Chapter 3

Robustness Properties of the Student t Based Pseudo Maximum Likelihood Estimator

In Chapter 2, some concepts from the robustness literature were introduced. An important concept was the influence function. In the present chapter, the influence function is used to assess the robustness properties of the Student t based pseudo maximum likelihood estimator with estimated degrees of freedom parameter. This estimator is often employed in the econometric literature as a first relaxation of the usual normality assumption (see, e.g., de Jong et al. (1992), Kleibergen and van Dijk (1993), Prucha and Kelejian (1984), and Spanos (1994)). In this chapter I show that the estimator is nonrobust in the sense that it has an unbounded influence function and an unbounded change-of-variance function[1] if the degrees of freedom parameter is estimated rather than fixed *a priori*. This result can already be established in the setting of the simple location/scale model and has obvious consequences for other robust estimators that estimate the tuning constant from the data.

At the basis of the above results lies the observation that the score function for the pseudo maximum likelihood estimator for the degrees of freedom parameter is unbounded. As a result, the influence functions of the degrees of freedom and scale estimator are also unbounded. In contrast, the influence function of the location parameter is bounded due to the block-diagonality of the Fisher information matrix under the assumption of symmetry. The change-of-variance function of the estimator for the location parameter, however, is unbounded, suggesting that standard inference procedures for the location parameter are nonrobust if they are based on the Student t pseudo maximum likelihood estimator with estimated degrees of freedom parameter. These results illustrate two basic points. First, one should carefully distinguish between parameters of interest and nuisance parameters when assessing the robustness properties of statistical procedures. Second, if a parameter can be estimated robustly in the sense that an estimator can be constructed with a bounded

[1] See Section 3.2 for a definition of the change-of-variance function.

influence function, this does not imply that the corresponding inference procedure for that parameter is also robust.

The chapter is set up as follows. Section 3.1 introduces the model and the pseudo maximum likelihood estimator based on the Student t distribution. Section 3.2 derives the influence function and change-of-variance function and provides a simple finite sample approximation to these functions. Section 3.3 provides a simulation based comparison of several robust and nonrobust estimators for the simple model of Section 3.1. Section 3.4 concludes this chapter.

3.1 The Model and Estimator

Consider the simple location/scale model

$$y_t = \mu + \sigma\varepsilon_t, \tag{3.1}$$

where μ is the location parameter, σ is the scale parameter, $\{\varepsilon_1, \ldots, \varepsilon_T\}$ is a set of i.i.d. drawings with unit scale, and T denotes the sample size. Model (3.1) is extremely simple, but it suffices to illustrate the problems studied in this chapter. The difficulties that arise for (3.1) also show up in more complicated models.

The usual way of estimating μ and σ in (3.1) is by means of ordinary least-squares (OLS). This produces the arithmetic sample mean and the sample standard deviation as estimators for μ and σ, respectively. As described in the previous chapter, the standard OLS estimator is sensitive to outliers in the data. One big outlier is enough to corrupt the estimates completely. In order to reduce the sensitivity of the results to outliers,[2] the class of M estimators was proposed by Huber (1964, 1981). In this chapter a specific element from this class is studied, namely the pseudo maximum likelihood estimator based on the Student t distribution (MLT estimator).

As was described in Section 2.3, an M estimator minimizes

$$\sum_{t=1}^{T} \rho(y_t; \mu, \sigma; \nu) \tag{3.2}$$

with respect to μ and σ, with ρ denoting some smooth function and $\nu > 0$ denoting a user specified tuning constant. In this chapter

$$\rho(y_t; \mu, \sigma; \nu) = -\ln\left(\frac{\Gamma(\frac{\nu+1}{2})}{\sigma\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}} \left(1 + \frac{(y_t-\mu)^2}{\nu\sigma^2}\right)^{-\frac{1}{2}(\nu+1)} \right), \tag{3.3}$$

[2] Methods for reducing the possibly bad effects of outliers have a long history, as can be seen from the references in Section 1.3 of Hampel et al. (1986). One of these references dates back to Greek antiquity.

with $\Gamma(\cdot)$ denoting the gamma function, such that ν can be interpreted as the degrees of freedom parameter of a Student t distribution. The first order conditions for the MLT estimators for μ and σ are $\sum_{t=1}^{T}\psi_\mu(y_t)=0$ and $\sum_{t=1}^{T}\psi_\sigma(y_t)=0$, respectively, with

$$\psi_\mu(y_t) = -\frac{(\nu+1)(y_t-\mu)}{\nu\sigma^2+(y_t-\mu)^2} \tag{3.4}$$

and

$$\begin{aligned}\psi_\sigma(y_t) &= \sigma^{-1}(1+(y_t-\mu)\psi_\mu(y_t)) \\ &= \frac{\nu}{\sigma}\cdot\frac{\sigma^2-(y_t-\mu)^2}{\nu\sigma^2+(y_t-\mu)^2}\end{aligned} \tag{3.5}$$

Although ν in the above setup can be regarded as a user specified tuning constant that determines the degree of robustness of the M estimator, it is not unusual to estimate ν together with μ and σ (see the references below). Several estimators for ν are available from the literature. Using the pseudo log likelihood ρ, the most obvious estimator for ν is given by the (pseudo) maximum likelihood (ML) estimator, i.e., the value $\hat{\nu}$ that solves $\sum_{t=1}^{T}\psi_\nu(y_t)=0$, with

$$\psi_\nu(y_t) = -\frac{1}{2}\left(\gamma(\frac{\nu+1}{2})-\gamma(\frac{\nu}{2})-\ln(1+\frac{(y_t-\mu)^2}{\nu\sigma^2})-\frac{\sigma}{\nu}\psi_\sigma(y_t)\right), \tag{3.6}$$

$\gamma(\nu)$ the digamma function ($\gamma(\nu)=d\ln(\Gamma(\nu))/d\nu$). , and $\Gamma(\cdot)$ the gamma function. This estimator is used in, e.g., Fraser (1976), Little (1988), and Lange, Little, and Taylor (1989). Spanos (1994) used an alternative estimator for ν based on the sample kurtosis coefficient. A third estimator used in the literature is the one proposed by Prucha and Kelejian (1984). They embed the family of Student t distributions in a more general class of distributions. Their estimator for ν uses an estimate of the first absolute moment of the disturbance term. Yet another possibility for estimating ν is by using tail-index estimators for the distribution of y_t (see, e.g., Groenendijk, Lucas, and de Vries (1995)).[3]

It is easily checked that the estimator of Spanos (1994) for ν is nonrobust. The nonrobustness of this estimator follows from the nonrobust estimation of the kurtosis coefficient. Similarly, the estimator used by Prucha and Kelejian

[3] It is perhaps useful to note here that (3.3) is closely linked to the assumption of i.i.d. Student t distributed errors in (3.1). Alternatively, one could study (3.1) under the assumption that $(\varepsilon_1,\ldots,\varepsilon_T)$ has a multivariate Student t distribution with diagonal precision matrix, such that the errors are uncorrelated rather than independent. Zellner (1976, p. 402) proved that in this setting β, σ, and ν cannot be estimated simultaneously by means of ML if only one realization of $\{y_t\}_{t=1}^{T}$ is available. One way to solve this problem is by using several realizations of $\{y_t\}_{t=1}^{T}$, as is possible in a panel data context. One can then construct a suitable estimator for the degrees of freedom parameter (see, e.g., Sutradhar and Ali (1986)).

(1984, p. 731) for estimating the first absolute moment is also not robust to outliers, which results in a nonrobust estimator for ν. Finally, tail-index estimators in their usual implementation are intrinsically nonrobust, because they concentrate on observations in the extreme quantiles of the distribution. So the only remaining candidate[4] for robust estimation of ν is the MLT estimator. The next section demonstrates, however, that the MLT estimator for ν is also nonrobust.

3.2 A Derivation of the Influence Function

Define $\theta = (\mu, \sigma, \nu)^\top$ and $\psi(y_t) = (\psi_\mu(y_t), \psi_\sigma(y_t), \psi_\nu(y_t))^\top$, then the MLT estimator $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\nu})^\top$ solves $\sum_{t=1}^{T} \psi(y_t) = 0$. This section presents the IF of the MLT estimator. First, some additional notation is needed. Let $\psi'(y_t) = \partial\psi(y_t)/\partial\theta^\top$, with

$$\psi'(y_t) = \begin{pmatrix} \psi_{\mu\mu}(y_t) & \psi_{\mu\sigma}(y_t) & \psi_{\mu\nu}(y_t) \\ \psi_{\mu\sigma}(y_t) & \psi_{\sigma\sigma}(y_t) & \psi_{\sigma\nu}(y_t) \\ \psi_{\mu\nu}(y_t) & \psi_{\sigma\nu}(y_t) & \psi_{\nu\nu}(y_t) \end{pmatrix},$$

and

$$\begin{aligned} \psi_{\mu\mu}(y_t) &= (\nu+1)(\nu\sigma^2 - (y_t-\mu)^2)/(\nu\sigma^2 + (y_t-\mu)^2)^2, \\ \psi_{\mu\sigma}(y_t) &= 2\nu\sigma(\nu+1)(y_t-\mu)/(\nu\sigma^2 + (y_t-\mu)^2)^2, \\ \psi_{\mu\nu}(y_t) &= (y_t-\mu)(\sigma^2 - (y_t-\mu)^2)/(\nu\sigma^2 + (y_t-\mu)^2)^2, \\ \psi_{\sigma\sigma}(y_t) &= ((y_t-\mu)\psi_{\mu\sigma}(y_t) - \psi_\sigma(y_t))/\sigma, \\ \psi_{\sigma\nu}(y_t) &= (y_t-\mu)\psi_{\mu\nu}(y_t)/\sigma, \end{aligned}$$

[4] Of course, one can object that the other estimators can easily be extended such that they become robust to at least some extent. For example, one can try to estimate the first absolute moment of the errors in a robust way, thus 'robustifying' the estimator of Prucha and Kelejian (1984). The problem with this approach is that one wants a consistent estimator for ν for the whole class of Student t distributions, as opposed to an estimator that is constistent for only one specific Student t distribution. If one constructs a robust estimator for the first absolute moment by downweighting extreme observations, one will end up estimating ν too high, in general. Therefore, the estimator based upon the robustly estimated first absolute moment has to be multiplied by a correction constant in order to make it a consistent estimator for ν. (This can be compared with the multiplication by 1.483 for the median absolute deviation in order to make it a consistent estimator for the standard deviation of the Gaussian distribution.) The problem now becomes one of choosing the appropriate multiplication constant. If one chooses this constant such that the estimator is consistent for, say, $\nu = 5$, the estimator will in general be inconsistent for all other values of ν.

A different objection can be raised against the use of tail-index estimators. In their usual implementation, these are based on averages of observations in the extreme quantiles of the sample. One can robustify these estimators to some extent by replacing the mean by, e.g., the median. As the number of observations on which the tail-index estimators are based is usually quite small, even such revised tail-index estimators have an intrinsically low breakdown point. Therefore, they are not considered further in this chapter.

$$\psi_{\nu\nu}(y_t) = \frac{1}{2}\left(\frac{1}{2}\gamma'(\frac{\nu}{2}) - \frac{1}{2}\gamma'(\frac{\nu+1}{2}) - \frac{(y_t-\mu)^2/\nu}{\nu\sigma^2+(y_t-\mu)^2} - \frac{\sigma}{\nu^2}\psi_\sigma(y_t) + \frac{\sigma}{\nu}\psi_{\sigma\nu}(y_t)\right),$$

with $\gamma'(\nu)$ the derivative of the digamma function. The following assumption is made.

Assumption 3.1 *$\{\varepsilon_t\}$ is an i.i.d. process; ε_t has a cumulative distribution function $F(\varepsilon_t)$ with probability density function $f(\varepsilon_t)$; the probability density function f is continuous and symmetric around zero; there exists an $\eta > 0$ such that $f(\varepsilon_t) = O(\varepsilon_t^{-(1+\eta)})$ for large values of ε_t; there exists a θ_0 such that $E(\psi(y_t)) = 0$ and $|E(\psi'(y_t))| \neq 0$, where the expectation is taken with respect to $f(\varepsilon_t)$.*

The following theorem is easily proved along the lines of Hampel, Ronchetti, Rousseeuw, and Stahel (1986, pp. 85, 101–102).

Theorem 3.1 *Given Assumption 3.1 and $\nu < \infty$, the IF of the MLT estimator $\hat{\theta}$ equals*

$$IF(y;\hat{\theta},F) = -\left(E(\psi'(y_t))\right)^{-1}\psi(y).$$

The IF of $\hat{\theta}$ is thus a linear transformation of the score vector $\psi(y_t)$. From (3.4) and (3.5) it follows that the score functions for μ and σ are bounded. The score function for ν, however, is unbounded, as is demonstrated by (3.6). The following result can be proved.

Theorem 3.2 *Given Assumption 3.1 and $\nu < \infty$, the IF of the MLT estimator for μ is bounded, while the IF's of the MLT estimators for σ and ν are unbounded.*

Proof. The unboundedness of the IF's of $\hat{\sigma}$ and $\hat{\nu}$ follows directly from the unboundedness of $\psi_\nu(y)$ and the fact that the $(2,3)$-element of $(E(\psi'(y_t)))^{-1}$ is nonzero, in general. Moreover, from the symmetry of $f(\varepsilon_t)$ it follows that $E(\psi_{\mu\sigma}(y_t)) = 0$ and $E(\psi_{\mu\nu}(y_t)) = 0$, implying that the IF of $\hat{\mu}$ only depends on y through the score function for μ. As the score function for μ is bounded for finite ν, the theorem is proved. □

Figure 3.1 displays some IF's evaluated at the standard Student t distribution ($\mu = 0$, $\sigma = 1$) with 1, 3, and 5 degrees of freedom (ν). The IF of $\hat{\mu}$ has a redescending shape: if $|y|$ becomes large, the IF approaches zero. This means that large outliers have a small influence on the MLT estimator for μ. The IF's of $\hat{\sigma}$ and $\hat{\nu}$ are negative and decreasing for sufficiently large values of y. This means, for example, that the estimate of ν is negatively biased if there is a single large outlier in the sample. The negative bias is more severe for higher values of ν. For $\nu = 1$, the IF of $\hat{\nu}$ is almost flat. This is due to the fact that the Cauchy distribution ($\nu = 1$) is already very fat-tailed. Observations as large as $y = 12$ are not unreasonable if a Cauchy distribution is generating the data. Therefore, the IF has a relatively flat shape for this low value of ν.

The effect of outliers on the scale estimator reveals a similar pattern. Finally, it is interesting to see that for small values of $|y|$ the estimators for both the scale and degrees of freedom parameter demonstrate a negative bias. This is due to the effect of centrally located observations (so called inliers, as opposed to outliers).

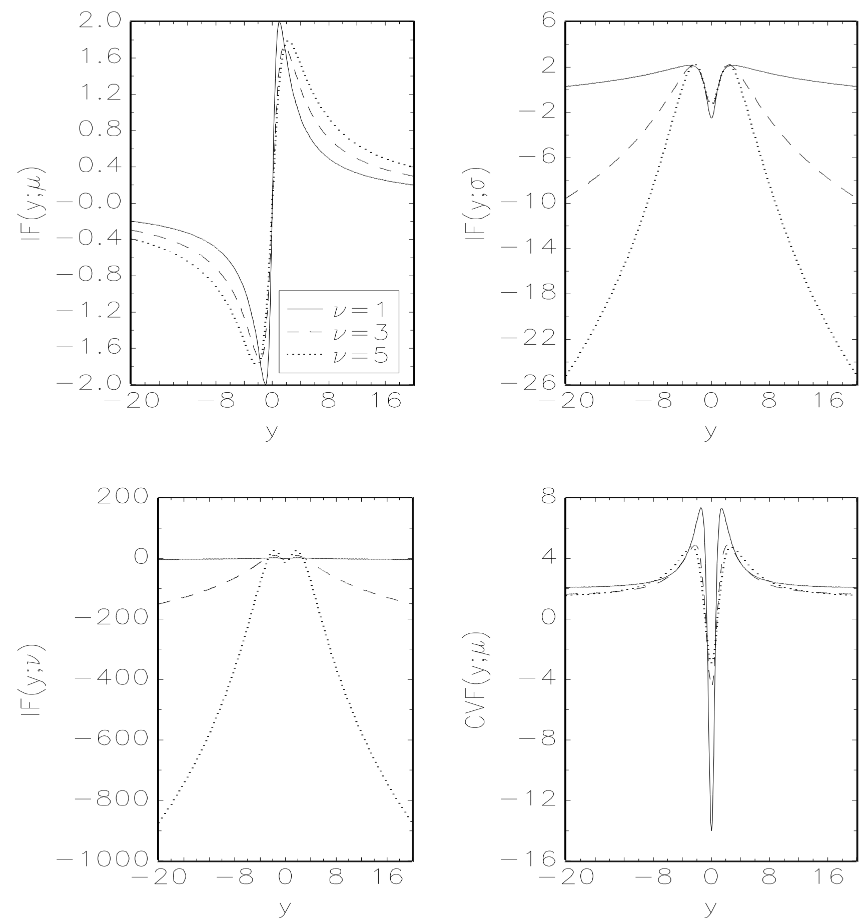

Figure 3.1.— Influence Functions and Change-of-Variance Function of the MLT Estimators for μ, σ, and ν

The unboundedness of the IF of the MLT estimator crucially depends on the fact that ν is estimated rather than fixed. If ν is considered as a tuning constant and fixed at a user specified value, the IF's of the MLT estimators for μ and σ are bounded. Therefore, it is important to specify the parameters of interest.

If ν is a parameter of interest, the nonrobustness of $\hat{\nu}$ is rather discomforting.[5] Solving the nonrobustness of the MLT estimator for ν or for any of the other estimators mentioned in Section 3.1, is, however, nontrivial. One possibility is to bound the score function for ν as in Theorem 1 of Hampel et al. (1986, p. 117). This requires that one specifies a central model for which the estimator must be consistent. As a result, one can only devise an outlier robust estimator for ν that is consistent for one specific Student t distribution, but inconsistent for any other Student t distribution (compare Footnote 4). If ν really is a parameter of interest, it seems undesirable to have an estimator that is only consistent for one specific, user specified value of ν.

In contrast, if only μ and σ are the parameters of interest, there are fewer difficulties. One can then fix ν and perform the remaining analysis conditional on ν. This corresponds to the strategy that is often followed in the robustness literature. Alternatively, one can estimate ν along with μ and σ and ignore the potential bias in the estimate of ν due to the occurrence of outliers. This second strategy is closely linked to the adaptive estimation literature (see, e.g., Hogg (1974)). An advantage of this strategy is that it allows a weighting of the observations conditional on the sample. If the sample is, for instance, approximately Gaussian, the estimate of ν will be very large and the estimator for μ will be close to the efficient estimator: the arithmetic mean. If, in contrast, there are some severe outliers in the sample, the estimate of ν will be lower and the extreme observations will be weighted less heavily.

If only μ is the parameter of interest, it is not only interesting to know whether μ can be estimated robustly, but also whether robust inference procedures can be constructed for this parameter. In order to answer this question, the sensitivity of the variance of $\hat{\mu}$ to outliers must be assessed. This can be done by means of the change-of-variance function (CVF), introduced by Rousseeuw (1981). The CVF is a similar concept as the IF. Whereas the IF measures the shift in an estimator due to an infinitesimal contamination, the CVF measures the corresponding shift in the variance of the estimator. If both the IF and the CVF of an estimator are bounded, robust inference procedures can be based on the estimator. It has already been shown that the only source of nonrobustness for the MLT estimator stems from the estimation of ν. If ν is fixed by the user, the MLT estimators for μ and σ have a bounded IF. It is rather straightforward to show that for fixed ν these estimators also have a bounded CVF. Moreover, even if ν is estimated, Theorem 3.2 states that $\hat{\mu}$ still has a bounded IF. The only interesting question that is left, therefore, is whether the estimation of ν affects the variance of the estimator $\hat{\mu}$.

In order to define the CVF of the MLT estimator, an expression for the asymptotic variance V of the estimator is needed. From Equation (4.2.2) of

[5]The nonrobustness of $\hat{\nu}$ is not alarming if the parameter is only used to check whether the sample contains outliers or exhibits leptokurtosis. In that case, the estimate of ν is only used as a diagnostic measure and has no intrinsic meaning. If, however, one is interested in ν as the degrees of freedom parameter of the Student t distribution, the nonrobustness of the estimator for ν is worrying.

Hampel et al. (1986), one obtains

$$V(\hat{\theta}, F) = E(IF(y_t; \hat{\theta}, F)IF(y_t; \hat{\theta}, F)^\top), \tag{3.7}$$

where the expectation is again taken with respect to $f(\varepsilon_t)$. Following Equation (2.5.11) of Hampel et al. (1986, p. 128), the CVF of the MLT estimator is defined as

$$CVF(y; \hat{\theta}, F) = \left. \frac{\partial V(\hat{\theta}, F^\eta)}{\partial \eta} \right|_{\eta=0}, \tag{3.8}$$

where F^η is the contaminated cumulative distribution function $F^\eta(\varepsilon_t) = (1 - \eta)F(\varepsilon_t) + \eta 1_{\{\varepsilon_t \geq y+\mu\}}(\varepsilon_t)$, with $1_A(\cdot)$ the indicator function of the set A. The following theorem establishes the unboundedness of the CVF of $\hat{\mu}$ if ν is estimated by means of the MLT estimator.

Theorem 3.3 *Let the conditions of Assumption 3.1 be satisfied and $\nu < \infty$. If θ is estimated with the MLT estimator $\hat{\theta}$, then the CVF of $\hat{\mu}$ is unbounded.*

Proof. The asymptotic variance of $\hat{\mu}$ is the $(1,1)$-element from the matrix $V(\hat{\theta}, F)$. Therefore, it only has to be shown that the $(1,1)$-element from $CVF(y; \hat{\theta}, F)$ is an unbounded function of y. Define the matrices B_1 and B_2 as

$$\begin{aligned} B_1(F^\eta) &= \int_{-\infty}^{\infty} \tilde{\psi}(y_t)(\tilde{\psi}(y_t))^\top dF^\eta(\varepsilon_t), \\ B_2(F^\eta) &= \int_{-\infty}^{\infty} \tilde{\psi}'(y_t) dF^\eta(\varepsilon_t), \end{aligned}$$

with $\tilde{\psi}$ and $\tilde{\psi}'$ defined as ψ and ψ', respectively, only with θ replaced by the functional $\hat{\theta}(F^\eta)$, where $\hat{\theta}(F^\eta)$ is the MLT estimator evaluated at the distribution F^η. Note that $\hat{\theta}(F^0) = (\mu, \sigma, \nu)^\top$. The asymptotic variance of the MLT estimator is now equal to

$$V(\hat{\theta}, F^\eta) = (B_2(F^\eta))^{-1} B_1(F^\eta)(B_2(F^\eta))^{-1}.$$

The $(1,1)$-element of the CVF of $\hat{\theta}$ is equal to the $(1,1)$-element of the matrix

$$\begin{aligned} -V_0 \left. \frac{dB_2(F^\eta)}{d\eta} \right|_{\eta=0} (B_2(F))^{-1} - (B_2(F))^{-1} \left. \frac{dB_2(F^\eta)}{d\eta} \right|_{\eta=0} V_0 \\ + (B_2(F))^{-1} \left. \frac{dB_1(F^\eta)}{d\eta} \right|_{\eta=0} (B_2(F))^{-1}, \end{aligned} \tag{3.9}$$

with $V_0 = V(\hat{\theta}, F^0)$. Due to the symmetry of $f(\varepsilon_t)$, it is easily checked that both V_0 and $B_2(F)$ are block-diagonal, with the blocks being the $(1,1)$-element and the lower-right (2×2) block. Therefore, it only has to be shown that either the $(1,1)$-element of $dB_1(F^\eta)/d\eta|_{\eta=0}$ or that of $dB_2(F^\eta)/d\eta|_{\eta=0}$ is unbounded. These elements are given by

$$(\psi_\mu(y))^2 - e_1^\top B_1(F) e_1 + E(2\psi_\mu(y_t)\psi_{\mu\mu}(y_t)) IF(y; \hat{\mu}, F) +$$

$$E(2\psi_\mu(y_t)\psi_{\mu\sigma}(y_t))IF(y;\hat{\sigma},F) + E(2\psi_\mu(y_t)\psi_{\mu\nu}(y_t))IF(y;\hat{\nu},F), \tag{3.10}$$

and

$$\psi_{\mu\mu}(y) - e_1^\top B_2(F)e_1 + E(\psi_{\mu\mu\mu}(y_t))IF(y;\hat{\mu},F) +$$
$$E(\psi_{\mu\mu\sigma}(y_t))IF(y;\hat{\sigma},F) + E(\psi_{\mu\mu\nu}(y_t))IF(y;\hat{\nu},F), \tag{3.11}$$

respectively, with $e_1^\top = (1,0,0)$ and three indices denoting third order partial derivatives, e.g., $\psi_{\mu\mu\mu}(y_t) = \partial\psi_{\mu\mu}(y_t)/\partial\mu$. Using (3.10) and (3.11), it is evident that without further assumptions the score function for ν is, in general, present in (3.9) with a nonzero loading factor. This causes the CVF of $\hat{\mu}$ to be unbounded. □

Theorem 3.3 only discusses the unboundedness of the CVF in the general case. An interesting question concerns the behavior of the CVF if the true distribution actually belongs to the Student t class. The following corollary gives the result.

Corollary 3.1 *Given the conditions of Theorem 3.3 and given that the ε_t's follow a Student t distribution with ν degrees of freedom, then the CVF of $\hat{\mu}$ is bounded.*

Proof. Without loss of generality, set $\sigma = 1$ and $\mu = 0$. It is tedious, but straightforward to show that

$$E(\psi_\mu(y_t)\psi_{\mu\sigma}(y_t)) = E(\psi_{\mu\mu\sigma}(y_t)) = -2\frac{(\nu+1)(\nu+2)}{(\nu+3)(\nu+5)},$$

and

$$E(\psi_\mu(y_t)\psi_{\mu\nu}(y_t)) = E(\psi_{\mu\mu\nu}(y_t)) = \frac{2(\nu-1)}{\nu(\nu+3)(\nu+5)}.$$

The result now follows directly from (3.9), (3.10), and (3.11). □

Figure 3.1 shows the CVF of $\hat{\mu}$ evaluated at several Student t distributions. As Corollary 3.1 predicts, this CVF is bounded. Centrally located values of y, i.e., inliers, cause a downward bias in the standard error of $\hat{\mu}$, while outliers result in a (bounded) upward bias.

Both Theorem 3.2 and 3.3 lead to the conclusion that estimating ν leads to nonrobust statistical procedures. Both the IF and the CVF are, however, defined in an asymptotic context. It might well be the case that the unboundedness of the IF and the CVF is less important in finite samples. In the next section, this is investigated by means of a Monte Carlo simulation experiment. Here, a much simpler strategy is used. Let $\{y_1,\ldots,y_{25}\}$ be a set of i.i.d. drawings from a Student t distribution with location zero, scale one, and degrees of freedom parameter ν. Construct the symmetrized sample $\{\tilde{y}_1,\ldots,\tilde{y}_{50}\}$, with $\tilde{y}_{2k} = -y_k$ and $\tilde{y}_{2k-1} = y_k$ for $k = 1,\ldots,25$. Let y be some real number and enlarge the sample with $\tilde{y}_{51} = y$. For the sample $\{\tilde{y}_t\}_{t=1}^{51}$ the MLT estimates of μ, σ, and ν can be computed, together with an estimate of the asymptotic variance of $\hat{\mu}$. This can be done for several values of ν. Figure 3.2 displays the difference between the estimated and true values of the parameters for several values of y.

The curves for $\hat{\mu}$ and $\hat{\sigma}$ reveal a qualitatively similar picture as the IF's in Figure 3.1. Large outliers have a small effect on $\hat{\mu}$, while causing a downward bias in $\hat{\sigma}$. Note that Figure 3.2 gives the finite sample approximation to the IF of $1/\hat{\nu}$ instead of $\hat{\nu}$ in order to facilitate the presentation. For $\nu = 5$, for example, the bias in $1/\hat{\nu}$ for $y = 0$ is approximately -0.1, implying that the estimate of ν is approximately 10. The curve for $1/\hat{\nu}$ shows the same pattern as the IF in Figure 3.1. Large outliers cause an upward bias in $1/\hat{\nu}$ and, thus, a downward bias in $\hat{\nu}$. Finally, also the shape of the curve showing the discrepancy between the estimated and the asymptotic standard error of $\hat{\mu}$ corresponds to the shape of the CVF of $\hat{\mu}$. Again it is seen that for $\nu = 1$ and $\nu = 3$ moderate outliers have a larger (absolute) effect on the standard error than extreme outliers.

3.3 A Numerical Illustration

This section presents the results of a small simulation experiment conducted in order to obtain insight into the finite sample behavior of the MLT estimator and several alternative estimators in a variety of circumstances. The model is always (3.1) with $\mu = 0$ and $\sigma = 1$. The estimators that are used are discussed in Subsection 3.3.1, while the different distributions for ε_t can be found in Subsection 3.3.2. Subsection 3.3.3 discusses the results. This discussion centers around the behavior of the estimators for μ, as μ has a similar interpretation for all error distribution except the χ^2 distribution. In contrast, the estimators for σ have different probability limits for different error distributions. Therefore, these estimators cannot be compared directly. This should be kept in mind when interpreting the results in Subsection 3.3.3.

3.3.1 Estimators

I consider the following seven estimators.

The first estimator uses the arithmetic mean and the ordinary standard deviation to estimate μ and σ, respectively. The standard error of the mean is estimated by the standard deviation divided by the square root of the sample size.

The second estimator uses the median and the median absolute deviation to estimate μ and σ, respectively. The median absolute deviation is multiplied by 1.483 in order to make it a consistent estimator for the standard deviation of a Gaussian distribution. The asymptotic standard error of the median is estimated by $(2\hat{f}(\tilde{\mu}))^{-1}$ (compare Hampel et al. (1986, p. 109), with $\tilde{\mu}$ denoting the median of y_t and $\hat{f}(\cdot)$ denoting a sample based estimate of the density function of the y_t's. The kernel estimator used to construct this density estimate is described when discussing the seventh estimator.

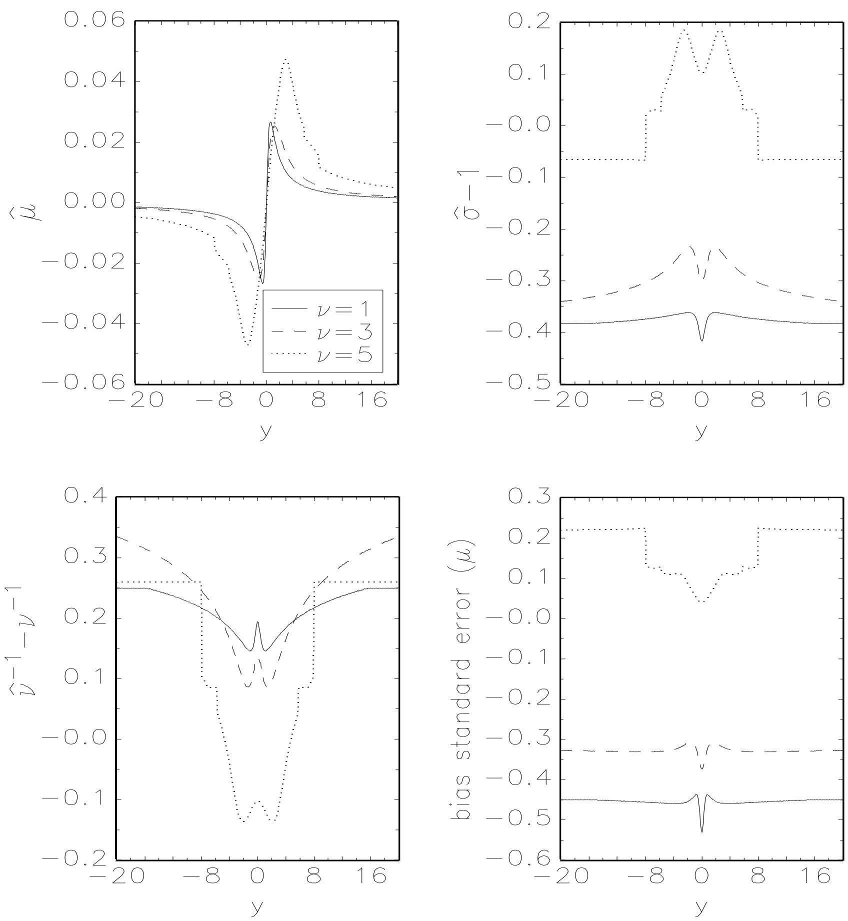

Figure 3.2.— Finite Sample Influence Curves for $\hat{\mu}$, $\hat{\sigma}$, $1/\hat{\nu}$, and the standard error of $\hat{\mu}$

The third estimator uses the MLT estimators for μ and σ with a fixed degrees of freedom parameter $\nu = 5$. This estimator is computed by means of an iteration scheme. The starting values used for μ and σ are the median and the median absolute deviation, respectively. Also the fourth through the seventh estimator below are computed by means of iteration schemes. For all estimator, the starting values mentioned above are used.

The fourth estimator is the same as the third estimator, only with $\nu = 1$ instead of $\nu = 5$.

The fifth estimator is the MLT estimator with estimated degrees of freedom parameter using (3.6). The MLT estimator for ν is restricted to the interval $[0.5, 694.678]$ in order to avoid nonconvergent behavior of the estimator.[6]

The sixth estimator uses the MLT score functions for μ and σ, but employs a different method for fixing the degrees of freedom parameter. The idea for determining ν is inspired by a method for determining the optimal amount of trimming for the trimmed mean (see Andrews et al. (1972), Hogg (1974)). The estimator is, therefore, called an adaptive estimator. For a given sample and a given value of ν, one can obtain an estimate of μ and σ and of the standard error of $\hat{\mu}$. The value of ν is then chosen such that the estimated standard error of $\hat{\mu}$ is minimized.

The seventh estimator is also adaptive, but in a more general sense, because it treats the whole error distribution $f(\varepsilon_t)$ as an (infinite-dimensional) nuisance parameter (see Manski (1984)). It can, therefore, also be called a semiparametric estimator. The ideas for this estimator are taken from Manski (1984), although the actual implementation differs in certain details, e.g., the choice of the bandwidth parameter and the kernel estimator. Given a preliminary estimate of μ, an estimate of the density function is constructed. The estimated density is then used to obtain a (nonparametric) maximum likelihood estimate of μ. The empirical mean of the squared estimated score is used to estimate the asymptotic standard error of this estimator. I use the median as the preliminary estimator.

The estimates of the density and the score function are constructed in the following way. Let $\{y_t\}_{t=1}^{T}$ denote the observed sample and let $\tilde{\mu}$ denote the preliminary estimate of μ. Construct $\tilde{u}_t = y_t - \tilde{\mu}$ and let $\tilde{u}_{t:T}$ denote the corresponding ascending order statistics. In order to protect the procedure against outliers, I remove the upper and lower 0.05-quantiles of the sample.

[6] The upper bound of 694.678 is due to the algorithm that was used for computing the estimator. The half-line $0.5 \leq \nu \leq \infty$ was mapped in a one-to-one way to the interval $0.5 \leq g(\nu) \leq 2$, with $g(\nu) = \nu$ for $0.5 \leq \nu \leq 1$ and $g(\nu) = 2 - \nu^{-1}$ for $\nu > 1$. Next, $g(\nu)$ was estimated using a golden search algorithm with endpoints 0.5 and 2, while the estimate of ν was set equal to $g^{-1}(\widehat{g(\nu)})$, with g^{-1} the inverse mapping of g. Due to the fact that the golden search algorithm uses a positive tolerance level for checking for convergence, the largest estimated value of $g(\nu)$ was approximately 1.998560484. This corresponds to the maximal value for ν of 694.678.

Let $\hat{T} = T - 2 \cdot \lfloor 0.05T \rfloor$ and $\hat{u}_t = \tilde{u}_{(t+\lfloor 0.05T \rfloor):T}$ for $t = 1, \ldots, \hat{T}$, then

$$\hat{f}(u) = \hat{T}^{-1} \sum_{t=1}^{\hat{T}} h^{-1} \phi((u - \hat{u}_t)/h), \tag{3.12}$$

with h a bandwidth parameter, $\phi(u) = (2\pi)^{-1/2} \exp(-u^2/2)$ the standard normal density function, and u an arbitrary real value (see Hendry (1995, p. 695) and Silverman (1986)).[7] The bandwidth parameter is set equal to $1.06T^{-0.2}$ times the median absolute deviation, which is again multiplied by 1.483. The score function is estimated by $\hat{f}'(u)/\hat{f}(u)$ for $\hat{f}(u) \geq 10^{-3}$ and zero otherwise, with

$$\hat{f}'(u) = \hat{T}^{-1} \sum_{t=1}^{\hat{T}} h^{-2} \phi'((u - \hat{u}_t)/h), \tag{3.13}$$

and $\phi'(u) = -u\phi(u)$.

In order to obtain the (nonparametric) maximum likelihood estimate based on $\hat{f}(u)$, the minimum of the function

$$\left(\sum_{t=1}^{T} \hat{f}'(y_t - \mu) \right)^2$$

with respect to μ is determined using a golden search algorithm with endpoints μ^l and μ^u. In order to avoid nonconvergent behavior of the estimator in the simulations, I set $\mu^l = \tilde{\mu} - \tilde{\sigma}$ and $\mu^r = \tilde{\mu} + \tilde{\sigma}$, with $\tilde{\sigma}$ the median absolute deviation of the y_t's, multiplied by 1.483.

3.3.2 Error Distributions

The performance of the above seven estimators is compared for several data generating processes. As mentioned earlier, the model is always (3.1), so only the error density $f(\varepsilon_t)$ is varied. The following seven choices for the error distribution are considered.

First, f is set equal to the standard Gaussian density. This density serves as a benchmark. For the Gaussian density, the mean is optimal. The other estimators, therefore, have a larger variance in this setting. It is interesting to know whether the increase in variance is acceptable compared with the reduction in sensitivity of the robust estimators to alternative error distributions.

Second, f is set equal to the Student t distribution with three degrees of freedom. This distribution still has finite second moments, so the mean should still be well behaved. The third order moment, however, does not exist, which implies an unstable behavior of the standard deviation as an estimator for the scale.

[7] Note that $\hat{f}(u)$ in fact estimates the density of $\hat{u}_t + \bar{u}_t$, with $\bar{u}_t$ a Gaussian random variable with mean zero and variance h^2, and $\bar{u}_t$ independent of $\hat{u}_t$.

Third, f is set equal to the (symmetrically) truncated Cauchy distribution. The truncation is performed in such a way that 95% of the original probability mass of the Cauchy is retained. In simulations in the time series context, this distribution is very useful for demonstrating the superiority of robust procedures in situations where all moments of the error distribution are still finite (see the Chapters 6 and 7).

Fourth, f is set equal to the standard Cauchy distribution. The first and higher order moments of this distribution do not exist.

Fifth, f is set equal to the slash distribution (see Andrews et al. (1972)). Drawings (ε_t) from this distributions are constructed by letting $\varepsilon_t = u_t/v_t$, with u_t a standard Gaussian random variable and v_t a uniform $[0, 1]$ random variable, independent of u_t. The first and higher order moments of this distribution do not exist.

The sixth distribution is added in order to illustrate the effect of asymmetry. In this case f is set equal to the recentered χ^2 distribution with two degrees of freedom. The distribution was centered as to have mean zero. Note that the robust estimators now have a different probability limit than the mean. This should not be taken as an argument against robust estimators. Robust estimators just estimate a different quantity if the error distribution is asymmetric (see, e.g., Hogg (1974, Section 7)). The real question is whether one is more interested in the mean or in the probability limit (or population version) of the robust estimator, e.g., the median. Using numerical integration, one can show that the probability limit of the median for this distribution is approximately -0.614, while the probability limits of the MLT estimators with $\nu = 5$ and $\nu = 1$ equal -0.506 and -0.773, respectively.

The seventh distribution, a mixture of normals, is added to illustrate the effect of outliers on the estimators for ν. With probability 0.9, a drawing is made from a normal with mean zero and variance 1/9, and with probability 0.1, a drawing is made from a normal with mean zero and variance 9. The variance of this mixture distribution is one. It can be expected that the robust estimators estimate the parameters from the normal with mean zero and variance 1/9 instead of the parameters of the mixture of normals. This, however, does not hold for the degrees of freedom parameter. As the largest component (approximately 90% of the observations) of the mixture is a normal distribution with mean zero and variance 1/9, one expects a robust estimator for ν to produce a very high estimate. It follows from Section 3.2, however, that the presence of the second component of the mixture causes a downward bias in the estimators of ν that are discussed in this chapter.

3.3.3 Results

As (3.1) is an extremely simple model, I only consider small samples of size $T = 25$. The means and standard deviations of the Monte-Carlo estimates over 400 replications are presented in Table 3.1. The median absolute deviations (multiplied by 1.483) and medians are presented in Table 3.2.

TABLE 3.1
Monte-Carlo Means and Standard Deviations for Several Estimators for the Location/Scale Model

	$\hat{\mu}$	$\hat{\sigma}_\mu$	$\hat{s}_\mu$	$\hat{\sigma}_{s\mu}$	$\hat{\sigma}$	$\hat{\sigma}_\sigma$	$\hat{\nu}$	$\hat{\sigma}_\nu$
				standard Gaussian				
mean	-0.007	0.205	0.198	0.028	0.992	0.141		
med	-0.008	0.248	0.236	0.105	0.966	0.236		
mlt5	-0.009	0.213	0.204	0.036	0.834	0.127		
mlt1	-0.009	0.259	0.240	0.095	0.576	0.117		
mlt	-0.009	0.209	0.189	0.030	0.919	0.167	499.157	308.508
adapt	-0.007	0.241	0.176	0.039	0.771	0.265	344.191	345.286
npml	-0.007	0.249	0.211	0.061				
				Student $t(3)$				
mean	-0.002	0.340	0.316	0.139	1.580	0.697		
med	0.002	0.264	0.321	0.162	1.117	0.293		
mlt5	-0.002	0.251	0.244	0.052	1.097	0.233		
mlt1	0.000	0.260	0.255	0.095	0.682	0.154		
mlt	-0.001	0.258	0.232	0.057	1.015	0.286	168.535	292.740
adapt	0.002	0.262	0.210	0.059	0.816	0.297	88.152	227.472
npml	0.002	0.276	0.252	0.074				
				truncated Cauchy				
mean	0.002	0.523	0.517	0.143	2.583	0.713		
med	0.003	0.305	0.513	0.322	1.383	0.430		
mlt5	-0.006	0.336	0.341	0.096	1.680	0.462		
mlt1	-0.006	0.287	0.280	0.106	0.884	0.242		
mlt	-0.004	0.306	0.285	0.093	1.167	0.426	31.018	137.405
adapt	-0.005	0.293	0.252	0.086	0.944	0.340	9.827	75.582
npml	0.000	0.333	0.335	0.107				
				standard Cauchy				
mean	0.232	29.121	4.482	28.805	22.408	144.025		
med	-0.005	0.317	0.601	0.390	1.489	0.469		
mlt5	-0.008	0.383	0.384	0.121	2.231	0.846		
mlt1	-0.008	0.280	0.292	0.105	0.974	0.279		
mlt	-0.010	0.289	0.291	0.104	1.059	0.393	11.257	81.527
adapt	-0.009	0.289	0.267	0.090	1.043	0.375	3.418	37.948
npml	-0.005	0.328	0.388	0.129				
				slash				
mean	-0.071	23.398	5.115	22.839	25.577	114.196		
med	0.021	0.525	1.264	0.706	2.182	0.610		
mlt5	0.046	0.552	0.541	0.163	3.069	1.148		
mlt1	0.015	0.474	0.453	0.157	1.420	0.370		
mlt	0.011	0.477	0.442	0.135	1.585	0.503	16.930	101.751
adapt	0.015	0.481	0.406	0.123	1.581	0.532	6.772	59.372
npml	0.029	0.525	0.570	0.176				

$\hat{\mu}$, $\hat{s}_\mu$, $\hat{\sigma}$, and $\hat{\nu}$ are the Monte-Carlo means of the estimators for μ, for the standard error of the estimator for μ, for σ, and for ν, respectively. The corresponding Monte-Carlo standard errors are $\hat{\sigma}_\mu$, $\hat{\sigma}_{s\mu}$, $\hat{\sigma}_\sigma$, and $\hat{\sigma}_\nu$ for $\hat{\mu}$, $\hat{s}_\mu$, $\hat{\sigma}$, and $\hat{\nu}$, respectively. The estimators are described in Subsection 3.3.1, while the error distributions are discussed in Subsection 3.3.2.

TABLE 3.1
(Continued)

	$\hat{\mu}$	$\hat{\sigma}_\mu$	$\hat{s}_\mu$	$\hat{\sigma}_{s\mu}$	$\hat{\sigma}$	$\hat{\sigma}_\sigma$	$\hat{\nu}$	$\hat{\sigma}_\nu$
				$\chi^2(2)-2$				
mean	-0.005	0.400	0.385	0.107	1.924	0.537		
med	-0.545	0.397	0.481	0.278	1.358	0.404		
mlt5	-0.370	0.360	0.296	0.072	1.336	0.300		
mlt1	-0.711	0.380	0.312	0.137	0.788	0.211		
mlt	-0.480	0.476	0.272	0.080	1.130	0.408	143.209	277.654
adapt	-0.595	0.451	0.250	0.079	0.970	0.381	80.893	219.260
npml	-0.466	0.423	0.346	0.113				
				$0.9 \cdot N(0,1/9)+0.1 \cdot N(0,9)$				
mean	-0.005	0.206	0.184	0.093	0.922	0.467		
med	-0.000	0.088	0.034	0.016	0.366	0.090		
mlt5	-0.000	0.081	0.079	0.020	0.409	0.143		
mlt1	-0.002	0.085	0.085	0.030	0.227	0.051		
mlt	-0.001	0.079	0.076	0.020	0.278	0.068	79.262	217.611
adapt	-0.001	0.084	0.070	0.019	0.269	0.087	57.433	188.017
npml	-0.001	0.089	0.091	0.028				

For the Gaussian error distribution, the mean is the most efficient estimator for μ, at least if we consider the Monte-Carlo standard deviation of the estimator (see the $\hat{\sigma}_\mu$ column). The mean is closely followed in terms of efficiency ($\hat{\sigma}_\mu$) by the MLT estimator with estimated ν (mlt) and the MLT estimator with ν fixed at 5 (mlt5). The remaining estimators perform much worse in terms of $\hat{\sigma}_\mu$. The standard errors of all estimators for μ ($\hat{s}_\mu$) seem to underestimate the true variability of the estimators ($\hat{\sigma}_\mu$) over the simulations. This holds in particular for the adaptive estimator. The MLT estimator of ν has a very high mean (see $\hat{\nu}$ column). If we consider the median of the MLT estimator for ν, it is at its upper boundery. The corresponding median absolute deviation reveals that for at least half of the simulations, the estimate of ν was at this boundary value. The adaptive estimate of ν is much lower (see especially the value in Table 3.2). The scale estimates ($\hat{\sigma}$) vary considerably over the different estimators. This is due to the fact that except for the mean, the median, and the MLT estimator with estimated ν, the estimators are estimating different quantities (compare Subsection 3.3.2). The adaptive estimator for σ has the highest variance ($\hat{\sigma}_\sigma$).

For the Student t distribution with three degrees of freedom, the mean performs much worse in terms of $\hat{\sigma}_\mu$. Now the MLT estimators perform best on the basis of Table 3.1, while on the basis of Table 3.2 the MLT estimators with fixed ν and the median perform best. The mean estimate of ν ($\hat{\nu}$) is again fairly high. The median estimate of ν, however, is much closer to the true value 3 for the MLT estimator. The adaptive estimator again underestimates ν. The discrepancy between the Monte-Carlo mean and median estimate of

TABLE 3.2
Monte-Carlo Medians and Median Absolute Deviations for Several Estimators for the Location/Scale Model

	$\hat{\mu}$	$\hat{\sigma}_\mu$	$\hat{s}_\mu$	$\hat{\sigma}_{s\mu}$	$\hat{\sigma}$	$\hat{\sigma}_\sigma$	$\hat{\nu}$	$\hat{\sigma}_\nu$
	standard Gaussian							
mean	-0.001	0.200	0.198	0.028	0.988	0.141		
med	-0.006	0.244	0.222	0.100	0.959	0.236		
mlt5	-0.001	0.206	0.203	0.037	0.831	0.131		
mlt1	-0.007	0.256	0.228	0.088	0.568	0.118		
mlt	-0.002	0.203	0.189	0.029	0.930	0.162	694.678	0.000
adapt	-0.001	0.226	0.179	0.039	0.816	0.333	49.035	71.513
npml	-0.003	0.247	0.205	0.056				
	Student $t(3)$							
mean	-0.006	0.320	0.288	0.076	1.441	0.378		
med	0.010	0.255	0.292	0.141	1.098	0.278		
mlt5	-0.000	0.251	0.241	0.050	1.070	0.225		
mlt1	0.007	0.255	0.237	0.088	0.671	0.146		
mlt	0.001	0.268	0.230	0.055	0.991	0.282	3.811	3.203
adapt	0.003	0.266	0.209	0.059	0.758	0.325	1.754	1.414
npml	0.003	0.264	0.245	0.072				
	truncated Cauchy							
mean	-0.006	0.509	0.515	0.150	2.575	0.748		
med	-0.004	0.295	0.425	0.229	1.308	0.380		
mlt5	-0.001	0.312	0.326	0.086	1.624	0.427		
mlt1	-0.012	0.267	0.265	0.096	0.847	0.220		
mlt	-0.013	0.279	0.273	0.083	1.091	0.361	1.763	0.613
adapt	-0.009	0.283	0.243	0.082	0.889	0.306	0.855	0.080
npml	-0.006	0.326	0.326	0.103				
	standard Cauchy							
mean	-0.069	1.327	1.095	0.814	5.474	4.070		
med	-0.003	0.300	0.513	0.300	1.424	0.430		
mlt5	-0.024	0.343	0.365	0.104	2.058	0.650		
mlt1	-0.001	0.270	0.279	0.097	0.950	0.261		
mlt	-0.011	0.285	0.280	0.096	1.005	0.327	1.075	0.383
adapt	-0.008	0.278	0.259	0.087	0.972	0.338	0.978	0.263
npml	-0.004	0.300	0.377	0.127				
	slash							
mean	0.020	1.881	1.456	1.161	7.279	5.803		
med	0.030	0.512	1.120	0.589	2.105	0.598		
mlt5	0.059	0.541	0.514	0.143	2.828	0.967		
mlt1	0.023	0.500	0.433	0.144	1.382	0.359		
mlt	0.012	0.497	0.428	0.128	1.520	0.427	1.173	0.470
adapt	0.011	0.490	0.399	0.113	1.526	0.535	1.225	0.629
npml	0.020	0.501	0.546	0.172				

$\hat{\mu}$, $\hat{s}_\mu$, $\hat{\sigma}$, and $\hat{\nu}$ are the Monte-Carlo medians of the estimators for μ, for the standard error of the estimator for μ, for σ, and for ν, respectively. The corresponding Monte-Carlo median absolute deviations, multiplied by 1.483, are $\hat{\sigma}_\mu$, $\hat{\sigma}_{s\mu}$, $\hat{\sigma}_\sigma$, and $\hat{\sigma}_\nu$ for $\hat{\mu}$, $\hat{s}_\mu$, $\hat{\sigma}$, and $\hat{\nu}$, respectively. The estimators are described in Subsection 3.3.1, while the error distributions are discussed in Subsection 3.3.2.

TABLE 3.2
(Continued)

	$\hat{\mu}$	$\hat{\sigma}_\mu$	$\hat{s}_\mu$	$\hat{\sigma}_{s\mu}$	$\hat{\sigma}$	$\hat{\sigma}_\sigma$	$\hat{\nu}$	$\hat{\sigma}_\nu$
				$\chi^2(2)-2$				
mean	-0.050	0.386	0.369	0.097	1.844	0.484		
med	-0.584	0.405	0.415	0.223	1.303	0.374		
mlt5	-0.401	0.363	0.289	0.069	1.300	0.289		
mlt1	-0.744	0.399	0.290	0.120	0.768	0.202		
mlt	-0.496	0.487	0.267	0.075	1.086	0.419	2.576	1.791
adapt	-0.628	0.478	0.247	0.075	0.909	0.377	1.949	1.703
npml	-0.506	0.422	0.334	0.112				
				$0.9\cdot N(0,1/9)+0.1\cdot N(0,9)$				
mean	-0.010	0.186	0.172	0.102	0.858	0.512		
med	-0.003	0.087	0.032	0.015	0.365	0.093		
mlt5	-0.002	0.079	0.077	0.016	0.379	0.107		
mlt1	-0.003	0.087	0.081	0.028	0.224	0.048		
mlt	0.000	0.080	0.075	0.019	0.275	0.067	1.669	0.752
adapt	0.002	0.084	0.069	0.019	0.264	0.095	2.177	1.899
npml	-0.001	0.085	0.088	0.027				

ν is due to large outlying values of $\hat{\nu}$ to the right. These are caused by the fact that for some samples of size 25 it is hardly possible to distinguish the Gaussian ($\nu = \infty$) from the Student t distribution. Further note that the scale estimator for the mean ($\hat{\sigma}$), i.e., the ordinary sample standard deviation, has a high variability ($\hat{\sigma}_\sigma$). This is caused by the nonexistence of the fourth moment of the distribution.

I now turn to the truncated Cauchy. For this distribution, the MLT estimator with $\nu = 1$ performs best in terms of $\hat{\sigma}_\mu$. This can be expected, as this estimator resembles the maximum likelihood estimator for this distribution. Only the standard error ($\hat{s}_\mu$) of the adaptive estimator seriously underestimates the true variability of the estimator ($\hat{\sigma}_\mu$). The median estimates of ν ($\hat{\nu}$) are now very low, explaining the relatively good performance of these estimators in terms of $\hat{\sigma}_\mu$.

For the standard Cauchy distribution, the mean is by far the worst estimator (see the $\hat{\mu}$ and the $\hat{\sigma}_\mu$ columns). The MLT estimator with $\nu = 1$, which is now exactly equal to the maximum likelihood estimator, performs best. The median estimates of ν are in the neighborhood of one for both the MLT and the adaptive estimator. Again the median estimate of ν obtained with the adaptive estimator is somewhat below that obtained with the MLT estimator. The second and third best estimators for μ are the adaptive estimator and the MLT estimator with estimated ν, respectively. Also note that the standard error of the adaptive estimator ($\hat{s}_\mu$) again seriously underestimates the true variability of the estimator ($\hat{\sigma}_\mu$).

The results for the slash distribution are similar to those for the Cauchy.

The variability of all estimators ($\hat{\sigma}_\mu$) appears to have increased with respect to the Cauchy case. Again the median estimates of ν are fairly close to one.

For the recentered χ^2 distribution the different probability limits of the various estimators for asymmetric error distributions are apparent. It is interesting to note that the two MLT estimators with fixed ν have better efficiency properties ($\hat{\sigma}_\mu$) than the mean, at least for Table 3.1. Moreover, based on Table 3.1, this distribution is the first one for which the nonparametric maximum likelihood estimator (npml) has a lower variability ($\hat{\sigma}_\mu$) than the estimators that only estimate ν instead of the whole error distribution. For Table 3.2, this also appeared for the Student t distribution with three degrees of freedom. The median estimates of ν ($\hat{\nu}$) are again quite low, with the adaptive estimate of ν below the MLT estimate.

Finally, for the mixture of normals, the mean performs worst in terms of $\hat{\sigma}_\mu$. This is due to the fact that the mean takes all observations into account, including the ones from the mixture component with variance 9. The MLT estimators with $\nu = 5$ and ν estimated perform best in terms of $\hat{\sigma}_\mu$. The standard error ($\hat{s}_\mu$) of the median seriously underestimates the true variability of the estimator ($\hat{\sigma}_\mu$) for this distribution. As expected from Section 3.2, the estimate of ν ($\hat{\nu}$) is biased towards zero. Note that the median estimate of ν for the adaptive estimator is now for the first time above that for the MLT estimator with estimated ν.

Summarizing, the following conclusions can be drawn from the simulation experiment.

1. In the considered experiment, the mean is only the best estimator in terms of $\hat{\sigma}_\mu$ if the errors are Gaussian. Even in this situation, the MLT estimator for μ with estimated ν performs approximately the same in terms of $\hat{\sigma}_\mu$.

2. The standard error of the adaptive estimator ($\hat{s}_\mu$) underestimates the true variability of the estimator ($\hat{\sigma}_\mu$) for all error distributions considered.

3. The nonparametric maximum likelihood estimator performs worse than or approximately the same as the median. The standard errors ($\hat{s}_\mu$) of both estimators in several cases grossly under- or overestimate the true variability ($\hat{\sigma}_\mu$) of the estimator for μ. Therefore, one can better use the median if one is only interested in a (robust) point estimate of μ, because this estimator is much easier to compute. If one also wishes to perform inference on μ, however, it is questionable whether any of these two estimators can be advised for practical use.

4. The estimators that only treat ν as a nuisance parameter (adapt and mlt) perform better in terms of $\hat{\sigma}_\mu$ for all symmetric distributions considered than the estimator that treats the whole error distribution as an (infinite-dimensional) nuisance parameter (npml). The only exception to this statement can be found in Table 3.2 for the Student t distribution with three degrees of freedom.

5. The nonrobustness of the MLT estimator for ν is evident if there are outliers in the data, as in the case of the mixture of normals. The nonrobustness of estimating ν along with μ and σ, however, seems to be either advantageous or negligible for estimation of and inference on μ.

6. One should be careful in defining the parameters of interest, because different estimators can have different probability limits for different error distributions. For σ and ν, this appears from all the experiments, while for μ it is illustrated by the experiment with the recentered χ^2 distribution.

3.4 Concluding Remarks

In this chapter I considered the simple location/scale model. For this model, I have demonstrated that the influence functions (IF) of the MLT estimators for the degrees of freedom parameter (ν) and the scale parameter (σ) are unbounded. The IF of the MLT estimator for the location parameter (μ) is bounded, but its change-of-variance (CVF) function is unbounded if the central or uncontaminated distribution does not belong to the Student t class. The easiest solution to the unboundedness of the IF's and the CVF is to fix the degrees of freedom parameter ν at a user specified value. This value can be chosen such that the estimator is reasonably efficient at a central distribution, e.g., the Gaussian distribution.

The unboundedness of the IF's and the CVF is, however, only a qualitative result. For example, the rate at which the IF diverges is very slow. Therefore, the practical implications of the unboundedness of the IF's and the CVF seems to be limited. This was illustrated in Section 3.3 by means of simulations. The MLT estimator with estimated degrees of freedom parameter seemed to perform as well as or better than the MLT estimators with fixed ν. Only the interpretation of the estimate of ν as the degrees of freedom parameter from an underlying Student t distribution seems to be incorrect if there are outliers in the data. Inference on μ, however, remains valid for most situations of practical interest.