Tilburg University

Local Sensitivity and Diagnostic Tests

Magnus, J.R.; Vasnev, A.L.

Publication date:
2004

No. 2004–105

LOCAL SENSITIVITY AND DIAGNOSTIC TESTS

By J.R. Magnus, A.L. Vasnev

October 2004

TILBURG ◆ UNIVERSITY

Local sensitivity and diagnostic tests[*]

Jan R. Magnus

CentER and Department of Econometrics and Operations Research,

Tilburg University, The Netherlands

and

Andrey L. Vasnev

CentER, Tilburg University, The Netherlands

October 28, 2004

<u>Title:</u> Local sensitivity and diagnostic tests

<u>Proposed running head:</u> Local sensitivity and diagnostic tests

<u>Corresponding author:</u>
Jan R. Magnus
CentER, Tilburg University
P.O. Box 90153
5000 LE Tilburg
The Netherlands

phone: +31-13-466-3092
fax: +31-13-466-3280
email: magnus@uvt.nl

<u>Keywords:</u> Maximum likelihood, Sensitivity, Diagnostic test, Regression

<u>JEL Codes:</u> C12, C22, C51, C52.

<u>Abstract</u>: In this paper we confront sensitivity analysis with diagnostic testing. Every model is misspecified, but a model is useful if the parameters of interest (the focus) are not sensitive to small perturbations in the underlying assumptions. The study of the effect of these violations on the focus is called sensitivity analysis. Diagnostic testing, on the other hand, attempts to find out whether a nuisance parameter is 'large' or 'small'. Both aspects are important, but traditional applied econometrics tends to use only diagnostics and forget about sensitivity analysis. We develop a theory of sensitivity in a maximum likelihood framework, propose a sensitivity test, give conditions under which the diagnostic and sensitivity tests are asymptotically independent, and demonstrate with three core examples that this independence is the rule rather than the exception, thus underlying the importance of sensitivity analysis.

1   Motivation and background

Suppose we wish to estimate $\boldsymbol{\beta}$ from the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \theta\boldsymbol{z} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon}|(\boldsymbol{X}, \boldsymbol{z}) \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_n),$$

where the regressor $\boldsymbol{z}$ may or may not be included in the model. In the restricted model (where $\theta = 0$), we estimate $\boldsymbol{\beta}$ by $\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, while the least-squares estimators for $\boldsymbol{\beta}$ and θ in the unrestricted model are

$$\widehat{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{z}(\boldsymbol{z}'\boldsymbol{M}\boldsymbol{z})^{-1}\boldsymbol{z}'\boldsymbol{M}\boldsymbol{y}, \quad \widehat{\theta} = \frac{\boldsymbol{z}'\boldsymbol{M}\boldsymbol{y}}{\boldsymbol{z}'\boldsymbol{M}\boldsymbol{z}},$$

where $\boldsymbol{M} := \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. The difference between $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ is thus given by

$$\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}} = -(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{z}\widehat{\theta}. \tag{1}$$

Now consider the function $\widetilde{\boldsymbol{\beta}}(\theta) := (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{y} - \theta\boldsymbol{z})$, which estimates $\boldsymbol{\beta}$ for each fixed value of θ. In particular we have $\widehat{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}(\widehat{\theta})$ and $\widetilde{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}(0)$. A Taylor expansion gives

$$\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}(\widehat{\theta}) - \widetilde{\boldsymbol{\beta}}(0) = \left.\frac{\partial\widetilde{\boldsymbol{\beta}}(\theta)}{\partial\theta}\right|_{\theta=0} \widehat{\theta} + O_p(1/n), \tag{2}$$

where — in this simple case — the remainder term is identically zero. We see from (1) and (2) that the difference between $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ factorizes as $\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}} = \boldsymbol{S}\widehat{\theta}$, where $\boldsymbol{S}$ denotes the *sensitivity*

$$\boldsymbol{S} := \left.\frac{\partial\widetilde{\boldsymbol{\beta}}(\theta)}{\partial\theta}\right|_{\theta=0} = -(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{z}.$$

We may think of $\widehat{\theta}$ as the 'magnitude' and of $\boldsymbol{S}$ as the 'direction' of the impact of the misspecification on $\widetilde{\boldsymbol{\beta}}$. Seen in this light it is reasonable to investigate the conditions under which $\widehat{\theta}$ and $\boldsymbol{S}$ are independent. If our focus is not to estimate $\boldsymbol{\beta}$ but, say, to forecast y, then the magnitude of the impact does not change, but the direction does change. This simply reflects the fact that a model may be a good approximation for one focus, but not for another.

In applied econometrics the choice between $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ is almost always based exclusively on the t-statistic $t_\theta := (\boldsymbol{z}'\boldsymbol{M}\boldsymbol{z})^{-1/2}\boldsymbol{z}'\boldsymbol{M}\boldsymbol{y}$ or on a simple transformation thereof, such as the Wald statistic $\mathrm{W} = t_\theta^2$. In other words, the choice is based on a *diagnostic*, answering the question whether $\widehat{\theta}$ is 'large' or 'small'. Since θ is a nuisance parameter, we are not primarily interested

in whether $\widehat{\theta}$ is large or small; our interest is in $\boldsymbol{\beta}$. It may very well be that $\widehat{\theta}$ is large, but that nevertheless the difference between $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ is small, a frequent observation in econometric practice, which occurs if the sensitivity is small. A proper choice between the estimators should therefore be based on both factors: the diagnostic *and* the sensitivity. If the diagnostic and the sensitivity would be highly correlated, then ignoring the sensitivity might not matter. However, as we shall demonstrate, the more common situation is that the sensitivity and the diagnostic are (asymptotically) independent. Then sensitivity analysis matters.

The current paper is motivated by the above highly stylized and simplified example. Sensitivity analysis matters in this example, and it equally matters in more complex examples and in other contexts. The aim of this paper is threefold. First, we develop the theory of sensitivity analysis in a general maximum likelihood context. Second, we give conditions under which the sensitivity is asymptotically independent of the diagnostic test. Third, we demonstrate that these conditions are satisfied in three important directions: mean misspecification, variance misspecification, and distribution misspecification.

(Local) sensitivity analysis thus studies the effect of (small) changes in the underlying assumptions on the output of the system. It plays an important role in (non)linear programming (Gal and Greenberg, 1997), chemistry and physics (Saltelli, Chan, and Scott, 2000), and other disciplines (Kleijnen, 1997). In econometrics, the 'output' is the statistic of interest, such as an estimator, a forecast, or a policy recommendation.[1]

There are two branches of sensitivity analysis: data perturbation and model perturbation. In data perturbation one may perturb the location of the regressors, or the location or the scale of the dependent variable in a regression context. This branch is associated mainly with the work of Huber (2004, first edition 1980) and Cook (1979, 1986).[2] In contrast, model perturbation considers the effects on the parameter of interest (or any other focus) of small deviations from the hypothesized model, such as the deletion of relevant regressors, the misspecification of the variance matrix, or deviations

---

[1]See Chatterjee and Hadi (1988) for an early summary of sensitivity analysis in linear regression, and Saltelli, Chan, and Scott (2000) for applications in many different directions.

[2]See also Cook and Weisberg (1982). Cook's 'likelihood displacement' method has been applied to elliptical disturbances (Galea, Paula, and Bolfarine (1997) and Liu (2000)), multivariate regression (Fung and Tang, 1997), growth curve models (Pan, Fang, and von Rosen, 1997), ridge regression (Shi and Wang, 1999), dropout models (Verbeke, Molenberghs, Thijs, Lesaffre, and Kenward, 2001), multivariate elliptical linear regression (Liu, 2002), and prediction (Hartless, Booth, and Littell, 2003). See Liu (2002) for further references. In Section 3 we shall relate Cook's method to our definition of sensitivity.

from normality. This branch plays a role in Bayesian statistics, in particular the effect of misspecifying the prior distribution (Leamer (1978, 1984), Polasek (1984)), but also in classical econometrics (Banerjee and Magnus (1999, 2000)).

In the current paper, our interest lies in the perturbation of models, and 'sensitivity analysis' will be understood to mean the study of the effect of small changes in model assumptions on an estimator of a parameter of interest. The paper is organized as follows. The formal maximum likelihood framework and the notation is explained in Section 2, where we state the assumptions used and obtain two asymptotic results concerning maximum likelihood estimation and diagnostic tests. In Section 3 we introduce the sensitivity statistic, derive the sensitivity test, and obtain its asymptotic distribution. Some often-occurring special cases are considered as well. The conditions under which the sensitivity test and the diagnostic test are asymptotically independent are studied in Section 4. This completes the theoretical part of the paper. In Sections 5–7 we investigate three important directions of model misspecification: the mean, the variance, and the error distribution. The finite sample performance is illustrated with Monte Carlo simulations. Section 8 concludes.

2   Maximum likelihood and diagnostic tests

The observations consist of the first n terms of a sequence of random vectors $(y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), \ldots$, not necessarily independent or identically distributed, where we think of y as the dependent variable and of $\boldsymbol{x}$ as the vector of explanatory variables.[3] The joint density, denoted by $f(\cdot; \boldsymbol{\delta})$, is assumed to be known, except for the values of a finite and fixed number of parameters $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)' \in \mathcal{D} \subset \mathbb{R}^p$. The log-likelihood function is

$$\ell(\boldsymbol{\delta}) := \ell(\boldsymbol{\delta}; (y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)) := \log f((y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n); \boldsymbol{\delta}).$$

We denote the true (but unknown) value of $\boldsymbol{\delta}$ by $\boldsymbol{\delta}_0$. All probabilities and expectations are taken with respect to the true underlying distribution. We impose a set of relatively weak conditions on the data to ensure the existence of certain expansions and the proper behavior of maximum likelihood (ML) estimators. All limits are taken for $n \to \infty$.

Assumption 1:
(a) the parameter space $\mathcal{D}$ is a compact subset of $\mathbb{R}^p$,

---

[3] We follow the notation proposed in Abadir and Magnus (2002).

(b) $\boldsymbol{\delta}_0$ lies in the interior $\mathcal{D}^0$ of $\mathcal{D}$,
(c) $\ell(\boldsymbol{\delta})$ is continuous on $\mathcal{D}$,
(d) $\ell(\boldsymbol{\delta})$ is two times continuously differentiable on $\mathcal{D}^0$.

Assumption 2: Let $k := k(\boldsymbol{\delta}, \boldsymbol{\delta}_0) := -\operatorname{E}\left(\ell(\boldsymbol{\delta}) - \ell(\boldsymbol{\delta}_0)\right)$ denote the absolute value of the Kullback-Leibler information. Then,
(a) $k(\boldsymbol{\delta}, \boldsymbol{\delta}_0) \to \infty$ for every $\boldsymbol{\delta} \neq \boldsymbol{\delta}_0$,
(b) $(1/k^2)\operatorname{var}(\ell(\boldsymbol{\delta}) - \ell(\boldsymbol{\delta}_0)) \to 0$,
(c) for every $\boldsymbol{\delta} \neq \boldsymbol{\delta}_0 \in \mathcal{D}$ there exists a neighborhood $N(\boldsymbol{\delta})$ of $\boldsymbol{\delta}$ such that

$$\Pr\left((1/k) \sup_{\boldsymbol{\phi} \in N(\boldsymbol{\delta})} (\ell(\boldsymbol{\phi}) - \ell(\boldsymbol{\delta})) < 1\right) \to 1.$$

Assumption 2(c) ensures that the normalized log-likelihood ratio is locally equicontinuous in probability. This condition is weaker than the more common condition that $(1/n)\ell(\boldsymbol{\delta})$ converges uniformly in probability. Assumptions 1(a), 1(c), and 2 together guarantee that the ML estimator $\widehat{\boldsymbol{\delta}}$ of $\boldsymbol{\delta}$ exists and is consistent, see Heijmans and Magnus (1986a).

We now define the score and the Hessian matrix as

$$\boldsymbol{q}(\boldsymbol{\delta}) := \frac{1}{\sqrt{n}} \frac{\partial \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}, \quad \mathcal{H}(\boldsymbol{\delta}) := \frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \, \partial \boldsymbol{\delta}'},$$

where we note that these are normalized in order to ensure stable variates.

Assumption 3:
(a) $\boldsymbol{q}(\boldsymbol{\delta}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\mathcal{I}}(\boldsymbol{\delta}_0))$,
(b) $\mathcal{H}(\boldsymbol{\delta}_0) \xrightarrow{p} -\boldsymbol{\mathcal{I}}(\boldsymbol{\delta}_0)$,
(c) $\boldsymbol{\mathcal{I}}(\boldsymbol{\delta})$ is continuous on $\mathcal{D}^0$ and $\boldsymbol{\mathcal{I}}(\boldsymbol{\delta}_0)$ is positive definite,
(d) for every $\epsilon > 0$ there exists a neighborhood $N(\boldsymbol{\delta}_0)$ of $\boldsymbol{\delta}_0$ such that

$$\Pr\left(\sup_{\boldsymbol{\delta} \in N(\boldsymbol{\delta}_0)} |\mathcal{H}_{ij}(\boldsymbol{\delta}) + \boldsymbol{\mathcal{I}}_{ij}(\boldsymbol{\delta})| > \epsilon\right) \to 0 \quad (i, j = 1, \ldots, p).$$

We notice that condition 3(d) is weaker than the more common assumptions requiring uniform convergence in probability of $\mathcal{H}(\boldsymbol{\delta})$ or uniform boundedness of third-order derivatives. Under Assumptions 1–3 the ML estimator $\widehat{\boldsymbol{\delta}}$ is first-order efficient and asymptotically normal in the sense that

$$\sqrt{n}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{\mathcal{I}}(\boldsymbol{\delta}_0)^{-1}\right),$$

see Heijmans and Magnus (1986b). Other sets of assumptions, such as those of Andrews (1998, p. 170), are of course possible.

We wish to think of the parameter vector $\boldsymbol{\delta}$ as consisting of two parts, namely a focus parameter $\boldsymbol{\beta} \in \mathbb{R}^k$ and a nuisance parameter $\boldsymbol{\theta} \in \mathbb{R}^m$, where $k + m = p$. Our interest lies in the estimation of the focus parameter $\boldsymbol{\beta}$. In the *unrestricted* model both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are estimated, and the ML estimators are denoted $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$, respectively. In the *restricted* model we impose the restriction $\boldsymbol{\theta} = \mathbf{0}$, so that only $\boldsymbol{\beta}$ is estimated; the restricted ML estimator is denoted $\widetilde{\boldsymbol{\beta}}$.

The score, Hessian, and information matrix can be evaluated at $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\theta})$ (generic), at $\widehat{\boldsymbol{\delta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}})$ (unrestricted ML estimator), at $\widetilde{\boldsymbol{\delta}} = (\widetilde{\boldsymbol{\beta}}, \mathbf{0})$ (restricted ML estimator), or at $\boldsymbol{\delta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ (true value). We follow standard notation by writing $\boldsymbol{\mathcal{I}}, \widehat{\boldsymbol{\mathcal{I}}}, \widetilde{\boldsymbol{\mathcal{I}}}$, and $\boldsymbol{\mathcal{I}}^0$ to indicate at which point the information matrix is evaluated; similar notation is adopted for the score and Hessian matrix. We partition

$$\boldsymbol{q} = \begin{pmatrix} \boldsymbol{q}_{\boldsymbol{\beta}} \\ \boldsymbol{q}_{\boldsymbol{\theta}} \end{pmatrix}, \quad \boldsymbol{\mathcal{H}} = \begin{pmatrix} \boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\boldsymbol{\theta}} \\ \boldsymbol{\mathcal{H}}_{\boldsymbol{\theta}\boldsymbol{\beta}} & \boldsymbol{\mathcal{H}}_{\boldsymbol{\theta}\boldsymbol{\theta}} \end{pmatrix},$$

and similarly for the information matrix $\boldsymbol{\mathcal{I}}$. Then, the unrestricted ML estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ satisfy the first-order conditions

$$\boldsymbol{q}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) = \mathbf{0}, \quad \boldsymbol{q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) = \mathbf{0}, \tag{3}$$

while the restricted ML estimator $\widetilde{\boldsymbol{\beta}}$ satisfies

$$\boldsymbol{q}_{\boldsymbol{\beta}}(\widetilde{\boldsymbol{\beta}}, \mathbf{0}) = \mathbf{0}. \tag{4}$$

To test the null hypothesis $\mathrm{H}_0 : \boldsymbol{\theta} = \mathbf{0}$, several statistics are available. The three classical tests are the Wald (W), the likelihood ratio (LR), and the Lagrange multiplier (LM) test. The latter, also known as the score test, is the most natural diagnostic in our context, and takes the form

$$\mathrm{LM} = \widetilde{\boldsymbol{q}}_{\boldsymbol{\theta}}' \left(\widetilde{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \widetilde{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\theta}\boldsymbol{\beta}} \widetilde{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} \widetilde{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\beta}\boldsymbol{\theta}} \right)^{-1} \widetilde{\boldsymbol{q}}_{\boldsymbol{\theta}}. \tag{5}$$

The W and LR tests are asymptotically equivalent to the LM test, and hence all asymptotic results hold for these two tests as well.

In order to establish asymptotic behavior, we need the first-order expansions:

$$\mathbf{0} = \boldsymbol{q}(\widehat{\boldsymbol{\delta}}) = \boldsymbol{q}(\boldsymbol{\delta}_0) + \boldsymbol{\mathcal{H}}^0 \sqrt{n}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) + O_p(1/\sqrt{n}), \tag{6}$$

$$\mathbf{0} = \boldsymbol{q}_{\boldsymbol{\beta}}(\widetilde{\boldsymbol{\beta}}, \mathbf{0}) = \boldsymbol{q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \mathbf{0}) + \boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0 \sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + O_p(1/\sqrt{n}), \tag{7}$$

and

$$\widetilde{\boldsymbol{q}}_{\boldsymbol{\theta}} = \boldsymbol{q}_{\boldsymbol{\theta}}(\widetilde{\boldsymbol{\beta}}, \mathbf{0}) = \boldsymbol{q}_{\boldsymbol{\theta}}(\boldsymbol{\beta}_0, \mathbf{0}) + \boldsymbol{\mathcal{H}}_{\boldsymbol{\theta}\boldsymbol{\beta}}^0 \sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + O_p(1/\sqrt{n}). \tag{8}$$

The following theorem can then be established and will serve as our point of departure.

Theorem 1 (Asymptotic distribution of ML estimators and LM test): Given Assumptions 1–3,
(a) the unrestricted ML estimator $\widehat{\boldsymbol{\beta}}$ is consistent and asymptotically normal such that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}\right), \quad \boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} = \left(\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}}^0 - \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\theta}}^0 (\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta\theta}}^0)^{-1} \boldsymbol{\mathcal{I}}_{\boldsymbol{\theta\beta}}^0\right)^{-1},$$

(b) the restricted ML estimator $\widetilde{\boldsymbol{\beta}}$ is consistent and asymptotically normal with asymptotic variance $(\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}}^0)^{-1}$,
(c) under the null hypothesis $\mathrm{H}_0 : \boldsymbol{\theta} = \mathbf{0}$, the LM statistic (5) follows asymptotically a $\chi^2(m)$-distribution.

Proof: We use the expansions (6), (7), and (8), which give

$$\sqrt{n}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) = (\boldsymbol{\mathcal{I}}^0)^{-1} \boldsymbol{q}(\boldsymbol{\delta}_0) + O_p(1/\sqrt{n}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, (\boldsymbol{\mathcal{I}}^0)^{-1}),$$

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = (\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}}^0)^{-1} \boldsymbol{q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \mathbf{0}) + O_p(1/\sqrt{n}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, (\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}}^0)^{-1}),$$

and

$$\widetilde{\boldsymbol{q}}_{\boldsymbol{\theta}} = \boldsymbol{q}_{\boldsymbol{\theta}}^0 - \boldsymbol{\mathcal{I}}_{\boldsymbol{\theta\beta}}^0 (\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}}^0)^{-1} \boldsymbol{q}_{\boldsymbol{\beta}}^0 + O_p(1/\sqrt{n})$$
$$\xrightarrow{d} \mathrm{N}(0, \boldsymbol{\mathcal{I}}_{\boldsymbol{\theta\theta}}^0 - \boldsymbol{\mathcal{I}}_{\boldsymbol{\theta\beta}}^0 (\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}}^0)^{-1} \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\theta}}^0),$$

and the results follow. $\parallel$

The expansions (6)–(8) also imply the following basic orthogonality result.

Theorem 2 (Asymptotic independence of $\widetilde{\boldsymbol{\beta}}$ and LM test): Given Assumptions 1–3 and under the null hypothesis $\mathrm{H}_0 : \boldsymbol{\theta} = \mathbf{0}$,

$$\sqrt{n} \operatorname{cov}\left(\widetilde{\boldsymbol{q}}_{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \to \mathbf{O},$$

and hence the restricted estimator $\widetilde{\boldsymbol{\beta}}$ and the LM test are asymptotically independent.

Proof: For two random vectors $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, let $\boldsymbol{z}_1 \approx \boldsymbol{z}_2$ denote 'asymptotic equality' in the sense that $\boldsymbol{z}_1 = \boldsymbol{z}_2 + O_p(1/\sqrt{n})$. Since $\widetilde{\boldsymbol{q}}_{\boldsymbol{\theta}} \approx \boldsymbol{q}_{\boldsymbol{\theta}}^0 -$

$\mathcal{I}_{\boldsymbol{\theta\beta}}^0(\mathcal{I}_{\boldsymbol{\beta\beta}}^0)^{-1}\boldsymbol{q}_{\boldsymbol{\beta}}^0$, we obtain

$$\operatorname{cov}\left(\widetilde{\boldsymbol{q}}_{\boldsymbol{\theta}}, \boldsymbol{q}_{\boldsymbol{\beta}}^0\right) \approx \operatorname{E}\left(\boldsymbol{q}_{\boldsymbol{\theta}}^0\boldsymbol{q}_{\boldsymbol{\beta}}^{0\,\prime}\right) - \mathcal{I}_{\boldsymbol{\theta\beta}}^0(\mathcal{I}_{\boldsymbol{\beta\beta}}^0)^{-1}\operatorname{E}\left(\boldsymbol{q}_{\boldsymbol{\beta}}^0\boldsymbol{q}_{\boldsymbol{\beta}}^{0\,\prime}\right)$$
$$\approx \mathcal{I}_{\boldsymbol{\theta\beta}}^0 - \mathcal{I}_{\boldsymbol{\theta\beta}}^0(\mathcal{I}_{\boldsymbol{\beta\beta}}^0)^{-1}\mathcal{I}_{\boldsymbol{\beta\beta}}^0 = \boldsymbol{O},$$

using the fact that $\mathcal{I}^0 \approx \operatorname{E}(\boldsymbol{q}^0\boldsymbol{q}^{0\,\prime})$. Hence, the LM test is asymptotically uncorrelated with $\boldsymbol{q}_{\boldsymbol{\beta}}^0$, and, because of the asymptotic normality, asymptotically independent. In addition, $\boldsymbol{q}_{\boldsymbol{\beta}}^0$ is asymptotically equal to $\mathcal{I}_{\boldsymbol{\beta\beta}}^0\sqrt{n}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$. The result follows. $\|$

Theorem 2 provides a generalization of the following well-known fact from least-squares theory. Let $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim \operatorname{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$. The estimator of $\boldsymbol{\theta}$ in the unrestricted model is $\widehat{\boldsymbol{\theta}} = (\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{y}$, where $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Under the restriction that $\boldsymbol{\theta} = \boldsymbol{0}$, the estimator of $\boldsymbol{\beta}$ in the restricted model is $\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$. Clearly, $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\beta}}$ are independent, because $\boldsymbol{M}\boldsymbol{X} = \boldsymbol{O}$. Moreover, $\widetilde{\boldsymbol{q}}_{\boldsymbol{\theta}} := \boldsymbol{Z}'\boldsymbol{M}\boldsymbol{y}/\widetilde{\sigma}^2$ is also independent of $\widetilde{\boldsymbol{\beta}}$, because both $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{y}$ and $\widetilde{\sigma}^2 := \boldsymbol{y}'\boldsymbol{M}\boldsymbol{y}/n$ are independent of $\widetilde{\boldsymbol{\beta}}$.

3 Sensitivity

While a diagnostic test answers the question 'Is it true?' (that the nuisance parameter is not zero), a sensitivity statistic answers the question 'Does it matter?' A diagnostic test such as the LM test may reject the null hypothesis that $\boldsymbol{\theta} = \boldsymbol{0}$, but this does not mean that the estimator of the focus parameter $\boldsymbol{\beta}$ will be sensitive to deviations of $\boldsymbol{\theta}$ from $\boldsymbol{0}$. In fact, Banerjee and Magnus (1999) found in the special case of AR(1) errors that the diagnostic test tells you very little about the sensitivity.[4] The essential difference between a diagnostic test and a sensitivity statistic is graphed in Figure 1, where we assume for simplicity that $k = m = 1$; hence there is one focus parameter β and one nuisance parameter θ. Figure 1 is a generalization of the well-known picture of the three classical tests, see for example Ruud (2000, p. 390). At $(\widehat{\beta}, \widehat{\theta})$ we obtain the maximum of the likelihood $\widehat{\ell}$, while at $(\widetilde{\beta}, 0)$, we obtain the restricted maximum $\widetilde{\ell}$. For every fixed value of θ, let $\widetilde{\beta}(\theta)$ denote the value of β which maximizes the (restricted) likelihood. The locus of all constrained maxima is the curve $C := \left(\widetilde{\beta}(\theta), \theta, \ell(\widetilde{\beta}(\theta), \theta)\right)$. In particular, the points $(\widetilde{\beta}, 0, \widetilde{\ell})$ and $(\widehat{\beta}, \widehat{\theta}, \widehat{\ell})$ are on this curve.

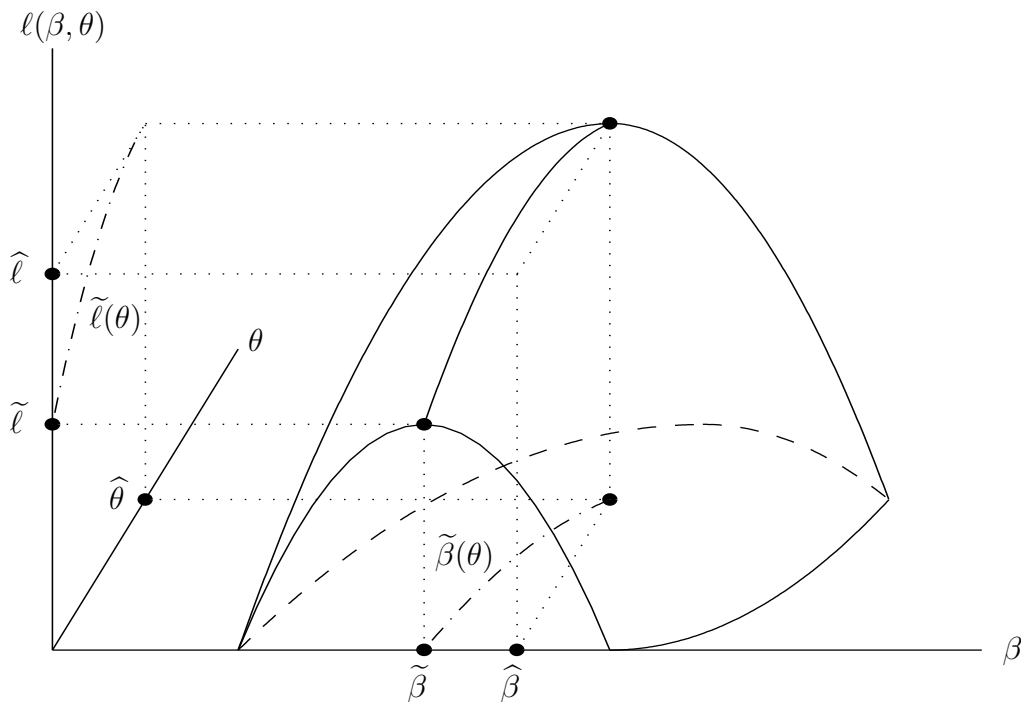[4]See also Helton and Davis (2000, p. 126).

Figure 1. Diagnostic test and sensitivity

The $\widetilde{\beta}(\theta)$-curve is thus the projection of the curve C onto the (β, θ)-plane; we shall call this projection the *sensitivity curve*. In contrast, if we project C onto the (θ, ℓ)-plane, we obtain the curve $\widetilde{\ell}$ defined as

$$\widetilde{\ell}(\theta) := \ell(\widetilde{\beta}(\theta), \theta), \tag{9}$$

which we shall call the *diagnostic curve*. The diagnostic curve $\widetilde{\ell}$ in the (θ, ℓ)-plane contains all relevant information needed to perform the usual diagnostic tests. In particular, the LR test is based on $\widetilde{\ell}(\widehat{\theta}) - \widetilde{\ell}(\widetilde{\theta})$, the Wald test is based on $\widehat{\theta}$, and the LM test is based on the derivative of $\widetilde{\ell}(\theta)$ at $\theta = 0$. The last statement follows from the fact that, at $(\beta, \theta) = (\widetilde{\beta}, 0)$,

$$\frac{\partial \widetilde{\ell}(\theta)}{\partial \theta} = \frac{\partial \ell(\widetilde{\beta}(\theta), \theta)}{\partial \theta} = \frac{\partial \ell(\beta, \theta)}{\partial \beta} \frac{\partial \widetilde{\beta}(\theta)}{\partial \theta} + \frac{\partial \ell(\beta, \theta)}{\partial \theta} = \frac{\partial \ell(\beta, \theta)}{\partial \theta},$$

since $\partial \ell(\beta, \theta)/\partial \beta = 0$ at the restricted maximum, by (4).

Analogous to the LM test in the (θ, ℓ)-plane, the sensitivity of $\widetilde{\beta}$ is the derivative of $\widetilde{\beta}(\theta)$ at $\theta = 0$ in the (β, θ)-plane. The sensitivity thus measures the effect of small changes in θ on the restricted ML estimator $\widetilde{\beta}$.

We now formally introduce the sensitivity statistic. The sensitivity curve contains all restricted ML estimators $\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$, that is, the collection of estimators satisfying the first-order condition

$$\boldsymbol{q}_{\boldsymbol{\beta}}(\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}) = \boldsymbol{0}. \tag{10}$$

The difference between the two estimators $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ can be approximated by the first term of a Taylor expansion,

$$\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}) - \widetilde{\boldsymbol{\beta}}(\boldsymbol{0}) = \left.\frac{\partial \widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{0}} \widehat{\boldsymbol{\theta}} + O_p(1/n).$$

Thus motivated we propose the following definition.

Definition 1 (Sensitivity): The (local) sensitivity of an estimator $\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta})$ to the nuisance parameter $\boldsymbol{\theta}$ at the point $\boldsymbol{0}$ is

$$\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}} := \left.\frac{\partial \widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{0}}. \tag{11}$$

Differentiating the first-order condition (10) with respect to $\boldsymbol{\theta}$ we obtain

$$\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}} = -\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\beta}}^{-1}\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\theta}}. \tag{12}$$

In some cases, it is convenient to scale the first-order condition (10) by a constant $c(\boldsymbol{\beta}, \boldsymbol{\theta})$. This scaling does not affect the sensitivity.

Our definition should be compared with Cook's (1986) definition of 'likelihood displacement'. Let us define the likelihood ratio function by

$$\mathrm{LR}(\boldsymbol{\theta}) = 2\left(\ell(\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}) - \ell(\widetilde{\boldsymbol{\beta}}(\boldsymbol{0}), \boldsymbol{0})\right),$$

so that the usual LR-statistic is given by $\mathrm{LR}(\widehat{\boldsymbol{\theta}})$. Cook's likelihood displacement is closely related to the LR function. In our context it can be defined as

$$\mathrm{LD}(\boldsymbol{\theta}) = -2\left(\ell(\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{0}) - \ell(\widetilde{\boldsymbol{\beta}}(\boldsymbol{0}), \boldsymbol{0})\right).$$

The first derivative of the LD at $\boldsymbol{\theta} = \boldsymbol{0}$ vanishes, because LD reaches its maximum at $\boldsymbol{\theta} = \boldsymbol{0}$, and the second derivative ('Cook's curvature') at $\boldsymbol{\theta} = \boldsymbol{0}$ is

$$\boldsymbol{C}_{\widetilde{\boldsymbol{\beta}}} := \left.\frac{\partial^2 \mathrm{LD}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{0}} = -2n\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\theta\beta}}\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\beta}}^{-1}\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\theta}} = 2n\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\theta\beta}}\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}. \tag{13}$$

Cook's curvature is thus closely related to our sensitivity statistic. The $m \times m$ matrix $\boldsymbol{C}_{\widetilde{\boldsymbol{\beta}}}$ is positive semidefinite, and we can find it largest eigenvalue and the associated eigenvector. This eigenvector gives the 'most sensitive direction' of the LD-curve. While the LD-curve is a general feature, the sensitivity is more informative in specific directions such as we are studying here; see also Cook (1986, p. 136).

The stochastic properties of $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}$ (or $\boldsymbol{C}_{\widetilde{\boldsymbol{\beta}}}$) have not been investigated in the literature. Neither has the relationship between $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}$ (or $\boldsymbol{C}_{\widetilde{\boldsymbol{\beta}}}$) and diagnostic testing been investigated.[5] To these issues we now turn. We shall need some further smoothness conditions on the Hessian matrix.

Assumption 4: There exists a finite $k(k+m) \times k$ matrix $\boldsymbol{R}^0$ and a finite $k(k+m) \times k(k+m)$ positive semidefinite matrix $\boldsymbol{G}^0$ such that

$$\sqrt{n} \operatorname{vec}\left((\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\beta}} : \widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\theta}}) - (\boldsymbol{\mathcal{H}}^0_{\boldsymbol{\beta\beta}} : \boldsymbol{\mathcal{H}}^0_{\boldsymbol{\beta\theta}}) \right) = \boldsymbol{R}^0 \sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + O_p(1/\sqrt{n})$$

and

$$\sqrt{n} \operatorname{vec}\left((\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\beta}} : \widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\theta}}) + (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}} : \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\theta}}) \right) \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \boldsymbol{G}^0).$$

We remark that Assumption 4 can be formulated in terms of 'deeper' assumptions, involving three times continuous differentiability and uniform boundedness, as in Assumption 3 of Newey and Smith (2004, p. 226), but there is no need to do so here.

Theorem 3 (Asymptotic behavior of the sensitivity): Given Assumptions 1–4 and under the null hypothesis $\mathrm{H}_0 : \boldsymbol{\theta} = \boldsymbol{0}$, the sensitivity $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}$ satisfies

$$\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}} \xrightarrow{p} -(\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1} \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\theta}}$$

and

$$\sqrt{n} \operatorname{vec}\left(\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}} + (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1} \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\theta}} \right) \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \boldsymbol{V_S}),$$

where

$$\boldsymbol{V_S} = \begin{pmatrix} (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1} \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\theta}} \otimes (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1} \\ -\boldsymbol{I}_m \otimes (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1} \end{pmatrix}' \boldsymbol{G}^0 \begin{pmatrix} (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1} \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\theta}} \otimes (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1} \\ -\boldsymbol{I}_m \otimes (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1} \end{pmatrix}.$$

In the special case where $\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\theta}} = \boldsymbol{O}$, we obtain $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}} \xrightarrow{p} \boldsymbol{O}$ and $\sqrt{n} \operatorname{vec} \boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}} \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \boldsymbol{V_S})$, with

$$\boldsymbol{V_S} = \left(\boldsymbol{I}_m \otimes (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1} \right) \left(\lim_{n\to\infty} \operatorname{var}(\sqrt{n} \operatorname{vec} \widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\theta}}) \right) \left(\boldsymbol{I}_m \otimes (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1} \right).$$

---

[5]The only exception seems to be Schwarzmann (1991) who shows — for the case of location perturbation of the dependent variable — that the eigenvector associated with the largest eigenvalue of Cook's curvature is proportional to the vector of residuals.

Proof: The first result follows from $\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\beta}} \overset{p}{\longrightarrow} -\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}}$ and $\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\theta}} \overset{p}{\longrightarrow} -\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\theta}}$. In order to obtain the asymptotic distribution of the sensitivity, we write

$$
\sqrt{n}\left(\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}} + (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1}\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\theta}}\right)
$$
$$
= (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1}\left(-\sqrt{n}(\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\theta}} + \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\theta}}) + \sqrt{n}(\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\beta}} + \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})(\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1}\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\theta}}\right)
$$
$$
+ O_p(1/\sqrt{n}).
$$

Vectorizing and using the assumed asymptotic distribution of $(\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\beta}} : \widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta\theta}})$, gives the required result. $\|$

In special cases substantial simplifications occur, as we shall see in Sections 5–7. Theorem 3 allows us to perform a sensitivity test (ST), based on the fact that

$$
\text{ST} := n\left(\text{vec}\left(\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}} + (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1}\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\theta}}\right)\right)' \boldsymbol{V}^+_{\boldsymbol{S}} \left(\text{vec}\left(\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}} + (\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\beta}})^{-1}\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\theta}}\right)\right)
$$

converges to a $\chi^2(r)$ distribution, where r denotes the rank of $\boldsymbol{V_S}$. In particular, when $\boldsymbol{V_S}$ is nonsingular, this gives us an asymptotic $\chi^2(mk)$-test for sensitivity.

Sometimes we are only interested in a subvector of $\boldsymbol{\beta}$ or, more generally, in a function $\boldsymbol{g}(\boldsymbol{\beta})$. We can easily extend the definition by defining the sensitivity of $\boldsymbol{g}(\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}))$ to the nuisance parameter $\boldsymbol{\theta}$ at the point $\boldsymbol{0}$ by

$$
\boldsymbol{S}_{\boldsymbol{g}(\widetilde{\boldsymbol{\beta}})} = \frac{\partial \boldsymbol{g}(\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}))}{\partial \boldsymbol{\beta}'} \frac{\partial \widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\boldsymbol{0}}. \tag{14}
$$

In particular, if we choose $\boldsymbol{g}(\boldsymbol{\beta}) = 2\sqrt{n}\boldsymbol{q}_{\boldsymbol{\theta}}(\boldsymbol{\beta}, \boldsymbol{0})$, then the sensitivity is equal to Cook's curvature of the LD function. Cook's curvature can thus be interpreted as the speed with which $\boldsymbol{q}_{\boldsymbol{\theta}}(\widetilde{\boldsymbol{\beta}}, \boldsymbol{0})$ changes along the sensitivity curve.

The special case where the focus parameter is partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and we are only interested in the sensitivity of $\widetilde{\boldsymbol{\beta}}_1$ to the nuisance parameter $\boldsymbol{\theta}$, is of particular importance as we shall see in Sections 5–7. By (10) we have

$$
\boldsymbol{q}_{\boldsymbol{\beta}_1}(\widetilde{\boldsymbol{\beta}}_1(\boldsymbol{\theta}), \widetilde{\boldsymbol{\beta}}_2(\boldsymbol{\theta}), \boldsymbol{\theta}) = \boldsymbol{0}, \quad \boldsymbol{q}_{\boldsymbol{\beta}_2}(\widetilde{\boldsymbol{\beta}}_1(\boldsymbol{\theta}), \widetilde{\boldsymbol{\beta}}_2(\boldsymbol{\theta}), \boldsymbol{\theta}) = \boldsymbol{0},
$$

and hence, upon differentiating, at $\boldsymbol{\theta} = \boldsymbol{0}$,

$$
\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1}\frac{\partial \widetilde{\boldsymbol{\beta}}_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta}_1\boldsymbol{\beta}_2}\frac{\partial \widetilde{\boldsymbol{\beta}}_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta}_1\boldsymbol{\theta}} = \boldsymbol{O}
$$
$$
\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta}_2\boldsymbol{\beta}_1}\frac{\partial \widetilde{\boldsymbol{\beta}}_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta}_2\boldsymbol{\beta}_2}\frac{\partial \widetilde{\boldsymbol{\beta}}_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta}_2\boldsymbol{\theta}} = \boldsymbol{O}.
$$

This gives

$$S_{\widetilde{\beta}_1} = -\widetilde{\underline{\mathcal{H}}}_{\beta_1\beta_1}^{-1} \widetilde{\underline{\mathcal{H}}}_{\beta_1\theta},$$

where

$$\widetilde{\underline{\mathcal{H}}}_{\beta_1\beta_1} := \widetilde{\mathcal{H}}_{\beta_1\beta_1} - \widetilde{\mathcal{H}}_{\beta_1\beta_2}\widetilde{\mathcal{H}}_{\beta_2\beta_2}^{-1}\widetilde{\mathcal{H}}_{\beta_2\beta_1},$$

and

$$\widetilde{\underline{\mathcal{H}}}_{\beta_1\theta} := \widetilde{\mathcal{H}}_{\beta_1\theta} - \widetilde{\mathcal{H}}_{\beta_1\beta_2}\widetilde{\mathcal{H}}_{\beta_2\beta_2}^{-1}\widetilde{\mathcal{H}}_{\beta_2\theta}.$$

If we define similarly

$$\underline{\mathcal{I}}^0_{\beta_1\beta_1} := \mathcal{I}^0_{\beta_1\beta_1} - \mathcal{I}^0_{\beta_1\beta_2}(\mathcal{I}^0_{\beta_2\beta_2})^{-1}\mathcal{I}^0_{\beta_2\beta_1},$$

and

$$\underline{\mathcal{I}}^0_{\beta_1\theta} := \mathcal{I}^0_{\beta_1\theta} - \mathcal{I}^0_{\beta_1\beta_2}(\mathcal{I}^0_{\beta_2\beta_2})^{-1}\mathcal{I}^0_{\beta_2\theta},$$

then we find $S_{\widetilde{\beta}_1} \overset{p}{\longrightarrow} -(\underline{\mathcal{I}}^0_{\beta_1\beta_1})^{-1}\underline{\mathcal{I}}^0_{\beta_1\theta}$.

In the special case where $\mathcal{I}^0_{\beta_1\beta_2} = \mathbf{O}$, we obtain $S_{\widetilde{\beta}_1} \overset{p}{\longrightarrow} -(\mathcal{I}^0_{\beta_1\beta_1})^{-1}\mathcal{I}^0_{\beta_1\theta}$, and the asymptotic variance of $S_{\widetilde{\beta}_1}$ is based on the relationship

$$\begin{aligned}
&\sqrt{n}\left(S_{\widetilde{\beta}_1} + (\mathcal{I}^0_{\beta_1\beta_1})^{-1}\mathcal{I}^0_{\beta_1\theta}\right) \\
&= (\mathcal{I}^0_{\beta_1\beta_1})^{-1}\big(-\sqrt{n}(\widetilde{\mathcal{H}}_{\beta_1\theta} + \mathcal{I}^0_{\beta_1\theta}) + \sqrt{n}(\widetilde{\mathcal{H}}_{\beta_1\beta_1} + \mathcal{I}^0_{\beta_1\beta_1})(\mathcal{I}^0_{\beta_1\beta_1})^{-1}\mathcal{I}^0_{\beta_1\theta} \\
&\quad + \sqrt{n}(\widetilde{\mathcal{H}}_{\beta_1\beta_2})(\mathcal{I}^0_{\beta_2\beta_2})^{-1}\mathcal{I}^0_{\beta_2\theta}\big) + O_p(1/\sqrt{n})
\end{aligned}$$

together with the joint asymptotic distribution of

$$\sqrt{n}(\widetilde{\mathcal{H}}_{\beta_1\theta} + \mathcal{I}^0_{\beta_1\theta}), \quad \sqrt{n}(\widetilde{\mathcal{H}}_{\beta_1\beta_1} + \mathcal{I}^0_{\beta_1\beta_1}), \quad \sqrt{n}(\widetilde{\mathcal{H}}_{\beta_1\beta_2}).$$

4   Asymptotic independence

We recall from Theorem 2 that the LM test is asymptotically independent of $\widetilde{\beta}$. This does not, however, imply that the LM test is asymptotically independent of the *direction* of $\widetilde{\beta}(\theta)$ at $\theta = \mathbf{0}$, that is, of the sensitivity. It is this type of independence that we address in this section.

We already know that the LM test is based on the score,

$$\widetilde{q}_\theta \approx q^0_\theta - \mathcal{I}^0_{\theta\beta}(\mathcal{I}^0_{\beta\beta})^{-1}q^0_\beta, \tag{15}$$

and that the sensitivity statistic satisfies

$$\begin{aligned}
&\sqrt{n}\left(S_{\widetilde{\beta}} + (\mathcal{I}^0_{\beta\beta})^{-1}\mathcal{I}^0_{\beta\theta}\right) \\
&\approx (\mathcal{I}^0_{\beta\beta})^{-1}\left(-\sqrt{n}(\widetilde{\mathcal{H}}_{\beta\theta} + \mathcal{I}^0_{\beta\theta}) + \sqrt{n}(\widetilde{\mathcal{H}}_{\beta\beta} + \mathcal{I}^0_{\beta\beta})(\mathcal{I}^0_{\beta\beta})^{-1}\mathcal{I}^0_{\beta\theta}\right). \tag{16}
\end{aligned}$$

Based on these facts we wish to demonstrate the following central result.

Theorem 4 (Asymptotic independence of sensitivity and LM test): Let Assumptions 1–4 hold. Then, under the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{0}$, the LM test and the sensitivity statistic $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}$ are asymptotically independent if and only if the correlation between

$$\boldsymbol{q}_{\boldsymbol{\theta}}^0 - \boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}\boldsymbol{\beta}}^0(\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0)^{-1}\boldsymbol{q}_{\boldsymbol{\beta}}^0,$$

and

$$\sqrt{n}(\boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0 + \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0) - \sqrt{n}(\boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0 + \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0)(\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0)^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0$$

vanishes asymptotically.

In the special case where $\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0 = \boldsymbol{O}$, asymptotic independence of the LM test and the sensitivity statistic occurs if and only if the correlation between $\boldsymbol{q}_{\boldsymbol{\theta}}^0$ and $\sqrt{n}\boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0$ approaches zero. The latter condition is satisfied if $\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}\boldsymbol{\theta}}^0$ does not depend on $\boldsymbol{\beta}$ and the correlation between $\boldsymbol{q}_{\boldsymbol{\beta}}^0$ and $\sqrt{n}\boldsymbol{q}_{\boldsymbol{\theta}}^0\boldsymbol{q}_{\boldsymbol{\theta}}^{0\prime}$ approaches zero.

Proof: In view of the joint asymptotic normality, the LM test and the sensitivity statistic will be asymptotically independent if and only if the correlation between

$$\boldsymbol{q}_{\boldsymbol{\theta}}^0 - \boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}\boldsymbol{\beta}}^0(\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0)^{-1}\boldsymbol{q}_{\boldsymbol{\beta}}^0,$$

and

$$\sqrt{n}(\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta}\boldsymbol{\theta}} + \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0) - \sqrt{n}(\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta}\boldsymbol{\beta}} + \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0)(\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0)^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0$$

vanishes asymptotically. Now write

$$\mathrm{vec}\left((\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta}\boldsymbol{\beta}} : \widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta}\boldsymbol{\theta}}) - (\boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0 : \boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0)\right) \approx \boldsymbol{R}^0(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0).$$

Then,

$$\sqrt{n}\,\mathrm{vec}\left((\widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta}\boldsymbol{\beta}} : \widetilde{\boldsymbol{\mathcal{H}}}_{\boldsymbol{\beta}\boldsymbol{\theta}}) + (\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0 : \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0)\right)$$
$$\approx \boldsymbol{R}^0\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \sqrt{n}\,\mathrm{vec}\left((\boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0 : \boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0) + (\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0 : \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0)\right)$$
$$\approx \boldsymbol{R}^0(\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0)^{-1}\boldsymbol{q}_{\boldsymbol{\beta}}^0 + \sqrt{n}\,\mathrm{vec}\left((\boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0 : \boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0) + (\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0 : \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0)\right).$$

Since $\boldsymbol{q}_{\boldsymbol{\beta}}^0$ is asymptotically independent of $\boldsymbol{q}_{\boldsymbol{\theta}}^0 - \boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}\boldsymbol{\beta}}^0(\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}^0)^{-1}\boldsymbol{q}_{\boldsymbol{\beta}}^0$, by Theorem 2, the first result follows. In the special case $\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}\boldsymbol{\theta}}^0 = \boldsymbol{O}$, the result follows from the fact that

$$\frac{\partial}{\partial\boldsymbol{\beta}'}\,\mathrm{vec}\,\mathrm{E}(\boldsymbol{q}_{\boldsymbol{\theta}}^0\boldsymbol{q}_{\boldsymbol{\theta}}^{0\prime}) = \sqrt{n}\,\mathrm{E}\left((\boldsymbol{I}_m \otimes \boldsymbol{q}_{\boldsymbol{\theta}}^0 + \boldsymbol{q}_{\boldsymbol{\theta}}^0 \otimes \boldsymbol{I}_m)\boldsymbol{\mathcal{H}}_{\boldsymbol{\theta}\boldsymbol{\beta}}^0\right)$$
$$+ \sqrt{n}\,\mathrm{E}((\mathrm{vec}\,\boldsymbol{q}_{\boldsymbol{\theta}}^0\boldsymbol{q}_{\boldsymbol{\theta}}^{0\prime})\boldsymbol{q}_{\boldsymbol{\beta}}^{0\prime}). \parallel$$

The condition for independence is essentially a third-moment condition in the same spirit as Corollary 4.4 in Newey and Smith (2004, p. 229). We will see in the examples of Sections 5–7 that the condition is satisfied for a wide class of situations. The special case where $\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta\theta}} = \mathbf{O}$ and $\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\theta\theta}}$ does not depend on $\boldsymbol{\beta}$ occurs often, for example in the case of variance misspecification; see Magnus (1978, Theorem 3, p. 288) and Section 6 below. It is not true that the sensitivity and the diagnostic are *always* independent. For example, the sensitivity of $\widetilde{\sigma}^2$ is typically not independent of the diagnostic test, as we shall see in Section 6.

Again we consider separately the important special case where the focus parameter $\boldsymbol{\beta}$ is partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, and we are only interested in the sensitivity of $\widetilde{\boldsymbol{\beta}}_1$ to the nuisance parameter $\boldsymbol{\theta}$. The condition in Theorem 4 then concerns the correlation between

$$\underline{\boldsymbol{q}}^0_{\boldsymbol{\theta}} - \underline{\boldsymbol{\mathcal{I}}}^0_{\boldsymbol{\theta\beta}_1}(\underline{\boldsymbol{\mathcal{I}}}^0_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1})^{-1}\underline{\boldsymbol{q}}^0_{\boldsymbol{\beta}_1}, \tag{17}$$

and

$$\sqrt{n}(\underline{\boldsymbol{\mathcal{H}}}^0_{\boldsymbol{\beta}_1\boldsymbol{\theta}} + \underline{\boldsymbol{\mathcal{I}}}^0_{\boldsymbol{\beta}_1\boldsymbol{\theta}}) - \sqrt{n}(\underline{\boldsymbol{\mathcal{H}}}^0_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1} + \underline{\boldsymbol{\mathcal{I}}}^0_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1})(\underline{\boldsymbol{\mathcal{I}}}^0_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1})^{-1}\underline{\boldsymbol{\mathcal{I}}}^0_{\boldsymbol{\beta}_1\boldsymbol{\theta}}, \tag{18}$$

where

$$\underline{\boldsymbol{q}}^0_{\boldsymbol{\beta}_1} := \boldsymbol{q}^0_{\boldsymbol{\beta}_1} - \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta}_1\boldsymbol{\beta}_2}(\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta}_2\boldsymbol{\beta}_2})^{-1}\boldsymbol{q}^0_{\boldsymbol{\beta}_2}, \quad \underline{\boldsymbol{q}}^0_{\boldsymbol{\theta}} := \boldsymbol{q}^0_{\boldsymbol{\theta}} - \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\theta\beta}_2}(\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta}_2\boldsymbol{\beta}_2})^{-1}\boldsymbol{q}^0_{\boldsymbol{\beta}_2}.$$

The special case where $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ are asymptotically independent is especially useful.

Theorem 5 (Asymptotic independence of sensitivity and LM test, special case): Consider the case where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and we are interested in the sensitivity of $\widetilde{\boldsymbol{\beta}}_1$ to $\boldsymbol{\theta}$. Let Assumptions 1–4 hold and let $\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta}_1\boldsymbol{\beta}_2} = \mathbf{O}$. Then, under the null hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$, the LM test and the sensitivity statistic $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}_1}$ are asymptotically independent if and only if the correlation between

$$\boldsymbol{q}^0_{\boldsymbol{\theta}} - \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\theta\beta}_1}(\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1})^{-1}\boldsymbol{q}^0_{\boldsymbol{\beta}_1} - \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\theta\beta}_2}(\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta}_2\boldsymbol{\beta}_2})^{-1}\boldsymbol{q}^0_{\boldsymbol{\beta}_2}$$

and

$$\sqrt{n}(\boldsymbol{\mathcal{H}}^0_{\boldsymbol{\beta}_1\boldsymbol{\theta}} + \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta}_1\boldsymbol{\theta}}) - \sqrt{n}(\boldsymbol{\mathcal{H}}^0_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1} + \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1})(\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1})^{-1}\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta}_1\boldsymbol{\theta}}$$
$$- \sqrt{n}\boldsymbol{\mathcal{H}}^0_{\boldsymbol{\beta}_1\boldsymbol{\beta}_2}(\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta}_2\boldsymbol{\beta}_2})^{-1}\boldsymbol{\mathcal{I}}^0_{\boldsymbol{\beta}_2\boldsymbol{\theta}}$$

vanishes asymptotically.

Proof: If $\boldsymbol{\mathcal{I}}^0_{\beta_1\beta_2} = \mathbf{O}$, we find $\underline{\boldsymbol{\mathcal{I}}}^0_{\beta_1\beta_1} = \boldsymbol{\mathcal{I}}^0_{\beta_1\beta_1}$, $\underline{\boldsymbol{\mathcal{I}}}^0_{\beta_1\theta} = \boldsymbol{\mathcal{I}}^0_{\beta_1\theta}$, and $\underline{\boldsymbol{q}}^0_{\beta_1} = \boldsymbol{q}^0_{\beta_1}$. In addition,

$$\underline{\widetilde{\boldsymbol{\mathcal{H}}}}_{\beta_1\beta_1} = \widetilde{\boldsymbol{\mathcal{H}}}_{\beta_1\beta_1} + O_p(1/n),$$

$$\underline{\widetilde{\boldsymbol{\mathcal{H}}}}_{\beta_1\theta} = \widetilde{\boldsymbol{\mathcal{H}}}_{\beta_1\theta} - \frac{1}{\sqrt{n}}(\sqrt{n}\widetilde{\boldsymbol{\mathcal{H}}}_{\beta_1\beta_2})(\boldsymbol{\mathcal{I}}^0_{\beta_2\beta_2})^{-1}\boldsymbol{\mathcal{I}}^0_{\beta_2\theta} + O_p(1/n),$$

while

$$\underline{\boldsymbol{q}}^0_{\boldsymbol{\theta}} = \boldsymbol{q}^0_{\boldsymbol{\theta}} - \boldsymbol{\mathcal{I}}^0_{\boldsymbol{\theta}\beta_2}(\boldsymbol{\mathcal{I}}^0_{\beta_2\beta_2})^{-1}\boldsymbol{q}^0_{\beta_2}.$$

is not simplified. The result now follows from (17) and (18). ∥

This completes the theoretical part of the paper. We now turn to three examples.

5   Misspecification in the mean

Our first example is the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon}|(\boldsymbol{X}, \boldsymbol{Z}) \sim \mathrm{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n),$$

where we consider $(\boldsymbol{\beta}, \sigma^2)$ as the focus parameter, and $\boldsymbol{\theta}$ as the nuisance parameter. We are interested in the sensitivity of $\boldsymbol{\beta}$ with respect to $\boldsymbol{\theta}$. The likelihood is the product of the conditional likelihood (conditional on $(\boldsymbol{X} : \boldsymbol{Z})$) and the likelihood of $(\boldsymbol{X} : \boldsymbol{Z})$. The conditional log-likelihood is given by

$$\ell = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\theta}).$$

The score vector is

$$\boldsymbol{q}^0 = \begin{pmatrix} \boldsymbol{q}^0_{\boldsymbol{\beta}} \\ q^0_{\sigma^2} \\ \boldsymbol{q}^0_{\boldsymbol{\theta}} \end{pmatrix} = \frac{1}{\sqrt{n}}\begin{pmatrix} \boldsymbol{X}'\boldsymbol{\varepsilon}/\sigma^2_0 \\ (\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - n\sigma^2_0)/(2\sigma^4_0) \\ \boldsymbol{Z}'\boldsymbol{\varepsilon}/\sigma^2_0 \end{pmatrix}$$

and the Hessian matrix is

$$\boldsymbol{\mathcal{H}}^0 = -\frac{1}{n}\begin{pmatrix} \boldsymbol{X}'\boldsymbol{X}/\sigma^2_0 & \boldsymbol{X}'\boldsymbol{\varepsilon}/\sigma^4_0 & \boldsymbol{X}'\boldsymbol{Z}/\sigma^2_0 \\ * & (\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - n\sigma^2_0/2)/\sigma^6_0 & \boldsymbol{Z}'\boldsymbol{\varepsilon}/\sigma^4_0 \\ * & * & \boldsymbol{Z}'\boldsymbol{Z}/\sigma^2_0 \end{pmatrix}.$$

We notice that $\boldsymbol{\mathcal{H}}^0_{\beta\sigma^2}$ and $\boldsymbol{\mathcal{H}}^0_{\sigma^2\boldsymbol{\theta}}$ are both of the order $O_p(1/\sqrt{n})$, reflecting the fact that $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}})$ is asymptotically independent of $\widehat{\sigma}^2$. Hence, according

to Theorem 5, the LM test and the sensitivity $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}$ are independent if the two expressions

$$\boldsymbol{q}_{\boldsymbol{\theta}}^0 - \boldsymbol{\mathcal{I}}_{\boldsymbol{\theta\beta}}^0 (\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}}^0)^{-1} \boldsymbol{q}_{\boldsymbol{\beta}}^0$$

and

$$\sqrt{n}(\boldsymbol{\mathcal{H}}_{\boldsymbol{\beta\theta}}^0 + \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\theta}}^0) - \sqrt{n}(\boldsymbol{\mathcal{H}}_{\boldsymbol{\beta\beta}}^0 + \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}}^0)(\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}}^0)^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\theta}}^0$$

are asymptotically uncorrelated. Now, the first expression is asymptotically proportional to $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{\varepsilon}/\sqrt{n}$, where $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. The second expression depends only on $\boldsymbol{X}$ and $\boldsymbol{Z}$, and has finite variance by Assumption 4. Hence they are asymptotically uncorrelated due to the regression condition $\mathrm{E}(\boldsymbol{\varepsilon}|\boldsymbol{X}, \boldsymbol{Z}) = \boldsymbol{0}$.

The restricted estimator is $\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, the LM test takes the form

$$\mathrm{LM} = \frac{\boldsymbol{y}'\boldsymbol{M}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y}/n},$$

the sensitivity in this example is $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}} = -(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}$, and we have shown that $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}$ and LM are asymptotically independent. In this case we can prove a stronger result: $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}$ and LM are independent in finite samples as well. This follows from the fact that the Wald test in this case is proportional to an F-distribution. As shown by Godfrey (1988, p. 51), the LM and LR tests are related to the Wald test by

$$\mathrm{LM} = \frac{\mathrm{W}}{1 + \mathrm{W}/n}, \quad \mathrm{LR} = n\log(1 + \mathrm{W}/n),$$

and hence the distribution of LM (and W and LR) does not depend on $(\boldsymbol{X}, \boldsymbol{Z})$. Thus, for any two measurable functions ϕ and ψ,

$$\mathrm{E}\left(\phi(\mathrm{LM})\psi(\boldsymbol{X}, \boldsymbol{Z})\right) = \mathrm{E}\left(\mathrm{E}(\phi(\mathrm{LM})|\boldsymbol{X}, \boldsymbol{Z})\psi(\boldsymbol{X}, \boldsymbol{Z})\right)$$
$$= \mathrm{E}\left(\phi(\mathrm{LM})\right)\mathrm{E}\left(\psi(\boldsymbol{X}, \boldsymbol{Z})\right).$$

Not only are LM and $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}$ uncorrelated, but any two measurable functions of LM and $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}$ are uncorrelated as well. Then, by Doob (1953, p. 92), LM and $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}$ are independent, and the same holds for the Wald and LR tests. We note, however, that $\widetilde{\boldsymbol{q}}_{\boldsymbol{\theta}}$ and $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}$ are only asymptotically independent, because the conditional distribution of $\widetilde{\boldsymbol{q}}_{\boldsymbol{\theta}}$ does depend on $(\boldsymbol{X}, \boldsymbol{Z})$.

6   Misspecification in the variance

Our second example concerns the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon}|\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{\Omega}(\theta)),$$

where $\boldsymbol{\Omega}(0) = \boldsymbol{I}_n$. Again we regard $(\boldsymbol{\beta}, \sigma^2)$ as the focus parameter and θ as the single nuisance parameter. Extensions to more than one nuisance parameter and to the nonlinear regression model are straightforward.[6] We are interested in the sensitivity of $\boldsymbol{\beta}$ to θ. The log-likelihood (conditional on $\boldsymbol{X}$) is

$$\ell = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \log |\boldsymbol{\Omega}| - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X\beta})' \boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}).$$

Letting $\boldsymbol{V}_1 := \partial \boldsymbol{\Omega}(\theta)/\partial \theta$ and $\boldsymbol{V}_2 := \partial^2 \boldsymbol{\Omega}(\theta)/\partial \theta^2$, both at $\theta = 0$, the score vector is given by

$$\boldsymbol{q}^0 = \begin{pmatrix} \boldsymbol{q}_{\boldsymbol{\beta}}^0 \\ \boldsymbol{q}_{\sigma^2}^0 \\ \boldsymbol{q}_{\theta}^0 \end{pmatrix} = \frac{1}{\sqrt{n}} \begin{pmatrix} \boldsymbol{X}'\boldsymbol{\varepsilon}/\sigma_0^2 \\ (\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - n\sigma_0^2)/(2\sigma_0^4) \\ (\boldsymbol{\varepsilon}'\boldsymbol{V}_1\boldsymbol{\varepsilon} - \sigma_0^2 \operatorname{tr} \boldsymbol{V}_1)/(2\sigma_0^2) \end{pmatrix}$$

and the Hessian matrix is

$$\boldsymbol{\mathcal{H}}^0 = -\frac{1}{n} \begin{pmatrix} \boldsymbol{X}'\boldsymbol{X}/\sigma_0^2 & \boldsymbol{X}'\boldsymbol{\varepsilon}/\sigma_0^4 & \boldsymbol{X}'\boldsymbol{V}_1\boldsymbol{\varepsilon}/\sigma_0^2 \\ * & (\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - n\sigma_0^2/2)/\sigma_0^6 & \boldsymbol{\varepsilon}'\boldsymbol{V}_1\boldsymbol{\varepsilon}/(2\sigma_0^4) \\ * & * & * \end{pmatrix},$$

with

$$\boldsymbol{\mathcal{H}}_{\theta\theta}^0 = -\frac{1}{2n\sigma_0^2} \left((2\boldsymbol{\varepsilon}'\boldsymbol{V}_1^2\boldsymbol{\varepsilon} - \sigma_0^2 \operatorname{tr} \boldsymbol{V}_1^2) - (\boldsymbol{\varepsilon}'\boldsymbol{V}_2\boldsymbol{\varepsilon} - \sigma_0^2 \operatorname{tr} \boldsymbol{V}_2) \right).$$

In this example, both $\boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\sigma^2}^0$ and $\boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\theta}^0$ are of the order $O_p(1/\sqrt{n})$, reflecting the fact that $\widehat{\boldsymbol{\beta}}$ is asymptotically independent of $(\widehat{\theta}, \widehat{\sigma}^2)$. Hence Theorem 5 implies that the LM test and the sensitivity $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}$ are independent if the correlation between

$$\boldsymbol{q}_{\theta}^0 - \boldsymbol{\mathcal{I}}_{\theta\sigma^2}^0 (\boldsymbol{\mathcal{I}}_{\sigma^2\sigma^2}^0)^{-1} \boldsymbol{q}_{\sigma^2}^0$$

and

$$\sqrt{n}\boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\theta}^0 - \sqrt{n}\boldsymbol{\mathcal{H}}_{\boldsymbol{\beta}\sigma^2}^0 (\boldsymbol{\mathcal{I}}_{\sigma^2\sigma^2}^0)^{-1} \boldsymbol{\mathcal{I}}_{\sigma^2\theta}^0$$

approaches zero. (We could also have employed the fact that $\boldsymbol{\mathcal{I}}_{\theta\theta}^0$ does not depend on $\boldsymbol{\beta}$, and apply Theorem 4.) The first expression depends only on the two quadratic forms $\boldsymbol{\varepsilon}'\boldsymbol{V}_1\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$, while the second expression depends only on the linear forms $\boldsymbol{X}'\boldsymbol{V}_1\boldsymbol{\varepsilon}$ and $\boldsymbol{X}'\boldsymbol{\varepsilon}$. Hence they are asymptotically independent if both expressions have finite variances in the limit. This is guaranteed if $\operatorname{tr} \boldsymbol{V}_1$, $\operatorname{tr} \boldsymbol{V}_1^2$, $\boldsymbol{X}'\boldsymbol{V}_1\boldsymbol{X}$, and $\boldsymbol{X}'\boldsymbol{X}$ are all of the order $O_p(n)$. This, in turn, is implied by Assumptions 3(c) and 4.

---

[6]Magnus (1978) provides the relevant framework and formulae for this case.

Letting $\boldsymbol{M} := \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, the restricted estimator and the sensitivity are

$$\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \quad \boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}_1\boldsymbol{M}\boldsymbol{y},$$

while the LM test takes the form

$$\mathrm{LM} = \frac{n}{2\,\mathrm{tr}\,\boldsymbol{V}_1^2/n} \left(\frac{\boldsymbol{y}'\boldsymbol{M}\boldsymbol{V}_1\boldsymbol{M}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y}} - \frac{\mathrm{tr}\,\boldsymbol{V}_1}{n} \right)^2,$$

which confirms that the LM test is a quadratic function of $\boldsymbol{\varepsilon}$ while the sensitivity is a linear function.

We notice that LM and $S_{\widetilde{\sigma}^2}$ are not independent in this case, because $S_{\widetilde{\sigma}^2} = -\boldsymbol{y}'\boldsymbol{M}\boldsymbol{V}_1\boldsymbol{M}\boldsymbol{y}/n$, which is strongly correlated with LM.

The fact that asymptotically LM and $\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}}$ are independent, does not tell us how fast the convergence takes place. Thus, we perform a Monte Carlo experiment in the same spirit as Banerjee and Magnus (1999). For a given value of n, we generate five regressors: constant, time trend, normal distribution $N(0,9)$, lognormal distribution $\log N(0,9)$, and uniform distribution $U[-2,2]$. Based upon these five regressors we consider ten data sets: five with two regressors and five with three regressors, as follows:

1: constant, linear trend   6: constant, linear trend, $N(0,9)$
2: constant, $N(0,9)$ 7: constant, linear trend, $\log N(0,9)$
3: constant, $\log N(0,9)$ 8: constant, $\log N(0,9)$, $U[-2,2]$
4: $N(0,9)$, $U[-2,2]$ 9: $N(0,9)$, $\log N(0,9)$, $U[-2,2]$
5: linear trend, $N(0,9)$ 10: linear trend, $N(0,9)$, $U[-2,2]$.

Our assumed alternative is the AR(1) model with parameter θ. Assuming that the null hypothesis that $\theta = 0$ is true, we calculate critical values ST^* and LM^* such that

$$\Pr(\mathrm{ST} > \mathrm{ST}^*) = \Pr(\mathrm{LM} > \mathrm{LM}^*) = 0.05,$$

where ST refers to the (one-dimensional) sensitivity *test* rather than the multi-dimensional sensitivity statistic $\boldsymbol{S}$. If ST and LM are independent, then the conditional probability $\Pr(\mathrm{ST} < \mathrm{ST}^*|\mathrm{LM} \geq \mathrm{LM}^*)$ will be equal to 0.95. If, on the other hand, ST and LM are perfectly dependent, then the conditional probability will be zero.[7] We performed 100,000 Monte Carlo simulations for each of the ten models and for each of $n = 25, 50, 100, 250, 500$, and 1000. Figure 2a demonstrates that the convergence to independence

---

[7]We look at the conditional probabilities rather than at the correlations, because this combines the convergence of the relevant random variables to normality with the convergence of the correlations to zero. The convergence of the correlations to zero is more rapid than the converge to normality.
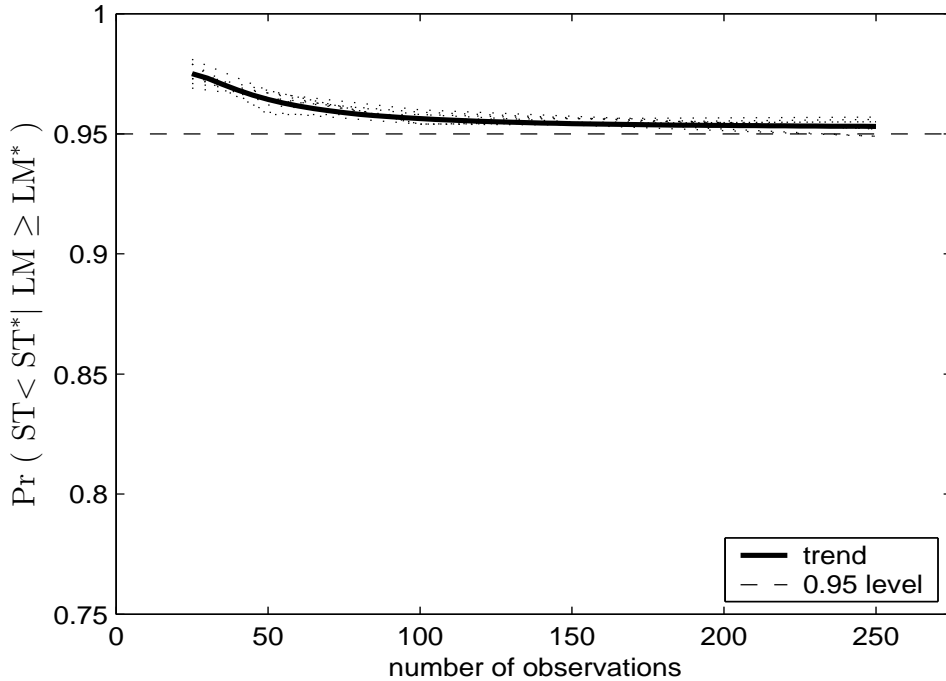
Figure 2a. Nonscalar variance: independence of LM test and sensitivity

is fast, and that the behavior for each of the ten data sets is similar. Interestingly, the LM test and the sensitivity test are *negatively* correlated in this case.[8]

We have chosen the LM test as our diagnostic test. The LR test and the Wald tests are asymptotically the same as the LM test, but not in finite samples. Hence, the LR and Wald tests will also be asymptotically independent of the sensitivity test, but the speed of convergence could be different. This is analyzed in Figures 2b and 2c. All three tests converge quickly to the 95% line; the Wald test is the slowest. The Wald test and the LR test are both positively correlated with the sensitivity test.

7   Misspecification in the distribution

Our third and final example concerns the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \sigma\boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon}|\boldsymbol{X} \sim \mathrm{D}(\boldsymbol{0}, \boldsymbol{I}_n),$$

---

[8]On average the correlation ρ between the LM test and the sensitivity test is $\rho = -0.06$ for $n = 25$, $\rho = -0.03$ for $n = 50$, and $\rho = -0.01$ for $n = 100$.
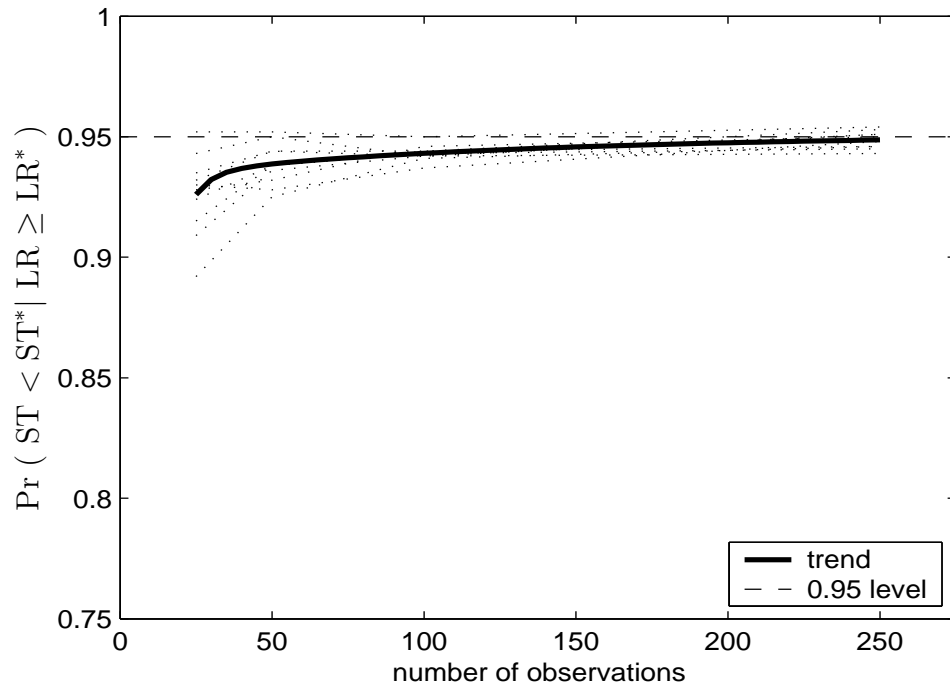
Figure 2b. Nonscalar variance: independence of LR test and sensitivity
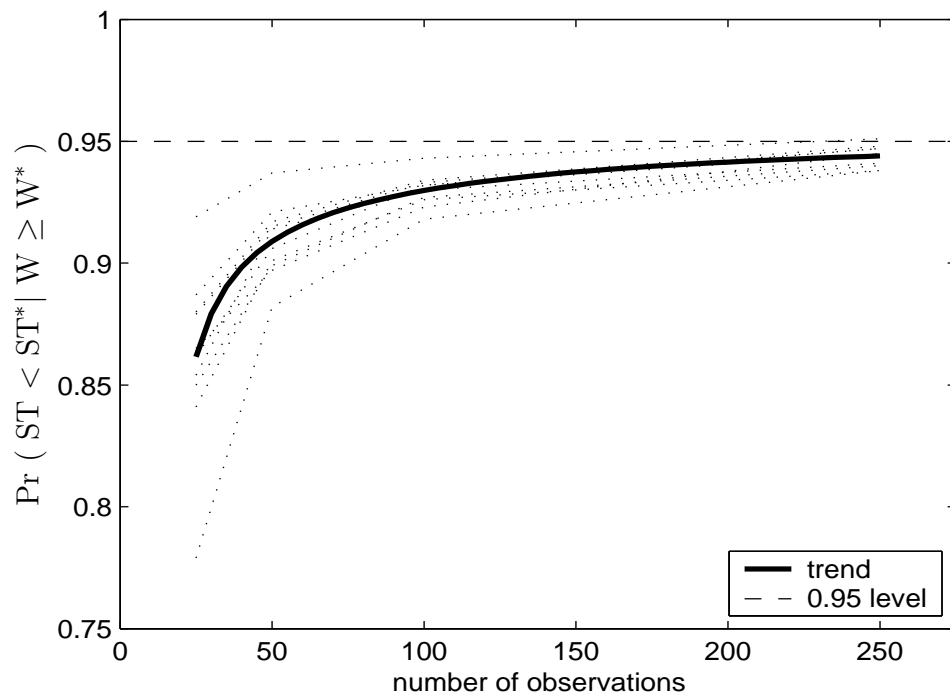
Figure 2c. Nonscalar variance: independence of Wald test and sensitivity

where the $\varepsilon_1, \ldots, \varepsilon_n$, conditional on $\boldsymbol{X}$, are i.i.d. with mean zero and variance one. We do not assume that the distribution D is normal. Instead we assume that the ε_i follow a general Pearson distribution defined implicitly by

$$\frac{\mathrm{d}\log f(\varepsilon_i)}{\mathrm{d}\varepsilon_i} = \frac{\theta_1 - \varepsilon_i}{1 - \theta_1\varepsilon_i + \theta_2(\varepsilon_i^2 - 3)},$$

see Kendall and Stuart (1976, p. 159). Notice that in our formulation of the Pearson family the random variable is scaled so that its variance is one. We regard $(\boldsymbol{\beta}, \sigma^2)$ as the focus parameter and $\boldsymbol{\theta} = (\theta_1, \theta_2)$ as the nuisance parameter. At $\boldsymbol{\theta} = \boldsymbol{0}$ we obtain $\mathrm{d}\log f(\varepsilon_i)/\mathrm{d}\varepsilon_i = -\varepsilon_i$ which defines the $\mathrm{N}(0,1)$ distribution.

We are interested in the sensitivity of $\boldsymbol{\beta}$ to $\boldsymbol{\theta}$. The log-likelihood conditional on $\boldsymbol{X}$ is given by

$$\ell = -\frac{n}{2}\log\sigma^2 + \sum_{i=1}^{n}\log f(\varepsilon_i), \quad \varepsilon_i = \frac{y_i - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma},$$

where y_i denotes the i-th component of $\boldsymbol{y}$ and $\boldsymbol{x}_i'$ denotes the i-th row of $\boldsymbol{X}$.

The score vector is given by

$$\boldsymbol{q}^0 = \begin{pmatrix} \boldsymbol{q}_{\boldsymbol{\beta}}^0 \\ \boldsymbol{q}_{\sigma^2}^0 \\ \boldsymbol{q}_{\boldsymbol{\theta}}^0 \end{pmatrix} = \frac{1}{\sqrt{n}}\begin{pmatrix} \boldsymbol{X}'\boldsymbol{\varepsilon}/\sigma_0 \\ (\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - n)/(2\sigma_0^2) \\ * \end{pmatrix},$$

where

$$\boldsymbol{q}_{\boldsymbol{\theta}}^0 = \frac{1}{\sqrt{n}}\begin{pmatrix} \frac{1}{3}\sum_i \varepsilon_i(3 - \varepsilon_i^2) \\ \frac{1}{4}\sum_i(\varepsilon_i^4 - 6\varepsilon_i^2 + 3) \end{pmatrix},$$

and the Hessian matrix is

$$\boldsymbol{\mathcal{H}}^0 = -\frac{1}{n}\begin{pmatrix} \boldsymbol{X}'\boldsymbol{X}/\sigma_0^2 & \boldsymbol{X}'\boldsymbol{\varepsilon}/\sigma_0^3 & * \\ * & (\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - n/2)/\sigma_0^4 & * \\ * & * & * \end{pmatrix},$$

with

$$\boldsymbol{\mathcal{H}}_{\boldsymbol{\theta}\boldsymbol{\beta}}^0 = -\frac{1}{n\sigma_0}\begin{pmatrix} \sum_i(1 - \varepsilon_i^2)\boldsymbol{x}_i' \\ \sum_i \varepsilon_i(\varepsilon_i^2 - 3)\boldsymbol{x}_i' \end{pmatrix}, \quad \boldsymbol{\mathcal{H}}_{\boldsymbol{\theta}\sigma^2}^0 = -\frac{1}{2n\sigma_0^2}\begin{pmatrix} \sum_i \varepsilon_i(1 - \varepsilon_i^2) \\ \sum_i \varepsilon_i^2(\varepsilon_i^2 - 3) \end{pmatrix},$$

and

$$\boldsymbol{\mathcal{H}}_{\boldsymbol{\theta}\boldsymbol{\theta}}^0 = -\frac{1}{n}\begin{pmatrix} \frac{1}{6}\sum_i(3\varepsilon_i^4 - 6\varepsilon_i^2 + 1) & \frac{1}{15}\sum_i(-6\varepsilon_i^5 + 35\varepsilon_i^3 - 45\varepsilon_i) \\ * & \frac{1}{6}\sum_i(2\varepsilon_i^6 - 18\varepsilon_i^4 + 54\varepsilon_i^2 - 21) \end{pmatrix}.$$

In this example, three blocks of the Hessian matrix, namely $\mathcal{H}^0_{\beta\sigma^2}$, $\mathcal{H}^0_{\beta\theta}$, and $\mathcal{H}^0_{\sigma^2\theta}$ are all of the order $O_p(1/\sqrt{n})$, reflecting the fact that $\widehat{\beta}$, $\widehat{\sigma}^2$, and $\widehat{\theta}$ are asymptotically independent. (In fact, since $-\mathrm{E}(\mathcal{H}^0_{\theta\theta}) = \mathrm{E}(q^0_\theta q^0_\theta) = \mathrm{diag}(2/3, 3/2)$, $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are asymptotically independent as well.) Hence Theorem 5 implies that the LM test and the sensitivity $S_{\widetilde{\beta}}$ are independent if and only if the correlation between q^0_θ and $\sqrt{n}\mathcal{H}^0_{\beta\theta}$ approaches zero. In general, this is not the case. We have

$$\sqrt{n}\,\mathrm{E}\left(q^0_{\theta_1}\mathcal{H}^0_{\beta\theta_1}\right) \to \mathbf{0}, \quad \sqrt{n}\,\mathrm{E}\left(q^0_{\theta_2}\mathcal{H}^0_{\beta\theta_2}\right) \to \mathbf{0}, \quad \sqrt{n}\,\mathrm{E}\left(q^0_{\theta_2}\mathcal{H}^0_{\beta\theta_1}\right) \to \mathbf{0},$$

but

$$\mathrm{corr}\left(\sqrt{n}\mathcal{H}^0_{\beta\theta_2}, q^0_{\theta_1}|\mathbf{X}\right) = \frac{1}{\sqrt{n}}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'\boldsymbol{\imath},$$

where $\boldsymbol{\imath}$ denotes the $n \times 1$ vector of ones. Hence, the LM test and the sensitivity $S_{\widetilde{\beta}}$ are independent if and only if $\sqrt{n}\mathcal{H}^0_{\beta\theta_2}$ and $q^0_{\theta_1}$ are asymptotically uncorrelated, that is, if and only if $\boldsymbol{\imath}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\imath}/n \xrightarrow{p} 0$.

We now distinguish between two cases. First, if the regression contains a constant term, then there is no loss in generality in taking the other regressors in deviation from their respective means. The sensitivity of the slope parameters is then asymptotically independent of the LM test. The sensitivity of the constant term itself will be correlated with the LM test because $\boldsymbol{\imath}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\imath}/n = 1$ for $\mathbf{X} = \boldsymbol{\imath}$.

Second, if the regression contains no constant term, then the sensitivity of the parameters will in general be correlated with the LM test, unless the regressors happen to be centered at zero. For example, if $\mathbf{x}$ is a sample from a distribution with mean μ and variance σ^2, then $\boldsymbol{\imath}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\boldsymbol{\imath}/n \xrightarrow{p} \mu^2/(\mu^2+\sigma^2)$, which is zero if and only if $\mu/\sigma = 0$. If $\mathbf{x}$ is the trend $1, 2, \ldots, n$, then $\boldsymbol{\imath}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\boldsymbol{\imath}/n \to 3/4$.

The restricted estimators are

$$\widetilde{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \quad \widetilde{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y}/n,$$

where $\mathbf{M} := \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Letting $\widetilde{\varepsilon} = \mathbf{M}\mathbf{y}/\widetilde{\sigma}$, the sensitivity of $\widetilde{\beta}$ to θ_1 and θ_2 is

$$S_{\widetilde{\beta}} = -\widetilde{\sigma}(\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{i=1}^n (1-\widetilde{\varepsilon}_i^2)\mathbf{x}_i : \sum_{i=1}^n \widetilde{\varepsilon}_i^3 \mathbf{x}_i\right) + O_p(1/n),$$

while the LM test is the Jarque-Bera test,[9]

$$\mathrm{LM} = n\left(\frac{\widetilde{\mu}_3^2}{6} + \frac{(\widetilde{\mu}_4 - 3)^2}{24}\right),$$

---

[9]See Jarque and Bera (1980, 1987).

where

$$\widetilde{\mu}_3 = \frac{1}{n}\sum_{i=1}^{n}(\widetilde{\varepsilon}_i - \widetilde{\mu}_1')^3, \quad \widetilde{\mu}_4 = \frac{1}{n}\sum_{i=1}^{n}(\widetilde{\varepsilon}_i - \widetilde{\mu}_1')^4, \quad \widetilde{\mu}_1' = \frac{1}{n}\sum_{i=1}^{n}\widetilde{\varepsilon}_i.$$

Instead of calculating the sensitivities of $\widetilde{\boldsymbol{\beta}}$ with respect to θ_1 and θ_2, we can also calculate the sensitivities with respect to the skewness μ_3 and the kurtosis μ_4, using the relationships

$$\mu_3 = \frac{2\theta_1}{4\theta_2 - 1}, \quad \mu_4 - 3 = \frac{6(\theta_1^2 - 4\theta_2^2 + \theta_2)}{(4\theta_2 - 1)(5\theta_2 - 1)}.$$

The sensitivity of $\widetilde{\boldsymbol{\beta}}$ to μ_3 and μ_4 is then

$$\boldsymbol{S}_{\widetilde{\boldsymbol{\beta}}} = -\widetilde{\sigma}(\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\frac{1}{2}\sum_{i=1}^{n}(\widetilde{\boldsymbol{\varepsilon}}_i^2 - 1)\boldsymbol{x}_i : \frac{1}{6}\sum_{i=1}^{n}\widetilde{\boldsymbol{\varepsilon}}_i^3\boldsymbol{x}_i\right) + O_p(1/n).$$

In Figure 3 we present the probability that the estimator $\widetilde{\boldsymbol{\beta}}$ is not sensitive
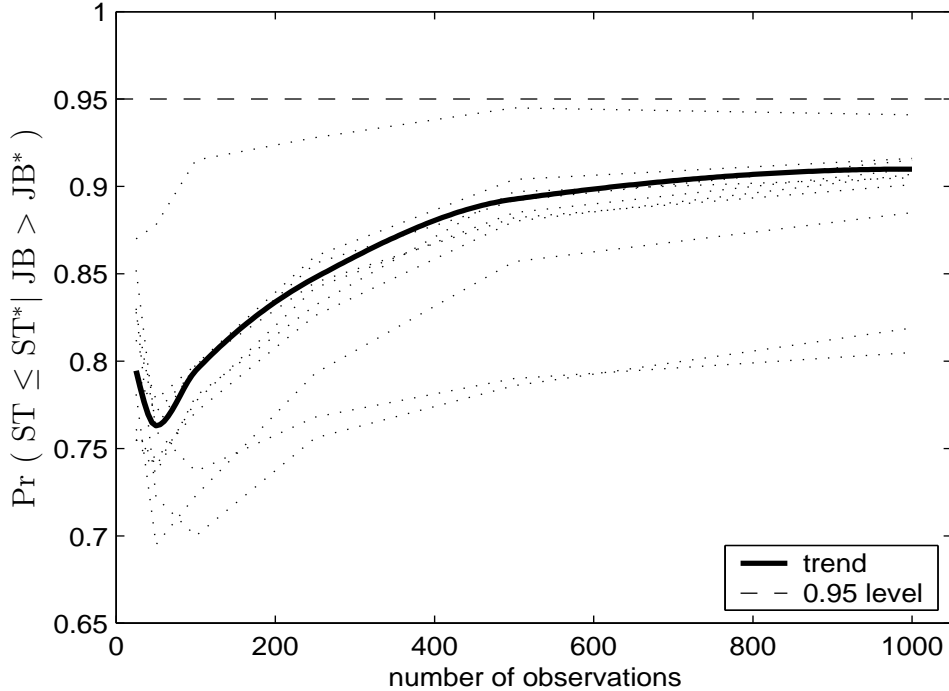

Figure 3. Nonnormality: (in)dependence of JB-test and sensitivity

to non-normality ($\mathrm{ST} \leq \mathrm{ST}^*$), while the Jarque-Bera test rejects the null

hypothesis of normality (JB > JB*). The sensitivity test is for the slope parameters only. Data sets 1–3 and 6–8 contain a constant term and the conditional probability therefore converges to 95%. The other four data sets do not contain a constant term, and indeed the probability in data sets 5 and 10 does not converge to 95%. Data set 4 contains no constant term, but the two regressors are both centered at zero; hence the probability also converges to 95%. Finally, data set 9 contains two regressors that are centered at zero, and one (a sample from the lognormal distribution) which is not centered at zero. Nevertheless it looks as if the probability converges here also to 95%. The reason is that for a regressor sampled from the $\log \mathrm{N}(0,9)$-distribution, the correlation between $\sqrt{n}\mathcal{H}^0_{\beta\theta_2}$ and $\boldsymbol{q}^0_{\theta_1}$ converges to $\mathrm{e}^{-9/2} \approx 0.01$, which is not zero, but close to zero. A special case of particular importance is
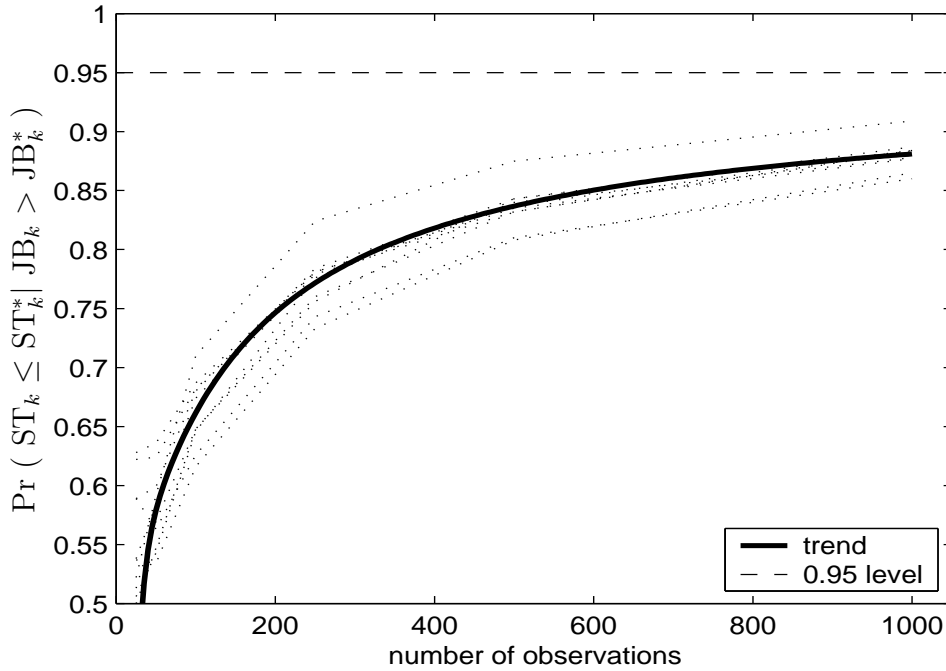

Figure 4. Kurtosis: independence of JB-test and sensitivity

the case where $\theta_1 = 0$, so that there is no skewness. In this case — which corresponds to the (scaled) t-distribution — we can test for kurtosis. Since $\sqrt{n}\,\mathrm{E}\left(\boldsymbol{q}^0_{\theta_1}\,\mathcal{H}^0_{\beta\theta_1}\right) \to \boldsymbol{0}$, the JB-test for kurtosis and the sensitivity test are asymptotically independent in this case, whether the regressors are measured in deviations or not. Figure 4 shows that the convergence to independence is rather slow however.

8 Conclusion

Sensitivity analysis matters. We argue that the usual diagnostic test provides only half the information required to decide whether a restricted estimator is good enough to learn about the focus parameters in the model; the other half is provided by a sensitivity test.

The results in this paper can be generalized from maximum likelihood to extremum estimators, and applied to a great variety of situations that are more complex than the relatively simple ones considered. Also, other characteristics of the sensitivity curve (apart from the derivative at zero) can be considered.

Sensitivity analysis is also important for its own sake, not in combination with diagnostics. It will help to expose the weakest link in a project, be it the model formulation, the data, the estimation method, or something else. This is our ultimate goal: to learn from a simple model in which direction we should generalize. The current paper is just a small step in this direction.

References

Abadir, K. M. and J. R. Magnus (2002). Notation in econometrics: A proposal for a standard, *The Econometrics Journal*, 5, 76–90.

Andrews, D. W. K. (1998). Hypothesis testing with a restricted parameter space, *Journal of Econometrics*, 84, 155–199.

Banerjee, A. N. and J. R. Magnus (1999). The sensitivity of OLS when the variance matrix is (partially) unknown, *Journal of Econometrics*, 92, 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*, 95, 157–176.

Chatterjee, S. and A. S. Hadi (1988). *Sensitivity Analysis in Linear Regression*, Wiley Series in Probability and Statistics, John Wiley, New York.

Cook, R. D. (1979). Influential observations in linear regression, *Journal of the American Statistical Association*, 74, 169-174.

Cook, R. D. (1986). Assessment of local influence (with discussion), *Journal of the Royal Statistical Society, Series B*, 48, 133-169.

Cook, R. D. and S. Weisberg (1982). *Residuals and Influence in Regression*, Chapman and Hall, New York.

Doob, J. L. (1953). *Stochastic Processes*, John Wiley, New York.

Fung, W. K. and M. K. Tang (1997). Assessment of local influence in multivariate regression analysis, *Communications in Statistics : Theory and Methods*, 26, 821–837.

Gal, T. and H. Greenberg, Eds (1997), *Advances in Sensitivity Analysis and Parametric Programming*, Kluwer, Dordrecht.

Galea, M., G. A. Paula, and H. Bolfarine (1997). Local influence in elliptical linear regression models, *The Statistician*, 46, 71–79.

Godfrey, L. G. (1988). *Misspecification Tests in Econometrics*, Cambridge University Press, Cambridge.

Hartless, G., J. G. Booth, and R. C. Littell (2003). Local influence of predictors in multiple linear regression, *Technometrics*, 45, 326–332.

Heijmans, R. D. H. and J. R. Magnus (1986a). Consistent maximum-likelihood estimation with dependent observations, *Journal of Econometrics*, 32, 253–285. Corrigenda in *Journal of Econometrics*, 35, 395.

Heijmans, R. D. H. and J. R. Magnus (1986b). On the first-order efficiency and asymptotic normality of maximum likelihood estimators obtained from dependent observations, *Statistica Neerlandica*, 40, 169–188.

Helton, J. C. and F. J. Davis (2000). Sampling-based methods, in: *Sensitivity Analysis* (Eds A. Saltelli, K. Chan, and E. M. Scott), Wiley Series in Probability and Statistics, John Wiley, Chichester.

Huber, P. J. (2004). *Robust Statistics*, Wiley Series in Probability and Statistics, John Wiley, Hoboken, NJ.

Jarque, C. M. and A. K. Bera (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Economics Letters*, 6, 255–259.

Jarque, C. M. and A. K. Bera (1987). A test for normality of observations and regression residuals, *International Statistical Review*, 55, 163–172.

Kendall, M. and A. Stuart (1976). *The Advanced Theory of Statistics*, Volume 1, Fourth Edition, Charles Griffin & Co Ltd, London.

Kleijnen, J. P. C. (1997). Sensitivity analysis and related analyses: A review of some statistical techniques, *Journal of Statistical Computation and Simulation*, 57, 111–142.

Leamer, E. E. (1978). *Specification Searches*, John Wiley, New York.

Leamer, E. E. (1984). Global sensitivity results for generalized least squares estimates, *Journal of the American Statistical Association*, 79, 867–870.

Liu, S. (2000). On local influence for elliptical linear models, *Statistical Papers*, 41, 211–224.

Liu, S. (2002). Local influence in multivariate elliptical linear regression models, *Linear Algebra and Its Applications*, 354, 159–174.

Magnus, J. R. (1978). Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. *Journal of Econometrics*, 7, 281–312.

Newey, W. K. and R. J. Smith (2004). Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica*, 72, 219–255.

Pan, J.-X., K.-T. Fang, and D. von Rosen (1997). Local influence assessment in the growth curve model with unrestricted covariance, *Journal of Statistical Planning and Inference*, 62, 263–278.

Polasek, W. (1984). Regression diagnostics for general linear regression models, *Journal of the American Statistical Association*, 79, 336–340.

Ruud, P. A. (2000). *An Introduction to Classical Econometric Theory*, Oxford University Press, New York/Oxford.

Saltelli, A., K. Chan, and E. M. Scott, Eds (2000). *Sensitivity Analysis*, Wiley Series in Probability and Statistics, John Wiley, Chichester.

Schwarzmann, B. (1991). A connection between local-influence analysis and residual diagnostics, *Technometrics*, 33, 103–104.

Shi, L. and X. Wang (1999). Local influence in ridge regression, *Computational Statistics and Data Analysis*, 31, 341–353.

Verbeke, G., G. Molenberghs, H. Thijs, E. Lesaffre, and M. G. Kenward (2001). Sensitivity analysis for nonrandom dropout: A local influence approach, *Biometrics*, 57, 7–14.