

ASYMPTOTIC PROPERTIES OF IDENTIFICATION OF HAMMERSTEIN MODELS WITH INPUT SATURATION

Dietmar Bauer ^{*,1} Brett Ninness^{**}

** Institute f. Econometrics, Operations Research and System
Theory,*

TU Wien, Austria,

Tel.: ++43 +1 58801/11944, Fax.: ++43 +1 58801 11999,

e-mail: Dietmar.Bauer@tuwien.ac.at

*** Department of Electrical and Computer Engineering,
University of Newcastle, Australia*

Abstract: This paper considers the estimation of Hammerstein models with input saturation. These models are characterised by a linear dynamical model acting on an input sequence which is affected by a hard saturation of unknown level. The main result of the paper lies in a specification of a set of sufficient conditions on the input sequence in order to ensure that a non-linear least-squares approach enjoys properties of consistency and asymptotic normality and furthermore, that an estimate of the parameter covariance matrix is also consistent. The set of assumptions is specified using the concept of near epoch dependence, which has been developed in the econometrics literature. Indeed, one purpose of this paper is to highlight the usefulness of this concept in the context of analysing estimation procedures for nonlinear dynamical systems.

Keywords: Nonlinear system identification, Hammerstein models, asymptotic properties

1. INTRODUCTION

This paper deals with the identification of models which might appear to be very specific. However, this class of systems have been analysed and applied in many different settings (Tao and Kokotović, 1996; Rangan *et al.*, 1995; Suykens and Wandewalle, 1998).

Indeed, the hard saturation considered in this paper seems to be a realistic assumption in many applications, since usually the control input signals can only be applied in a certain interval due to technical limitations. Examples include

the angle of a rudder, the opening of a valve, the power input to an electrical motor (Tao and Kokotović, 1996).

In one sense, the results considered in this paper are not particularly new. For example, (L.Ljung, 1978) states conditions, under which nonlinear least squares estimators are consistent in a very general framework.

However, the difference between these pre-existing results and the results of this paper are that here, by virtue of a particular stochastic framework applied, the structure of the required assumptions is significantly different in a way that may be more natural in applications of the results.

¹ Some of the work for this paper has been done while this author was visiting the Department of Automatic Control, Linköping University, Sweden.

Specifically, whereas (L.Ljung, 1978) uses assumption imposed on the measured input and output data which imply certain properties of the non-linearities, our assumptions are imposed directly on the the nature of the non-linearities.

Since we are dealing with consistency and asymptotic normality we need the notion of a true system. However, we want to emphasize that results in the spirit of (L.Ljung, 1978) in which no such true system exists can also be handled by the framework used in this paper.

A key feature of this paper is the introduction to the engineering community of tools provided in (Pötscher and Prucha, 1997) in the context of non-linear models arising in econometrics and, in particular, the idea of ' L_p near epoch dependence' proves to be of great utility.

This latter concept formularises the dependence of a process on some other underlying process, and it is made powerful by the fact that some of the properties of the underlying process can be transferred to the process under investigation and furthermore, these properties are retained under a wide range of non-linear and dynamic transformations.

It is the aim of this paper to demonstrate the application of these tools and concepts in order to find sufficient conditions on the additive noise and the input sequence in order to guarantee consistency, asymptotic normality and consistency of an estimate of the corresponding covariance matrix for nonlinear least squares estimates.

The paper is organised as follows: In the next section we describe the model set and present the necessary concepts. Section 3 then provides some facts for the concept of L_p near epoch dependency. In section 4 the main results of the paper are stated. Section 5 finishes with a discussion of the obtained results.

2. MODEL SET AND ESTIMATION CRITERIA

In this paper we deal with discrete time Hammerstein models with input saturation which can be described in the following form:

$$\begin{aligned} v_{t,i} &= f(u_{t,i}, \alpha_{1,i}, \alpha_{2,i}) \\ x_{t+1} &= Ax_t + Bv_t \\ y_t &= Cx_t + n_t \end{aligned} \quad (1)$$

where

$$f(u, \alpha_1, \alpha_2) = \begin{cases} \alpha_1 & ; u \leq \alpha_1 \\ u & ; \alpha_1 \leq u \leq \alpha_2 \\ \alpha_2 & ; u \geq \alpha_2 \end{cases}$$

and

$$\begin{aligned} y_t \in \mathbf{R}^p, \quad u_t \mathbf{R}^m, \quad v_t \in \mathbf{R}^m, \quad x_t \in \mathbf{R}^n, \quad n_t \in \mathbf{R}^p \\ A \in \mathbf{R}^{n \times n}, \quad B \in \mathbf{R}^{n \times m}, \quad C \in \mathbf{R}^{p \times n}. \end{aligned}$$

It will always be assumed that $\underline{c} \leq \alpha_{1,i} \leq \alpha_{2,i} \leq \bar{c}$ for some $-\infty < \underline{c} < \bar{c} < \infty$. Let $\alpha \in \Lambda \subset \mathbf{R}^{2m}$ denote the vector of stacked saturation points. In the following we will also use the notation $v_t(u_t, \alpha)$ to denote the vector of functions $f(u_{t,i}, \alpha_{1,i}, \alpha_{2,i})$. It will also be assumed that $(A, B, C) = (A(\beta), B(\beta), C(\beta))$, i.e. that the system is parametrised by some parameter vector $\beta \in \Phi$. The parameter set Φ is assumed to be compact and $\{(A(\beta), B(\beta), C(\beta)) : \beta \in \Phi\}$ is assumed to contain only stable transfer functions. We will also use the notation $G(q, \theta) = C(qI - A)^{-1}B$. The direct feedthrough term is only neglected for notational reasons and a term $D(\theta)$ could be included without changing any of the results of this paper. The full parameter vector θ then is defined as $\theta = (\alpha, \beta) \in \Theta$. Here the parameter set Θ is again assumed to be compact. In this set there might be nonminimal systems in the following sense: If e.g. the first column of B is zero, then the saturation of the first component of u_t cannot be determined, since this coordinate of the input does not contribute at all to the output. Therefore all saturation parameters in this coordinate correpond to the same output behaviour and thus the level of saturation cannot be identified. Throughout the paper a system will be called minimal, if each single input system $G(q, \theta)e_i, 1 \leq i \leq m$ is nonzero, where e_i denotes the i -th vector of the canonical basis.

The estimation criterion we will consider is nonlinear least squares, i.e. the estimate $\hat{\theta}$ is defined as the minimising argument of

$$\hat{\theta} \triangleq \arg \min_{\theta \in \Theta} \frac{1}{T} \text{Tr} \left[\Sigma \sum_{t=1}^T e_t(\theta) e_t(\theta)' \right]$$

where $\Sigma \in \mathbf{R}^{p \times p}$ is user defined and $e_t(\theta) = y_t - \hat{y}_t(\theta)$ denotes the one step ahead prediction error according to the model specified by θ as:

$$\hat{y}_t(\theta) = G(q, \theta)v_t(u_t, \theta)$$

This implies that we do not model the noise n_t . Similar results can be obtained if the noise is suitably modeled as an ARMA process and estimation is performed by pseudo maximum likelihood estimation (i.e. using the Gaussian density as the criterion function). The choice of the user defined parameter Σ influences the estimation accuracy. In the case, where n_t is white noise with zero mean and variance equal to Ω then the optimal choice is equal to $\Sigma = \Omega^{-1}$, and then the estimates are asymptotically efficient and equivalent to ML estimates. In general however n_t will be correlated and the usual choice of Σ will be the identity.

3. NEAR EPOCH DEPENDENCE

The assumptions on the input and the noise will be stated in the concept of L_p near epoch dependency which is defined as follows:

Definition 1. A scalar process $\{y_t\}$ is called L_p ($p > 1$) near epoch dependent (n.e.d.) on some basis process $\{e_t\}$ of size $-q$, $q > 0$, if

$$\sup_{t \in \mathbb{N}} \mathbb{E}\{(\{y_t - \mathbb{E}\{y_t \mid e_{t+m}, \dots, e_{t-m}\}\}^p)^{1/p} = \phi(m)$$

where $\phi(m)/m^q \rightarrow 0$. A vector process is called L_p n.e.d. of size $-q$ on $\{e_t\}$, if each component is L_p n.e.d. of size $-q$ on $\{e_t\}$.

Thus the size of the n.e.d. gives a hint on the magnitude of the influence of the underlying process $\{e_t\}$ for times far apart. The index p indicates the norm in which the deviations are measured.

In the following we will use a more restrictive framework in order to make the exposition simpler. Throughout the paper we will assume, that the underlying process $\{e_t\}$ is i.i.d. It will also be assumed, that a L_p n.e.d. process $\{y_t\}$ is strictly stationary. Both additional assumptions are not minimal in the sense that similar results can be obtained for much weaker assumptions. However for notational simplicity we choose this framework.

To give an example of a L_p n.e.d. process consider the process

$$X_t = \sum_{j=0}^{\infty} K_j e_{t-j}$$

where $\{e_t\}$ is i.i.d. with finite second moment. If $\sum_{j=m+1}^{\infty} \|K_j\|_2^2 \leq m^{-2q}$, then $\{X_t\}$ is L_2 n.e.d. on $\{e_t\}$ of size $-r$ for all $r > q$. Thus if the filter $\{K_j\}$ corresponds to an ARMA system, then $\{X_t\}$ is L_2 n.e.d. on $\{e_t\}$ of any size q .

Some of the properties of L_p n.e.d. processes are collected in the next lemma. For proofs and additional references see (Pötscher and Prucha, 1997).

Lemma 2. Suppose that

- (1) $g(s)$ fulfills the following Lipschitz type of condition on each compact subset Ξ of its domain: For every compact Ξ there exists a constant $C(\Xi)$ such that for $s, s' \in \Xi$ we have that for some $p > 1$ it holds that

$$\|g(s) - g(s')\|_p \leq C(\Xi) \|s - s'\|_p$$

- (2) $\{X_t\}$ is a scalar strictly stationary process such that $\mathbf{E}\{|g(X_t)|^p\} < \infty$ for some $p > 1$.
- (3) $\{X_t\}$ is L_p n.e.d. on $\{e_t\}$ of size $-q$.

Then the following holds:

- i) The process $\{g(X_t)\}$ is also L_p n.e.d. on $\{e_t\}$, but maybe of a different size. If $C(\Xi) = C$

independent of Ξ then $\{g(X_t)\}$ is L_p n.e.d. of size $-q$.

- ii) For every finite integer $k > 0$ the process $\{\{X'_t, X'_{t-1}, \dots, X'_{t-k}\}'\}$ is L_p n.e.d. on $\{e_t\}$ of size $-q$.
- iii) If $\{X_t\}$ is L_p n.e.d. on $\{e_t^X\}$ of size $-q$ and $\{Y_t\}$ is L_p n.e.d. on $\{e_t^Y\}$ of size $-q$ then $\{\{X'_t, Y'_t\}'\}$ is L_p n.e.d. of size $-q$ on $\{\{(e_t^X)', (e_t^Y)'\}'\}$.
- iv) If $\{X_t\}$ is a vector, such that each component is L_p n.e.d. on $\{e_t\}$ of size $-q$ and β is of suitable dimension, then $\{\beta' X_t\}$ is L_p n.e.d. of size $-q$ on $\{e_t\}$.
- v) $\{X_t\}$ is L_p n.e.d. on $\{e_t\}$ of size $-r$, $r > q$.
- vi) $\{X_t\}$ is $L_{p'}$ n.e.d. on $\{e_t\}$ of size $-q$ for all $p' < p$.
- vii) If $\mathbf{E}\{|X_t|^{p'}\} < \infty$, then $\{X_t\}$ is L_s n.e.d. on $\{e_t\}$ of size $-q$ for all $s < p'$.
- viii) If $Y_{t+k} = g(Y_t, X_t, \dots, X_{t+k})$, where

$$\begin{aligned} & |g(Y_t, X_t, \dots, X_{t+k}) - g(\bar{Y}_t, \bar{X}_t, \dots, \bar{X}_{t+k})| \leq \\ & d_y |Y_t - \bar{Y}_t| + d_x \|(X_t - \bar{X}_t, \dots, X_{t+k} - \bar{X}_{t+k})\|, \\ & 0 < d_y < 1, d_x < \infty \end{aligned}$$

then $\{Y_t\}$ is L_p n.e.d. on $\{e_t\}$ of size $-q$.

- ix) If $\mathbf{E}\{\|X_t\|_1^{1+\varepsilon}\} < \infty$, $\varepsilon > 0$ and if $\{e_t\}$ is i.i.d. then $\{X_t\}$ fulfills a weak law of large numbers, i.e. $N^{-1} \sum_{j=1}^N X_t \rightarrow \mathbf{E}X_t$, where convergence is in probability with $N \rightarrow \infty$.
- x) Let $\{X_t\}$ be L_2 n.e.d. on $\{e_t\}$, which is i.i.d., of size -1 and satisfies $\mathbf{E}X_t = 0$, $\mathbf{E}\{|X_t|^r\} < \infty$, $r > 2$. Further define

$$P_N = \frac{1}{N} \mathbf{E} \left\{ \sum_{t=1}^N X_t \left(\sum_{t=1}^N X_t \right)' \right\}$$

Assume, that $P_N \rightarrow P$. Then

$$N^{-1/2} \sum_{t=1}^N X_t \rightarrow \mathcal{N}(0, P)$$

as $N \rightarrow \infty$, where convergence is in distribution and $\mathcal{N}(0, P)$ stands for normal distribution with zero mean and variance matrix P .

Point i) states the invariance of the concept under nonlinear transformations fulfilling Lipschitz type of conditions, and these apply for a large class of practical relevance such as wavelets, sigmoidal functions used in neural networks and polynomials.

If the Lipschitz bound is required to be independent of the set, then polynomials can no longer be used without further requirements. These additional assumptions include higher moment conditions, see (Pötscher and Prucha, 1997, Chapter 6) for a discussion on this topic. ii) - iv) show that the concepts are robust with respect to linear transformations. v) - vii) provide an ordering of sizes and norms used. viii) shows, that some

dynamic transformations are allowed with n.e.d. processes. Crucial for the argument here is the restriction $d_y < 1$. However the condition is not as restrictive as it might seem on first sight, since the condition has to hold only for some k . Consider a multivariate autoregression of first order $x_{t+1} = Ax_t + Bu_t$. Even for stable A the norm $\|A\|_2$ can be larger than 1. However, since all eigenvalues of A lie strictly inside the unit circle due to the assumed stability, then there exists a k such that $\|A^k\|_2 < 1$. Considering $x_{t+k} = A^k x_t + \sum_{j=0}^{k-1} A^j B u_{t+k-j}$ then shows the n.e.d. of $\{x_t\}$ on $\{u_t\}$. Finally ix) and x) will be central for the investigation of asymptotic properties, since they allow one to transfer law of large numbers and central limit theorems from the underlying process to the n.e.d. process by checking simple moment conditions.

These tools will be applied in the following in the analysis of the estimates using the model structure described in section 2.

4. ASYMPTOTIC PROPERTIES

In this section the asymptotic properties of the least squares estimate $\hat{\theta}$ are analysed. The discussion will provide a consistency result, the asymptotic normality of the parameter vector and finally also a procedure to assess the variance matrix of the limiting normal distribution will be given and consistency thereof stated.

Theorem 3. (Consistency). Let $\{y_t\}$ be generated by a system of the form (1). Assume that the additive noise $\{n_t\}$ is a strictly stationary process, which is $L_{2+\varepsilon}$ ($\varepsilon > 0$) n.e.d. on $\{e_t^n\}$ (i.i.d.), where $\mathbf{E}n_t = 0$, $\mathbf{E}\{\|n_t\|^{2+\varepsilon}\} < \infty$ and $\mathbf{E}\{\|e_t^n\|^{2+\varepsilon}\} < \infty$. Let the input $\{u_t\}$ be a strictly stationary process, which is $L_{2+\varepsilon}$ ($\varepsilon > 0$) n.e.d. on $\{e_t^u\}$ (i.i.d.), where $\mathbf{E}\{\|u_t\|^{2+\varepsilon}\} < \infty$ and $\mathbf{E}\{\|e_t^u\|^{2+\varepsilon}\} < \infty$. The input is assumed to be independent of $\{n_t\}$. Assume, that the model order is known. The parameter set is assumed to be compact. Corresponding to the input we assume, that the stationary distribution of $[u_t', \dots, u_{t-3n}']'$ has a density, which is strictly greater than zero on $[-\underline{c}, \bar{c}]^{3n+1}$. Also assume that the order of the system n is known. Finally assume, that the true system is minimal. Then the estimate $\hat{\theta} \in \Theta$ converges in probability to the true parameter value θ_\circ .

PROOF. The proof consists of showing all the conditions given in (Bauer and Ninness, 1999), Theorem 5.1. The assumptions stated there include the parametrisation of the linear model structure, which was done in the same way as in the present contribution. The assumptions on the nonlinear model structure are easily verified, as

the function $f(u_t, \alpha)$ clearly is Lipschitz continuous with uniform Lipschitz bound one. Assumption set 3 in (Bauer and Ninness, 1999) includes the assumption of a true model, assumptions on the noise and the input process identical to the present one, except for the condition on the density of the input. The moment bounds are obvious from the nature of the nonlinearities and the uniformity over the parameter set is ensured by the compactness assumption on Λ . Thus it only remains to verify the identifiability condition. Since it is clear that the asymptotic criterion function $\bar{V}(\theta) = \lim_{N \rightarrow \infty} V_T(\theta)$ is minimised at θ_\circ , it is only necessary to show that $\hat{y}_t(\theta_1) = \hat{y}_t(\theta_\circ)$ for all t with probability one if and only of $\theta_1 = \theta_\circ$. The above equality can be reformulated, using the vector $Y_t^k(\theta_i) = [\hat{y}_{t+k}(\theta_i)', \dots, \hat{y}_t(\theta_i)']'$, as $Y_t^k(\theta_1) = Y_t^k(\theta_\circ)$. It follows from the recursions defining the linear dynamical system that

$$\Pi Y_t^k(\theta_i) = \Gamma_i \begin{bmatrix} v_{t+k}(u_{t+k}, \theta_i) \\ \vdots \\ v_t(u_t, \theta_i) \end{bmatrix}.$$

Here Π is a projection onto the orthogonal complement of the span of the columns of the two observability matrices. Γ_i is Π times an upper triangular matrix of rank at least $(k+1-n)p$ and thus Γ_i has rank at least $(k+1-n)p - 2n$. The identifiability condition is equivalent to the condition $Y_t^k(\theta_i) = Y_t^k(\theta_\circ)$, $k > 0$ with probability one. Now assume that $\alpha^1 \neq \alpha^\circ$. If e.g. $\alpha_{1,1}^1 < \alpha_{1,1}^\circ$, due to the assumption on the input density for $k = 3n$ then with nonzero probability we can construct sequences, lying in some set \mathcal{S} say, such that all other variables are saturated and $\alpha_{1,1}^1 < u_{t,1} < \alpha_{1,1}^\circ$. Thus $f(u_{t,1}, \alpha_{1,1}^1) = u_{t,1}$, $f(u_t, \alpha_{1,1}^\circ) = \alpha_{1,1}^\circ$. Then it follows from the nonminimality that on \mathcal{S} the equality $Y_t^{3n}(\theta_1) = Y_t^{3n}(\theta_\circ)$ cannot hold, since there $Y_t^{3n}(\theta_\circ)$ is constant, whereas $Y_t^{3n}(\theta_1)$ is not. This shows the identifiability condition and concludes the proof.

Similar arguments would be possible for strong consistency results involving higher moment conditions on $\{n_t\}$ and $\{u_t\}$ and also a restriction on the size of the n.e.d. It is remarkable that in the nonlinear case, assumptions on the second moments alone are not sufficient for consistency. This can be seen easily by using discrete distributions. The condition on the density is fulfilled (for example) if

$$u_t = \sum_{j=0}^{\infty} K_j e_{t-j}^u$$

where $\{e_t^u\}$ is i.i.d. with a density, where the support is equal to \mathbf{R}^m and K_0 is nonsingular. Thus e.g. if $\{u_t\}$ is an ARMA process driven by Gaussian noise, the condition of the Theorem holds.

In order to show asymptotic normality only a few more assumptions are needed:

Theorem 4. (Asymptotic Normality). Let the assumptions of Theorem 3 hold. Furthermore let θ_o be an interior point of Θ and let the size q of the n.e.d. of both $\{n_t\}$ and $\{u_t\}$ be such that $q > 1$. Let $u_t = e_t^u + \sum_{j=1}^{\infty} K_j e_{t-j}^u$, where e_t^u has density with support \mathbf{R}^m .

Then with the definition of the Hessian matrix

$$R(\theta_o) = \lim_{N \rightarrow \infty} \mathbf{E} \frac{d^2}{d\theta\theta'} \Big|_{\theta=\theta_o} V_N(\theta)$$

it holds that $R(\theta_o) > 0$. Further under these conditions

$$\sqrt{N}(\hat{\theta} - \theta_o) \rightarrow \mathcal{N}(0, P)$$

where convergence is in distribution and $P_o = R(\theta_o)^{-1} Q_o R(\theta_o)^{-1}$ denotes the asymptotic variance matrix.

PROOF. The proof is a consequence of Theorem 6.1. of (Bauer and Ninness, 1999). The assumptions which are needed in excess of the ones of Theorem 3 mostly concern derivatives, which are no issue here, since the first order derivative of $f(u_t, \alpha)$ is piecewise constant, taking on only two values: zero or one. The points of nondifferentiability are a nullset and thus can be neglected. Therefore the second derivative is zero almost everywhere. Here the existence of a density for u_t is essential.

The last statement, which has to be shown is that the asymptotic Hessian $R(\theta_o)$ is nonsingular. To this end consider

$$\mathbf{E} V_N''(\theta) = \mathbf{E} \left(\frac{1}{N} \sum_{t=1}^N \psi_t(\theta_o) \Sigma \psi_t(\theta_o)' + \frac{d^2}{d\theta\theta'} e_t(\theta_o)' \Sigma n_t \right)$$

where the second summand is zero due to the zero mean assumption on n_t and the independence of input and additive noise. Here $\psi_t(\theta_o)$ denotes the matrix, whose j -th column is equal to the transpose of

$$\psi_t^j(\theta_o) = \left[\frac{d}{d\theta_j} e_t(\theta_o) \right]$$

Due to the stationarity assumption it is enough to consider only one term. That is, singularity of $\mathbf{E} V_N''(\theta)$ implies that there exists a vector x such that $x' \psi_t(\theta_o) = 0$ for almost all input sequences. In order to deal with this term, examine the derivative with respect to the truncation bounds more closely. For given input u the function $f(u, \alpha_1, \alpha_2)$ is equal to α_1 for $u < \alpha_1$, equal to u for $u \geq \alpha_1$. Thus the derivative with respect to α_1 is equal to one and zero respectively on the

respective areas. Analogously the derivative with respect to the parameters α_2 can be derived. Since the input has a density, which is nonzero on $[\underline{c}, \bar{c}]$, it is possible to design input sequences with positive probability, which are not saturated in any variable. Therefore all the derivatives with respect to parameters in α are zero. From this it follows from similar arguments as have been used in the proof of Theorem 3 that the summand due to the linear subsystem must be zero, which in turn implies that the parametrisation of the linear subsystem is nonminimal, which is a contradiction. Thus the coordinates of x corresponding to the linear system have to be zero. Similar arguments show the nonsingularity by choosing inputs, which lead to saturations in all but one coordinate. This shows, that $x = 0$ and thus the nonsingularity of $R(\theta_o)$. This finally concludes the proof.

The authors want to emphasize, that the conditions given in the theorem are only sufficient, but not necessary. It seems to be plausible, that weaker assumptions on the input sequence also suffice. This is a topic of further research.

Finally the estimation of P will also be considered. Again we follow the suggestions of (Pötscher and Prucha, 1997). It follows from the proof of the central limit theorem 4 that P is equal to the limit for $N \rightarrow \infty$ of

$$R(\theta_o)^{-1} \left(\frac{1}{N} \mathbf{E} \sum_{s,t=1}^N \psi_t^i(\theta_o)' \Sigma n_t \psi_s^j(\theta_o)' \Sigma n_t \right)_{i,j} R(\theta_o)^{-1}$$

From the definition of $R(\theta_o)$ it follows, that this quantity can be estimated consistently from the data as

$$\frac{d^2}{d\theta\theta'} V_N(\theta) \Big|_{\theta=\hat{\theta}}$$

In fact this is shown in the proof of Theorem 4. Therefore it remains to show the convergence of the expectation given above and to provide a technique to estimate this limit. Thus consider

$$\begin{aligned} & \frac{1}{N} \mathbf{E} \sum_{t=1}^N \psi_t^i(\theta_o)' \Sigma n_t \sum_{s=1}^N \psi_s^j(\theta_o)' \Sigma n_t = \\ & \sum_{l=1-T}^{T-1} \frac{(N-|l|)}{N} \text{Tr} \left[\mathbf{E} \psi_l^j(\theta_o) \psi_0^i(\theta_o)' \Sigma \mathbf{E} n_0 n_l' \Sigma \right] \rightarrow \\ & \sum_{l=-\infty}^{\infty} \text{Tr} \left[\mathbf{E} \psi_l^j(\theta_o) \psi_0^i(\theta_o)' \Sigma \mathbf{E} n_0 n_l' \Sigma \right] \end{aligned}$$

Here the convergence of the sum has to be ensured by conditions on n_t and $\psi_t^i(\theta_o)$ to have covariance sequences tending to zero sufficiently fast. A straightforward idea would be to use the second expression and replace true covariances with estimated ones. However this does not lead to consistent estimators, since the estimates of variances with high lags are known to be very poor. Therefore the accuracy of the above esti-

mator might be increased by introducing weights (see Pötscher and Prucha, 1997; Hjalmarsson *et al.*, 1994). For $l \geq 0$ let

$$\hat{\gamma}_l(\psi, i, j) = \frac{1}{N} \sum_{t=1}^{T-l} \psi_t^i(\hat{\theta}) \psi_{t+l}^j(\hat{\theta})'$$

$$\hat{\gamma}_l(n) = \frac{1}{N} \sum_{t=1}^{T-l} \hat{n}_t(\hat{\theta}) \hat{n}_{t+l}(\hat{\theta})'$$

where $\hat{n}_t(\hat{\theta}) = y_t - \hat{y}_t(\hat{\theta})$. Then consider the estimate

$$\hat{Q}_N = \sum_{l=1}^{T-1} w(l, N) \text{Tr}[\hat{\gamma}_l(\psi, i, j) \Sigma \hat{\gamma}_l(n) \Sigma]$$

The estimate differs from the true quantity given above in that true covariances are replaced with estimates and that the weights $(N - |l|)/N$ are replaced with $w(l, N)$. In the framework of Theorem 4 it can be shown, that this estimate \hat{Q}_N converges in probability to Q_\circ .

Theorem 5. Let the conditions of Theorem 4 hold and let the size of the n.e.d. of $\{n_t\}$ and $\{u_t\}$ be $-2(r-1)/(r-2)$ for some $r > 2$, such that $\mathbf{E}\{\|n_t\|^{2r+\varepsilon}\} < \infty$. Then if $w(l, N)$ satisfies

$$\max_{N \in \mathbf{N}, 0 \leq l \leq N} |w(l, N)| < \infty$$

$$\lim_{N \rightarrow \infty} w(l, N) \rightarrow 1$$

$$\lim_{N \rightarrow \infty} \sum_{j=1}^N l |w(l, N)| < \infty$$

the estimate $\hat{Q}_N \rightarrow Q_\circ$ in probability.

The proof follows in a straightforward fashion from the verification of the assumptions of Theorem 7.1. of (Bauer and Ninness, 1999), which is based on the corresponding theory developed in chapter 13 of (Pötscher and Prucha, 1997).

The question of the choice of the weighting sequence $w(l, N)$ is given a heuristic interpretation in (Pötscher and Prucha, 1997): From the definition of the estimated quantity Q_\circ it follows, that it essentially is the spectrum at frequency zero of a stationary process, which is estimated. Therefore the weighting $w(l, N)$ can be chosen according to the same rules, which govern the choice of spectral estimates.

5. CONCLUSIONS

In this paper the concept of near epoch dependence is applied to a rather simple model structure. It has been shown, that the use of this concept leads to the derivation of the usual asymptotic properties of least squares estimators based

on assumptions on the input and the model structure as well as the existence of a true system rather than on assumptions on the input, output data. The concepts used in this paper seem to be valuable tools for the analysis of nonlinear dynamical systems, as they take care of the fact, that for nonlinear systems different kinds of dependence structures have to be used than only relying on second moments, as is done for linear dynamical systems.

The analysis of the Hammerstein models resulted in the simple fact, that we are able to identify the true model (if it exists) consistently, with asymptotically normal parameter estimates, of which the accuracy can be estimated consistently from the data, based solely on the assumption, that the parameter space is compact (i.e. we have some kind of a priori knowledge on the location of the parameter) and that we are able to construct input sequences, which vary reasonably in this set and are not limited to a finite number of setpoints.

6. REFERENCES

- Bauer, D. and Brett Ninness (1999). Asymptotic properties of least-squares estimates of Hammerstein–Weiner model structures. Technical Report EE9947. Department of Electrical and Computer Engineering, University of Newcastle.
- Hjalmarsson, H., S. Gunnarsson and M. Gevers (1994). A convergent iterative restricted complexity control design scheme. Technical Report LiTH-ISY-I-1653. Department of Electrical Engineering, Linköping University, Sweden.
- L.Ljung (1978). ‘Convergence analysis of parametric identification methods’. *IEEE Transactions on Automatic Control* **AC-23**(5), 770–783.
- Pötscher, B. and I. Prucha (1997). *Dynamic Nonlinear Econometrics Models*. Springer-Verlag. Berlin-Heidelberg.
- Rangan, S., Greg Wolodkin and Kameshwar Poolla (1995). New results for hammerstein system identification. In ‘Proceedings of the 34th IEEE Conference on Decision and Control’. pp. 697–702.
- Suykens, J. and Joos Wandewalle (1998). *Nonlinear Modelling: Advanced black-box techniques*. Kluwer Academic Publishers.
- Tao, G. and Petar V. Kokotović (1996). *Adaptive Control of Systems with Actuator and Sensor Nonlinearities*. Adaptive and Learning Systems for Signal Processing, Communications and Control. John Wiley and Sons.