Asymptotic properties of Hammerstein model estimates

Dietmar Bauer
Institute f. Econometrics, Operations Research and System Theory,
TU Wien, Argentinierstr. 8, A-1040 Wien, Austria,
e-mail: Dietmar.Bauer@tuwien.ac.at

Brett Ninness
Department of Electrical and Computer Engineering,
University of Newcastle, University Drive, Callaghan, NSW 2308, Australia
e-mail: brett@ee.newcastle.edu.au

Abstract

This paper considers the estimation of Hammerstein models. The main result of the paper lies in a specification of a set of sufficient conditions on the input sequence, the noise (and the true system) 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. This setup is utilized in an example, where the static nonlinearity is due to input saturation.
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, e.g. 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 particularily new. For example, (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.

Specifically, whereas (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 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 (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 possibly 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, while section 5 presents the application of these results to the case of hard input saturation. Section 6 finishes with a discussion of the obtained results.

2 Model set and estimation criteria

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

$$\begin{aligned} v_t &= f(u_t, \alpha) \\ x_{t+1} &= Ax_t + Bv_t \\ y_t &= Cx_t + n_t \end{aligned} \tag{1}$$

where

$$\begin{array}{cccc} y_t \in \mathbf{R}^p, & u_t \in \mathbf{R}^{m_u}, & v_t \in \mathbf{R}^{m_v}, & x_t \in \mathbf{R}^n, \\ n_t \in \mathbf{R}^p, & A \in \mathbf{R}^{n\times n}, & B \in \mathbf{R}^{n\times m}, & C \in \mathbf{R}^{p\times n}. \end{array}$$

It will always be assumed that $\alpha \in \Theta_\alpha \subset \mathbf{R}^{d_\alpha}$, where Θ_α is compact. 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 $\beta \in \Theta_\beta$. The parameter set Θ_β is assumed to be compact and $\{(A(\beta), B(\beta), C(\beta)) : \beta \in \Theta_\beta\}$ is assumed to contain only stable transfer functions. We will also use the notation $G(q, \beta) = C(qI - A)^{-1}B$. The direct feedthrough term is only neglected for notational reasons and a term $D(\beta)$ 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.

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

$$\widehat{\theta} \triangleq \underset{\theta\in\Theta}{\arg\min} \frac{1}{N} \mathrm{Tr}\left[\Sigma \sum_{t=1}^{N} e_t(\theta) e_t(\theta)'\right]$$

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

$$\widehat{y}_t(\theta) = G(q, \theta) f(u_t, \alpha)$$

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 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*

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

where $\phi(m)/m^{-r} \to 0, r < q$. 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 1 *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 strictly stationary process such that $\mathbf{E}\{\|g(X_t)\|_2^p\} < \infty$.*

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 $\{[X_t', X_{t-1}', \cdots, X_{t-k}']'\}$ 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 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 $1 < p' < p$.*

vii) *If $\mathbf{E}\{\|X_t\|_2^{p'}\} < \infty$, then $\{X_t\}$ is L_s n.e.d. on $\{e_t\}$ of size $-q$ for all $1 < s < p'$.*

viii) *If $Y_{t+k} = g(Y_t, X_t, \cdots, X_{t+k})$, where*

$$\|g(Y_t, X_t, \cdots, X_{t+k}) - g(\bar{Y}_t, \bar{X}_t, \cdots, \bar{X}_{t+k})\|_2 \leq d_y\|Y_t - \bar{Y}_t\|_2 + d_x\|X_t - \bar{X}_t, \cdots, X_{t+k} - \bar{X}_{t+k}\|_2$$
$$0 < d_y < 1, d_x < \infty$$

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. $N^{-1}\sum_{t=1}^N X_t \to \mathbf{E}X_t$, where convergence is in probability with $N \to \infty$.*

x) *Let $\{X_t\}$ be L_2 n.e.d. on $\{e_t\}$, which is i.i.d., of size -1 and satisfy $\mathbf{E}X_t = 0, \mathbf{E}\{\|X_t\|_2^r\} < \infty$, $r > 2$. 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 $P_N \to P$. Then

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

as $N \to \infty$, 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 $x_{t+k} = A^k x_t + \sum_{j=0}^{k-1} A^j B u_{t+k-j-1}$ 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 laws 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 $\widehat{\theta}$ 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 1 (Consistency) *Assume that the process $\{y_t\}$ is generated by a system of the form (1), where the following assumptions hold:*

1. **Linear system:** *The linear part is assumed to be parametrised by $\beta \in \Theta_\beta$, where Θ_β is compact.*

The family of filters $G(q,\theta)$ is uniformely stable. The true system order n is known.

2. **Static Nonlinearity:** *$f(\cdot,\alpha)$ is parameterised by a vector $\alpha \in \Theta_\alpha \subseteq \mathbf{R}^{d_\alpha}$ where Θ_α is compact. The function $f(u,\alpha)$ is Lipschitz-continuous in its first argument on $\mathbf{R}^m$ ($\|f(u,\alpha)-f(u',\alpha)\| \leq C\|u-u'\|$) and is continuous in its second argument on Θ_α.*

3. **Input:** *$\{u_t\}$ is $L_{2+\varepsilon}$ near epoch dependent of size $-q$ (q arbitrary) on some independent process $\{e_t^u\}$ and is strictly stationary.*

4. **Measurement Noise:** *The measurement noise process $\{n_t\}$ is $L_{2+\varepsilon}$ near epoch dependent of size $-q$ (again q is arbitrary) on some independent process $\{e_t^n\}$ which is also independent of $\{e_t^u\}$, is strictly stationary, and satisfies*

$$\mathsf{E}\{n_t\}=0, \qquad \mathsf{E}\{\|n_t\|^{2+\varepsilon}\}<\infty$$

for some $\varepsilon>0$.

5. **Basis Processes:** *The independent basis processes $\{e_t^u\}$ and $\{e_t^n\}$ satisfy*

$$\mathsf{E}\left\{\|e_t^u\|^{2+\varepsilon'}\right\}<\infty, \qquad \mathsf{E}\left\{\|e_t^n\|^{2+\varepsilon'}\right\}<\infty$$

for some $\varepsilon'>0$.

6. **Effect of Non-linearities:** *For some $\epsilon>0$*

$$\mathsf{E}\left\{\left(\max_{\alpha\in\Theta_\alpha}\|f(u_t,\alpha)\|\right)^{2+\varepsilon}\right\}<\infty \qquad (2)$$

7. **Identifiability:** *Let $\widehat{y}_t(\theta) = G(q,\theta)f(u_t,\alpha)$. Then*

$$\widehat{y}_t(\theta_1)=\widehat{y}_t(\theta_2) \qquad (3)$$

with probability one for all t if, and only if $\theta_1=\theta_2$. Here the probability is introduced by the strictly stationary sequence $\{u_t\}$.

Then the nonlinear least squares estimate $\hat{\theta}$ converges in probability to the true vector $\theta_\circ$.

For a proof see (Bauer and Ninness, 1999). 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 for the input process.

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

Theorem 2 (Asymptotic Normality) *Let the assumptions of Theorem 1 hold. Furthermore let $\theta_\circ$ 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 $q>1$. Additionally we require the following assumptions:*

- *$f(u,\alpha)$ is a u-almost everywhere twice continuously differentiable function of α. It holds that*

$$\left\|\frac{\partial}{\partial\alpha}f(u,\alpha)-\frac{\partial}{\partial\alpha}f(u',\alpha)\right\| \leq C\|u-u'\|.$$

- *The linear dynamical parameterisations $(A(\beta),B(\beta),C(\beta),D(\beta))$ are twice continuously differentiable with respect to β.*

- *There exists a compact neighbourhood $\Theta_\alpha(\alpha_\circ)$ of $\alpha_\circ$ such that for some $\varepsilon>0$:*

$$\begin{aligned}\mathsf{E}\left\{\max_{\alpha\in\Theta_\alpha(\alpha_0)}\left\|\frac{\partial^2}{\partial\alpha^2}f(u_t,\alpha)\right\|^{1+\varepsilon}\right\} &< \infty\\ \mathsf{E}\left\{\max_{\alpha\in\Theta_\alpha(\alpha_0)}\left\|\frac{\partial}{\partial\alpha}f(u_t,\alpha)\right\|^{2+\varepsilon}\right\} &< \infty \quad (4)\end{aligned}$$

- *With the definition of the asymptotic Hessian $R(\theta_\circ)$ as*

$$R(\theta_\circ)\triangleq\lim_{N\to\infty}\mathsf{E}\left\{\frac{\mathrm{d}^2}{\mathrm{d}\theta\theta^T}\bigg|_{\theta=\theta_\circ}V_N(\theta)\right\}$$

it is assumed that $R(\theta_\circ)>0$.

Then

$$\sqrt{N}(\widehat{\theta}-\theta_\circ)\to\mathcal{N}(0,P_\circ)$$

where convergence is in distribution and $P_\circ = R(\theta_\circ)^{-1}Q_\circ R(\theta_\circ)^{-1}$ denotes the asymptotic variance matrix. Here

$$Q_\circ\triangleq\lim_{N\to\infty}N\mathsf{E}\left\{\frac{d}{d\theta}V_N(\theta_\circ)[\frac{d}{d\theta}V_N(\theta_\circ)]^T\right\}$$

The proof is a consequence of Theorem 6.1. of (Bauer and Ninness, 1999). The authors want to emphasize, that the conditions given in the theorem are only sufficient, but not necessary.

Next the estimation of $P_\circ$ will be considered. Again we follow the suggestions of (Pötscher and Prucha, 1997). It follows from the proof of the central limit theorem 2 that $P_\circ$ is equal to the limit for $N\to\infty$ of

$$R(\theta_\circ)^{-1}\left(\frac{1}{N}\mathbf{E}\sum_{s,t=1}^{N}\psi_t^i(\theta_\circ)^T\Sigma n_t\psi_s^j(\theta_\circ)^T\Sigma n_s\right)_{i,j}R(\theta_\circ)^{-1}$$

where $\psi_t^i(\theta) = \frac{d}{d\theta_i}\widehat{y}_t(\theta)$. From the definition of $R(\theta_\circ)$ it follows, that this quantity can be estimated consistently from the data as

$$\frac{d^2}{d\theta\theta^T} V_N(\theta)|_{\theta=\widehat{\theta}}$$

Therefore it remains to provide conditions for the convergence of the expectation given above and to provide a technique to estimate this limit. Thus consider

$$\frac{1}{N}\mathbf{E}\sum_{t=1}^N \psi_t^i(\theta_\circ)^T\Sigma n_t \sum_{s=1}^N \psi_s^j(\theta_\circ)^T\Sigma n_s = \sum_{l=1-N}^{N-1}\frac{(N-|l|)}{N}\mathrm{Tr}\left[\mathbf{E}\psi_l^j(\theta_\circ)\psi_0^i(\theta_\circ)^T\Sigma\mathbf{E}n_0n_l^T\Sigma\right] \to \sum_{l=-\infty}^{\infty}\mathrm{Tr}\left[\mathbf{E}\psi_l^j(\theta_\circ)\psi_0^i(\theta_\circ)^T\Sigma\mathbf{E}n_0n_l^T\Sigma\right]$$

Here the convergence of the sum has to be ensured by conditions on n_t and $\psi_t^i(\theta_\circ)$ 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 estimator might be increased by introducing weights (see Pötscher and Prucha, 1997; Hjalmarsson *et al.*, 1994). For $l \geq 0$ let

$$\begin{aligned}\widehat{\gamma}_l(\psi,i,j) &= \frac{1}{N}\sum_{t=1}^{N-l}\psi_t^i(\widehat{\theta})\psi_{t+l}^j(\widehat{\theta})' \\ \widehat{\gamma}_l(n) &= \frac{1}{N}\sum_{t=1}^{N-l}\widehat{n}_t(\widehat{\theta})\widehat{n}_{t+l}(\widehat{\theta})'\end{aligned}$$

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

$$\widehat{Q}_N = \sum_{l=1-N}^{N-1} w(l,N)\mathrm{Tr}\left[\widehat{\gamma}_l(\psi,i,j)\Sigma\widehat{\gamma}_l(n)\Sigma\right]$$

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 2 it can be shown, that this estimate $\widehat{Q}_N$ converges in probability to $Q_\circ$ under some mild additional assumptions:

Theorem 3 *Let the conditions of Theorem 2 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 $\mathbf{E}\{\|n_t\|^{2r+\varepsilon}\} < \infty$. Assume there exists a compact neighbourhood $\Theta_\alpha(\alpha_\circ)$ of $\alpha_\circ$ such that for some $\varepsilon > 0$.*

$$\begin{aligned}\mathbf{E}\left\{\max_{\alpha\in\Theta_\alpha(\alpha_\circ)}\left\|\frac{\partial^2}{\partial\alpha^2}f(u_t,\alpha)\right\|^{2+\varepsilon}\right\} &< \infty \\ \mathbf{E}\left\{\max_{\alpha\in\Theta_\alpha(\alpha_\circ)}\left\|\frac{\partial}{\partial\alpha}f(u_t,\alpha)\right\|^{2r+\varepsilon}\right\} &< \infty \qquad (5)\end{aligned}$$

Then if $w(l,N)$ satisfies

$$\begin{aligned}\max_{N\in\mathbf{N},0\leq l\leq N}|w(l,N)| &< \infty \\ \lim_{N\to\infty} w(l,N) &\to 1 \\ \lim_{N\to\infty}\sum_{l=1}^N l|w(l,N)| &< \infty\end{aligned}$$

the estimate $\widehat{Q}_N \to Q_\circ$ 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 Application to hard input saturation

In this section we apply the abstract concepts given in the last section to the Hammerstein model structure, where the static nonlinearity is given by hard saturation of unknown level, which is described as follows:

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

for scalar u. This function will be applied componentwise implying that there are $2m_u$ saturation points, which will be denoted as $\alpha_{1,i},\alpha_{2,i}$. 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 $\alpha \in \Theta_\alpha \subset \mathbf{R}^{2m}$ denote the vector of stacked saturation points. Corresponding to the linear parametrisation we introduce the notion of *nonredundancy*: A system $G(q,\beta)$ is called nonredundant, if all columns $G(q,\theta)_j \neq 0$. The reason for this notion is of course, that a redundant system prohibits the identification of the saturation points in at least one coordinate of u_t as the effects of the saturation do not appear in $\widehat{y}_t(\theta)$ in this case. Then we obtain the following result:

Theorem 4 *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* $\{n_t\}$. *The parameter set is assumed to be compact. Corresponding to the input we assume, that the stationary distribution of* $[u_t', \cdots, u_{t-3n}']'$ *has a density, which is strictly greater than zero on* $[-\underline{c} - \varepsilon, \bar{c} + \varepsilon]^{3n+1}$ *for arbitrary* $\varepsilon > 0$. *Also assume that the order of the system* n *is known. Finally assume, that the true system is nonredundant. Then the nonlinear least squares estimate* $\widehat{\theta} \in \Theta$ *converges in probability to the true parameter value* $\theta_\circ$.

Furthermore let $\theta_\circ$ *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* $R(\theta_\circ) > 0$ *and the asymptotic normality holds.*

Finally if the size conditions on n_t *and* u_t *and the moment condition for* n_t *stated in Theorem 3 hold, then the estimate* $\hat{Q}_N$ *converges in probability to* $Q_\circ$ *for all weighting sequences fulfilling the restrictions stated in Theorem 3*

Proof: The proof follows from a verification of the assumptions made in the theorems in the last section. The hard saturation ensures, that all needed moments exist. The assumptions on the inputs are imposed in order to ensure identifiability (consistency part) and nonsingularity of the asymptotic Hessian (asymptotic normality part). $\square$

The theorem imposes by no means minimal conditions. Especially the assumptions on the input might seem questionable. However we stress again, that the main purpose of the example is not to state new results in the estimation context but rather to show, that the framework provided by L_p near epoch dependency leads to convenient results for the asymptotic properties of nonlinear least squares estimates. Essentially the only restrictive assumptions, in the situation where the input can be chosen freely, include the known system order and the strict stationarity. Again we note, that the assumption of stationarity is only introduced for notational simplicity and can be replaced with weaker assumptions.

6 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 with the hard saturation nonlinearity 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 together with mild assumptions on the noise. These assumptions seem to be more readily justifiable than the analogous assumptions given in (Ljung, 1978).

References

Bauer, D. and Brett Ninness (1999). Asymptotic properties of least-squares estimates of Hammerstein–Wiener 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.

Ljung, L. (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.