A Nonparametric Simulated Maximum Likelihood Estimation Method*

Jean-David Fermanian† Bernard Salanié‡

May 23, 2001

Abstract

Existing simulation-based estimation methods are either general-purpose but asymptotically inefficient or asymptotically efficient but only suitable for restricted classes of models. This paper presents a new simulated maximum-likelihood method that rests on estimating the likelihood nonparametrically on a simulated sample. We prove that this method, which can be used on very general models, is consistent and asymptotically efficient. We also give some preliminary Monte-carlo simulation results on the stochastic volatility model.

Introduction

Many parametric estimation procedures in econometrics are based on the maximization of a criterion function. This may be the mean square error as for the least squares method, the likelihood function for maximum likelihood estimation, or the likelihood of a well-chosen pseudo-model for pseudo-maximum likelihood methods. Unfortunately, the criterion function sometimes does not have a closed-form expression. This is true, for instance, of limited-dependent variable models with lagged dependent variables, where the likelihod function and other competing criterion functions can only be written as integrals of large dimension (equal to the number of observations). Simulation-based estimation methods were devised precisely to circumvent this problem[1]. By replacing untractable expectations with their Monte-Carlo counterparts, they allow the relevant criterion functions to be computed, which has made it possible for econometricians to estimate new classes of models.

*We thank Christian Gouriéroux, Guy Laroque, Nour Meddahi, Alain Monfort, Eric Renault and Christian Robert for their comments. Remaining errors and imperfections are ours. This paper was written while Bernard Salanié was visiting the University of Chicago, which he thanks for its hospitality.

†ENSAE and CREST.

‡CREST, CNRS URA 2200 and CEPR.

[1] Early references are Lerman-Manski (1981), Pakes (1986), Laroque-Salanié (1989), McFadden (1989) and Pakes-Pollard (1989).

Simulation-based estimation methods belong to two general classes[2]. The first one consists of methods that are reasonably general-purpose but are not efficient asymptotically, even when the number of simulation draws is allowed to increase fast enough. The method of simulated moments (McFadden (1989), Pakes-Pollard (1989)) and the simulated pseudo-maximum likelihood methods (Laroque-Salanié (1989, 1993, 1994)) both belong to this class. As they rely on simulating the obvious mathematical expectation with its Monte-Carlo counterpart, they can be applied to a large class of models. However, they simulate criterion functions that (even with an infinite number of simulations) do not lead to efficient estimators. The indirect inference methods (Gouriéroux-Monfort-Renault (1993) and Smith (1993)) also belong to that first category. The second class of simulation-based estimation methods relies on simulating the likelihood function, so that the resulting estimators are asymptotically efficient (again, with an infinite number of simulations). The simulated likelihood methods (see e.g. Lee (1995)) and the method of simulated scores (Hajivassiliou-McFadden (1998)) are examples of such estimation methods. The diffciulty with these methods is that as the likelihood function usually cannot be written as a function of mathematical expectations, they can only be applied to restrictive classes of models. Thus there has been a lot of emphasis on the literature on dynamic LDV models, but the methods that have been proposed only apply to models defined by linear constraints, for which several classes of efficient simulators have been devised (see, e.g., Börsch-Supan and Hajivassiliou (1993)). To the best of our knowledge, there exists no currently available method that is both asymptotically efficient and applicable to a very wide class of econometric models[3].

The purpose of this paper is to present such a simulation-based estimation method, which we call the NonParametric Simulated Maximum Likelihood method, or NPSML for short. Start from a fully parametric model whose reduced form can be simulated (which is a very mild requirement). Then our method consists in approximating the unknown likelihood function with a kernel-based nonparametric estimator based on simulations of the endogenous variables of the model. Since this strategy is applicable to a very wide class of models, it provides a quasi-universal simulator. Moreover, we prove in this paper that in static models, it provides consistent, asymptotically normal and asymptotically efficient estimators when the number of simulations goes to infinity and the bandwidth of the kernel estimator goes to zero. We then argue that the method can be extended to dynamic models and explain how to do so.

Section 1 presents the basic idea of the NPSML estimation method, using a static (but very general) model as an application. It states our consistency and asymptotic efficiency theorems, which are proved in the appendix. Section 2 discusses the assumptions of these theorems. In section 3, we show how the NPSML method can be extended to fully

[2] Gouriéroux-Monfort (1996) survey the available methods. Hajivassiliou-Ruud (1994) focusses on limited-dependent variable models, while Stern (1997) concentrates on empirical applications.

[3] Although the efficient method of moments of Gallant-Tauchen (1996) comes close.

dynamic models.

1 NPSML for Static Models

Simulation-based methods are clearly most useful in dynamic settings. Nevertheless, it seems simpler to introduce the NPSML method and its asymptotic properties on a static model. Therefore, consider a model with reduced form

$$y = g(x, \theta, \varepsilon), \tag{1-1}$$

where

- θ, the parameter of interest, belongs to a compact set $\Theta \subset \mathbb{R}^q$,
- the observed endogenous variable y is a vector of $\mathbb{R}^m$,
- the exogenous variable x belongs to $\mathbb{R}^d$,
- $\varepsilon \in \mathbb{R}^e$ represents the disturbances.

We assume that both the function g and the distribution of the disturbances ε are fully known[4]. Thus this is a fully parametric model—only the estimation technique has a nonparametric element.

Let $(x_t, y_t)_{t=1,\ldots,T}$ be an i.i.d. sample. The associated loglikelihood then is

$$L_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \ln l_t(\theta),$$

denoting $l_t(\theta)$ the density of y_t knowing (x_t, θ). We assume

Assumption L1 : the maximum likelihood estimator $\tilde{\theta}_T$ is consistent, asymptotically normal and asymptotically efficient. The true parameter θ_0 belongs to the interior of Θ. More precisely, we assume that

$$-\frac{\partial^2 L_T}{\partial\theta\partial\theta'}(\theta^*) \xrightarrow[T\to\infty]{P} \Omega, \tag{1-2}$$

uniformly with respect to θ^* in a neighborhood of θ_0, and that

$$T^{1/2} \frac{\partial L_T}{\partial\theta}(\theta_0) \xrightarrow[T\to\infty]{D} \mathcal{N}(0, \Omega). \tag{1-3}$$

For the class of models we are interested in, the likelihood function $l_t(\theta)$ cannot be computed in a closed form, so that it is impossible to compute the maximum likelihood estimator $\tilde{\theta}_T$. We propose instead to approximate each term $l_t(\theta)$ by a kernel estimator based on some i.i.d. simulated sample $(\varepsilon_t^s)_{s=1,\ldots,S}$ drawn from the distribution[5] of ε.

[4] As usual, unknown parameters of the distribution of ε are integrated to θ. Moreover, lagged values of the observed endogenous variable can be subsumed within x in the usual manner. We will introduce lagged latent variables later in this paper.

[5] The (ε_t^s) can also be the same for each t; the proofs of asymptotic results go through in that case.

Thus, denoting $y_t^s(\theta) = g(x_t, \theta, \varepsilon_t^s)$, the likelihood $l_t(\theta)$ is estimated by

$$l^S(y_t|x_t, \theta) \equiv l_t^S(\theta) \equiv \frac{1}{Sh^m} \sum_{s=1}^{S} K\left(\frac{y_t - y_t^s(\theta)}{h}\right) \tag{1-4}$$

Here, h is a bandwidth such that $h \longrightarrow 0$ when $S \longrightarrow \infty$, and K is a kernel. Under technical conditions that are stated below, $l_t^S(\theta)$ converges to $l_t(\theta)$ when the number of simulations S goes to infinity. Thus a natural idea consists in defining the NPSML estimator as the global maximizer of

$$\tilde{L}_T^S(\theta) = \frac{1}{T} \sum_{t=1}^{T} \ln l_t^S(\theta) \tag{1-5}$$

on Θ.

For technical reasons, it is in fact necessary to trim the smallest values of l_t^S. This can be done by considering the nonparametric simulated loglikelihood

$$\tilde{L}_T^S(\theta) = \frac{1}{T} \sum_{t=1}^{T} \tau_S(l_t^S(\theta)) \ln l_t^S(\theta), \tag{1-6}$$

where τ_S is a sufficiently regular function[6] such that $\tau_S(x) = 0$ if $|x| < h^\delta$ and $\tau_S(x) = 1$ if $|x| > 2h^\delta$, with $\delta > 0$.

Thus we define the NPSML estimator by

$$\hat{\theta}_T^S = \arg\max_{\theta \in \Theta} \tilde{L}_T^S(\theta)$$

We now state a set of assumptions under which it is strongly consistent when T and S go to infinity and the bandwidth h goes to zero.

The first subset of assumptions concern the kernel. We assume

Assumption K: the kernel K is twice continuously differentiable and has compact support.

Let ρ be the order of the kernel, i.e. $\int x_1^{\alpha_1} \dots x_m^{\alpha_m} K(x)\, dx$ is zero if $0 < \sum_{j=1}^m \alpha_j < \rho$ and nonzero if $\sum_j \alpha_j \in \{0, \rho\}$. (Classically, $\rho = 2$ for positive symmetrical kernels).

We need more assumptions on the exact likelihood function. In addition to Assumption L1, we assume

Assumption L2: $l(y|x, \theta)$ is bounded above on $\mathbb{R}^d \times \mathbb{R}^m \times \Theta$.

[6] We consider in the paper the continuously differentiable function defined by

$$\tau_S(x) = 4(x - h^\delta)^3/h^{3\delta} - 3(x - h^\delta)^4/h^{4\delta}$$

when $x \in [h^\delta, 2h^\delta]$. Therefore, this function τ_S is piecewise polynomial and $||\tau_S'||_\infty = O(h^{-\delta})$.

Assumption L3: there exist some $\beta > 1$ and some constant C_0 such that a.e.

$$\sup_{T,\theta} \frac{1}{T} \sum_{t=1}^{T} |\ln l_t(\theta)|^{\beta} \leq C_0$$

Assumption L4:

$$\sup_{\theta} \left\| \frac{\partial l(Y|X,\theta)}{\partial \theta} \right\| \text{ belongs to } L^1.$$

Assumption L5: $\partial^{\rho} l(y|x,\theta)/\partial y^{\rho}$ is bounded above on $\mathbb{R}^d \times \mathbb{R}^m \times \Theta$.
We need a few technical conditions:

Assumption T1: there exists $\kappa > 0$ such that $S \leq T^{\kappa}$.
Assumption T2: there exists $\pi > 0$ such that $h \geq S^{-\pi}$.
Assumption T3: there exists ν such that

$$\ln h.P(\|X,Y\| > S^{\nu}) \underset{S\to\infty}{\longrightarrow} 0.$$

We also need one assumption about the reduced form of the model.

Assumption M1: there exist a function ϕ and some $s_0 \geq 0$ such that

$$h^{s_0} \sup_{\theta, \|x\| \leq S^{\nu}} \left\{ \left\| \frac{\partial g(x,\theta,\varepsilon)}{\partial \theta} \right\| + \left\| \frac{\partial g(x,\theta,\varepsilon)}{\partial x} \right\| \right\} \leq \phi(\varepsilon),$$

with $E[\phi(\varepsilon)] < \infty$, and where ν was introduced in Assumption T3.

Note that if $\partial g/\partial\theta$ and $\partial g/\partial x$ are bounded in norm, then it suffices to take $s_0 = 0$ in Assumption M1. Otherwise $s_0 > 0$ is necessary because the supremum is taken over a compact set that increases in size like S^{ν}.

We also need to put some assumptions on the order of the trimming function and the rate of convergence of the bandwidth to zero as the number of simulation draws goes to infinity.

Assumption R1: $\delta < \rho$.
Assumption R2: $Sh^{m+2\delta}/\ln S$ goes to infinity as S goes to infinity.
Assumption R3: there exists $p > 2$ such that

$$\sum_{S\geq 1} \left(\frac{\ln S}{S}\right)^{p/2-1} h^{-mp/2} < +\infty. \tag{1-7}$$

We can finally state our consistency theorem.

Theorem 1.1 *Under assumptions K, M1, L1-L5, R1-R3 and T1-T3, $\hat{\theta}_T^S$ is strongly consistent: almost everywhere,*

$$\hat{\theta}_T^S \underset{S,T\to\infty}{\longrightarrow} \theta_0.$$

It is easy to check that $\hat{\theta}_T^S$ is weakly consistent under the same assumptions except R3.

In order to prove the asymptotic normality and efficiency of the NPSML estimator, we need a few more assumptions. To state them, let V_0 denote a neighbourhood of θ_0. We first need to strengthen our assumptions on the reduced form of the model by adding

Assumption M2: for some $r_0 \geq 0$ and some $p_0 > 4$,

$$h^{r_0} \sup_{\theta \in V_0, \|x\| \leq S^\nu} \left\| \frac{\partial g(x, \theta, \varepsilon)}{\partial \theta} \right\| \leq \bar{\phi}(\varepsilon), \tag{1-8}$$

where $E[\bar{\phi}(\varepsilon)^{p_0}] < \infty$.

Assumption M3: there exists a function $\bar{\psi}$ and $s_1 \geq 0$ such that, for every $\varepsilon > 0$,

$$h^{s_1} \sup_{\theta \in V_0, \|x\| \leq S^\nu} \left\{ \left\| \frac{\partial^2 g(x, \theta, \varepsilon)}{\partial^2 \theta} \right\| + \left\| \frac{\partial^2 g(x, \theta, \varepsilon)}{\partial x \partial \theta} \right\| + \left\| \frac{\partial g(x, \theta, \varepsilon)}{\partial \theta} \right\|^2 \right.$$
$$\left. + \quad \left\| \frac{\partial g(x, \theta, \varepsilon)}{\partial x} \right\| \cdot \left\| \frac{\partial g(x, \theta, \varepsilon)}{\partial \theta} \right\| \right\} \leq \bar{\psi}(\varepsilon),$$

where $E[\bar{\psi}(\varepsilon)] < \infty$.

Again, if the derivatives of g are bounded, then one can take $r_0 = s_1 = 0$ (see the remark after Assumption M1).

We also need three more assumptions on the exact likelihood function:

Assumption L6:

$$\left\| \frac{\partial l(y|x, \theta)}{\partial \theta} \right\| \text{is bounded above on } \mathbb{R}^d \times \mathbb{R}^m \times V_0.$$

Assumption L7: there exists $\gamma > 1$ such that

$$\sup_{\theta \in V_0} T^{-1} \sum_{t=1}^{T} \left\| \frac{\partial \ln l_t(\theta)}{\partial \theta} \right\|^\gamma = O_P(1)$$

Assumption L8:

$$\frac{\partial^{\rho+1} l(y|x, \theta)}{\partial \theta \partial y^\rho} \text{is bounded on } \mathbb{R}^d \times \mathbb{R}^m \times V_0.$$

The assumptions on the rates of convergence also have to be strengthened.

Assumption R4:

$$T^{1/2} h^{\rho - \delta} \ln h \underset{S,T \to \infty}{\longrightarrow} 0$$

Assumption R5:

$$Th^{-2m-2\delta-2-2r_0}\ln^2 h \ln S/S \underset{S,T\to\infty}{\longrightarrow} 0$$

Assumption R6:

$$\left[T^{1/2}h^{-\delta}|\ln h| \sup_{\{(x,y,\theta)\in A_h\}}\left\|\frac{\partial l(y|x,\theta)}{\partial\theta}\right\| + T^{\gamma/2(\gamma-1)}\right] P_{\theta_0}(\inf_{\theta\in V_0} l_t(\theta)\le 2h^\delta) \underset{S,T\to\infty}{\longrightarrow} 0,$$

where $A_h = \{(x,y,\theta)|l(y|x,\theta)\in[h^\delta, 2h^\delta], \theta\in V_0\}$.

Finally, we need to add one technical condition.

Assumption T4:

$$\left[T^{1/2}h^{-\delta-m-1}|\ln h| + T^{\gamma/2(\gamma-1)}\right] P_{\theta_0}(\|x_t, y_t\| > S^\nu) \underset{S,T\to\infty}{\longrightarrow} 0$$

Recall that we denote Ω the asymptotic variance-covariance matrix of the exact maximum likelihood estimator. We can finally state our asymptotic efficiency theorem.

Theorem 1.2 *Under assumptions K, M1-M3, L1-L8, R1-R6 and T1-T4, $\hat\theta_T^S$ is asymptotically normal and asymptotically efficient:*

$$\sqrt{T}(\hat\theta_T^S - \theta_0) \underset{S,T\to\infty}{\overset{law}{\longrightarrow}} \mathcal{N}(0,\Omega). \tag{1-9}$$

To simplify the proof, we have used assumptions that are more restrictive than need be. For instance, Assumption L6 could be replaced with a condition on the moments of the derivative of the likelihood function. Also, the assumptions used imply that $\hat\theta_T^S$ is strongly consistent, whereas convergence in probability would suffice.

When this theorem applies, $\hat\theta_T^S$ has the same asymptotic variance as $\tilde\theta_T$, the exact maximum likelihood estimator of θ_0. Thus $\hat\theta_T^S$ is asymptotically efficient and Ω can be estimated by

$$\hat\Omega = \frac{1}{T}\sum_{t=1}^{T}\tau_S(l_t^S(\hat\theta_T^S))\cdot\left(\frac{\partial\ln l_t^S}{\partial\theta}(\hat\theta_T^S)\right)\cdot\left(\frac{\partial\ln l_t^S}{\partial\theta}(\hat\theta_T^S)\right)'. \tag{1-10}$$

2 Choice of the Parameters

To apply our estimation method in practice, it is necessary to fix the values of the parameters ρ, δ, K, h, ν and S. The main difficulty consists in choosing the rates of convergence of the number of simulations S to infinity and of the bandwidth h to zero so that the assumptions of Theorems 1.1 and 1.2 hold. Let us therefore assume that $S = K_1T^a$ and $h = K_2S^{-b}$ for some positive constants K_1 and K_2.

First examine the assumptions that imply that the NPSML estimator is consistent. We neglect assumptions T1 to T3, which are bound to hold if the density of (x, y) is not too thick-tailed[7]. Then the relevant assumptions are R1 to R3. These translate into

$$\delta < \rho \ , \ b < \frac{1}{m + 2\delta} \ ,$$

and

$$b < \frac{p-4}{mp} \text{ for some } p > 2.$$

Clearly, this last condition holds as soon as $b < 1/m$, which is implied by the second condition. Therefore it suffices to choose $\delta < \rho$ and $b < 1/(m+2\delta)$. In particular, take the usual case of a second-order kernel ($\rho = 2$). Then we can choose $\delta = 1$ for instance and the asymptotically optimal bandwidth selector for kernel density estimation, for which $b = 1/(m+4)$, fits the bill. Thus the most natural choice for the rate of convergence of h to zero yields a consistent NPSML estimator. Note that as expected, the speed of convergence of S to infinity is irrelevant for consistency, viz we only require that $a > 0$.

Now consider the assumptions for asymptotic normality. In addition to R1 to R3, R4 to R6 must also hold. R4 and R5 translate into

$$ab > \frac{1}{2(\rho - \delta)} \text{ and } 2ab(m + \delta + 1 + r_0) + 1 < a \qquad \text{(2-11)}$$

Now choose $\rho > \delta$ and some $K > \frac{1}{2(\rho-\delta)}$. Moving on the hyperbola $ab = K$ to the zone of large a and small b will satisfy all assumptions. Thus our conditions define a nondegenerate region of the (a, b) plane. One would hope that this region intersects the line $b = 1/(m+2\rho)$ that defines the usual asymptotically optimal bandwidth. It can be checked that such is the case if $(\rho - \delta)$ is large enough, which may imply using higher-order kernels ($\rho > 2$).

Assumption R6 is more problematic, as it should be checked on a case-by-case basis. Clearly, it holds when h goes to zero fast enough. Take for instance the simplest case, in which y is normally distributed with mean θ and unit variance. Then tedious calculations show that R6 holds if

$$T^{\frac{\gamma}{2(\gamma-1)}} \frac{h^{\delta}}{\sqrt{|\ln h|}} \longrightarrow 0,$$

which is satisfied for $ab > \gamma/2\delta(\gamma - 1)$. Unfortunately, it seems very difficult to find a more general sufficient condition for R6.

3 NPSML for the dynamic case

As mentioned before, models in which x contains lagged observable endogenous variables can be treated in exactly the same (provide that ε is not serially correlated). However, there is a class of models for which dynamic simulations are called for. This includes

[7]It should tend to zero more quickly than $||(x, y)||^{-k}$, for some $k \geq 0$, when $||(x, y)||$ tends to infinity.

- models with both lagged observable endogenous variables and serially correlated disturbances
- models with lagged latent variables.

An example of the latter is the stochastic volatility model, which can be written (in its simplest form)

$$\begin{cases} y_t = \exp(y_t^*/2)\varepsilon_{1t} \\ y_t^* = a + by_{t-1}^* + \sigma\varepsilon_{2t}, \end{cases}$$

where y_t represents the observed returns and y_t^* the latent volatility.

In these models, the likelihood function for observation t is a t-dimensional integral, which can very rarely be computed in closed-form. For very simple instances of these models, it is possible to apply clever tricks to use simulated maximum-likelihood or the method of simulated scores, but there exists as yet no fully general method. As we shall now see, it is easy to extend the NPSML method to these models in order to obtain consistent and asymptotically normal estimator, with some loss in asymptotic efficiency.

For such models, we rewrite the reduced form as

$$z_t = g(x_t, z_{t-1}, \theta_0, u_t),$$

where the vector of endogenous variables z_t may contain

- observable endogenous variables y_t
- latent endogenous variables y_t^*
- disturbances ε_t.

Now u_t represents the innovations in the disturbances and is still assumed to be drawn from a known distribution $\mathcal{L}$. For instance, we might have

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \sigma u_t$$

where both ρ_1 and σ are parameters to be estimated.

We now resort to dynamic simulations i.e., given an initializing scheme for $z_0^s(\theta)$, we compute for $s = 1, \dots, S$ and $t = 1 \dots, T$

$$z_t^s(\theta) = g(x_t, z_{t-1}^s(\theta), \theta, u_t^s),$$

where the u_t^s are drawn from $\mathcal{L}$.

As this may seem abstract, consider the stochastic volatility model as presented above. Denote $\theta = (a, b, \sigma)$. For each t, draw $(\varepsilon_{1t}^s, \varepsilon_{2t}^s)$ in the assumed distribution of $(\varepsilon_{1t}, \varepsilon_{2t})$ for $s = 1 \dots, S$. Also draw $y_0^{*s}(\theta)$ from the stationary distribution of the volatility process implied by θ. Then compute recursively

$$\begin{cases} y_t^s(\theta) = \exp(y_t^{*s}(\theta)/2)\varepsilon_{1t}^s \\ y_t^{*s}(\theta) = a + by_{t-1}^{*s}(\theta) + \sigma\varepsilon_{2t}^s, \end{cases}$$

Given the simulated paths of the observable endogenous variables $y_t^s(\theta)$, we could proceed exactly as for the static model. However, this will only approximate the marginals $l(y_t|x_t, \theta)$ of the likelihood function

$l(y_1, \ldots, y_T)$, and thus it will use only a small part of the information contained in the sample under study.

Therefore we use an idea due to Laroque-Salanié (1993): we generalize our earlier procedure by choosing some integer $k > 0$, defining $Y_t = (y_t, \ldots, y_{t-k})$ and approximating the likelihood of Y_t by

$$l_t^S(\theta) = \frac{1}{Sh^{k+1}} \sum_{s=1}^{S} K\left(\frac{Y_t - Y_t^s(\theta)}{h}\right)$$

where K and h are a well-chosen kernel and bandwidth, both $(k+1)$-dimensional this time. This will allow us to approximate the marginals $l(y_t, \ldots, y_{t-k})$ of the likelihood function, conditional to $(x_t, \ldots, x_{t-k})$.

The NPSML estimator $\hat{\theta}_T^S$ then is obtained as usual by maximizing

$$\sum_{t=1}^{T} \tau_S(l_t^S(\theta)) \ln l_t^S(\theta)$$

We cannot claim asymptotic efficiency this time, since we are only approximating k-order marginals of the likelihood function. Nevertheless, the NPSML estimator will be close to asymptotically efficient if these marginals contain about as much information as the full likelihood function. Of course, the curse of dimension will severely limit the possible values of k, to, say, 1 or 2. Therefore it is an empirical question whether the efficiency loss is large or not in any particular model.

Theorems 1.1 and 1.2 can be adapted to the dynamic framework. Indeed, the key tool of the proofs is lemma A.1, which can be extended relatively easily. Nonetheless, the notation becomes messier and the assumptions must be strengthened. There are in fact so many technical differences between the static and the dynamic cases that another paper seems to be necessary to expose and prove the asymptotic properties of the dynamic NPSML estimators. It seems more interesting to examine how well these methods perform in practice. To this end, we now turn to a Monte-Carlo simulation study of the finite-sample properties of the dynamic NPSML estimator as applied to the stochastic volatility model.

4 A Monte-Carlo Simulation Very Preliminary

As explained above, the stochastic volatility model has a lagged latent variable (the volatility), which makes the likelihood for observation t a t-dimensional integral. Several estimation methods have been proposed to circumvent this difficulty. Melino-Turnbull (1990) and other authors suggested to use a classical method of moments estimator ; but later simulations have shown that this is a rather inefficient procedure. Harvey-Ruiz-Shephard (1994) proposed a quasi-maximum likelihood estimator (QML) based on rewriting the model as a state-space model and approximating the distribution of the errors with Gaussian variables. Jacquier-Polson-Rossi (1994) presented a Bayesian approach based on a Monte Carlo Markov chain (MCMC). Danielsson (1994) showed how to obtain accurate approximations to the likelihood using importance sampling. Finally, Sandmann-Koopman (1998) have proposed a Monte Carlo maximum likelihood method (MCL).

Our objective here is not to debate the relative merits of the various approaches, but to see how our NPSML method fares in a small Monte Carlo simulation of the stochastic volatility model. This is made easy by the fact that Jacquier-Polson-Rossi (1994) (hereafter JPR) compared the finite sample performance of QML and MCMC and that later Sandmann-Koopman (1998) (hereafter SK) used the very same experimental setup to measure the performance of their MCL method.

We should emphasize here that researchers on the stochastic volatility model usually are not only interested in estimation, but also in filtering (how to get the best estimate of current volatility given the observed returns). We only focus here on estimation.

Let us write the stochastic volatilty model as[8]

$$\begin{cases} r_t &= \overline{\sigma} e^{h_t/2} \xi_t \\ h_t &= \phi h_{t-1} + \sigma_\eta \eta_t \end{cases}$$

Here r_t denotes the (residual) returns and h_t is the underlying latent volatility process. The errors ξ_t and η_t are assumed to be independent across time and to be uncorrelated centered normals with unit variance.

Let $u = Varr_t = \overline{\sigma}^2 e^{h_t}$ the variance of returns. Both JPR and SK specify nine Monte-Carlo designs by crossing two criteria. The first one is the coefficient of variation of u

$$CV = \frac{Vu}{(Eu)^2} = \exp\left(\frac{\sigma_\eta^2}{1-\phi^2}\right)$$

which they take to be equal to $0, 1$, 1 or 10. The second one is the value of the autocorrelation parameter ϕ, which they take to be equal to 0.9, 0.95 or 0.98. Finally, they fix the last parameter by setting the expected variance of returns

$$Eu = \overline{\sigma}^2 \exp\left(\frac{\sigma_\eta^2}{2(1-\phi^2)}\right)$$

equal to 0.0009.

Empirical studies of the stochastic volatility model show that estimates of CV cluster around one. Therefore, as already suggested by SK, we focus on that the case where $CV = 1$. At this very preliminary stage, we have only explored the value $\phi = 0.95$, which is case 5 in SK.

There are a few decisions to be made when applying NPSML. First, there is the number of lags k ; so far we only experimented with $k = 1$ and $k = 2$. Then there is the method used to nonparametrically estimate the density. We used the normal kernel[9]. We chose the bandwidth h by applying Silverman's rule for the bandwidth that minimizes the mean integrated square error for a normal density. Finally, we trim the 5% lowest values of the likelihood. Sometimes the algorithm wanders into regions where more than 5% of the simulated likelihoods are very close to

[8] We use the notation in SK.

[9] Since the stochastic volatility model generates returns that are not serially correlated, we did not prewhiten the data.

zero, which creates problems when taking the logarithm. To avoid this, we bound the likelihood below by 10^{-30}. We checked that this constraint is never binding at the optimum.

Our preliminary results are given in Table 1. The number of observations is $T = 500$ and we ran 500 replications of the stochastic volatility process. The parameters of interest are σ_η, ϕ and $\alpha = (1-\phi)\ln\overline{\sigma}^2$ which is just a simple transformation used to conform to the presentation of the results in JPR and SK.

In our first experiment, denoted NPSML1 in the table, NPSML was used with $k = 1$ and $S = 50$. One estimation takes about twenty seconds on average on a Pentium 300 microcomputer, which seems very competitive with alternative approaches. It turns out that the maximization does not always converge smoothly to reasonable estimates. There are two sorts of difficulties. First, sometimes the algorithm does not converge at all. That happened for 12 replications out of 500. In 8 of these cases the algorithm used (OPTMUM in GAUSS) reports that the secant update failed; in our experience, this problem would typically be solved by perturbing the parameter values slightly and restarting the algorithm. The four other cases report that the calculation of the function failed, which is a bit puzzling at this stage. To compute the summary statistics in Table 1, we eliminated all of these 12 problem cases. There remain 488 replications. Second, there are a notable number of very small estimates of σ_η : 45 are smaller than 0.001. At this stage, we do not know how to explain these outliers, nor whether other researchers have had similar problems. It is likely, though, that in an actual simulations one might find ways around these difficulties.

In addition to the estimates for σ_η, ϕ and α, we also report estimates for $\overline{\sigma}$, for the long-run volatility ($\sqrt{Eu} = \sqrt{Vr}$) and for the long-run standard error of the volatility

$$\sigma_h = \frac{\sigma_\eta}{1-\phi^2}$$

These cannot be found in the tables of JPR and SK, but they obviously are of interest.

The NPSML1 column shows that the estimates for ϕ and σ_η are about as good as the QML estimates, but are noticeably more dispersed than the MCMC and MCL estimates. The estimates for α are wide off the mark, however. Because $\alpha = (1-\phi)\ln\overline{\sigma}^2$ and $\ln\overline{\sigma}_0^2 \simeq -7.4$, any error on ϕ reflects in a large error on α, unless the estimate of $\overline{\sigma}$ covaries enough. This is also a problem for MCMC, but not for QML or MCL. Note, however, that the estimates for the volatility and for $\overline{\sigma}$ are rather good.

In order to try to improve on the NPSML1 results, we also ran a simulation with $k = 2$. Clearly, we need more simulations to overcome the curse of dimensionality. Therefore we chose $S = 500$ this time. This defines the NPSML2 experiment. For lack of time, we only ran 100 replications this time. Each replication takes about 4'40". This time there is only one non-convergence case, and two cases of low σ_η. The results are

Table 1: **Simulation of the stochastic volatility model**

Parameter	True value	QML	MCL	MCMC	NPSML1	NPSML2
σ_η	0.260	0.302	0.233	0.280	0.244	0.257
		(0.17)	(0.07)	(0.07)	(0.19)	(0.14)
ϕ	0.950	0.906	0.930	0.920	0.896	0.928
		(0.18)	(0.10)	(0.05)	(0.18)	(0.07)
α	−0.368	−0.368	−0.372	−0.560	−0.795	−0.536
		(0.01)	(0.01)	(0.34)	(1.42)	(0.52)
$\overline{\sigma}$	0.025	–	–	–	0.022	0.023
		–	–	–	(0.003)	(0.003)
Volatility	0.030	–	–	–	0.026	0.028
		–	–	–	(0.004)	(0.004)
σ_h	0.833	–	–	–	0.731	0.776
		–	–	–	(0.40)	(0.29)

strikingly better. Figure 1 (resp. 2) shows the distribution of the estimates of σ_η (resp. ϕ) for NPSML1 (hashed line[10]) and NPSML2 (full line). The bias on the estimate of σ_η is the lowest of all the methods, even though the standard error is still relatively large. The bias on the estimate on ϕ is comparable to the best existing method (QML), and the standard error is very low. The estimates for α are improved, but are still much worse than those of QML and MCL.

We find these results to be encouraging, for a very first pass. It is worth reminding the reader that contrary to QML, MCMC and MCL, NPSML does not exploit at all the peculiar features of the stochastic volatility model; in a sense, it is rather surprising that it performs so well. Moreover, we have not tried to refine the quality of the nonparametric density estimates that are central to the method. The bandwidth we use is only optimal in the L^2 sense for estimation of a normal density; but the likelihood function is surely not normal, and since we are estimating it in a fixed set of points, we can probably choose the bandwidth more efficiently, at a slight cost in computation time. This is left for further work.

References

Ai, C. (1997), "A Semiparametric Maximum Likelihood Estimator", *Econometrica*, 65, 933-963.

Börsch-Supan, A. and V. Hajivassiliou (1993), "Smooth Unbiased Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models", *Journal of Econometrics*, 58, 347-368.

Bosq, D. and J.-P. Lecoutre (1987), *Théorie de l'estimation fonctionnelle*, Economica, Paris.

[10] A couple of the estimates of ϕ by NPSML1 are actually beyond the left edge of the graph.

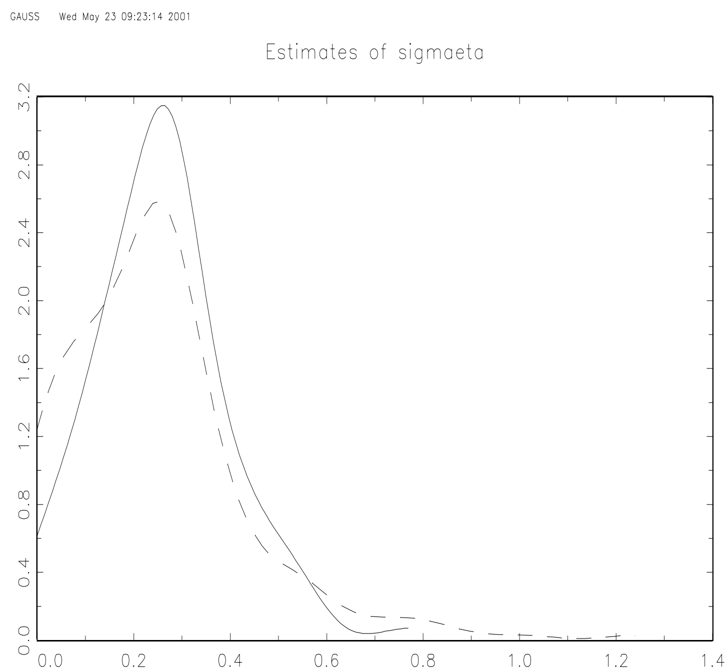

Figure 1: Estimates of σ_η

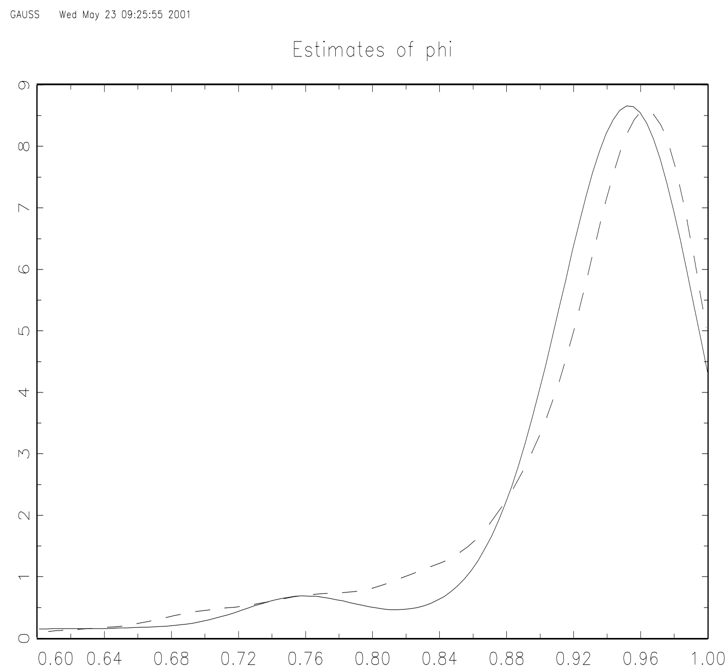

Figure 2: Estimates of ϕ

Gallant, R. and G. Tauchen (1996), "Which Moments to Match?", *Econometric Theory*, 12, 657-681.

Gouriéroux, C. and A. Monfort (1996), *Simulation-based Econometric Methods*, Oxford University Press.

Gouriéroux, C., A. Monfort and E. Renault (1993), "Indirect Inference", *Journal of Applied Econometrics*, 8, S85-S118.

Hajivassiliou, V. and D. McFadden (1998), "The Method of Simulated Scores for the Estimation of Limited-Dependent Variable Models", *Econometrica*, 66, 863-896.

Hajivassiliou, V. and P. Ruud (1994), "Classical Estimation Methods for LDV Models Using Simulation", in R. Engle and D. McFadden eds, *Handbook of econometrics, vol 4*, Elsevier.

Harvey, A., E. Ruiz and N. Shephard (1994), "Multivariate Stochastic Variance Models", *Review of Economic Studies*, 61, 247-264.

Jacquier, E., N. Polson and P. Rossi (1994), "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business and Economic Statistics*, 12, 371-417 (with discussion).

Laroque, G. and B. Salanié (1989), "Estimation of Multimarket Fix-price Models: An Application of Pseudo-Maximum Likelihood methods", *Econometrica*, 57, 831-860.

Laroque, G. and B. Salanié (1993), "Simulation-based Estimation of Models with Lagged Latent Variables", *Journal of Applied Econometrics*, 8, S119-S133.

Laroque, G. and B. Salanié (1994), "Estimating the Canonical Disequilibrium Model: Asymptotic Theory and Finite Sample Properties", *Journal of Econometrics*, 62, 165-210.

Lee, L.F. (1995), "Asymptotic Bias in Simulated Maximum Likelihood Estimation of Discrete Choice Models", *Econometric Theory*, 11, 437-483.

Lerman, S. and C. Manski (1981), "On the Use of Simulated Frequencies to Approximate Choice Probabilities", in C. Manski & D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press.

McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration", *Econometrica* , 57, 995-1026.

Melino, A. and S. Turnbull (1990), "Pricing Foreign Currency Options with Stochastic Volatility", *Journal of Econometrics*, 45, 239-265.

Pakes, A. (1986), "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks", *Econometrica*, 54, 755-84.

Pakes, A. and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, 57, 1027-1057.

Sandmann, G. and S. Koopman (1998), "Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood", *Journal of Econometrics*, 87, 271-301.

Smith, A. (1993), "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions", *Journal of Applied Econometrics*, 8, S63-S84.

Stern, S. (1997), "Simulation-based Estimation", *Journal of Economic Literature*, 35, 2006-2039.

Appendix: Proofs of the Asymptotic Results

Denote by Cst some "universal" positive constants (viz independent from every other quantities). By default, the expectations are taken with respects to the true law whose parameter is θ_0.

A Technical lemmas

The proofs of Theorems 1.1 and 1.2 rely on asymptotic convergence results of many kernel-based estimates. To establish them, we will repeatedly use an improved version of lemma B.1 of Ai (1997). The rate of convergence is a bit higher than in Ai's lemma and the result is true almost everywhere under the additional assumption (A-12) below.

Lemma A.1 *Let $(u_i)_{i\geq 1}$ be an i.i.d. sequence of realizations of a random variable u, and denote*

$$a_N(w) = N^{-1}\sum_{i=1}^{N} a_N(w, u_i)$$

a sample average of some real terms $a_N(w, u_i)$, $w \in \mathbb{R}^K$. Let h_N be a bandwidth sequence such that $h_N \to 0$ when $N \to \infty$ and such that $h_N > N^{-\pi}$ for some $\pi > 0$. Assume that for every (u, w, N, h_N),

i. $h_N^r |a_N(w, u)| < c_1(u)$ *and* $E[c_1^p(u)] < +\infty$ *for some* $r \geq 0$ *and* $p > 2$,

ii. $h_N^s \|\partial a_N(w, u)/\partial w\| < c_2(u)$ *and* $E[c_2(u)] < +\infty$ *for some* $s \geq 0$,

iii. $E[h_N^{2r} a_N^2(w, u)] \leq Cst.h_N^t$ *for some* $t \geq 0$.

Define $W_N = \{w \in \mathbb{R}^K; \|w\| \leq N^\nu\}$, $\nu > 0$. If

$$\sum_{N\geq 1}\left(\frac{\ln N}{N}\right)^{p/2-1} h_N^{-tp/2} < +\infty, \tag{A-12}$$

then there exists a constant C_0 such that almost everywhere, for every N,

$$\left(\frac{Nh^{2r-t}}{\ln N}\right)^{1/2} \mathbf{1}(w \in W_N)\, |a_N(w) - E[a_N(w)]| \leq C_0. \tag{A-13}$$

Moreover, replacing assumption (A-12) with

$$Nh_N^{tp/(p-2)}/\ln N \underset{N\to\infty}{\longrightarrow} +\infty, \tag{A-14}$$

a stronger result is true in probability, viz for every $\varepsilon > 0$,

$$P\left(\left(\frac{Nh^{2r-t}}{\ln N}\right)^{1/2} \mathbf{1}(w \in W_N)\, |a_N(w) - E[a_N(w)]| > \varepsilon\right) \underset{N\to\infty}{\longrightarrow} 0. \tag{A-15}$$

Proof of lemma A.1: To simplify the notation, we suppress most N subscripts from now on. The technique of proof is exactly the same as in Ai (1997) even if our result is a bit stronger and is stated a.e. Note that a.e. $\sup_N |N^{-1}\sum_{i=1}^N c_2(u_i)|$ is bounded. For some $M_N > 0$, define $d_i = \mathbf{1}(c_1(u_i) \leq M_N)$. Then $a(w) = a_1(w) + a_2(w)$ where $a_1(w)$ is a sample average of terms $(1-d_i)a(w,u_i)$ and $a_2(w)$ is a sample average of terms $d_i a(w,u_i)$, $i = 1,\dots,N$. Then

$$P\left(\sup_{w\in W_N} |a(w) - E[a(w)]| > \varepsilon\right) \leq P\left(\sup_{w\in W_N} |a_1(w) - E[a_1(w)]| > \varepsilon/2\right)$$
$$+ \quad P\left(\sup_{w\in W_N} |a_2(w) - E[a_2(w)]| > \varepsilon/2\right) \equiv p_1 + p_2.$$

Invoking condition i, we get

$$\begin{aligned} p_1 \leq P&\left(\sup_{w\in W_N} \frac{1}{N}\sum_{i=1}^N |(1-d_i)a(w,u_i)| > \frac{\varepsilon}{4}\right) \\ + \quad P&\left(\sup_{w\in W_N} E[(1-d_i)|a(w,u_1)|] > \frac{\varepsilon}{4}\right) \\ \leq \quad P&\left(\frac{1}{N}\sum_{i=1}^N (1-d_i)c_1(u_i) > h^r\frac{\varepsilon}{4}\right) \\ + \quad P&\left(E\left[(1-d_i)c_1(u_i)\right] > h^r\frac{\varepsilon}{4}\right). \end{aligned} \tag{A-16}$$

By Hölder's inequality, we have

$$E[(1-d_i)c_1(u_i)] \leq P(c_1(u_i) > M_N)^{1-1/p}\cdot E[c_1^p(u_i)]^{1/p} \leq E[c_1^p(u_i)]/M_N^{p-1}.$$

Therefore, the second term of equation (A-16) is zero for N sufficiently large if $M_N^{p-1}\varepsilon h^r$ tends to the infinity when N tends to the infinity. This assumption will be satisfied with our forthcoming choices (see below). Thus we obtain for N sufficiently large

$$p_1 \leq \frac{4E[(1-d_i)c_1(u_i)]}{\varepsilon h^r} = O\left(\frac{1}{M_N^{p-1}\varepsilon h^r}\right).$$

Moreover, cover W_N classically by b_N boxes W_{jN}, $j = 1,\dots,b_N$ of length δ_N. It is easy to choose the boxes so that $b_N \sim N^{\nu K}/\delta_N^K$. Denote w_j be the center of each box W_{jN}. If $w \in W_{jN}$, we get by assumption ii that

$$|a_2(w) - a_2(w_j)| \leq \frac{\|w - w_j\|}{Nh_N^s}\sum_{i=1}^N c_2(u_i) \leq \frac{C_1\delta_N}{h_N^s} \text{ a.e. },$$

for some constant C_1. Deduce that a.e.

$$\begin{aligned} \sup_{w\in W_N} |a_2(w) - E[a_2(w)]| &\leq \max\nolimits_{1\leq j\leq b_N}\left[\sup_{w\in W_{jN}} |a_2(w) - a_2(w_j)|\right. \\ &+ \left.\sup_{w\in W_{jN}} |E[a_2(w) - a_2(w_j)]| + |a_2(w_j) - E[a_2(w_j)]|\right] \\ &\leq \frac{2C_1\delta_N}{h_N^s} + \sup_j |a_2(w_j) - E[a_2(w_j)]|. \end{aligned}$$

Applying Bernstein's inequality, we get

$$
\begin{aligned}
p_2 &\le P\left(\frac{2C_1\delta_N}{h_N^s} > \varepsilon/4\right) + b_N \sup_{1\le j\le b_N} P\left(|a_2(w_j) - E[a_2(w_j)]| > \varepsilon/4\right) \\
&\le \quad P\left(C_1\frac{\delta_N}{h^s} > \varepsilon/8\right) + 2b_N \exp\left(-\frac{Nh^{2r}\varepsilon^2}{C_2 E[h^{2r}a_N^2(w,u)] + 16M_N\varepsilon h^r}\right) \\
&\le \quad P\left(C_1\frac{\delta_N}{h^s} > \varepsilon/8\right) + 2b_N \exp\left(-\frac{Nh^{2r-t}\varepsilon^2}{C_3 + 16M_N\varepsilon h^{r-t}}\right)
\end{aligned}
$$

for some positive constants C_1, C_2 and C_3. Choosing $\varepsilon^2 = C^* h^{t-2r}\ln N/N$ and $M_N = h^{t-r}\varepsilon^{-1}$, it is easy to verify that $M_N \underset{N\to\infty}{\longrightarrow} +\infty$ under assumption (A-12). Moreover, if $\delta_N = [\varepsilon h^s/(8C_1)] \wedge 1$, then $b_N = O(N^{\bar\pi})$, $\bar\pi > 0$ and

$$p_2 \le 0 + Cst.N^{\bar\pi}\exp(-Cst.C^*\ln N).$$

Thus, for C^* sufficiently large, $\sum_N p_2 < +\infty$. At last, since $M_N^{p-1}\varepsilon h^r = (\ln N/N)^{1-p/2}h^{tp/2}$, assumption (A-12) implies that $\sum_N p_1 < \infty$. Then, by Borel-Cantelli's lemma the strong uniform convergence is proved.

To state the convergence in probability (equation (A-15)), it is sufficient to prove that p_1 and p_2 tend to zero when N tends to the infinity. This is the case with our previous choices (ε, M_N) and under (A-14). $\square$

Lemma A.1 allows us to state the strong consistency of kernel estimates uniformly with respects to the parameter θ and to some increasing compact sets of observations. It will be used repeatedly in the following three lemmas.

Lemma A.2 *Under assumptions K, M1, L2, L5, R3 and T2, for every $\nu > 0$, a.e.*

$$\inf\left(h^{-\rho}, \left(\frac{Sh^m}{\ln S}\right)^{1/2}\right)\mathbf{1}(\|x,y\| \le S^\nu)\sup_\theta\left|l^S(y|x,\theta) - l(y|x,\theta)\right|$$

is bounded. Moreover, it tends to zero in probability replacing assumption R3 by assumption R2.

Proof of lemma A.2: Apply lemma A.1 with $w = (x,y,\theta)$, $u = \varepsilon$, $N = S$, $K = m + d + q$ and

$$a_N(w,u) = h^{-m}K\left(\frac{y - g(x,\theta,\varepsilon)}{h}\right).$$

Since K is bounded, we can choose $r = m$ and p arbitrarily large. Moreover

$$E[h^{2m}a_N^2(w,u)] = \int K^2\left(\frac{y - g(x,\theta,\varepsilon)}{h}\right)dP_\varepsilon = h^m\int K^2(t)l(y-ht|x,\theta)\,dt.$$

By assumption L2, we can choose $t = m$. Moreover

$$
\begin{aligned}
&\left\|h^{m+1+s_0}\frac{\partial a_N(w,u)}{\partial w'}\right\| = h^{s_0}\left\|K'\left(\frac{y-g(x,\theta,\varepsilon)}{h}\right).\left[-\frac{\partial g(x,\theta,\varepsilon)}{\partial x'}, 1, -\frac{\partial g(x,\theta,\varepsilon)}{\partial \theta'}\right]\right\| \\
&\le \quad Cst\|K'\|_\infty\left(1 + \phi(\varepsilon)\right)
\end{aligned}
$$

belongs to L^1. Hence, under R3,

$$\left(\frac{Sh^m}{\ln S}\right)^{1/2} \mathbf{1}(\|x,y\| \leq S^\nu) \sup_\theta \left\| l^S(y|x,\theta) - E[l^S(y|x,\theta)] \right\|$$

is bounded a.e. It remains to deal with the bias term. A Taylor expansion provides

$$E[l^S(y|x,\theta)] = l(y|x,\theta) + \frac{(-h)^\rho}{\rho!} \int \frac{\partial^\rho l(y-\theta_t^* ht|x,\theta)}{\partial^\rho y} K(t) t^\rho \, dt,$$

where $\theta_t^* \in [0,1]$. Since $d^\rho l(\cdot|x,\theta)$ is uniformly bounded (assumption L5), $\sup_{(x,y,\theta)} h^{-\rho} \left| E[l^S(y|x,\theta)] - l(y|x,\theta) \right|$ is bounded, proving the result. It is easy to check that assumption R2 implies A-14 and thus the convergence in probability by Lemma A.1. □

Lemma A.3 *Under assumptions K, M2, M3, L8 and T2, for every $\nu >$ 0, we have a.e.*

$$\inf\left(h^{-\rho}, \left(\frac{Sh^{2m+2+2r_0}}{\ln S}\right)^{1/2}\right) \mathbb{1}(\|x,y\| \leq S^\nu) \sup_\theta \left\| \frac{\partial l^S(y|x,\theta)}{\partial \theta} - \frac{\partial l(y|x,\theta)}{\partial \theta} \right\|$$

tends to zero in probability.

Proof of lemma A.3: Apply lemma A.1 with $w = (x,y,\theta)$, $u = \varepsilon$, $N = S$, $K = m + d + q$ and

$$\begin{aligned} a_N(w,u) &= h^{-m} \frac{\partial}{\partial \theta_k} K\left(\frac{y - g(x,\theta,\varepsilon)}{h})\right) \\ &= \quad -h^{-m-1} \frac{\partial g(x,\theta,\varepsilon)}{\partial \theta_k} K'\left(\frac{y-g(x,\theta,\varepsilon)}{h}\right), \end{aligned}$$

for each $k = 1,\ldots,q$. Set $r = m + 1 + r_0$ and $s = m + 2$. Since $h^{m+1+r_0}|a_N(w,u_i)| \leq \|K'\|_\infty \bar{\phi}(\varepsilon_i)$, and $E[\bar{\phi}(\varepsilon_i)^{p_0}] < \infty$, $p_0 > 2$, lemma A.1 is valid with $p = p_0$. Moreover,

$$\begin{aligned} E[h^{2m+2+2r_0} a_N^2(w,u)] &= h^{2r_0} \int K'\left(\frac{y-g(x,\theta,\varepsilon)}{h}\right)^2 \left(\frac{\partial g(x,\theta,\varepsilon)}{\partial \theta_k}\right)^2 dP_\varepsilon \\ &\leq \quad \|K'\|_\infty^2 E[\bar{\phi}(\varepsilon)^2] < \infty. \end{aligned}$$

Then, set $t = 0$. Clearly, assumption ii of lemma A.1 is satisfied with $s = m + 2 + s_1$. Hence, since $p_0 > 4$, we have

$$\sum_{S \geq 1} \left(\frac{\ln S}{S}\right)^{p_0/2-1} < +\infty, \tag{A-17}$$

Then (A-12) is satisfied and

$$\left(\frac{h^{2m+2+2r_0} S}{\ln S}\right)^{1/2} \mathbf{1}(\|x,y\| \leq S^\nu) \sup_\theta \left\| \frac{\partial l^S(y|x,\theta)}{\partial \theta} - E[\frac{\partial l^S(y|x,\theta)}{\partial \theta}] \right\|$$

is bounded a.e. To deal with the bias, a Taylor expansion provides as previously

$$E[\frac{\partial l^S(y|x,\theta)}{\partial\theta}] = \frac{\partial l(y|x,\theta)}{\partial\theta} + \frac{(-h)^\rho}{\rho!}\int \frac{\partial^{\rho+1}}{\partial\theta\partial^\rho y} l(y-\theta_t^* ht|x,\theta)K(t)t^\rho\, dt,$$

where $\theta_t^* \in [0,1]$. Since $\partial^{\rho+1}l(y|x,\theta)/\partial\theta\partial^\rho y$ is uniformly bounded,

$$\sup_{(x,y,\theta)} h^{-\rho}\left\| E[\frac{\partial l^S(y|x,\theta)}{\partial\theta}] - \frac{\partial l(y|x,\theta)}{\partial\theta}\right\| \text{ is bounded,}$$

proving the result. □

Lemma A.4 *Under assumptions L4, T1 and T2, we have a.e.*

$$\left(\frac{T}{\ln T}\right)^{1/2} \sup_\theta \left\{\frac{1}{T}\sum_{t=1}^{T}[1-\tau_S(l_t(\theta))] - E[1-\tau_S(l_t(\theta))]\right\} \textit{ is bounded .} \tag{A-18}$$

Proof of lemma A.4: Apply lemma A.1 with $w = \theta$, $u = (x,y)$, $N = T$, $K = q$ and

$$a_N(\theta) = T^{-1}\sum_{t=1}^{T} a_N(\theta,u_t), \qquad a_N(\theta,u_t) = [1-\tau_S(l_t(\theta))]\,.$$

Note that a_N depends on (T,h) only, despite the index S and even if there exists a relation between h and S. Therefore, $h = h(T)$ is a sequence which tends to zero when $T \longrightarrow +\infty$. Since a_N is bounded, choose $r = 0$ and p arbitrarily large. Moreover, we can choose $s = \delta$ since

$$\left\|\frac{\partial a_N(\theta,u)}{\partial\theta}\right\| = \left\|\tau_S'(l(y|x,\theta))\frac{\partial l(y|x,\theta)}{\partial\theta}\right\| \le Cst.h^{-\delta}\sup_\theta\left\|\frac{\partial l(y|x,\theta)}{\partial\theta}\right\|.$$

Finally, note that $E[h^{2r}a_N^2(\theta,u)] = O(1)$ and set $t = 0$. Then (A-12) is satisfied and a.e.

$$\left(\frac{T}{\ln T}\right)^{1/2} \sup_\theta \left\{T^{-1}\sum_{t=1}^{T}[1-\tau_S(l_t(\theta))] - E[(1-\tau_S(l_t(\theta)))]\right\} \text{ is bounded ,} \tag{A-19}$$

proving the result. □

B Proof of Theorem 1.1

In this proof, $\sup_\theta$ means the supremum over $\theta \in \Theta$. A simple splitting of the simulated loglikelihood provides

$$\tilde{L}_T^S(\theta) - L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left(\|x_t, y_t\| > S^\nu\right) \tau_S(l_t^S(\theta)) \ln l_t^S(\theta)$$

$$- \quad \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left(\|x_t, y_t\| > S^\nu\right) \ln l_t(\theta)$$

$$+ \quad \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left(\|x_t, y_t\| \leq S^\nu\right) \tau_S(l_t^S(\theta)) \left[\ln l_t^S(\theta) - \ln l_t(\theta)\right]$$

$$+ \quad \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left(\|x_t, y_t\| \leq S^\nu\right) \left[\tau_S(l_t^S(\theta)) - 1\right] \ln l_t(\theta) \equiv T_1 + T_2 + T_3 + T_4.$$

The proof is completed if we show that $\sup_{\theta\in\Theta} |\tilde{L}_T^S(\theta) - L_T(\theta)|$ tends to zero a.e. when S and T tend to the infinity.

Study of $T3$: Invoking lemma A.2, we have almost surely

$$\inf\left\{h^{-\rho}, \left(\frac{Sh^m}{\ln S}\right)^{1/2}\right\} \mathbb{1}(\|x_t, y_t\| \leq S^\nu) \sup_\theta \left|l_t^S(\theta) - l_t(\theta)\right|$$

is bounded. Note that

$$\tau_S(l_t^S(\theta)) \left|\ln l_t^S(\theta) - \ln l_t(\theta)\right| \leq \frac{1}{l_t^*(\theta)} |l_t^S(\theta) - l_t(\theta)|,$$

where $l_t^*(\theta)$ lies between $l_t^S(\theta)$ and $l_t(\theta)$. Moreover, if $l_t^S(\theta)$ tends to $l_t(\theta)$ faster than h^δ, then $\tau_S(l_t^S(\theta)) > 0$ implies that $|l_t^*(\theta)| \geq C.h^\delta$ for some constant C. Hence, since $\delta < \rho$ and $Sh^{m+2\delta}/\ln S \to \infty$, we have a.e. uniformly with respect to θ,

$$|T_3| \leq Cst.h^{-\delta}\left\{h^\rho \vee \left(\frac{Sh^m}{\ln S}\right)^{-1/2}\right\} \equiv O(u_S), \qquad \text{(B-20)}$$

which tends to zero when $S \to \infty$.

Study of T_4: Obviously, we have

$$T_4 = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left(\|x_t, y_t\| \leq S^\nu\right) \left[\tau_S(l_t^S(\theta)) - \tau_S(l_t(\theta))\right] \ln l_t(\theta)$$

$$+ \quad \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left(\|x_t, y_t\| \leq S^\nu\right) \left[\tau_S(l_t(\theta)) - 1\right] \ln l_t(\theta) \equiv T_{41} + T_{42}.$$

Since $h^\delta \|\tau_S'\|_\infty$ is bounded, deduce from lemma A.2 that a.e.

$$\sup_\theta \sup_{x_t, y_t, \|x_t, y_t\| \leq S^\nu} |\ \tau_S(l_t^S(\theta)) - \tau_S(l_t(\theta))\ | = O\left(h^{-\delta}\left\{h^\rho \vee \left(\frac{Sh^m}{\ln S}\right)^{-1/2}\right\}\right) = O(u_S),$$

which tends to zero. Hence, since $\sup_{\theta,T} T^{-1}\sum_{t=1}^{T}|\ln l_t(\theta)|$ is a.e. bounded as a consequence of Assumption L3, then a.e.

$$\sup_{\theta}|T_{41}| = O(u_S).\sup_{\theta,T}\frac{1}{T}\sum_{t=1}^{T}|\ln l_t(\theta)| = O(u_S)\underset{S\to\infty}{\longrightarrow}0.$$

Moreover, by Hölder's inequality, we have for each θ,

$$|T_{42}| \leq \left[\frac{1}{T}\sum_{t=1}^{T}[1-\tau_S(l_t(\theta))]\right]^{1/\alpha}\cdot\left[\frac{1}{T}\sum_{t=1}^{T}|\ln l_t(\theta)|^{\beta}\right]^{1/\beta}, \tag{B-21}$$

where $\alpha^{-1}+\beta^{-1}=1$, $\alpha>1$, $\beta>1$. Thus, by assumption L3 and lemma A.4, we have a.e.

$$\sup_{\theta}|T_{42}| \leq Cst.\left[\sup_{\theta}P(l_t(\theta)\leq 2h^{\delta})+\left(\frac{\ln T}{T}\right)^{1/2}\right]^{1/\alpha}, \tag{B-22}$$

which tends to zero when $h\longrightarrow 0$ and $T\longrightarrow+\infty$.

<u>Study of T_1:</u> Note that $|\tau_S(x)\ln x|\leq \mathbb{1}(x>h^{\delta})|\ln x|$ and that $l_t^S(\theta)\leq \|K\|/h^m$. Thus, since the logarithmic function is monotonic,

$$\sup_{\theta}\left|\tau_S(l_t^S(\theta))\ln l_t^S(\theta)\right| \leq \sup_{l_t^S(\theta)\in[h^{\delta},\|K\|/h^m]}\left|\tau_S(l_t^S(\theta))\ln l_t^S(\theta)\right| \leq \left|\ln\left(\frac{\|K\|}{h^m}\right)\right|\vee|\ln h^{\delta}| = O(\ln h).$$

Thus,

$$\sup_{\theta}|T_1| \leq Cst.|\ln h|\cdot\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\left(\|x_t,y_t\|>S^{\nu}\right).$$

Using Hoeffding's inequality (Bosq and Lecoutre (1987)), for every $\varepsilon>0$,

$$\begin{aligned}&P\left(\sup_{S\leq T^{\kappa}}\left|\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\left(\|x_t,y_t\|\geq S^{\nu}\right)-E[\mathbb{1}(\|x_t,y_t\|>S^{\nu})]\right|>\varepsilon\right)\\ &\leq\quad 2T^{\kappa}\sup_{S\leq T^{\kappa}}\exp(-2T\varepsilon^2).\end{aligned}$$

By Borel-Cantelli's lemma, and setting $\varepsilon^2=C^*\ln T/T$, it is easy to see that a.e.

$$\sup_{S\leq T^{\kappa}}\left|\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\left(\|x_t,y_t\|\geq S^{\nu}\right)-P_{\theta_0}\left(\|x_t,y_t\|>S^{\nu}\right)\right| = O\left(\left(\frac{\ln T}{T}\right)^{1/2}\right).$$

Because $h\geq T^{-\pi\kappa}$ by assumption T1 and T2, $\ln h=O(\ln T)$. Then, deduce from assumption T3 that a.e.

$$\sup_{\theta}|T_1|\underset{S,T\to\infty}{\longrightarrow}0.$$

<u>Study of T_2:</u> Note that, by Hölder's inequality, we have

$$|T_2| \leq \left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\left(\|x_t,y_t\|>S^{\nu}\right)\right]^{1/\alpha}\cdot\left[\frac{1}{T}\sum_{t=1}^{T}|\ln l_t(\theta)|^{\beta}\right]^{1/\beta}.$$

Then, invoking assumption L3, this term can be dealt like T_1, viz $\sup_{\theta}|T_2|$ tends to zero a.e. $\square$

C Proof of Theorem 1.2

Now, we seek to state the asymptotic normality of $\hat{\theta}_T^S$. Note that

$$\frac{\partial L_T}{\partial \theta}(\hat{\theta}_T^S) = \frac{\partial L_T}{\partial \theta}(\theta_0) + \frac{\partial^2 L_T}{\partial \theta \partial \theta'}(\theta^*)(\hat{\theta}_T^S - \theta_0) \text{ and } \frac{\partial \tilde{L}_T^S}{\partial \theta}(\hat{\theta}_T^S) = 0,$$

where θ^* lies between θ_0 and $\hat{\theta}_T^S$. Thus,

$$T^{1/2}(\hat{\theta}_T^S - \theta_0) = \left(-\frac{\partial^2 L_T}{\partial \theta \partial \theta'}(\theta^*)\right)^{-1} \cdot \left\{T^{1/2}\frac{\partial L_T}{\partial \theta}(\theta_0) + T^{1/2}\frac{\partial(\tilde{L}_T^S - L_T)}{\partial \theta}(\hat{\theta}_T^S)\right\}. \tag{C-23}$$

The assumptions of Theorem 1.2 contain those of Theorem 1.1, so that $\hat{\theta}_S^T$ is strongly consistent. Given assumption L1, it is sufficient to prove that

$$T^{1/2}\frac{\partial(\tilde{L}_T^S - L_T)}{\partial \theta}(\theta) \tag{C-24}$$

tends to zero in probability uniformly with respect to θ belonging to a neighborhood V_0 of θ_0, or more precisely that, for every $\varepsilon > 0$,

$$P\left(\sup_{\theta \in V_0}\left\|T^{1/2}\frac{\partial(\tilde{L}_T^S - L_T)(\theta)}{\partial \theta}\right\| > \varepsilon\right) \underset{S,T\to\infty}{\longrightarrow} 0. \tag{C-25}$$

In this proof, θ belongs to V_0. Particularly, $\sup_\theta$ means $\sup_{\theta \in V_0}$. It is sufficient to verify (C-25). Some obvious calculations provide

$$\begin{aligned} T^{1/2}\frac{\partial(\tilde{L}_T^S - L_T)(\theta)}{\partial \theta} &= T^{-1/2}\sum_{t=1}^T \tau_S(l_t^S(\theta))\left[\frac{\partial l_t^S(\theta)}{\partial \theta} - \frac{\partial l_t(\theta)}{\partial \theta}\right]\frac{1}{l_t^S(\theta)} \\ &+ \quad T^{-1/2}\sum_{t=1}^T \tau_S(l_t^S(\theta))\frac{(l_t - l_t^S)(\theta)}{l_t^S(\theta)} \cdot \frac{\partial \ln l_t(\theta)}{\partial \theta} \\ &+ \quad T^{-1/2}\sum_{t=1}^T [\tau_S(l_t^S(\theta)) - 1]\frac{\partial \ln l_t(\theta)}{\partial \theta} \\ &+ \quad T^{-1/2}\sum_{t=1}^T \tau_S'(l_t^S(\theta))\frac{\partial l_t^S(\theta)}{\partial \theta}\ln l_t^S(\theta) \equiv A_1 + A_2 + A_3 + A_4. \end{aligned}$$

Study of A_1: Note that, for every θ and every realization,

$$\left|\frac{\tau_S(l_t^S(\theta))}{l_t^S(\theta)}\right| \le h^{-\delta}.$$

Applying lemma A.3, we obtain for every $\varepsilon > 0$,

$$
\begin{aligned}
P(\sup_\theta \|A_1\| > \varepsilon) \leq P\left(\sup_\theta \left\| \sum_{t=1}^T \mathbb{1}(\|x_t, y_t\| \leq S^\nu) \tau_S(l_t^S(\theta)) \left[\frac{\partial l_t^S(\theta)}{\partial \theta} - \frac{\partial l_t(\theta)}{\partial \theta} \right] \frac{1}{l_t^S(\theta)} \right\| > T^{1/2}\varepsilon/2 \right) \\
+ \quad P\left(\sup_\theta \left\| \sum_{t=1}^T \mathbb{1}(\|x_t, y_t\| > S^\nu) \tau_S(l_t^S(\theta)) \left[\frac{\partial l_t^S(\theta)}{\partial \theta} - \frac{\partial l_t(\theta)}{\partial \theta} \right] \frac{1}{l_t^S(\theta)} \right\| > T^{1/2}\varepsilon/2 \right) \\
\leq \quad P\left(\sup_\theta \sup_{x_t, y_t, \|x_t, y_t\| \leq S^\nu} \left\| \frac{\partial l_t^S(\theta)}{\partial \theta} - \frac{\partial l_t(\theta)}{\partial \theta} \right\| > T^{-1/2} h^\delta \varepsilon/2 \right) \\
+ \quad P\left(\sup_\theta \sum_{t=1}^T \mathbb{1}(\|x_t, y_t\| > S^\nu) \left\| \frac{\partial l_t^S(\theta)}{\partial \theta} - \frac{\partial l_t(\theta)}{\partial \theta} \right\| > T^{1/2} h^\delta \varepsilon/2 \right) \\
\leq \quad P\left(Cst \left\{ h^\rho \vee \left(\frac{\ln S}{S h^{2m+2+2r_0}} \right)^{1/2} \right\} > T^{-1/2} h^\delta \varepsilon/2 \right) \\
+ \quad P\left(\sup_\theta \sum_{t=1}^T \mathbb{1}(\|x_t, y_t\| > S^\nu) \left\| \frac{\partial l_t^S(\theta)}{\partial \theta} - \frac{\partial l_t(\theta)}{\partial \theta} \right\| > T^{1/2} h^\delta \varepsilon/2 \right) \equiv P_{11} + P_{12}.
\end{aligned}
$$

The first term P_{11} is zero for T sufficiently large under assumptions R4 and R5. To deal with the second term, recall that, by assumption L6, $\|\partial l_t / \partial \theta\|$ is bounded and, by assumption M2,

$$
\sup_\theta \left\| \frac{\partial l_t^S}{\partial \theta} \right\| \leq \|K'\|_\infty h^{-m-1} \cdot \frac{1}{S} \sum_{s=1}^S \bar{\phi}(\varepsilon_t^s),
$$

where each $\bar{\phi}(\varepsilon_t^s) \in L^{p_0}$, $p_0 > 2$. Since the ε_t^s are independent from (x_t, y_t) for each s and t, this provides

$$
\begin{aligned}
P_{12} \leq P\left(\sum_{t=1}^T \mathbb{1}\left(\|x_t, y_t\| > S^\nu\right) \sum_{s=1}^S \bar{\phi}(\varepsilon_t^s) > Cst.T^{1/2} S h^{\delta+m+1} \varepsilon \right) \\
+ \quad P\left(\sum_{t=1}^T \mathbb{1}\left(\|x_t, y_t\| > S^\nu\right) > Cst.T^{1/2} h^\delta \varepsilon \right) \\
\leq \quad Cst.T^{1/2} \frac{E\left[\mathbb{1}(\|x_t, y_t\| > S^\nu) \bar{\phi}(\varepsilon_t^s)\right]}{h^{\delta+m+1}\varepsilon} + Cst.T^{1/2} \frac{E\left[\mathbb{1}(\|x_t, y_t\| > S^\nu)\right]}{h^\delta \varepsilon} \\
\leq \quad Cst \cdot \frac{T^{1/2}}{h^{\delta+m+1}\varepsilon} P\left(\|x_t, y_t\| > S^\nu\right) E[\bar{\phi}(\varepsilon_t^s)],
\end{aligned}
$$

which tends to zero by assumption T4 (using the independence between (x_t, y_t) and $(\varepsilon_t^s)_s$).

Study of A_2: For each θ and each realization, we have

$$
\begin{aligned}
\|A_2\| \leq T^{-1/2} \sum_{t=1}^T |(l_t^S - l_t)(\theta)| \mathbb{1}(\|x_t, y_t\| \leq S^\nu) \left| \frac{\tau_S(l_t^S(\theta))}{l_t^S(\theta)} \right| \cdot \left\| \frac{\partial \ln l_t(\theta)}{\partial \theta} \right\| \\
+ \quad T^{-1/2} \sum_{t=1}^T \left(|l_t^S(\theta)| + l_t(\theta) \right) \mathbb{1}(\|x_t, y_t\| > S^\nu) \left| \frac{\tau_S(l_t^S(\theta))}{l_t^S(\theta)} \right| \cdot \left\| \frac{\partial \ln l_t(\theta)}{\partial \theta} \right\| \equiv A_{21} + A_{22}.
\end{aligned}
$$

Applying lemma A.2, it is easy to see that, for all $\varepsilon > 0$,

$$
\begin{aligned}
P(\sup_\theta \|A_{21}\| \geq \varepsilon) &\leq P\left(T^{-1/2}\sup_\theta \sum_{t=1}^T |(l_t^S - l_t)(\theta)| \mathbb{1}(\|x_t, y_t\| \leq S^\nu)\cdot \left\|\frac{\partial \ln l_t(\theta)}{\partial\theta}\right\| > h^\delta \varepsilon\right) \\
&\leq P\left(T^{-1/2}\left\{h^{\rho-\delta} \vee \left(\frac{\ln S}{Sh^{m+2\delta}}\right)^{1/2}\right\}\sup_\theta \sum_{t=1}^T \left\|\frac{\partial \ln l_t(\theta)}{\partial\theta}\right\| > Cst.\varepsilon\right) \\
&\leq P\left(\sup_\theta\left(\frac{1}{T}\sum_{t=1}^T \left\|\frac{\partial \ln l_t(\theta)}{\partial\theta}\right\|^\gamma\right)^{1/\gamma} > Cst.\varepsilon\lambda_T^S\right)
\end{aligned}
$$

where $\lambda_T^S \to \infty$ when S and T tend to the infinity (by assumptions R4 and R5). Hence, by assumption L7, we have

$$
P(\sup_\theta \|A_{21}\| > \varepsilon) \underset{S,T\to\infty}{\longrightarrow} 0.
$$

Moreover,

$$
\begin{aligned}
\sup_\theta \|A_{22}\| &\leq \sup_\theta T^{-1/2}\sum_{t=1}^T \left[1 + Cst.h^{-\delta} l_t(\theta)\right] \mathbb{1}(\|x_t, y_t\| > S^\nu) \left\|\frac{\partial \ln l_t(\theta)}{\partial\theta}\right\| \\
&\leq T^{1/2}\left(\frac{1}{T}\sum_{t=1}^T \mathbb{1}(\|x_t, y_t\| > S^\nu)\right)^{1-1/\gamma}\cdot\left(\sup_\theta \frac{1}{T}\sum_{t=1}^T \left\|\frac{\partial \ln l_t(\theta)}{\partial\theta}\right\|^\gamma\right)^{1/\gamma} \\
&+ Cst.T^{1/2}h^{-\delta}\left(\frac{1}{T}\sum_{t=1}^T \mathbb{1}(\|x_t, y_t\| > S^\nu\right)\cdot\left(\sup_{x,y,\theta}\left\|\frac{\partial l(y|x,\theta)}{\partial\theta}\right\|\right).
\end{aligned}
$$

Therefore, using assumption L7 and the uniform boundedness of $\partial l_t(\theta)/\partial\theta$ (assumption L6),

$$
\begin{aligned}
P(\sup_\theta \|A_{22}\| > \varepsilon) &\leq P\left(\left(\frac{1}{T}\sum_{t=1}^T \mathbb{1}(\|x_t, y_t\| > S^\nu)\right)^{1-1/\gamma} > Cst.T^{-1/2}\varepsilon\right) \\
&+ P\left(\frac{1}{T}\sum_{t=1}^T \mathbb{1}(\|x_t, y_t\| > S^\nu) > Cst.T^{-1/2}h^\delta\varepsilon\right) \\
&\leq Cst\frac{T^{\gamma/2(\gamma-1)}}{\varepsilon^{\gamma/(\gamma-1)}}P\left(\|x_t, y_t\| > S^\nu\right) + Cst\cdot\frac{T^{1/2}}{h^\delta\varepsilon}P\left(\|x_t, y_t\| > S^\nu\right).
\end{aligned}
$$

Thus, invoking assumption T4, $\sup_\theta \|A_{22}\|$ tends to zero in probability.

Study of A_3: Thanks to assumption L7 and Hölder's inequality, note that

$$\begin{aligned}
\|A_3\| &\leq T^{-1/2}\sum_{t=1}^{T}\mathbb{1}(\|x_t,y_t\|\leq S^\nu)\left|\tau_S(l_t^S(\theta))-\tau_S(l_t(\theta))\right|.\left\|\frac{\partial\ln l_t(\theta)}{\partial\theta}\right\| \\
&+ \quad T^{-1/2}\sum_{t=1}^{T}\mathbb{1}(\|x_t,y_t\|>S^\nu)\left\|\frac{\partial\ln l_t(\theta)}{\partial\theta}\right\|+T^{-1/2}\sum_{t=1}^{T}[1-\tau_S(l_t(\theta))]\left\|\frac{\partial\ln l_t(\theta)}{\partial\theta}\right\| \\
&\leq \quad T^{1/2}\left\{Cst.h^{-\delta}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(\|x_t,y_t\|\leq S^\nu)\left|l_t^S(\theta)-l_t(\theta)\right|^{\gamma/(\gamma-1)}\right]^{1-1/\gamma}\right. \\
&+ \quad \left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(\|x_t,y_t\|>S^\nu)\right]^{1-1/\gamma} \\
&+ \quad \left.\left[\frac{1}{T}\sum_{t=1}^{T}[1-\tau_S(l_t(\theta))]\right]^{1-1/\gamma}\right\}.\left[\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{\partial\ln l_t(\theta)}{\partial\theta}\right\|^\gamma\right]^{1/\gamma} \qquad \text{(C-26)}
\end{aligned}$$

Applying lemma A.2, the first term is bounded in probability by

$$Cst.T^{1/2}h^{-\delta}\left[\left(\frac{\ln S}{Sh^m}\right)^{1/2}+h^\rho\right],$$

which tends to zero by assumptions R4 and R5. Moreover, note that

$$\sup_\theta\frac{1}{T}\sum_{t=1}^{T}[1-\tau_S(l_t(\theta))]\leq\sup_\theta\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(l_t(\theta)\leq 2h^\delta)\leq\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(l_t^*(\theta_0)\leq 2h^\delta), \qquad \text{(C-27)}$$

where $l_t^*(\theta_0)=\inf_{\theta\in V_0}l_t(\theta)$. Thus,

$$\begin{aligned}
&P\left(\sup_\theta T^{1/2}\left[\frac{1}{T}\sum_{t=1}^{T}[1-\tau_S(l_t(\theta))]\right]^{1-1/\gamma}>\varepsilon\right) \\
&\leq \quad P\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(l_t^*(\theta_0)\leq 2h^\delta)>(\varepsilon T^{-1/2})^{\gamma/(\gamma-1)}\right) \\
&\leq \quad \varepsilon^{-\gamma/(\gamma-1)}T^{\gamma/2(\gamma-1)}P(l_t^*(\theta_0)\leq 2h^\delta),
\end{aligned}$$

which tends to zero by assumption R6.

It remains to deal with the second term of (C-26), which is of the same order as

$$T^{1/2}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(\|x_t,y_t\|>S^\nu)\right]^{1-1/\gamma}.$$

But, for every $\eta>0$,

$$\begin{aligned}
&P\left(T^{1/2}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(\|x_t,y_t\|>S^\nu)\right]^{1-1/\gamma}>\eta\right)\leq\left(\eta T^{-1/2}\right)^{\gamma/(1-\gamma)}P(\|X,Y\|>S^\nu) \\
&= \quad O\left(T^{\gamma/(2\gamma-2)}P(\|X,Y\|>S^\nu)\right),
\end{aligned}$$

which tends to zero when $(S,T) \longrightarrow \infty$, by assumption T4.

Study of A_4: Let us split A_4 as

$$A_4 = T^{-1/2}\sum_{t=1}^T \tau_S'(l_t^S(\theta))\ln l_t^S(\theta)\left(\frac{\partial l_t^S(\theta)}{\partial\theta} - \frac{\partial l_t(\theta)}{\partial\theta}\right)\mathbb{1}(\|x_t,y_t\| \leq S^\nu)$$

$$+ \quad T^{-1/2}\sum_{t=1}^T \tau_S'(l_t^S(\theta))\ln l_t^S(\theta)\frac{\partial l_t(\theta)}{\partial\theta}\mathbb{1}(\|x_t,y_t\| \leq S^\nu)$$

$$+ \quad T^{-1/2}\sum_{t=1}^T \tau_S'(l_t^S(\theta))\ln l_t^S(\theta)\frac{\partial l_t^S(\theta)}{\partial\theta}\mathbb{1}(\|x_t,y_t\| > S^\nu) \equiv A_{41} + A_{42} + A_{43}.$$

Since τ_S' is a polynomial supported by $[h^\delta, 2h^\delta]$, we have for every $x > 0$,

$$0 \leq \tau_S'(x)|\ln x| \leq Cst.h^{-\delta}|\ln h|\bar\tau_S(x),$$

where $(\bar\tau_S)_{S\geq 1}$ is a bounded sequence of polynomials supported by $[h^\delta, 2h^\delta]$. Invoking lemma A.3, we obtain that

$$P(\sup_\theta \|A_{41}\| > \varepsilon) \leq P\left(\sup_\theta \sup_{x_t,y_t,\|x_t,y_t\|\leq S^\nu}\left\|\frac{\partial l_t^S(\theta)}{\partial\theta} - \frac{\partial l_t(\theta)}{\partial\theta}\right\| > Cst.\varepsilon T^{-1/2}h^\delta/|\ln h|\right)$$

$$\leq \quad P\left(\sup_\theta T^{1/2}|\ln h|\left\{h^{\rho-\delta} \vee \left(\frac{\ln S}{Sh^{2m+2\delta+2+2r_0}}\right)^{1/2}\right\} > Cst.\varepsilon\right)$$

which is zero for S sufficiently large, thanks to assumptions R4 and R5.

Since the functions $(\bar\tau_S)_{S\geq 1}$ can be dealt exactly like $(1-\tau_S)_{S\geq 1}$, the term A_{42} is bounded like A_3, replacing γ by $+\infty$. Therefore, for S sufficiently large

$$\|A_{42}\| \leq Cst.T^{1/2}h^{-\delta}|\ln h|\sup_\theta \frac{1}{T}\sum_{t=1}^T \bar\tau_S(l_t(\theta))\left\|\frac{\partial l(y_t|x_t,\theta)}{\partial\theta}\right\|$$

$$\leq \quad Cst.|\ln h|T^{1/2}h^{-\delta}\sup_{\{(x,y,\theta)\in A_h\}}\left\|\frac{\partial l(y_t|x_t,\theta)}{\partial\theta}\right\|\frac{1}{T}\sum_{t=1}^T \mathbb{1}(l_t^*(\theta_0) \leq 2h^\delta),$$

where $l_t^* = \inf_{\theta\in V_0} l_t(\theta)$. Thus, for every $\varepsilon > 0$,

$$P\left(\sup_\theta \|A_{42}\| > \varepsilon\right) \leq Cst.\varepsilon^{-1}|\ln h|T^{1/2}h^{-\delta}\sup_{\{(x,y,\theta)\in A_h\}}\left\|\frac{\partial l(y_t|x_t,\theta)}{\partial\theta}\right\|P_{\theta_0}(l_t^*(\theta_0) \leq 2h^\delta),$$

which tends to zero under R6.

Finally, note that

$$\sup_\theta \|A_{43}\| \leq Cst\cdot\frac{|\ln h|}{h^{m+\delta+1}}T^{-1/2}\sum_{t=1}^T \mathbb{1}(\|x_t,y_t\| > S^\nu)\cdot\frac{1}{S}\sum_{s=1}^S \bar\phi(\varepsilon_t^s),$$

and deduce that

$$P(\sup_\theta \|A_{43}\| > \varepsilon) \leq \frac{Cst.T^{1/2}|\ln h|}{h^{m+1+\delta}\varepsilon}E\left[\mathbb{1}(\|x_t,y_t\| > S^\nu)\bar\phi(\varepsilon_t^s)\right]$$

$$\leq \quad \frac{Cst.T^{1/2}|\ln h|}{h^{m+1+\delta}\varepsilon}P\left(\|x_t,y_t\| > S^\nu\right).E[\bar\phi(\varepsilon_t^s)]$$

which tends to zero by assumption T4, proving the result. □