



ELSEVIER

Journal of Econometrics 69 (1995) 27–59

**JOURNAL OF
Econometrics**

Classical and Bayesian aspects of robust unit root inference

Henk Hoek*, André Lucas, Herman K. van Dijk

Tinbergen Institute and Econometric Institute, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands

Abstract

This paper has two themes. First, we classify some effects which outliers in the data have on unit root inference. We show that, both in a classical and a Bayesian framework, the presence of additive outliers moves 'standard' inference towards stationarity. Second, we base inference on an independent Student- t instead of a Gaussian likelihood. This yields results that are less sensitive to the presence of outliers. Application to several time series with outliers reveals a negative correlation between the unit root and degrees of freedom parameter of the Student- t distribution. Therefore, imposing normality may incorrectly provide evidence against the unit root.

Key words: Outliers; Robustness; Unit root inference; Student- t distribution; Bayesian analysis

JEL classification: C11; C12; C22

1. Introduction

Graphical presentation of several economic time series reveals a number of common characteristics. Two of these characteristics are the presence of an upward trend and the presence of outlying observations. Both properties of economic time series have been extensively studied in the past decade.

First, following the seminal article of Nelson and Plosser (1982), an interesting debate is going on in the econometric literature whether most macroeconomic

* Corresponding author.

We thank Luc Bauwens, two anonymous referees, and seminar participants at Monash University and ESEM 1994 (Maastricht) for helpful comments on earlier versions of this paper. The usual disclaimer with respect to remaining errors applies.

time series contain a deterministic or a stochastic trend. The latter may correspond to a unit root in the autoregressive representation of the series. An overview of classical unit root analysis is provided by Campbell and Perron (1991). Bayesian contributions include, e.g., Sims (1988), DeJong and Whiteman (1991a,b), Kleibergen and van Dijk (1993), Schotman and van Dijk (1991a,b, 1993), Phillips (1991), and the references cited there.

Second, robust statistical inference has spawned an extensive literature on the treatment of outlying observations (see, e.g., Huber, 1981; Hampel et al., 1986). In particular, estimators are proposed that are robust with respect to contamination of the data. Robust estimators and statistics in a time series context can be found in, e.g., Künsch (1984), Martin (1981), Martin and Yohai (1986), and Lucas (1995a,b).

In the present paper we discuss some aspects of both fields. The aim is to classify the effect that outlying observations in the data have on both classical and Bayesian unit root inference. Next, we discuss a relatively simple way to reduce the effects of outliers. The basic idea is to base inference on an underlying Student- t error distribution instead of the usual Gaussian distribution (see, e.g., Kleibergen and van Dijk, 1993; Geweke, 1993). Note that we use the hypothesis of independently and identically distributed (iid) disturbances, which is different from the hypothesis of a multivariate Student- t error distribution, studied by Zellner (1976).

To motivate the use of the iid Student- t , some concepts from the robustness literature are used, in particular the notion of influence functions; see Section 2. These concepts are applied to classical unit root testing procedures in Section 3. We show that the often used Dickey–Fuller t statistic is sensitive to aberrant observations. Assuming a Student- t distribution with finite degrees of freedom, we show that the influence of an outlier on the maximum likelihood estimators (and indirectly on test statistics derived from these estimators) is bounded. Bayesian unit root inference using an iid Student- t likelihood is discussed in Section 4. Section 5 provides some empirical illustrations and discusses the relation of our work to previous results obtained in the literature. The effect which additive outliers have on the maximum likelihood estimates and the posterior means is shown. The posterior evidence, based on the analysis of series with outliers, suggests a negative correlation between the unit root parameter and the degrees of freedom parameter. We give some concluding remarks in Section 6. Finally, some technical results are given in the Appendix.

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 concepts from the literature

on robust statistical inference are introduced. First, we specify a model which generates outliers. A useful model is the general replacement model, given in Martin and Yohai (1986),

$$y_t = (1 - z_t)x_t + z_t w_t. \quad (1)$$

The binary 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 (cdf) $F_x(\cdot)$, while the *contaminating* process w_t has cdf $F_w(\cdot)$. Both of these processes can be non-iid. For instance, they may belong to the ARMA class. 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 cdf of the y_t process obviously depends on γ . We denote this by adding a superscript to its cdf $F_y^\gamma(\cdot)$.

Model (1) is easily recognized as a finite two-component mixture model, since $F_y^\gamma(\cdot) = (1 - \gamma)F_x(\cdot) + \gamma F_w(\cdot)$. Therefore, we could make use of the literature on modeling finite mixtures for constructing outlier robust inference procedures. In our parametric context, this requires the full specification of the w_t process. We refrain from this strategy and let the w_t process be (partially) unspecified; see (2) and (3) below. This allows us to specify procedures that are robust to more general forms of outlier behavior than those implied by the mixture of *two* stochastic processes. Informally, our aim 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 we do not need to specify a complete model for the w_t process. This automatically yields a more parsimonious model parameterization. Also, the estimation of a finite 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. We note that in the second stage of our analysis we make use of the Student- t distribution, which is an uncountable mixture of Gaussian distributions, with the inverted gamma-2 distribution as the mixing distribution (see Section 4). 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 as in the case of a finite mixture distribution.

By imposing certain structure on the z_t and w_t processes, model (1) can generate different types of outliers. For example, if the z_t are iid, model (1) 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 is illustrated using

an autoregressive (AR) process and the general replacement model (1). Let ξ_t be a process independent of x_t and z_t . AO's can now be modelled by specifying

$$\text{AO: } \phi(L)x_t = \varepsilon_t, \quad w_t = x_t + \xi_t, \quad t = 1, \dots, T, \quad (2)$$

and IO's by specifying

$$\text{IO: } \phi(L)x_t = \varepsilon_t, \quad w_t = x_t + \xi_t/\phi(L), \quad t = 1, \dots, T, \quad (3)$$

where $\phi(L)$ is a polynomial of order p in the lag operator L , $Lx_t = x_{t-1}$, and z_t is an iid process.

Using (1) and the AO specifications of x_t and w_t , we obtain that $y_t = x_t + z_t\xi_t$. Given that z_t equals 0 most of the times, y_t is in many cases 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 on the other hand we can write $\phi(L)y_t = \varepsilon_t + z_t\xi_t$. The additional error term now appears in the innovations that drive the time series.

The effect of both isolated AO's and IO's is visualized in Fig. 1. The figures show a realization of the y_t process for $\phi(L) = 1 - 0.9L$, $\gamma = 0.01$, $w_t \equiv 10$, and $\{\varepsilon_t\}$ a set of iid standard normal innovations. Also a scatter diagram of y_t versus y_{t-1} is presented. The same set of innovations is used to generate both time series. The outlier occurs at time $t = 25$.

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 y_{t-1} values, which all lie in the neighborhood of the line with slope $\phi = 0.9$. This last set of observations provides a strong signal for the true value of ϕ .

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

Outliers may seriously affect the 'usual' estimation and test procedures, like those based on ordinary least squares (OLS). If one is interested in describing the bulk of the data, then procedures that are less sensitive to the presence of aberrant observations are attractive. Several outlier robust estimators have been proposed in the literature (e.g., Huber, 1981; 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. Its finite sample approximation for the OLS estimator is closely related to the DF-BETA diagnostic of Belsley et al. (1980), which measures the standardized contribution of the r th observation to the estimator (see Hampel et al., 1986, Sec. 2.1.e). In a more formal context the IF is an

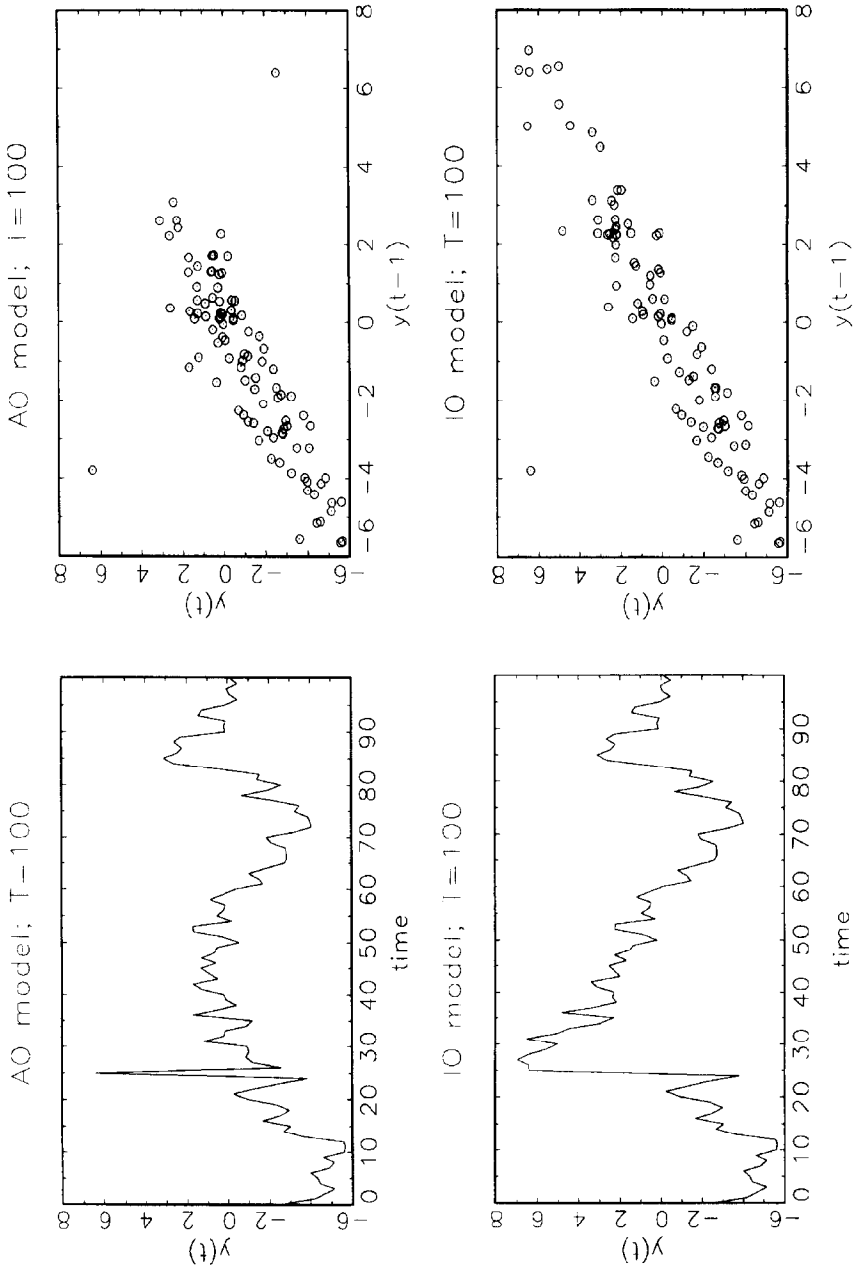


Fig. 1. Effects of isolated additive and innovative outliers.

asymptotic concept. It depends on the cdf and not on the sample. We define estimators $\hat{\phi}$ that can be considered as functionals on a space of cdf's, so $\hat{\phi} = \hat{\phi}(F_y^\gamma)$. Next, consider the change in $\hat{\phi}(\cdot)$ that is implied by increasing γ from 0 to some small positive number. So, we study $(\hat{\phi}(F_y^\gamma) - \hat{\phi}(F_y^0))$. If γ is very small, this difference is negligible. We therefore standardize by dividing by γ . The IF is now defined as the limit of this standardized difference as γ approaches 0, or

$$IF(\hat{\phi}, \{F_y^\gamma\}) = \lim_{\gamma \downarrow 0} \frac{\hat{\phi}(F_y^\gamma) - \hat{\phi}(F_y^0)}{\gamma}, \quad (4)$$

if this limit exists (Martin and Yohai, 1986). For more details on the IF we refer to, e.g., Hampel et al. (1986). By way of illustration, consider the IF of the mean.

Example 1. Let w_t , x_t , y_t , and z_t from (1) be iid processes. Furthermore, let $\hat{\phi}$ be the mean, so $\hat{\phi}(F_y^\gamma) = \int y dF_y^\gamma(y)$. Now consider a special choice for $F_y^\gamma(\cdot)$, namely $F_y^\gamma(y) = (1 - \gamma)F_y^0(y) + \gamma\Delta_\zeta(y)$, where $\Delta_\zeta(\cdot)$ is a cdf which jumps from 0 to 1 at ζ . If this special choice is made for $F_y^\gamma(\cdot)$, we use the term influence curve (IC) rather than an IF. Using the definition in (4), we obtain

$$\begin{aligned} \frac{\hat{\phi}(F_y^\gamma) - \hat{\phi}(F_y^0)}{\gamma} &= \frac{\int y d[(1 - \gamma)F_y^0(y) + \gamma\Delta_\zeta(y)] - \int y dF_y^0(y)}{\gamma} \\ &= \zeta - \hat{\phi}(F_y^0). \end{aligned} \quad (5)$$

If $F_y^0(\cdot)$ is the standard normal, the IC is equal to ζ . Note that this is an unbounded function in ζ .

Estimators with a bounded IF possess certain robustness properties and are therefore desirable (Hampel et al., 1986, Ch. 2; Martin and Yohai, 1986). The IF was originally developed to measure the change in the value of an estimator in the iid setting. Rousseeuw (1981) invented a similar concept to measure the change in the variance of an estimator: the change-of-variance function. A different line was followed by Künsch (1984) and Martin and Yohai (1986). They generalized the IF to situations with dependent observations.

Martin and Yohai (1986) prove that the OLS or conditional Gaussian ML estimator for AR models has an unbounded IC under AO contamination. This suggests that the estimator is not robust. The nonrobustness of the OLS estimator can be made explicit quite easily. This is shown in the next section.

3. Classical analysis

Following Nelson and Plosser (1982), the most popular classical unit root test has been the Dickey–Fuller t test (see Dickey and Fuller, 1979). The Dickey–Fuller t statistic (DF- t) is traditionally obtained by estimating an

autoregressive (AR) model by ordinary least squares. However, it was argued in the previous section that the OLS estimator is nonrobust against AO's. A test statistic based on this estimator might therefore also be nonrobust, as the following example illustrates.

Example 2. Consider the AR(1) $x_t = \phi x_{t-1} + \varepsilon_t$, with $t = 1, \dots, T$. The process ε_t is iid 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 (1) are defined. The OLS estimator of ϕ , calculated with the observed series y_t , equals

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

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 Fig. 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 $(T - 1)^{-1/2}(1 + o(1))$, so 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 Lucas (1995a, b) and Franses and Haldrup (1994). The intuition behind this phenomenon is straightforward: an AO is, by its definition, in contrast with the persistence of shocks implied by the unit root hypothesis. The consequences of IO's are less clear. It is known that the OLS estimator of the AR parameters is not very sensitive to the occurrence of IO's (see, e.g., Martin, 1981, and the remarks in Bustos and Yohai, 1986). However, using similar calculations as in Example 2, Lucas (1995a) shows 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, we look for an alternative estimator that possesses a bounded IF. One of the simplest alternatives is the conditional pseudo maximum likelihood estimator within the class of iid Student- t distributed innovations (MLT); compare Gouriéroux et al. (1984). 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

$$\sum \frac{\hat{\varepsilon}_t / \sigma}{1 + \hat{\varepsilon}_t^2 / \sigma^2 \nu} y_{t-1} = 0, \tag{6}$$

where $\hat{\varepsilon}_t = y_t - \hat{\phi} y_{t-1}$, σ is a scale parameter, and ν is the degrees of freedom parameter. It is easily seen that the MLT estimator for ϕ falls within the class of

M estimators.¹ It is known (see, e.g., Hampel et al., 1986) that for regression problems in the iid 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 often unbounded in the iid 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 1. Consider model (1); let x_t be an AR(p) process with iid innovations; let w_t be as in (2) and let z_t be a binary iid process with $P(z_t = 1) = \gamma$; then the MLT estimator has a bounded IF.

A more precise statement and a proof are given in the Appendix. Here, we present a heuristic derivation of a finite sample analogue of the IC for the MLT estimator. Consider the uncontaminated AR(1) series, $x_t = \phi x_{t-1} + \varepsilon_t$, which is observed from $t = 0, \dots, T$. Let $w_t = x_t + \zeta_t$, with $\zeta_t = 0$ for all $t \neq s$, and $\zeta_s = \zeta$ with $1 < s < T$. To avoid unnecessary complexities, we assume that $x_s = x_{s-1} = \varepsilon_s = \varepsilon_{s+1} = 0$, $\sigma = 1$, and that ζ^2/T is negligible. We now look at 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 by $\hat{\phi}$ and $\tilde{\phi}$, respectively. Note that the contamination parameter γ of (4) equals T^{-1} . We therefore look at the quantity $T(\tilde{\phi} - \hat{\phi})$. 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 (6) that $\tilde{\phi}$ solves

$$\sum \frac{\tilde{e}_t(\tilde{\phi})}{1 + \tilde{e}_t(\tilde{\phi})^2/v} y_{t-1} = 0. \tag{7}$$

Substituting $e_t(\hat{\phi})$ for $\tilde{e}_t(\tilde{\phi})$ and x_{t-1} for y_{t-1} in (7), we obtain a similar equation for $\hat{\phi}$. Taking a first-order Taylor expansion of the right-hand side (rhs) of (7) around $\hat{\phi}$ and omitting higher-order terms, we obtain

$$0 \approx \sum \frac{\tilde{e}_t(\hat{\phi})}{1 + \tilde{e}_t(\hat{\phi})^2/v} y_{t-1} - \left[T^{-1} \sum \frac{1 - \tilde{e}_t(\hat{\phi})^2/v}{(1 + \tilde{e}_t(\hat{\phi})^2/v)^2} y_{t-1}^2 \right] T(\tilde{\phi} - \hat{\phi}). \tag{8}$$

Next, we 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 (8) by $I_T(\hat{\phi})$. Using the estimator generating function of $\hat{\phi}$

¹ In the AR(1) context, the OLS estimator solves $\sum y_{t-1} \varepsilon_t = 0$. M estimators solve $\sum y_{t-1} \psi(\varepsilon_t) = 0$, where the function $\psi(\cdot)$ is chosen such that outliers are weighted less heavily. For examples, see Huber (1981) and Hampel et al. (1986).

$\hat{\phi}$ [compare (7)], the values of x_t and ε_t for t near s , formula (8) can be rewritten as

$$\begin{aligned} T(\bar{\phi} - \hat{\phi}) &\approx (I_T(\hat{\phi}))^{-1} \sum \frac{\tilde{\varepsilon}_t(\hat{\phi})}{1 + \tilde{\varepsilon}_t(\hat{\phi})^2/\nu} y_{t-1} \\ &= (I_T(\hat{\phi}))^{-1} \left[\frac{\tilde{\varepsilon}_{s+1}(\hat{\phi})}{1 + \tilde{\varepsilon}_{s+1}(\hat{\phi})^2/\nu} y_s + \sum \frac{e_t(\hat{\phi})}{1 + e_t(\hat{\phi})^2/\nu} x_{t-1} \right] \\ &= (I_T(\hat{\phi}))^{-1} \frac{-\hat{\phi} \zeta^2}{1 + \hat{\phi}^2 \zeta^2/\nu}. \end{aligned} \tag{9}$$

Note that this function is bounded in ζ . The OLS estimator is obtained by letting $\nu \rightarrow \infty$. It is easily seen that in that case (9) 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 cdf's F_y^0 . This suggests that the MLT estimator can still provide useful information about the true value of ϕ if the true cdf is in some neighborhood of the assumed 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.² 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 our knowledge not noted earlier in the literature. A few examples of IC's of the class of MLT estimators of ϕ in $y_t = \phi y_{t-1} + \varepsilon_t$ for the case of standard normal iid innovations are given in Fig. 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 in order to deal with outliers.

So far, we assumed that the scale parameter σ was known. Usually, it has to be estimated simultaneously with the autoregressive parameters ϕ_i , $i = 1, \dots, p$. This can be done by 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 $y_1 - \phi y_0$. However, there may arise some problems with the number of outliers this simultaneous estimation procedure can cope with (Maronna and Yohai, 1991).

² The function $\psi(\cdot)$ is called weakly redescending if $\lim_{x \rightarrow \pm\infty} \psi(x) = 0$. It is called strongly redescending if for some positive constant c it holds that $\psi(x) = 0$ for all $|x| \geq c$.

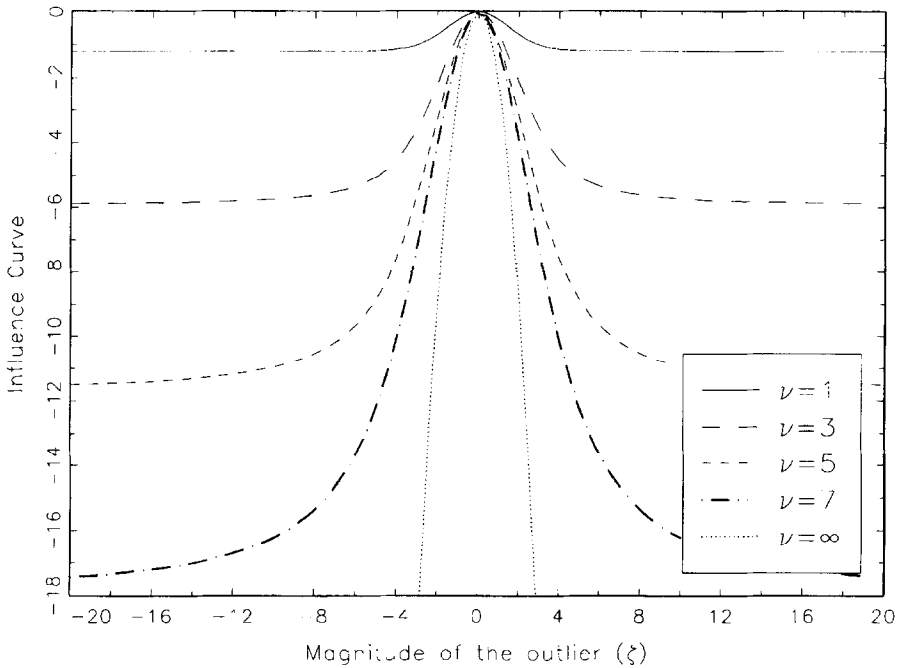


Fig. 2. Influence curve for $\phi = 0.8$ for several values of ν evaluated at $F_y^0 = \text{Gaussian}$.

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 (Lucas, 1992). Developing a bounded influence estimator for ν that is consistent for the family of t distributions is difficult. However, it was noted previously that the MLT estimator belongs to the class of M estimators. Most of these 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 t -distribution, but (s)he only uses the first-order conditions of the 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 t distributed innovations with a different degrees of freedom parameter, can be kept within bounds. However, some tradeoff has to be made between efficiency and robustness (Hampel et al., 1986, p. 44). Similar arguments are encountered in the literature on pseudo maximum likelihood estimators (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 to 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 1. Moreover, power can be gained by using the MLT methodology if the innovations are leptokurtic; see Lucas (1995b).

Upon closer inspection of (2), 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 time series. These tests can be found in the work of Phillips (1987) and Phillips and Perron (1988). However, the results in Lucas (1995a) 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 the present context. We use a similar simulation setup as in Fuller (1976). The asymptotic distribution of the DF- t based on the MLT estimator can be found in Lucas (1995a,b). We generate x_t from model (1) as a random walk of length 50, 100, or 200 with iid innovations. First we let $w_t \equiv x_t$ and estimate the regression models $y_t = \phi y_{t-1} + \varepsilon_t$, $y_t = \alpha + \phi y_{t-1} + \varepsilon_t$, and $y_t = \alpha + \beta t + \phi y_{t-1} + \varepsilon_t$ with the MLT estimator. We use several values for the degrees of freedom parameter ν . The DF- t for each of these models can be calculated. This process is repeated 1,000 times and the 50th-order statistic of the simulated DF- t values serves as an estimate of the 5% critical value. The standard error of $\hat{\phi}$ for the model without constant and trend is estimated by

$$\hat{\sigma}^2 \left[\sum_{t=2}^T y_{t-1}^2 \psi'(\hat{\varepsilon}_t/\hat{\sigma}) \right]^{-2} \left[\sum_{t=2}^T y_{t-1}^2 \psi(\hat{\varepsilon}_t/\hat{\sigma})^2 \right], \tag{10}$$

with $\psi'(x) = d\psi(x)/dx$, $\hat{\varepsilon}_t = y_t - \hat{\phi} y_{t-1}$, and $\hat{\sigma}$ is the estimate of the scale of ε_t (compare Hampel et al., 1986, p. 316). Similar formulas are used for the other two regression models. For $\hat{\sigma}$ we use the pseudo maximum likelihood estimator

under the Student- t distribution. The same value of ν is used for computing $\hat{\phi}$ and $\hat{\sigma}$. Note that (10) is in fact a kind of heteroskedasticity-consistent covariance estimator (compare White, 1980). This causes a discrepancy between the critical values tabulated by Fuller and the ones supplied in this paper for $\nu = \infty$. As noted by simulations in Lucas (1995b), the use of heteroskedasticity-consistent standard errors also helps to make the standard DF- t more robust. For completeness, we report the results for the standard DF- t with ($\nu = \omega$) and without ($\nu = \infty$) the heteroskedasticity correction for the standard error estimate. The critical values for the Gaussian random walk are presented in Table 1 under the heading 'clean'.

In order to illustrate the robustness aspects of the test, we perform similar simulations with z_t from (1) equal to an iid process and γ equal to 0.05. For the ξ_t process (2), we consider a sequence of iid Gaussian random variables with mean zero and standard deviation equal to five. The results of these simulations can be found in Table 1 under the heading 'outliers'.

The simulation experiment leads to the following four conclusions.

First, the size distortions due to the presence of AO's are increasing in ν , the tuning constant of the MLT estimator. This can be expected, because ν determines the degree of robustness of the estimator.

It is illustrative to consider an example of how Table 1 can be used. 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 we use the outlier generating scheme described below Table 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 of the exact model that generated the outliers. Therefore, he continues to use the critical value -3.527 , which for the present DGP gives a 51% level test. Similar arguments for the other two cases lead to the result that these tests have a size of 10% and 7.5%, respectively, for the DGP 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 in reducing the sensitivity of the size with respect to outliers. This finding was also noted in Lucas (1995b).

Second, the MLT based tests demonstrate a slower convergence to the asymptotic distribution. The different behavior for $\nu = 1$ and $\nu = \infty$ in Table 1 is considerable. Therefore, it seems necessary to use the finite sample critical values from Table 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 Fig. 3, that presents the cdfs of the unit root tests for

Table 1
Critical values at the 5% level for the unit root test with the MLT estimator for standard Gaussian iid innovations

v	n = 50		n = 100		n = 200	
	Clean	Outliers	Clean	Outliers	Clean	Outliers
<i>None</i>						
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
<i>Constant</i>						
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
<i>Trend</i>						
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 Gaussian random walk without outliers. For the entries under the heading 'outliers', the simulations are based on a random walk with 5% additive outliers. 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.

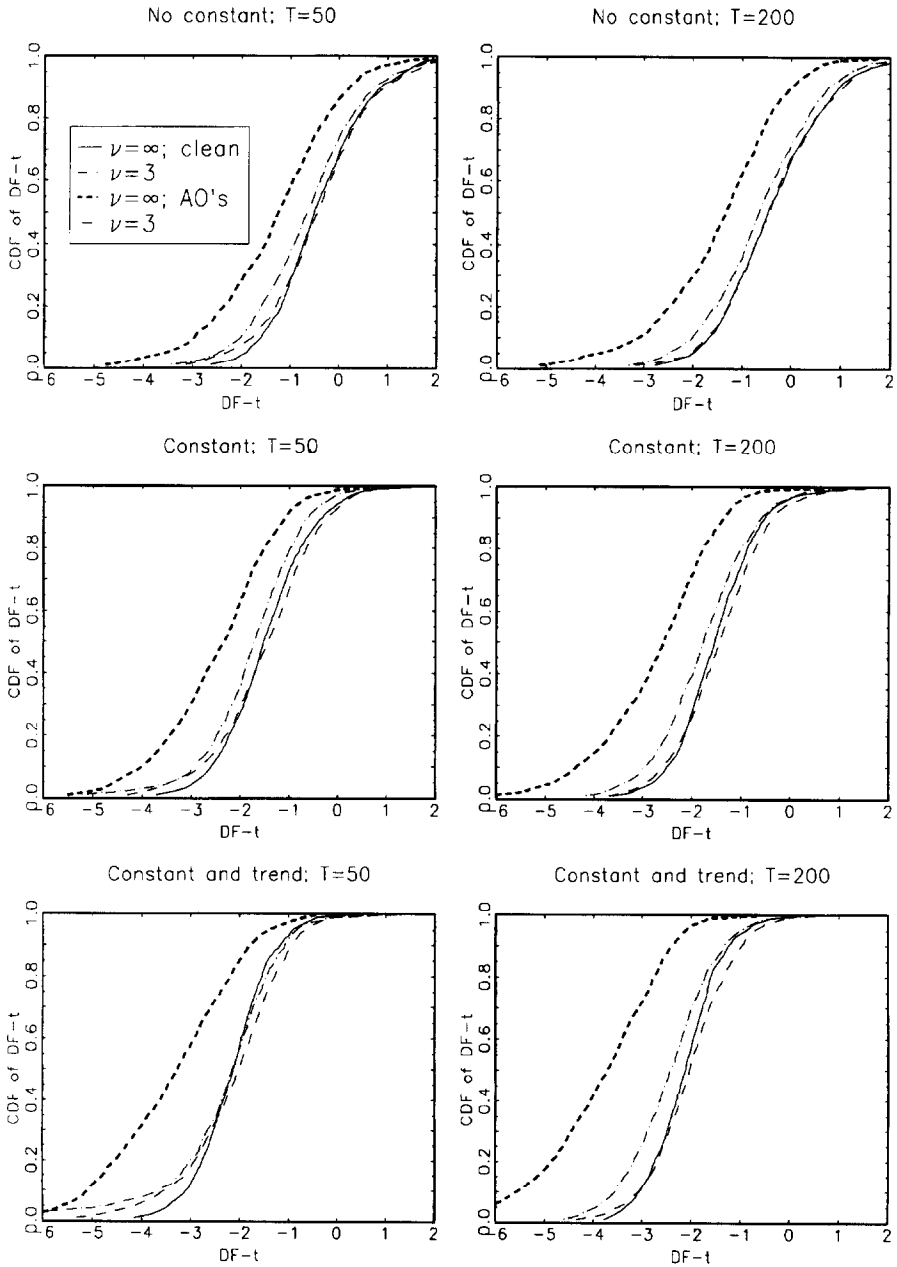


Fig. 3. CDF of the DF-t based on OLS and on MLT.

two values of ν . Especially for the regression model with trend the figure reveals that the cdf can change considerably if the sample is enlarged. Moreover, if outliers are added to the sample, the cdf 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 1, simulations were performed using a random walk with Student- t instead of Gaussian innovations. The result of using more heavily tailed innovations is that the critical values for all of the tests shift somewhat to the right,³ thus decreasing the type I error of the tests if the critical values of Table 1 are used. The robustness and convergence properties of the test remain similar.

Fourth, using a different distribution for generating the AO's does not alter the results. We 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 value for the MLT based test with $\nu = 3$ remains remarkably stable in the same setting and only shifts from -3.50 to -3.95 . Similar conclusions can be drawn from the remaining experiments.

4. Bayesian analysis

Recent years have seen a growing number of Bayesian studies on the possible presence of a unit root in macroeconomic time series (e.g., Sims, 1988; DeJong and Whiteman, 1991a,b; Phillips, 1991; Schotman and van Dijk, 1991a,b, 1993). The focus of these studies is on the specification of a prior distribution: posterior inference should, to a certain extent, not be sensitive 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. However, it is easily shown that also in a Bayesian framework AO's can seriously affect unit root inference.

Example 3. Consider the same processes x_t , w_t , and z_t as in Example 2. Define the parameter vector $\theta = (\phi, \sigma^2)$. Assume a diffuse prior $\pi(\theta) \propto \sigma^{-1}$ and a Gaussian likelihood. The marginal posterior for ϕ is a t density (see Judge et al., 1988, Sec. 7.4.4):

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

³The change is considerable if Cauchy innovations are used. This is in accordance with Knight (1991), who derives that for a certain class of infinite variance innovations the DF- t is again asymptotically normal, even if it is based on an M estimator.

where $\hat{\phi}_{OLS}$ and $\hat{\sigma}_{OLS}^2$ 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 mean μ , precision matrix Σ^{-1} , and degrees of freedom ν . Looking at the extreme case of an infinitely large AO ($\zeta \rightarrow \infty$), we obtain

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

Therefore, a sufficiently large additive outlier pushes the posterior away from the unit root, even if the true value of ϕ is unity.

Assuming normality in 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, Sec. 4.7). To model outliers, fat-tailed distributions have been suggested in the Bayesian literature as well. For example, Leamer (1978, Sec. 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 t likelihood possess certain robustness properties. Intuitively, these properties are passed on to results obtained from a Bayesian posterior analysis that uses a Student- t likelihood. Analogously, Example 3 shows that the nonrobustness of OLS estimators is 'inherited' by Bayesian inference using a Gaussian likelihood. We note that in order to evaluate the robustness properties in a Bayesian framework with the iid Student- t distribution the concept of the (posterior) score function has been used by Smith (1981, Sec. 5) and West (1984). We only analyze the effect of the Student- t on unit root inference, since that is our parameter of interest.

Geweke (1993) 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 iid 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 lead us to choose an iid Student- t likelihood⁴ instead of a Gaussian one and to label our Bayesian analysis of the unit root hypothesis as being *robust*.

⁴ 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. Also, the posterior for the location parameters is identical to the posterior under a Gaussian likelihood.

The likelihood function for the general linear model $y = X\beta + e$, where e is a vector of iid Student- t disturbances, is proportional to

$$\ell(\beta, \sigma^2, \nu | y) \propto \prod_{i=1}^T \frac{\Gamma((\nu + 1)/2)}{\Gamma(\nu/2)(\nu\sigma^2)^{1/2}} \left[1 + \frac{(y_i - x_i'\beta)^2}{\nu\sigma^2} \right]^{-(\nu+1)/2}, \tag{11}$$

where y_i and x_i' are the i th row of y and X , respectively. In our case, x_i contains a constant (and possibly a time trend) as well as lagged endogenous variables. This leads to the same expression as (11). Note that in time series models one usually conditions on a fixed initial value y_0 .

In order to examine the effect of the iid Student- t error distribution on unit root inference, we 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. We assume a flat prior for the parameters of the model, including ν ,

$$\pi(\mu, \phi, \sigma^2, \nu) \propto \sigma^{-1}. \tag{12}$$

An improper uniform prior for ν imposes normality, because the prior odds in favor of normality are infinite (see Geweke, 1993).⁵ Geweke proposes, therefore, to use an exponential prior for ν with parameter λ . The sensitivity of the results with respect to the choice of λ is examined by varying this parameter. The computations can be performed using the Gibbs sampler if one exploits the equivalence of a normal linear model with heteroskedasticity and the 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 paper we proceed as follows. We note that the likelihood specified in (11) approaches the Gaussian likelihood for large values of ν . Only for low values of ν the difference between the two is substantial. We therefore choose to specify a uniform prior 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:

$$\begin{aligned} \pi(\nu) &= 1/\nu^* \quad \text{for } 0 < \nu \leq \nu^*, \\ &= 0 \quad \text{elsewhere.} \end{aligned} \tag{13}$$

A sensitivity analysis can be performed by considering different values of ν^* . Note that it is argued in Section 3 that simultaneously estimating $\{\beta, \sigma^2, \nu\}$ results in a unbounded influence function. *A priori* imposing bounds on ν does not solve this problem.

The posterior resulting from (11), (12), and (13) is difficult to handle analytically. In order to perform posterior analysis, we apply importance sampling

⁵ Normality is equivalent to $\nu = \infty$.

using the SISAM program (see Hop and van Dijk, 1992). A multivariate t density is 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 ($v = v^*$ versus $v < v^*$) can be tested. Some empirical results are presented in the next section.

5. Empirical illustration

To illustrate the use of the iid Student- t in a Classical and Bayesian unit root framework, several time series are analyzed. First, we consider the Finland/US real exchange rate, studied by 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 Fig. 4 clearly shows, this series 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, we consider the following model for this series:

$$y_t = \mu + \rho y_{t-1} + \phi_1 \Delta y_{t-1} + \varepsilon_t, \quad (14)$$

where ρ denotes the unit root parameter, defined as the sum of the original AR coefficients.⁶

Next, we consider a well-known series from the marketing literature, the Lydia Pinkham annual advertising series (see Fig. 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, we consider the regression:

$$y_t = \mu + \rho y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \varepsilon_t. \quad (15)$$

Finally, as in Schotman and van Dijk (1991b), we consider the extended Nelson/Plosser data (see also Lucas, 1995a). To study these data, we consider an AR(3) model with linear time trend

$$y_t = \mu + \delta t + \rho y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \varepsilon_t. \quad (16)$$

All three models are first estimated using a maximum likelihood procedure under iid Student- t errors. Table 2 presents results for some values of v .

With the exception of the series of employment, velocity, and S&P 500, the estimates of ρ increase as v decreases, which is an indication of negative correlation between the maximum likelihood estimators for ρ and v . The most

⁶So in the case of the AR(1) model considered in the examples, $\rho = \phi$.

Table 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 normality. Heteroskedasticity-consistent standard errors are given between brackets. The final column gives the ML estimate of the degrees of freedom parameter ν .

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 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 2 can also be used to compute the DF- t statistic ($(\hat{\rho} - 1)/\hat{s}_\rho$). 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 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 error, 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 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 iid innovations.

The Bayesian posterior results, reported in Table 3, are comparable. 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 Figs. 4 and 5. Also, for both series the posterior for ρ shifts to the right when the restriction $\nu = \infty$ is dropped. The case against iid normality is the 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.

Our 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 statistic increases from -5.74 to -2.65 . However, the inclusion of dummy variables requires pretesting for the presence and the location of outliers. Our approach does not need such a first round and may therefore be easier to implement.

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 Fig. 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. The contour plot of the bivariate posterior and the entry in the final column of Table 3 give no indication of a substantial correlation between ρ and ν . For the interest rate series, shown in Fig. 7, the posterior for ρ clearly shifts to the right. Also, this posterior is somewhat less concentrated than the 'normal' posterior.

Table 3
 Posterior moments for the linear autoregressive model

Series	$E_N(\rho)$	$E_t(\rho)$	$E_t(v)$	$R_{\rho,v}$
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

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

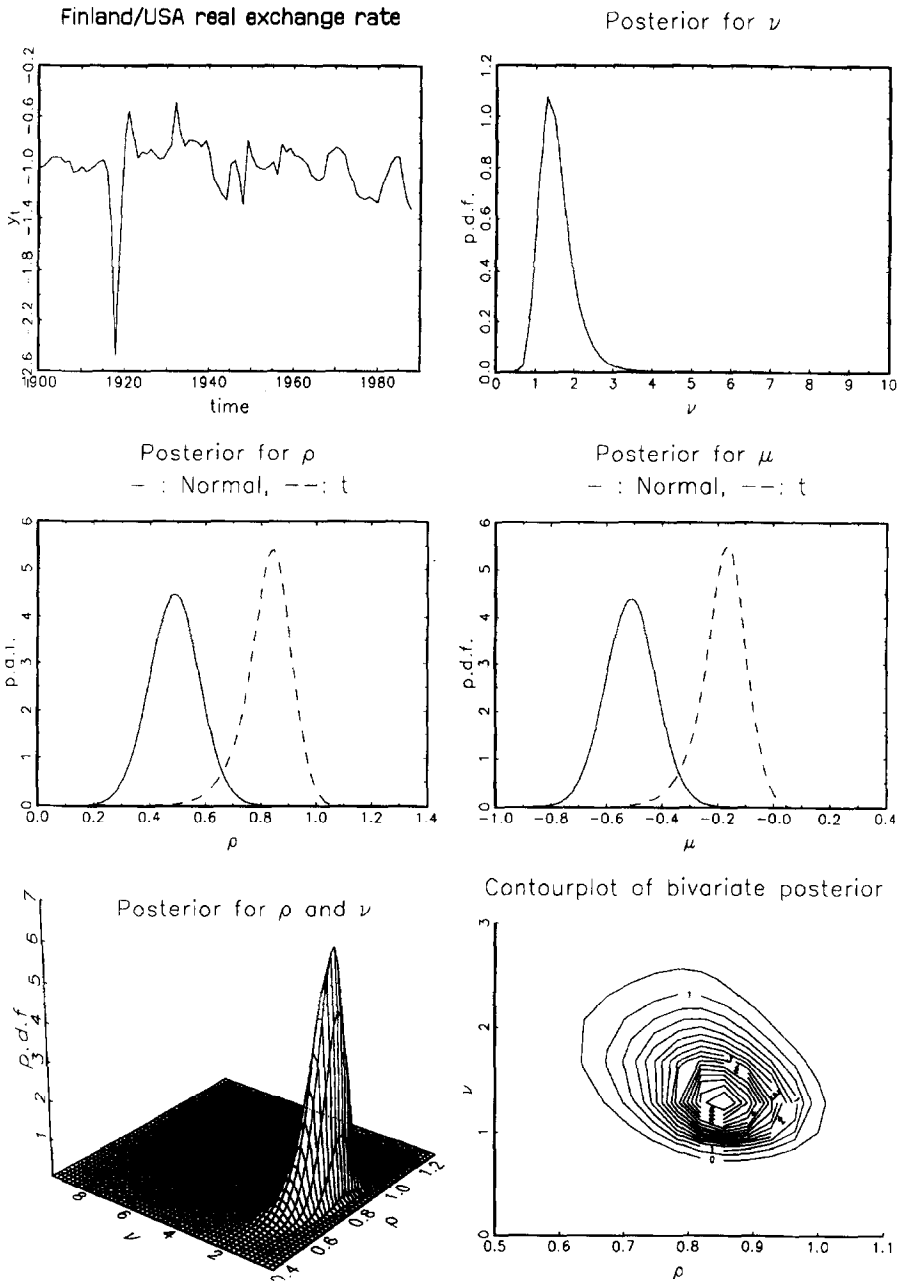


Fig. 4. Finland/US real exchange rate: Posterior results, linear model.

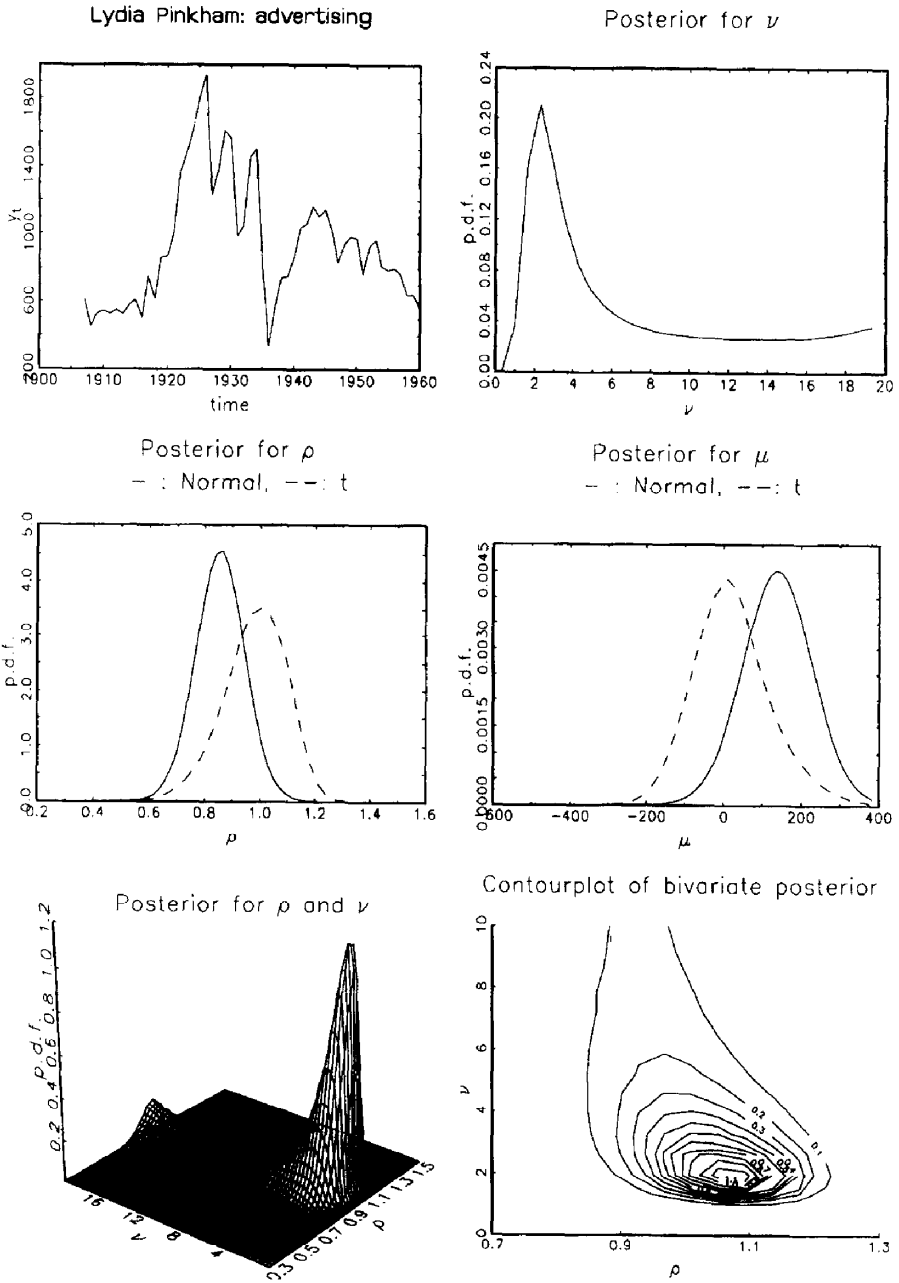


Fig. 5. Lydia Pinkham advertising: Posterior results, linear model.

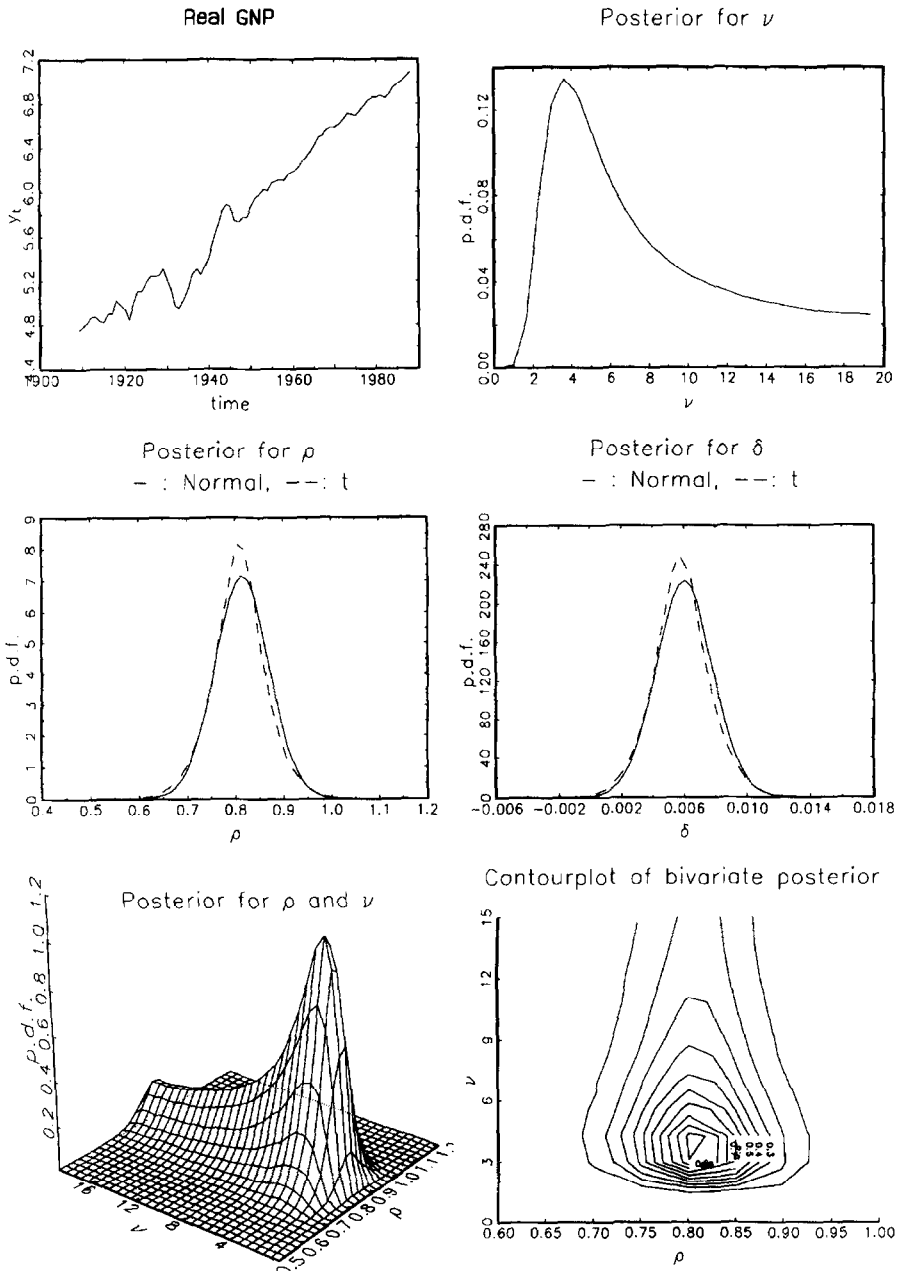


Fig. 6. Real GNP: Posterior results, linear model.

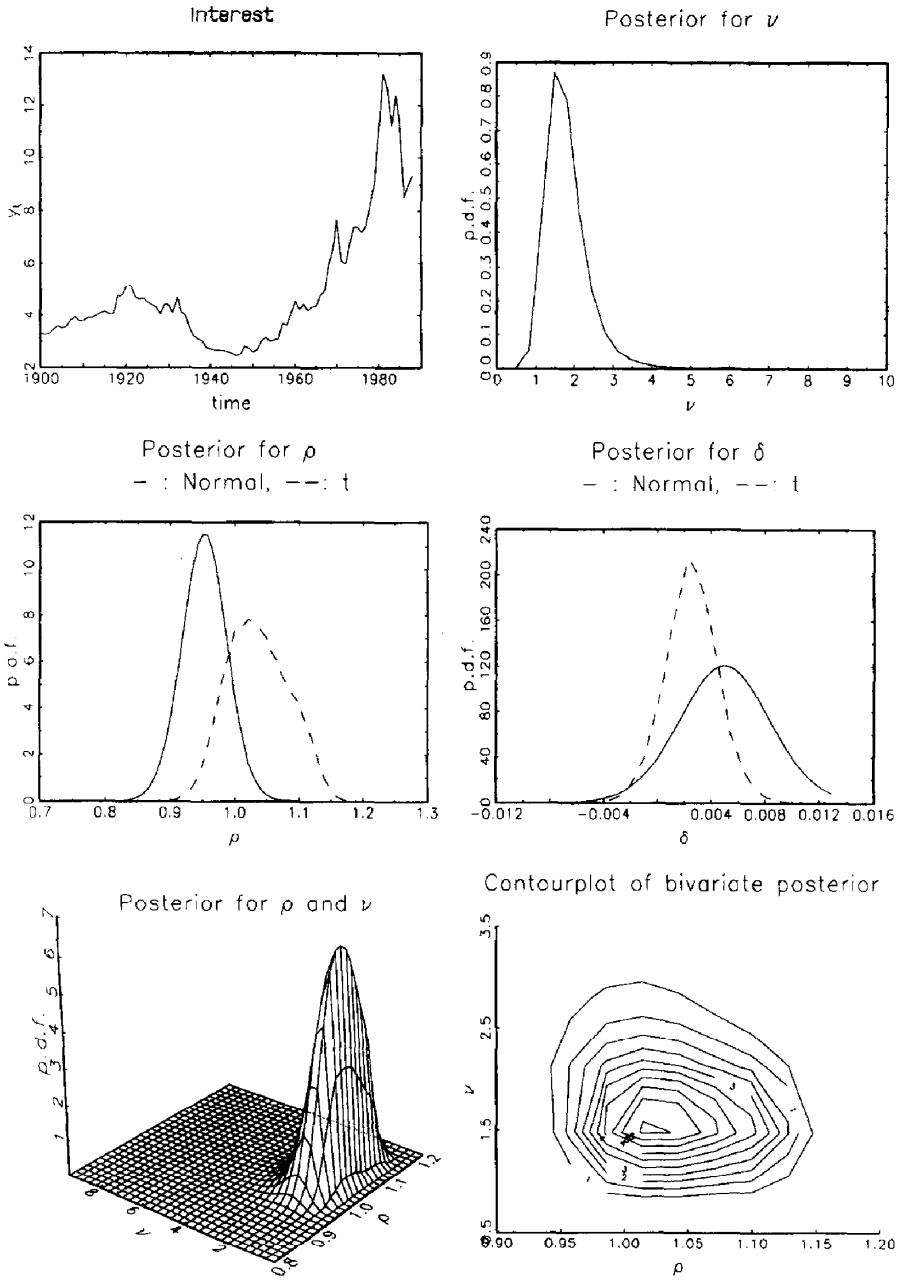


Fig. 7. Interest: Posterior results, linear model.

The posterior for δ shifts to the left. The correlation between v and ρ , given in Table 3, is negative. Also, the posterior for v is concentrated on the interval (0, 5). Negligible posterior weight is given to values of v exceeding 5, providing strong evidence against the assumption of iid normal innovations. With the exception of the velocity series all Nelson/Plosser series reveal either a negligible or a negative correlation between ρ and v .

The Nelson/Plosser series have also been analyzed by Lucas (1995a) and Geweke (1993). Lucas, using the MM estimator (Yohai, 1987), obtains mixed results. The unit root hypothesis is rejected for 4 of the 14 series. An explanation of the difference between his results and the ones obtained here is as follows. Most of the Nelson/Plosser series are characterized by the presence of a structural break (see, e.g., Perron, 1989; Zivot and Phillips, 1991). The MM estimator can cope with a large number of outliers. 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 also differ from our 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 and model specification. To analyze the differences in some detail, consider the 'structural' model parameterization used by Geweke (which is due to Schotman and van Dijk, 1993):

$$y_t - \mu - \delta t = u_t, \quad \alpha(L)u_t = \varepsilon_t, \quad (17)$$

for $t = 1, \dots, T$, where $\alpha(L)$ is a polynomial in the lag operator L , and $\{\varepsilon_t\}$ is a set of iid Student- t innovations with v degrees of freedom and variance parameter σ^2 . Geweke (1993) shows that an equivalent assumption is given by $\varepsilon_t \sim N(0, \sigma^2 \omega_t)$ with $v/\omega_t \sim \chi^2(v)$, $t = 1, \dots, T$.

In case of an AR(3) process, the polynomial $\alpha(L)$ can be decomposed as $\alpha(L) = (1 - \rho)L + (1 - \phi_1 L - \phi_2 L^2)(1 - L)$. Now, combining the prior specification of Geweke (1993)⁷ with respect to v and the prior specification for the remaining parameters of Schotman and van Dijk (1991a,b, 1993), we obtain:

$$\begin{aligned} \pi(\sigma) &\propto \sigma^{-1}, \\ \pi(\rho) &= \begin{cases} \frac{1}{2} A^{-1}, & \rho \in [1 - A, 1), \quad A > 0, \\ \frac{1}{2}, & \rho = 1, \end{cases} \\ \pi(\delta, \phi_1, \phi_2) &\propto 1, \end{aligned} \quad (18)$$

⁷The complete specification of Geweke differs by considering an AR(5) model with an increasing prior on ρ and normal priors on the other location parameters.

$$\pi(\mu | \rho, \sigma, \omega_0) \sim N(y_0, \sigma^2 \omega_0 / (1 - \rho^2)),$$

$$\pi(v/\omega_t | v) \sim \chi^2(v), \quad t = 0, \dots, T.$$

For a discussion and justification of the first four priors, we refer to Schotman and van Dijk (1991a, b). Note that the prior for ρ enables us to compute the posterior odds in favor of difference stationarity. To avoid the Lindley paradox, A is chosen such that $[1 - A, 1)$ is the 99% HPD region under stationarity. Prior odds are equal to one. Further, to take account of the fact that we consider the Student- t instead of the Gaussian likelihood, the scaling factor ω_0 is included in the conditional prior for μ . Finally, we consider the following prior for the degrees of freedom parameter:

$$\pi(v) \propto \exp(-\lambda v), \quad v \in (0, v^*]. \quad (19)$$

If $\lambda = 0$, this corresponds to the flat prior considered in (13). The posterior, obtained by multiplying the likelihood of (17) with the prior densities (18) and (19), is evaluated using the Gibbs sampler. Results, both for the original (1970) and the extended (1988) Nelson/Plosser series, are reported in Table 4, for the case of $\lambda = 0$. Results (not reported here) are qualitatively similar if we consider the exponential prior for the degrees of freedom parameter: $\pi(v) \propto \exp(-v)$. Of course, this prior results in lower posterior means for v .

For seven series, real GNP, real GNP per capita, employment, consumer price index, wages, velocity, and S&P 500, dropping the normality assumption results in a decrease of the posterior odds in favor of difference stationarity, both for the 1970 and the 1988 sample. For the 1970 sample the posterior correlation between v and ρ for these seven series, with the exception of wages, is positive. This supports Geweke's findings. For the series of unemployment, the GNP deflator, and the interest rate, abandoning the assumption $v = \infty$ results in a larger odds ratio for both samples. For the other series, the results are mixed. However, considering the 1988 instead of the 1970 sample results in a (stronger) negative correlation between v and ρ for almost all series. As the results for the employment and interest rate series show, this is not always accompanied by a larger posterior odds ratio.

We conclude from Table 4 that the difference between Geweke's and our results can, to a large extent, be attributed to the different samples under consideration and much less to the model specification. This is also demonstrated by Fig. 8. This figure gives the interest rate series as well as the posterior expectation of the vector of scale factors, ω_t . We see that the largest scale factors, corresponding to large residuals, are concentrated in the final years of the sample. This graphically explains the large shift in the posterior mean of v from 8.86 to 1.68, reported in Table 4.

Our results, 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

Table 4
Posterior moments and posterior odds, structural representation of the autoregressive model

Series	1970					1988						
	$E_t(\rho)$	$E_t(v)$	$R_{\rho,v}$	K_t	$E_N(\rho)$	K_N	$E_t(\rho)$	$E_t(v)$	$R_{\rho,v}$	K_t	$E_N(\rho)$	K_N
Real GNP	0.86 (0.07)	10.19 (5.07)	0.01	0.52	0.87 (0.07)	0.61	0.86 (0.06)	7.41 (4.50)	0.04	0.35	0.87 (0.06)	0.43
Nominal GNP	0.94 (0.04)	6.31 (4.31)	-0.01	1.24	0.93 (0.04)	1.29	0.97 (0.02)	3.80 (2.29)	-0.07	1.99	0.96 (0.03)	1.87
Real GNP per capita	0.86 (0.07)	10.70 (5.02)	0.03	0.52	0.86 (0.07)	0.66	0.85 (0.06)	7.72 (4.82)	-0.02	0.28	0.86 (0.06)	0.45
Industrial production	0.87 (0.06)	7.38 (4.15)	-0.07	0.48	0.87 (0.06)	0.53	0.89 (0.06)	6.89 (3.98)	-0.16	0.81	0.88 (0.06)	0.47
Employment	0.89 (0.05)	10.18 (5.25)	0.05	0.51	0.90 (0.05)	0.74	0.89 (0.05)	6.77 (4.30)	-0.04	0.37	0.89 (0.05)	0.48
Unemployment	0.84 (0.08)	9.99 (4.98)	-0.14	0.74	0.80 (0.09)	0.53	0.86 (0.07)	6.24 (4.13)	-0.24	0.46	0.80 (0.08)	0.34
GNP deflator	0.97 (0.02)	2.52 (1.13)	-0.03	3.07	0.94 (0.03)	1.26	0.99 (0.01)	2.93 (1.26)	-0.12	2.80	0.98 (0.01)	2.35
Consumer price index	0.99 (0.01)	2.09 (0.86)	0.14	1.96	0.98 (0.01)	2.05	0.99 (0.00)	2.71 (0.98)	-0.00	2.16	0.99 (0.01)	2.47
Wages	0.94 (0.04)	5.32 (4.45)	-0.04	0.88	0.94 (0.04)	1.20	0.97 (0.02)	2.76 (1.93)	-0.10	1.43	0.96 (0.02)	1.54
Real wages	0.90 (0.06)	10.75 (5.24)	0.08	1.16	0.91 (0.06)	1.42	0.96 (0.03)	12.15 (4.73)	-0.12	2.57	0.96 (0.03)	1.92
Money	0.93 (0.03)	5.15 (3.76)	0.13	0.33	0.93 (0.03)	0.55	0.96 (0.02)	5.88 (3.67)	-0.03	0.74	0.95 (0.02)	0.65
Velocity	0.96 (0.03)	11.16 (5.00)	-0.04	1.61	0.96 (0.03)	1.87	0.97 (0.02)	6.94 (4.57)	0.11	1.51	0.98 (0.02)	2.39
Interest	0.98 (0.02)	8.86 (5.12)	-0.04	4.17	0.98 (0.02)	3.54	0.98 (0.01)	1.68 (0.47)	-0.15	2.88	0.96 (0.03)	1.89
S&P 500	0.95 (0.04)	11.50 (4.54)	-0.05	1.43	0.94 (0.04)	1.79	0.96 (0.03)	11.26 (4.34)	0.04	1.59	0.96 (0.03)	2.07

$E_t(\rho)$ and $E_t(v)$ are respectively the posterior expectation of the unit root parameter and the degrees of freedom parameter. The heading K indicates the posterior odds ratio in favor of difference stationarity. The posterior correlation between ρ and v is given in the column labeled $R_{\rho,v}$. Subindices 'N' and 't' denote results based on the Normal ($v = \infty$) respectively the Student-t ($v \leq 20$) likelihood.

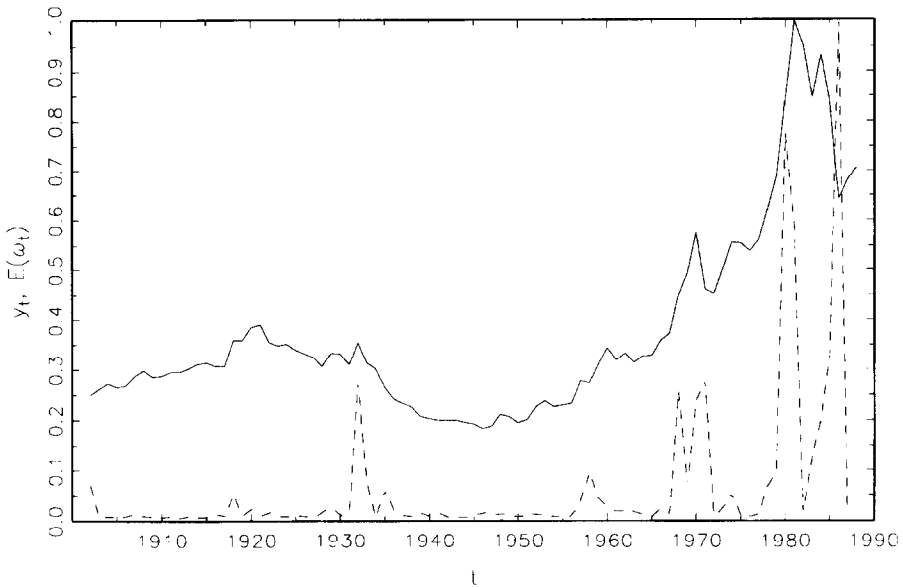


Fig. 8. Interest: Series and weights (scaled).

finding supports our theoretical results of Sections 2 and 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 we reach the conclusion that, both for the 1970 and the 1988 sample, for six series the posterior odds ratio indicate (trend) stationarity and for eight series they favor a unit root. These results are relatively robust with respect to the specification of the prior for ν .

6. Concluding remarks

In this paper we examined the effect of outliers in the data on unit root inference. 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 under a Gaussian likelihood is caused by the nonrobustness of the OLS estimator (which equals the posterior mean in a flat prior Bayesian analysis). This estimator has an unbounded influence function. We showed that

the influence function of the maximum likelihood estimator which is based on a Student- t likelihood with finite degrees of freedom is bounded. Therefore, the Dickey–Fuller t test based on this estimator is less sensitive to aberrant observations than the DF t test based on OLS. Critical values for the test are computed by means of simulation. Its (in)sensitivity to outliers is illustrated by using both simulated and empirical data. As an additional result we have 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, we argued that replacing the Gaussian likelihood by an iid Student- t likelihood results in posteriors that are less sensitive to outlying observations in the data. A proper uniform prior for the degrees of freedom parameter is 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 our theoretical results. For these series, a negative correlation between the degrees of freedom parameter and the unit root parameter was found. The results for the Nelson/Plosser series are less clear. The Bayesian analysis of these series yields the result that six out of fourteen series are probably stationary. The dynamics of these series have yet to be investigated for the presence of structural changes next to the presence of outliers. We attribute the difference between our result and the positive correlation between the unit root and the degrees of freedom parameter found by Geweke (1993) to the different samples that are used in both studies.

Finally, we note that our 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 procedures can be assessed. Moreover, the maximum likelihood estimator based upon the Student- t distribution has a bounded influence function, but only just. An alternative might be to consider the fraction of outliers an estimator can cope with. This leads to the consideration of high breakdown estimators, as is done in Lucas (1995a). However, the present (low breakdown) estimator has its own merits. It is easily calculated and provides at least some protection against outliers. In a Bayesian framework it is of interest to compare the iid Student- t model with a finite mixture model which is only partially specified. This is also a topic for further research.

Appendix

This appendix discusses the boundedness of the influence curve (IC) 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 iid. For ease of exposition it is assumed that the mean and variance of ε_t are known and equal to 0 and 1, respectively. If the variance parameter is unknown, estimation strategies similar to the ones used in the iid setting can be used (Hampel et al., 1986, p. 105). Further, consider the AO model (2) with $\xi_t \equiv \zeta$ and z_t iid. Since the formal definition of the IC is of an asymptotic nature (Martin and Yohai, 1986), we define the MLT estimator as the functional $\hat{\phi}(F_y^\gamma) \rightarrow \mathbf{R}$ that solves

$$\int \frac{v\varepsilon}{v + \varepsilon^2} y_0 dF_y^\gamma(y) = 0, \tag{20}$$

with $\varepsilon = y_1 - \hat{\phi}(F_y^\gamma)y_0$, $y = (y_1, y_0)'$, and F_y^γ the cumulative distribution function of y , given that $P(z_t = 1) = \gamma$. The IC of the MLT estimator is given by the limit in (4). We assume that the regularity conditions given in Martin and Yohai (1986) hold. We can prove the following proposition:

Proposition 1'. The IC of the estimator $\hat{\phi}$, implicitly defined in (20), under the isolated AO model with $\xi_t \equiv \zeta$, equals

$$IC(\zeta, \hat{\phi}\{F_y^\gamma\}) = -C^{-1} \int \frac{\varepsilon_1 - \phi\zeta}{v + (\varepsilon_1 - \phi\zeta)^2} (y_0 + \zeta) dF_y^0(y),$$

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

$$C = \int \frac{\varepsilon_1^2 - v}{[\varepsilon_1^2 + v]^2} y_0^2 dF_y^0(y).$$

Proof. It is easily verified that the MLT estimator defined above belongs to the class of $\tilde{\Psi}$ estimators used by Martin and Yohai (1986). It is also straightforward to check that the G_j^k quantities from their formula satisfy the following restrictions: $G_0^1 = 0$ and $G_j^1 = 0$ for all $j \geq 2$. More

$$G_1^1 = \int \frac{\varepsilon_1 - \phi\zeta}{v + (\varepsilon_1 - \phi\zeta)^2} (y_0 + \zeta) dF_y^0(y).$$

Therefore, the proposition follows straightforwardly by applying their Theorem 4.2. \square

References

Belsley, David A., Edwin Kuh, and Roy E. Welsch, 1980, Regression diagnostics: Identifying influential data and sources of collinearity (Wiley, New York, NY).
 Berger, James O., 1985, Statistical decision theory and Bayesian analysis (Springer Verlag, Berlin).

- Bustos, Oscar H. and Victor J. Yohai, 1986, Robust estimates for ARMA models, *Journal of the American Statistical Association* 81, 155–168.
- Campbell, John Y. and Pierre Perron, 1991, Pitfalls and opportunities: What macroeconomists should know about unit roots, *NBER Macroeconomics Annual*, 141–200.
- DeJong, David N. and Charles H. Whiteman, 1991a, Reconsidering 'Trends and random walks in macroeconomic time series', *Journal of Monetary Economics* 28, 221–254.
- DeJong, David N. and Charles H. Whiteman, 1991b, The temporal stability of dividend and stock prices: Evidence from the likelihood function, *American Economic Review* 81, 600–617.
- Dickey, David A. and Wayne A. Fuller, 1979, Distribution of estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* 74, 427–431.
- Franses, Philip-Hans and Niels Haldrup, 1994, The effects of additive outliers on tests for unit roots and cointegration, *Journal of Business and Economic Statistics* 12, 471–478.
- Fuller, Wayne A., 1976, *Introduction to statistical time series* (Wiley, New York, NY).
- Geweke, John, 1993, Bayesian treatment of the independent Student-*t* linear model, *Journal of Applied Econometrics*, forthcoming.
- Gouriéroux, Christian, Alain Montfort, and A. Trognon, 1984, Pseudo maximum likelihood methods: Theory, *Econometrica* 52, 681–700.
- Hampel, Frank R., Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel, 1986, *Robust statistics: The approach based on influence functions* (Wiley, New York, NY).
- Helmer, Richard M. and Johny K. Johansson, 1977, An exposition of the Box–Jenkins transfer function analysis with an application to the advertising–sales relationship, *Journal of Marketing Research* 14, 227–239.
- Hop, J. Peter and Herman K. van Dijk, 1992, SISAM and MIXIN: Two algorithms for the computation of posterior moments and densities using Monte Carlo integration, *Computer Science on Economics and Management* 5, 183–220.
- Huber, Peter J., 1981, *Robust statistics* (Wiley, New York, NY).
- Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lütkepohl, and Tsoung-Chao Lee, 1988, *Introduction to the theory and practice of econometrics* (Wiley, New York, NY).
- Kleibergen, Frank and Herman K. van Dijk, 1993, Nonstationarity in GARCH models: A Bayesian analysis, *Journal of Applied Econometrics* 8, S41–S61.
- Knight, Keith, 1991, Limit theory for *M*-estimates in an integrated infinite variance process, *Econometric Theory* 7, 200–212.
- Künsch, Hans, 1984, Infinitesimal robustness for autoregressive processes, *Annals of Statistics* 12, 843–863.
- Leamer, Edward E., 1978, *Specification searches* (Wiley, New York, NY).
- Lucas, André, 1992, Nonrobustness of the maximum likelihood estimator using *t* distributions, Unpublished document (Erasmus University, Rotterdam).
- Lucas, André, 1995a, An outlier robust unit root test with an application to the extended Nelson–Plosser data, *Journal of Econometrics* 66, 153–173.
- Lucas, André, 1995b, Unit root tests based on *M* estimators, *Econometric Theory* 11, 331–346.
- Maronna, Ricardo A. and Victor J. Yohai, 1991a, The breakdown point of simultaneous general *M* estimates of regression and scale, *Journal of the American Statistical Association* 86, 699–703.
- Martin, R. Douglas, 1981, Robust methods for time series, in: David F. Findley, ed., *Applied time series analysis II* (Academic Press, New York, NY) 683–759.
- Martin, R. Douglas and Victor J. Yohai, 1986, Influence functionals for time series, *Annals of Statistics* 14, 781–818.
- Nelson, Charles R. and Charles I. Plosser, 1982, Trends and random walks in macroeconomic time series: Some evidence and implications, *Journal of Monetary Economics* 10, 139–162.
- Perron, Pierre, 1989, The great crash, the oil price shock, and the unit root hypothesis, *Econometrica* 57, 1361–1401.

- Perron, Pierre and Timothy J. Vogelsang, 1992, Nonstationarity and level shifts with an application to purchasing power parity, *Journal of Business and Economic Statistics* 10, 301–320.
- Phillips, Peter C.B., 1987, Time series regression with a unit root, *Econometrica* 55, 277–301.
- Phillips, Peter C.B., 1991, To criticize the critics: An objective Bayesian analysis of stochastic trends, *Journal of Applied Econometrics* 6, 333–364.
- Phillips, Peter C.B. and Pierre Perron, 1988, Testing for a unit root in time series regression, *Biometrika* 75, 335–346.
- Prucha, Ingmar R. and Harry H. Kelejian, 1984, The structure of simultaneous equation estimators: A generalization towards nonnormal disturbances, *Econometrica* 52, 721–736.
- Rousseeuw, Peter J., 1981, A new infinitesimal approach to robust estimation, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 56, 127–132.
- Schotman, Peter C. and Herman K. van Dijk, 1991a, On Bayesian routes to unit roots, *Journal of Applied Econometrics* 6, 387–401.
- Schotman, Peter C. and Herman K. van Dijk, 1991b, A Bayesian analysis of the unit root in real exchange rates, *Journal of Econometrics* 49, 195–238.
- Schotman, Peter C. and Herman K. van Dijk, 1993, Posterior analysis of possibly integrated time series with an application to real GNP, in: P. Caines, J. Geweke, and M. Taqqu, eds., *New directions in time series analysis, Part II* (Springer Verlag, Berlin).
- Sims, Christopher A., 1988, Bayesian scepticism on unit root econometrics, *Journal of Economic Dynamics and Control* 12, 463–474.
- Smith, Adrian F.M., 1981, Bayesian approaches to outliers and robustness, in: Jean-Pierre Florens, Michel Mouchart, J.P. Raoult, and L. Simar, eds., *Specifying statistical models* (Springer Verlag, Berlin) 13–35.
- West, Mike, 1984, Outlier models and prior distributions in Bayesian linear regression, *Journal of the Royal Statistical Society Series B* 46, 431–439.
- White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817–838.
- Yohai, Victor J., 1987, High breakdown-point and high efficiency robust estimates for regression, *Annals of Statistics* 15, 642–656.
- Zellner, Arnold, 1976, Bayesian and non-Bayesian analysis of the regression model with multivariate Student-*t* error terms, *Journal of the American Statistical Association* 71, 400–405.
- Zivot, Eric and Peter C.B. Phillips, 1991, A Bayesian analysis of trend determination in economic time series, Discussion paper 1002 (Cowles Foundation, Yale University, New Haven, CT).