains constant. An analogous result holds for all other values of the time series. The influence of y_{t_0} on y_t for $t > t_0$ decreases with t for the deterministic trend model ($|\phi| < 1$), whereas it remains constant for the unit root model ($\phi = 1$). Alternatively, one can consider the impact of the disturbances ε_t on the time series y_t. (4.2′) and (4.3′) demonstrate that shocks have a temporary effect in the deterministic trend model and a permanent effect in the unit root model. Therefore, the unit root model and the unit root hypothesis are closely linked to the phenomenon of shocks having a permanent rather than a transitory effect.

The second important difference between (4.2′) and (4.3′) emerges by considering the deviations from the deterministic trend line. For (4.2′), one obtains the deviations $u_t = \sum_{i=1}^{t} \phi^{t-i}\varepsilon_i$. Note that $u_t = \phi u_{t-1} + \varepsilon_t$, with $u_0 = 0$ and $|\phi| < 1$. Therefore, $\{u_t\}$ is an asymptotically stationary process.[2] For (4.3′), in contrast, one obtains $\tilde{u}_t = \sum_{i=1}^{t} \varepsilon_i$ as the deviation from the deterministic trend line. The $\tilde{u}_t$ process is a random walk and, therefore, clearly nonstationary as its variance increases linearly over time. This different behavior of u_t and $\tilde{u}_t$ has important consequences for statistical inference procedures. If (4.2) is the actual data generating mechanism, one can use standard statistical tools and standard asymptotic distribution theory to perform inference on the parameters. If (4.3) generated the data, however, nonstandard asymptotic distribution theory is needed. This will become apparent in the subsequent chapters of this thesis.

The unit root hypothesis is concerned with the question whether (4.2) or (4.3) provides a better statistical representation of the data. From (4.2) and (4.3) it is evident that this hypothesis can be investigated by testing whether ϕ in (4.2) is significantly different from unity. One can question the importance of this specific value of the parameter ϕ. The above discussion illustrates, however, that there are important differences between the two competing models, both from the point of statistical analysis and from the point of (economic) interpretation. This motivates the subsequent investigations.

4.1.2 Relevance of Unit Roots for Economics and Econometrics

I now turn to the economic and econometric relevance of unit roots. Following the seminal article of Nelson and Plosser (1982), an interesting debate has been going on in the econometric literature whether most macroeconomic time series contain a deterministic or a stochastic trend. Especially the lasting effect of shocks under the unit root hypothesis, the difference between modeling levels versus differences, and the consequences of unit roots for forecasting seem to be of major economic and econometric importance. Nelson and Plosser (1982) found that from the fourteen economic time series they considered, thirteen were best described by models of the form (4.3) rather than (4.2). They concluded

> ...that macroeconomic models that focus on monetary disturbances as a source of purely transitory fluctuations may never be successful

[2] u_t can be made stationary by redefining $u_0 = \varepsilon_0/(1-\phi^2)^{1/2}$.

> in explaining a large fraction of output variation and that stochastic variation due to real factors is an essential element of any model of macroeconomic fluctuations.

So, based upon a simple univariate analysis of the unit root hypothesis, the authors drew conclusions that are of major importance for economic modeling.

The results of Nelson and Plosser (1982) have been thoroughly put to the test over the last decade. It turns out that for several economic time series the arguments in favor of the stochastic trend model are not as strong as suggested by Nelson and Plosser (see, e.g., Perron (1989)). The arguments against the unit root model and against the unit root testing problem in general can be classified into four catergories, namely: the unit root model is unrealistic for most economic phenomena, the unit root tests have low power against certain stationary alternatives, the unit root tests are sensitive to outliers and structural changes, and the unit root model generates only one particular type of nonstationarity. Below, I discuss each of these issues in more detail.

First, unit root models are sometimes called unrealistic from an economic perspective. Nelson and Plosser (1982), for example, found that the stochastic trend or unit root model provides a better statistical description of the interest rate r_t than the deterministic trend model. If one asserts that the interest rate has a unit root and if one abstracts from the possibility of a (constant) deterministic time trend in r_t, then the dominating driving factor of r_t is a random walk process. This implies that, with probability one, the interest rate will exceed any given threshold value (see, e.g., Feller (1968, Chapter 14)). Therefore, if a statistical test procedure fails to reject the unit root hypothesis for r_t, one may still be reluctant to conclude that r_t has a unit root, because the implications of such a statement are economically implausible.

Against such a reasoning one can argue that there are economic time series for which the unit root model is perfectly plausible. Perhaps the most well known series are those of exchange rate returns. Based on simple arbitrage arguments, one can show that such series will demonstrate a stochastic trend behavior (de Vries (1994)). Moreover, it can be shown that the expected waiting time for r_t to cross the above mentioned threshold, is infinite (see Feller (1968)). Thus, the stochastic trend model need not generate unrealistic predictions for r_t if the model is not extrapolated too far into the future. Finally, the unit root model can be viewed as a statistical device for modeling the salient features of the data. Such a strategy can be very useful, even if one does not believe that a unit root is actually present in the data generating mechanism. For example, for processes with roots less than, but close to unity, the statistical theory for unit processes can provide a better guide for inference procedures in finite samples than standard central limit theory based on asymptotic normality arguments (see, e.g., Phillips (1988), Magnus and Rothenberg (1988), and Stock (1994)).

A second point that is often raised against the use of unit root tests, is that these tests have low power in finite samples against certain stationary alternatives. A simple illustration of this argument is presented by Cochrane

(1991). Consider the model

$$(1-\phi L)y_t = \varepsilon_t, \qquad t = 1, 2, \ldots \tag{4.4}$$

with $\{\varepsilon_t\}_{t=0}^{\infty}$ a Gaussian i.i.d. process, $|\phi| < 1$, and $y_0 = \varepsilon_0/(1-\phi^2)^{1/2}$. (4.4) obviously describes a stationary process, but for a fixed sample size this process cannot be distinguished from a unit root process if ϕ is sufficiently close to one. Consequently, in finite samples stationary processes can be found that are arbitrarily close to unit root processes. Similarly,

$$(1-L)y_t = (1-\theta L)\varepsilon_t, \tag{4.5}$$

with $|\theta| < 1$, describes a unit root process. Assuming that $\varepsilon_t = 0$ for $t \leq 0$, (4.5) can be rewritten as

$$y_t = y_0 + \theta\varepsilon_t + (1-\theta)\sum_{i=1}^{t}\varepsilon_i. \tag{4.6}$$

By setting θ arbitrarily close to unity, a process generated by (4.5) cannot be distinguished from a white noise process in a fixed, finite sample.

Both of the above comments are valid and are of some concern to the applied researcher. They illustrate that a statistical test cannot always be conclusive in finite samples and that both severe size distortions and a low power are in a certain sense inherent to the unit root testing problem. Some of the properties of models like (4.4) and (4.5) have already been investigated. Phillips (1987b) considered models of the form (4.4) with $\phi = 1 - c/T$, where c is some constant and T denotes the sample size. Such processes describe autoregressive models with roots that are close to unity. Phillips proved that the statistical inference procedures based on unit root asymptotics are more suitable for processes with roots close to the unit circle, even though these processes may be (asymptotically) stationary. This illustrates the usefulness of studying unit root models even if these models are not strictly applicable (see also Campbell and Perron (1991)). The properties of unit root tests for models like (4.5) are also well documented in the literature. The major conclusion that emerges from this literature is that standard unit root tests like the Dickey-Fuller t-test (see Dickey and Fuller (1979)) suffer from severe size distortions if the moving average part $((1-\theta L)\varepsilon_t)$ is not taken into account (see Herce (1994), Schwert (1987), and Chapters 5 and 6). These size distortions are not satisfactorily solved in finite samples by employing correction techniques that are guaranteed to work asymptotically, like the nonparametric corrections put forward by Phillips (1987) and Phillips and Perron (1988).

A third point of critique to the application of standard unit root tests like those of Dickey and Fuller (1979, 1981), Phillips (1987), and Phillips and Perron (1988), is that they are sensitive to the occurrence of atypical events. These events might have a different effect on unit root inference, depending on whether their impact has a temporary or permanent character. Perron

(1989) showed that standard unit root tests break down if there is a structural break in the data generating process, e.g., a level shift. The intuitive idea is that the unit root hypothesis is closely associated with shocks having a permanent effect. A structural break essentially corresponds to a shock with a lasting effect on the time series (Perron and Vogelsang (1992)). If this shock is not explicitly taken into account, standard unit root tests will mistake the structural break for a unit root. Some entries to the literature on structural breaks are Perron (1989), Stock (1994), and the special issue of the *Journal of Business and Economic Statistics* (1992, volume 10).

Another type of atypical event is the additive outlier. This is an event with a large, temporary effect on the series. In certain cases, this effect dominates the remaining information contained in the series and, thus, biases unit root inference towards rejection of the unit root hypothesis. Relevant references are Franses and Haldrup (1994) and Chapters 4 through 6. More on additive outliers can be found in Section 4.2. Additive outliers can be easily dealt with using outlier robust estimation procedures.

The fourth and final critique to the application of unit root tests concerns the fact that the unit root model generates only one particular form of nonstationarity. A rejection of the unit root hypothesis does not necessarily mean that the data are stationary. A simple counterexample is given by the model $y_t = (\sigma_1 + \sigma_2 1_{\{t \geq \lfloor T/2 \rfloor\}})\varepsilon_t$, where ε_t is standard Gaussian white noise, and 1_A is the indicator function for the set A. This model generates a variance change in the series y_t. It is easily shown that the standard t-test for $\hat{b} = 1$ in the regression model $y_t = by_{t-1} + e_t$ will reject almost surely when the sample size tends to infinity.

The notions of a unit root and nonstationarity are almost used as synonyms in the contemporary econometric literature. As the example above demonstrates, however, one should carefully distinguish between the two concepts. The presence of a unit root implies nonstationarity, but one cannot go much further than that. On the one hand, a faillure to reject the unit root hypothesis does not imply that the series contains a unit root, as is nicely exemplified by the literature on structural breaks (see, e.g., Perron (1989)). On the other hand, a rejection of the unit root hypothesis does not imply that the series is stationary (see the example above with the variance shift).

For more details on unit roots and economic time series, the reader is referred to Diebold and Nerlove (1990), Campbell and Perron (1991), and Stock (1994) for reviews on classical unit root analysis. For Bayesian contributions, some references are Sims (1988), DeJong and Whiteman (1991a,b), Phillips (1991b), Schotman and van Dijk (1991a,b, 1993), and Kleibergen and van Dijk (1993).

4.1.3 Some Standard Unit Root Tests

In this subsection, two of the most well known univariate unit root testing procedures are discussed, namely those put forward by Dickey and Fuller (1979) (see also Fuller (1976)). Outlier robust variants of these testing procedures are discussed in Chapters 4 through 6.

Before giving the details of the tests, I first introduce some additional terminology that is often used in the unit root context. Define the dth order difference operator Δ^d as $(1-L)^d$. Then a process $\{y_t\}$ is said to be *integrated* of order d, denoted as $y_t \sim I(d)$, if $\{\Delta^k y_t\}$ is nonstationary for $k = 0, \ldots, d-1$, and stationary for $d = k$. This means that dth order differencing suffices to make the series stationary, but that lower order differencing does not suffice. The two orders of integration that are mostly used in economics are zero and one. A series is $I(1)$ if it has a unit root that can be removed by differencing. It is $I(0)$ if the process is (asymptotically) stationary, while its partial sums constitute an $I(1)$ process.[3] The highest order of integration encountered in economics is two, which is sometimes used in the context of monetary variables (see, e.g., Juselius (1995)).

In order to test whether a series is $I(1)$ or $I(0)$, Dickey and Fuller (1979) proposed the following procedure (see also Fuller (1976)). They started with an autoregressive representation of a time series y_t, namely

$$\phi^*(L)y_t = (1 - \phi_1^* L - \ldots - \phi_p^* L^p)y_t = \varepsilon_t, \tag{4.7}$$

where $\{\varepsilon_t\}$ is a white noise process with finite variance, L is the lag operator, and $\phi^*(z) = 0$ for $z \in \mathbb{C}$ implies either $|z| > 1$ or $z = 1$. Using the decomposition of Beveridge and Nelson (1981) of the polynomial $\phi^*(L)$, (4.7) can be rewritten as

$$(\phi(L)\Delta + \phi^*(1)L)y_t = \varepsilon_t, \tag{4.8}$$

with $\phi(L) = 1 - \phi_1 L - \ldots - \phi_{p-1}L^{p-1}$, $\phi_i = -\sum_{j=i+1}^{p} \phi_j^*$, and $\phi_0 = -\phi^*(1)$. Obviously, $z = 1$ is a root of $\phi^*(z)$ if and only if $\phi^*(1) = 0$. So the unit root hypothesis can be tested by considering whether $\phi_0 = 0$ in the regression model implied by (4.8):

$$\Delta y_t = \phi_0 y_{t-1} + \phi_1 \Delta y_{t-1} + \ldots + \phi_{p-1}\Delta y_{t-p+1} + \varepsilon_t. \tag{4.9}$$

If $\phi_0 = 0$, (4.9) is an AR($p-1$) model in first differences, while if $\phi_0 \neq 0$ (subject to the restrictions mentioned below (4.7)), (4.9) describes a stationary AR(p) process. Dickey and Fuller proposed to estimate the parameters in (4.9) with OLS. The unit root hypothesis can then be tested either using the ordinary t-test statistic for ϕ_0, hereafter denoted as the Dickey-Fuller t-test, or using the statistic $T\hat{\phi}_0$, where $\hat{\phi}_0$ is the estimate of ϕ_0. Under the null hypothesis of a unit root ($\phi_0 = 0$), these test statistics have nonstandard limiting distributions. For example, the t-test for $\phi_0 = 0$ in (4.9) is not asymptotically normally distributed under the null hypothesis. Instead, its limiting distribution can be expressed in terms of a functional of certain stochastic processes, namely Brownian motions. An explicit form of the distribution can be derived using the results of Evans and Savin (1981, 1984) or Abadir (1992).

[3]This last condition is needed in order to exclude processes of the form $y_t = \varepsilon_t - \varepsilon_{t-1}$ to be called $I(0)$ processes, where $\{\varepsilon_t\}$ is an i.i.d. process.

There are several extensions of the original Dickey-Fuller procedure. Said and Dickey (1984) extended the test for dealing with general ARMA processes. Phillips (1987) proposed simple modifications of the Dickey-Fuller tests that have the same limiting distributions under a wide variety of data generating processes. These modified tests allow for quite heterogeneous and temporally dependent error processes $\{\varepsilon_t\}$. Also models of the form

$$\phi^*(L)y_t = \alpha_0 + \alpha_1 t + \ldots + \alpha_q t^q + \varepsilon_t \tag{4.10}$$

have been investigated (see, e.g., Fuller (1976), Phillips and Perron (1988)). It turns out that the incorporation of deterministic functions of time as additional regressors has major consequences for the asymptotic distributions of the unit root test statistics (see also Chapter 6).

Upon closing this introduction, it is useful to note that the test for $\phi_0 = 0$ in (4.9) only considers one special type of unit root. It does not, for example, test for complex unit roots of modulus one. Complex unit roots are encountered if one investigates the order of integration for time series that exhibit seasonality. Straightforward extensions of the Dickey-Fuller methodology to tests for seasonal unit roots can be found in Hylleberg et al. (1990).

4.2 Some Robustness Concepts

In order to analyze the effects of outliers on unit root inference and to propose methods that are less sensitive in this respect, some additional concepts from the literature on robust statistical inference are introduced. First, I discuss a model that generates outliers in a time series context. A useful model is the general replacement model, given in Martin and Yohai (1986):

$$y_t = (1 - z_t)x_t + z_t w_t. \tag{4.11}$$

The Bernoulli random variable z_t equals 1 with probability γ and is 0 otherwise. The *core* or *outlier free* process, $\{x_t\}$, has cumulative distribution function (c.d.f.) $F_x(\cdot)$, while the *contaminating* process, $\{w_t\}$, has c.d.f. $F_w(\cdot)$. Both of these processes can be non-i.i.d. For instance, they may belong to the class of autoregressive moving average (ARMA) processes. Finally, the realization of the y_t process contains the actually observed values of the time series. Note that $y_t = w_t$ with probability γ and $y_t = x_t$, otherwise. The parameter γ controls the amount of contamination. Usually γ is small, typically 0.05 to 0.15. The c.d.f. of the y_t process obviously depends on γ. This is denoted by adding a superscript to it: $F_y^\gamma(\cdot)$.

Model (4.11) is easily recognized as a two-component mixture model, since $F_y^\gamma(\cdot) = (1-\gamma)F_x(\cdot) + \gamma F_w(\cdot)$. Therefore, the literature on modeling finite components mixtures could be used for constructing outlier robust inference procedures. In the present parametric context, this requires the full specification of the w_t process. I refrain from this strategy and let the w_t process be (partially) unspecified, see (4.12) and (4.13), below. This allows me to specify procedures that are robust to more general types of outliers than those implied by the mixture of *two* stochastic processes.

Informally, the aim in this chapter is to develop procedures that are nearly optimal for $\gamma = 0$ and 'satisfactory' for γ equal to some small, positive number. The advantage of this approach is that it is not necessary to specify a complete model for the w_t process. This automatically yields a more parsimonious model parameterization. Also, the estimation of a finite components mixture model may be problematic if there are only few outliers. In that case there is little information in the data to identify the values of the parameters of those components of the mixture that correspond to the small group of outliers.

It is worth mentioning at this stage that in the subsequent analysis the Student t distribution is used for dealing with the outlier problem. The Student t distribution is an uncountable mixture of Gaussian distributions, with the χ^2 distribution as the mixing distribution. The advantage of the Student t is that it has attractive robustness properties while one does not have to specify the number of components in the mixture distribution. This contrasts with the finite components mixture model.

By imposing a certain structure on the z_t and w_t processes, model (4.11) can generate different types of outliers. For example, if the z_t process is i.i.d., (4.11) generates isolated outliers. If the z_t are intertemporally dependent, patches of outliers can occur (compare the examples in Martin and Yohai (1986)). The two types of outliers usually encountered in the literature are additive outliers (AO's) and innovative outliers (IO's). The difference between these two types is most easily illustrated using an AR process and the general replacement model (4.11). Let $\{\xi_t\}$ be a process independent of $\{x_t\}$ and $\{z_t\}$. AO's can now be modeled by specifying

$$\text{AO:} \qquad \begin{aligned} \phi(L)x_t &= \varepsilon_t \\ w_t &= x_t + \xi_t \end{aligned} \qquad t = 1, \ldots, T \tag{4.12}$$

and IO's by specifying

$$\text{IO:} \qquad \begin{aligned} \phi(L)x_t &= \varepsilon_t \\ w_t &= x_t + \xi_t/\phi(L) \end{aligned} \qquad t = 1, \ldots, T, \tag{4.13}$$

where $\phi(L)$ is a polynomial of order p in the lag operator L, $\{z_t\}$ and $\{\varepsilon_t\}$ are i.i.d. processes, and $\{\xi_t\}$ is an i.i.d. process that is independent of the other processes.

Using (4.11) and the AO specifications of x_t and w_t, one obtains that $y_t = x_t + z_t\xi_t$. Given that z_t equals 0 most of the time, y_t is mostly equal to the uncontaminated AR process x_t. Now and then, i.e., with probability γ, y_t is observed with a measurement error ξ_t. For IO's, in contrast, one can write $\phi(L)y_t = \varepsilon + z_t\xi_t$. The additional error term now appears in the innovations that drive the time series.

The effects of both isolated AO's and IO's are visualized in Figure 4.1. The two upper panels in the figure show a realization of the y_t process for

$\phi(L) = 1 - 0.9L$, $\gamma = 0.01$, $w_t \equiv 10$, $T = 100$, and $\{\varepsilon_t\}_{t=1}^{T}$ a set of i.i.d. standard normal innovations. The lower two panels present scatter diagrams of y_t versus y_{t-1} for both time series. The same set of innovations was used to generate both series. The outlier occurs at time $t = 25$.

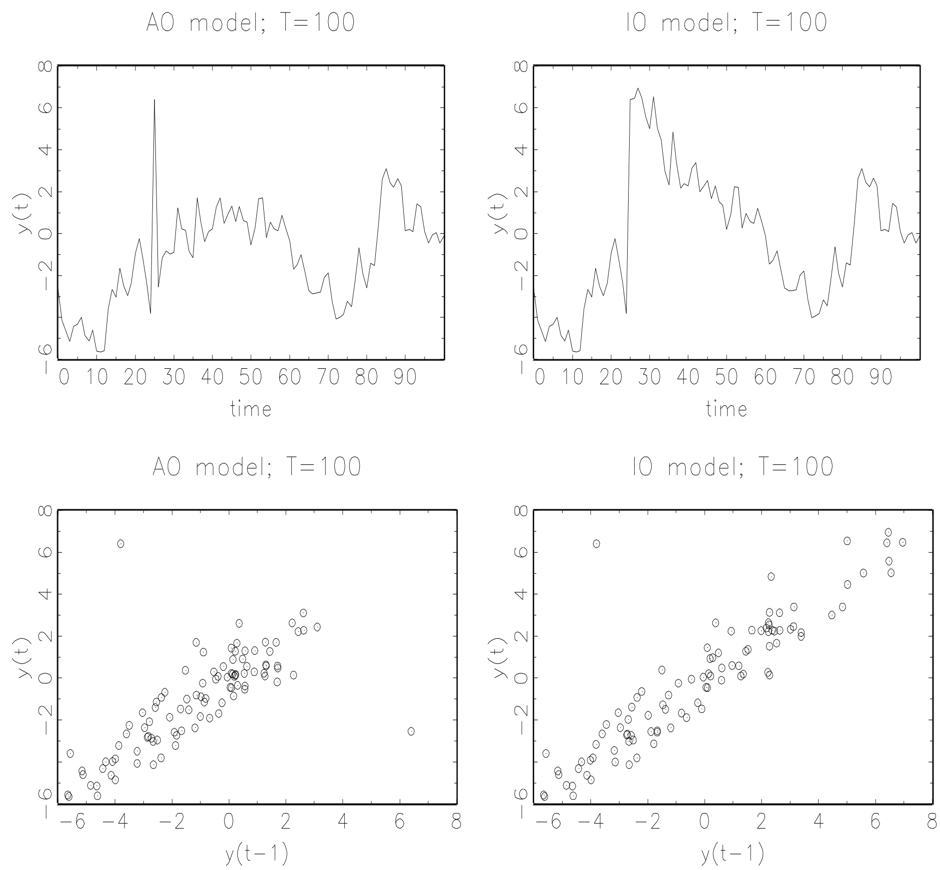

Figure 4.1.— Effects of Isolated Additive and Innovative Outliers

It is clearly seen that the AO only causes a single departure from the normal pattern of the time series. The series jumps upward at the time the outlier occurs and immediately jumps back the period afterwards. As can be seen in the scatter diagram this causes two outliers in the (y_{t-1}, y_t)-plane.

The IO also causes the series to jump upward at the time of the outlier. Afterwards, however, the series gradually adjusts to its normal pattern. In the scatter diagram this results in one vertical outlier, followed by a set of points with large y_t and large y_{t-1} values that all lie in the neighborhood of the line with slope $\phi = 0.9$. These last observations provide a strong signal of the true value ϕ.

Note that if $\phi(L) = 1 - L$, the AO pattern remains comparable to the one shown in Figure 4.1, whereas the IO results in a level shift.

Outliers may seriously affect the 'usual' estimation and testing procedures, like those based on ordinary least-squares (OLS). If one is interested in describing the bulk of the data, then procedures have to be developed that are less sensitive to the presence of aberrant observations. Several outlier robust estimators have been proposed in the literature (see, e.g., Huber (1981) and Hampel et al. (1986)). In order to evaluate the properties of such estimators, different concepts are available. Among these, the influence function (IF) plays a prominent role. Heuristically, the IF measures the change in the value of an estimator when a few outliers are added to the sample (see also Chapter 2). Its finite sample approximation for the OLS estimator is closely related to the DFBETA diagnostic of Belsley, Kuh, and Welsch (1980), which measures the standardized contribution of the tth observation to the estimator (see Hampel et al. (1986, Section 2.1.e)).

Formally, the IF is an asymptotic concept. For the i.i.d. regression setting, it was already discussed in Chapter 2. Defining the IF in the more general context of dependent observations is less trivial (see Künsch (1984) and Martin and Yohai (1986)). Here, I provide the definition of Martin and Yohai (1986). I consider estimators $\hat{\phi}$ that can be considered as functionals on the space of c.d.f.'s, so $\hat{\phi} = \hat{\phi}(F_y^\gamma)$. Consider the change in $\hat{\phi}(\cdot)$ that is implied by increasing γ from 0 to some small positive number, so, $(\hat{\phi}(F_y^\gamma) - \hat{\phi}(F_y^0))$. If γ is very small, this difference is negligible if $\hat{\phi}(\cdot)$ is continuous. Therefore, the difference is standardized by dividing it by γ. The IF is now defined as the limit of this standardized difference as γ approaches 0:

$$IF(\hat{\phi}, \{F_y^\gamma\}) = \lim_{\gamma \downarrow 0} \frac{\hat{\phi}(F_y^\gamma) - \hat{\phi}(F_y^0)}{\gamma}, \tag{4.14}$$

if this limit exists.

Note that estimators with a bounded IF are desirable, because outliers only have a bounded influence on such estimators. Martin and Yohai (1986) proved that the OLS or conditional Gaussian maximum likelihood (ML) estimator for AR models has an unbounded IF under AO contamination. This suggests that the estimator is not robust.[4] The nonrobustness of the OLS estimator can be made explicit quite easily. This is done in the next section.

4.3 Classical Analysis

Following Nelson and Plosser (1982), the most popular classical unit root test has been the Dickey-Fuller t-test (see Fuller (1976)). As explained in Sub-

[4]Some care has to be taken when concluding the nonrobustness of an estimator from the unboundedness of its influence function (see Davies (1995)).

section 4.1.3, the Dickey-Fuller t-test statistic (DF-t) is usually obtained by estimating an AR model with OLS. It was argued in the previous section, however, that the OLS estimator is nonrobust to AO's. A test statistic based on this estimator might, therefore, also be nonrobust, as the following example illustrates.

Example 4.1 Consider the AR(1), $x_t = \phi x_{t-1} + \varepsilon_t$, with $t = 1, \ldots, T$. The process ε_t is i.i.d. with finite variance. Let $z_t = 0$ for all $t \neq s$ and $z_s = 1$ for some $1 < s < T$. Finally, $w_s = x_s + \zeta$. Now all variables in (4.11) are defined. The OLS estimator of ϕ, calculated with the observed series y_t, equals

$$\hat{\phi} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} = \frac{\zeta(x_{s-1} + x_{s+1}) + \sum_{t=2}^{T} x_t x_{t-1}}{\zeta^2 + 2\zeta x_s + \sum_{t=2}^{T} x_{t-1}^2} = O(\zeta^{-1}).$$

A large AO corresponds to a large value of ζ. Therefore, such an outlier causes the OLS estimator to be biased towards zero. This is easily understood by considering the bad leverage point in the lower-right corner of the scatter diagram for the AO series in Figure 4.1. As the OLS estimator takes all observations of the sample into account, this outlier causes a flatter regression line to be fitted and, thus, a smaller value of $|\hat{\phi}|$ to be estimated. It can also be shown that the standard error of the OLS estimator is $(T-1)^{-1/2}(1+o(1))$, such that the DF-t tends towards $-\sqrt{(T-1)}$ for large values of ζ. Hence, rejection of the unit root hypothesis seems likely in the case of a large AO. △

'Overrejection' of the unit root hypothesis due to large AO's is reported by Franses and Haldrup (1994) and in Chapters 5 and 6. The intuition behind this phenomenon is straightforward: an AO is, by its definition, at odds with the persistence of shocks implied by the unit root hypothesis. The consequences of IO's are less clear. It is known that the point estimates of the AR parameters obtained with the OLS estimator are not very sensitive to the occurrence of IO's, but that the variance of the OLS estimator increases rapidly if the innovations become nonnormal (see, e.g., Martin (1981) and the remarks in Bustos and Yohai (1986)). Using similar calculations as in Example 4.1, it is shown in Example 5.1 in Chapter 5 that large IO's can either cause overrejection or underrejection of the unit root hypothesis, depending on the true value of the autoregressive coefficient.

As was noted previously, the OLS estimator has an unbounded IF. Therefore, I consider an alternative estimator that has a bounded IF. One of the simplest alternatives is the conditional pseudo maximum likelihood estimator based upon the Student t distribution (MLT estimator), compare Chapter 3. As an example, consider the AR(1) model $y_t = \phi y_{t-1} + \varepsilon_t$. The MLT estimator of ϕ is defined as the value $\hat{\phi}$ that solves

$$\sum_{t=2}^{T} \frac{\hat{\varepsilon}_t/\sigma}{1 + \hat{\varepsilon}_t^2/(\sigma^2\nu)} y_{t-1} = 0, \tag{4.15}$$

where $\hat{\varepsilon}_t = y_t - \hat{\phi} y_{t-1}$, σ is a scale parameter, and ν is the degrees of freedom parameter.[5] It is easily seen that the MLT estimator for ϕ falls within the class of M estimators (see Subsection 2.3.1). It is known (see, e.g., Hampel et al. (1986)) that for regression problems in the i.i.d. setting the IF of this class of estimators can be decomposed into two parts. One part measures the influence of large residuals, while the other measures the influence of the design. Ordinary M estimators impose a bound on the former, but leave the latter untouched. Therefore, the compound IF for these estimators is usually unbounded in the i.i.d. setting. However, the following proposition shows that ordinary M estimators like the MLT estimator can have a bounded IF in the time series context.

Proposition 4.1 *Consider model (4.11); let x_t be an AR(p) process with i.i.d. innovations; let w_t be as in (4.12) and let z_t be an i.i.d. Bernoulli process with $P(z_t = 1) = \gamma$; then the MLT estimator has a bounded IF.*

A more precise statement of the proposition is given in the appendix. Here, I present a heuristic derivation of a finite sample analogue of the IF in order to illustrate the boundedness of the IF for the MLT estimator. Consider the uncontaminated AR(1) series, $x_t = \phi x_{t-1} + \varepsilon_t$, which is observed from $t = 0, \ldots, T$. Let $w_t = x_t + \xi_t$, with $\xi_t = 0$ for all $t \neq s$ and $\xi_s = \zeta$, with $1 < s < T$. In order to simplify the exposition and to avoid unnecessary complexities, assume that $x_s = x_{s-1} = \varepsilon_s = \varepsilon_{s+1} = 0$, $\sigma = 1$, and that ζ^2/T is negligible. The key quantity to look at is the difference between the MLT estimator based on the clean or outlier free sample (with $\zeta = 0$) and on the contaminated sample ($\zeta \neq 0$). Denote these estimators by $\hat{\phi}$ and $\tilde{\phi}$, respectively. Note that the contamination parameter γ of (4.14) equals T^{-1}. Define $\tilde{e}_t(\phi) = y_t - \phi y_{t-1}$ and $e_t(\phi) = x_t - \phi x_{t-1}$. It is obvious from (4.15) that $\tilde{\phi}$ solves

$$\sum_{t=1}^{T} \frac{\tilde{e}_t(\tilde{\phi})}{1 + \tilde{e}_t(\tilde{\phi})^2/\nu} y_{t-1} = 0. \tag{4.16}$$

Substituting $e_t(\hat{\phi})$ for $\tilde{e}_t(\tilde{\phi})$ and x_{t-1} for y_{t-1} in (4.16), one obtains a similar equation for $\hat{\phi}$. Taking a first order Taylor expansion of the right hand side of (4.16) around $\hat{\phi}$ and omitting higher order terms, one obtains

$$0 \approx \sum_{t=1}^{T} \frac{\tilde{e}_t(\hat{\phi})}{1 + \tilde{e}_t(\hat{\phi})^2/\nu} y_{t-1} - \left[T^{-1} \sum_{t=1}^{T} \frac{1 - \tilde{e}_t(\hat{\phi})^2/\nu}{(1 + \tilde{e}_t(\hat{\phi})^2/\nu)^2} y_{t-1}^2 \right] T(\tilde{\phi} - \hat{\phi}). \tag{4.17}$$

Next, notice that $\tilde{e}_t(\hat{\phi}) = e_t(\hat{\phi})$ for all $t \neq s, s+1$. Denote the factor between square brackets in (4.17) by $I_T(\hat{\phi})$. Using the analogue of (4.16) for $\hat{\phi}$ and the

[5] The parameter σ is always estimated in the present chapter. The parameter ν, in contrast, can be either fixed or estimated. If ν is estimated from the data, this is denoted by using $\hat{\nu}$ rather than ν.

values of x_t and ε_t for t near s, (4.17) can be rewritten as

$$\begin{aligned} T(\tilde{\phi}-\hat{\phi}) &\approx (I_T(\hat{\phi}))^{-1}\sum_{t=1}^{T}\frac{\tilde{e}_t(\hat{\phi})}{1+\tilde{e}_t(\hat{\phi})^2/\nu}y_{t-1} \\ &= (I_T(\hat{\phi}))^{-1}\left[\frac{\tilde{e}_{s+1}(\hat{\phi})}{1+\tilde{e}_{s+1}(\hat{\phi})^2/\nu}y_s+\sum_{t=1}^{T}\frac{e_t(\hat{\phi})}{1+e_t(\hat{\phi})^2/\nu}x_{t-1}\right] \\ &= (I_T(\hat{\phi}))^{-1}\frac{-\hat{\phi}\zeta^2}{1+\hat{\phi}^2\zeta^2/\nu}. \end{aligned} \tag{4.18}$$

Note that this function is bounded in ζ. The OLS estimator is obtained by letting $\nu \to \infty$. It is easily seen that in that case (4.18) becomes unbounded in ζ. A more formal statement of these results is given in the appendix.

It follows from the proposition that the IF of the MLT estimator is bounded for a wide variety of uncontaminated c.d.f.'s F_y^0. This suggests that the MLT estimator can still provide useful information about the true value of ϕ if the true c.d.f. is in some neighborhood of the assumed Student t distribution with ν degrees of freedom. A necessary condition for the boundedness of an M estimator in a time series context is that the estimator is defined by a weakly redescending ψ function.[6] More specifically, $\psi(\varepsilon_t)$ must be $O(\varepsilon_t^{-1})$ for large values of ε_t. In this sense, the MLT estimator forms a borderline case, because for this estimator $\psi(\varepsilon_t) = (\varepsilon_t/\sigma)/(1+\varepsilon_t^2/\nu\sigma^2) = O(\varepsilon_t^{-1})$ for large ε_t. The finding that ordinary M estimators can have a bounded IF in the time series context was, to my knowledge, not explicitly noted earlier in the literature. A few examples of IF's of MLT estimators for ϕ in the model $y_t = \phi y_{t-1} + \varepsilon_t$, with ε_t standard normal white noise, are given in Figure 4.2.

The bounded IF property also holds for the general AR(p). Incorporating deterministic functions of time, like a linear time trend or a constant, causes no special problems. As long as they are generated correctly, they will not be outlying in the space of explanatory (or predetermined) variables. Therefore the use of an ordinary M estimator is enough for dealing with outliers.

So far, the scale parameter σ was assumed to be known. Usually, it has to be estimated along with the autoregressive parameters ϕ_i, $i = 1, \ldots, p$. This can be done using the first order conditions of the MLT estimator for both ϕ and σ. Such an estimator for σ has a bounded IF, because its corresponding ψ function is bounded and only depends upon the true process through $y_1 - \phi y_0$ (see Chapter 3). There may be some problems, however, with the number of outliers such a simultaneous estimation procedure can cope with (see Maronna and Yohai (1991)).

A harder problem has to be faced if one wants to estimate the degrees of freedom parameter ν simultaneously. Simply using the first order conditions of the MLT estimator for ϕ, σ, and ν results in an unbounded IF for both $\hat{\sigma}$ and $\hat{\nu}$ for all finite values of ν (see Chapter 3). Developing a bounded influence

[6]The function $\psi(\cdot)$ is called weakly redescending if $\lim_{x\to\pm\infty}\psi(x) = 0$. It is called strongly redescending if for some positive constant c it holds that $\psi(x) = 0$ for all $|x| \geq c$.

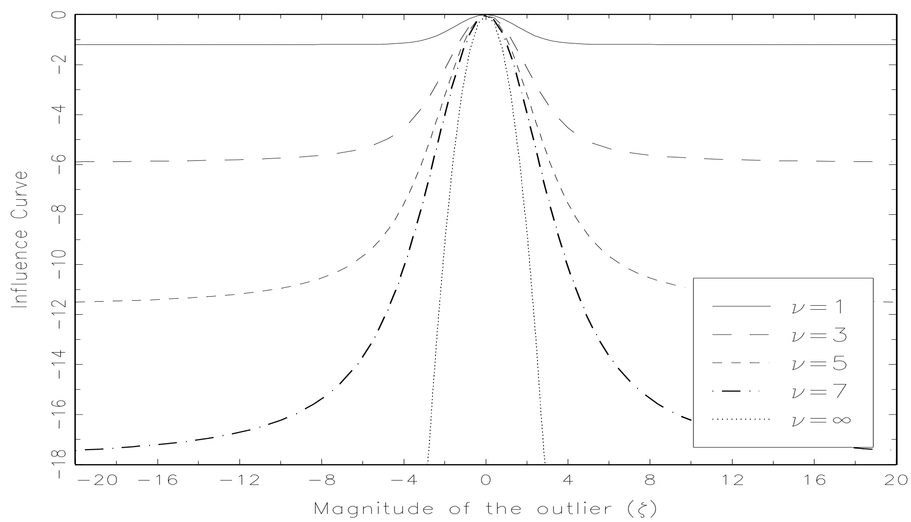

Figure 4.2.— Some Influence Functions of Several MLT estimators Evaluated at a Gaussian AR(1) Process ($\phi = 0.8$)

estimator for ν that is consistent for the family of Student t distributions is difficult. It was noted previously, however, that the MLT estimator belongs to the class of M estimators. Most of these estimators make use of a tuning constant, which is fixed at a value prespecified by the user. The parameter ν can also be treated as such a tuning constant. In this case, the user does not believe that the innovations are actually drawn from the prespecified Student t distribution, but (s)he only uses the first order conditions of the pseudo likelihood under this distribution in order to obtain a certain degree of robustness. The efficiency loss caused by fixing ν if the sample is actually driven by Student t distributed innovations with a different degrees of freedom parameter, can be kept within bounds. Some tradeoff has to be made, however, between efficiency and robustness (compare Hampel et al. (1986, p. 44)). Similar arguments are encountered in the literature on pseudo maximum likelihood estimators (see Gouriéroux et al. (1984)), a class of estimators that also comprises the MLT estimator.

Another point is that for $\phi = 1$ the IF of both the OLS estimator and the MLT estimator are identically equal to zero. This can be seen from the formulas in the appendix. There the scalar C, which is closely related to the second unconditional moment of y_t, diverges towards infinity if ϕ approaches unity. As a result, one might think that for integrated processes there is no need to use the more complicated MLT estimator in order to obtain robustness. Two things can be said about this. First, the MLT estimator for a fixed value of the degrees of freedom parameter ν is not more difficult to compute than the OLS estimator. It can be obtained by using an iterative weighted least-squares algorithm (Prucha and Kelejian (1984)). Second, under the alternative hypothesis of stationarity, the MLT estimator is more robust than its OLS counterpart according to proposition 4.1. Moreover, power can be gained by using the MLT methodology if the innovations are leptokurtic, see Chapters 6 and 7.

Upon closer inspection of (4.12), the AO model can also be regarded as a measurement error model, with x_t the clean process and $z_t\xi_t$ the measurement error. Therefore, under this type of contamination, y_t in fact follows an ARMA instead of a pure AR process. Consequently, the nonrobustness of the OLS based DF-t might be repaired by constructing a test that takes into account the temporal dependence of the disturbances that drive the times series. These tests can be found in Phillips (1987) and Phillips and Perron (1988). The results in Chapter 5, however, show that the approach of Phillips and Perron is not outlier robust in finite samples, as opposed to the results based on robust estimation procedures.

A practical problem with the use of different estimators for testing the unit root hypothesis is that each time new critical values have to be tabulated. This also holds for the MLT estimator used in this chapter. I used a similar simulation setup as in Fuller (1976). The asymptotic distribution of the DF-t based on the MLT estimator can be found in Chapters 5 and 6. I generated x_t from model (4.11) as a random walk of length 50, 100, or 200 with i.i.d. standard Gaussian innovations. First, I let $w_t \equiv x_t$ and estimated the regression models $y_t = \phi y_{t-1} + \varepsilon_t$, $y_t = \alpha + \phi y_{t-1} + \varepsilon_t$, and $y_t = \alpha + \beta t + \phi y_{t-1} + \varepsilon_t$ using the MLT estimator. Several values for the degrees of freedom parameter ν were used. The DF-t for each of these models was calculated and this process repeated 1,000 times. The 50th order statistic of the simulated DF-t values was used as an estimate of the 5% critical value. The standard error of $\hat{\phi}$ for the model without constant and trend was estimated by

$$\hat{\sigma}^2 \left[\sum_{t=2}^{T} y_{t-1}^2 \psi'(\hat{\varepsilon}_t/\hat{\sigma}) \right]^{-2} \left[\sum_{t=2}^{T} y_{t-1}^2 \psi(\hat{\varepsilon}_t/\hat{\sigma})^2 \right], \tag{4.19}$$

with $\psi'(x) = d\psi(x)/dx$, $\hat{\varepsilon}_t = y_t - \hat{\phi} y_{t-1}$, and $\hat{\sigma}$ the estimate of the scale of ε_t (compare Hampel et al. (1986, p. 316)). Similar formulas were used for the other two regression models. For $\hat{\sigma}$ the pseudo maximum likelihood estimator under the Student t distribution was used. The same value of ν was used for computing $\hat{\phi}$ and $\hat{\sigma}$. Note that (4.19) is in fact a kind of heteroskedasticity consistent covariance estimator as in White (1980). This causes a discrepancy between the critical values tabulated by Fuller and the ones supplied here for $\nu = \infty$. As noted by simulations in Chapter 6, the use of heteroskedasticity consistent standard errors helps to make the standard DF-t more robust. For completeness, I report the results for the standard DF-t *with* ($\nu = \omega$) and *without* ($\nu = \infty$) the heteroskedasticity correction for the standard errors. The critical values for the Gaussian random walk are presented in Table 4.1 under the heading 'clean.'

TABLE 4.1
5% Critical Values for MLT Based Dickey-Fuller Tests

ν	n=50		n=100		n=200	
	clean	outliers	clean	outliers	clean	outliers
			none			
1	-2.891	-2.918	-2.415	-2.596	-2.216	-2.321
2	-2.353	-2.554	-2.176	-2.441	-2.091	-2.345
3	-2.270	-2.447	-2.054	-2.485	-2.040	-2.477
4	-2.273	-2.414	-2.065	-2.469	-2.026	-2.571
5	-2.184	-2.441	-2.064	-2.486	-2.045	-2.615
7	-2.116	-2.458	-2.081	-2.494	-2.069	-2.685
10	-2.098	-2.471	-2.081	-2.497	-2.051	-2.752
ω	-2.047	-2.749	-2.086	-2.736	-2.009	-2.818
∞	-1.885	-3.517	-2.034	-3.990	-2.001	-3.863
			constant			
1	-4.009	-4.370	-3.545	-3.764	-3.117	-3.233
2	-3.458	-3.642	-3.262	-3.447	-3.004	-3.305
3	-3.429	-3.393	-3.097	-3.358	-2.974	-3.395
4	-3.312	-3.404	-3.061	-3.315	-2.932	-3.398
5	-3.254	-3.449	-3.060	-3.304	-2.920	-3.425
7	-3.279	-3.461	-3.076	-3.226	-2.931	-3.442
10	-3.224	-3.532	-3.082	-3.189	-2.912	-3.446
ω	-3.297	-3.782	-3.046	-3.475	-2.915	-3.373
∞	-2.920	-4.566	-2.856	-5.058	-2.806	-4.859
			trend			
1	-5.498	-6.713	-4.280	-4.242	-3.661	-3.840
2	-4.420	-5.119	-3.760	-3.868	-3.498	-3.939
3	-4.081	-4.636	-3.661	-3.864	-3.502	-3.940
4	-3.973	-4.695	-3.663	-3.850	-3.443	-3.935
5	-3.944	-4.676	-3.659	-3.868	-3.439	-3.982
7	-3.807	-4.898	-3.664	-3.786	-3.466	-3.975
10	-3.743	-5.033	-3.684	-3.777	-3.453	-3.958
ω	-3.836	-5.425	-3.742	-4.204	-3.450	-3.988
∞	-3.438	-5.586	-3.527	-5.885	-3.349	-6.238

The 5% critical values under the heading 'clean' are based on simulations that use a random walk without standard Gaussian innovations. For the entries under the heading 'outliers', the simulations are based on a random walk with 5% additive outiers. The outliers are generated by adding drawings from a normal with zero mean and standard deviation 5 to a randomly chosen subset of 5% of the original observations. The headings 'none', 'constant' and 'trend' refer to the deterministic components that are incorporated in the regression model. The degrees of freedom parameter ν is used as a tuning constant in this table and is not estimated from the data.

In order to illustrate the robustness aspects of the tests, I performed similar simulations with z_t from (4.11) equal to an i.i.d. process and γ equal to 0.05. For the ξ_t process (4.12), I considered a sequence of Gaussian i.i.d. random variables with mean zero and standard deviation five. The results of these simulations can be found in Table 4.1 under the heading 'outliers.'

The simulation experiments lead to the following four conclusions.

First, the absolute difference between the 'clean' critical value and the 'outlier' critical value is increasing in ν, the tuning constant of the MLT estimator. This is to be expected, because ν determines the degree of robustness of the estimator.

As a second conclusion from the simulations, one finds that the MLT based tests demonstrate a slower convergence to the asymptotic distribution. This can be seen by looking at the behavior of the 'clean' critical values for varying sample sizes for a specific choice of ν. For example, for the case with a trend the difference between the 'clean' critical value for $n = 50$ and $n = 200$ is much larger for $\nu = 1$ than for $\nu = 10$. The different behavior for $\nu = 1$ and $\nu = \infty$ in Table 4.1 is considerable. Therefore, it seems necessary to use the finite sample critical values from Table 4.1 for MLT based unit root tests in small samples. Also note that setting $\nu = 1$ in samples of this size does not always yield the maximum protection against outliers. The convergence behavior and robustness properties are further illustrated in Figure 4.3, which presents the c.d.f.'s of the unit root tests for two values of ν. Especially for the regression model with trend the figure reveals that the c.d.f. can change considerably if the sample is enlarged. Moreover, if outliers are added to the sample, the c.d.f. of the test based on the OLS estimator ($\nu = \infty$) shifts more to the left than the one based on the MLT estimator with $\nu = 3$.

Third, in addition to the results reported in Table 4.1, unreported simulations were performed using random walks with Student t instead of Gaussian innovations. The results of using fat-tailed innovations are that the critical values for all of the tests shift somewhat to the right, thus decreasing the type I error of the tests if the critical values of Table 4.1 are used.[7] The robustness and convergence properties of the tests remain similar.

As a fourth conclusion from the simulations, it turns out that using a different distribution for generating the AO's does not alter the results. I used the Cauchy and the symmetric delta distribution. The latter generates the values 5 and -5 with equal probability. The Cauchy AO's result in a very large shift to the left of the 5% critical values of the OLS based DF-t, e.g., from -3.45 to -8.46 for $\nu = \omega$, $T = 200$, and a regression model with trend. The critical values for the MLT based tests with $\nu = 3$ remain remarkably stable in the same setting and only shift from, e.g., -3.50 to -3.95. Similar conclusions can be drawn from the remaining experiments.

It is illustrative to consider an example of how Table 4.1 can be used in

[7] The change is considerable if Cauchy innovations are used. This is in accordance with the findings of Knight (1991), who derives that for a certain class of infinite variance innovations the DF-t is asymptotically normally distributed, even if it is based on an M estimator.

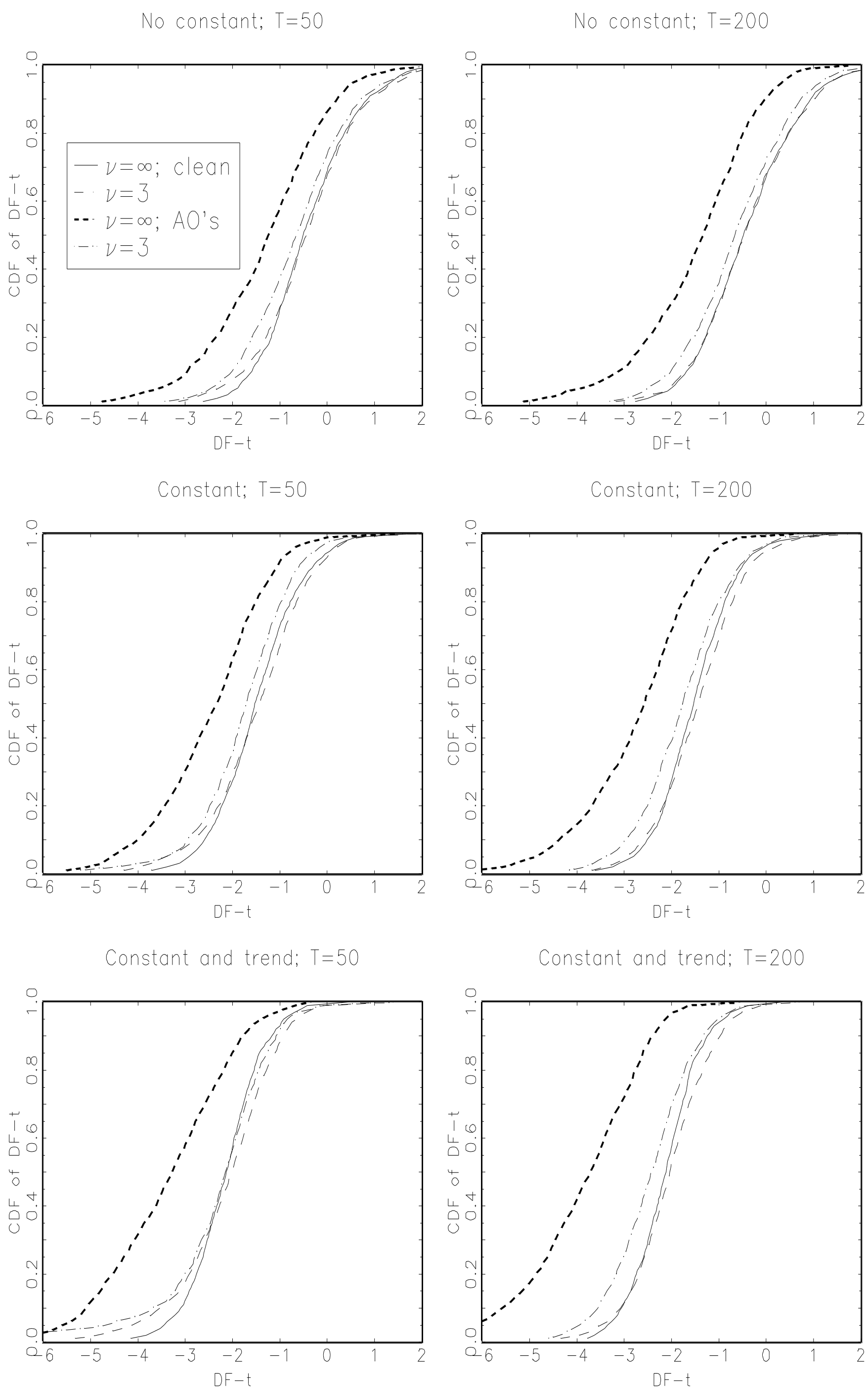

Figure 4.3.— C.D.F.'s of the DF-t Based on the OLS and MLT Estimator

practice. Consider the case $T = 100$ and a regression model with trend. If the standard DF-t is used at the 5% level, the appropriate critical value is -3.527. For the DF-t with heteroskedasticity consistent standard errors ($\nu = \omega$), the appropriate value is -3.742, while for the MLT based test with $\nu = 3$, it equals -3.661. These values are to be used by the applied researcher. Now consider the effect of additive outliers on each of these three tests. If one uses the outlier generating scheme described below Table 4.1, the actual critical value for a 5% level test with the ordinary DF-t for a sample with outliers is -5.885. However, the researcher is unaware of this exact value, because (s)he, in general, is ignorant about the exact model that generated the outliers. Therefore, (s)he continues to use the critical value -3.527, which for the present data generating process gives a 51% level test. Similar arguments for the other two tests lead to the result that these tests have a size of 10% and 7.5%, respectively, for the data generating process with 5% AO's. The size distortion of the MLT based test is the smallest. Note, however, that using heteroskedasticity consistent standard errors also helps to reduce the sensitivity of the size of the standard DF-t to outliers. This finding is also noted in Chapter 6.

4.4 A Bayesian Analysis

Recent years have seen a growing number of Bayesian studies on the possible presence of a unit root in macroeconomic time series (see, e.g., Sims (1988), DeJong and Whiteman (1991a,b), Phillips (1991b), Schotman and van Dijk (1991a,b, 1993), and Kleibergen and van Dijk (1993)). The focus of these studies is on the specification of a prior distribution: posterior inference should, to a certain extent, be insensitive to the choice of the prior. It is this kind of robustness that is usually studied in the Bayesian literature. Robustness with respect to 'irregularities' in the data has received less attention. It is easily shown, however, that also in a Bayesian framework AO's can seriously affect unit root inference.

Example 4.2 Consider the same processes x_t, w_t and z_t as in Example 4.1. Define the parameter vector $\theta = (\phi, \sigma^2)$. Assume a diffuse prior $\pi(\theta) \propto \sigma^{-1}$ and a Gaussian likelihood. The marginal posterior of ϕ is a Student t density (see Judge et al. (1988, Section 7.4.4)):

$$p(\phi|y_1, \ldots, y_T) = t\left(\phi; \hat{\phi}_{OLS}, \hat{\sigma}^2_{OLS}\left[\sum (y_{t-1})^2\right]^{-1}, T-1\right),$$

where $\hat{\phi}_{OLS}$ and $\hat{\sigma}^2_{OLS}$ are the OLS estimates of ϕ and σ^2, respectively. Furthermore, $t(\cdot; \mu, \Sigma, \nu)$ denotes the density function of the Student t distribution with location μ, precision matrix Σ^{-1}, and degrees of freedom parameter ν. Looking at the extreme case of an infinitely large AO ($\zeta \to \infty$), one obtains from the previous section that

$$p(\phi|y_1, \ldots, y_T) = t(\phi; 0, (T-1)^{-1}, T-1).$$

Therefore, a sufficiently large additive outlier pushes the posterior away from the unit root, even if the data is a random walk. △

Assuming normality in the case of (additive) outliers in the data can be regarded as a misspecification of the likelihood. Therefore, robustness with respect to outliers can be linked to the issue of insensitivity of posterior results to misspecification of the likelihood function (see Berger (1985, Section 4.7)). To model outliers, fat-tailed distributions have been suggested in the Bayesian literature as well. For example, Leamer (1978, Section 8.2), Smith (1981), and West (1984) mention the use of the independent Student t to 'robustify' the posterior results. This robustness property of the Bayesian posterior results that are based upon the Student t likelihood can be explained by the close link between the posterior and the likelihood, in particular if 'data dominated' priors are employed. As argued in the previous section, maximum likelihood results based on a Student t likelihood possess certain robustness properties. Intuitively, these properties are passed on to the results obtained from a Bayesian posterior analysis that uses a Student t likelihood. Analogously, Example 4.2 shows that the nonrobustness of OLS estimators is 'inherited' by Bayesian inference procedures that use a Gaussian likelihood. In order to evaluate the robustness properties in a Bayesian framework with the i.i.d. Student t distribution, the concept of the (posterior) score function has been used by Smith (1981, Section 5) and West (1984). In this chapter, I only analyze the effect of the Student t on unit root inference.

Geweke (1995) also presents a Bayesian analysis of the Student t linear model. Concentrating on computational issues, in particular the implementation of Gibbs sampling techniques to compute posterior results, he shows that the assumption of an i.i.d. Student t distribution is equivalent to the introduction of a certain type of heteroskedasticity. Hence, outlying observations (large residuals) are weighted less heavily. The ability of the Student t to model heteroskedasticity of the GARCH type is also discussed by Kleibergen and van Dijk (1993). All these arguments motivate the choice of an i.i.d. Student t likelihood[8] instead of a Gaussian one.

The likelihood function for the general linear model $y_t = x_t'\beta + \varepsilon_t$, where $\{\varepsilon_t\}$ is an Student t i.i.d. process, is proportional to

$$\ell(\beta, \sigma^2, \nu | y) \propto \prod_{t=1}^{T} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)(\nu\sigma^2)^{1/2}} \left[1 + \frac{(y_t - x_t'\beta)^2}{\nu\sigma^2} \right]^{-(\nu+1)/2}, \tag{4.20}$$

In the present context of unit root testing, x_t contains a constant (and possibly a time trend) as well as lagged endogenous variables. Note that in time series models, one usually conditions on some fixed initial values $y_0, \ldots, y_{-p+1}$.

[8]Zellner (1976) analyzes the case of a multivariate Student t, where the innovations are uncorrelated, but not independent. In a flat prior Bayesian framework, an identification problem arises. As Zellner shows, specification of a flat prior for the degrees of freedom parameter ν leads to a flat marginal posterior on that parameter. Also, the posterior for the location parameters is identical to the posterior under a Gaussian likelihood.

In order to examine the effect of the i.i.d. Student t likelihood on unit root inference, I consider the following parameterization of the AR(1) model with unknown mean: $y_t = \mu + \phi y_{t-1} + \varepsilon_t$. The generalization to the AR(p) model is straightforward. I assume a flat prior for the parameters of the model, including ν,

$$\pi(\mu, \phi, \sigma^2, \nu) \propto \sigma^{-1}. \tag{4.21}$$

An improper uniform prior for ν essentially imposes normality, because the prior odds in favor of normality are infinite (see Geweke (1995)).[9] Geweke, therefore, proposes the use of an exponential prior for ν with parameter λ. The sensitivity of the results with respect to the choice of λ can, of course, be examined by varying this parameter. The computations can be performed using the Gibbs sampler if one exploits the equivalence of a heteroskedastic Gaussian linear model with a homoskedastic Student t linear model. The exponential priors considered by Geweke give more prior weight to low values of ν. As a result, the hypothesis of a Student t distribution is a priori more likely than the hypothesis of normality.

In this chapter, a different procedure is followed. The difference between the likelihood specified in (4.20) and the Gaussian likelihood is only substantial for low values of ν. If ν tends to infinity, the difference disappears. Therefore, a uniform prior is specified for ν on the interval $(0, \nu^*]$, where ν^* is such that the Student t distribution with this degrees of freedom parameter is 'sufficiently' close to the normal. One obtains

$$\pi(\nu) = 1/\nu^* \text{ for } 0 < \nu \leq \nu^*, \ 0 \text{ elsewhere.} \tag{4.22}$$

A sensitivity analysis can be performed by considering different values of ν^*. Note that it was argued in Section 4.3 that simultaneously estimating β, σ^2, and ν results in an unbounded influence function for the estimators for σ^2 and ν. A priori imposing bounds on ν does not solve this problem.

The posterior resulting from (4.20) through (4.22) is difficult to handle analytically. In order to perform a posterior analysis, importance sampling techniques are applied using the SISAM program (see Hop and van Dijk (1992)). A multivariate Student t density was used as the importance function. Given the posterior densities, one may test for the presence of a unit root, $\phi = 1$. Also the (approximate) normality assumption, $\nu = \nu^*$ versus $\nu < \nu^*$, can be tested. Some empirical results are presented in the next section.

4.5 Empirical Illustration

To illustrate the use of the i.i.d. Student t for classical and Bayesian unit root testing procedures, several time series are analyzed. First, the Finland/US real exchange rate is studied (see Perron and Vogelsang (1992) and Franses and Haldrup (1994)). This series is obtained by deflating the nominal exchange rate by a consumer price index. As Figure 4.4 clearly shows, this series

[9] Normality is equivalent to $\nu = \infty$.

is characterized by the presence of outliers. The outliers appear to be additive: the series almost immediately returns to the 'normal' pattern. Following Franses and Haldrup, the model for this series is

$$y_t = \mu + \rho y_{t-1} + \phi_1 \Delta y_{t-1} + \varepsilon_t, \tag{4.23}$$

where ρ denotes the unit root parameter, defined as the sum of the original AR coefficients (compare (4.9)).

Next, a well known series from the marketing literature is considered, namely the Lydia Pinkham annual advertising series (see Figure 4.5). The first difference of this series contains several (additive) outliers, in particular in the middle of the sample. These additive outliers in the first differences correspond to innovative outliers in the level of the series. Helmer and Johansson (1977) specified an AR(2) model for the first differences of the series, implying a unit root in the level. To test for this unit root, I consider the model

$$y_t = \mu + \rho y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \varepsilon_t. \tag{4.24}$$

Finally, as in Schotman and van Dijk (1991b), the extended Nelson-Plosser data are considered (see also Chapter 5). To study these data, an AR(3) model with linear time trend is used:

$$y_t = \mu + \delta t + \rho y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \varepsilon_t. \tag{4.25}$$

The AR(3) model is the autoregressive model of the lowest order that can generate a stochastic trend and a cycle, simultaneously.

All three models are first estimated using the MLT estimator. Table 4.2 presents the results for several values of ν.

With the exception of the employment series, velocity, and S&P 500, the estimates of ρ increase as ν decreases. This indicates a negative correlation between the maximum likelihood estimators for ρ and ν. The most remarkable result is obtained for the Finland/US real exchange rate. Assuming normality gives an estimate of 0.49 for the unit root parameter, while under a Student t distribution with one degree of freedom, this estimate equals 0.88. Note the sharp decline in the estimated standard deviation of this estimate, a result that is also obtained for the Lydia Pinkham advertising series. For the extended Nelson-Plosser series this relation is less clear.

The entries in Table 4.2 can also be used to compute the DF-t statistic $((\tilde{\rho} - 1)/\tilde{s}_\rho)$. For the Finland/US real exchange rate series, for example, this statistic moves from -5.74 ($\nu = \infty$) to -3.21 ($\nu = \omega$), and finally, to -2.00 ($\nu = 1$). Using Table 4.1 with n = 100, the first t-statistic is significant at the 5% level using both the 'clean' and 'outliers' critical values (respectively -2.856 and -5.058). The second statistic, using the heteroskedasticity consistent standard errors, is significant when using the 'clean' critical value (-3.046), but insignificant using the 'outliers' critical value (-3.475). Finally, the third, more robust statistic is insignificant under both processes, the critical values being -3.545 and -3.764, respectively. This example shows that taking account of

TABLE 4.2
MLT Estimates of ρ with ν Degrees of Freedom

Series	$\nu = \omega$	$\nu = 10$	$\nu = 5$	$\nu = 3$	$\nu = 1$	$\hat{\nu}$
Finland/US	0.489 (0.159)	0.505 (0.244)	0.630 (0.216)	0.743 (0.163)	0.876 (0.062)	1.23
Advertising	0.856 (0.115)	0.941 (0.145)	1.018 (0.108)	1.066 (0.072)	1.120 (0.053)	1.81
Real GNP	0.813 (0.055)	0.813 (0.055)	0.815 (0.058)	0.821 (0.067)	0.893 (0.070)	3.62
Nominal GNP	0.944 (0.039)	0.960 (0.031)	0.965 (0.030)	0.969 (0.030)	0.972 (0.035)	2.42
Real GNP per capita	0.803 (0.056)	0.802 (0.055)	0.806 (0.059)	0.814 (0.071)	0.879 (0.046)	3.55
Industrial production	0.826 (0.055)	0.829 (0.052)	0.840 (0.052)	0.852 (0.054)	0.885 (0.068)	3.93
Employment	0.864 (0.049)	0.860 (0.049)	0.861 (0.046)	0.866 (0.042)	0.863 (0.039)	2.51
Unemployment	0.744 (0.066)	0.779 (0.064)	0.801 (0.063)	0.821 (0.065)	0.885 (0.048)	3.47
GNP deflator	0.966 (0.025)	0.985 (0.016)	0.989 (0.015)	0.993 (0.015)	0.999 (0.012)	2.38
Consumer price index	0.994 (0.010)	0.994 (0.009)	0.994 (0.008)	0.995 (0.008)	1.001 (0.006)	1.73
Wages	0.939 (0.032)	0.941 (0.028)	0.943 (0.025)	0.947 (0.024)	0.959 (0.020)	1.74
Real wages	0.935 (0.040)	0.947 (0.041)	0.957 (0.041)	0.967 (0.041)	0.987 (0.043)	∞^*
Money	0.941 (0.024)	0.949 (0.022)	0.952 (0.022)	0.953 (0.022)	0.960 (0.039)	3.35
Velocity	0.968 (0.025)	0.963 (0.026)	0.959 (0.027)	0.955 (0.027)	0.947 (0.030)	2.73
Interest	0.953 (0.053)	0.977 (0.067)	0.992 (0.073)	0.998 (0.077)	0.996 (0.017)	1.42
S&P 500	0.932 (0.032)	0.936 (0.034)	0.935 (0.035)	0.930 (0.037)	0.897 (0.040)	7.01

The first five columns report MLT estimates of the unit root parameter ρ using a Student t likelihood with ν degrees of freedom. The heading $\nu = \omega$ denotes the Gaussian pseudo maximum likelihood estimator. Heteroskedasticity consistent standard errors are given between parentheses. The final column gives the ML estimate of the degrees of freedom parameter ν. *: ν was in fact estimated using a grid search method. For the real wage series, the estimate of ν was at its upper bound, which was equal to some large number.

outliers in the series can lead to nonrejection of the, otherwise rejected, unit root hypothesis. Finally, except for the real wage series, the MLT estimates of ν are all relatively small, providing some evidence against the assumption of Gaussian i.i.d. innovations.

The results for the Finland/US real exchange rate series correspond to the results of Franses and Haldrup (1994). By using dummy variables, their estimate of ρ increases from 0.49 to 0.81. The corresponding DF-t statistics increase from -5.74 to -2.65. The inclusion of dummy variables, however, requires pretesting for the presence and the location of outliers. The present approach does not need such a first round and may, therefore, be easier to implement.

The Bayesian posterior results,[10] reported in Table 4.3, are comparable to the classical results. A negative correlation between ρ and ν is found for the Finland/US real exchange rate series and, even stronger, for the Lydia Pinkham advertising series. This is clearly demonstrated in the contour plots in Figures 4.4 and 4.5. Also, for both series the posterior for ρ shifts to the right when the restriction $\nu = \infty$ is dropped. The case against i.i.d. normality is strongest for the exchange rate series: all posterior mass for ν is concentrated on the interval (0,4). For the advertising series, the posterior has a mode near $\nu = 2$, but it is skewed to the right.

The results for the extended Nelson-Plosser series are less clear. For half of the series the marginal posterior for ρ changes only slightly when the assumption of normality is dropped. Take as an example the Real GNP series. Some marginal posteriors for this series are plotted in Figure 4.6. The effect of dropping the normality assumption on the posteriors for ρ and δ is very small. The posterior density of ν has a mode around 4, but is skewed to the right. Both the contour plot of the bivariate posterior and the entry in the final column of Table 4.3 give no indication of a substantial correlation between ρ and ν for Real GNP. For the interest rate series, shown in Figure 4.7, the posterior for ρ clearly shifts to the right. Moreover, this posterior is somewhat less concentrated than the 'normal' posterior. The posterior for δ shifts to the left. The correlation between ν and ρ, given in Table 4.3, is negative and the posterior for ν is concentrated on the interval (0,5). Negligible posterior weight is given to values of ν exceeding 5, providing strong evidence against the assumption of i.i.d. normal innovations. With the exception of the velocity series, all Nelson-Plosser series reveal either a negligible or a negative correlation between ρ and ν.

The Nelson-Plosser series are also analyzed in Chapter 5 and in a paper of Geweke (1995). In Chapter 5, using the MM estimator of Yohai (1987), mixed results are obtained. The unit root hypothesis is rejected for 4 of the 14 series. An explanation of the difference between these results and the ones obtained here is as follows. Some of the Nelson-Plosser series are characterized by the presence of a structural break (see, e.g., Perron (1989) and Zivot and Phillips (1991)). The MM estimator can cope with a large number of outliers.

[10] The empirical application uses an upper bound of $\nu^* = 20$ for the prior on ν.

TABLE 4.3
Posterior Results for $\nu^* = 20$

Series	$E_N(\rho)$	$E_t(\rho)$	$E_t(\nu)$	$R_{\rho,\nu}$
Finland/US	0.489 (0.090)	0.829 (0.085)	1.52 (0.45)	-0.39
Advertising	0.856 (0.089)	0.993 (0.116)	6.98 (5.65)	-0.53
Real GNP	0.814 (0.056)	0.813 (0.058)	7.97 (4.89)	-0.03
Nominal GNP	0.944 (0.032)	0.966 (0.031)	3.74 (2.50)	-0.11
Real GNP per capita	0.803 (0.058)	0.802 (0.066)	8.04 (5.06)	-0.03
Industrial production	0.826 (0.053)	0.838 (0.058)	7.22 (4.36)	-0.14
Employment	0.864 (0.048)	0.862 (0.048)	6.95 (5.00)	-0.04
Unemployment	0.749 (0.071)	0.798 (0.079)	6.90 (4.58)	-0.22
GNP deflator	0.966 (0.021)	0.991 (0.017)	4.81 (4.35)	-0.15
Consumer price index	0.994 (0.011)	0.995 (0.008)	3.51 (3.18)	-0.10
Wages	0.939 (0.029)	0.946 (0.020)	12.7 (6.87)	-0.12
Real wages	0.935 (0.045)	0.945 (0.051)	12.1 (4.86)	-0.11
Money	0.941 (0.024)	0.951 (0.024)	7.03 (4.76)	-0.10
Velocity	0.968 (0.025)	0.960 (0.027)	7.10 (5.29)	0.14
Interest	0.953 (0.035)	1.040 (0.048)	1.80 (0.63)	-0.17
S&P 500	0.932 (0.036)	0.935 (0.040)	7.97 (4.89)	0.01

E(ρ) and E(ν) are, respectively, the posterior expectation of the unit root parameter and the degrees of freedom parameter. Posterior standard errors are between parentheses. The posterior correlation between ρ and ν is given in the column labeled $R_{\rho,\nu}$. Subindices 'N' and 't' denote results based on the Normal ($\nu = \infty$) and the Student t ($\nu \leq 20$) likelihood, respectively.

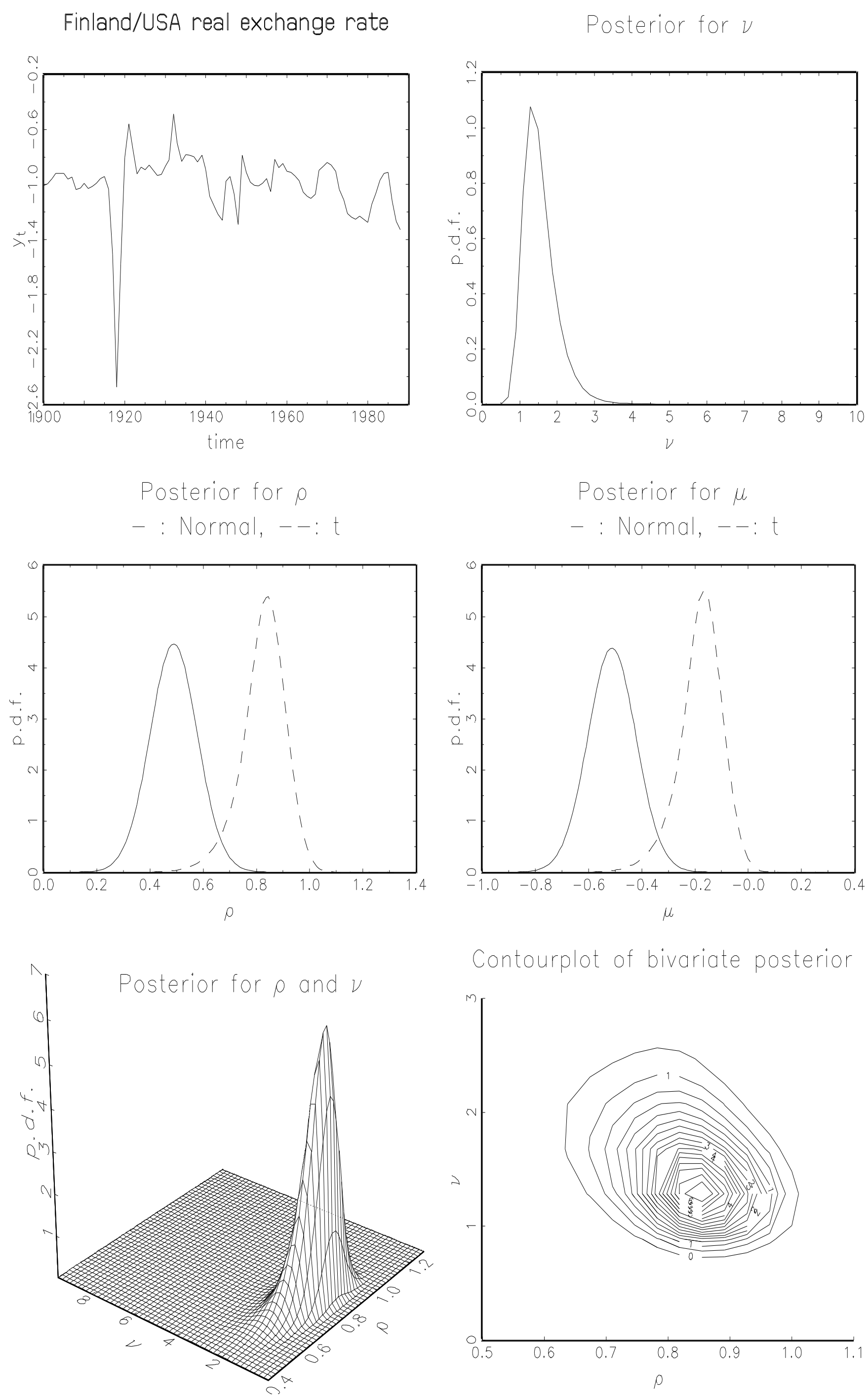

Figure 4.4.— Finland/US real exchange rate: posterior results, linear model

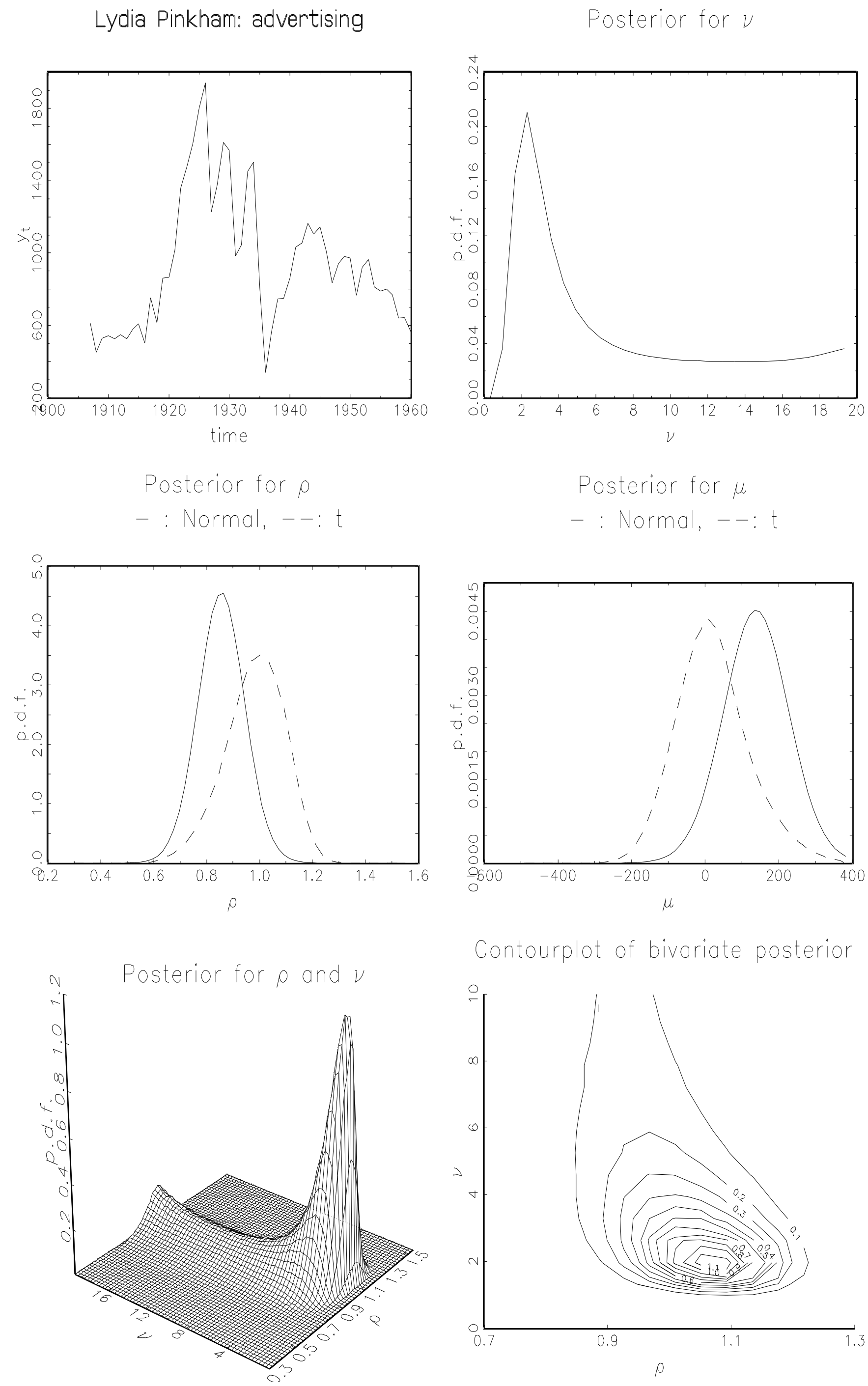

Figure 4.5.— Lydia Pinkham advertising: posterior results, linear model

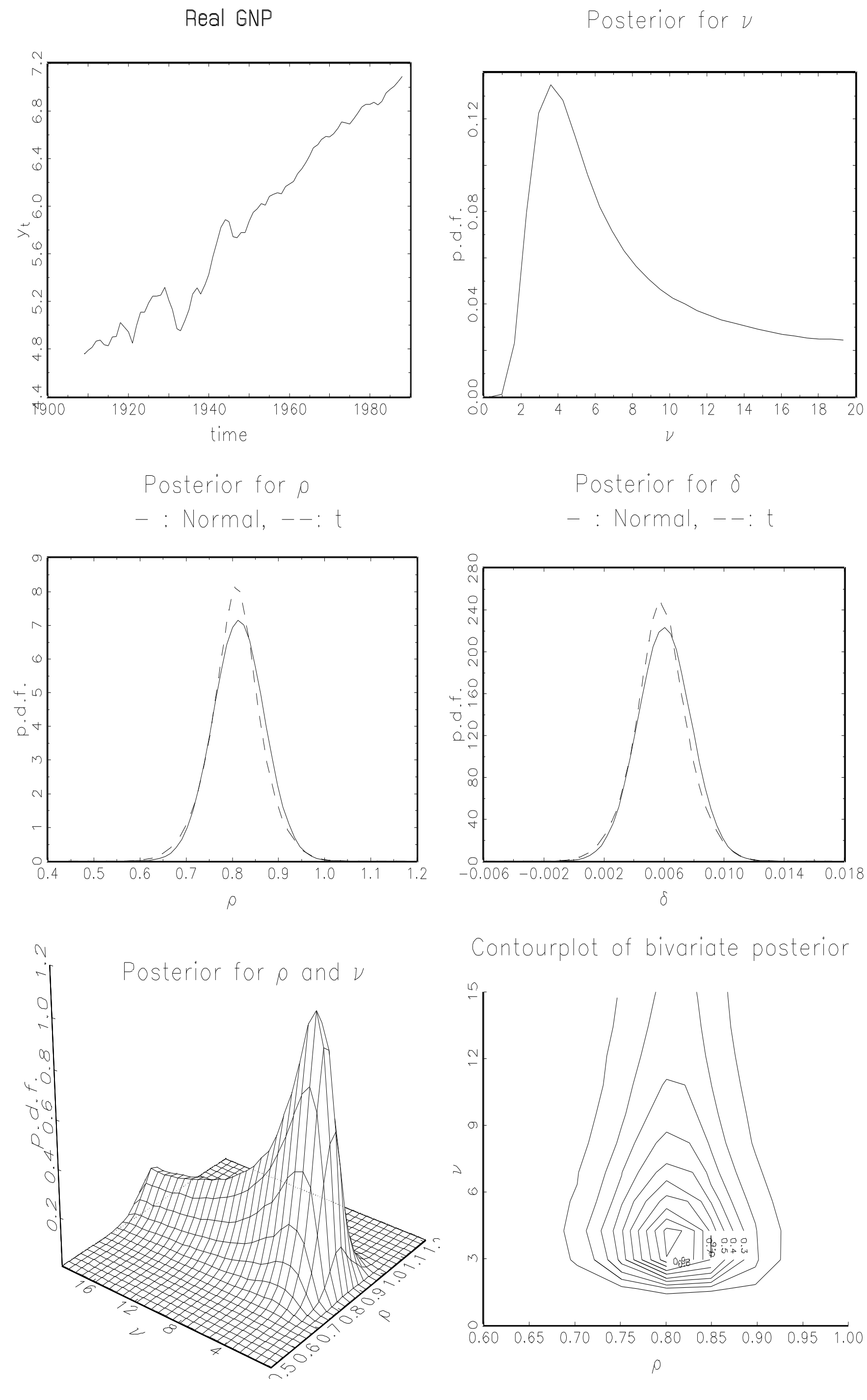

Figure 4.6.— Real GNP: posterior results, linear model

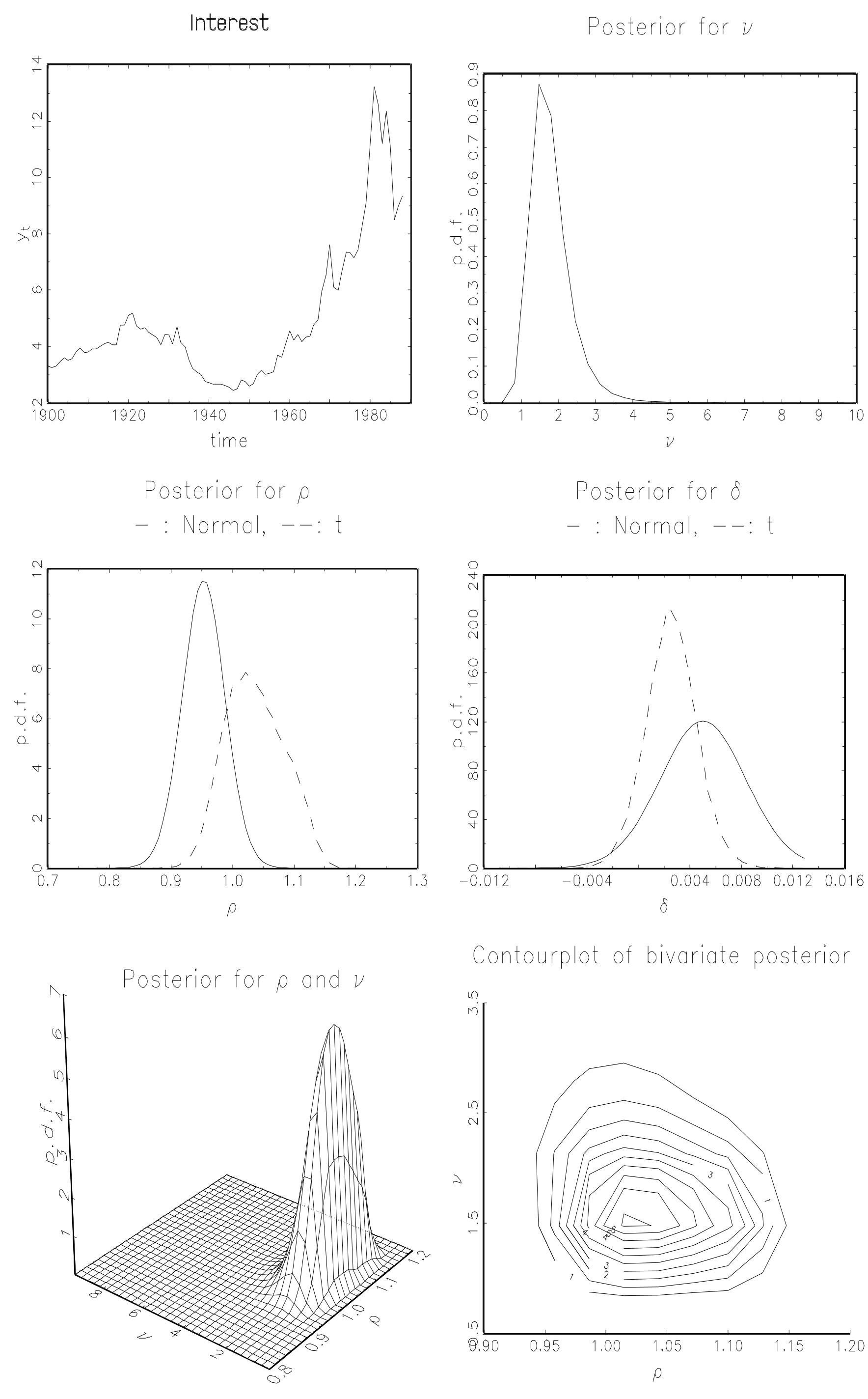

Figure 4.7.— Interest: posterior results, linear model

In contrast, the MLT estimator is only robust to a few outlying observations. Therefore, the MM estimator is probably more 'robust' to structural changes than the MLT estimator. Loosely speaking, the MM estimator discards the observations before or after the structural break, depending on their number.

The Bayesian results of Geweke (1995) also differ from the present results. Considering the original Nelson-Plosser data (until 1970), Geweke finds that the degrees of freedom parameter, the unit root parameter, and the posterior odds ratio in favor of difference stationarity are positively correlated. However, Geweke uses a different prior, a different model specification, and a smaller data set (the original Nelson-Plosser data). It is argued in Hoek et al. (1995) that the difference in outcomes can, to a large extent, be attributed to the different samples that are used, as opposed to to the different model specifications that are employed.

The results of this section, in particular for the Lydia Pinkham advertising and the Finland/US real exchange rate series, indicate that the degrees of freedom parameter and the unit root parameter are negatively correlated. This finding supports the theoretical results of Sections 4.2 and 4.3 and is also obtained by Kleibergen and van Dijk (1993) in their analysis of the US treasury bill rate series. As a result of this negative correlation, maintaining the assumption of normality ($\nu = \infty$) in the analysis of series containing outlying observations may incorrectly provide evidence against the unit root hypothesis. With respect to the Nelson-Plosser data series, the posterior odds ratios indicate that six series are (trend)stationarity. These results are relatively robust with respect to the specification of the prior for ν (see Hoek et al. (1995)).

4.6 Concluding Remarks

In this chapter the effect of outliers in the data on unit root inference was examined. It was shown that additive outliers provide evidence against the unit root hypothesis, even if the bulk of the data is described by a difference stationary model.

The outlier sensitivity of the standard Dickey-Fuller statistic and of Bayesian inference procedures under a Gaussian likelihood is caused by the nonrobustness of the OLS estimator (which equals the posterior mean in a Bayesian analysis with flat priors). This estimator has an unbounded influence function. The influence function of the maximum likelihood estimator based on a Student t likelihood with finite degrees of freedom, was shown to be bounded. Therefore, the Dickey-Fuller t-test based on this estimator is less sensitive to aberrant observations than the Dickey-Fuller t-test based on OLS. Critical values for the test were computed by means of simulation. The (in)sensitivity of the tests to outliers was illustrated using both simulated and empirical data. As an additional result, it was shown that the use of heteroskedasticity consistent standard errors in the computation of the OLS based Dickey-Fuller t-test also provides some protection against the distortional effects of additive outliers.

In a Bayesian context, it was argued that replacing the Gaussian likelihood by an i.i.d. Student t likelihood results in posteriors that are less sensitive to outlying observations. A proper uniform prior for the degrees of freedom parameter was proposed.

The analysis of several time series, in particular the Finland/US real exchange rate and the Lydia Pinkham advertising series, provided empirical support for the theoretical results. For these series, a negative correlation between the degrees of freedom parameter and the unit root parameter was found. This also holds for most of the Nelson-Plosser series.

Finally, the present robustification of the Dickey-Fuller t-test is only a first step towards the creation of an outlier resistant unit root test. The influence function is only one out of several concepts by which the robustness of statistical procedures can be assessed. Moreover, the maximum likelihood estimator based upon the Student t distribution has a bounded influence function, but only just. As an alternative to the influence function one might consider the fraction of outliers an estimator can cope with. This leads to the consideration of high breakdown estimators, as is done in the next chapter. The present (low breakdown) estimators have their own merits. They are easily calculated and provide at least some protection against outliers.

4.A Proof of the Proposition

This appendix discusses the boundedness of the influence function (IF) of the MLT estimator for autoregressive models under isolated additive outlier (AO) contamination. As the results for the AR(p) are qualitatively similar to the results for the AR(1), only the latter case is dealt with in detail. Let x_t be a stationary AR(1) process, $x_t = \phi x_{t-1} + \varepsilon_t$. The ε_t process is i.i.d. with zero mean and variance one. If the variance is unknown, one can use the techniques in, e.g., Hampel et al. (1986, p. 105) to estimate it. Further, consider the AO model (4.12) with $\xi_t \equiv \zeta$ and $\{z_t\}$ an i.i.d. process. The MLT estimator can be defined as the functional $\hat{\phi}(F_y^\gamma) \to \mathbb{R}$ that solves

$$\int_{-\infty}^{\infty} \frac{\varepsilon_1}{1+\varepsilon_1^2/\nu} y_0 dF_y^\gamma(y) = 0, \tag{4.26}$$

with $\varepsilon_1 = y_1 - \hat{\phi}(F_y^\gamma)y_0$, $y = (y_1, y_0)'$, and F_y^γ the cumulative distribution function of y, given that $P(z_t = 1) = \gamma$. The IF of the MLT estimator under the present form of additive outlier contamination is given by (4.14). Given the regularity conditions in Martin and Yohai (1986), the following proposition follows directly from their Theorem 4.2.

Proposition 4.2 *The IF of the estimator $\hat{\phi}$, implicitly defined in (4.26), under the isolated AO model with $\xi_t \equiv \zeta$, equals*

$$IF(\zeta, \hat{\phi}, \{F_y^\gamma\}) = -C^{-1} \int_{-\infty}^{\infty} \frac{\varepsilon_1 - \phi\zeta}{\nu + (\varepsilon_1 - \phi\zeta)^2} (y_0 + \zeta) dF_y^0(y),$$

with $\varepsilon_1 = y_1 - \phi y_0$ *and*

$$C = \int_{-\infty}^{\infty} \frac{\varepsilon_1^2 - \nu}{(\varepsilon_1^2 + \nu)^2} y_0^2 dF_y^0(y).$$