Proceedings of the
American Control Conference
San Francisco, California • June 1993

Statistical Analysis of Least-Squares Identification for Robust Control Design: Output Error Case With Affine Parametrization

Robert L. Kosut*
Integrated Systems, Inc.
3260 Jay St.
Santa Clara. CA 95054, U.S.A.

Brian D. O. Anderson†
Dept. Of Systems Engineering
The Australian National University
Canberra, ACT 2601, Australia

Abstract Precise, finite-data statistical properties are determined using a *least-squares estimator* based on an output error model with an affine parameter representation where the *true system is of output error form, but is not in the model set.* The purpose of the analysis is to show the effect of unmodeled dynamics on the resulting closed-loop system designed on the basis of the estimated transfer function. This simple problem set-up is prototypical of the interplay between system identification and robust control design.

Introduction

The problem addressed is the following: given a finite collection of sensed sampled input/output data from an unknown system, what level of confidence can be assigned to a feedback controller design or modification.

To make the problem both representative and analytically tractable, the following *a priori qualitative* data is assumed:

(a1) The system which is generating the data is a discrete *linear-time-invariant system in output error* form, i.e.,

$$y_t = (Gu)_t + e_t \tag{1}$$

where t is the sampling time, u and y are the sensed input and output sequences, respectively, and e is an unpredictable output disturbance. The operator G is linear-time-invariant with unknown transfer function $G(z)$ and corresponding impulse response sequence g. Thus,

$$(Gu)_t = \sum_{k=1}^{\infty} g_k u_{t-k} \tag{2}$$

(a2) $G(z)$ is stable, *i.e.*, all the poles of $G(z)$ are strictly inside the unit circle. Hence, there exist positive constants $M \geq 1$ and $\rho < 1$ such that

$$|g_k| \leq M\rho^{k-1}, \forall k \geq 1 \tag{3}$$

*Supported by AFOSR, Directorate of Mathematical and Computer Sciences, Contract No. F49620-90-C-0064, NSF U.S./Australia Cooperative Research Program, Travel Grant INT-9014152, and the Australian National University as a Visiting Fellow.

†Supported by the Cooperative Research Center for Robust and Adaptive Control.

(a3) The unpredictable sequence e is zero-mean gaussian i.i.d. with unknown variance λ_e.

(a4) The input sequence u is deterministic, hence, independent of e.

It is important to emphasize that none of the parameters that appear in the above assumptions are *assumed to be known*; they are *only known* to exist. Hence, there is no *quantitative a priori* knowledge about M, ρ, or λ_e.

The above qualitative assumptions do, however, impose varying degrees of restrictiveness. Assumption (a1) imposes an LTI structure, which by itself is not necessarily restrictive, however, the output error form is very specific. This latter restriction, together with the gaussian assumption (a3) makes the statistical analysis easier without resorting to a central limit theorem or a law of large numbers. Assumption (a4) implies that the system is operating in open-loop, for otherwise u would have a component which is correlated with e.

For control design it is desireable to obtain an estimate of $G(z)$. It is standard practice to form a parametric model $G(z, \theta)$ and estimate the free parameter θ. Although many parametric forms are possible, *e.g.*, [4], for ease of analysis we choose the following *affine* FIR paramtrization:

$$G(z, \theta) = \sum_{k=1}^{n} \theta_k z^{-k} \tag{4}$$

Thus, the problem is to estimate the first n impulse response coefficients $\{g_1, \cdots, g_n\}$. Although we specialize to the FIR modeling case, all the results apply *mutadis mutandis* to any other affine model of $G(z)$, *e.g.*, Laguerre or Kautz models as described in [5]. The essence of the problem addressed here is, in our opinion, the motivation for *the work* described in the recent special issue [6] on system identification for robust control design. In comparison with [2], the smoothness parameters M, ρ are not estimated by modeling the tail of the impulse response $\{g_{n+1}, g_{n+2}, \ldots\}$ as a random variable. Our attempt here is to precisely determine the *effect* of the unmodeled dynamics, *i.e.*, the tail of the impulse response, on a least-squares parameter estimator, without any further prior assumptions.

Least-Squares Estimation

In this section we use least-squares on the measured data to estimate the first n impulse response coefficients $\{g_1, g_2, \ldots\}$ in (2). Towards this end, the unknown impulse response parameters $\{g_1, \ldots, g_L\}$ are partitioned into the (finite) parameter vector to be estimated,

$$\alpha = \begin{bmatrix} g_1 \\ \vdots \\ g_n \end{bmatrix} \in \mathbb{R}^n \quad (5)$$

which consists of the first n impulse response coefficients, and the (infinite) parameter vector

$$\beta = \begin{bmatrix} g_{n+1} \\ g_{n+2} \\ \vdots \end{bmatrix} \in \mathbb{R}^\infty \quad (6)$$

which is the remainder of the impulse response. These parameters – the "tail" of the impulse response, $\{g_{n+1}, g_{n+2}, \ldots\}$ – can significantly bias the estimate of the "head," namely, $\{g_1, \ldots, g_n\}$. Statisticians refer to β as a "nuisance" parameter. Note that because G is stable, $\|\beta\|$ is not only finite, but decreases exponentially as n increases. That is, using (3),

$$\|\beta\|^2 = \sum_{k=n+1}^{\infty} g_k^2 \le \frac{M^2 \rho^{2n}}{1-\rho^2} \quad (7)$$

Using the definition of α and β together with (1) gives,

$$Y = X\alpha + \widetilde{X}\beta + E \quad (8)$$

where

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^N, \; E = \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix} \in \mathbb{R}^N \quad (9)$$

$$X = \begin{bmatrix} u_0 & \cdots & u_{1-n} \\ \vdots & & \vdots \\ u_{N-1} & \cdots & u_{N-n} \end{bmatrix} \in \mathbb{R}^{N\times n} \quad (10)$$

$$\widetilde{X} = \begin{bmatrix} u_{-n} & u_{-n-1} & \cdots \\ \vdots & & \vdots \\ u_{N-n-1} & u_{N-n-2} & \cdots \end{bmatrix} \in \mathbb{R}^{N\times\infty} \quad (11)$$

Assuming that $X'X \in \mathbb{R}^{n\times n}$ is non-singular, *i.e.*, u is persistently exciting of order n, the least-squares estimate of α is given by the well known formula:

$$\widehat{\alpha} = \begin{bmatrix} \widehat{g}_1 \\ \vdots \\ \widehat{g}_n \end{bmatrix} = \arg\min_{\theta\in\mathbb{R}^n} \|Y - X\theta\|^2 = (X'X)^{-1}X'Y \quad (12)$$

where $\{\widehat{g}_k \mid k = 1:n\}$ can be thought of as estimates of $\{g_k \mid k = 1:n\}$. We also take the estimate of λ_e, the output error variance, as the sample-variance,

$$\widehat{\lambda}_e = \frac{1}{N}\|Y - X\widehat{\alpha}\|^2 \quad (13)$$

When $\beta = 0$, it is well known that $\widehat{\alpha}$ and $\widehat{\lambda}_e$ are the maximum likelihood estimates of α and λ_e, respectively, *e.g.*, [1]. In our case, $\beta \neq 0$, and its effect on the estimates is the subject of the next section.

Statistical Analysis

In this section we analyze the effect of the nuisance parameter β on the estimates $\widehat{\alpha}$ and $\widehat{\lambda}_e$ of α and λ_e, respectively. We use the standard notation $\mathcal{N}(\mu, \Sigma)$ to denote a gaussian distribution with mean μ and variance Σ. Likewise, $\chi^2(m)$ denotes a chi-squared distribution with m degrees of freedom. Recall that if $q \in \mathbb{R}^m$ is drawn from $\mathcal{N}(0, R)$ with R non-singular, then $q'R^{-1}q \in \chi^2(m)$. We also use $\chi^2(m, \tau)$ to denote a *non-central* chi-squared distribution with m degrees of freedom and non-centrality parameter τ. To fix the definition of the non-centrality parameter, if $q \in \mathbb{R}^m$ is drawn from $\mathcal{N}(\mu, R)$, then $q'R^{-1}q \in \chi^2(m, \tau)$ with $\tau = \mu'R^{-1}\mu$. From [3], we also use: as either[1] m or $\tau \to \infty$, $\chi^2(m, \tau) \to \mathcal{N}(m + \tau, 2(m + 2\tau))$. Hence, $\chi^2(m, 0) = \chi^2(m)$ and as $m \to \infty$, $\chi^2(m) \to \mathcal{N}(m, 2m)$.

It is convenient to define the "covariance" matrices,[2]

$$\Sigma_{11} = \frac{1}{N}X'X \in \mathbb{R}^{n\times n} \quad (14)$$

$$\Sigma_{12} = \frac{1}{N}X'\widetilde{X} \in \mathbb{R}^{n\times\infty} \quad (15)$$

$$\Sigma_{22} = \frac{1}{N}\widetilde{X}'\widetilde{X} \in \mathbb{R}^{\infty\times\infty} \quad (16)$$

Observe that only Σ_{11} can be formed from the data and by assumption is invertible.

The following theorem describes the distributions of the key random variables.

Theorem 1 *Define the parameter error,*

$$\widetilde{\alpha} = \widehat{\alpha} - \alpha. \quad (17)$$

and the output error,

$$\widehat{E} = Y - X\widehat{\alpha} \quad (18)$$

Under assumptions (a1)-(a4),

(i) The parameter error $\widetilde{\alpha}$ and the residual $\widehat{E}$ are independent and normally distributed as follows:

$$\widetilde{\alpha} \in \mathcal{N}\left(\Sigma_{11}^{-1}\Sigma_{12}\beta, \frac{\lambda_e}{N}\Sigma_{11}^{-1}\right) \quad (19)$$

$$\widehat{E} \in \mathcal{N}\left(\Gamma\widetilde{X}\beta, \lambda_e \cdot \Gamma\right) \quad (20)$$

where $\Gamma \in \mathbb{R}^{N\times N}$, given by,

$$\Gamma = I_N - X(X'X)^{-1}X' \quad (21)$$

has rank $N - n$ and is idempotent, i.e., $\Gamma = \Gamma^2$.

[1] It can be shown that this result is also true if both m or $\tau \to \infty$.

[2] Although the matrices Σ_{12}, Σ_{22} are infinite dimensional, they always appear multiplying β. Hence, these terms are bounded because the elements in β decay exponentially.

(ii) $\frac{N\widehat{\lambda}_e}{\lambda_e}$ *and* $\frac{N}{\lambda_e}\widetilde{\alpha}'\Sigma_{11}\widetilde{\alpha}$ *have the following non-central chi-squared distributions:*

$$\frac{N\widehat{\lambda}_e}{\lambda_e} \in \chi^2\left(N-n,\ \frac{N}{\lambda_e}\delta\right) \quad (22)$$

$$\frac{N}{\lambda_e}\widetilde{\alpha}'\Sigma_{11}\widetilde{\alpha} \in \chi^2\left(n,\ \frac{N}{\lambda_e}\gamma\right) \quad (23)$$

where

$$\gamma = \beta'\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12}\beta \quad (24)$$

$$\delta = \beta'\Sigma_{22}\beta - \gamma = \beta'\left(\Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12}\right)\beta \quad (25)$$

(iii) As $N \to \infty$,

$$\widehat{\lambda}_e \to \mathcal{N}\left(\lambda_e + \delta,\ \frac{2\lambda_e}{N}(\lambda_e + 2\delta)\right) \quad (26)$$

$$\widetilde{\alpha}'\Sigma_{11}\widetilde{\alpha} \to \mathcal{N}\left(\frac{n\lambda_e}{N} + \gamma,\ \frac{2\lambda_e}{N}\left(\frac{n\lambda_e}{N} + 2\gamma\right)\right) \quad (27)$$

The results in part(i) follow directly from the underlying assumptions and definitions of the variables, and except for the non-zero bias terms, are standard, *e.g.*, [1]. Part (ii) is non-standard, in that the error statistics involve non-central chi-square distributions. These results are obtained by direct appeal to the relation between a normally distributed random variable and the non-central chi-squred statistic as stated in the introduction to this section. The asymptotic results in part(iii) follow from the asymptotic normal approximation to a non-central chisquare distribution as stated in the introduction to this section.

In part (iii) of the theorem, the asymptotic variances decay as $1/N$. Hence, for sufficiently large N, the random variable approaches the mean with high probability. This leads directly to the following:

Approximation 1 *For sufficiently large N, the following approximations hold with high probability,*

$$\widehat{\lambda}_e \approx \lambda_e + \delta \quad (28)$$

$$\widetilde{\alpha}'\Sigma_{11}\widetilde{\alpha} \approx \frac{n}{N}\lambda_e + \gamma \quad (29)$$

Observe that for large N, the variance estimate $\widehat{\lambda}_e$ tends to over-estimate the true variance λ_e. In addition, the errors $\widetilde{\alpha}$ and $\widehat{\lambda}_e - \lambda_e$ are driven by the "nuisence" parameter β, *i.e.*, the tail of the impulse response.

A special case of interest is when the input u is white, *i.e.*,

$$\Sigma_{11} = \lambda_u \cdot I_n,\ \Sigma_{12} = 0,\ \Sigma_{22} = \lambda_u \cdot I_\infty \quad (30)$$

Theorem 2 *If u is white, i.e., (30) holds, then:*

$$\frac{N\widehat{\lambda}_e}{\lambda_e} \in \chi^2\left(N-n,\ \frac{N}{\lambda_e}\lambda_u\|\beta\|^2\right) \quad (31)$$

$$\frac{N}{\lambda_e}\lambda_u\|\widetilde{\alpha}\|^2 \in \chi^2(n) \quad (32)$$

In addition, as $N \to \infty$,

$$\widehat{\lambda}_e \to \mathcal{N}\left(\lambda_e + \lambda_u\|\beta\|^2,\ \frac{2\lambda_e}{N}\left(\lambda_e + 2\lambda_u\|\beta\|^2\right)\right) \quad (33)$$

The asymptotic part of the above theorem leads to the following:

Approximation 2 *For sufficiently large N, if u is white, i.e., (30) holds, then with high probability:*

$$\widehat{\lambda}_e \approx \lambda_e + \lambda_u\|\beta\|^2 \quad (34)$$

$$\|\widetilde{\alpha}\|^2 \leq \frac{3n}{N}\frac{\lambda_e}{\lambda_u} \quad (35)$$

Large N and High Probability

When the input is white, "large N" can be taken as,

$$N >> \frac{2(1+2\eta)}{(1+\eta)^2},\quad \eta = \frac{\lambda_u\|\beta\|^2}{\lambda_e} \quad (36)$$

where η is the ratio of the energy in the tail to the output error energy. Typical values of N, *e.g.*, 500-1000, will always be well in excess of variations caused by η. Moreover, from central and non-central chi-square tables (*e.g.*, [3]), values of $N \geq 100$ and $n \geq 20$ make the normal approximations very accurate. In consequence, "high probability" is in excess of 99.95% for typical data lengths and model orders. Similar numbers hold for the general case with a non-white input.

Frequency Response Estimation

The results of the previous section can be used to analyze the errors in frequency response estimation. Towards this end, express $G(z)$, the true transfer function as,

$$G(z) = D(z)'\alpha + \widetilde{D}(z)'\beta \quad (37)$$

where

$$D(z) = \begin{bmatrix} z^{-1} \\ \vdots \\ z^{-n} \end{bmatrix},\quad \widetilde{D}(z) = \begin{bmatrix} z^{-(n+1)} \\ z^{-(n+2)} \\ \vdots \end{bmatrix} \quad (38)$$

Let $\widehat{G}(z)$ denote the transfer function estimate of $G(z)$ defined as

$$\widehat{G}(z) = D(z)'\widehat{\alpha} \quad (39)$$

where $\widehat{\alpha}$ is the least-squares parameter estimate from (12) of the the first n impulse response coefficients of $G(z)$. Let $\Delta(z)$ denote the transfer function error defined as,

$$\Delta(z) = G(z) - \widehat{G}(z) \quad (40)$$

$$= -D(z)'\widetilde{\alpha} + \widetilde{D}(z)'\beta \quad (41)$$

where

$$D(z)'\widetilde{\alpha} = \sum_{k=1}^{n}(\widehat{g}_k - g_k)z^{-k},\quad \widetilde{D}(z)'\beta = \sum_{k=n+1}^{\infty} g_k z^{-k} \quad (42)$$

with $\widetilde{\alpha}$ the parameter error from (17).

From Theorem 1 the following result is obtained.

Theorem 3 *The following results hold at each frequency* ω:

(i) Normal distribution

$$\Delta(e^{j\omega}) \in \mathcal{N}\left(F(e^{j\omega})'\beta,\ \frac{\lambda_e}{N}D(e^{j\omega})^*\Sigma_{11}^{-1}D(e^{j\omega})\right) \quad (43)$$

where

$$F(z)' = \widetilde{D}(z)' - D(z)'\Sigma_{11}^{-1}\Sigma_{12} \quad (44)$$

(ii) Non-central chi-squared distribution

$$\frac{|\Delta(e^{j\omega})|^2}{\frac{\lambda_e}{N}D(e^{j\omega})^*\Sigma_{11}^{-1}D(e^{j\omega})} \in \chi^2(1, \epsilon(\omega)) \quad (45)$$

with non-centrality parameter,

$$\epsilon(\omega) = \frac{|F(e^{j\omega})'\beta|^2}{\frac{\lambda_e}{N}D(e^{j\omega})^*\Sigma_{11}^{-1}D(e^{j\omega})} \quad (46)$$

(iii) Asymptotic Normality
As $N \to \infty$,

$$\frac{|\Delta(e^{j\omega})|^2}{\frac{\lambda_e}{N}D(e^{j\omega})^*\Sigma_{11}^{-1}D(e^{j\omega})} \to \mathcal{N}(1+\epsilon(\omega),\ 2(1+2\epsilon(\omega))) \quad (47)$$

Part (iii) leads to the following result.

Approximation 3 *For sufficiently large N, the following approximation holds with high probability at each frequency ω:*

$$|\Delta(e^{j\omega})|^2 \approx \frac{\lambda_e}{N}D(e^{j\omega})^*\Sigma_{11}^{-1}D(e^{j\omega}) + |F(e^{j\omega})'\beta|^2 \quad (48)$$

Observe that if u is white (30) then

$$\begin{aligned} D(e^{j\omega})^*\Sigma_{11}^{-1}D(e^{j\omega}) &= D(e^{j\omega})^*\left(\frac{1}{\lambda_u}I_n\right)D(e^{j\omega}) \\ &= \frac{1}{\lambda_u}D(e^{j\omega})^*D(e^{j\omega}) = \frac{n}{\lambda_u} \end{aligned}$$

This leads to the following:

Theorem 4 *If u is white, i.e., (30) holds, then at each frequency ω:*

(i) Normal distribution

$$\Delta(e^{j\omega}) \in \mathcal{N}\left(\widetilde{D}(e^{j\omega})'\beta,\ \frac{n}{N}\frac{\lambda_e}{\lambda_u}\right) \quad (49)$$

(ii) Non-central chi-squared distribution

$$\frac{|\Delta(e^{j\omega})|^2}{\left(\frac{n}{N}\frac{\lambda_e}{\lambda_u}\right)} \in \chi^2(1, \epsilon(\omega)) \quad (50)$$

with non-centrality parameter

$$\epsilon(\omega) = \frac{|\widetilde{D}(e^{j\omega})'\beta|^2}{\left(\frac{n}{N}\frac{\lambda_e}{\lambda_u}\right)} \quad (51)$$

(iii) Asymptotic Normality
As $N \to \infty$,

$$\frac{|\Delta(e^{j\omega})|^2}{\left(\frac{n}{N}\frac{\lambda_e}{\lambda_u}\right)} \to \mathcal{N}(1+\epsilon(\omega),\ 2(1+2\epsilon(\omega))) \quad (52)$$

Part (iii) together with Approximation 2 leads to:

Approximation 4 *If u is white, i.e., (30) holds, then for sufficiently large N, the following approximation holds with high probability at each frequency ω:*

$$|\Delta(e^{j\omega})|^2 \approx \frac{n}{N}\frac{\lambda_e}{\lambda_u} + |\widetilde{D}(e^{j\omega})'\beta|^2 \quad (53)$$

Robust Control Analysis

In this section, we use the asymptotic frequency domain bounds to evaluate controller robustness. The goal of control is to reduce the output variance. Consider the LTI feedback controller

$$u = -\widehat{K}y \quad (54)$$

where $\widehat{K}$ stabilizes the "estimated" FIR system

$$y = \widehat{G}u + e,\ \widehat{G}(z) = \sum_{k=1}^{n} g_k z^{-k} \quad (55)$$

Applying the control (54) to the actual system (1) yields the closed-loop system

$$y = \frac{\widehat{T}}{1+\widehat{Q}\Delta}e,\ u = -\frac{\widehat{Q}}{1+\widehat{Q}\Delta}e \quad (56)$$

where

$$\widehat{T} = \frac{1}{1+\widehat{G}\widehat{K}},\ \widehat{Q} = \frac{\widehat{K}}{1+\widehat{G}\widehat{K}} \quad (57)$$

with Δ the estimation error as defined in (40). Since the nominal system is stable, it follows that Δ, $\widehat{T}$, and $\widehat{Q}$ are stable transfer functions. Hence, the closed-loop system is stable if and only if,

$$|1 + \widehat{Q}(e^{j\omega})\Delta(e^{j\omega})| > 0,\ \forall|\omega| \le \pi \quad (58)$$

If this holds, then the spectrum of y, under closed-loop –*not during identification*– is given by:

$$\Phi_y(\omega) = \left|\frac{\widehat{T}(e^{j\omega})}{1+\widehat{Q}(e^{j\omega})\Delta(e^{j\omega})}\right|^2 \lambda_e \quad (59)$$

Suppose that u, *during identification*, is white, i.e., (30) holds. To establish stability, observe that a sufficient condition for stability is that,

$$|\widehat{Q}(e^{j\omega})| \cdot |\Delta(e^{j\omega})| < 1,\ \forall|\omega| \le \pi \quad (60)$$

Using the expression for $|\Delta(e^{j\omega})|$ in Approximation 4 and substituting for λ_s from (34). it follows that for large N, the closed-loop system is stable, with high probability, if,

$$|\widehat{Q}(e^{j\omega})|^2 \left[\frac{3n}{N}\left(\frac{\widehat{\lambda}_e}{\lambda_u} - \|\beta\|^2\right) + |\widetilde{D}(e^{j\omega})'\beta|^2\right] < 1,\ \forall|\omega| \le \pi \tag{61}$$

Hence, using the large N approximations, with high probability, the output spectrum is bounded as follows:

$$\Phi_y(\omega) \le \frac{|\widehat{T}(e^{j\omega})|^2\left(\widehat{\lambda}_e - \lambda_u\|\beta\|^2\right)}{\left(1 - |\widehat{Q}(e^{j\omega})|\left[\frac{3n}{N}\left(\frac{\widehat{\lambda}_e}{\lambda_u} - \|\beta\|^2\right) + |\widetilde{D}(e^{j\omega})'\beta|^2\right]^{1/2}\right)^2} \tag{62}$$

The only unknown quantity is β. From (34), we also know with high probability that,

$$\lambda_e \approx \widehat{\lambda}_e - \lambda_u\|\beta\|^2$$

Since λ_e must be positive, it follows that

$$\|\beta\|^2 \le \widehat{\lambda}_e/\lambda_u \tag{63}$$

provides a worst-case upper bound. Observe that this bound is known because $\widehat{\lambda}_e$ is the computed variance estimate and λ_u is selected by the user as the input variance. As a practical matter, it is unlikely that β will achieve this bound. If it did, then the noise variance $\lambda_e \approx 0$, which for large N, will almost never occur.

Using (3), we get

$$|\widetilde{D}(e^{j\omega})'\beta| = |\sum_{k=n+1}^{\infty} g_k e^{-j\omega k}| \le \frac{M\rho^n}{1-\rho}$$

Hence, for large N, the closed-loop system is stable with high probability if,

$$|\widehat{Q}(e^{j\omega})|^2 \left[\frac{3n}{N}\frac{\widehat{\lambda}_e}{\lambda_u} + \frac{M^2\rho^{2n}}{(1-\rho)^2}\right] < 1,\ \forall|\omega| \le \pi \tag{64}$$

The constants M and ρ are unknown, so in order to evaluate the above robustness condition, either we require *a priori* knowledge or infer the values from the first n impulse response coefficients $\widehat{\alpha}' = [\widehat{g}_1 \cdots \widehat{g}_n]$. That is, define the estimates $\widehat{M}, \widehat{\rho}$ via

$$|\widehat{g}_k| \le \widehat{M}\widehat{\rho}^{k-1},\ \forall k \in [1, n] \tag{65}$$

and replace M, ρ with $\widehat{M}, \widehat{\rho}$. This leads to the robustness test:

$$|\widehat{Q}(e^{j\omega})|^2 \left[\frac{3n}{N}\frac{\widehat{\lambda}_e}{\lambda_u} + \frac{\widehat{M}^2\widehat{\rho}^{2n}}{(1-\widehat{\rho})^2}\right] < 1,\ \forall|\omega| \le \pi \tag{66}$$

Now, suppose that the closed-loop system is stable and the above inequality holds. Then the spectrum of y is bounded, with high probability, by: of y and u are given, respectively, by:

$$\Phi_y(\omega) \le \frac{|\widehat{T}(e^{j\omega})|^2\widehat{\lambda}_e}{\left(1 - |\widehat{Q}(e^{j\omega})|\left[\frac{3n}{N}\frac{\widehat{\lambda}_e}{\lambda_u} + \frac{\widehat{M}^2\widehat{\rho}^{2n}}{(1-\widehat{\rho})^2}\right]^{1/2}\right)^2} \tag{67}$$

The above bound gives an indication of the trade between bias and variance as the model order varies – all results being valid for data length $N \ge 500$ with probability in excess of 99.95

Concluding Remarks

Using an output error linear plant, we have shown that with gaussian noise and affine models, there is a very rich structure in the analysis of standard least-squares estimation of the first n impulse response coefficients. The remaining coefficients bias the estimate in a precisely defined way involving non-central chi-squared statistics. These appear to be extremely useful in predicting model error for robust control design from finite data records. Much still remains to be done even for this restricted and analytically tractable case, particularly in finding a means to bound the effect of the bias (the tail of the impulse response) without having to perform additional identification with ever larger parameter orders. This ultimately may involve additional *a priori* quantitative knowledge. We feel that this paper indicates a first step towards the more difficult problem of model structures which account for non-white noise, *e.g.*, ARX or ARMAX models.

*

References

[1] G.C. Goodwin and R.L. Payne, *Dynamic System Identification: Experiment Design and data Analysis*, Academic Press, New York, 1977.

[2] G.C. Goodwin, M. Gevers, and B. Ninness, "Quantifying the error in estimated transfer functions with applications to model order selection," *IEEE Trans. on Automatic Control*, vol. 37, no.7, pp. 913-928, July 1992.

[3] N.L. Johnson and S. Kotz, *Continuous Univariate Distributions*, 1970.

[4] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Inc., 1987.

[5] B. Wahlberg, "System identification using Laguerre models," *IEEE Trans. on Automatic Control*, vol. 36, no.5, pp. 551-562, 1991.

[6] "Special Issue on System Identification for Robust Control Design," edited by R.L. kosut, G.C. Goodwin, and M.P. Polis, *IEEE Trans. on Automatic Control*, vol. 37, no.7, pp. 899-1008, July 1992.