Simultaneous Choice of Design and Estimator in Nonlinear Regression with Parameterized Variance

Andrej Pázman[1] and Luc Pronzato[2]

[1] Comenius University, Bratislava pazman@center.fmph.uniba.sk
[2] Laboratoire I3S, CNRS-Université de Nice Sophia Antipolis
pronzato@i3s.unice.fr

Summary. In some nonlinear regression problems with parameterized variance both the design and the method of estimation have to be chosen. We compare asymptotically two methods of estimation: the penalized weighted LS (PWLS) estimator, which corresponds to maximum likelihood estimation (MLE) under the assumption of normal errors, and the two-stage LS (TSLS) estimator. We show that when the kurtosis κ of the distribution of the errors is zero, the asymptotic covariance matrix of the estimator is smaller for PWLS than for TSLS, which may not be the case when κ is not zero. We then suggest to construct two optimum designs, one for PWLS under the assumption $\kappa = 0$, the other for TSLS (with arbitrary κ), and compare their properties for different values of κ. All developments are made under the assumption of a randomized design, which allows rigorous proofs for the asymptotic properties of the estimators while avoiding the technical difficulties encountered in classical references such as Jennrich (1969) (finite tail product of the regression function and its derivatives, etc.).

Key words: Nonlinear regression, LS estimation, penalized LS, two-stage LS, iterative LS, robustness, kurtosis

1 Introduction

We consider a nonlinear regression problem, with observations

$$Y_k = \eta(x_k, \bar{\theta}) + \varepsilon_k \, , \ \mathrm{E}\{\varepsilon_k\} = 0 \, , \quad k = 1, \ldots, N \, , \tag{1}$$

where $\bar{\theta}$ denotes the unknown true value of the model parameters. The observation errors $\varepsilon_k = \varepsilon(x_k)$ are assumed to be independently distributed. It frequently happens that the full parameterized probability distribution of the errors ε_k is not available, whereas their variance is a known function of the design variable x and of (some of) the parameters θ of the mean response, that is,

$$\mathtt{E}\{\varepsilon_k^2\} = \lambda(x_k, \bar{\theta}) , \quad k = 1, \ldots, N . \tag{2}$$

The parameter estimation problem in this case is called method of fitting expectations in Jennrich and Ralston (1979), see also Maljutov (1988). The (ordinary) LS estimator, which ignores the information contained in the variance function, is strongly consistent and asymptotically normally distributed under standard assumptions. However, using the information on θ provided by the variance may yield a more precise estimation, hence the importance of choosing a suitable estimation method. We consider two approaches (Section 3), first a penalized weighted least-squares (PWLS) estimator, which corresponds to maximum likelihood estimation under the assumption of normal errors, second a two-stage least-squares (TSLS) estimator, where ordinary LS is used at the first stage, and the estimator is plugged in the variance function, to be used for weighted LS estimation at the second stage. The asymptotic properties of the estimators are obtained under the assumption of a *randomized design*, introduced in Section 2, which permits to maintain the proofs relatively simple although rigorous. Section 4 presents a design strategy based on the asymptotic properties derived in Section 3: we compute an optimum design for PWLS estimation under the assumption of a zero kurtosis for the errors ε_k, together with an optimum design for TSLS estimation; then we compare their performance (and that of the associated estimation methods) when the kurtosis varies. We show that in some particular situations the conclusion is design-free: one estimation method is always preferable, independently of the design, depending on the value of the kurtosis and the magnitude of the errors.

2 Randomized designs and uniform strong law of large numbers

Definition 1. *We call* randomized design *with measure ξ on the design space $\mathcal{X}$, $\int_{\mathcal{X}} \xi(dx) = 1$, a sequence $\{x_i\}$ of design points independently sampled from the mesure ξ on $\mathcal{X}$.*

The following assumptions will be used throughout the paper.
H1 Θ is a compact subset of $\mathbb{R}^d$, $\bar{\theta} \in \Theta$.
H2 $\eta(x,\theta)$ and $\lambda(x,\theta)$ are continuous functions of $\theta \in \Theta$ for any $x \in \mathcal{X}$, with $\eta(x,\theta)$ and $\lambda^{-1}(x,\theta)$ bounded on $\mathcal{X} \times \Theta$, $\lambda(x,\bar{\theta})$ bounded on $\mathcal{X}$.
H3 $\eta(x,\theta)$ and $\lambda(x,\theta)$ are two times continuously differentiable with respect to $\theta \in \mathrm{int}(\Theta)$ for any $x \in \mathcal{X}$, their derivatives are bounded on $\mathcal{X} \times \mathrm{int}(\Theta)$.

Our proofs are based on uniform convergence with respect to θ of the criterion function $J_N(\theta)$ defining the estimator $\hat{\theta}^N = \arg\min_\theta J_N(\theta)$. We shall thus need a uniform Strong Law of Large Numbers (SLLN). Note that the proper definition of the estimator as a random variable is ensured by Lemma

2 in Jennrich (1969) (see also Bierens (1994), p. 16). In the following $\hat{\theta}^N$ will refer to the measurable choice from $\arg\min_{\theta\in\Theta} J_N(\theta)$. The asymptotic results of the next sections are based on the following lemma, which is a simplified version of Theorem 2.7.1 in Bierens (1994).

Lemma 1 (Uniform SLLN). *Let $\{z_i\}$ be a sequence of i.i.d. random vectors, and $a(z,\theta)$ be a Borel measurable real function of $(z,\theta) \in \mathbb{R}^r \times \Theta$, continuous in θ for any z, with Θ a compact subset of $\mathbb{R}^p$. Suppose that*

$$\mathtt{E}[\sup_{\theta\in\Theta} |a(z,\theta)|] < \infty \,, \tag{3}$$

then $\mathtt{E}[a(z,\theta)]$ is continuous in $\theta \in \Theta$ and $\frac{1}{N}\sum_{i=1}^{N} a(z_i,\theta) \overset{\theta}{\rightrightarrows} \mathtt{E}[a(z,\theta)]$ a.s. when $N \to \infty$, where $\overset{\theta}{\rightrightarrows}$ means uniform convergence with respect to θ.

Once the almost sure uniform convergence of the criterion function $J_N(\cdot)$ is obtained, the almost sure convergence of the estimator will follow from the next lemma. The proof is a straightforward application of the continuity and uniform convergence properties.

Lemma 2 (Consistency from uniform convergence of the estimation criterion). *Assume that the sequence of functions $\{J_N(\theta)\}$ converges uniformly on Θ to the function $J(\theta)$, with $J_N(\theta)$ continuous with respect to $\theta \in \Theta$ for any N, Θ a compact set of $\mathbb{R}^p$, and $J(\theta)$ such that*

$$J(\bar{\theta}) = \min_{\theta\in\Theta} J(\theta) \text{ and } J(\theta) > J(\bar{\theta}) \;\forall\theta \neq \bar{\theta} \in \Theta \,.$$

Then $\lim_{N\to\infty} \hat{\theta}^N = \bar{\theta}$, where $\hat{\theta}^N \in \arg\min_{\theta\in\Theta} J_N(\theta)$. When the functions $J_N(\cdot)$ are random, and the uniform convergence to $J(\cdot)$ is almost sure, then convergence of $\hat{\theta}^N$ to $\bar{\theta}$ is also almost sure.

3 Penalized weighted LS and two-stage LS estimation

Since the optimum weights $w(x) = \sigma^{-2}(x,\bar{\theta}) = \lambda^{-1}(x,\bar{\theta})$ cannot be used for weighted LS estimation ($\bar{\theta}$ is unknown), it is tempting to use the weights $\lambda^{-1}(x,\theta)$, that is, to choose $\hat{\theta}^N$ that minimises the criterion

$$J_N(\theta) = \frac{1}{N}\sum_{k=1}^{N} \frac{[y(x_k) - \eta(x_k,\theta)]^2}{\lambda(x_k,\theta)} \,. \tag{4}$$

However, this approach in not recommended since $\hat{\theta}^N$ is generally not consistent. Indeed, using Lemma 1 with $z_k = (x_k,\varepsilon_k)$, $a(z_k,\theta) = [\eta(x_k,\bar{\theta}) + \varepsilon_k - \eta(x_k,\theta)]\lambda^{-1}(x_k,\theta)$ and Lemma 2 we easily obtain the following.

Theorem 1. *Let $\{x_i\}$ be a randomized design with measure ξ on $\mathcal{X} \subset \mathbb{R}^d$. Assume that H1 and H2 are satisfied. Then the estimator $\hat{\theta}^N_{LS}$ that minimises (4) in the model (1,2) converges a.s. to the set $\bar{\Theta}$ of values of θ that minimise*

$$J(\theta) = \int_{\mathcal{X}} \lambda(x,\bar{\theta})\lambda(x,\theta)^{-1}\,\xi(dx) + \int_{\mathcal{X}} \lambda(x,\theta)^{-1}[\eta(x,\theta)-\eta(x,\bar{\theta})]^2\,\xi(dx)\,.$$

Notice that, in general, $\bar{\theta} \notin \bar{\Theta}$.

3.1 Penalized weighted LS estimation

Consider now the following modification of the criterion (4),

$$J_N(\theta) = \frac{1}{N}\sum_{k=1}^{N}\frac{[y(x_k)-\eta(x_k,\theta)]^2}{\lambda(x_k,\theta)} + \frac{1}{N}\sum_{k=1}^{N}\log\lambda(x_k,\theta)\,. \tag{5}$$

$\hat{\theta}^N_{PWLS}$ that minimises (5) can be considered as a *penalized* weighted LS (PWLS) estimator, where the term $(1/N)\sum_{k=1}^N \log\lambda(x_k,\theta)$ penalizes large variances. It corresponds to maximum likelihood estimation under the assumption that the errors ε_k are normally distributed. It can be obtained numerically by direct minimization of (5) using a nonlinear optimisation method, or by solving an infinite sequence of weighted LS problems as suggested in Downing et al. (2001).

Next theorem shows that this estimator is strongly consistent and asymptotically normally distributed without the assumption of normal errors. The proof follows directly from Lemmas 1 and 2 and the Central Limit Theorem (CLT).

Theorem 2. *Let $\{x_i\}$ be a randomized design with measure ξ on $\mathcal{X} \subset \mathbb{R}^d$. Assume that H1 and H2 are satisfied and that for any θ, θ' in Θ,*

$$\left.\begin{array}{r}\int_{\mathcal{X}} \lambda^{-1}(x,\theta)[\eta(x,\theta)-\eta(x,\theta')]^2\,\xi(dx)=0\\ \int_{\mathcal{X}} |\lambda^{-1}(x,\theta)\lambda(x,\theta')-1|\,\xi(dx)=0\end{array}\right\} \Leftrightarrow \theta=\theta'\,. \tag{6}$$

Then the estimator $\hat{\theta}^N_{PWLS}$ that minimises (5) in the model (1,2) converges a.s. to $\bar{\theta}$. If, moreover, the errors ε_k have finite fourth-order moment $\mathtt{E}\{\varepsilon_i^4\}$ and $\mathtt{E}\{\varepsilon_i^3\}=0$ for all k, $\bar{\theta}\in\mathrm{int}(\Theta)$, H3 is satisfied and the matrix

$$\begin{aligned}\mathbf{M}_1(\xi,\bar{\theta}) = &\int_{\mathcal{X}} \lambda^{-1}(x,\bar{\theta})\frac{\partial\eta(x,\theta)}{\partial\theta}_{|\bar{\theta}}\frac{\partial\eta(x,\theta)}{\partial\theta^\top}_{|\bar{\theta}}\xi(dx)\\ &+\frac{1}{2}\int_{\mathcal{X}} \lambda^{-2}(x,\bar{\theta})\frac{\partial\lambda(x,\theta)}{\partial\theta}_{|\bar{\theta}}\frac{\partial\lambda(x,\theta)}{\partial\theta^\top}_{|\bar{\theta}}\xi(dx)\end{aligned} \tag{7}$$

is nonsingular, then $\hat{\theta}^N_{PWLS}$ satisfies

$$\sqrt{N}(\hat{\theta}^N_{PWLS}-\bar{\theta})\xrightarrow{\mathrm{d}}z \sim \mathcal{N}(0,\mathbf{M}_1^{-1}(\xi,\bar{\theta})\mathbf{M}_2(\xi,\bar{\theta})\mathbf{M}_1^{-1}(\xi,\bar{\theta}))$$

as $N \to \infty$*, with*

$$\mathbf{M}_2(\xi,\bar\theta) = \mathbf{M}_1(\xi,\bar\theta) + \frac{1}{4}\int_{\mathcal{X}} \lambda^{-2}(x,\bar\theta)\, \frac{\partial\lambda(x,\theta)}{\partial\theta}_{|\bar\theta}\, \frac{\partial\lambda(x,\theta)}{\partial\theta^\top}_{|\bar\theta}\, \kappa(x)\,\xi(dx) \quad (8)$$

where $\kappa(x) = \mathtt{E}\{\varepsilon^4(x)\}\lambda^{-2}(x,\bar\theta) - 3$ *is the kurtosis of the distribution of* $\varepsilon(x)$*.*

One may notice that when the errors ε_k are normally distributed, $\kappa(x) = 0$, $\mathbf{M}_2(\xi,\bar\theta) = \mathbf{M}_1(\xi,\bar\theta)$ and

$$\sqrt{N}(\hat\theta^N_{PWLS} - \bar\theta) \xrightarrow{\text{d}} z \sim \mathcal{N}(0, \mathbf{M}_1^{-1}(\xi,\bar\theta))\,,\ N \to \infty\,. \quad (9)$$

3.2 Two-stage LS estimation

By two stage LS, we mean using first some estimator $\hat\theta^N_1$, and then plugging the estimate into the weight function $\lambda(x,\theta)$. The second-stage estimator $\hat\theta^N_{TSLS}$ is then obtained by minimizing

$$J_N(\theta, \hat\theta^N_1) = \frac{1}{N}\sum_{k=1}^{N} \frac{[y(x_k) - \eta(x_k,\theta)]^2}{\lambda(x_k,\hat\theta^N_1)} \quad (10)$$

with respect to $\theta \in \Theta$. Again, using Lemmas 1 and 2 and the CLT we can show that $\hat\theta^N_{TSLS}$ is consistent when $\hat\theta^N_1$ converges (it does not need to be consistent, that is, convergence to $\bar\theta$ is not required), and, when $\hat\theta^N_1$ is $\sqrt{N}$-consistent, that is, when $\sqrt{N}(\hat\theta^N_1 - \bar\theta)$ is bounded in probability (that is, $\forall\epsilon > 0\ \exists A$ and N_0 such that $\forall N > N_0$, $\text{Prob}\{\sqrt{N}(\hat\theta^N_1 - \bar\theta) > A\} < \epsilon$), $\hat\theta^N_{TSLS}$ is asymptotically normally distributed.

Theorem 3. *Let* $\{x_i\}$ *be a randomized design with measure* ξ *on* $\mathcal{X} \subset \mathbb{R}^d$*. Assume that H1 and H2 are satisfied, that* $\hat\theta^N_1$ *converges to some* $\hat\theta \in \Theta$ *and that for any* θ, θ' *in* Θ*,*

$$\int_{\mathcal{X}} \lambda^{-1}(x,\hat\theta)[\eta(x,\theta) - \eta(x,\theta')]^2\,\xi(dx) = 0 \Leftrightarrow \theta = \theta'\,. \quad (11)$$

Then the estimator $\hat\theta^N_{TSLS}$ *that minimises (10) in the model (1,2) converges a.s. to* $\bar\theta$*. If, moreover, H3 is satisfied, the matrix*

$$\mathbf{M}(\xi,\bar\theta) = \int_{\mathcal{X}} \lambda^{-1}(x,\bar\theta)\, \frac{\partial\eta(x,\theta)}{\partial\theta}_{|\bar\theta}\, \frac{\partial\eta(x,\theta)}{\partial\theta^\top}_{|\bar\theta}\, \xi(dx) \quad (12)$$

is nonsingular and the first-stage estimator $\hat\theta^N_1$ *plugged in (10) is* $\sqrt{N}$*-consistent, with* $\bar\theta \in \text{int}(\Theta)$*, then* $\hat\theta^N_{TSLS}$ *satisfies*

$$\sqrt{N}(\hat\theta^N_{TSLS} - \bar\theta) \xrightarrow{\text{d}} z \sim \mathcal{N}(0, \mathbf{M}^{-1}(\xi,\bar\theta))\,,\ N \to \infty\,.$$

Note that a natural candidate for the first-stage estimator $\hat{\theta}_1^N$ is the ordinary LS estimator, which is $\sqrt{N}$-consistent under the assumptions of Theorem 3. Also note that $\mathbf{M}^{-1}(\xi,\bar{\theta})$ is the asymptotic covariance matrix of the weighted LS estimator in the case where *the variance function (2) is known.*

Increasing the number of stages leads to iteratively re-weighted LS estimation, which relies on sequence of estimators constructed as follows:

$$\hat{\theta}_k^N = \arg\min_{\theta\in\Theta} J_N(\theta,\hat{\theta}_{k-1}^N)\,,\;\; k=1,2\ldots \tag{13}$$

where $J_N(\theta,\theta')$ is defined by (10) and where $\hat{\theta}_1^N$ can be taken equal to the LS estimator. Using Theorem 3, a simple induction shows that, for any *fixed* k, $\hat{\theta}_k^N$ is strongly consistent and asymptotically normally distributed, $\sqrt{N}(\hat{\theta}_k^N - \bar{\theta})\stackrel{\rm d}{\to} z \sim \mathcal{N}(0,\mathbf{M}^{-1}(\xi,\bar{\theta}))$. We simply mention the following property, which relies on a classical result in fixed point theory, see Stoer and Bulirsch (1993) p. 267, and states that the recursion (13) converges a.s. for N large enough.

Theorem 4. *If the conditions of Theorem 3 are satisfied, the iteratively reweighted LS estimator defined by (13) in the model (1,2) converges a.s. for N large enough:* $\Pr\{\forall N > N_0\,,\; \lim_{k\to\infty}\hat{\theta}_k^N = \hat{\theta}_\infty^N\} \to 1$ *when* $N_0\to\infty$.

When the errors ε_k are normally distributed, the asymptotic covariance matrix of the two-stage LS estimator, see (12), is larger than that of the penalized WLS estimator, see (7). This advantage of $\hat{\theta}_{PWLS}^N$ over $\hat{\theta}_{TSLS}^N$ may disappear when the distribution of the errors has a positive kurtosis. In general, the conclusion depends on the design ξ, which raises the issue of choosing simultaneously the method of estimation and the design. This is discussed in the next section.

4 Choosing the design and the estimator

In the rest of the paper we assume that the kurtosis κ is constant (it does not depend on x). It corresponds to the rather common situation where the distributions of the errors at different x are similar and only differ by a scaling factor.

We start by a simple example where the choice between $\hat{\theta}_{PWLS}^N$ and $\hat{\theta}_{TSLS}^N$ does not depend on the design: depending on the kurtosis and the magnitude of the errors, one estimation method is uniformly better than the other.

Example 1. Suppose that in the model (1) the variance of the errors satisfies $\mathrm{Var}(\varepsilon_k)=\lambda(x_k,\theta)=\alpha[\eta(x_k,\theta)+\beta]^2$, for any k. In particular, when $\beta=0$ it corresponds to the situation where the relative precision of the observations is constant (and equal to $\alpha>0$). Direct calculations then gives for $\hat{\theta}_{PWLS}^N$

$$\mathbf{M}_1^{-1}(\xi,\bar{\theta})\mathbf{M}_2(\xi,\bar{\theta})\mathbf{M}_1^{-1}(\xi,\bar{\theta}) = \rho(\kappa,\alpha)\mathbf{M}^{-1}(\xi,\bar{\theta})\,,$$

where $\rho(\kappa,\alpha) = (1+2\alpha+\kappa\alpha)/(1+2\alpha)^2$, $\mathbf{M}_1(\xi,\bar{\theta})$ and $\mathbf{M}_2(\xi,\bar{\theta})$ are given by (7) and (8), and $\mathbf{M}^{-1}(\xi,\theta)$ is the asymptotic covariance matrix for $\hat{\theta}^N_{TSLS}$, see (12). Whatever the design ξ, TSLS should thus be preferred to PWLS when $\rho(\kappa,\alpha) > 1$, that is, when $\kappa > 2$ and $\alpha < \alpha^* = (\kappa-2)/4$, and vice-versa otherwise. Generally speaking, it means that for observations with constant relative precision the two-stage procedure is always preferable when $\kappa > 2$, provided that the errors are small enough. For instance, when the errors have the exponential distribution $\varphi(\varepsilon) = (1/2)\exp(-|\varepsilon|)$, $\kappa = 3$ and the limiting value for α is $\alpha^* = 1/4$. Figure 1 gives the evolution of $\rho(\kappa,\alpha)$ as a function of α for $\kappa = 3$.

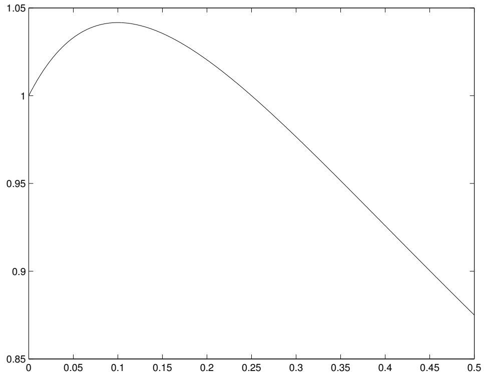

Fig. 1. Evolution of $\rho(\kappa,\alpha)$ as a function of α in Example 1: $\kappa = 3$ and $\hat{\theta}^N_{TSLS}$ should be preferred to $\hat{\theta}^N_{PWLS}$ for $\alpha < 1/4$

In more general situations the estimator and the design must be chosen simultaneously. The following approach may be used: (i) determine the optimum designs ξ^*_{PWLS} for the PWLS estimator *under the assumption of zero kurtosis* and ξ^*_{TSLS} for the TSLS estimator; (ii) compare the values of the design criteria for both estimators at different values of the kurtosis κ. Note that the asymptotic covariance matrix of $\hat{\theta}_{PWLS}$ is linear in κ. Therefore, for any design criterion $\phi(\cdot)$ such that $\phi(\mathbf{M})$ is monotonic in $\mathbf{M}$, a value κ^* exists such that $(\xi^*_{TSLS}, \hat{\theta}_{TSLS})$ should be preferred to $(\xi^*_{PWLS}, \hat{\theta}_{PWLS})$ for $\kappa > \kappa^*$.

The optimum design for TSLS estimation corresponds to a standard design problem with homoscedastic errors where the derivative $\partial\eta(x,\theta)/\partial\theta$ is replaced by $(1/\sqrt{\lambda(x,\theta)})\partial\eta(x,\theta)/\partial\theta$, see (12). In the case of PWLS with $\kappa = 0$, the information matrix to be used is $\mathbf{M}_1(\xi,\theta)$, see (9). Optimum designs for such situations can be constructed with the same algorithms as for the standard case, see Downing et al. (2001). In particular, when $\lambda(x,\theta) = g[\eta(x,\theta)]$, where $g(\cdot)$ is a differentiable function from $\mathbb{R}$ into $\mathbb{R}^+$ with derivative $g'(\cdot)$, $\mathbf{M}_1(\xi,\theta) = \int_{\mathcal{X}} G(x,\theta)[\partial\eta(x,\theta)/\partial\theta][\partial\eta(x,\theta)/\partial\theta^\top]\xi(dx)$ with $G(x,\theta) = \lambda^{-1}(x,\theta) + \{g'[\eta(x,\theta)]\}^2\lambda^{-2}(x,\theta)/2$, and standard approaches for optimum design with homoscedastic errors can be used.

Example 2. We take $\eta(x,\theta) = \theta_0 + \theta_1 x + \theta_2 x^2$, $\lambda(x,\theta) = 5/\eta(x,\theta)^2$, $x \in \mathcal{X} = [-1,1]$. Although this is a linear regression model, heteroscedasticity implies that numerical values must be specified for the parameters θ to design an optimum experiment. We take $\theta^\top = (\theta_0,\theta_1,\theta_2) = (2,1,1/2)$. The design criterion is D-optimality. The D-optimum design ξ^*_{TSLS} for the TSLS estimator has three support points $-1, 0.256, 1$ that receive equal weight $1/3$. The D-optimum design ξ^*_{PWLS} for the PWLS estimator with normal errors is supported at $-1, 0.078, 1$, with weight $1/3$ at each point. Figure 2 shows that TSLS estimation with ξ^*_{TSLS} should be preferred to PWLS estimation with ξ^*_{PWLS} when $\kappa > \kappa^* \simeq 4.09$.

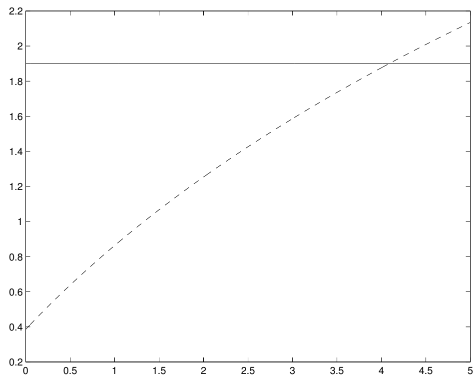

Fig. 2. Evolution of $\log\det[\mathbf{M}_1^{-1}(\xi^*_{PWLS},\theta)\mathbf{M}_2(\xi^*_{PWLS},\theta)\mathbf{M}_1^{-1}(\xi^*_{PWLS},\theta)]$ as a function of κ (dashed line) and $\log\det\mathbf{M}^{-1}(\xi^*_{TSLS},\theta)$ (full line) in Example 2

References

H.J. Bierens. *Topics in Advanced Econometrics.* Cambridge University Press, Cambridge, 1994.

D. Downing, V.V. Fedorov, and S. Leonov. Extracting information from the variance function: optimal design. In A.C. Atkinson, P. Hackl, and W.G. Müller, editors, *mODa6 – Advances in Model–Oriented Design and Analysis*, pages 45–52. Physica Verlag, Heidelberg, 2001.

R.I. Jennrich. Asymptotic properties of nonlinear least squares estimation. *Annals of Math. Stat.*, 40:633–643, 1969.

R.I. Jennrich and M.L. Ralston. Fitting nonlinear models to data. *Annals Rev. Biophys. Bioeng.*, 8:195–238, 1979.

M.B. Maljutov. Design and analysis in generalized regression model F. In V.V. Fedorov and H. Laüter, editors, *Model-Oriented Data Analysis, Proceedings 1st IIASA Workshop, Eisenach (GDR), March 1987*, pages 72–76. Springer, Berlin, 1988.

J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis (2nd Edition).* Springer, Heidelberg, 1993.