Econometrics Journal (2007), volume **10**, pp. 166–192.
doi: 10.1111/j.1368-423X.2007.00204.x

Local sensitivity and diagnostic tests

JAN R. MAGNUS AND ANDREY L. VASNEV

Center and Department of Econometrics and Operations Research, Tilburg University, The Netherlands

E-mail: magnus@uvt.nl; a.vasnev@uvt.nl

Received: March 2006

Summary In this paper, we confront sensitivity analysis with diagnostic testing. Every model is misspecified (in the sense that no model coincides with the data-generating process), 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 (statistically) '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, give conditions under which the diagnostic and the sensitivity 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.

Keywords: *Maximum likelihood*, *Sensitivity*, *Diagnostic test*, *Regression*.

1. MOTIVATION AND BACKGROUND

Since models are often compared to maps, let us consider the map of Europe in Figure 1. At first glance, the map appears to be fine. But upon further investigation, it becomes clear that the map is not useful when planning a trip from Palermo to Lisbon, from Ibiza to Dublin, or from Zagreb to Ajaccio. In fact, the map fails in eight important respects: Ireland seems to have been replaced by Iceland, Normandy is where Brittany used to be and vice versa, Sardinia and Corsica have changed places, Sicily has disappeared, the Balearic Islands have moved from the south-east of Spain to the north-west, Portugal and north-west Spain have disappeared completely, and so have the Baltic States, Albania and former Yugoslavia, so that Austria and Hungary now border the Adriatic Sea. As a map of Europe, this is not a good one. Nevertheless, for many itineraries the map is perfectly adequate. The route from Paris to Warsaw, for example, is correctly specified.

The example illustrates the well-known fact that inadequate models may still be useful in certain important directions, and it leads to the distinction between diagnostic testing and sensitivity analysis. A diagnostic test asks the question whether the model is correct (and would be rejected in our map example), while sensitivity analysis asks whether deviations from the truth are important in the direction that we are interested in (which is not the case if we travel from Paris to Warsaw). These are different questions and the second question (the sensitivity) seems to be the

Figure 1. Map of Europe.

more important. In econometric practice, however, diagnostic testing is the norm and sensitivity analysis the exception. If diagnostic testing and sensitivity analysis were closely correlated, then this would not be a problem. The main result of this paper, however, is that the diagnostic test and the sensitivity are (asymptotically) independent. Hence sensitivity analysis does matter.

In econometric terms, suppose we wish to estimate $\boldsymbol{\beta}$ from the linear regression model

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

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 $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, while the least-squares estimators for $\boldsymbol{\beta}$ and θ in the unrestricted model are

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

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

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

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

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

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

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

Hence, the larger is $\widetilde{\theta}$ (the diagnostic), the more important is the sensitivity. We may think of $\widetilde{\theta}$ as the 'magnitude' and of $\boldsymbol{S}$ as the 'direction' of the impact of the misspecification on $\widehat{\boldsymbol{\beta}}$. (Figure 2 in Section 3 further clarifies this interpretation.) Seen in this light it becomes important to investigate the relationship between $\widetilde{\theta}$ and $\boldsymbol{S}$. 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 $\widetilde{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}$ 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 $\widetilde{\theta}$ is (statistically) 'large' or 'small'. Since θ is a nuisance parameter, we are not primarily interested in whether $\widetilde{\theta}$ is large or small (relative to its precision); our interest is in $\boldsymbol{\beta}$. It may very well be that $\widetilde{\theta}$ is large, but that nevertheless the difference between $\widetilde{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}$ 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 will 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 idea of checking the sensitivity of objects of interest to restrictions underlies also Hausman's (1978) specification test, which builds on Durbin (1954). Hausman is not interested in whether certain restrictions on the parameters hold, but rather whether the parameters are estimated consistently. Hausman's test is, however, not a sensitivity; it is a diagnostic as Holly (1987, p. 68) has demonstrated. Our paper is closer to the idea of 'local inconsistency' contained in Kiefer and Skoog (1984). We will return to these papers after we have formally introduced sensitivity.

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 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 and obtain its asymptotic distribution. Some often-occurring special cases are considered as well. The conditions under which the sensitivity and the diagnostic 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

[1] See Chatterjee and Hadi (1988) for an early summary of sensitivity analysis in linear regression, and Saltelli *et al.* (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; 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 *et al.* 2001), multivariate elliptical linear regression (Liu, 2002), and prediction (Hartless, Booth and Littell 2003). See Liu (2002) for further references. In Section 3, we will relate Cook's method to our definition of sensitivity.

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

on the data to ensure the existence of certain expansions and the proper behaviour 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 $\mathbb{R}^p$,*
(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 $\kappa := \kappa(\boldsymbol{\delta}, \boldsymbol{\delta}_0) := -\mathrm{E}\,(\ell(\boldsymbol{\delta}) - \ell(\boldsymbol{\delta}_0))$ denote the absolute value of the Kullback–Leibler information. Then,*

(a) *$\kappa(\boldsymbol{\delta}, \boldsymbol{\delta}_0) \to \infty$ for every $\boldsymbol{\delta} \neq \boldsymbol{\delta}_0$,*
(b) *$(1/\kappa^2)\,\mathrm{var}(\ell(\boldsymbol{\delta}) - \ell(\boldsymbol{\delta}_0)) \to 0$,*
(c) *for every $\boldsymbol{\delta} \neq \boldsymbol{\delta}_0 \in \mathcal{D}$ there exists a neighbourhood $N(\boldsymbol{\delta})$ of $\boldsymbol{\delta}$ such that*

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

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 $\widetilde{\boldsymbol{\delta}}$ of $\boldsymbol{\delta}$ exists and is consistent (see Heijmans and Magnus 1986a).

We now define the score and the Hessian matrix as

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

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) $\boldsymbol{\mathcal{H}}(\boldsymbol{\delta}_0) \xrightarrow{p} -\boldsymbol{\mathcal{I}}(\boldsymbol{\delta}_0)$,
(c) *the information matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\delta})$ is continuous on $\mathcal{D}^0$ and $\boldsymbol{\mathcal{I}}(\boldsymbol{\delta}_0)$ is positive definite,*
(d) *for every $\epsilon > 0$ there exists a neighbourhood $N(\boldsymbol{\delta}_0)$ of $\boldsymbol{\delta}_0$ such that*

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

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

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

(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 by $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$, 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 by $\widehat{\boldsymbol{\beta}}$.

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

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

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

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

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

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

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

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

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 behaviour, we need the first-order expansions:

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

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

and

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

The following well-known 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 $\widetilde{\boldsymbol{\beta}}$ is consistent and asymptotically normal such that*

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

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

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

$$\begin{aligned}\sqrt{n}(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) &= (\mathcal{I}^0)^{-1}\boldsymbol{q}(\boldsymbol{\delta}_0) + O_p(1/\sqrt{n}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, (\mathcal{I}^0)^{-1}),\\ \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= \left(\mathcal{I}^0_{\beta\beta}\right)^{-1}\boldsymbol{q}_\beta(\boldsymbol{\beta}_0, \mathbf{0}) + O_p(1/\sqrt{n}) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \left(\mathcal{I}^0_{\beta\beta}\right)^{-1}\right),\end{aligned}$$

and

$$\begin{aligned}\widehat{\boldsymbol{q}}_\theta &= \boldsymbol{q}^0_\theta - \mathcal{I}^0_{\theta\beta}\left(\mathcal{I}^0_{\beta\beta}\right)^{-1}\boldsymbol{q}^0_\beta + O_p(1/\sqrt{n})\\ &\xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \mathcal{I}^0_{\theta\theta} - \mathcal{I}^0_{\theta\beta}\left(\mathcal{I}^0_{\beta\beta}\right)^{-1}\mathcal{I}^0_{\beta\theta}\right),\end{aligned}$$

and the results follow. □

The expansions (6)–(8) also imply the following basic orthogonality result, which is related to the arguments in Kiviet and Phillips (1986, section 5) and Davidson and MacKinnon (1987, section 4); (see also proposition 1 in Bickel *et al.* 1993).

Theorem 2. *(Asymptotic independence of* $\widehat{\boldsymbol{\beta}}$ and LM test): *If Assumptions 1–3 hold and* $\boldsymbol{\theta}_0 = \mathbf{0}$, *then*

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

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

Proof. For two random vectors z_1 and z_2, let $z_1 \approx z_2$ denote 'asymptotic equality' in the sense that $z_1 = z_2 + O_p(1/\sqrt{n})$. Since $\widehat{\boldsymbol{q}}_\theta \approx \boldsymbol{q}^0_\theta - \mathcal{I}^0_{\theta\beta}(\mathcal{I}^0_{\beta\beta})^{-1}\boldsymbol{q}^0_\beta$, we obtain

$$\begin{aligned}\mathrm{cov}\left(\widehat{\boldsymbol{q}}_\theta, \boldsymbol{q}^0_\beta\right) &\approx \mathrm{E}\left(\boldsymbol{q}^0_\theta \boldsymbol{q}^{0\prime}_\beta\right) - \mathcal{I}^0_{\theta\beta}(\mathcal{I}^0_{\beta\beta})^{-1}\mathrm{E}\left(\boldsymbol{q}^0_\beta \boldsymbol{q}^{0\prime}_\beta\right)\\ &\approx \mathcal{I}^0_{\theta\beta} - \mathcal{I}^0_{\theta\beta}\left(\mathcal{I}^0_{\beta\beta}\right)^{-1}\mathcal{I}^0_{\beta\beta} = \mathbf{O},\end{aligned}$$

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

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

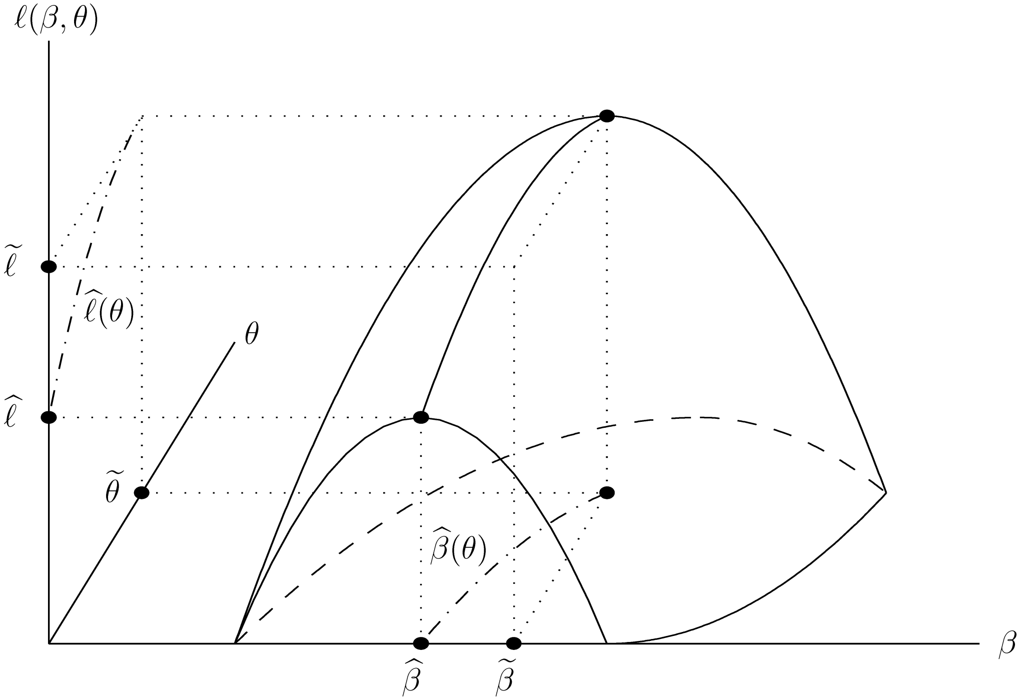

Figure 2. Diagnostic test and sensitivity.

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 $\boldsymbol{\theta} = \mathbf{0}$, but this does not mean that the estimator of the focus parameter $\boldsymbol{\beta}$ will be sensitive to deviations of $\boldsymbol{\theta}$ from $\mathbf{0}$. 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 2, where we assume for simplicity that $k = m = 1$; hence there is one focus parameter β and one nuisance parameter θ. Figure 2 is a generalization of the well-known picture of the three classical tests (see for example Ruud 2000, p. 390). At $(\widetilde{\beta}, \widetilde{\theta})$, we obtain the maximum of the likelihood $\widetilde{\ell}$, while at $(\widehat{\beta}, 0)$, we obtain the restricted maximum $\widehat{\ell}$. For every fixed value of θ, let $\widehat{\beta}(\theta)$ denote the value of β which maximizes the (restricted) likelihood. The locus of all constrained maxima is the curve $C := (\widehat{\beta}(\theta), \theta, \ell(\widehat{\beta}(\theta), \theta))$. In particular, the points $(\widehat{\beta}, 0, \widehat{\ell})$ and $(\widetilde{\beta}, \widetilde{\theta}, \widetilde{\ell})$ are on this curve.

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

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

curve $\widehat{\ell}(\theta)$ defined as

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

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

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

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 $\widehat{\beta}$ is the derivative of $\widehat{\beta}(\theta)$ at $\theta = 0$ in the (β, θ)-plane. The sensitivity thus measures the effect of small changes in θ on the restricted ML estimator $\widehat{\beta}$. The difference between sensitivity and all diagnostic tests (including the Hausman test based on $\widetilde{\beta} - \widehat{\beta}$) is now clear. The diagnostic tests are designed to test how close the restricted maximum $(\widehat{\beta}, 0, \widehat{\ell})$ and the global maximum $(\widetilde{\beta}, \widetilde{\theta}, \widetilde{\ell})$ are on the curve C. Different diagnostic tests look at this closeness from different angles. Sensitivity, in contrast, measures the local behaviour of $\widehat{\beta}(\theta)$ at $\theta = 0$.

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

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

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

$$\widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\widetilde{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\beta}}(\mathbf{0}) = \left.\frac{\partial \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\mathbf{0}} \widetilde{\boldsymbol{\theta}} \; + \; O_p(1/n). \tag{11}$$

Thus motivated we propose the following definition.

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

$$\boldsymbol{S}_{\widehat{\beta}} := \left.\frac{\partial \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\mathbf{0}}. \tag{12}$$

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

$$\boldsymbol{S}_{\widehat{\beta}} = -\widehat{\boldsymbol{\mathcal{H}}}_{\beta\beta}^{-1}\widehat{\boldsymbol{\mathcal{H}}}_{\beta\boldsymbol{\theta}}. \tag{13}$$

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. Also, we might wish to study sensitivity with respect to a function of $\boldsymbol{\theta}$ rather than with respect to $\boldsymbol{\theta}$ itself. The generalization is straightforward.

Let us now return to the Hausman (1978) specification test, which is based on $\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}})$. We distinguish between two cases. If $\boldsymbol{\mathcal{I}}_{\theta\beta}^{0} = \mathbf{O}$, then Theorem 1 shows that the two estimators $\widetilde{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}$ have the same asymptotic distribution, so that Hausman's test is not applicable. Sensitivity analysis, however, goes one step further, because it analyzes the term of order $1/n$ in the expansion (11). If $\mathcal{I}^0_{\theta\beta} \neq \mathbf{0}$, then Hausman's test is equivalent to the LM test (a diagnostic), because the probability limit of the sensitivity is nonzero. Thus, Hausman's test does not measure sensitivity.

Kiefer and Skoog (1984, p. 876) introduce 'local inconsistency', an idea related to Theil (1957), and provide (nonlinear) examples. This comes close to sensitivity analysis. Kiefer and Skoog (1984) do not, however, study properties of the 'local inconsistency' and its relation to the diagnostic.

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

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

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

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

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

$$C_{\widehat{\beta}} := \left.\frac{\partial^2 \mathrm{LD}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\mathbf{0}} = -2n\widehat{\mathcal{H}}_{\theta\beta}\widehat{\mathcal{H}}^{-1}_{\beta\beta}\widehat{\mathcal{H}}_{\beta\theta} = 2n\widehat{\mathcal{H}}_{\theta\beta}S_{\widehat{\beta}}. \tag{14}$$

Cook's curvature is thus closely related to our sensitivity statistic. The $m \times m$ matrix $C_{\widehat{\beta}}$ is positive semidefinite, and we can find its 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 $S_{\widehat{\beta}}$ (or $C_{\widehat{\beta}}$) have not been investigated in the literature. Neither has the relationship between $S_{\widehat{\beta}}$ (or $C_{\widehat{\beta}}$) and diagnostic testing been investigated.[5] To these issues we now turn. We will 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*

$$\sqrt{n}\,\mathrm{vec}\left((\widehat{\mathcal{H}}_{\beta\beta} : \widehat{\mathcal{H}}_{\beta\theta}) - \left(\mathcal{H}^0_{\beta\beta} : \mathcal{H}^0_{\beta\theta}\right)\right) = R^0\sqrt{n}(\widehat{\beta} - \beta_0) + O_p(1/\sqrt{n})$$

and

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

The assumption may appear a little opaque because the matrices R^0 and G^0 are left undefined. These matrices can be obtained (somewhat tediously) from Bartlett identities relating expected derivatives of the log-likelihood function, as in Chesher (1983) and Lancaster (1984). Assumption 4 can also be formulated in terms of 'deeper' assumptions, involving three times continuous

[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.

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 behaviour of the sensitivity): If Assumptions 1–4 hold and* $\boldsymbol{\theta}_0 = \mathbf{0}$, *then the sensitivity* $S_{\widehat{\beta}}$ *satisfies*

$$S_{\widehat{\beta}} \xrightarrow{p} -\left(\mathcal{I}^0_{\beta\beta}\right)^{-1}\mathcal{I}^0_{\beta\boldsymbol{\theta}}$$

and

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

where

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

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

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

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

$$\begin{aligned}
&\sqrt{n}\,\left(S_{\widehat{\beta}} + \left(\mathcal{I}^0_{\beta\beta}\right)^{-1}\mathcal{I}^0_{\beta\boldsymbol{\theta}}\right) \\
&\quad = \left(\mathcal{I}^0_{\beta\beta}\right)^{-1}\left(-\sqrt{n}\left(\widehat{\mathcal{H}}_{\beta\boldsymbol{\theta}} + \mathcal{I}^0_{\beta\boldsymbol{\theta}}\right) + \sqrt{n}\left(\widehat{\mathcal{H}}_{\beta\beta} + \mathcal{I}^0_{\beta\beta}\right)\left(\mathcal{I}^0_{\beta\beta}\right)^{-1}\mathcal{I}^0_{\beta\boldsymbol{\theta}}\right) \\
&\qquad + O_p(1/\sqrt{n}).
\end{aligned}$$

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

In Theorem 3, we have chosen to analyze sensitivity at the point where the nuisance parameter $\boldsymbol{\theta}$ is zero, because this corresponds with our key question: what sort of error does the practitioner make (in terms of sensitivity) when choosing the small model instead of the large model. It is possible to evaluate the sensitivity at some other value of the nuisance parameter and under some other null hypothesis, but we will not pursue this further.

In special cases substantial simplifications occur, as we will see in Sections 5–7. 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 $\boldsymbol{g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}))$ to the nuisance parameter $\boldsymbol{\theta}$ at the point $\mathbf{0}$ by

$$\boldsymbol{S}_{\boldsymbol{g}(\widehat{\beta})} = \left.\frac{\partial \boldsymbol{g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}))}{\partial \boldsymbol{\beta}'}\frac{\partial \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\mathbf{0}}. \tag{15}$$

In particular, if we choose $\boldsymbol{g}(\boldsymbol{\beta}) = 2\sqrt{n}\boldsymbol{q}_{\boldsymbol{\theta}}(\boldsymbol{\beta}, \mathbf{0})$, 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 $\boldsymbol{q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\beta}}, \mathbf{0})$ changes along the sensitivity curve.

These expressions are particularly useful when one is interested in prediction or forecasting. In the case of the linear model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, the predictor is obtained by writing $\boldsymbol{g}(\boldsymbol{\beta}) = \boldsymbol{X\beta}$ and is studied in detail in Banerjee and Magnus (1999). The same paper also studies the sensitivity of the variance, which is the main building block for interval estimation.

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 $\widehat{\boldsymbol{\beta}}_1$ to the nuisance parameter $\boldsymbol{\theta}$, is of particular importance as we will see in Sections 5–7. By (10), we have

$$\boldsymbol{q}_{\beta_1}(\widehat{\boldsymbol{\beta}}_1(\boldsymbol{\theta}), \widehat{\boldsymbol{\beta}}_2(\boldsymbol{\theta}), \boldsymbol{\theta}) = \mathbf{0}, \quad \boldsymbol{q}_{\beta_2}(\widehat{\boldsymbol{\beta}}_1(\boldsymbol{\theta}), \widehat{\boldsymbol{\beta}}_2(\boldsymbol{\theta}), \boldsymbol{\theta}) = \mathbf{0},$$

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

$$\begin{aligned}
\widehat{\mathcal{H}}_{\beta_1\beta_1}\frac{\partial\widehat{\boldsymbol{\beta}}_1(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \widehat{\mathcal{H}}_{\beta_1\beta_2}\frac{\partial\widehat{\boldsymbol{\beta}}_2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \widehat{\mathcal{H}}_{\beta_1\boldsymbol{\theta}} &= \mathbf{O} \\
\widehat{\mathcal{H}}_{\beta_2\beta_1}\frac{\partial\widehat{\boldsymbol{\beta}}_1(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \widehat{\mathcal{H}}_{\beta_2\beta_2}\frac{\partial\widehat{\boldsymbol{\beta}}_2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \widehat{\mathcal{H}}_{\beta_2\boldsymbol{\theta}} &= \mathbf{O}.
\end{aligned}$$

This gives

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

where

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

and

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

If we define similarly

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

and

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

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

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

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

together with the joint asymptotic distribution of

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

4. ASYMPTOTIC INDEPENDENCE

We recall from Theorem 2 that the LM test is asymptotically independent of $\widehat{\beta}$. This does not, however, imply that the LM test is asymptotically independent of the *direction* of $\widehat{\beta}(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \mathbf{0}$, 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,

$$\widehat{\boldsymbol{q}}_{\boldsymbol{\theta}} \approx \boldsymbol{q}_{\boldsymbol{\theta}}^0 - \mathcal{I}_{\theta\beta}^0 (\mathcal{I}_{\beta\beta}^0)^{-1} \boldsymbol{q}_{\beta}^0, \tag{16}$$

and that the sensitivity statistic satisfies

$$\begin{aligned} &\sqrt{n}\ \left(S_{\widehat{\beta}} + (\mathcal{I}_{\beta\beta}^0)^{-1} \mathcal{I}_{\beta\theta}^0\right) \\ &\quad \approx (\mathcal{I}_{\beta\beta}^0)^{-1} \left(-\sqrt{n}(\widehat{\mathcal{H}}_{\beta\theta} + \mathcal{I}_{\beta\theta}^0) + \sqrt{n}(\widehat{\mathcal{H}}_{\beta\beta} + \mathcal{I}_{\beta\beta}^0)(\mathcal{I}_{\beta\beta}^0)^{-1} \mathcal{I}_{\beta\theta}^0\right). \end{aligned} \tag{17}$$

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

Theorem 4. *(Asymptotic independence of sensitivity and LM test): If Assumptions 1–4 hold and $\boldsymbol{\theta}_0 = \mathbf{0}$, then the LM test and the sensitivity statistic $\boldsymbol{S}_{\widehat{\beta}}$ are asymptotically independent if and only if the correlation between*

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

and

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

vanishes asymptotically.

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

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

and

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

vanishes asymptotically. Now write

$$\operatorname{vec}\left((\widehat{\mathcal{H}}_{\beta\beta} : \widehat{\mathcal{H}}_{\beta\theta}) - (\mathcal{H}_{\beta\beta}^0 : \mathcal{H}_{\beta\theta}^0)\right) \approx R^0(\widehat{\beta} - \beta_0).$$

Then,

$$\begin{aligned} &\sqrt{n}\operatorname{vec}\left((\widehat{\mathcal{H}}_{\beta\beta} : \widehat{\mathcal{H}}_{\beta\theta}) + (\mathcal{I}_{\beta\beta}^0 : \mathcal{I}_{\beta\theta}^0)\right) \\ &\quad \approx R^0\sqrt{n}(\widehat{\beta} - \beta_0) + \sqrt{n}\operatorname{vec}\left((\mathcal{H}_{\beta\beta}^0 : \mathcal{H}_{\beta\theta}^0) + (\mathcal{I}_{\beta\beta}^0 : \mathcal{I}_{\beta\theta}^0)\right) \\ &\quad \approx R^0(\mathcal{I}_{\beta\beta}^0)^{-1}\boldsymbol{q}_{\beta}^0 + \sqrt{n}\operatorname{vec}\left((\mathcal{H}_{\beta\beta}^0 : \mathcal{H}_{\beta\theta}^0) + (\mathcal{I}_{\beta\beta}^0 : \mathcal{I}_{\beta\theta}^0)\right). \end{aligned}$$

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

$$\frac{\partial}{\partial\boldsymbol{\beta}'}\mathrm{vecE}\big(\boldsymbol{q}_{\theta}^{0}\boldsymbol{q}_{\theta}^{0\prime}\big)=\sqrt{n}\mathrm{E}\big(\big(I_m\otimes\boldsymbol{q}_{\theta}^{0}+\boldsymbol{q}_{\theta}^{0}\otimes I_m\big)\mathcal{H}_{\theta\beta}^{0}\big)$$
$$+\sqrt{n}\mathrm{E}\big(\big(\mathrm{vec}\boldsymbol{q}_{\theta}^{0}\boldsymbol{q}_{\theta}^{0\prime}\big)\boldsymbol{q}_{\beta}^{0\prime}\big).\qquad\square$$

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 $\mathcal{I}_{\beta\theta}^{0}=\mathbf{O}$ and $\mathcal{I}_{\theta\theta}^{0}$ 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 $\widehat{\sigma}^2$ is typically not independent of the diagnostic test, as we will 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 $\widehat{\boldsymbol{\beta}}_1$ to the nuisance parameter $\boldsymbol{\theta}$. The condition in Theorem 4 then concerns the correlation between

$$\underline{\boldsymbol{q}}_{\theta}^{0}-\underline{\mathcal{I}}_{\theta\beta_1}^{0}\big(\underline{\mathcal{I}}_{\beta_1\beta_1}^{0}\big)^{-1}\underline{\boldsymbol{q}}_{\beta_1}^{0},\tag{18}$$

and

$$\sqrt{n}\big(\underline{\mathcal{H}}_{\beta_1\theta}^{0}+\underline{\mathcal{I}}_{\beta_1\theta}^{0}\big)-\sqrt{n}\big(\underline{\mathcal{H}}_{\beta_1\beta_1}^{0}+\underline{\mathcal{I}}_{\beta_1\beta_1}^{0}\big)\big(\underline{\mathcal{I}}_{\beta_1\beta_1}^{0}\big)^{-1}\underline{\mathcal{I}}_{\beta_1\theta}^{0},\tag{19}$$

where

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

The special case where $\widetilde{\boldsymbol{\beta}}_1$ and $\widetilde{\boldsymbol{\beta}}_2$ 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 $\widehat{\boldsymbol{\beta}}_1$ to $\boldsymbol{\theta}$. If Assumptions 1–4 hold, $\boldsymbol{\theta}_0=\mathbf{0}$, and, in addition, $\mathcal{I}_{\beta_1\beta_2}^{0}=\mathbf{O}$, then the LM test and the sensitivity statistic $\boldsymbol{S}_{\widehat{\beta}_1}$ are asymptotically independent if and only if the correlation between*

$$\boldsymbol{q}_{\theta}^{0}-\mathcal{I}_{\theta\beta_1}^{0}\big(\mathcal{I}_{\beta_1\beta_1}^{0}\big)^{-1}\boldsymbol{q}_{\beta_1}^{0}-\mathcal{I}_{\theta\beta_2}^{0}\big(\mathcal{I}_{\beta_2\beta_2}^{0}\big)^{-1}\boldsymbol{q}_{\beta_2}^{0}$$

and

$$\sqrt{n}\big(\mathcal{H}_{\beta_1\theta}^{0}+\mathcal{I}_{\beta_1\theta}^{0}\big)-\sqrt{n}\big(\mathcal{H}_{\beta_1\beta_1}^{0}+\mathcal{I}_{\beta_1\beta_1}^{0}\big)\big(\mathcal{I}_{\beta_1\beta_1}^{0}\big)^{-1}\mathcal{I}_{\beta_1\theta}^{0}$$
$$-\sqrt{n}\mathcal{H}_{\beta_1\beta_2}^{0}\big(\mathcal{I}_{\beta_2\beta_2}^{0}\big)^{-1}\mathcal{I}_{\beta_2\theta}^{0}$$

vanishes asymptotically.

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

$$\begin{aligned}\underline{\widehat{\mathcal{H}}}_{\beta_1\beta_1} &= \widehat{\mathcal{H}}_{\beta_1\beta_1} + O_p(1/n),\\ \underline{\widehat{\mathcal{H}}}_{\beta_1\theta} &= \widehat{\mathcal{H}}_{\beta_1\theta} - \frac{1}{\sqrt{n}}(\sqrt{n}\widehat{\mathcal{H}}_{\beta_1\beta_2})\big(\mathcal{I}^0_{\beta_2\beta_2}\big)^{-1}\mathcal{I}^0_{\beta_2\theta} + O_p(1/n),\end{aligned}$$

while

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

is not simplified. The result now follows from (18) and (19). □

We comment briefly on the question how sensitivity can be studied for the more general class of extremum estimators. An extremum estimator $\overline{\boldsymbol{\delta}}$ of $\boldsymbol{\delta}$ maximizes $Q(\boldsymbol{\delta})$ over $\boldsymbol{\delta}$ for some real-valued function Q. The regularity conditions for Q are typically weaker than those for the log-likelihood function. As a result, the asymptotic distribution of $\overline{\boldsymbol{\delta}}$ might be non-normal and the convergence rate might be different from the traditional $\sqrt{n}$-convergence rate (for example, Manski's (1975) maximum score estimator has a convergence rate of $n^{-1/3}$ and a nonstandard limiting distribution; see Kim and Pollard 1990). If Q is twice differentiable, sensitivity may be defined similar to (13) as

$$\boldsymbol{S}_{\widehat{\beta}} = -\widehat{Q}^{-1}_{\beta\beta}\widehat{Q}_{\beta\theta},$$

where

$$\widehat{Q}_{\beta\beta} = \left.\frac{\partial^2 Q}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right|_{\beta=\widehat{\beta},\theta=\mathbf{0}} \quad \text{and} \quad \widehat{Q}_{\beta\theta} = \left.\frac{\partial^2 Q}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}'}\right|_{\beta=\widehat{\beta},\theta=\mathbf{0}}.$$

The asymptotic variances of the estimators and the sensitivity statistic will change because the information equality $\mathrm{E}(\boldsymbol{q}^0_{\delta}\boldsymbol{q}^{0'}_{\delta}) = -\mathrm{E}(\mathcal{H}^0_{\delta\delta})$ does not necessarily hold. Diagnostic tests are usually considered only for specific subclasses of extremum estimators. For example, in the generalized method of moments, Wald, gradient and distance difference tests are comparable to the usual Wald, LM and LR tests (see Ruud 2000, chapter 22). In other subclasses, the asymptotic independence between sensitivity and well-defined diagnostics may still hold as long as the asymptotic expansions remain valid.

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

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

where we consider $(\boldsymbol{\beta}, \sigma^2)$ as the focus parameter and $\boldsymbol{\theta}$ as the nuisance parameter. The linearity is chosen for simplicity of exposition only. The non-linear regression model can be treated similarly, for example, along the lines of Kiefer and Skoog's (1984) analysis of the probit model.

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

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

The score vector is

$$\boldsymbol{q}^0 = \begin{pmatrix} \boldsymbol{q}^0_{\beta} \\ \boldsymbol{q}^0_{\sigma^2} \\ \boldsymbol{q}^0_{\theta} \end{pmatrix} = \frac{1}{\sqrt{n}} \begin{pmatrix} \boldsymbol{X}'\boldsymbol{\epsilon}/\sigma_0^2 \\ \left(\boldsymbol{\epsilon}'\boldsymbol{\epsilon} - n\sigma_0^2\right)/\left(2\sigma_0^4\right) \\ \boldsymbol{Z}'\boldsymbol{\epsilon}/\sigma_0^2 \end{pmatrix}$$

and the Hessian matrix is

$$\mathcal{H}^0 = -\frac{1}{n}\begin{pmatrix} \boldsymbol{X}'\boldsymbol{X}/\sigma_0^2 & \boldsymbol{X}'\boldsymbol{\epsilon}/\sigma_0^4 & \boldsymbol{X}'\boldsymbol{Z}/\sigma_0^2 \\ * & \left(\boldsymbol{\epsilon}'\boldsymbol{\epsilon} - n\sigma_0^2/2\right)/\sigma_0^6 & \boldsymbol{Z}'\boldsymbol{\epsilon}/\sigma_0^4 \\ * & * & \boldsymbol{Z}'\boldsymbol{Z}/\sigma_0^2 \end{pmatrix}.$$

We note that $\mathcal{H}^0_{\beta\sigma^2}$ and $\mathcal{H}^0_{\sigma^2\theta}$ are both of the order $O_p(1/\sqrt{n})$, reflecting the fact that $(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\theta}})$ is asymptotically independent of $\widetilde{\sigma}^2$. Hence, according to Theorem 5, the LM test and the sensitivity $\boldsymbol{S}_{\widehat{\beta}}$ are independent if the two expressions

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

and

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

are asymptotically uncorrelated. Now, the first expression is asymptotically proportional to $\boldsymbol{Z}'\boldsymbol{M\epsilon}/\sqrt{n}$, 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 $\mathrm{E}(\boldsymbol{\epsilon} \mid \boldsymbol{X}, \boldsymbol{Z}) = \boldsymbol{0}$.

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

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

the sensitivity in this example is $\boldsymbol{S}_{\widehat{\beta}} = -(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}$, and we have shown that $\boldsymbol{S}_{\widehat{\beta}}$ and LM are asymptotically independent. In this case, we can prove a stronger result: $\boldsymbol{S}_{\widehat{\beta}}$ 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 (X, Z). Thus, for any two measurable functions ϕ and ψ,

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

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

6. MISSPECIFICATION IN THE VARIANCE

Our second example concerns the linear regression model

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

where $\boldsymbol{\Omega}(0) = I_n$. 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

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

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

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

and the Hessian matrix is

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

with

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

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

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

and

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

approaches zero. (We could also have employed the fact that $\boldsymbol{\mathcal{I}}^0_{\theta\theta}$ does not depend on $\boldsymbol{\beta}$, and apply Theorem 4.) The first expression depends only on the two quadratic forms $\boldsymbol{\epsilon}'\boldsymbol{V}_1\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}'\boldsymbol{\epsilon}$, while the second expression depends only on the linear forms $\boldsymbol{X}'\boldsymbol{V}_1\boldsymbol{\epsilon}$ and $\boldsymbol{X}'\boldsymbol{\epsilon}$. Hence they are

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

asymptotically independent if both expressions have finite variances in the limit. This is guaranteed if $\mathrm{tr}\boldsymbol{V}_1$, $\mathrm{tr}\boldsymbol{V}_1^2$, $\boldsymbol{X}'\boldsymbol{V}_1\boldsymbol{X}$ and $\boldsymbol{X}'\boldsymbol{X}$ are all of the order $O_p(n)$. This, in turn, is implied by Assumptions 3(c) and 4.

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

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

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 $\boldsymbol{\epsilon}$ while the sensitivity is a linear function. We note that LM and $S_{\widehat{\sigma}^2}$ are not independent in this case, because $\boldsymbol{S}_{\widehat{\sigma}^2} = -\boldsymbol{y}'\boldsymbol{M}\boldsymbol{V}_1\boldsymbol{M}\boldsymbol{y}/n$, which is strongly correlated with LM.

The fact that asymptotically LM and $\boldsymbol{S}_{\widehat{\beta}}$ 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: a constant, a time trend, and series from the normal distribution N(0, 9), lognormal distribution log N(0, 9), and uniform distribution r[−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), r[−2, 2]
4: N(0, 9), r[−2, 2]	9: N(0, 9), log N(0, 9), r[−2, 2]
5: linear trend, N(0, 9)	10: linear trend, N(0, 9), r[−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 SS* and LM* such that

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

where we use the one-dimensional 'scaled sensitivity'

$$\mathrm{SS} := n(\mathrm{vec}\,\boldsymbol{S}_{\widehat{\beta}})'\widehat{\boldsymbol{V}}_{\boldsymbol{S}}^{+}(\mathrm{vec}\,\boldsymbol{S}_{\widehat{\beta}}), \tag{20}$$

rather than the multi-dimensional sensitivity statistic $\boldsymbol{S}_{\widehat{\beta}}$. If SS and LM are independent, then the conditional probability $\Pr(\mathrm{SS} < \mathrm{SS}^*|\mathrm{LM} \geq \mathrm{LM}^*)$ will be equal to 0.95. If, on the other hand, SS 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 3 demonstrates that the convergence to independence is fast, and that the behaviour for each of the 10 data sets is similar. Interestingly, the LM test and the (scaled) sensitivity are *negatively* correlated in this case.[8]

[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.

[8] On average, the correlation ρ between the LM test and the scaled sensitivity is $\rho = -0.06$ for $n = 25$, $\rho = -0.03$ for $n = 50$ and $\rho = -0.01$ for $n = 100$.

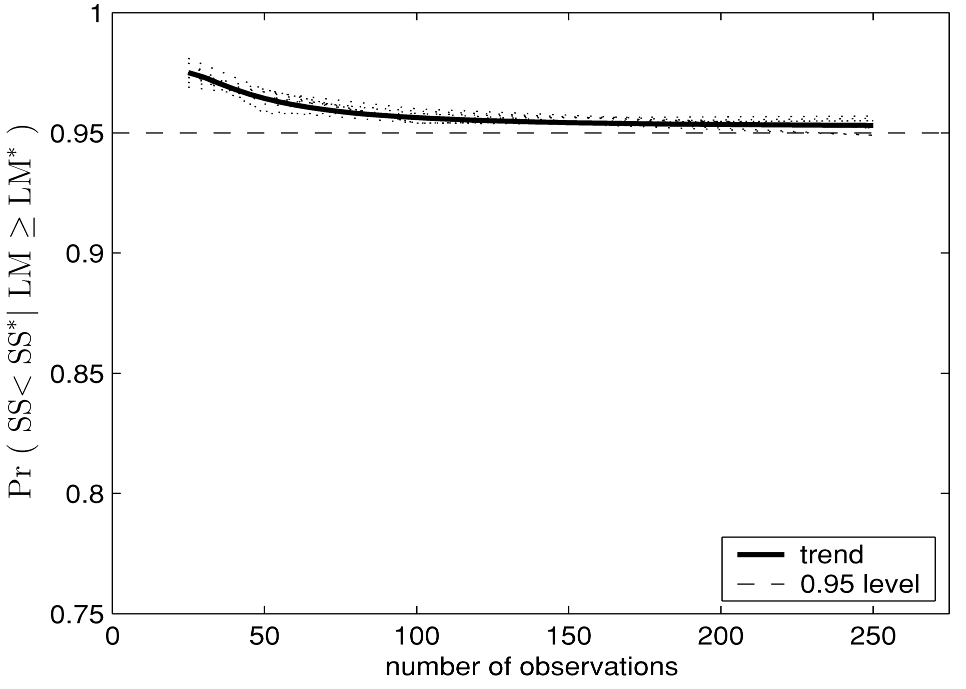

Figure 3. Nonscalar variance: independence of LM test and sensitivity.

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 scaled sensitivity, but the speed of convergence could be different. This is analyzed in Figures 4 and 5. 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 scaled sensitivity.

7. MISSPECIFICATION IN THE DISTRIBUTION

Our third and final example concerns the linear regression model

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

where the $\epsilon_1, \ldots, \epsilon_n$, conditional on 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

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

(see Kendall and Stuart 1976, p. 159). Note 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

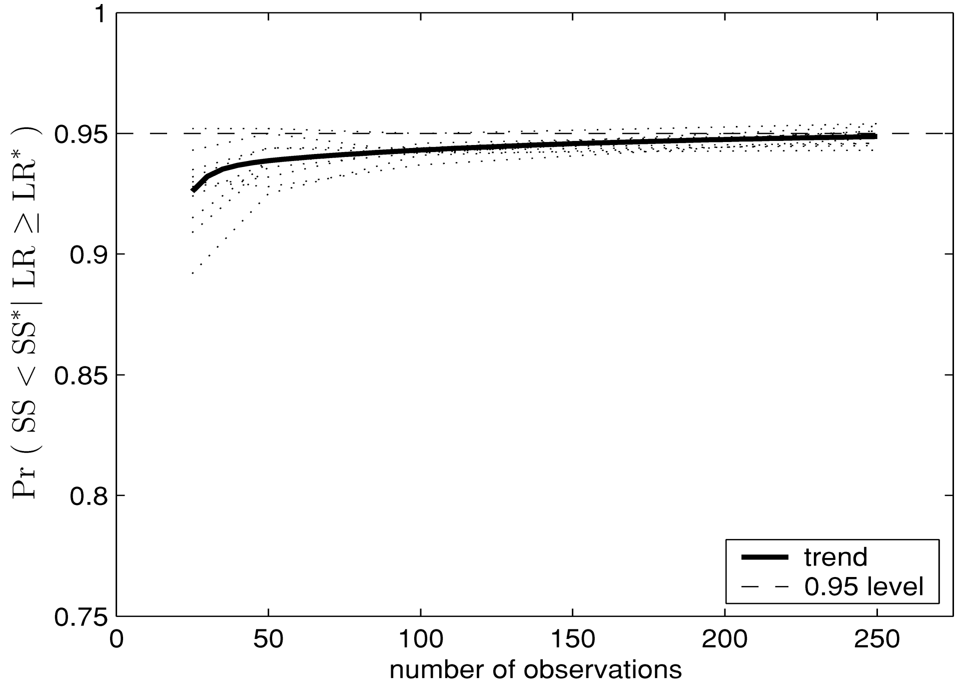

Figure 4. Nonscalar variance: independence of LR test and sensitivity.

and $\boldsymbol{\theta} = (\theta_1, \theta_2)$ as the nuisance parameter. At $\boldsymbol{\theta} = \mathbf{0}$ we obtain $\mathrm{d}\log f(\epsilon_i)/\mathrm{d}\epsilon_i = -\epsilon_i$ which defines the N(0, 1) distribution. We are interested in the sensitivity of $\boldsymbol{\beta}$ to $\boldsymbol{\theta}$. The log-likelihood conditional on X is given by

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

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

The score vector is given by

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

where

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

and the Hessian matrix is

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

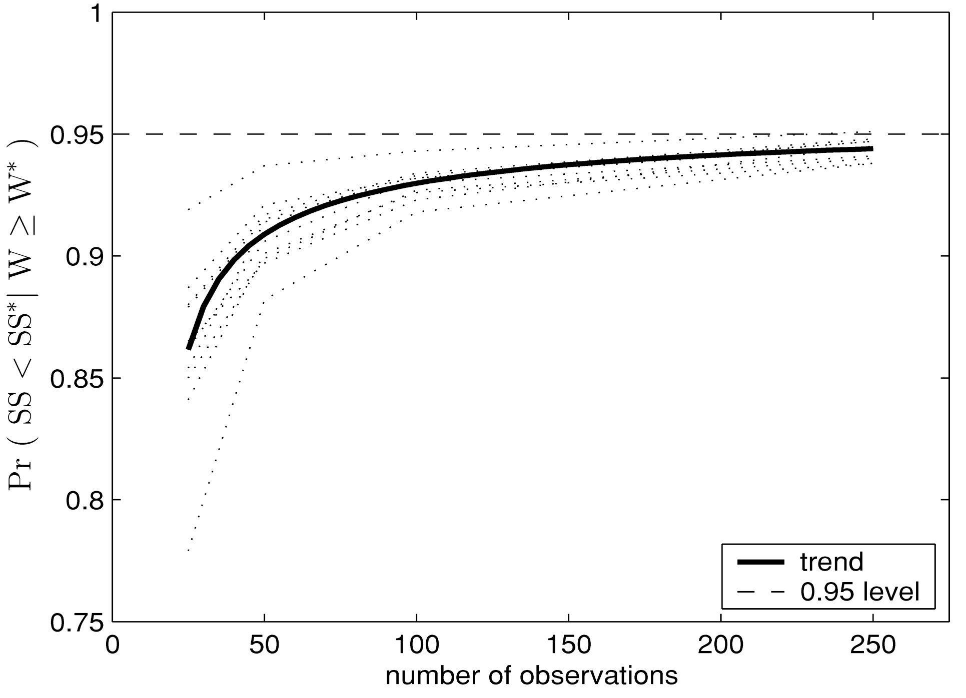

Figure 5. Nonscalar variance: independence of Wald test and sensitivity.

with

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

and

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

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 $\widetilde{\boldsymbol{\beta}}$, $\widetilde{\sigma}^2$ and $\widetilde{\boldsymbol{\theta}}$ are asymptotically independent. (In fact, since $-\mathrm{E}(\mathcal{H}^0_{\theta\theta}) = \mathrm{E}(\boldsymbol{q}^0_{\theta}\boldsymbol{q}^0_{\theta}) = \mathrm{diag}(2/3, 3/2)$, $\widetilde{\theta}_1$ and $\widetilde{\theta}_2$ are asymptotically independent as well.) Hence Theorem 5 implies that the LM test and the sensitivity $\boldsymbol{S}_{\widehat{\beta}}$ are independent if and only if the correlation between $\boldsymbol{q}^0_{\theta}$ and $\sqrt{n}\mathcal{H}^0_{\beta\theta}$ approaches zero. In general, this is not the case. We have

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

but

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

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

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 $\boldsymbol{\imath}'X(X'X)^{-1}X'\boldsymbol{\imath}/n = 1$ for $\boldsymbol{X} = \boldsymbol{\imath}$.

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 centred at zero. For example, if $\boldsymbol{x}$ is a sample from a distribution with mean μ and variance σ^2, then $\boldsymbol{\imath}'\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{\imath}/n \xrightarrow{p} \mu^2/(\mu^2+\sigma^2)$, which is zero if and only if $\mu/\sigma = 0$. If $\boldsymbol{x}$ is the trend $1, 2, \dots, n$, then $\boldsymbol{\imath}'\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{\imath}/n \to 3/4$.

The restricted estimators are

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

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

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

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

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

where

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

Instead of calculating the sensitivities of $\widehat{\boldsymbol{\beta}}$ 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

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

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

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

In Figure 6, we present the probability that the estimator $\widehat{\boldsymbol{\beta}}$ is not sensitive to non-normality ($\mathrm{SS} \leq \mathrm{SS}^*$), while the Jarque–Bera test rejects the null hypothesis of normality ($\mathrm{JB} > \mathrm{JB}^*$). The sensitivity is for the slope parameters only. Data sets 1–3 and 6–8 contain a constant term and

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

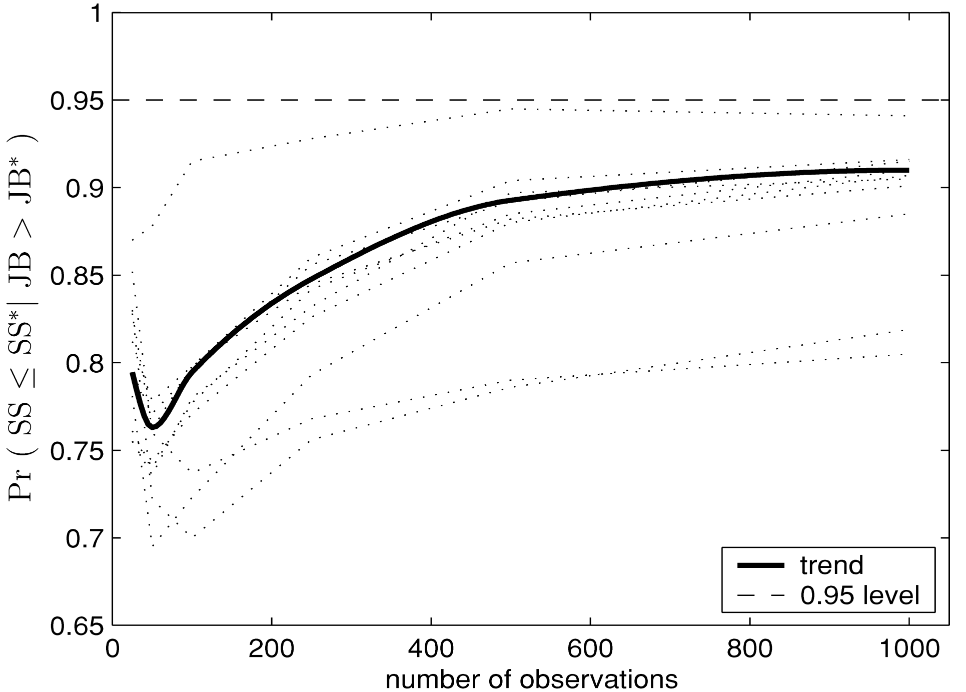

Figure 6. Non-normality: (in)dependence of JB-test and sensitivity.

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 centred at zero; hence the probability also converges to 95%. Finally, data set 9 contains two regressors that are centred at zero, and one (a sample from the lognormal distribution) which is not centred 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 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 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 $\sqrt{n}\mathrm{E}\left(\boldsymbol{q}^0_{\theta_1}\mathcal{H}^0_{\beta\theta_1}\right) \to \mathbf{0}$, the JB-test for kurtosis and the scaled sensitivity are asymptotically independent in this case, whether the regressors are measured in deviations or not. Figure 7 shows that the convergence to independence is rather slow however.

8. CONCLUSION

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 the sensitivity. Knowledge of one gives very little or no information on the other. Hence sensitivity analysis matters.

Suppose one assumes normality. One may test whether normality is rejected or not (diagnostic) or one may ask whether one step away from normality leads to 'large' changes in the estimates

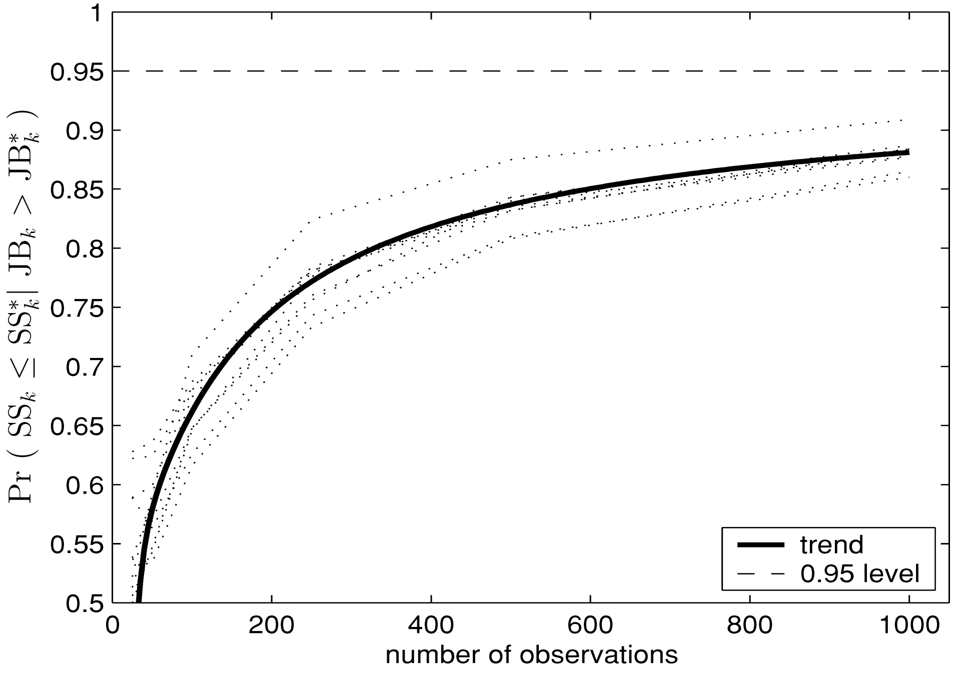

Figure 7. Kurtosis: independence of JB-test and sensitivity.

of the focus parameters (sensitivity). The paper shows that these two questions are essentially orthogonal. Hence the diagnostic may firmly reject the hypothesis of normality, in practice it may matter little.

The sensitivity (in a specific direction) can usually be calculated. But how does one decide whether the sensitivity is large or small? This is similar to asking whether the diagnostic is large or small. The latter question has a quantitative statistical answer in terms of 'statistical significance' but also a common-sense qualitative answer in terms of 'importance'. In the same manner, a sensitivity test can be based on the statistical significance of the scaled sensitivity (20) (the B1-statistic of Banerjee and Magnus (1999) in the case of variance misspecification), but also (and perhaps more importantly) on the qualitative importance.

The application of sensitivity analysis depends on the context. In some cases, it may be sufficient to know whether the sensitivity statistic itself is large or small. In other cases, the statistical properties of the sensitivity statistic should play a role. Sometimes the sign of the sensitivity can provide important information, for example if one distinguishes between directions of change in $\widehat{\boldsymbol{\beta}}$. Sometimes 'relative sensitivity' $S_{\widehat{\beta}_j}/\widehat{\beta}_j$ might be of interest, showing the percentage change in the estimator of β_j when the nuisance parameter increases by one unit. (Note that relative sensitivity is also asymptotically independent of the diagnostic tests.) Finally, sensitivity may be used to discriminate between directions: the higher the sensitivity, the more precise this direction should be estimated.

Under suitable regularity conditions, 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.

ACKNOWLEDGEMENTS

We are grateful to seminar participants at Tilburg University, the University of Exeter and the Econometric Society World Congress, London, August 2005 for their constructive comments. We are particularly grateful to Pavel Čížek, Jan Kiviet, two referees and the co-editor for their helpful and positive comments.

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,* 155–99.

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,* 157–76.

Bickel, P. J., C. A. J. Klaassen, Y. Ritov and J. A. Wellner (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. London: Johns Hopkins Press.

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

Chesher, A. (1983). The information matrix test: Simplified calculation via a score test interpretation. *Economics Letters 13,* 45–48.

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

Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society: Series B 48,* 133–69.

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

Davidson, R. and J. G. MacKinnon (1987). Implicit alternatives and the local power of test statistics. *Econometrica 55,* 1305–29.

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

Durbin, J. (1954). Errors in variables. *Review of the International Statistical Institute 22,* 23–32.

Fung, W. K. and M. K. Tang (1997). Assessment of local influence in multivariate regression analysis. *Communications in Statistics: Theory and Methods 26,* 821–37.

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

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: Cambridge University Press.

Hartless, G., J. G. Booth and R. C. Littell (2003). Local influence of predictors in multiple linear regression. *Technometrics 45,* 326–32.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica 46,* 1251–71.

Heijmans, R. D. H. and J. R. Magnus (1986a). Consistent maximum-likelihood estimation with dependent observations. *Journal of Econometrics 32,* 253–85. [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,* 169–88.

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

Holly, A. (1987). Specification tests: An overview. Chapter 2. In Bewley, T. F., (Ed.), *Advances in Econometrics*, Vol. 1, pp. 59–97. Cambridge: Cambridge University Press.

Huber, P. J. (2004). *Robust Statistics*. Hoboken, NJ: John Wiley.

Jarque, C. M. and A. K. Bera (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters 6,* 255–59.

Jarque, C. M. and A. K. Bera (1987). A test for normality of observations and regression residuals. *International Statistical Review 55,* 163–72.

Kendall, M. and A. Stuart (1976). *The Advanced Theory of Statistics*, Volume 1, 4th Edition. London: Charles Griffin & Co Ltd.

Kiefer, N. M. and G. R. Skoog (1984). Local asymptotic specification error analysis. *Econometrica 52,* 873–86.

Kim, J. and D. Pollard (1990). Cube root asymptotics. *Annals of Statistics 18,* 191–219.

Kiviet, J. F. and G. D. A. Phillips (1986). Testing strategies for model selection. *Applied Mathematics and Computations 20,* 237–69.

Kleijnen, J. P. C. (1997). Sensitivity analysis and related analyses: A review of some statistical techniques. *Journal of Statistical Computation and Simulation 57,* 111–42.

Lancaster, T. (1984). The covariance matrix of the information matrix test. *Econometrica 52,* 1051–54.

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

Leamer, E. E. (1984). Global sensitivity results for generalized least squares estimates. *Journal of the American Statistical Association 79,* 867–70.

Liu, S. (2000). On local influence for elliptical linear models. *Statistical Papers 41,* 211–24.

Liu, S. (2002). Local influence in multivariate elliptical linear regression models. *Linear Algebra and Its Applications 354,* 159–74.

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.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3,* 205–28.

Newey, W. K. and R. J. Smith (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica 72,* 219–55.

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,* 263–78.

Polasek, W. (1984). Regression diagnostics for general linear regression models. *Journal of the American Statistical Association 79,* 336–40.

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

Saltelli, A., K. Chan and E. M. Scott, Editors (2000). *Sensitivity Analysis*. Chichester: John Wiley.

Schwarzmann, B. (1991). A connection between local-influence analysis and residual diagnostics. *Technometrics 33,* 103–4.

Shi, L. and X. Wang (1999). Local influence in ridge regression. *Computational Statistics and Data Analysis 31,* 341–53.

Theil, H. (1957). Specification errors and the estimation of economic relationships. *Review of the International Statistical Institute 25,* 41–51.

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.