One Step Towards a Less Sensitive Statistic

Anurag N. Banerjee[1]

Department of Economics and Finance,

Durham Business School,

Durham University.

September 17, 2008

[1]I would like to thank ESRC (grant number RES-000-22-0646) and the British Academy for the financial support.

Abstract

Econometric estimates or decisions are applicable if they are not sensitive to small changes of "nuisance" parameters. The statistic which measures these violations is called a sensitivity measure. In this article, we give a sufficient condition for a statistic to be sensitive, leading to the definition of the sensitivity measure. The measure suggests when the sensitivity statistic is "small", we need not bother about the nuisance parameters. The main contribution of this article is to propose a correction factor for the statistic of interest if we find "large" sensitivity. We define such an estimator and analyse its properties. As an application we analyse the sensitivity of the slope and variance estimates of the linear model with respect to the presence of associated nuisance parameters in the variance covariance matrix. We also give an example where the risk function is parametrised by a nuisance parameter.

JEL: C12; C22; C51; C52

Keywords: Estimation, model sensitivity, nuisance parameters, linear regression, risk functions.

1 Introduction

Let us suppose we want to estimate an income elasticity using a given set of time-series observations. We use a simple model, ignoring serial correlation or the income distribution. We thus obtain an estimate (say, by ordinary least squares) of the required elasticity. Many will argue that the model is far removed from reality. But this does not mean that our current estimate cannot be close to the estimate in a bigger, more realistic model. Analysing when this is likely to be the case and if it is indeed the case how to rectify it, is the purpose of the proposed research.

Economic and econometric models are used as simple metaphors for the real economic process. There are several reasons to study economic models- to justify some economic hypothesis, to get a mathematical model of the economy, but mostly to study a particular feature or features of the real economy. When we are interested in estimating or inferring a particular feature of the economy, anything else in the model or the estimation process is only important in so far as it may influence the study of the desired feature. For example, we may be interested in estimating an income elasticity as in the previous paragraph or in testing an economic hypothesis like constant returns to scale. In that case, the estimate of the elasticity or the test statistic is our statistic of interest. The other features of the models, like autocorrelation or hetroscedasticty are only at best of secondary interest or indeed there as nuisance parameters. We are only interested to know what the effects of the changes of these unknown parameters[1] have on the estimates of the parameters we are interested in i.e. the elasticity or the test statistic. For this purpose the idea of sensitivity statistics was proposed. This is as follows:

Let $S(\mathbf{y},\theta)$ be the relevant statistic where $\mathbf{y}=(y_1,..,y_t,..,y_T)$, be a data point and θ is the nuisance parameter either from the data generating process or a part of the estimation method, for example a parameter of the risk function. Developing $S(\mathbf{y},\theta)$ in a Taylor

[1] The nuisance parameter θ may not be a part of the Data Genterating Process at all, it can be a parameter of a optimisation criteria to generate the estimate.

expansion (with respect to θ) gives

$$S(\mathbf{y},\theta) = S(\mathbf{y},\theta_0) + \sum_{i=1}^{p} (\theta_i - \theta_{0,i}) \left. \frac{\partial S(\mathbf{y},\theta)}{\partial \theta_i} \right|_{\theta=\theta_0} + \cdots . \tag{1.1}$$

where θ_0 is a known point and the summation is over the number of nuisance parameters p. We would consider $S(\mathbf{y},\theta)$ and $S(\mathbf{y},\theta_0)$ to be "almost equal" if

$$\sum_{i=1}^{p} (\theta_i - \theta_{0,i}) \left. \frac{\partial S(\mathbf{y},\theta_0)}{\partial \theta_i} \right|_{\theta=0} \approx 0$$

and a sufficient condition for this is that

$$\left. \frac{\partial S(\mathbf{y},\theta_0)}{\partial \theta_i} \right|_{\theta=\theta_0} = 0 \quad (i = 1, \ldots, p).$$

The sensitivity statistics are based on this simple observation, so that if the statistic shows "less" sensitivity at θ_0, then the use of $S(\mathbf{y},\theta_0)$ is "good enough" and the value of nuisance parameter need not be changed.

The method of sensitivity analysis was developed by Banerjee and Magnus (1999), who investigated the sensitivity of the OLS estimators to the white noise assumption of the error process. The standard linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ was considered assuming $\mathbf{u} \sim N(0, \sigma^2 \boldsymbol{\Omega}(\theta))$, where σ^2 and θ are unknown. They proposed a pair of sensitivity statistics $B1$ and $D1$ to measure the sensitivity of $\hat{\beta}(\theta)$ the slope estimator and the variance estimator $\hat{\sigma}^2(\theta)$ respectively, when the autocorrelation parameter θ of a stationary AR(1) disturbance process moves away from zero. The authors derived the distribution of $B1$ and showed that $D1$ has the same form as the Durbin-Watson (DW) test statistic. This implies that the sensitivity measures for certain statistics are equivalent to a test statistic of the nuisance parameter. This idea has been extended to t-statistic and F-statistic (Banerjee and Magnus, 2000), unit root forecasts (Banerjee, 2001), Maximum Likelihood Estimates (Magnus and Vasnev, 2007), Impact Factors (Omtzigt and Paruolo, 2005), Restricted Least Squares (Wan et. al. 2007) and other statistic(s) of interest.

The question remains, what should we do if the sensitivity is "large" and we must conclude that $S(\mathbf{y},\theta_0)$ is sensitive to misspecification of the nuisance parameters ? One possible solution is to use (1.1) and write the modified statistic as:

$$S\left(\mathbf{y},\hat{\theta}_S\right) \approx S\left(\mathbf{y},\theta_0\right) + \sum_{i=1}^{p}\left(\hat{\theta}_{S,i}-\theta_{0,i}\right)\left.\frac{\partial S\left(\mathbf{y},\theta\right)}{\partial\theta_i}\right|_{\theta=\theta_0}, \tag{1.2}$$

where $\hat{\theta}_S$ is some consistent estimate of θ. What are the possible problems of this approach ?

First, how do we find the point $\hat{\theta}_S$. Second, and most important, we would expect that this estimator would make the statistic $S\left(\mathbf{y},\hat{\theta}_S\right)$ "insensitive" at $\hat{\theta}_S$. If we use any consistent estimator this may not be very good in reducing sensitivity of the statistic of interest. We discuss cases for which $\hat{\theta}_S$ is indeed consistent, but it is possible that inconsistent estimators may be better at reducing sensitivity of $S\left(\mathbf{y},\theta\right)$. Finally, consistency of $\hat{\theta}_S$ might not have any economic or statistical meaning when θ is a part of the estimation method (for example a parametrised risk-function).

An analogy of this idea comes from the literature on shrinkage estimation where requiring unbiasedness of the estimator reduces its accuracy. The shrinkage estimator is then a function of the loss function. Likewise, here $\hat{\theta}_S$ the estimator of θ will be function of $S\left(\mathbf{y},\theta\right)$. As in shrinkage estimation, the idea is not to get the "right" estimate of the true value of θ (which anyway is not the parameter that we are interested in) but to get an estimate $\hat{\theta}_S$ which would reduce the sensitivity of the statistic $S\left(\mathbf{y},\theta\right)$.

We define the least sensitive estimate $\hat{\theta}_S$ by using a one step Newton-Raphson type method as:

$$\hat{\theta}_{S,i} = \theta_{0,i} - \frac{\left\langle \frac{\partial S(\mathbf{y},\theta_0)}{\partial\theta_i}, \frac{\partial^2 S(\mathbf{y},\theta_0)}{\partial\theta_i^2}\right\rangle}{\left\|\frac{\partial^2 S(\mathbf{y},\theta_0)}{\partial\theta_i^2}\right\|^2}\ \forall\ i.$$

where $\langle\rangle$ is any inner product and $\|\|$ the associated norm. We can the plug in the values in equation (1.2) and find the least sensitive estimator. Notice that correction factors are additive hence is an advantage when there is a large number of unknown nuisance parameters. We can add more correction terms without changing the other corrections already done.

After establishing our estimator, we then analyse the small and large sample properties of $\hat{\theta}_S$ and $S\left(\mathbf{y},\hat{\theta}_S\right)$, although the study of $\hat{\theta}_S$ is just incidental to the study of $S\left(\mathbf{y},\theta\right)$. We provide two application of our method. First, following Banerjee and Magnus (1999) we investigate the sensitivity corrections of the OLS estimators when the nuisance parameters are in the variance component. We consider both time-series (ARFIMA) and cross-sectional nuisance parameters. Secondly we consider estimation of central tendency under asymmetric risk function parametrised by a nuisance parameter and show conditions under which our methodology fails.

The paper is organised as follows: Section 2 defines a sensitivity statistic, provides a justification of the definition and develops the least sensitive estimate. Section 3 provides some small and large sample properties of these estimators. In Section 4 we analyse the slope and variance estimates of the linear model under covariance misspecification. We also analyse special cases for the linear model to illustrate our method. Section 5 we analyse the central tendency measure under a parametrised risk function and state conditions of failure of our method. Section 6 concludes.

2 Sensitivity of a Statistic Under Nuisance Parameters.

Let $\mathbf{y} = (y_1, .., y_t, ..y_T)$ be a data point from a Data Generating Process. We are interested in the $K \times 1$ vector of statistic $S(\mathbf{y}, \theta) = (S_1(\mathbf{y}, \theta), .., S_k(\mathbf{y}, \theta), .., S_K(\mathbf{y}, \theta))$ which are parametrised by θ. The statistic $S(\mathbf{y}, \theta)$ which we are interested in are estimates or inferences of parameters of interest (or some functions of them) at $\theta \in \times_{i=1}^{p} \left[\underline{\theta}_i, \overline{\theta}_i\right]$ which are nuisance parameters. Therefore, any inference or estimation (such as pretesting) on the nuisance parameter θ is only incidental. Banerjee and Magnus (1999, 2000), give an alternative to pretesting for θ. Their study is based on sensitivity analysis of the statistic around θ_0. Their investigation considered, not how to test for "true" θ_0 but whether the difference $S(\mathbf{y}, \theta) - S(\mathbf{y}, \theta_0)$ is large or small when θ is a value different from θ_0, the value under which the estimation is taking place (most often $\theta_0 = 0$). The difference $S(\mathbf{y}, \theta) - S(\mathbf{y}, \theta_0)$ is approximated using a Taylor series approximation.

Suppose $S(\mathbf{y}, \theta)$ can be expanded in a Taylor series expansion,

$$S(\mathbf{y}, \theta) - S(\mathbf{y}, \theta_0) = \sum_{i=1}^{p} (\theta_i - \theta_{0,i}) \left.\frac{\partial S(\mathbf{y}, \theta)}{\partial \theta_i}\right|_{\theta=\theta_0} + \cdots \tag{2.3}$$

If higher order differences are ignored, then we would consider $S(\mathbf{y}, \theta_1)$ and $S(\mathbf{y}, \theta_0)$ to be "almost equal" if

$$\nabla_i^{(1)} S(\mathbf{y}, \theta_0) = \left.\frac{\partial S(\mathbf{y}, \theta)}{\partial \theta_i}\right|_{\theta=\theta_0} \approx 0 \ , \ \ (i = 1, \ldots, p) \tag{2.4}$$

the derivative of the statistic $S(\mathbf{y}, \theta_0)$ evaluated at θ_0 will be formally defined as a sensitivity measure.

A contribution of this paper is the following formal sufficiency condition for the justification of the use of the derivative (2.4) as a measure of locally "almost equal" which is stated in following theorem:

Theorem 1 *Suppose there exists a* $\lambda_i > 0$ *such that*

$$\left.\frac{\partial S(\mathbf{y}, \theta)}{\partial \theta_i}\right|_{\theta=\theta_0} > \lambda_i$$

and

$$\left.\frac{\partial^2 S(\mathbf{y}, \theta)}{\partial \theta_i^2}\right|_{\theta=\theta_0} > 0, \tag{2.5}$$

then

$$\Pr_{\theta} \left(S(\mathbf{y}, (\theta_{0,i} + \eta \ , \theta_{0,-i})) - S(\mathbf{y}, \theta_0) > \eta \lambda_i \right) \geq \Pr_{\theta} \left(\left.\frac{\partial S(\mathbf{y}, \theta)}{\partial \theta_i}\right|_{\theta=\theta_0} > \lambda_i \right)$$

for $\theta \in [\theta_0, (\theta_{0,i} + \eta, \theta_{0,-i}))$ *for some* $\eta > 0$.

Proof of Theorem 1: See Appendix.

Remark 1 *This can also be proved for* $\theta \in (\theta_0 - \eta, \theta_0]$ *and* $\frac{\partial S(\mathbf{y}, \boldsymbol{\theta})}{\partial \theta_i} < \lambda_i < 0$. *This implies that we have* $\left| \frac{\partial S(\mathbf{y}, \boldsymbol{\theta})}{\partial \theta_i} \right| > \lambda_i$ *for some interval around* θ_0 *such that*

$$\Pr_{\theta_0} \left(\left| S(\mathbf{y}, \theta_0 + \eta) - S(\mathbf{y}, \theta_0) \right| > \varepsilon \right) > \xi > 0.$$

The theorem and the remark imply that if the derivative is "large" there is a significant probability of a large deviation in statistic $S(\mathbf{y}, \theta_0)$ with respect to a small change in the parameter. This leads to our definition of sensitivity measure.

Definition 1 (Sensitivity Measure) *Define a sensitivity measure of $S(\mathbf{y}, \theta)$ with respect to θ_i:*

$$\nabla_i^{(1)} S(\mathbf{y}, \theta_0) = \left. \frac{\partial S(\mathbf{y}, \theta)}{\partial \theta_i} \right|_{\theta=\theta_0} (i = 1, \ldots, p).$$

We should note that this is a sufficient condition not a necessary one. Banerjee (2001) shows that sensitivity measures of forecasts of unit root process with deterministic trend is zero. Whereas the forecasts are sensitive to near unit root violations. Therefore, we can extend our analysis to higher order terms

$$\nabla_i^{(k)} S(\mathbf{y}, \theta_0) = \left. \frac{\partial^k S(\mathbf{y}, \theta)}{\partial \theta_i^k} \right|_{\theta=\theta_0}, k = 2, ... \tag{2.6}$$

of the Taylor series expansion (2.3) and define higher order sensitivity measures if required (Banerjee 2001).

If the statistic $\nabla_i^{(1)} S(\mathbf{y}, \theta_0)$ shows little sensitivity by being close to zero, then we use the estimator/statistic $S(\mathbf{y}, \theta_0)$. But how do we know $\nabla_i^{(1)} S(\mathbf{y}, \theta_0)$ is close to zero ?

The sensitivity measure $\nabla^{(1)} S(\mathbf{y}, \theta_0)$ will generally be random variables, therefore Theorem (1) provides a basis for the study of the following probabilities as a measure of "closeness" to zero

$$\pi_\alpha(\theta : S(\mathbf{y}, \theta_0)) = \Pr_\theta \left(\left\| \nabla_i^{(1)} S(\mathbf{y}, \theta_0) \right\| \geq S_\alpha \right) \tag{2.7}$$

where $\Pr_\theta$ is the probability measure associated with the random variable $\mathbf{y}$ at θ and $\|.\|$ is an appropriate norm. The cutoff point S_α is obtained from the equation,

$$\Pr_{\theta_0} \left(\left\| \nabla_i^{(1)} S(\mathbf{y}, \theta_0) \right\| \geq S_\alpha \right) = \alpha \ , \ 0 < \alpha < 1. \tag{2.8}$$

The level α can be interpreted as "allowable" sensitivity of the statistic $S(\mathbf{y}, \theta_0)$.

Definition 2 (Sensitivity Function) *The statistic $\pi_\alpha(\theta : S(\mathbf{y}, \theta_0))$ defined in (2.7) and (2.8) is the sensitivity function of the statistic $S(\mathbf{y}, \theta_0)$ measured at level α, where the distribution $\mathbf{y}$ is evaluated at θ.*

What should we do if $\nabla_i^{(1)} S(\mathbf{y}, \theta_0)$ is "large", and we must conclude that $S(\mathbf{y},\theta_0)$ is sensitive to misspecification of the nuisance parameters ?

One possible solution is to use the next term in the Taylor's expansion and solve for $\widehat{\theta}_{S,i}$ conditional on $\widehat{\theta}_{S,-i} = \theta_{0,-i}$, such that $\nabla_i^{(1)} S\left(\mathbf{y},\widehat{\theta}_S\right) = 0$. This leads to a definition of the insensitivity point.

Definition 3 (Local Insensitivity) *A statistics $S\left(\mathbf{y},\hat{\theta}_S\right)$ is locally insensitive in a neighborhood $\hat{\theta}_S$, when*

$$\nabla_i^{(1)} S\left(\mathbf{y}, \left(\hat{\theta}_{S,i}, \hat{\theta}_{S,-i}\right)\right) = 0 \text{ for all } i = 1, ..., p. \tag{2.9}$$

and $\hat{\theta}_S$ is called the insensitivity point of $S(\mathbf{y},\theta)$.

The following procedure gives us a method to compute the insensitivity point. Given a starting value θ_0 we solve for $\hat{\theta}_{S,i}$ componentwise as:

$$\hat{\theta}_{S,i} = \arg_{\theta_{-i}=} \nabla_i^{(1)} S\left(\mathbf{y}, (\theta_i, \theta_{0,-i})\right) = 0. \tag{2.10}$$

We use a One-Step Newton-Raphson procedure by approximating at $\hat{\theta}_{S,i}$ near θ_0. Using a Taylor's expansion $\nabla_i^{(1)} S\left(\mathbf{y},\widehat{\theta}_S\right)$ around θ_0 and ignoring the higher orders we have,

$$\begin{aligned} \nabla_i^{(1)} S\left(\widehat{\theta}_S\right) &= \nabla_i^{(1)} S(\mathbf{y}, \theta_0) + \left(\hat{\theta}_{S,i} - \theta_{0,i}\right) \nabla_i^{(2)} S(\mathbf{y}, \theta_0) \qquad (2.11) \\ 0 &= \nabla_i^{(1)} S(\mathbf{y}, \theta_0) + \left(\hat{\theta}_{S,i} - \theta_{0,i}\right) \nabla_i^{(2)} S(\mathbf{y}, \theta_0) \\ -\nabla_i^{(1)} S(\mathbf{y}, \theta_0) &= \left(\widehat{\theta}_{S,i} - \theta_{0,i}\right) \nabla_i^{(2)} S(\mathbf{y}, \theta_0). \end{aligned}$$

Using appropriate inner product we can find $\hat{\theta}_{S,i}$ as

$$\hat{\theta}_{S,i} = \theta_{0,i} - \frac{\left\langle \nabla_i^{(1)} S(\mathbf{y}, \theta_0), \nabla_i^{(2)} S(\mathbf{y}, \theta_0) \right\rangle}{\left\| \nabla_i^{(2)} S(\mathbf{y}, \theta_0) \right\|^2}. \tag{2.12}$$

Note that if $S(\mathbf{y}, \theta_0)$ is a scaler (2.12) simplifies to

$$\hat{\theta}_{S,i} = \theta_{0,i} - \frac{\nabla_i^{(1)} S(\mathbf{y}, \theta_0)}{\nabla_i^{(2)} S(\mathbf{y}, \theta_0)}. \tag{2.13}$$

Finally, we can define the sensitivity adjusted estimator $S\left(\mathbf{y},\widehat{\theta}_S\right)$ of $S(\mathbf{y},\theta)$ near θ_0 as:

$$\begin{aligned} S\left(\mathbf{y},\widehat{\theta}_S\right) &= S(\mathbf{y},\theta_0) + \sum_{i=1}^{p}\left(\hat{\theta}_{S,i} - \theta_{0,i}\right)\nabla_i^{(1)} S(\mathbf{y},\theta_0) \\ &= S(\mathbf{y},\theta_0) - \sum_{i=1}^{p}\frac{\left\langle \nabla_i^{(1)} S(\mathbf{y},\theta_0), \nabla_i^{(2)} S(\mathbf{y},\theta_0)\right\rangle}{\left\|\nabla_i^{(2)} S(\mathbf{y},\theta_0)\right\|^2}\nabla_i^{(1)} S(\mathbf{y},\theta_0). \end{aligned} \tag{2.14}$$

Again, if $S(\mathbf{y},\theta)$ is a scaler, (2.14) simplifies to:

$$S\left(\mathbf{y},\widehat{\theta}_S\right) = S(\mathbf{y},\theta_0) - \sum_{i=1}^{p}\frac{\left[\nabla_i^{(1)} S(\mathbf{y},\theta_0)\right]^2}{\nabla_i^{(2)} S(\mathbf{y},\theta_0)}. \tag{2.15}$$

One obvious assumption we made during this discussion is $\left\|\nabla_i^{(2)} S(\mathbf{y},\theta_0)\right\|^2 > 0$. If $\left\|\nabla_i^{(2)} S(\mathbf{y},\theta_0)\right\|^2 = 0$ then we shall use the higher order terms of the Taylor series expansion of $\nabla_i^{(1)} S(\mathbf{y},\theta)$ around θ_i. In section 5, we give such an example.

3 Properties of Least Sensitive Estimator

Existence of least sensitive estimator has to be dealt conditionally for each θ_i given θ_{-i}. From the previous section equation(s), (2.14) and (2.15) we see that adjustment of the statistic $S(\mathbf{y},\theta_0)$ is separable, thus for further analysis it is enough to study one nuisance parameter, i.e. $p = 1$.

Motivated by Theorem (1), existence and uniqueness of the solution (2.12) will be guaranteed if the following assumption holds:

Assumption 1 *Without loss of generality we assume for all* $\mathbf{y}$, $\nabla^{(1)} S(\mathbf{y},\theta)$ *is increasing in* θ *i.e.*

$$\nabla^{(2)} S(\mathbf{y},\theta) > 0 \tag{3.16}$$

and

$$\nabla^{(1)} S(\mathbf{y},\underline{\theta}) < 0 < \nabla^{(1)} S\left(\mathbf{y},\overline{\theta}\right). \tag{3.17}$$

The second part of the assumption is a single crossing property which along with continuity of $\nabla^{(1)} S_k(\mathbf{y},\theta)$ ensures the existence of a solution. The first part of the assumption guarantees that the solution is unique. It might be possible that in certain instances Assumption (3.16) is not satisfied and the use of local solutions are needed. That is why from a practical point of view, our investigation concerns local sensitivity of $S(\mathbf{y},\theta)$ by defining appropriate neighborhoods of θ_0 and modifying assumption to ensure local existence (Theorem 1 condition (2.5)). Thus, though we assume a global single crossing property, we can modify our assumption to take into consideration finitely many multiple roots as well. Subsequent results can be easily generalised in the case of multiple crossing (though finitely many) of the $\nabla^{(1)} S(\mathbf{y},\theta)$ function.

An interesting property of statistical estimation are the estimators which are scale and location invariant. We give conditions under which $\hat{\theta}_S$ is scale and location invariant.

Theorem 2 *If*

$$S(a+b\mathbf{y},\theta) = a' + b' S(\mathbf{y},\theta), \tag{3.18}$$

where a' is a constant vector and b' is scaler constant independent of θ, then $\widehat{\theta}_S$ the least sensitive estimate is scale and location invariant.

Proof of Theorem 2: Note that $\nabla^{(k)} S(a+b\mathbf{y},\theta) = b'\nabla^{(k)} S(\mathbf{y},\theta)$, $k=1,2$. Therefore,

$$\hat{\theta}_{S,i} - \theta_{0,i} = -\frac{\left\langle \nabla_i^{(1)} S(a+b\mathbf{y},\theta_0), \nabla_i^{(2)} S(a+b\mathbf{y},\theta_0) \right\rangle}{\left\| \nabla_i^{(2)} S(a+b\mathbf{y},\theta_0) \right\|^2} = -\frac{\left\langle \nabla_i^{(1)} S(\mathbf{y},\theta_0), \nabla_i^{(2)} S(\mathbf{y},\theta_0) \right\rangle}{\left\| \nabla_i^{(2)} S(\mathbf{y},\theta_0) \right\|^2}.$$

QED.

As an example if $S(\mathbf{y},\theta)$ is linear in $\mathbf{y}$ then $\widehat{\theta}_S$ is scale and location invariant. Also if $S(\mathbf{y},\theta)$ is a homogenous function in $\mathbf{y}$ then $\widehat{\theta}_S$ is scale invariant. More cases of invariance of $S(\mathbf{y},\theta)$ are discussed in the next section when we study the linear model.

3.1 Asymptotics

As traditionally the case, in order to study the properties of the locally least sensitive estimates like the π_α curves and $S\left(\mathbf{y},\widehat{\theta}_S\right)$, we need the distribution of $\widehat{\theta}_S$. Though we are interested in finite sample properties which we shall study in the next section for estimates of linear models, we shall look at some asymptotic properties of the estimates θ from a general point of view. The following theorem gives us conditions of consistency of the estimator $\widehat{\theta}_S$.

Theorem 3 *Given assumption (1), if*

$$\operatorname*{plim}_{T\to\infty} \nabla^{(k)} S(\mathbf{y},\theta) = \nabla^{(k)} S(\theta), \ \ (k = 1, 2) \tag{3.19}$$

exists uniformly for all θ, *then there exist* θ_S^* *such that:*

1. $\nabla^{(1)} S(\theta_S^*) = 0$,

2. θ_S^* *is unique if* $\nabla^{(2)} S(\theta) > 0$.

3. $\operatorname{plim}_{T\to\infty} \widehat{\theta}_S = \theta_S^*$ *if* $\theta_0 = \theta_S^*$.

Proof of Theorem 3: Consider the statistic $S(\mathbf{y},\theta)$ satisfying assumption (1). Taking limits (3.17) we have

$$\begin{aligned} \operatorname*{plim}_{T\to\infty} \nabla^{(1)} S(\mathbf{y},\underline{\theta}) &< 0 < \operatorname*{plim}_{T\to\infty} \nabla^{(1)} S\left(\mathbf{y},\overline{\theta}\right) \\ \nabla^{(1)} S(\underline{\theta}) &< 0 < \nabla^{(1)} S\left(\overline{\theta}\right). \end{aligned} \tag{3.20}$$

Using (3.16) and (3.19) we have $\nabla^{(2)} S(\theta) > 0$. Therefore there exists θ_S^* such that $\nabla^{(1)} S(\theta_S^*) = 0$. Equation (3.20) implies uniqueness because of the single crossing property.

Taking limits on both sides of (2.12)

$$\begin{aligned} \operatorname*{plim}_{T\to\infty} \widehat{\theta}_S &= \theta_0 - \operatorname*{plim}_{T\to\infty} \frac{\left\langle \nabla^{(1)} S(\mathbf{y},\theta_0), \nabla^{(2)} S(\mathbf{y},\theta_0)\right\rangle}{\left\|\nabla^{(2)} S(\mathbf{y},\theta_0)\right\|^2} \\ \operatorname*{plim}_{T\to\infty} \widehat{\theta}_S &= \theta_0 - \frac{\left\langle \operatorname{plim}_{T\to\infty} \nabla^{(1)} S(\mathbf{y},\theta_0), \operatorname{plim}_{T\to\infty} \nabla^{(2)} S(\mathbf{y},\theta_0)\right\rangle}{\left\|\operatorname{plim}_{T\to\infty} \nabla^{(2)} S(\mathbf{y},\theta_0)\right\|^2} \\ &= \theta_0 - \frac{\left\langle \nabla^{(1)} S(\theta_0), \nabla^{(2)} S(\theta_0)\right\rangle}{\left\|\nabla^{(2)} S(\theta_0)\right\|^2}. \end{aligned}$$

Further,

$$\operatorname*{plim}_{T\to\infty} S\left(\mathbf{y},\widehat{\theta}_S\right) = S(\theta_0) - \frac{\left\langle \nabla^{(1)}S(\theta_0), \nabla^{(2)}S(\theta_0)\right\rangle}{\left\|\nabla^{(2)}S(\theta_0)\right\|^2}\nabla^{(1)}S(\theta_0)$$

and therefore if $\theta_0 = \theta_S^*$ then $\nabla^{(1)}S(\theta_S^*) = 0$. Hence,

$$\operatorname*{plim}_{T\to\infty} \widehat{\theta}_S = \theta_S^*.$$

QED.

Theorem (3) also implies

$$\operatorname*{plim}_{T\to\infty} S\left(\mathbf{y},\widehat{\theta}_S\right) = S(\theta_S^*).$$

Hence, $S(\mathbf{y},\theta_S^*)$ is insensitive to perturbations in the neighborhood of θ_S^*.

Example 1 (Consistancy of $\widehat{\theta}_S$) *Let* $S(\mathbf{y},\theta) = \frac{1}{T}\sum \mathfrak{L}(y_t,\theta)$*where* $\mathfrak{L}$ *is a loss-function and* $S(\theta) = E(\mathfrak{L}(y_t,\theta))$. *As* $\operatorname{plim}_{T\to\infty} S(\mathbf{y},\theta) \to S(\theta)$ *uniformly for all* $\theta \in \Theta$ *and almost every sequence* $(y_1,..,y_t,..,y_T)$. *Let* $\widehat{\theta}_S$ *minimises* $S(\mathbf{y},\theta)$ *(i.e.* $\nabla^{(1)}S\left(\mathbf{y},\widehat{\theta}_S\right) = 0$*), and* θ_S^* *uniquely minimises* $S(\theta)$ *(i.e.* $\nabla^{(1)}S(\theta_S^*) = 0$*) then* $S(\mathbf{y},\theta) \overset{a.s.}{\to} S(\theta)$*, implies* $\widehat{\theta}_S \overset{a.s.}{\to} \theta_S^*$ *(Lemma 2.2 White (1980)).*

These estimates can be (non)-linear least square estimates (White, 1981), Maximum Likelihood Estimates (White, 1982) or most estimates which are obtained by minimising some objective function which have uniform convergence properties.

The consistency of the estimates of the nuisance parameters have implications on the sensitivity function $\pi_\alpha(\theta : S(\mathbf{y},\theta_0))$. The following theorem shows that if $\theta_0 = \theta_S^*$ the sensitivity of $S(\mathbf{y},\theta)$ is "low".

Theorem 4 *Given assumption (1) and condition (3.19) of Theorem (3) hold, we have:*

1. $\lim_{T\to\infty} \pi(\theta : S(\mathbf{y},\theta_0)) = \pi_\alpha(\theta : S(\theta_0))$.

2. *Further if* $\nabla^{(2)}S(\theta_S^*) > 0$ *then there exists* $S^* > 0$ *such that,*

$$\lim_{T\to\infty} \Pr\left(\left|\widehat{\theta}_S - \theta_S^*\right| \geq S^*\right) = 0$$

and for $\theta_0 \notin (\theta_S^* - \eta, \theta_S^* + \eta)$ *some* $\eta > 0$,

$$\lim_{T\to\infty} \Pr\left(\left|\widehat{\theta}_S - \theta_0\right| \geq S^*\right) = 1.$$

Proof of Theorem 4: Part 1) follows from the fact that convergence in probability implies convergence in distribution. 2) Since$\nabla^{(2)} S(\mathbf{y}, \theta_0) > 0$, because of the single crossing property (Assumption (1)), there exists $\theta_0 > \theta_S^* + \eta$, $\eta > 0$ such that $\Pr\left(\nabla^{(1)} S(\mathbf{y}, \theta_0) > S^* \left\|\nabla^{(2)} S(\mathbf{y}, \theta_0)\right\|\right) = 1$ for some $S^* > 0$. Note that

$$\begin{aligned}
\Pr\left(\left|\widehat{\theta}_S - \theta_0\right| \geq S^*\right) &= \Pr\left(\left|\frac{\left\langle \nabla^{(1)} S(\mathbf{y}, \theta_0), \nabla^{(2)} S(\mathbf{y}, \theta_0)\right\rangle}{\left\|\nabla^{(2)} S(\mathbf{y}, \theta_0)\right\|^2}\right| \geq S^*\right) \\
&\geq \Pr\left(\left\|\nabla^{(1)} S(\mathbf{y}, \theta_0)\right\| > S^* \nabla^{(2)} S(\mathbf{y}, \theta_0)\right).
\end{aligned}$$

Therefore $\lim_{T\to\infty} \Pr\left(\left|\widehat{\theta}_S - \theta_0\right| \geq S^*\right) = 1$ and $\lim_{T\to\infty} \Pr\left(\left|\widehat{\theta}_S - \theta_S^*\right| \geq S^*\right) = 0$.

QED.

This is particularly useful result when we present our results using the probability of sensitivity which is defined as

$$\widehat{\pi}(\theta_0 : S(\mathbf{y}, \theta_0)) = \Pr_{\theta_0}\left(\nabla^{(1)} S(\mathbf{y}, \theta_0) \leq \nabla^{(1)} S(\mathbf{y}_{obs}, \theta_0)\right), \tag{3.21}$$

where $\nabla^{(1)} S(\mathbf{y}, \theta_0)$ is the relevant sensitivity measure and $\nabla^{(1)} S(\mathbf{y}_{obs}, \theta_0)$ is the value $\nabla^{(1)} S(\mathbf{y}, \theta_0)$ computed at the observed data point $\mathbf{y}_{obs}$. The probability $\widehat{\pi}(\theta_0 : S(\mathbf{y}, \theta_0))$ is interpreted as how sensitive the estimate $S(\mathbf{y}, \theta_0)$ is to the specifications of the nuisance parameters given the observed value $\mathbf{y}_{obs}$. The larger the value of $\widehat{\pi}$, the more sensitive $S(\mathbf{y}_{obs}, \theta_0)$ is to misspecification of the nuisance parameters, since small changes in $\theta_{0,i}$ would give large changes in $S(\mathbf{y}_{obs}, \theta_0)$.

Note that, if $\theta_0 = \theta_S^*$ from theorem (3) we have:

$$\lim_{T\to\infty} \widehat{\pi}(\theta_S^* : S(\mathbf{y}, \theta_S^*)) = 1.$$

Hence, if $\widehat{\theta}_S$ is a consistent estimator of $\theta_0 = \theta_S^*$, we have lower sensitivity at θ_S^* but a higher sensitivity from a neighborhood slightly away from θ_S^*. Similarly $\theta_0 = \theta_{\overline{S}}^*$ where $\overline{S}$ is a statistic different from S and also if $\lim \widehat{\theta}_{\overline{S}} = \theta^*_{\overline{S}}$. Then it is possible that $S\left(\mathbf{y}, \theta^*_{\overline{S}}\right)$ is less sensitive.

This has been observed in the case of Durbin-Watson statistic under ARMA misspecification (Banerjee and Magnus, 1999). Recall that $DW \approx 2\left(1 - \widehat{\theta}_{MLE}\right)$ where $\widehat{\theta}_{MLE}$ (the estimated AR(1) parameter) is obtained from minimising sum of squares of error in a linear regression model (Durbin and Watson, 1971) and also as an approximation to the Maximum Likelihood Estimator (MLE) of θ. This will be in general true for any MLE as well. But when they looked at the sensitivity of the slope we found that there is less sensitivity.

Though we are not interested in asymptotics, these observations give us some indicator of sensitivity of $S(\mathbf{y}, \theta_0)$ in small samples.

Asymptotics of large dimensional estimates

In the previous discussion we assumed the dimension of $S(\mathbf{y}, \theta)$ fixed (i.e. K). Now suppose that $\lim_{T\to\infty} K \to \infty$, for example $S(\mathbf{y}, \theta)$ is a predictor of a linear model, then theorem (3) may not be valid. Analysing such cases requires us to impose some more structures on the statistic $S(\mathbf{y}, \theta)$ which we define below.

Suppose $S(\mathbf{y}, \theta) = (S_1(\mathbf{y}, \theta), ..., S_k(\mathbf{y}, \theta), ..., S_K(\mathbf{y}, \theta))$ such that

$$S_k(\mathbf{y}, \theta) = s\left(y_{\pi_k}; \theta\right), \tag{3.22}$$

where ${\pi_k}^2 \in \{(t_1, t_2, ..., t_m) : t_1 < t_2 < ... < t_m\}$ all possible subsets of $\{1, .., T\}$ of size m and $s(., \theta) : \mathbb{R}^m \to \mathbb{R}$ is symmetric in its y_{π_k} arguments for every θ and twice differentiable w.r.t. θ. Note that $K = \binom{T}{m}$.

From (2.12) we have

$$\widehat{\theta}_S - \theta_0 = -\frac{\sum_{k=1}^{K} \nabla^{(1)} s\left(y_{\pi_k}; \theta_0\right) \nabla^{(2)} s\left(y_{\pi_k}; \theta_0\right)}{\sum_{k=1}^{K} \left(\nabla^{(2)} s\left(y_{\pi_k}; \theta_0\right)\right)^2}.$$

The following theorem summarises the asymptotic analysis of $\widehat{\theta}_S$.

[2]This is a slight abuse of notation, not to be confused with π_α probabilities.

Theorem 5 *Let* $\mathbf{y}$ *be a vector of i.i.d random variables also* $E_\theta\left(\nabla^{(1)}s\left(y_{\pi_1};\theta\right)\right)^4<\infty$ *and* $E_\theta\left(\nabla^{(2)}s\left(y_{\pi_1};\theta\right)\right)^4<\infty$ *then:*

1.
$$\operatorname*{plim}_{T\to\infty}\widehat{\theta}_S-\theta_0\to\frac{\mu_{(12)}\left(\theta_0\right)}{\mu_{(22)}\left(\theta_0\right)} \tag{3.23}$$

2.
$$\sqrt{T}\left(\widehat{\theta}_S-\theta_0\right)\to N\left(\frac{\mu_{(12)}\left(\theta_0\right)}{\mu_{(22)}\left(\theta\right)},\frac{m^2\sigma^2_{(12)}\left(\theta_0\right)}{\left[\mu_{(22)}\left(\theta\right)\right]^2}\right), \tag{3.24}$$

where $\mu_{(12)}\left(\theta\right)=E_\theta\left(\nabla^{(1)}s\left(y_{\pi_1};\theta\right)\nabla^{(2)}s\left(y_{\pi_1};\theta\right)\right)$, $\mu_{(22)}\left(\theta\right)=E_\theta\left(\nabla^{(2)}s\left(y_{\pi_1};\theta\right)\right)^2$ *and* $\sigma^2_{(12)}\left(\theta\right)=Var_{y_1}(E_{y_{-1}}(\nabla^{(1)}s\left(y_{\pi_1};\theta\right)\nabla^{(2)}s\left(y_{\pi_1};\theta\right))$

Proof of Theorem 5: Since $\nabla^{(1)}s\left(y_{\pi_k};\theta_0\right)$ and $\nabla^{(2)}s\left(y_{\pi_k};\theta_0\right)$ are also symmetric in y_{π_k}, it is easily seen $\sum_{k=1}^{K}\nabla^{(1)}s\left(y_{\pi_k};\theta_0\right)\nabla^{(2)}s\left(y_{\pi_k};\theta_0\right)$ and $\sum_{k=1}^{K}\left(\nabla^{(2)}s\left(y_{\pi_k};\theta_0\right)\right)^2$ are U-Statistics. It follows from Hoeffding (1948):

$$\lim_{T\to\infty}T^{-1/2}\sum_{k=1}^{K}\nabla^{(1)}s\left(y_{\pi_k};\theta_0\right)\nabla^{(2)}s\left(y_{\pi_k};\theta_0\right)\overset{D}{=}N(\mu^{(12)}\left(\theta\right),m^2\sigma^2_{(12)}\left(\theta\right))$$
$$\text{plim}_{T\to\infty}T^{-1}\sum_{k=1}^{K}\left[\nabla^{(2)}s\left(y_{\pi_k};\theta_0\right)\right]^2=\mu_{(22)}\left(\theta\right)$$

respectively. Therefore (3.24) and (3.23) follows.
QED.

Remark 2 *The theorem (5) can be generalised by assuming stationarity of* y_t*. See for example:* $\beta-mixing$ *(Arcones, 1995),* $\alpha-mixing$ *or* $m-dependence$ *(Lee, 1990).*

Remark 3 *The theorem (5) is also valid for weighted U-statistics as well (Hsing and Wu 2004).*

Theorem (5) states that, $\widehat{\theta}_S\to\theta_0$ if and only if $\mu^{(12)}\left(\theta_0\right)=0$. Therefore, we make following analogous but a weaker form of (1). The following assumption guarantees existence and uniqueness of the least sensitive θ.

Assumption 2 *Assume $\mu^{(12)}(\theta)$ is strictly monotone and*

$$\mu^{(12)}(\underline{\theta}) < 0 < \mu^{(12)}\left(\overline{\theta}\right).$$

This implies that $S(\mathbf{y}, \theta_S^*)$ is the least sensitive estimator where θ_S^* is such that $\mu^{(12)}(\theta_S^*) = 0$. Otherwise we need to modify this estimator componentwise as,

$$\begin{aligned} S_k\left(\mathbf{y}, \widehat{\theta}_S\right) &= s\left(y_{\pi_k}; \widehat{\theta}_S\right) \\ &= s\left(y_{\pi_k}; \theta_0\right) + \frac{\sum_{k=1}^K \nabla^{(1)} s\left(y_{\pi_k}; \theta_0\right) \nabla^{(2)} s\left(y_{\pi_k}; \theta_0\right)}{\sum_{k=1}^K \left(\nabla^{(2)} s\left(y_{\pi_k}; \theta_0\right)\right)^2} \nabla^{(1)} s\left(y_{\pi_k}; \theta_0\right). \end{aligned}$$

Asymptotically,

$$\operatorname*{plim}_{T\to\infty} s\left(y_{\pi_k}; \widehat{\theta}_S\right) = s\left(y_{\pi_k}; \theta_0\right) + \frac{\mu_i^{(12)}(\theta_0)}{\mu_i^{(22)}(\theta_0)} \nabla^{(1)} s\left(y_{\pi_k}; \theta_0\right).$$

Then we obtain $S\left(\mathbf{y}, \widehat{\theta}_S\right)$ which is now less sensitive than the original statistic $S(\mathbf{y}, \theta_0)$. In the next section, we shall consider the statistics generated by linear models as an example of our method described in this section.

4 Least Sensitive Estimates of the Slope and Variance of the Linear Model

In the previous section we have considered a general sensitivity statistic, analysed its properties and proposed how to modify the statistic of interest. In this section we shall consider the standard linear regression model and its estimates as an application of our method. Consider the linear model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{4.25}$$

where $\mathbf{y}$ is an $T \times 1$ random vector of observations, $\mathbf{X}$ a non-random $T \times l$ matrix of regressors, $\boldsymbol{\beta}$ a $l \times 1$ vector of unknown parameters and u an $T \times 1$ vector of random disturbances. We assume that $\mathbf{X}$ has full column-rank l and that $\mathbf{u}$ follows,

$$\mathbf{u} \sim (0, \sigma^2 \boldsymbol{\Omega}(\theta)), \tag{4.26}$$

where $\sigma^2 > 0$ and $\mathbf{\Omega}(\theta)$ is a matrix function of the $p \times 1$ parameter vector $\theta = (\theta_1, \ldots, \theta_p)'$, positive definite and differentiable at least in a neighborhood of $\theta = 0$. Without loss of generality we may assume that

$$\mathbf{\Omega}(0) = \mathbf{I}. \tag{4.27}$$

Our parameters of statistics of interest are the estimates of the slope parameter $\boldsymbol{\beta}$ and the variance σ^2.

As before, we shall analyse the model for $p = 1$ and for simplicity of notation assume $\theta_0 = 0$. We define the $T \times T$ symmetric matrix as:

$$\mathbf{A}^{(k)} = \left.\frac{\partial^k \mathbf{\Omega}^{-1}(\theta)}{\partial \theta^k}\right|_{\theta=0}, k = 1, 2. \tag{4.28}$$

Finally we define,

$$\begin{aligned}
\mathbf{C}_\theta^{(1)} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}^{(1)}, \\
\mathbf{C}_\theta^{(2)} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{A}^{(2)} - 2\mathbf{A}^{(1)}(\mathbf{I} - \mathbf{M})\mathbf{A}^{(1)}\right),
\end{aligned}$$

where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the usual idempotent matrix.

Banerjee and Magnus (1999) define the sensitivity statistic for the predictor $\hat{\mathbf{y}}(\theta)$ (with respect to θ) as: $\mathbf{z}^{(1)} = \nabla^{(1)}\hat{\mathbf{y}}(0)$ and variance $\widehat{\sigma^2}(\theta)$ as $\lambda^{(1)} = \nabla^{(1)}\ \widehat{\sigma^2}(0)$. In this section we shall consider more general linear combinations of the slope estimates $\mathbf{R}\hat{\boldsymbol{\beta}}(\theta)$ where $\mathbf{R}$ $(r \times l)$ is a fixed matrix of full rank.

Define

$$\mathbf{w}_\theta^{(k)} = \mathbf{R}\nabla^{(k)}\widehat{\boldsymbol{\beta}}(0), k = 1, 2 \tag{4.29}$$

the k^{th} order the sensitivity of $\mathbf{R}\widehat{\beta}(\theta)$ (with respect to θ).

In order to use the $r \times 1$ vector as a sensitivity statistic, we can transform $\mathbf{w}^{(1)}$ into a quadratic form in the usual way as:

$$B = \frac{\mathbf{w}_\theta^{(1)\prime}\left(\mathbf{C}_\theta^{(1)}\mathbf{R}\mathbf{R}'\mathbf{C}_\theta^{(1)\prime}\right)^{-}\mathbf{w}_\theta^{(1)}}{(T-k)\hat{\sigma}^2(0)}, \tag{4.30}$$

where $\hat{\sigma}^2(0)$ is the OLS estimator of the variance.

Asking a similar question on our variance estimate $\hat{\sigma}^2(\theta)$, we define the sensitivity of $\hat{\sigma}^2(\theta)$ (with respect to θ) as

$$\lambda_\theta^{(k)} = \nabla^{(k)}\hat{\sigma}^2(0), \;\; k = 1, 2. \tag{4.31}$$

The following theorem obtains the closed form expressions of $\mathbf{w}_\theta^{(k)}$ and $\lambda_\theta^{(k)}$.

Theorem 6 *At $\theta = 0$,the sensitivity measures of $\mathbf{R}\widehat{\boldsymbol{\beta}}(\theta)$ is given by*

1)

$$\mathbf{w}_\theta^{(1)} = \mathbf{R}\mathbf{C}_\theta^{(1)}\widehat{\mathbf{e}} \text{ and } \mathbf{w}_\theta^{(2)} = \mathbf{R}\mathbf{C}_\theta^{(2)}\widehat{\mathbf{e}},$$

where $\widehat{\mathbf{e}} = \mathbf{M}\mathbf{y}$.

2) The sensitivity of the variance is given by,

$$\lambda_\theta^{(1)} = \frac{\widehat{\mathbf{e}}'\mathbf{A}^{(1)}\widehat{\mathbf{e}}}{T-l} \text{ and } \lambda_\theta^{(2)} = \frac{\widehat{\mathbf{e}}'\left(\mathbf{A}^{(2)} - 2\mathbf{A}^{(1)}\left(\mathbf{I}-\mathbf{M}\right)\mathbf{A}^{(1)}\right)\widehat{\mathbf{e}}}{T-l}.$$

Proof of Theorem 6: See Appendix.

Using the equation (2.12) we can estimate $\widehat{\theta}_{R\beta}$ which gives the least sensitive estimate of $\mathbf{R}\widehat{\boldsymbol{\beta}}(\theta)$ as

$$\begin{aligned} \widehat{\theta}_{R\beta} &= \frac{\mathbf{w}_\theta^{(1)\prime}\mathbf{w}_\theta^{(2)}}{\mathbf{w}_\theta^{(2)\prime}\mathbf{w}_\theta^{(2)}} \\ &= \frac{\widehat{\mathbf{e}}'\mathbf{C}_\theta^{(1)\prime}\mathbf{R}'\mathbf{R}\mathbf{C}_\theta^{(2)}\widehat{\mathbf{e}}}{\widehat{\mathbf{e}}'\mathbf{C}_\theta^{(2)\prime}\mathbf{R}'\mathbf{R}\mathbf{C}_\theta^{(1)}\widehat{\mathbf{e}}}. \end{aligned} \tag{4.32}$$

We transform $\widehat{\theta}_{R\beta}$ as:

$$\widehat{\theta}_{R\beta} = \frac{1}{2}\frac{\widehat{\mathbf{e}}'\left[\mathbf{C}_\theta^{(2)\prime}\mathbf{R}'\mathbf{R}\mathbf{C}^{(1)} + \mathbf{C}_\theta^{(1)\prime}\mathbf{R}'\mathbf{R}\mathbf{C}^{(2)}\right]\widehat{\mathbf{e}}}{\widehat{\mathbf{e}}'\mathbf{C}_\theta^{(2)\prime}\mathbf{R}'\mathbf{R}\mathbf{C}_\theta^{(2)}\widehat{\mathbf{e}}},$$

such that $\widehat{\theta}_{R\beta}$ is a ratio of a symmetric quadratic form to a non-negative definite quadratic form.

As a special case, consider a linear combination of $\boldsymbol{\beta}'s$ i.e. $\mathbf{r}\widehat{\boldsymbol{\beta}}(\theta)$ where $\mathbf{r}$ $(1 \times l)$ vector, then from (2.13), we have

$$\widehat{\theta}_{R\beta} = \frac{\mathbf{r}\mathbf{C}^{(1)\prime}\widehat{\mathbf{e}}}{\mathbf{r}\mathbf{C}^{(2)}\widehat{\mathbf{e}}}. \tag{4.33}$$

Also using (2.13) we can estimate $\widehat{\theta}_\sigma$, which gives the least sensitive estimate of $\hat{\sigma}^2(\theta)$ as

$$\begin{aligned}\widehat{\theta}_\sigma &= \frac{\lambda^{(1)}}{\lambda^{(2)}} \\ &= \frac{\widehat{\mathbf{e}}'\mathbf{A}^{(1)}\widehat{\mathbf{e}}}{\widehat{\mathbf{e}}'\left[\mathbf{A}^{(2)} - 2\mathbf{A}^{(1)}\left(\mathbf{I}-\mathbf{M}\right)\mathbf{A}^{(1)}\right]\widehat{\mathbf{e}}},\end{aligned}$$

in case of no independent variables $\widehat{\theta}_{R\beta}$ remains undefined and

$$\widehat{\theta}_\sigma = \frac{\mathbf{y}'\mathbf{A}^{(1)}\mathbf{y}}{\mathbf{y}'\mathbf{A}^{(2)}\mathbf{y}} \tag{4.34}$$

Properties:

1. The sensitivity statistic $\mathbf{w}_\theta^{(k)}$ and $\lambda_\theta^{(k)}$ are independent of $\boldsymbol{\beta}$.

2. Since $\mathbf{R}\widehat{\beta}(\theta)$ and $\hat{\sigma}^2(\theta)$ are in homogenous of order one and two in $\mathbf{y}$ respectively, then by Theorem (2) $\widehat{\theta}_{R\beta}$ and $\widehat{\theta}_\sigma$ are scale invariant.

3. Further as $\widehat{\mathbf{e}} = \mathbf{My}$, it follows that $\widehat{\theta}_{R\beta}$ and $\widehat{\theta}_\sigma$ are regression invariant as well.

4. Since $\widehat{\mathbf{e}} = \mathbf{Mu}$ and $\mathbf{u} \sim (0, \sigma^2\mathbf{I})$ when $\theta = 0$, $\widehat{\theta}_{R\beta}$ and $\widehat{\theta}_\sigma$ are ratios of quadratic forms of independent random variables. Thus both converges to a distribution at rate of $\sqrt{T}$. Whittle (1964) gives sufficient conditions for such limiting distributions to be normal even when $\mathbf{u}$ is not normal.

5. We can then use methods like IMHOF to compute the finite sample distribution of $\widehat{\theta}_{R\beta}$, assuming normality of $\mathbf{u}$.

6. We can use similar methods for $\widehat{\theta}_\sigma$ as well if $\mathbf{A}^{(2)} - 2\mathbf{A}^{(1)}\left(\mathbf{I}-\mathbf{M}\right)\mathbf{A}^{(1)}$ is positive semi-definite, otherwise we shall have to use simulation.

A closed form distribution of $\widehat{\theta}_{R\beta}$ when $\mathbf{r}\boldsymbol{\beta}$ is scalar can be obtained in the following theorem:

Theorem 7 *The distribution of $\widehat{\theta}_{R\beta}$ defined by (4.33) (i.e. when $\mathbf{r}$ is a $(1 \times l)$ vector) and $\mathbf{u} \sim N(0, \sigma^2\boldsymbol{\Omega}(\theta))$ is given by*

$$F_\theta\left(\widehat{\theta}_{R\beta}\right) = \frac{1}{2} + \frac{\sqrt{2}}{\sqrt{\pi}} \arctan\left(\frac{\sigma_2\widehat{\theta}_{R\beta} - \rho\sigma_1}{\sigma_1\sqrt{1-\rho^2}}\right),$$

where

$$\begin{aligned}\sigma_1^2 &= \sigma^2 \mathbf{r}'\mathbf{C}_\theta^{(1)\prime}\mathbf{M}\boldsymbol{\Omega}(\theta)\mathbf{M}\mathbf{C}_{\boldsymbol{\theta}}^{(\mathbf{1})}\mathbf{r} \\ \sigma_2^2 &= \sigma^2 \mathbf{r}'\mathbf{C}_\theta^{(2)\prime}\mathbf{M}\boldsymbol{\Omega}(\theta)\mathbf{M}\mathbf{C}_{\boldsymbol{\theta}}^{(\mathbf{2})\prime}\mathbf{r}\end{aligned}$$

and

$$\rho = \sigma^2 \frac{\mathbf{r}'\mathbf{C}_\theta^{(1)\prime}\mathbf{M}\boldsymbol{\Omega}(\theta)\mathbf{M}\mathbf{C}_{\boldsymbol{\theta}}^{(\mathbf{2})}\mathbf{r}}{\sigma_1\sigma_2}$$

Proof of Theorem 7: See Appendix.

Let us consider some specific examples. The first example is of a two-error components model where we shall exactly calculate the correction factors and show $\widehat{\theta}_{R\beta}$ is constant when there is no intercept term in the model. The other is the standard ARFIMA process, for which we calculate the obtain closed form parameter estimates of the correction factor. We also show the relation between the parameter estimates and the usual autocorrelation functions.

Two-error components model.

Consider the two component model

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \\ \mathbf{u} &\simeq (0, \sigma^2 (\mathbf{E}+\theta(\mathbf{I}-\mathbf{E}))),\end{aligned}$$

where

$$\mathbf{E} = \frac{1}{T}\mathbf{1}\mathbf{1}', \ \mathbf{1} = (1, 1, \ldots, 1)'.$$

Differentiating we have

$$\mathbf{A}^{(1)} = (\mathbf{I}-\mathbf{E}) \text{ and } \mathbf{A}^{(2)} = \mathbf{0}.$$

If the regression contains an intercept, so that $\mathbf{M1} = 0$, then it is easy to see that

$$\mathbf{w}_\theta^{(1)} = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left[\mathbf{I}-\mathbf{E}\right]\mathbf{M}\mathbf{y} = \mathbf{0}.$$

In fact, $\mathbf{R}\widehat{\boldsymbol{\beta}}(\theta)$ do not depend on θ at all in this case, because the two-error components model (with constant term) is one example where GLS = OLS, that is, $(\mathbf{X}'\boldsymbol{\Omega}^{-1}(\theta)\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}(\theta)\mathbf{y}$ = $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ for every θ. This has been shown in Banerjee and Magnus (1999).

Theorem 8 *In case of a two component model without the constant i.e.* $\mathbf{M1} \neq 0$, $\widehat{\theta}_{R\beta}$ *and* $\widehat{\theta}_{\sigma}$ *is given by:*

$$\widehat{\theta}_{R\beta} = \frac{1}{2\mu} \text{ and } \widehat{\theta}_{\sigma} = \frac{1}{2(1-\mu)} \frac{\widehat{\mathbf{e}}'(\mathbf{I}-\mathbf{E})\widehat{\mathbf{e}}}{\widehat{\mathbf{e}}'\mathbf{E}\widehat{\mathbf{e}}},$$

respectively where $\mu = \frac{1}{T}\mathbf{1}'\mathbf{M1}$.

Proof of Theorem 8: See Appendix.

Notice that the best estimate of the nuisance parameter when adjusting for the slope is constant whereas for the variance is stochastic. Also the distribution of $\widehat{\theta}_{\sigma}$ can be calculated using IMHOF method since $\widehat{\mathbf{e}}'\mathbf{E}\widehat{\mathbf{e}}$ is positive semidefinite.

Using Theorem (8) we obtain the least sensitive estimate of $\mathbf{R}\boldsymbol{\beta}$ by adjusting the OLS estimate $\mathbf{R}\hat{\boldsymbol{\beta}}(0)$ as

$$\mathbf{R}\hat{\boldsymbol{\beta}}(\widehat{\theta}_{R\beta}) = \mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{I} - \frac{1}{2\mu}\mathbf{EM}\right)\mathbf{y}.$$

The modified least sensitive estimate of the variance is given by,

$$\widehat{\sigma}^2(\widehat{\theta}_{\sigma}) = \widehat{\sigma}^2(0) + \frac{1}{2(1-\mu)} \frac{\left[\widehat{\mathbf{e}}'(\mathbf{I}-\mathbf{E})\widehat{\mathbf{e}}\right]^2}{\widehat{\mathbf{e}}'\mathbf{E}\widehat{\mathbf{e}}}.$$

ARFIMA Time series.

If the y_t process is covariance stationary, then $\boldsymbol{\Omega}(\theta)$ can be written as

$$\boldsymbol{\Omega}(\theta) = \mathbf{I} + \sum_{h=1}^{T-1} \omega_h(\boldsymbol{\theta})\mathbf{T}^{(h)}, \tag{4.35}$$

where $\mathbf{T}^{(h)}, 0 \leq h \leq T-1$, the $T \times T$ symmetric Toeplitz matrix with

$$\mathbf{T}^{(h)}(i,j) = \begin{cases} 1 & \text{if } |i-j| = h, \\ 0 & \text{otherwise.} \end{cases}$$

If the errors u_t are distributed as ARFIMA(p,d,q) process:

$$\Delta^d u_t = \sum_{i=1}^{p} \phi_i \Delta^d u_{t-1} + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \varepsilon_t \tag{4.36}$$

with innovations $\varepsilon_1, \ldots, \varepsilon_T \sim$i.i.d. $(0, \sigma^2)$ and $\theta = (\phi_1, .., \phi_p, d, \psi_1, ..., \psi_q)$.

Theorem 9 *Let $\sigma^2\mathbf{\Omega}(\theta)$ be the covariance matrix of $u_1, \ldots, u_n$. Then,*

$$\left.\frac{\partial\mathbf{\Omega}\left(\theta\right)}{\partial\phi_i}\right|_{\theta=0} = \left.\frac{\partial\mathbf{\Omega}(\theta)}{\partial\psi_i}\right|_{\theta=0} = \mathbf{T}^{(i)}, \tag{4.37}$$

$$\left.\frac{\partial^2\mathbf{\Omega}\left(\theta\right)}{\partial\phi_i^2}\right|_{\theta=0} = 2\left(\mathbf{I}+\mathbf{T}^{(2i)}\right), \tag{4.38}$$

$$\left.\frac{\partial^2\mathbf{\Omega}(\theta)}{\partial\psi_i^2}\right|_{\theta=0} = 2\mathbf{I} \tag{4.39}$$

and

$$\left.\frac{\partial\mathbf{\Omega}\left(\theta\right)}{\partial d}\right|_{\theta=0} = \sum_{t=1}^{T-1}\frac{1}{t}\mathbf{T}^{(t)}, \tag{4.40}$$

$$\left.\frac{\partial^2\mathbf{\Omega}\left(\theta\right)}{\partial d^2}\right|_{\theta=0} = \frac{\pi^2}{3}\mathbf{I}+\sum_{t=1}^{T-1}\frac{4\left(t\Psi(t)+\gamma\right)+2}{t^2}\mathbf{T}^{(t)}, \tag{4.41}$$

where $\Psi(t) = \frac{\Gamma'(t)}{\Gamma(t)}$ is the polygamma function and γ is the Euler's constant.

Proof of Theorem 9: See Appendix.

From (4.37 to 4.38) and since $\mathbf{M}$ is idempotent we have the least sensitive AR estimates for $\mathbf{R}\boldsymbol{\beta}$ as

$$\widehat{\phi}_{i,R\beta} = \frac{\widehat{\mathbf{e}}'\mathbf{C}_{\phi}^{(1)\prime}\mathbf{R}'\mathbf{R}\mathbf{C}_{\phi}^{(2)}\widehat{\mathbf{e}}}{\widehat{\mathbf{e}}'\mathbf{C}_{\phi}^{(2)\prime}\mathbf{R}'\mathbf{R}\mathbf{C}_{\phi}^{(2)}\widehat{\mathbf{e}}}$$

and the MA estimates as

$$\widehat{\psi}_{i,R\beta} = \frac{\widehat{\mathbf{e}}'\mathbf{C}_{\psi}^{(1)\prime}\mathbf{R}'\mathbf{R}\mathbf{C}_{\psi}^{(2)}\widehat{\mathbf{e}}}{\widehat{\mathbf{e}}'\mathbf{C}_{\psi}^{(2)}\mathbf{R}'\mathbf{R}\mathbf{C}_{\psi}^{(2)}\widehat{\mathbf{e}}},$$

where

$$\begin{aligned}
\mathbf{C}_{\phi}^{(1)} &= \mathbf{C}_{\psi}^{(1)} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{T}^{(i)}, \\
\mathbf{C}_{\phi}^{(2)} &= 2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{T}^{(2i)} - \mathbf{C}_{\psi}^{(2)}, \\
\mathbf{C}_{\psi}^{(2)} &= 2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{T}^{(i)}\left(\mathbf{I}-\mathbf{M}\right)\mathbf{T}^{(i)}.
\end{aligned}$$

We obtain the least sensitive AR and MA estimates for the variance using (4.39 to 4.38) as

$$\widehat{\phi}_{i,\sigma} = \frac{1}{2}\frac{\widehat{\mathbf{e}}'\mathbf{T}^{(i)}\widehat{\mathbf{e}}}{\widehat{\mathbf{e}}'\mathbf{T}^{(2i)}\widehat{\mathbf{e}}+\widehat{\mathbf{e}}'\left[\mathbf{I}-\mathbf{T}^{(i)}\left(\mathbf{I}-\mathbf{M}\right)\mathbf{T}^{(i)}\right]\widehat{\mathbf{e}}}$$

and

$$\widehat{\psi}_{i,\sigma} = \frac{1}{2}\frac{\widehat{\mathbf{e}}'\mathbf{T}^{(i)}\widehat{\mathbf{e}}}{\widehat{\mathbf{e}}'\left[\mathbf{I}-\mathbf{T}^{(i)}\left(\mathbf{I}-\mathbf{M}\right)\mathbf{T}^{(i)}\right]\widehat{\mathbf{e}}},$$

respectively. Notice that $\widehat{\phi}_{i,\sigma}$ and $\widehat{\psi}_{i,\sigma}$ can be written as:

$$\widehat{\phi}_{i,\sigma} = \frac{\rho_{(i)}}{1 + 2\rho_{(2i)} - \overline{\rho_{(i)}}} \text{ and } \widehat{\psi}_{i,\sigma} = \frac{\rho_{(i)}}{1 - \overline{\rho_{(i)}}},$$

where $\rho_{(i)} = \frac{1}{2}\frac{\widehat{\mathbf{e}}'\mathbf{T}^{(i)}\widehat{\mathbf{e}}}{\widehat{\mathbf{e}}'\widehat{\mathbf{e}}}$ is the i^{th} autocorrelation function and $\overline{\rho_{(i)}} = \frac{\widehat{\mathbf{e}}'\mathbf{T}^{(i)}(\mathbf{I}-\mathbf{M})\mathbf{T}^{(i)}\widehat{\mathbf{e}}}{\widehat{\mathbf{e}}'\widehat{\mathbf{e}}}$. If $\mathbf{X} = \mathbf{1}$ then

$$\overline{\rho_{(i)}} = \frac{\left[\sum_t \left(\widehat{e}_{t-1} + \widehat{e}_{t+1}\right)\right]^2}{\sum_t \widehat{e}_t^2}.$$

The distribution of $\widehat{\phi}_{i,\sigma}$ and $\widehat{\psi}_{i,\sigma}$ at $\theta = 0$ is given by the following theorem.

Theorem 10 *Given $\theta = 0$,*

$$\sqrt{T}\widehat{\phi}_{i,\sigma} \overset{asy}{=} \sqrt{T}\widehat{\psi}_{i,\sigma} \overset{asy}{\sim} N(0,1).$$

for all i.

Proof of Theorem 10: See Appendix.

The least sensitive variance estimates of MA and AR parameters also gives us an elegant relationship between them as

$$\frac{1}{\widehat{\psi}_{i,\sigma}} - \frac{1}{\widehat{\phi}_{i,\sigma}} = 2\frac{\rho_{(2i)}}{\rho_{(i)}}. \tag{4.42}$$

This leads us to the following theorem, which helps us to quantify the difference of influence on the variance sensitivity due to the AR and MA parameters.

Theorem 11 *Let $(u_1, ..., u_t, ...u_T)$ be i.i.d random variables then,*

$$\lim_{T\to\infty} \Pr\left(\frac{1}{2}\left(\frac{1}{\widehat{\psi}_{i,\sigma}} - \frac{1}{\widehat{\phi}_{i,\sigma}}\right) \leq x\right) = \frac{1}{2} + \sqrt{\frac{2}{\pi}}\arctan(x).$$

Proof of Theorem 11: Follows easily from the asymptotic distribution of $\rho_{(i)}$ and $\rho_{(2i)}$.

Further simplification of the estimates of the ARFIMA process $\widehat{\theta}_\sigma = (\widehat{\phi}_{1,\sigma}, ..., \widehat{\phi}_{p,\sigma}, \widehat{d}_\sigma, \widehat{\psi}_{1,\sigma}, ..., \widehat{\psi}_{i,\sigma})$ can be done in the absence of any regressors. Using (4.34) we have

$$\begin{aligned} \widehat{\phi}_{i,\sigma} &= \frac{\rho_{(i)}}{1 + 2\rho_{(2i)}}, \\ \widehat{\psi}_{i,\sigma} &= \rho_{(i)}. \end{aligned}$$

The estimate of the long memory parameter is given by:

$$\widehat{d}_\sigma = \frac{\sum_{t=1}^{T-1} \frac{1}{t}\rho_{(t)}}{\frac{\pi^2}{6} + \sum_{t=1}^{T-1} \frac{4(t\Psi(t)+\gamma)+2}{t^2}\rho_{(t)}}.$$

We find the least sensitive estimate $\hat{\sigma}^2(\widehat{\theta}_\sigma)$ by adjusting the OLS estimate $\hat{\sigma}^2(0)$ as

$$\hat{\sigma}^2(\widehat{\theta}_\sigma) = \hat{\sigma}^2(0) + 2\hat{\sigma}^2(0)\left(\sum_{i=1}^{q} \widehat{\phi}_{i,\sigma}\rho_{(i)} + \sum_{i=1}^{q} \widehat{\phi}_{i,\sigma}\rho_{(i)} + \widehat{d}_\sigma \sum_{t=1}^{T-1} \frac{1}{t}\rho_{(t)}\right)$$

This corrected estimator for σ^2 is similar to a Newey-West (Newey and West, 1987) long run variance estimator where the weights are the least sensitive estimates of the nuisance parameters. Therefore we can interpret $\hat{\sigma}^2(\widehat{\theta}_\sigma)$ as the most stable long-run variance estimator.

5 Risk functions and sensitivity of central tendency estimates.

In this section we provide an example where our method of correction might fail.

We consider the problem of estimation of the central tendency μ of the data under general risk function $R(\mu, \theta) = E\mathfrak{L}(y - \mu, \theta)$ where θ is nuisance parameter comes from the loss function $\mathfrak{L}(y - \mu, \theta)$. The corresponding empirical risk function for the central tendency parameter μ is given by:

$$\widehat{R}(\mu, \theta) = \frac{1}{T}\sum_{t=1}^{T} \mathfrak{L}(y_t - \mu, \theta)$$

As θ varies so will the estimator of μ which is obtained by minimising the risk function $R(\mu, \theta)$ with respect to μ. That is the estimate $\widehat{\mu}(\theta)$ is a solution to the equation:

$$\widehat{R}_\mu(\widehat{\mu}(\theta), \theta) = 0 \tag{5.43}$$

where $\widehat{\mu}(\theta)$ is the minimiser.

Following our discussion on sensitivity of the estimate with respect to nuisance parameter θ, we obtain the sensitivity of $\widehat{\mu}(\theta)$ by implicitly differentiating (5.43) with respect to θ as:

$$\widehat{\mu}_\theta(\theta) = -\frac{\widehat{R}_{\mu\theta}(\widehat{\mu}(\theta), \theta)}{\widehat{R}_{\mu\mu}(\widehat{\mu}(\theta), \theta)}. \tag{5.44}$$

The following theorem obtains the expressions of $\widehat{\theta}_\mu$ the value of θ for which the estimate of the central tendency $\widehat{\mu}(\theta)$ has the least sensitivity.

Theorem 12 *The the least sensitive estimate of θ is:*

$$\widehat{\theta}_\mu = -\frac{\widehat{R}_{\mu\theta}(\widehat{\mu}(\theta),\theta)\left[\widehat{R}_{\mu\mu}(\widehat{\mu}(\theta),\theta)\right]^2}{\widehat{R}_{\mu\mu\mu}(\widehat{\mu}(\theta),\theta)\left[\widehat{R}_{\mu\theta}(\widehat{\mu}(\theta),\theta)\right]^2 - \widehat{R}_{\mu\theta\mu}(\widehat{\mu}(\theta),\theta)\widehat{R}_{\mu\theta}(\widehat{\mu}(\theta),\theta) + \widehat{R}_{\mu\theta\theta}(\widehat{\mu}(\theta),\theta)\widehat{R}_{\mu\mu}(\widehat{\mu}(\theta),\theta)}.$$

Proof of Theorem 12**:** See Appendix.

Linex loss function

As an example we consider the Linex loss function introduced by Varian (1975) which is given by:

$$\mathfrak{L}(y,\theta) = \frac{\exp(\theta y) - \theta y - 1}{\theta^2}.$$

Note that $\lim_{\theta\to 0}\mathfrak{L}(y,\theta) = y^2$ the usual quadratic loss function and also $\lim_{\theta\to 0}\widehat{\mu}(\theta) = \overline{y}$ and $\lim_{\theta\to 0}\widehat{R}(\mu,\theta) = \widehat{m}^{(2)}(y) = Var(y)$ the usual estimates of mean and variance. These observations makes $\theta_0 = 0$ an interesting point to measure the sensitivity of the sample mean $\overline{y}$ against small asymmetric risk measures and what are the pitfalls of sensitivity correction. The following theorem gives us the relevant expressions of the sensitivity $\widehat{\mu}_\theta(\theta)$ and the proposed corrections in form of $\widehat{\theta}_\mu$.

Theorem 13 *The sensitivity measure of the estimate of central tendency when the linex loss function is:*

1) $\lim_{\theta\to 0}\widehat{\mu}_\theta(\theta) = \frac{1}{2}\widehat{m}^{(2)}(y)$

the least sensitive asymmetry parameter value is given by

2) $\lim_{\theta\to 0}\widehat{\theta}_\mu = \frac{3}{2}\frac{\widehat{m}^{(2)}(y)}{\widehat{m}^{(3)}(y)}$

and the corrected central tendency measure is given by

3) $\widehat{\mu}\left(\widehat{\theta}_\mu\right) = \overline{y} + \frac{3}{4}\frac{\left[\widehat{m}^{(2)}(y)\right]^2}{m^{(3)}(y)}$,

where $\widehat{m}^{(2)}(y)$ and $\widehat{m}^{(3)}(y)$ is the 2nd and 3rd central moment from the data respectively.

Proof of Theorem13: See Appendix.

Part 1) of theorem (13) shows that the sensitivity measure is proportional to the least-squares risk. Part 2) and 3) gives us the asymmetry parameter of the loss-function in turns of the skewness of the data and shows how to modify the central tendency measure using the skewness. This is similar to the expression obtained by Zellner (1986, pp 447 eq (3.4)) but in this case the asymmetry parameter θ is data dependent and and gives the direction of modification.

Using Weak Law of Large Numbers we have asymptotically

$$\begin{aligned} \underset{T\to\infty}{\text{plim}} \lim_{\theta\to 0} \widehat{\theta}_\mu &= \frac{3}{2}\frac{m^{(2)}}{m^{(3)}} \\ \underset{T\to\infty}{\text{plim}}\, \widehat{\mu}\left(\widehat{\theta}_\mu\right) &= \overline{y} + \frac{3}{4}\frac{\left(m^{(2)}\right)^2}{m^{(3)}} \end{aligned}$$

where $m^{(i)}$ are the central moments of the distribution.

Remark 4 *Note that is the distribution is symmetric then* $\text{plim}_{T\to\infty}\, \widehat{\mu}_{\theta\theta}(\theta) = -\frac{1}{3}m^{(3)} = 0$ *this violates our assumption in theorem (1) and therefore we cannot modify the distribution by the method we proposed, atleast the asymptotic results does not hold. We would need the higher terms in the Taylor expansion as defined in equation (2.11). The other theorems stated will also need to be modified given the presence of higher terms.*

6 Conclusion

The idea of sensitivity was introduced by Banerjee and Magnus (1999) and has been discussed is various contexts by others. One of the unresolved problems of sensitivity analysis was how to correct of sensitiveness if a statistic is found to be sensitive to changes in nuisance parameters.

In this article we provide a sufficiency condition (a NP type lemma) leading to the definition of the sensitivity measure at θ_0 :

$$\nabla_i^{(1)} S(\mathbf{y}, \theta_0) = \left.\frac{\partial S(\mathbf{y}, \theta)}{\partial \theta_i}\right|_{\theta=\theta_0}.$$

We then propose a one-step Newton-Raphson method to decrease sensitivity of the statistic $S(\mathbf{y},\theta)$ to the variations of the unknown nuisance parameters θ near some known value of $\theta = \theta_0$. In doing so we provide a one-step "estimate" of the nuisance parameters $\widehat{\theta}_S$. The correction is done using $\widehat{\theta}_S$'s , with the corrected statistic being

$$S\left(\mathbf{y},\widehat{\theta}_S\right) = S(\mathbf{y},\theta_0) + \sum_{i=1}^{p}\left(\widehat{\theta}_{S,i} - \theta_{0,i}\right)\nabla_i^{(1)} S(\mathbf{y},\theta_0).$$

The separability of the correction factors is an advantage when there is a large number of unknown nuisance parameters. We can add more correction terms without changing the other corrections already done.

We analysed the properties of $\widehat{\theta}_S$ and hence the correction factors under various assumptions in large samples and small samples. Specifically, we studied the slope and the variance estimates of the linear model in presence of nuisance parameters to the covariance matrix. We provide nuisance parameter estimates of the two-error component model and ARFIMA process. The proposed procedure is broad enough to incorporate other non linear models with nuisance parameters like the Heckman Estimator or Tobit models. We also, studied the estimation of central tendency measure under asymmetric risk function prametrised by a nuisance parameter and show conditions under which our methodology fails.

Further extension of the one-step method can be done by using a recursive procedure such as:

$$S\left(\mathbf{y},\hat{\theta}^{(j)}\right) = S\left(\mathbf{y},\hat{\theta}^{(j-1)}\right) + \sum_{i=1}^{p}\left(\hat{\theta}_i^{(j)} - \hat{\theta}_i^{(j-1)}\right)\nabla_i^{(1)} S\left(\mathbf{y},\hat{\theta}^{(j-1)}\right),$$

where $\hat{\theta}^{(j)}$ the j^{th} iterate is obtained conditional on the $j-1^{th}$ iterate $\hat{\theta}^{(j-1)}$ such that $\hat{\theta}^{(0)} = \theta_0$. This iteration is computed using the following equation:

$$\hat{\theta}_i^{(j)} = \hat{\theta}_i^{(j-1)} - \frac{\left\langle \nabla_i^{(1)} S\left(\mathbf{y},\hat{\theta}^{(j-1)}\right), \nabla_i^{(2)} S\left(\mathbf{y},\hat{\theta}^{(j-1)}\right)\right\rangle}{\left\|\nabla_i^{(2)} S\left(\mathbf{y},\hat{\theta}^{(j-1)}\right)\right\|^2},\ i = 1,..,p.$$

Spall (2000) provides conditions for convergence of such stochastic Newton-Raphson method.

Appendix

Proof of Theorem 1: [3]

Since $\left.\frac{\partial^2 S(\mathbf{y},\theta)}{\partial\theta_i^2}\right|_{\theta=\theta_0} > 0$ there exists a neighborhood $[\theta_0, \theta_0 + \eta\ )$ for some $\eta > 0$ such that

$$\left.\frac{\partial S(\mathbf{y},\theta)}{\partial\theta_i}\right|_{\theta=\theta_0} > \frac{\partial S(\mathbf{y},\theta)}{\partial\theta_i}, \tag{6.45}$$

for all $\theta \in [\theta_0, \theta_0 + \eta)$. Now consider the mean value theorem

$$S(\mathbf{y},\theta_0+\eta) - S(\mathbf{y},\theta_0) = \eta \left.\frac{\partial S(\mathbf{y},\theta)}{\partial\theta_i}\right|_{\theta=\xi} \text{ for some } \xi \in [\theta_0, \theta_0+\eta)\,.$$

From (6.45) this implies

$$S(\mathbf{y},\theta_0+\eta) - S(\mathbf{y},\theta_0) = \eta \left.\frac{\partial S(\mathbf{y},\theta)}{\partial\theta_i}\right|_{\theta=\xi} > \eta \left.\frac{\partial S(\mathbf{y},\theta)}{\partial\theta_i}\right|_{\theta=\theta_0} > \eta\lambda > 0.$$

Hence,

$$\Pr_\theta\left(S(\mathbf{y},\theta_0+\eta) - S(\mathbf{y},\theta_0) > \eta\lambda\right) \geq \Pr_\theta\left(\left.\frac{\partial S(\mathbf{y},\theta)}{\partial\theta_i}\right|_{\theta=\theta_0} > \lambda\right),$$

for all $\theta \in [\theta_0, \theta_0 + \eta\ )$.

QED.

Proof of Theorem 6:

Define $\frac{\partial^k \mathbf{\Omega}(\theta)^{-1}}{\partial\theta^k} = \mathbf{A}^{(k)}(\theta)$ and $\widehat{\mathbf{e}}(\theta) = \left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\theta)\right)$. Then,

$$\frac{\partial\widehat{\boldsymbol{\beta}}(\theta)}{\partial\theta} = -\left(\mathbf{X}'\mathbf{\Omega}(\theta)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(1)}(\theta)\,\widehat{\mathbf{e}}(\theta)\,.$$

Hence, at $\theta = 0$

$$\mathbf{b}^{(1)} = -\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{A}^{(1)}\widehat{\mathbf{e}}.$$

[3] For notational simplicity we shall prove the theorm assuming $p = 1$. Generalisation to partial derivatives is trivial.

$$
\begin{aligned}
\frac{\partial^2\widehat{\boldsymbol{\beta}}(\theta)}{\partial\theta^2} &= -\left[\frac{\partial\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}}{\partial\theta}\mathbf{X}\mathbf{A}^{(1)}\left(\theta\right)+\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(2)}\left(\theta\right)\right]\widehat{\mathbf{e}}\left(\theta\right)\\
&\quad+\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(1)}\left(\theta\right)\mathbf{X}\frac{\partial\widehat{\boldsymbol{\beta}}(\theta)}{\partial\theta}\\
&= -\left[\begin{array}{c}\frac{\partial\left(\mathbf{X}'\boldsymbol{\Omega}(\theta)^{-1}\mathbf{X}\right)^{-1}}{\partial\theta}\mathbf{X}\mathbf{A}^{(1)}\left(\theta\right)+\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(2)}\left(\theta\right)\\+\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(1)}\left(\theta\right)\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(1)}\left(\theta\right)\end{array}\right]\widehat{\mathbf{e}}\left(\theta\right)\\
&= -\left[\begin{array}{c}-\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{A}^{(1)}\left(\theta\right)\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(1)}\left(\theta\right)\\+\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(2)}\left(\theta\right)\\-\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(1)}\left(\theta\right)\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(1)}\left(\theta\right)\end{array}\right]\widehat{\mathbf{e}}\left(\theta\right)
\end{aligned}
$$

then at $\theta = 0$

$$
\begin{aligned}
\mathbf{b}^{(2)} &= -\left[\begin{array}{c}-\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{A}^{(1)}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(1)}+\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(2)}\\-\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(1)}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(1)}\end{array}\right]\widehat{\mathbf{e}}\\
&= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left[\mathbf{A}^{(2)}-2\mathbf{A}^{(1)}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(1)}\right]\widehat{\mathbf{e}}=\mathbf{C}_\theta^{(2)}\widehat{\mathbf{e}}.
\end{aligned}
$$

Variance:

$$
\begin{aligned}
(T-l)\lambda^{(1)}\left(\theta\right) &= (T-l)\frac{\partial\widehat{\sigma}^2\left(\theta\right)}{\partial\theta}=\frac{\partial\widehat{\mathbf{e}}'\left(\theta\right)\boldsymbol{\Omega}\left(\theta\right)^{-1}\widehat{\mathbf{e}}\left(\theta\right)}{\partial\theta}\\
&= -2\widehat{\mathbf{e}}'\left(\theta\right)\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\frac{\partial\widehat{\boldsymbol{\beta}}(\theta)}{\partial\theta}+\widehat{\mathbf{e}}'\left(\theta\right)\mathbf{A}^{(1)}\left(\theta\right)\widehat{\mathbf{e}}\left(\theta\right)\\
&= -2\widehat{\mathbf{e}}'\left(\theta\right)\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{A}^{(1)}\left(\theta\right)\widehat{\mathbf{e}}\left(\theta\right)+\widehat{\mathbf{e}}'\left(\theta\right)\mathbf{A}^{(1)}\left(\theta\right)\widehat{\mathbf{e}}\left(\theta\right)\\
&= \widehat{\mathbf{e}}'\left(\theta\right)\mathbf{A}^{(1)}\left(\theta\right)\widehat{\mathbf{e}}\left(\theta\right)
\end{aligned}
$$

and

$$
\begin{aligned}
(T-l)\lambda^{(2)}\left(\theta\right) &= \frac{\partial\lambda_i^{(1)}\left(\theta\right)}{\partial\theta_i}=\frac{\partial\widehat{\mathbf{e}}'\left(\theta\right)\mathbf{A}^{(1)}\left(\theta\right)\widehat{\mathbf{e}}\left(\theta\right)}{\partial\theta_i}\\
&= -2\widehat{\mathbf{e}}'\left(\theta\right)\mathbf{A}^{(1)}\left(\theta\right)\mathbf{X}\frac{\partial\widehat{\boldsymbol{\beta}}(\theta)}{\partial\theta}+\widehat{\mathbf{e}}'\left(\theta\right)\mathbf{A}^{(2)}\left(\theta\right)\widehat{\mathbf{e}}\left(\theta\right)\\
&= \widehat{\mathbf{e}}'\left(\theta\right)\mathbf{A}^{(2)}\left(\theta\right)\widehat{\mathbf{e}}\left(\theta\right)-2\widehat{\mathbf{e}}'\left(\theta\right)\mathbf{A}^{(1)}\left(\theta\right)\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(1)}\left(\theta\right)\widehat{\mathbf{e}}\left(\theta\right)\\
&= \widehat{\mathbf{e}}'\left(\theta\right)\left[\mathbf{A}^{(2)}\left(\theta\right)-2\mathbf{A}^{(1)}\left(\theta\right)\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Omega}\left(\theta\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{A}^{(1)}\left(\theta\right)\right]\widehat{\mathbf{e}}\left(\theta\right).
\end{aligned}
$$

Thus,

$$
(T-l)\lambda^{(2)}=\widehat{\mathbf{e}}'\left[\mathbf{A}^{(2)}-2\mathbf{A}^{(1)}\left(\mathbf{I}-\mathbf{M}\right)\mathbf{A}^{(1)}\right]\widehat{\mathbf{e}}
$$

QED.

Proof of Theorem 7: Following Hinkley (1969) if $(u_1, u_2) \sim BN(0, 0, \sigma_1^2, \sigma_2^2, \rho)$ and let $u = \frac{u_1}{u_2}$, the the distribution of w is given by

$$F(u) = 2\mathcal{L}\left(\frac{\sigma_2 u - \rho\sigma_1}{\sqrt{u^2\sigma_2^2 - 2u\sigma_2\rho\sigma_1 + \sigma_1^2}}\right),$$

where

$$\mathcal{L}(\theta) = \frac{1}{4} + \frac{1}{\sqrt{2\pi}} \arctan\left(\sqrt{\frac{\theta^2}{1-\theta^2}}\right).$$

Therefore,

$$F(u) = \frac{1}{2} + \sqrt{\frac{2}{\pi}} \arctan\left(\sqrt{\frac{(\sigma_2 u - \rho\sigma_1)^2}{(1-\rho^2)\,\sigma_1^2}}\right).$$

QED.

Proof of Theorem (8): Using (4.29) we have,

$$\begin{aligned}
\mathbf{w}^{(1)} &= -\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}-\mathbf{E})\mathbf{M}\mathbf{y} \\
\mathbf{w}^{(2)} &= -2\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left((\mathbf{I}-\mathbf{E})\left(\mathbf{I}-\mathbf{M}\right)\left(\mathbf{I}-\mathbf{E}\right)\right)\mathbf{M}.
\end{aligned}$$

Since $\mathbf{X}'\mathbf{M} = 0$. Therefore

$$\begin{aligned}
\mathbf{w}^{(1)} &= \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}\mathbf{M}\mathbf{y} \\
\mathbf{w}^{(2)} &= 2\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'(\mathbf{I}-\mathbf{E})\left(\mathbf{I}-\mathbf{M}\right)\mathbf{E}\mathbf{M}\mathbf{y} \\
&= 2\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{E}\mathbf{M}\mathbf{y} - \mathbf{2}\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-\mathbf{1}}\mathbf{X}'\mathbf{E}\left(\mathbf{I}-\mathbf{M}\right)\mathbf{E}\mathbf{M}\mathbf{y} \\
&= 2\mathbf{w}^{(1)} - 2\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-\mathbf{1}}\mathbf{X}'\mathbf{E}\left(\mathbf{I}-\mathbf{M}\right)\mathbf{E}\mathbf{M}\mathbf{y} \\
&= \mathbf{w}^{(1)} - 2\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-\mathbf{1}}\mathbf{X}'\frac{1}{T}\mathbf{1}\left(\frac{\mathbf{1}'\left(\mathbf{I}-\mathbf{M}\right)\mathbf{1}}{T}\right)\frac{1}{T}\mathbf{1}'\mathbf{M}\mathbf{y} \\
&= 2\left(1 - \frac{\mathbf{1}'\left(\mathbf{I}-\mathbf{M}\right)\mathbf{1}}{T}\right)\mathbf{w}^{(1)} = 2\mu\mathbf{w}^{(1)},
\end{aligned}$$

where $\mu = \frac{1}{T}\mathbf{1}'\mathbf{M}\mathbf{1}$. Then $\widehat{\theta}_{R\beta}$ is obtained as

$$\widehat{\theta}_{R\beta} = -- \frac{\mathbf{w}^{(1)\prime}\mathbf{w}^{(2)}}{\mathbf{w}^{(2)\prime}\mathbf{w}^{(2)}} = -\frac{1}{2\mu}.$$

The sensitivity of the variance estimate is given by:

$$\begin{aligned}
\lambda^{(1)} &= -\widehat{\mathbf{e}}'(\mathbf{I}-\mathbf{E})\widehat{\mathbf{e}}, \\
\lambda^{(2)} &= 2\widehat{\mathbf{e}}'\left[(\mathbf{I}-\mathbf{E})(\mathbf{I}-\mathbf{M})(\mathbf{I}-\mathbf{E})\right]\widehat{\mathbf{e}}.
\end{aligned}$$

The estimate of θ when for the least sensitive variance estimate can then be obtained as

$$\begin{aligned}
\widehat{\theta}_\sigma &= -\frac{\widehat{\mathbf{e}}'(\mathbf{I}-\mathbf{E})\widehat{\mathbf{e}}}{2\widehat{\mathbf{e}}'\left[(\mathbf{I}-\mathbf{E})(\mathbf{I}-\mathbf{M})(\mathbf{I}-\mathbf{E})\right]\widehat{\mathbf{e}}} \\
&= -\frac{\widehat{\mathbf{e}}'(\mathbf{I}-\mathbf{E})\widehat{\mathbf{e}}}{2\widehat{\mathbf{e}}'\left[\mathbf{E}(\mathbf{I}-\mathbf{M})\mathbf{E}\right]\widehat{\mathbf{e}}} \\
&= \frac{1}{2(1-\mu)}\frac{\widehat{\mathbf{e}}'(\mathbf{I}-\mathbf{E})\widehat{\mathbf{e}}}{\widehat{\mathbf{e}}'\mathbf{E}\widehat{\mathbf{e}}}.
\end{aligned}$$

QED.

<u>Proof of Theorem 9:</u> Following Harvey (1993, p. 29) we introduce the autocovariance generating function

$$g_\theta(z) = \sum_{h=-\infty}^{\infty} \omega_h(\theta) z^h,$$

where $\omega_h(\theta)$ is the autocovariance at lag h and z is a complex number. Note $g_\theta(L) = \mathbf{\Omega}(\theta)$ where L is the lag-operator. For the ARFIMA(p,d,q) model we have

$$g_\theta(z) = \frac{\psi(z)(1-z)^{-d}(1-z^{-1})^{-d}\psi(z^{-1})}{\phi(z)\phi(z^{-1})},$$

where $\psi(z) = 1 + \sum_{h=1}^{q} \psi_h z^h$ and $\phi(z) = 1 - \sum_{h=1}^{q} \phi_h z^h$ and $\theta = (\phi_1, .., \phi_p, d, \psi_1, ..., \psi_q)$. Note that $g_0(z) = 1$. The first two derivatives $g_\theta(z)$ with respect to ψ_h give

$$\begin{aligned}
\frac{\partial g_\theta(z)}{\partial \psi_h} &= \frac{(1-z)^{-d}(1-z^{-1})^{-d}}{\phi(z)\phi(z^{-1})}(\psi(z^{-1})z^h + \psi(z)z^{-h}) \\
\frac{\partial^2 g_\theta(z)}{\partial \psi_h^2} &= \frac{(1-z)^{-d}(1-z^{-1})^{-d}}{\phi(z)\phi(z^{-1})}(z^{-h}z^h + z^h z^{-h}) \\
&= 2\frac{(1-z)^{-d}(1-z^{-1})^{-d}}{\phi(z)\phi(z^{-1})}.
\end{aligned}$$

The first two derivatives $g_\theta(z)$ with respect to ϕ_h give

$$\begin{aligned}
\frac{\partial g_\theta(z)}{\partial \phi_h} &= g_\theta(z)\frac{(\phi(z^{-1})z^h + \phi(z)z^{-h})}{\phi(z)\phi(z^{-1})} \\
\frac{\partial^2 g_\theta(z)}{\partial \phi_h^2} &= \frac{\partial g_\theta(z)}{\partial \phi_h}\frac{(\phi(z^{-1})z^h + \phi(z)z^{-h})}{\phi(z)\phi(z^{-1})} \\
&\quad + g_\theta(z)\left(z^h\frac{z^h}{\phi(z)^2} + z^{-h}\frac{z^{-h}}{\phi(z)^2}\right).
\end{aligned}$$

The first two derivatives $g_\theta(z)$ with respect to d gives

$$\begin{aligned}
\frac{\partial g_\theta(z)}{\partial d} &= -g_\theta(z)\left[\log\left(1-z^{-1}\right) + \log\left(1-z^{-1}\right)\right] \\
\frac{\partial^2 g_\theta(z)}{\partial d^2} &= -\frac{\partial g_\theta(z)}{\partial d}\left[\log\left(1-z^{-1}\right) + \log\left(1-z^{-1}\right)\right]
\end{aligned}$$

and hence, at $\theta = 0$,

$$\begin{aligned}
\left.\frac{\partial g_\theta(z)}{\partial \psi_h}\right|_{\theta=0} &= (z^h + z^{-h}), \\
, \quad \left.\frac{\partial^2 g_\theta(z)}{\partial \psi_h^2}\right|_{\theta=0} &= 2, \\
\left.\frac{\partial g_\theta(z)}{\partial \phi_h}\right|_{\theta=0} &= (z^h + z^{-h}) \text{ and} \\
\left.\frac{\partial^2 g_\theta(z)}{\partial \phi_h^2}\right|_{\theta=0} &= (z^h + z^{-h})(z^h + z^{-h}) + (z^{2s} + z^{-2s}) \\
&= 2 + 2\left(z^{2s} + z^{-2s}\right)
\end{aligned}$$

$$\begin{aligned}
\left.\frac{\partial g_\theta(z)}{\partial d}\right|_{\theta=0} &= \log(1-z)\left(1-z^{-1}\right) \\
&= \sum_{h=1}^{\infty}\frac{1}{h}z^h \\
\left.\frac{\partial^2 g_\theta(z)}{\partial d^2}\right|_{\theta=0} &= \left[\log(1-z)\left(1-z^{-1}\right)\right]^2 \\
&= \frac{\pi^2}{3} + \sum_{h=1}^{\infty}\frac{2 + 4h\left(\Psi(h)+\gamma\right)}{h^2}\left(z^h + z^{-h}\right).
\end{aligned}$$

The last two expressions are obtained from lemma (1).

QED.

Lemma 1 *Let z be a complex number such that $|z| < 1$, then the series expansion of*

$$\log(1-z)\left(1-z^{-1}\right) = \sum_{h=1}^{\infty} \frac{1}{h} z^h$$

and

$$\left[\log(1-z)\left(1-z^{-1}\right)\right]^2 = \frac{\pi^2}{3} + \sum_{h=1}^{\infty}\left[\frac{2+4h\left(\Psi(h)+\gamma\right)}{h^2}\right] z^h,$$

where $\Psi(h) = \frac{\Gamma'(h)}{\Gamma(h)}$ is the polygamma function and γ is the Euler's constant.

<u>Proof of Lemma 1</u>: Let $z = \exp(-\boldsymbol{i}x)$, then $(1-z)(1-z^{-1}) = 2\sin^2\left(\frac{x}{2}\right)$ where $\boldsymbol{i}=\sqrt{-1}$. Consider the function $\left[(1-z)(1-z^{-1})\right]^{-d}$. Let the power series expansion be

$$\left[(1-z)\left(1-z^{-1}\right)\right]^{-d} = \sum_{h=0}^{\infty} \omega_h(d)\exp(\boldsymbol{i}xh).$$

Note that $\omega_h(d)$ can be obtained from by using the Integral transform:

$$\begin{aligned} \frac{1}{2\pi}\int_0^{2\pi}\left[(1-z)\left(1-z^{-1}\right)\right]^{-d}\exp(-\boldsymbol{i}xh)\,dx &= \omega_h(d) \\ \frac{4^{-d}}{\pi}\int_0^{\pi}\sin^{-2d}(x)\exp(-\boldsymbol{i}xh)\,dx &= \omega_h(d), \end{aligned} \tag{6.46}$$

since the Fourier coefficients of the sine part of the transform is zero, then

$$\frac{4^{-d}}{\pi}\int_0^{\pi}\sin^{-2d}(x)\cos(2hx) = \omega_h(d).$$

Then by (Erdélyi et.al (1953, p. 12) we have

$$\omega_h(d) = \frac{(-1)^h\Gamma(1-2d)}{\Gamma(-d-h+1)\Gamma(-d+h+1)}. \tag{6.47}$$

Now note that

$$\left.\frac{\partial^k\left[(1-z)\left(1-z^{-1}\right)\right]^{-d}}{\partial d^k}\right|_{d=0} = \left[\log(1-z)\left(1-z^{-1}\right)\right]^k,\ k=,1,2.$$

Differentiation under the integral sign in (6.46) on both sides gives

$$\begin{aligned} \frac{1}{2\pi}\int_0^{2\pi}\left[\log(1-z)\left(1-z^{-1}\right)\right]^k\exp(-\boldsymbol{i}xh)\,dx &= \left.\frac{\partial^k\omega_h(d)}{\partial d^k}\right|_{d=0},\ k=1,2. \\ \frac{1}{2\pi}\int_0^{2\pi}\left[\log 2\sin^2\left(\frac{x}{2}\right)\right]^k\exp(-\boldsymbol{i}xh)\,dx &= \left.\frac{\partial^k\omega_h(d)}{\partial d^k}\right|_{d=0},\ k=1,2. \end{aligned}$$

From (6.47) we get

$$\left.\frac{\partial\omega_h(d)}{\partial d}\right|_{d=0} = \begin{cases} 0, & \text{if } h=0 \\ \frac{1}{h}, & h>0 \end{cases}$$

and

$$\left.\frac{\partial^2\omega_h(d)}{\partial d^2}\right|_{d=0} = \begin{cases} \frac{\pi^2}{3}, & \text{if } h=0 \\ \frac{2+4h(\Psi(h)+\gamma)}{h^2} & h>0 \end{cases}.$$

QED.

Proof of Theorem 10:

Note that $\mathbf{MT}^{(i)}(\mathbf{I}-\mathbf{M})\mathbf{T}^{(i)}\mathbf{M}$ is a symmetric matrix of rank l. Let the eigen values of the matrix be $eig_1, eig_2, .., eig_l$ then $\widehat{\mathbf{e}}'\mathbf{T}^{(i)}(\mathbf{I}-\mathbf{M})\mathbf{T}^{(i)}\widehat{\mathbf{e}} = eig_1v_1^2+eig_2v_2^2+..+eig_lv_l^2$ where $v_i^2, i = 1, ...l$ are bounded random variables. Then $\text{plim}_{T\to\infty}\frac{\widehat{\mathbf{e}}'\mathbf{T}^{(i)}(\mathbf{I}-\mathbf{M})\mathbf{T}^{(i)}\widehat{\mathbf{e}}}{T} = 0$ and $\text{plim}_{T\to\infty}\frac{\widehat{\mathbf{e}}'\widehat{\mathbf{e}}}{T} < \infty$. Therefore $\text{plim}_{T\to\infty}\overline{\rho_{(i)}} = \frac{\widehat{\mathbf{e}}'\mathbf{T}^{(i)}(\mathbf{I}-\mathbf{M})\mathbf{T}^{(i)}\widehat{\mathbf{e}}}{\widehat{\mathbf{e}}'\widehat{\mathbf{e}}} = 0$. Also $\sqrt{T}\rho_{(i)} \overset{asy}{\sim} N(0,1)$.

QED.

Proof of Theorem 12**:** Implicitly differentiating $R_{\mu\mu}(\mu,\theta)\widehat{\mu}^{(1)}(\theta) + R_{\mu\theta}(\widehat{\mu}(\theta),\theta) = 0$ again with respect to θ we get

$$\begin{gathered} R_{\mu\mu\mu}(\widehat{\mu}(\theta),\theta)\left[\widehat{\mu}^{(1)}(\theta)\right]^2 + R_{\mu\mu}(\widehat{\mu}(\theta),\theta)\widehat{\mu}^{(2)}(\theta) \\ +R_{\mu\theta\mu}(\widehat{\mu}(\theta),\theta)\widehat{\mu}^{(1)}(\theta) + R_{\mu\theta\theta}(\widehat{\mu}(\theta),\theta) = 0 \end{gathered}$$

This implies

$$\widehat{\mu}^{(2)}(\theta) = -\frac{R_{\mu\mu\mu}(\widehat{\mu}(\theta),\theta)\left[\widehat{\mu}^{(1)}(\theta)\right]^2 + R_{\mu\theta\mu}(\widehat{\mu}(\theta),\theta)\widehat{\mu}^{(1)}(\theta) + R_{\mu\theta\theta}(\widehat{\mu}(\theta),\theta)}{R_{\mu\mu}(\widehat{\mu}(\theta),\theta)}$$

The the least sensitive estimate of θ is:

$$\begin{aligned} \widehat{\theta}_\mu &= -\frac{\widehat{\mu}^{(1)}(\theta)}{\widehat{\mu}^{(2)}(\theta)} \\ &= -\frac{R_{\mu\theta}(\widehat{\mu}(\theta),\theta)\left[R_{\mu\mu}(\widehat{\mu}(\theta),\theta)\right]^2}{R_{\mu\mu\mu}(\widehat{\mu}(\theta),\theta)\left[\widehat{\mu}^{(1)}(\theta)\right]^2 + R_{\mu\theta\mu}(\widehat{\mu}(\theta),\theta)\widehat{\mu}^{(1)}(\theta) + R_{\mu\theta\theta}(\widehat{\mu}(\theta),\theta)} \\ &= -\frac{R_{\mu\theta}(\widehat{\mu}(\theta),\theta)\left[R_{\mu\mu}(\widehat{\mu}(\theta),\theta)\right]^2}{R_{\mu\mu\mu}(\widehat{\mu}(\theta),\theta)\left[R_{\mu\theta}(\widehat{\mu}(\theta),\theta)\right]^2 - R_{\mu\theta\mu}(\widehat{\mu}(\theta),\theta)R_{\mu\theta}(\widehat{\mu}(\theta),\theta) + R_{\mu\theta\theta}(\widehat{\mu}(\theta),\theta)R_{\mu\mu}(\widehat{\mu}(\theta),\theta)} \end{aligned}$$

QED.

Proof of Theorem13: Using (5.44) we have:

$$\lim_{\theta\to 0}\widehat{\mu}_{\theta}(\theta)=-\lim_{\theta\to 0}\frac{R_{\mu\theta}\left(\widehat{\mu}(\theta),\theta\right)}{R_{\mu\mu}\left(\widehat{\mu}(\theta),\theta\right)}=\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(y_t-\overline{y}\right)^2$$

$$\lim_{\theta\to 0}R_{\mu\mu}\left(\widehat{\mu}(\theta),\theta\right)=\frac{1}{T}\sum_{t=1}^{T}\lim_{\theta\to 0}\frac{\partial^2\mathfrak{L}\left(y_t-\mu,\theta\right)}{\partial\mu\partial\mu}=1$$

$$\begin{aligned}
\lim_{\theta\to 0}R_{\mu\theta}\left(\widehat{\mu}(\theta),\theta\right) &= \frac{1}{T}\sum_{t=1}^{T}\lim_{\theta\to 0}\left.\frac{\partial^2\mathfrak{L}\left(y_t-\mu,\theta\right)}{\partial\theta\partial\mu}\right|_{\mu=\widehat{\mu}(\theta)}\\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{2}\left(\overline{y}-y_t\right)\left(\overline{y}-y_t\right)-\overline{y}\left(\overline{y}-y_t\right)+y_t\left(\overline{y}-y_t\right)\right)\\
&= -\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(y_t-\overline{y}\right)^2
\end{aligned}$$

$$\begin{aligned}
\lim_{\theta\to 0}R_{\mu\theta\mu}\left(\widehat{\mu}(\theta),\theta\right) &= \frac{1}{T}\sum_{t=1}^{T}\lim_{\theta\to 0}\left.\frac{\partial^2\mathfrak{L}\left(y_t-\mu,\theta\right)}{\partial\theta\partial\mu\partial\mu}\right|_{\mu=\widehat{\mu}(\theta)}\\
&= \frac{1}{T}\sum_{t=1}^{T}\left(y_t-\overline{y}\right)=0
\end{aligned}$$

$$\lim_{\theta\to 0}R_{\mu\mu\mu}\left(\widehat{\mu}(\theta),\theta\right)=\frac{1}{T}\sum_{t=1}^{T}\lim_{\theta\to 0}\left.\frac{\partial^2\mathfrak{L}\left(y_t-\mu,\theta\right)}{\partial\mu\partial\mu\partial\mu}\right|_{\mu=\widehat{\mu}(\theta)}=0$$

$$\begin{aligned}
\lim_{\theta\to 0}R_{\mu\theta\theta}\left(\widehat{\mu}(\theta),\theta\right) &= \frac{1}{T}\sum_{t=1}^{T}\lim_{\theta\to 0}\left.\frac{\partial^2\mathfrak{L}\left(y_t-\mu,\theta\right)}{\partial\mu\partial\theta\partial\theta}\right|_{\mu=\widehat{\mu}(\theta)}\\
&= \frac{1}{T}\sum\overline{y}^2\left(\overline{y}-y_t\right)+y_t^2\left(\overline{y}-y_t\right)+\frac{1}{3}\left(\overline{y}-y_t\right)\left(\overline{y}-y_t\right)^2\\
&\quad -2\overline{y}\left(\overline{y}-y_t\right)\left(\overline{y}-y_t\right)+y_t\left(\overline{y}-y_t\right)\left(\overline{y}-y_t\right)-2\overline{y}y_t\left(\overline{y}-y_t\right)\\
&= \frac{1}{3}\frac{1}{T}\sum_{t=1}^{T}\left(\overline{y}-y_t\right)^3
\end{aligned}$$

$$\begin{aligned}
\widehat{\mu}_{\theta\theta}(\theta) &= -\frac{R_{\mu\mu\mu}\left(\widehat{\mu}(\theta),\theta\right)\left[\widehat{\mu}^{(1)}(\theta)\right]^2+R_{\mu\theta\mu}\left(\widehat{\mu}(\theta),\theta\right)\widehat{\mu}^{(1)}(\theta)+R_{\mu\theta\theta}\left(\widehat{\mu}(\theta),\theta\right)}{R_{\mu\mu}\left(\widehat{\mu}(\theta),\theta\right)}\\
&= -\frac{1}{3}\frac{1}{T}\sum_{t=1}^{T}\left(y_t-\overline{y}\right)^3
\end{aligned}$$

$$\lim_{\theta \to 0} \widehat{\theta}_\mu = -\lim_{\theta \to 0} \frac{\widehat{\mu}^{(1)}(\theta)}{\widehat{\mu}^{(2)}(\theta)} = \frac{3\frac{1}{T}\sum_{t=1}^{T}(y_t - \overline{y})^2}{2\sum_{t=1}^{T}(y_t - \overline{y})^3}$$

so the corrected mean is

$$\widehat{\mu}\left(\widehat{\theta}_\mu\right) = \overline{y} + \frac{3}{4}\frac{\left(\frac{1}{T}\sum_{t=1}^{T}(y_t - \overline{y})^2\right)^2}{\frac{1}{T}\sum_{t=1}^{T}(y_t - \overline{y})^3}$$

QED.

References

Arcones, M.A. (1995), *On the central limit theorem for U-statistics under absolute regularity*, Statistics and Probability Letters, Vol 24(3), pp 245-249.

Banerjee, A.N. and J.R. Magnus (1999), *The sensitivity of OLS when the variance matrix is (partially) know*, Journal of Econometrics, Vol 92(2), pp 295-323.

Banerjee, A.N. and J.R. Magnus (2000), *On the sensitivity of the usual t- and F-tests to covariance misspecification*, Journal of Econometrics, Vol 95(10), pp 157-176.

Banerjee, A.N. (2001), *Sensitivity of Univariate AR(1) Time-series Forecasts near the Unit Root*, Journal of Forecasting, Vol 20(3), pp 203-229.

Durbin, J. and S. Watson (1971), *Testing for serial correlation in least squares regression, III*, Biometrika Vol 58, pp 1-19.

Erdélyi, A., W. Magnus, F. Oberhettinger, F.G. Tricomi, and H. Bateman (1953), Higher transcendental functions (Vol. 1), New York.

Hinkley, D.V. (1969), *On the Ratio of Two Correlated Normal Random Variables*, Biometrika, Vol. 56(3), pp 635-639 .

Hsing, T and W.B. Wu (2004), *On weighted U-statistics for stationary processes,* Annals of Probability. Vol 32(2), 1600-1631.

Lee, A.J. (1990), U-statistics, Theory and Practice, Marcel Dekker, New York.

Magnus, J.R and A.L. Vasnev (2007), *Local sensitivity and diagnostic tests*, Econometrics Journal, Vol 10(1), pp. 166-192.

Newey, W. K. and K.D. West (1987), *A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix*, Econometrica 55, 703-708.

Omtzigt P. and P Paruolo (2005), *Impact factors*, Journal of Econometrics, Vol 128(1), pp 31-68.

Spall, J.C (2000), *Adaptive Stochastic Approximation by the Simultaneous Perturbation Method*, IEEE Transactions on Automatic Control, Vol 45(10), pp 1839-1853.

Varian, H. R (1975), *A Bayesian approach to real estate assessment*, Studies Bayesian Econometrics and Statistics in Honor of Leonard J. Savage.Fienberg, S. E. and Zellner, A., Eds.; North Holland: Amsterdam, pp 195-208.

Wagner, D.H (1969), *Nonlinear Functional Versions of the Neyman-Pearson Lemma,* Siam Review, Vol. 11(1), pp. 52-65.

Wan A.T.K., G. Zou and H Qin (2007), *On the sensitivity of the restricted least squares estimators to covariance misspecification*, Econometrics Journal, Vol 10(3), pp 471-487.

White, H. (1981), *Consequences and Detection of Misspecified Nonlinear Regression Models*, Journal of the American Statistical Association, Vol. 76(374), pp. 419-433.

White, H. (1982), *Maximum Likelihood Estimation of Misspecified Models*, Econometrica, Vol 50(1), pp 1-25.

White, H. (1980c), *Nonlinear Regression on Cross-Section Data*, Econometrica, Vol 48(3), pp 721-746.

Whittle, P (1964), *On the Convergence to Normality of Quadratic Forms in Independent Variables*, Theory of Probability and its Applications, pp 103-108.

Zellner, A (1986), *Bayesian Estimation and Prediction Using Asymmetric Loss Functions*, Journal of the American Statistical Association, Vol 81(394), pp. 446-451.