Chapter 5

High Breakdown Unit Root Tests

In this chapter I suggest another alternative to the OLS based unit root test of Dickey and Fuller (1979). The test departs from the one suggested in Chapter 4 in that the present test can cope with a larger number of outliers. The behavior of this alternative test is studied using simulated data as well as the fourteen economic time series considered by Nelson and Plosser (1982) and extended by Schotman and van Dijk (1991a). The chapter largely draws from the material presented in Lucas (1995a).

The setup is as follows. Section 5.1 introduces the problem and motivates the choice of high breakdown point (HBP) estimators for testing the unit root hypothesis. Section 5.2 discusses the outlier mechanism and the MM estimator that is used. A preliminary asymptotic analysis of unit root tests based on M estimators can be found in Section 5.3. A full discussion of the appropriate asymptotics can be found in Chapter 6. Section 5.4 compares the performance of the HBP unit root test with that of the standard Dickey-Fuller test by means of simulations. Section 5.5 presents the results of the robust and nonrobust unit root tests for an empirical data set, namely the extended Nelson-Plosser series. Section 5.6 concludes this chapter.

5.1 Introduction

In Chapter 4 I have discussed the outlier sensitivity of the standard Dickey-Fuller t-test (DF-t) for a unit root. The solution proposed in that chapter was to replace the OLS estimator in the Dickey-Fuller procedure by an M estimator, in particular, by a pseudo maximum likelihood estimator based on the Student t distribution. This procedure went some way in making the DF-t less sensitive to anomalous observations. It is, however, well known in the robustness literature that M estimators can only cope with a limited number of outliers. In the i.i.d. regression setting, one extreme outlier is enough to corrupt the results obtained with an M estimator (see Hampel et al. (1986, Chapter 6)). In order to remedy this problem, the class of generalized M

(GM) estimators was introduced. GM estimators have a breakdown point of at most $(1+p)^{-1}$, where p is the number of regressors that is used (see Maronna (1976) and Maronna and Yohai (1981, 1991)). The percentage of outliers GM estimators can cope with (approximately $100/(1+p)$), therefore, decreases with the number of included regressors.

The use of GM estimators for testing the unit root hypothesis has two major drawbacks. First, GM estimators introduce weights for the regressors in the model. In order to obtain these weights, the regressors are standardized by some type of scaling matrix that mostly has the interpretation of a covariance matrix. In a typical unit root regression model, at least one of the regressors is integrated of order one. Consequently, one cannot use the standard GM procedure, as integrated regressors cannot easily be standardized. Moreover, it was shown in Chapter 4 that ordinary M estimators sometimes suffice for giving protection against outliers, at least in the sense that they have a bounded influence function for autoregressive (AR) time series models under additive outlier contamination. This suggests that additional weight functions for the regressors are not always needed in a time series context in order to obtain robustness.

The second drawback of using GM estimators for testing the unit root hypothesis, is that the maximum percentage of outliers of $100(1+p)^{-1}$ for GM estimators was derived in the context of i.i.d. regression. In the time series setting, additional complications arise. Consider a standard AR(p) process

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t,$$

with ε_t Gaussian white noise and p known. Furthermore, assume that one of the points y_s is replaced by the outlier $y_s + \zeta$ for some s and some large ζ. Ignoring endpoint effects, the outlying observation will enter the model $p+1$ times: first as the left-hand side variable and afterwards p times as a right-hand side variable. One outlying value of y_t can thus cause $p+1$ outlying $(p+1)$-tuples $(y_t, y_{t-1}, \ldots, y_{t-p})$, which are precisely the $(p+1)$-tuples that are used to calculate the estimates of the AR parameters. So if there is a single additive outlier (see Section 4.2) in the data, there are already $(p+1)$ outliers in a regression sense. This was illustrated for the case $p = 1$ in Figure 4.1. Regarding the occurrence of several additive outliers as a realistic possibility, the upper bound of $100(1+p)^{-1}$ per cent outliers for GM estimators is thus quickly reached (compare the arguments in Martin and Yohai (1991) and Rousseeuw and Leroy (1987, Section 7.2)).

In order to construct a unit root test that can deal with several additive outliers simultaneously, I propose the use of an HBP estimator. HBP estimators are designed to have a high breakdown point (see Section 2.2), ideally of[1] 1/2. This means that they can still provide useful information about the characteristics of the bulk of the data in the presence of a large number of

[1] It is worthwhile to note here that Davies (1993) conjectures that the maximum breakdown point for affine equivariant estimators in a regression context is 1/3 rather than 1/2. However, Davies considers full neighborhoods of a central model by defining a metric for the

anomalous observations. It should be kept in mind that statistical breakdown is mostly concerned with extreme outlier configurations. This might be deemed unrealistic. Moreover, the breakdown point was initially introduced for the i.i.d. regression setting. Defining a breakdown point for dependent observations is much more difficult (see Boente, Fraiman, and Yohai (1987) and Papantoni-Kazakos (1984)).

Several HBP estimators for the regression setting have been proposed in the literature: repeated medians (Siegel (1982)), least median of squares and least trimmed squares (Rousseeuw (1984)), S estimators (Rousseeuw and Yohai (1984)), MM estimators (Yohai (1987)), τ estimators (Yohai and Zamar (1988)), high breakdown GM estimators (Simpson, Ruppert, and Caroll (1992) and Coakley and Hettmansperger (1993)), projection based estimators (see, e.g., Maronna and Yohai (1993) and Tyler (1994)), generalized S estimators (Croux, Rousseeuw, and Hössjer (1994)), and rank based estimators (Hössjer (1994)). All of these estimators can attain a breakdown point of approximately 1/2 in large samples. Especially the earlier HBP estimators have a poor (relative) asymptotic efficiency if there are no outliers in the data. In this chapter I use the MM estimator. This estimator can achieve a high efficiency and a high breakdown point at the same time. Another advantage of the MM estimator as opposed to the GM estimator is that it needs no weights for the regressors. As explained earlier, this is very important in the context of integrated regressors.

The number of applications of outlier robust techniques to time series problems is limited. Apart from the material in the previous chapter, there are some examples of the use of low breakdown point estimators for time series problems. Martin (1979, 1981), for example, suggests the use of GM and conditional mean M estimators (see also Schick and Mitter (1994)), Bustos and Yohai (1986) use outlier robust estimates of residual autocovariances to estimate the parameters of ARMA models, and Allende and Heiler (1992) employ recursive GM estimators for ARMA models. There are also some applications of HBP estimators in a time series context. Lucas (1992, Chapter 4), for example, and Rousseeuw and Leroy (1987, Chapter 7) discuss the least median of squares estimator for ARMA and AR models, respectively, while Martin and Yohai (1991) use a combination of S estimators and Kalman filter equations to obtain HBP estimates for AR models. All of these articles deal with stationary time series. For nonstationary processes, some references are Herce (1993), Knight (1989, 1991), and Chapters 3 and 6.

To conclude this introduction, I briefly comment on the relation between the outlier robust unit root tests developed in Chapters 4 and 5 and the work of Andrews (1993). Andrews discusses exactly median-unbiased estimation in the AR(1) model. The term *median* might suggest a link to the robustness literature and to the present approach of using robust estimators in time series models. The present approach and that of Andrews are, however, quite dis-

space of probability measures. Moreover, Davies requires the estimators to be (uniquely) defined in this full neighborhood. This approach differs from the finite sample breakdown point approach of Donoho and Huber (1983).

tinct. Andrews proposes methods to correct for the downward small sample bias of the (nonrobust) OLS estimator in the AR(1) model with positive AR coefficient. His methods work well even in certain innovative outlier situations (see Section 4.2), as is demonstrated by his Table 4. In other settings, however, Andrews' approach is quite sensitive to the occurrence of discordant observations, especially additive outliers. In principle, Andrews' methods can also be used to correct for the finite sample bias of the robust estimators used in the present setup. His approach, therefore, complements rather than substitutes for the present procedure of using robust estimators to test for unit roots.

5.2 Outliers and Robust Estimators

Modeling outliers in a time series context is more complicated than in the ordinary regression case. This is due to the fact that for time series not only the magnitude of the outliers matters, but also their place in time and their correlation structure over time. As explained in Section 4.2, there are two outlier generating models for time series that are often employed in the literature, namely the additive outlier (AO) model and the innovative outlier (IO) model (see (4.12) and (4.13), respectively). For large AO's it was shown in Example 4.1 that the DF-t tends to $-(T-1)^{1/2}$, implying a rejection of the unit root hypothesis for sufficiently large values of the sample size T. Below, I give a similar derivation for a large, isolated IO.

Example 5.1 Let

$$y_t = \phi y_{t-1} + \varepsilon_t + \xi_t \tag{5.1}$$

for $t = 1, \ldots, T$, where ε_t is white noise, ξ_t is a contamination error, and $|\phi| \leq 1$. Further assume that $y_0 = 0$. I only consider one outlier that lies approximately halfway the sample, say at $t = s$, thus avoiding the intricacies of endpoint effects. The outlier enters the model by assuming that $\xi_t = 0$ for $t \neq s$, and $\xi_s = \zeta$, where ζ is some large number. The series y_t now contains an IO (see (4.13)). Solving the difference equation (5.1) subject to $y_0 = 0$, one obtains

$$y_t = \sum_{i=0}^{t-1} \phi^i (\varepsilon_{t-i} + \xi_{t-i}). \tag{5.2}$$

Let $c(\phi) = \sum_{t=s}^{T-1} \phi^{2(t-s)}$, then

$$\sum_{t=1}^{T} y_{t-1}^2 = \zeta^2 c(\phi) + O(\zeta), \tag{5.3}$$

and

$$\sum_{t=1}^{T} y_t y_{t-1} = \phi \zeta^2 c(\phi) + O(\zeta). \tag{5.4}$$

So, if ζ tends to infinity, then the OLS estimator $\hat{\phi}$ tends to the true value ϕ. This corresponds with the right panels of Figure 4.1, where there is only one vertical outlier in the (y_t, y_{t-1})-plane and a large set of high-leverage points that are close to the line with slope ϕ. Notice that

$$\begin{aligned} Ts^2 &= \sum_{t=1}^{T}(y_t^2 - 2\hat{\phi}y_t y_{t-1} + \hat{\phi}^2 y_{t-1}^2) \\ &= (\phi^{2(T-s)} + c(\phi) - 2\hat{\phi}\phi c(\phi) + \hat{\phi}^2 c(\phi))\zeta^2 + O(\zeta). \end{aligned} \tag{5.5}$$

Combining (5.5) with (5.3), one obtains that the standard error of the OLS estimator tends to

$$g(\phi) = T^{-1/2}(1 - \phi^2 + \phi^{2(T-s)}/c(\phi))^{1/2} \tag{5.6}$$

if ζ tends to infinity, implying that the DF-t tends to

$$\begin{aligned} \text{DF-t} &= \frac{\hat{\phi} - 1}{g(\phi) + o(1)} \\ &\overset{\zeta\to\infty}{\longrightarrow} -\left(T\frac{(1-\phi)(1-\phi^{2(T-s)})}{1+\phi}\right)^{1/2}. \end{aligned} \tag{5.7}$$

Thus, the DF-t might or might not reject, depending on the true value of ϕ and on the sample size. It is clear from (5.7), however, that if ϕ is close to but less than one, then large sample sizes (T) are needed for the DF-t to reject the unit root hypothesis. This indicates that the occurrence of IO's may weaken the power properties of the OLS based DF-t. This claim is substantiated in Chapters 6 and 7. △

The main conclusion from Examples 4.1 and 5.1 is that if the OLS based DF-t is used to test the unit root hypothesis, one extreme AO results in rejection of the hypothesis even if $\phi = 1$, while one extreme IO results in either rejection or non-rejection, depending on the true value ϕ and the sample size. Franses and Haldrup (1994), using a somewhat more general and less extreme form of contamination, generalize the result for AO's and provide some asymptotic results.

The examples indicate that there is a need for robustifying the DF-t to outliers. Therefore, I replace the nonrobust OLS estimator in the Dickey-Fuller procedure by the MM estimator of Yohai (1987). The MM estimator can simultaneously attain a high breakdown point and a high efficiency. The estimator starts with a (low efficiency) HBP estimator, like the LMS or an S estimator (see Rousseeuw (1984) and Rousseeuw and Yohai (1984)). This low efficiency HBP estimator produces initial estimates of the parameters, which can be used to obtain a high breakdown scale estimate. The HBP estimates of the regression parameters and of the scale parameter can be used as starting values for a k-step M estimation procedure. This M estimator inherits the breakdown behavior of the initial estimator, but can at the same time be chosen such that it is highly efficient at some central model, e.g., the Gaussian distribution. In the remainder of this section I discuss some more details of the S and MM estimators that are employed in this chapter.

Rousseeuw and Yohai (1984) introduced the class of S estimators. This estimator can attain a breakdown point of 1/2, is regression equivariant, and, in constrast to the LMS estimator of Rousseeuw (1984), enjoys $T^{1/2}$-consistency. Rousseeuw and Yohai showed that S estimators satisfy the same type of first order conditions as ordinary M estimators. In order to calculate an S estimator, one proceeds as follows. Assume for the remainder of this section that the model is $y_t = \phi y_{t-1} + \varepsilon_t$. The S and MM estimator are, of course, also defined for more general regression models. One starts by choosing a bounded function ρ_1 (compare Section 2.3). For a given value of ϕ, define the scale estimate $\sigma_S(\phi) > 0$ as the solution of

$$T^{-1} \sum_{t=1}^{T} \rho_1((y_t - \phi y_{t-1})/\sigma_S) = 0.5 \sup\{\rho_1(x) | x \in \mathbb{R}\}. \tag{5.8}$$

The S estimator of ϕ is the minimand of $\sigma_S(\phi)$. The OLS estimator is obtained by setting $\rho_1(x) = x^2$ and replacing the right-hand side of (5.8) by $(T-1)/T$.

Because ρ_1 has to be bounded, there can be many local optima. A global search algorithm has to be employed to find the global minimum. I used the following procedure. First, I computed the LMS estimator by means of the random subsampling technique described in Rousseeuw and Leroy (1987). In the simulations, I used 1,000 random subsamples, while the number of subsamples for the empirical data sets in Section 5.5 was 10,000. Starting from the LMS estimates, I used a Newton-Raphson algorithm to locally improve the objective function $\sigma_S(\phi)$ of the S estimator subject to the constraint (5.8). As the objective function is only implicitly defined by the constraint (5.8), I used the implicit function theorem to obtain the necessary derivatives of $\sigma_S(\phi)$ with respect to ϕ. A step optimizer was included in the algorithm in order to ensure that the objective function was decreased during each iteration.[2] Especially the initial global random search is rather computer intensive. This seems to be a generic problem for HBP estimators, although recently some less time consuming algorithms have been put forward (Atkinson (1994) and Woodruff and Rocke (1994)).

A disadvantage of the S estimator is that its high breakdown point is counterbalanced by a low efficiency at the Gaussian model (see Rousseeuw and Yohai (1984)). Therefore, Yohai (1987) proposed to use the S estimates as starting values for an ordinary M estimator. This procedure works as follows. Starting from the HBP estimate of the AR parameter ϕ, one first calculates a HBP scale estimate, σ_{MM}. If the S estimator is used as the initial estimator, the final objective function of the S estimator can be used as the scale estimate.

[2] In fact, small increases in the objective function were allowed if the algorithm got stuck near a local optimum.

The HBP scale estimate is kept fixed during the remaining calculations. Next, an M estimator is computed based on the objective function

$$\min_{\phi} \sum_{t=1}^{T} \rho_2((y_t - \phi y_{t-1})/\sigma_{MM}), \tag{5.9}$$

where ρ_2 is again a bounded function. This produces the MM estimate ϕ_{MM}.

So far, I have said nothing about the functions ρ_1 and ρ_2, except that they have to be bounded. Naturally, there are some additional conditions that have to be met in order for the MM estimator to be consistent and asymptotically normal. These conditions can be found in Yohai (1987). The specifications for ρ_1 and ρ_2 that are used in this chapter are

$$\rho_i(x) = \begin{cases} (3c_i^4 x^2 - 3c_i^2 x^4 + x^6)/6 & \text{for } |x| \leq c_i, \\ c_i^6/6 & \text{for } |x| > c_i, \end{cases} \tag{5.10}$$

with $i = 1, 2$, $c_1 = 1.547$, and $c_2 = 4.685$ (Rousseeuw and Yohai (1984), Yohai (1987)). The derivative of $\rho_i(x)$ produces the bisquare ψ function of Beaton and Tukey (1974). The constant c_1 ensures that the S estimator has a breakdown point of approximately 1/2. The value of c_2 serves to improve the efficiency of the S estimator at the central Gaussian model. Using $c_2 = 4.685$, the MM estimator has an asymptotic efficiency of 95% with respect to the maximum likelihood estimator for Gaussian i.i.d. innovations.

A final point to note is that the MM estimator has not got a bounded influence function (IF). This might seem a bit worrying. The following can, however, be proved. Let ϵ denote a small, but *positive* fraction of contamination. Moreover, let $B(\epsilon)$ denote the bias of the MM estimator under this fraction of contamination. Then for every small, but positive ϵ, $B(\epsilon)/\epsilon$ is finite, meaning that the bias standardized by the fraction of contamination is bounded for strictly positive amounts of contamination (Yohai (1987, p. 650)).[3] Perhaps the unboundedness of the IF for the MM estimator is related to that of the LMS estimator. The LMS estimator can be shown to have an unbounded IF for inliers, i.e., observations that lie near the center of the data (see Hettmansperger and Sheather (1992) and Davies (1993)).

5.3 Asymptotic Distribution Theory for I.I.D. Innovations

In this section I discuss the asymptotic distribution of the DF-t for M estimators in the context of the AR(1) with i.i.d. innovations. A more elaborate exposition of the asymptotic theory for unit root tests based on M estimators can be found in the next chapter.

[3] In fact, the bias curve of the MM estimator (see Subsection 2.2.4 for a definition) is finite for values of ϵ below the breakdown point, but it passes vertically through the origin.

The asymptotics of the OLS estimator for the autoregressive parameter of an AR(1) under the assumption of a unit root have been extensively studied in the literature (see the references in Diebold and Nerlove (1990)). Here, this literature is extended to the DF-t based on an M estimator. We can state the following objective for the outlier robust unit root test. At the central model, that is in a situation without outliers, the robust and OLS based DF-t must be comparable in terms of size and power. In general, the power of the robust test will be somewhat below that of the classical test at the central model. This is a result of the tradeoff between efficiency and robustness for robust estimators (see Hampel et al. (1986, Section 2.1.c)). The lack of power of the robust test at the central model must be compensated for by a gain in insensitivity of size and power to departures from the central model. Whereas the size and power of the OLS based DF-t can be spoilt completely by one extreme additive outlier, its robust counterpart is less sensitive to anomalous observations.

In this section the asymptotic behavior of the robust DF-t is studied at the central model. The effects of contamination by outliers is studied in the next section by means of several simulation experiments. Consider the simple AR(1) case, $y_t = \phi y_{t-1} + \varepsilon_t$. Assume that the variance of the innovations σ^2 is known. If σ^2 is unknown, it can be estimated along with the other unknown parameters. As long as one uses a consistent estimator, the results of this section are not changed (compare Phillips (1987), Knight (1989, Section 4.3)).

Following Section 2.3, the M estimator $\hat{\phi}_M$ is a solution of

$$\sum_{t=1}^{T} \psi((y_t - \hat{\phi}_M y_{t-1})/\sigma) y_{t-1} = 0, \tag{5.11}$$

where ψ is a function satisfying the conditions below. The unit root corresponds to the case $\phi = 1$. In order to calculate the DF-t, one needs an estimate of the variance of $\hat{\phi}_M$. Following Hampel et al. (1986, Section 2.3), I use

$$s_\phi^2 = \sigma^2 \left(\sum_{t=1}^{T} \psi^2((y_t - \hat{\phi}_M y_{t-1})/\sigma) y_{t-1}^2 \right) / \left(\sum_{t=1}^{T} \psi'((y_t - \hat{\phi}_M y_{t-1})/\sigma) y_{t-1}^2 \right)^2, \tag{5.12}$$

where $\psi'(x) = d\psi(x)/dx$ (compare (4.19)). The DF-t is now given by

$$t_\phi = (\hat{\phi}_M - 1)/s_\phi. \tag{5.13}$$

I make use of the following assumptions.

Assumption 5.1 *The errors ε_t are i.i.d. with mean zero and positive finite variance σ^2; Δy_t is stationary.*

Assumption 5.2 *$\psi(\cdot)$ is bounded; $\psi'(\cdot)$ is bounded and Lipschitz continuous; $E(\psi(\varepsilon_t/\sigma)) = 0$; $\sigma_\psi^2 = E(\psi(\varepsilon_t/\sigma)^2)$ and $0 < \sigma_\psi^2 < \infty$; $\mu_\psi = E(\psi'(\varepsilon_t/\sigma))$ and $0 < \mu_\psi < \infty$.*

All expectations are taken with respect to the central model. Assumption 5.1 can be relaxed by allowing more general innovation sequences like mixing processes (see Chapter 6) at the cost of additional complexity. The results obtained here partly overlap those of Knight (1989). The infinite variance assumption that is used in his paper, however, simplifies the final result. Moreover, it is illustrative to provide the asymptotic distribution of the DF-t for i.i.d. innovations explicitly before turning to the more general results of the next chapter.

The asymptotic distrubution of t_ϕ is expressed as a functional of Brownian motions. I define the partial sum processes $B_T(s) = T^{-1/2}\sum_{t=1}^{\lfloor sT \rfloor}\varepsilon_t$ and $S_T(s) = T^{-1/2}\sum_{t=1}^{\lfloor sT \rfloor}\psi(\varepsilon_t/\sigma)$, for all $s \in [0,1]$, where $\lfloor x \rfloor$ is the integer part of x. The following lemma follows directly from Corollary 2.2 of Phillips and Durlauf (1986).

Lemma 5.1 *If Assumption 5.1 holds, then*

$$(B_T(\cdot), S_T(\cdot)) \Rightarrow (\sigma B(\cdot), \sigma_\psi S(\cdot)),$$

with B and S denoting two (correlated) standard Brownian motions with $E(B(s)S(s)) = s \cdot r$ for $0 \le s \le 1$, and r the correlation between ε_t and $\psi(\varepsilon_t/\sigma)$.

The next theorem states the main asymptotic result for the i.i.d. case. Its proof can be derived from the proof of Theorem 6.1 in the next chapter.

Theorem 5.1 *If Assumptions 5.1 and 5.2 are satisfied and if $(\hat{\phi}_M - 1) = o_p(T^{-1/2})$, then under the null hypothesis $H_0:\ \phi = 1$,*

$$T^{-1}\sum_{t=1}^{T}\psi(\Delta y_t/\sigma)y_{t-1} \Rightarrow \sigma_\psi\sigma\int BdS \tag{5.14}$$

$$T^{-2}\sum_{t=1}^{T}\psi^2(\Delta y_t/\sigma)y_{t-1}^2 \Rightarrow \sigma_\psi^2\sigma^2\int B^2 \tag{5.15}$$

$$T^{-3/2}\sum_{t=1}^{T}\psi'(\Delta y_t/\sigma)y_{t-1}^2 \Rightarrow \mu_\psi\sigma^2\int B^2 \tag{5.16}$$

$$t_\phi \Rightarrow \int BdS\left(\int B^2\right)^{-1/2} \tag{5.17}$$

The condition $(\hat{\phi}_M - 1) = o_p(T^{-1/2})$ may seem strange at first sight. It ensures that the 'right' solution of (5.11) is chosen. This is important if there are multiple solutions to (5.11), as is the case for S and MM estimators. Because the ρ functions that define these estimators have to be bounded (see, e.g., (5.10)), their derivatives, $\psi(x)$, vanish for large values of x. Therefore, (5.11) can be satisfied for values $\hat{\phi}_M$ that are completely different from the true value

1. By imposing the additional condition in the theorem, these complications are excluded.

The final result in (5.17) is similar to the expression found in Phillips (1987), except that the numerators of the two expressions differ. This can be seen by choosing $\psi(x) = x$. In that case, $\sigma_\psi^2 = 1$ and $S = B$ in Lemma 5.1, yielding $t_\phi \Rightarrow \int BdB(\int B^2)^{-1/2}$. This last expression equals part (e) from Theorem 3.1 of Phillips (1987) for the i.i.d. case.

The relation between the DF-t based on an M estimator (t_M) and on OLS (t_{OLS}) can be made more explicit. If r is defined as the correlation between ε_t and $\psi(\varepsilon_t/\sigma)$, $r = E(\varepsilon_t\psi(\varepsilon_t/\sigma))/(\sigma_\psi\sigma)$, then it is proved in the next chapter that $t_M - rt_{OLS} \Rightarrow \sqrt{1-r^2}N(0,1)$, where $N(0,1)$ is a standard normal random variable. Estimators that, like the MM estimator, have a high efficiency at the Gaussian distribution, also have values of r close to unity. Therefore, the differences between the critical values of t_M and t_{OLS} are expected to be small for these estimators.

Theorem 5.1 is of little practical importance, because it only states a result for the AR(1) case. It can, however, easily be shown that under the present assumptions the augmented DF-t has the same limiting distribution. The augmented DF-t is the t-test statistic for $\phi = 1$ in the AR(p) regression model

$$y_t = \phi y_{t-1} + \phi_1 \Delta y_{t-1} + \ldots + \phi_{p-1}\Delta y_{t-p+1} + \varepsilon_t \tag{5.18}$$

(compare Section 4.1 and Fuller (1976, p. 374)). This follows from the fact that $T^{-3/2}\sum_{t=1}^{T} y_{t-1}\Delta y_{t-k} = o_p(1)$ for $k = 1, \ldots, p-1$.

It is well known (see, e.g., Fuller (1976, p. 378) and the next chapter) that including a constant or linear time trend into (5.18) changes the asymptotic distribution of the OLS based DF-t. This also holds in the present context of DF-t statistics based on M estimators. Theorem 5.1 can be generalized to deal with such complications (see Theorem 6.1 in Chapter 6).

A last point to mention here is that Theorem 5.1 presents a result for M estimators. The theorem is, however, equally valid for both S and MM estimators. This may seem a bit contradictory, as the breakdown behavior of ordinary M estimators is completely different from that of the other two estimators. This suggests that the derivations given above cannot cover all three estimators at once. Rousseeuw and Yohai (1984) and Yohai (1987) show, however, that both S and MM estimators satisfy the same type of first order conditions as ordinary M estimators. Because only the first order condition is relevant for the asymptotic distribution of the DF-t, the results hold for all three classes of estimators.

5.4 Size and Power Robustness of the Unit Root Tests

In this section, the model

$$z_t = \phi z_{t-1} + \mu + \gamma t + \varepsilon_t \tag{5.19}$$

is considered. A constant and trend are included in (5.19), because they are needed in the empirical illustration contained in the next section. It was shown in the previous section for $\phi = 1$ that the DF-t converges in distribution to the random variable given in (5.17). In this section I focus on the finite sample distributions of the (non)robust DF-t statistics. The critical values for these test statistic are obtained by means of a simulation experiment. The behavior of the tests is also studied in situations with AO's or IO's, both for $\phi = 1$ and $\phi \neq 1$.

The simulation experiment was set up as follows. I set $\gamma = \mu = z_0 = 0$. Next, T standard normal i.i.d. variables ε_t were generated.[4]

For the series with IO's, a new series $\tilde{\varepsilon}_t$ was constructed using random numbers v_t that were uniformly distributed over the interval $[0, 1]$. The variable $\tilde{\varepsilon}_t$ was set equal to ε_t if $v_t \geq \eta$, with η a prespecified constant in the unit interval. If $v_t < \eta$, $\tilde{\varepsilon}_t = \varepsilon_t + \xi_t$, where ξ_t was a random variable with distribution function G. Using this new innovations sequence $\tilde{\varepsilon}_t$, a time series y_t was constructed according to the model $y_t = \phi y_{t-1} + \tilde{\varepsilon}_t$ and $y_0 = 0$.

For series with AO's, the original sequence ε_t was used to construct an uncontaminated time series z_t using (5.19). Next, a new time series y_t was constructed using uniform random numbers v_t with $y_t = z_t + \xi_t$ if $v_t < \eta$, and $y_t = z_t$, otherwise.

I used several choices for the contaminating distribution G of ξ_t. The first one, $G(\xi_t) = 1_{\{\xi_t \geq 0\}}(\xi_t)$, generated no outliers. Here, $1_A(\cdot)$ denotes the indicator function of the set A. The second distribution generated outliers of equal size 5, $G(\xi_t) = 1_{\{x \geq 5\}}(\xi_t)$. The third and fourth distribution were a Gaussian with mean zero and variance 9, and a standard Cauchy distribution, respectively. For each of these choices of G I used $\eta = 0.05$. Note that the Cauchy distribution violates assumption 5.1. It is only included to illustrate the effect of extreme outliers.

In order to estimate the critical values of the robust and nonrobust unit root tests at the 5% significance level, the possibly contaminated series y_t was used along with a constant and a trend in order to compute both the OLS and the MM estimator. The values of the associated DF-t statistics were stored over 10,000 replications and the final estimate of the critical value was the 0.05th-quantile of these 10,000 values, i.e., the 500th observation. In order to study the effect of the sample size, the simulations were carried out for $T = 50, 100, 200$. The results are presented in[5] Table 5.1.

Before discussing the details of Table 5.1, something must be said about the illustrative purpose of its entries (compare Section 4.5). When testing the unit root hypothesis for some empirical time series, a researcher, in general, does

[4] I used the uniform random number generator drand48(), provided in the standard library of the C programming language on a SUN Sparc Workstation (System IV). This generator has a period of approximately $2^{31} - 1$.

[5] Estimated standard errors of the entries in Table 5.1 are in general below 0.05. The exceptions are the entry for the MM estimator for $T = 50$ and δ_5 AO contamination, and the entries for the OLS estimator for $T = 100, 200$ with $C(0, 1)$ AO contamination. The estimated standard errors for these three entries is approximately 0.10.

TABLE 5.1
Upper 5% Points for the Robust and Nonrobust DF-t
for Contaminated and Uncontaminated Processes

Contaminating distribution	$T = 50$ OLS	$T = 50$ MM	$T = 100$ OLS	$T = 100$ MM	$T = 200$ OLS	$T = 200$ MM
			No outliers			
δ_0	-3.50	-4.03	-3.46	-3.67	-3.42	-3.52
			Additive outliers			
δ_5	-5.44	-6.10	-5.72	-3.61	-5.85	-3.36
$N(0,9)$	-4.61	-4.43	-4.63	-3.79	-4.62	-3.70
$C(0,1)$	-6.66	-4.25	-9.04	-3.70	-12.77	-3.63
			Innovative outliers			
δ_5	-3.46	-3.11	-3.51	-3.47	-3.44	-3.08
$N(0,9)$	-3.45	-3.46	-3.54	-3.86	-3.44	-3.33
$C(0,1)$	-3.54	-3.91	-3.49	-3.45	-3.45	-3.16

The table contains the 0.05-quantiles of 10,000 simulated DF-t tests. Standard errors are, in general, below 0.05. The contaminating distributions are δ_x (a point mass at x), $N(0,9)$ (a normal distribution with mean 0 and variance 9), and $C(0,1)$ (the standard Cauchy distribution).

not know which observations are outliers with respect to the chosen model. Neither does (s)he know the kind of the contamination (s)he is dealing with (IO's or AO's). Therefore, (s)he will use the critical values simulated for an uncontaminated error process. For example, for a time series of length 100 (s)he can use -3.46 for the nonrobust test and -3.67 for the robust test, suggesting a size of 5%. If, however, the considered time series contains several outliers, the actual size of the test can be far different from 5%. This is, of course, undesirable. One would like to have a test with an approximately constant size under a variety of contaminations. It is the aim of the table to show that the MM estimator is more suited to meet this objective than the OLS estimator.

The entries under the heading OLS in the first row of Table 5.1, i.e., for uncontaminated processes, are similar to the entries in the table of Fuller (1976, p. 373).[6] The absolute critical values for the robust test are somewhat larger in this case than those for the nonrobust test. The difference, however, decreases for increasing values of the sample size T. This can be expected in view of the high efficiency of the MM estimator at the central (Gaussian) model. For the uncontaminated process, I also performed simulations with $T = 400$, resulting in two almost identical critical values of -3.41 and -3.42 for the OLS and MM estimator, respectively.

The remaining entries in Table 5.1 demonstrate the size robustness or nonrobustness of the DF-t based on the MM and the OLS estimators. Except for small samples, $T = 50$, the critical values of the robust unit root tests do not

[6] In contrast to the simulations presented in Section 4.3, no heteroskedasticity consistent standard errors are used for the OLS based DF-t.

fluctuate much under different forms of additive outlier contamination. This stands in sharp contrast to the results based on the traditional OLS estimator. Consider the case $T = 100$ with critical value -3.46. If one uses this critical value, one believes that the significance level is approximately 5%. If, however, the time series is contaminated with 5% outliers, the actual size of the test is 27% if $N(0, 9)$ is used as a contaminating distribution, or even 57% if δ_5 is used. Note that even in the extreme case of a contaminating Cauchy distribution, the critical values of the robust DF-t at a 5% significance level are very similar to the ones for the uncontaminated process, while those for the nonrobust estimator are very much affected. Also note that the critical values of the OLS based DF-t decrease for increasing sample size T for δ_5 and $C(0, 1)$, whereas they increase for the MM estimator.

The results are reversed if one looks at IO situations. In these cases, the size of the traditional OLS based DF-t seems more stable than that of its robust counterpart. An important difference, however, is that the size of the robust test at the fixed critical value given in the top row of Table 5.1 is often decreased by the occurrence of the IO's, while that of the nonrobust test was increased by AO's.

A final note on the entries in Table 5.1 concerns the results for the additive outliers. When looking at the AO generating model, one sees that the true model is given by $y_t = y_{t-1} + \varepsilon_t + \xi_t - \xi_{t-1}$, which is an ARIMA(0,1,1) model. Therefore, it would be more appropriate to use the Phillips-Perron (1988) $Z(t_\alpha)$ test instead of the ordinary DF-t, because the $Z(t_\alpha)$ test corrects for the MA behavior of the disturbance term. Using the same simulation framework as above, it turned out that the results obtained with this latter test statistic were comparable to those obtained with the DF-t.

I now turn to the power behavior of the robust and nonrobust DF-t. I focus on the IO case. In order to obtain some insight, a small simulation study is set up to calculate the size adjusted power of the robust and nonrobust DF-t. Again μ, γ, z_0, and y_0 are set equal to zero. For several values of ϕ between 0.5 and 1.10 a (possibly contaminated) series y_t of length 100 was constructed using the same method as for the level simulations that were described earlier in this section. Next, the OLS and MM estimates were calculated using the first lag of y_t, a constant, and a trend as regressors. This procedure was repeated 100 times.[7] Using the critical values from Table 5.1, a record was kept of the number of rejections of the null hypothesis of a unit root for both the robust and nonrobust DF-t. The rejection percentages are plotted in Figure 5.1 for the uncontaminated process and a process with innovative outliers using the contaminating distribution $N(0, 9)$.

The first thing to note in Figure 5.1 is the similarity in power behavior of the robust and nonrobust test for uncontaminated processes. This is due to

[7] This relatively small number of replications was used, because the MM estimator is time consuming to compute. The number of 100 simulations is, however, sufficient to illustrate the qualitative difference between the power behavior of the DF-t based on the MM estimator and the OLS estimator.

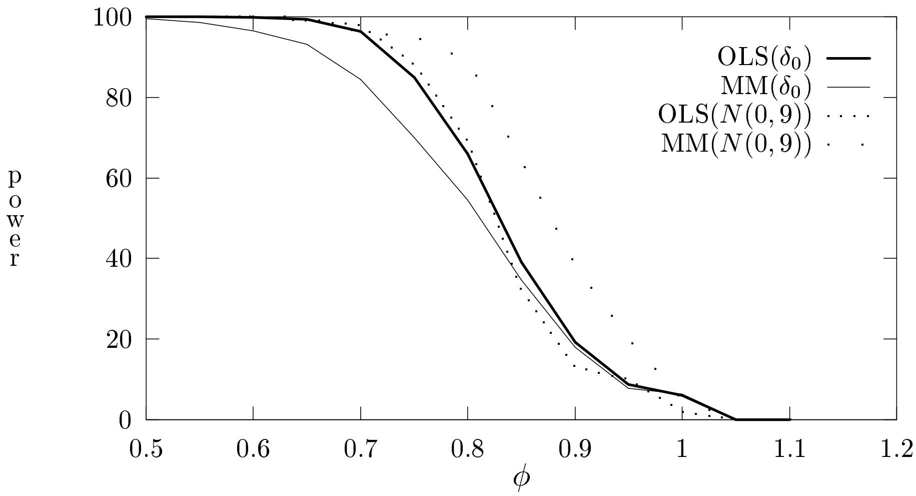

Figure 5.1.— Simulated Size Adjusted Power Functions for Processes Containing Innovative Outliers or No Outliers

the high efficiency of the MM estimator at the central model. Furthermore the figure clearly demonstrates that the size adjusted power of the robust test is superior to that of the traditional test in the IO case. This finding is in itself of little practical importance, because in reality one will not be working with a size adjusted critical value, but just with the top entries of Table 5.1. This entails a poorer power behavior of the robust test for at least some values of ϕ, because Table 5.1 indicates that the distribution of the robust test statistic has shifted to the left in the IO situation. The results for IO's with contaminating distribution δ_5 are similar to the ones described above. From the simulations in the next chapter, however, it appears that if the innovations are fat-tailed, there is a fairly large region over which the robust DF-t outperforms the nonrobust DF-t in terms of power.

Unreported additional simulations revealed that the size adjusted power of the robust test is somewhat inferior to that of the nonrobust test for series with AO's. Again, the same argument of practical relevance can be raised, now favoring the use of the robust estimator over the nonrobust one. Finally, it is worth mentioning that for $\phi = 0.5$, the size adjusted power of the traditional test in the extreme case of $C(0,1)$ additive outliers is only 0.15, while that of the robust test is as high as 0.95 (compare Section 6.4).

5.5 Application to the Extended Nelson-Plosser Data

Nelson and Plosser (1982) looked at the behavior of fourteen economic time series. Using the testing methodology of Dickey and Fuller, they showed that thirteen of these series were better described by an AR process with a unit root than by a stationary AR process around a deterministic trend. By now, most authors agree that their findings were rather premature. Several researchers were able to obtain different conclusions in a Bayesian (see for example the Oct-Dec 1991 issue of the *Journal of Applied Econometrics*) or nonparametric framework (see, e.g., Bierens (1992)). In the classical parametric context, Perron (1989) argued that most of the series considered by Nelson and Plosser could be regarded as stationary fluctuations around a linear trend if allowance was made for several structural breaks in this trend function. A number of authors have criticized Perron's findings. They indicated that if one appropriately accounts for the fact that the date of the structural break is fixed on the basis of the data, the evidence against the unit root hypothesis is much weaker (see the July 1992 issue of the *Journal of Business and Economic Statistics*).

It is interesting to note that there exists a link between outliers and structural breaks. This is clearly exhibited in Section 2 of Perron and Vogelsang (1992). Both additive and innovative outliers can cause structural breaks under the unit root hypothesis. Perron (1989) suggested to remedy this problem by introducing the appropriate dummy variables. As will be clear from the previous sections, one could also try to use robust estimators to eliminate the effect of outlying observations (or innovations). In this way one can circumvent the problem of choosing the specific points in time for the dummy to equal one. This should overcome the size problem of the test of Perron signaled by Christiano (1992). A noticeable difference between the robust estimation approach and the approach of Perron, however, is that when using robust estimators to deal with the outliers, the structural breaks can be present only under the null hypothesis of a unit root. Under the alternative hypothesis the effect of outliers is only temporary due to the stationary character of the series in that case.[8]

Nelson and Plosser used yearly data up to 1970 for fourteen economic time series. In this chapter I use the extended Nelson-Plosser data of Schotman and van Dijk (1991a), which contain observations up to 1988. The first difficulty arises in fixing the order of the AR polynomial for each series. I use the model selection strategy as suggested by Perron (1989) and Perron and Vogelsang (1992). First an AR(8) is estimated, with a trend and a constant included in the regression. If the coefficient of the longest lag differs significantly from zero at the 5% significance level, the AR(8) model is used for testing the unit root hypothesis. Otherwise the order of the AR polynomial is reduced by 1 and the selection strategy is applied to this lower order AR model. It is clear that the modeling strategy described above may give rise to different orders of the AR polynomials, depending on whether the OLS or the MM estimator is used. Therefore, I present the results for both models.

Let model I be the model chosen with the OLS estimator and model II the

[8]It is of course possible to extend the robust estimation approach with dummy type regressors that capture level shifts and/or trend breaks. This is not done in the present chapter.

TABLE 5.2
Tests for Autoregressive Unit Roots in Model I

Series	T	p	$\hat{\phi}_{OLS}$	$t_{\phi_{OLS}}$	outliers	$\hat{\phi}_M$	t_{ϕ_M}	outliers
Real GNP	80	2	0.82	-3.45	1	0.81	-4.70*	4
Nominal GNP	80	2	0.94	-2.02	1	0.96	-1.38	5
Real per capita GNP	80	2	0.82	-3.52*	1	0.80	-4.77*	4
Industrial production	129	6	0.85	-2.49	1	0.89	-1.71	9
Employment	99	2	0.85	-3.41	1	0.84	-3.27	8
Unemployment rate	99	4	0.72	-3.92*	1	0.75	-3.80*	8
GNP deflator	100	2	0.97	-1.59	2	0.99	-0.88	5
Consumer prices	129	2	0.99	-1.01	3	0.99	-1.33	10
Wages	89	7	0.91	-2.47	0	0.95	-1.00	12
Real wages	89	2	0.93	-1.68	0	0.94	-1.56	1
Money stock	100	2	0.94	-2.86	1	0.95	-2.13	3
Velocity	120	1	0.96	-1.60	0	0.94	-2.39	6
Interest rate	89	6	0.96	-1.11	2	0.90	-1.97	11
Common stock prices	118	2	0.92	-2.41	1	0.92	-2.44	2

The table contains the OLS ($\hat{\phi}_{OLS}$) and the MM ($\hat{\phi}_M$) estimates of the unit root parameter in an autoregressive model of order p. The corresponding values of the unit root tests are $t_{\phi_{OLS}}$ and t_{ϕ_M}, respectively. * denotes significance at the 5% level using the first row of Table 5.1. T denotes the sample size. Outliers gives the number of observations with standardized absolute residuals greater than 3. The model selection for this table was based on the OLS estimator.

one chosen with the MM estimator. The results are shown in Tables 5.2 and 5.3. The length of the time series is given by T and the chosen order of the AR polynomial by p. Furthermore, the number of points with scaled residuals greater than 3 in absolute value is counted for both the OLS and the MM estimator and presented under the label outliers. The scaling is done using an estimate of the standard error of the regression. For the OLS estimator, the usual residual standard deviation is used, while for the MM estimator the HBP scale estimate σ_{MM} is used (see Section 5.2).

In Section 5.4 the benefits of using robust estimators for unit root testing were demonstrated using simulated contaminated time series. In the present section, using real economic data, it can once again be seen that results obtained with robust estimators are less sensitive to outliers. Consider for example the series of common stock prices, which illustrates the effect of outliers on model selection. Using the model selection strategy described above and the OLS estimator, one chooses a second order AR polynomial. If one uses the MM estimator instead, one ends up with a first order AR model. The cause of the difference between the two selected models can be visualized nicely by looking at added variable plots (see Cook and Weisberg (1991)). Let y_t denote the stock prices at time t and let P be the projection matrix, projecting on

TABLE 5.3
Tests for Autoregressive Unit Roots in Model II

Series	T	p	$\hat{\phi}_{OLS}$	$t_{\phi_{OLS}}$	outliers	$\hat{\phi}_M$	t_{ϕ_M}	outliers
Real GNP	80	8	0.75	-3.22	1	0.68	-4.83*	6
Nominal GNP	80	6	0.95	-1.47	1	0.97	-1.37	9
Real per capita GNP	80	8	0.73	-3.39	1	0.65	-4.38*	7
Industrial production	129	4	0.81	-3.38	3	0.83	-3.17	9
Employment	99	4	0.84	-3.43	0	0.85	-4.13*	12
Unemployment rate	99	4	0.72	-3.92*	1	0.75	-3.80*	8
GNP deflator	100	6	0.96	-1.51	2	0.98	-1.45	5
Consumer prices	129	4	0.99	-1.20	3	0.99	-1.50	9
Wages	89	2	0.94	-2.36	0	0.94	-2.84	12
Real wages	89	1	0.96	-0.97	0	0.97	-0.90	1
Money stock	100	2	0.94	-2.86	1	0.95	-2.13	3
Velocity	120	1	0.96	-1.60	0	0.94	-2.39	6
Interest rate	89	8	0.96	-0.67	2	1.04	1.53	13
Common stock prices	118	1	0.94	-1.82	1	0.93	-2.44	2

The table contains the OLS ($\hat{\phi}_{OLS}$) and the MM ($\hat{\phi}_M$) estimates of the unit root parameter in an autoregressive model of order p. The corresponding values of the unit root tests are $t_{\phi_{OLS}}$ and t_{ϕ_M}, respectively. * denotes significance at the 5% level using the first row of Table 5.1. T denotes the sample size. Outliers gives the number of observations with standardized absolute residuals greater than 3. The model selection for this table was based on the MM estimator.

the space spanned by the columns of X, where X is a $T \times 3$ matrix with the tth row of X equal to $(1, t, y_{t-1})$. The added variable plot in Figure 5.2 plots $\hat{e}_t^{(1)} = (I - P)y_t$ versus $\hat{e}_t^{(2)} = (I - P)\Delta y_{t-1}$. Three outliers can be observed in this figure. Performing an OLS regression using all points in Figure 5.2, one obtains the solid line. This line has a significant positive slope coefficient. This is largely due to the outlying observations for 1931 and 1932, which are marked with a solid circle. Omitting these observations and recalculating the OLS estimator, one obtains the dotted line, which is markedly flatter than the solid one and has an insignificant slope coefficient. Even if one also discards the leverage point 1933, indicated by the open circle in the figure, the regression line is flatter than the one based on the full sample. The MM estimator automatically recognizes the three points as being outliers and accordingly assigns less weight to them. Therefore the MM estimator is more suited for model selection in an outlier context.

The results of the DF-t tests are quite remarkable. While there is some doubt about the stationarity of real GNP and GNP per capita if one uses the OLS estimator, the DF-t obtained with the MM estimator strongly rejects the unit root hypothesis at the 5% significance level, both in model I and II. Similar findings hold for the employment series, although rejection of the unit

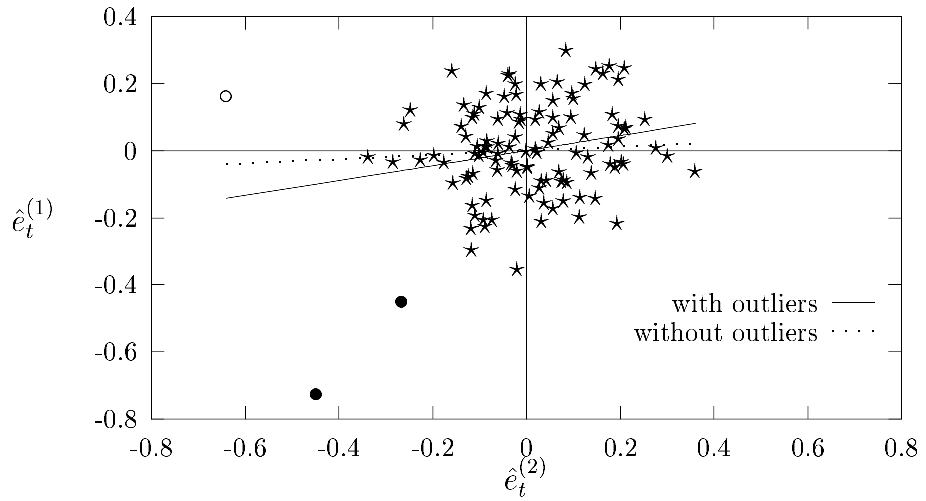

Figure 5.2.— Added Variable Plot for Lagged Differenced Common Stock Prices

root hypothesis is now only obtained in model II. Another result is the larger value of the test statistic for industrial production in model II. Although the unit root cannot be rejected at the 5% significance level, this series is more a borderline case than suggested by the results obtained with OLS. For the remaining series, there is no difference in the conclusions that are obtained with the OLS and the MM estimator.

Concerning the nature of the outliers in the series, most of them appear to be of the innovative type. As explained in the introduction, an isolated additive outlier is directly followed by a number of additional outliers. These 'patches' of outliers rarely occur in the fourteen Nelson-Plosser series. Therefore, one can suspect that the outliers are more often of the innovative than of the additive type. In order to visualize the outliers, one can use the added variable plots discussed earlier. Consider the (log) real GNP series. Let P denote the projection matrix on the space spanned by the columns of X, with the tth row of X now equal to $(1, t, \Delta y_{t-1}, \ldots, \Delta y_{t-7})$, and y_t denoting the real GNP series. Figure 5.3 plots $\hat{e}_t^{(1)} = (I - P)y_t$ (vertical axis) versus $\hat{e}_t^{(2)} = (I - P)y_{t-1}$ (horizontal axis). The 6 outliers are depicted by the circles (both open and solid). It is easily seen that these observations, even when using an OLS projection, are outliers. If the real GNP series contains a unit root, the points in the scatter diagram should lie around a line with slope one. The OLS regression lines computed with and without the outlying observations are presented by the solid and dotted line, respectively. One sees that the outliers cause the slope of the regression line to be pulled towards one. More important, however, is the increase caused by the outliers in the estimated variance of the error term. This increased estimated error variance results in an increase in the standard errors of the parameter estimates. Therefore,

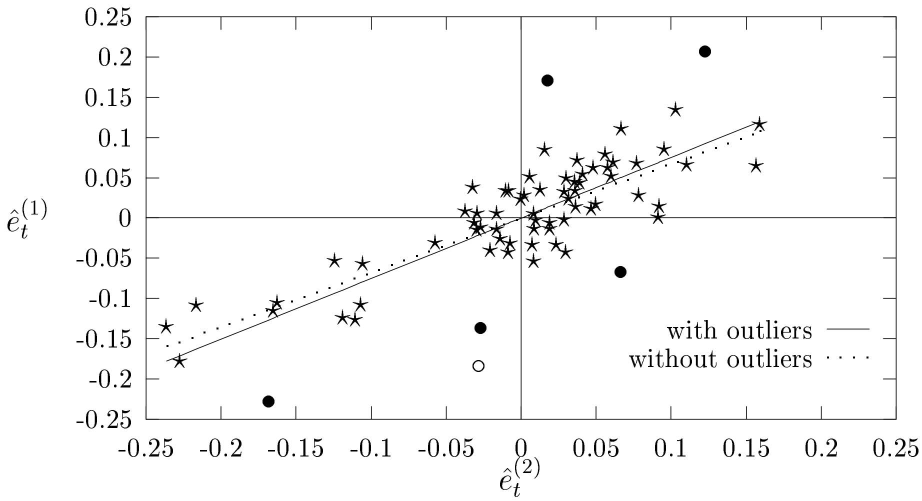

Figure 5.3.— Added Variable Plot for the Lagged Logarithm of Real GNP

it is not remarkable that downweighting or omitting these anomalous data points, as is done by the MM estimator, leads to a stronger rejection of the unit root hypothesis. It is also worth noting that with the OLS estimator only the observation for 1932, the open circle in Figure 5.3, is detected as being an outlier. This can be compared to the performance of the robust MM estimator, which has the advantage that several outlying observations are detected in just one estimation run.

A final point to note is that the MM estimator indicates a structural break for the interest rate series after 1980. This conclusion is based the observation that seven of the thirteen outliers occur during the last nine years of the sample. All the standardized residuals of these outliers exceed 4.5 in absolute value. In contrast, the OLS estimator only indicates the year 1980 as being an outlier with an absolute standardized residual of about 3. It is indeed true that the interest rate series shows an increase in the error variance after 1980 (see Figure 4.7). This suggests that the MM estimator is also more suited than the OLS estimator for detecting specific model inadequacies.

5.6 Conclusions

In this chapter an outlier robust alternative was considered for the well known Dickey-Fuller t-test (DF-t) for testing the unit root hypothesis. This alternative test was obtained by replacing the nonrobust ordinary least squares estimator in the Dickey-Fuller procedure by the high breakdown point MM estimator. A simulation study revealed that the size of this robust test was much less influenced by the occurrence of additive outliers than that of its nonrobust counterpart. The fact that the use of robust estimators may lead to different outcomes, both for model selection procedures and for the unit root test, was demonstrated using the macroeconomic time series considered in Nelson and Plosser (1982) and extended by Schotman and van Dijk (1991a). For one of these series, namely the interest rate, it appeared that the robust estimator was also more suited for detecting model inadequacy.

The results are promising, although several problems remain to be tackled. First, the robust estimation of ARMA models could be considered as an alternative to the AR models discussed in this chapter. High breakdown estimation for ARMA models is, however, still a largely open area (see, e.g., Lucas (1992, Chapter 4)). Another option is to consider alternative robust estimators, as is done in Chapters 4 and 6, or to extend the present testing methodology to the multivariate context of cointegration (see Chapters 7 and 8).

Second, the interpretation of the outliers is important. Are there reasons to expect certain years not to fit into the pattern set out by the bulk of the data? If there are, one can feel fairly comfortable when discarding these observations. Otherwise, it might prove difficult to accept the fact that these data points are not fully taken into account when estimating the model. If there are several outliers for which one cannot find an explanation, respecifying the model might ultimately be deemed a more plausible solution than assigning less weight to the discordant observations.

Finally, it is an interesting topic for further research to compare the performance of the present method with that of the skilled econometrician. One can argue that every skilled econometrician, possibly aided by some statistical test procedure, can (almost) always spot influential observations and correct for them by introducing dummies (compare Franses and Haldrup (1994)). As Hampel et al. (1986, Section 1.4) note, this procedure is also a form of robust estimation. Therefore a simulation experiment could be performed to find out whether the procedure described in this paper performs significantly better than the traditional approach mentioned above. Mechanizing the subjective judgement of the (skilled) econometrician is not an easy job, however, and until this can be done in a satisfactory way, such an experiment must be postponed.