Variance Error Quantifications that are Exact for Finite Model Order

Brett Ninness*
School of of Electrical Engineering &
Computer Science, University of Newcastle,
Australia.
brett@ee.newcastle.edu.au

Håkan Hjalmarsson**
Department of Sensors, Signals and Systems,
The Royal Institute of Technology, S-100 44
Stockholm, Sweden.
hakan.hjalmarsson@s3.e.kth.se

***Abstract*— This paper is concerned with the frequency domain quantification of noise induced errors in dynamic system estimates. Preceding seminal work on this problem provides general expressions that are approximations whose accuracy increases with observed data length *and* model order. In the interests of improved accuracy, this paper provides new expressions whose accuracy depends *only* on data length. They are therefore 'exact' for arbitrarily small true model order and apply to the general cases of Output-Error and Box–Jenkins model structures.**

I. Introduction

If the widely used prediction-error method with a quadratic criterion is employed [4], then a seminal result is that under open-loop conditions the noise-induced estimation error, as measured by the variability of the ensuing frequency response estimate $G(e^{j\omega}, \widehat{\theta}_N^n)$, obeys [6], [9], [4], [8]

$$\lim_{m\to\infty}\lim_{N\to\infty}\frac{N}{m}\mathrm{Var}\{G(e^{j\omega},\widehat{\theta}_N^n)\} = \frac{\Phi_\nu(\omega)}{\Phi_u(\omega)}. \qquad (1)$$

Here Φ_ν and Φ_u are, respectively, the measurement noise and input excitation power spectral densities, and $\widehat{\theta}_N^n$ is the prediction error estimate based on N observed data points of a vector $\theta^n \in \mathbf{R}^n$ that parameterises a model structure $G(q, \theta^n)$ for which (essentially) the model order $m = \dim \theta^n/(2^d)$ where d is the number of denominator polynomials to be estimated in the model structure.

Although this result is asymptotic in both data length N and model order m, it suggests the very well known approximation for finite data and model order of

$$\mathrm{Var}\{G(e^{j\omega},\widehat{\theta}_N^n)\} \approx \frac{m}{N}\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)}. \qquad (2)$$

Apart from its simplicity, a key factor underlying the importance and popularity of the quantification (2) is that, according to its derivation [6], [9], [4], [8], the expression (1) applies for a very wide class of so-called 'shift invariant' model structures. For example, all the well known FIR, ARX, ARMAX, Output–Error and Box–Jenkins structures are shift invariant [6].

However, a fundamental aspect of the approximation (2) is that (since it is derived from (1) which is asymptotic in model order m) its accuracy for realistic finite model orders is not guaranteed [2], [13].

This paper extends these and related pre-existing results appearing in [16], [14], [15], [12], [11] for the fixed denominator FIR-like model class to the much more general cases of Output-Error and Box–Jenkins modelling. In all cases, and in the interests of accuracy, expressions are derived here that hold exactly for finite model orders.

Among others, a key conclusion arising from these new results is that, while the recent work [16] suggests that when substantial errors occur in the quantification (2) they are due to special factors such as '*erroneous noise models, coloured inputs and fixed poles*', the new expressions developed here establish that these substantial errors in (2) are a much more general phenomenon that are not predicated on these special factors.

Underpinning the work in this paper is a new approach to the variance quantification issue that involves the use of what is known as a 'reproducing kernel' for a space.

Finally, the expanded companion paper [10] addresses certain extensions to this work regarding how the model class relates to the true underlying system. Furthermore, all proofs for the results presented here may be found in [10].

II. Motivation

In the interests of illustrating the practical (and theoretical) significance of the analysis undertaken here, we provide a brief illustrative simulation example in which the following simple system $G(q) = 0.05/(q - 0.95)$ is used to generate an $N = 10000$ sample input-output data record where the output $\{y_t\}$ is corrupted by white Gaussian noise of variance $\sigma^2 = 10$, and where the input $\{u_t\}$ is a realisation of a stationary, zero mean, unit variance white Gaussian process.

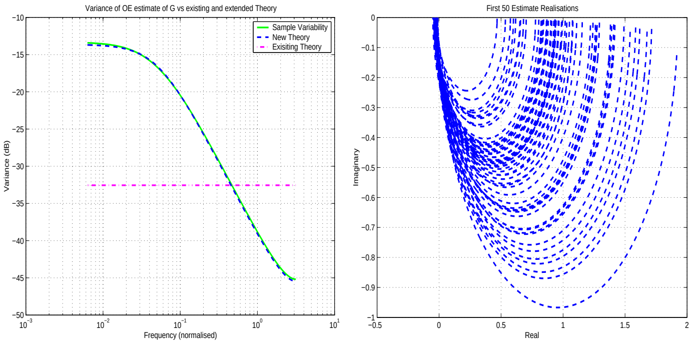
On the basis of this observed data, a first order Output-Error model structure $G(q, \widehat{\theta}_N^n)$ is estimated, and the sample mean square error in this estimate over 1000 estimation experiments with different input and noise realisations is used as an estimate of $\mathrm{Var}\{G(e^{j\omega}, \widehat{\theta}_N^n)\}$ which is plotted as a solid line in Figure 1(a).

The 'classical' approximation (2) is shown as a dash-dot line in that same figure, and is clearly a poor approximation to the true variability.

In particular, it is quantitatively misleading in a rather dramatic way, in that it is inaccurate at low and high frequencies by orders of magnitude. If it were used to perform the very common procedure of judging the radius of error bounds on estimated Nyquist plots, then those radii would be approximately one hundred times too small near zero frequency, and more than ten times too large at the folding frequency.

In recognition of this, a main contribution of this paper is to establish that under certain assumptions on the input and

*This work was supported by the Australian Research Council.
**Work by this author done while visiting University of Newcastle.



(a) True Variability vs. Theory

(b) Sample Estimate Realisations

Fig. 1. *Figures illustrating variability of Output–Error Estimates: Figure (a) shows the true variability vs. theoretically derived approximations. There the solid line is the Monte–Carlo estimate of the true variability, the dash-dot line is the pre-existing approximation (2) which does not account for system poles or model structure. The dashed line is the new approximation (subsequently) presented in this paper whereby estimated system pole positions $\{\xi_1, \cdots, \xi_m\}$ and the fact that an Output–Error structure is employed are both accounted for. Figure (b) shows the first 50 (of 1000) estimate realisations to give a sense of the scale of the variability being quantified in Figure (a).*

measurement noise, then in the Output-Error case considered in this section

$$\lim_{N\to\infty} N\cdot\mathrm{Var}\{G(e^{j\omega},\widehat{\theta}^n_N)\} = 2\cdot\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)}\sum_{k=1}^{m}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2} \quad (3)$$

where the $\{\xi_k\}$ are the poles of the underlying true system. Note that, like the recent results in [16], where more restrictive model structures with fixed denominator were considered, the expression (3) also applies for finite model order m, and hence the ensuing approximation suggested in this paper of

$$\mathrm{Var}\{G(e^{j\omega},\widehat{\theta}^n_N)\} \approx \frac{2}{N}\cdot\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)}\sum_{k=1}^{m}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2} \quad (4)$$

does not depend on the model order m being large. This is in contrast to the work underlying the well known pre-existing approximation (2).

In a sense then, the right hand side of (4) is 'exact' for finite model order m, as illustrated by the dashed line in Figure 1(a) which is the expression (4) and exactly matches the true variability shown as the solid line.

Finally, Figure 1(b) illustrates the first fifty (of one thousand) estimate realisations (represented via the corresponding Nyquist plot of $G(e^{j\omega},\widehat{\theta}^n_N)$) that are averaged to produce an estimate of the true variability shown as the solid line in Figure 1(a). This is shown to emphasise the possible large scale of the error that can be accurately quantified by the results of this paper. That is, the results here are not restricted to the evaluation of minor effects.

III. Technical Preliminaries

In this paper, the idea of what is called a 'Reproducing Kernel' for a space will prove to be a vital tool that allows for the direct simplification of complicated quantities via what is essentially a geometric principle. In relation to this, consider a subspace X_n of L_2 defined by a sequence $\{g_k\}$ of elements of L_2 as

$$X_n \triangleq \mathrm{Span}\{g_1,\cdots,g_n\}. \quad (5)$$

The so-called 'reproducing kernel' $\varphi_n(\lambda,\omega) : [-\pi,\pi]\times[-\pi,\pi]\to\mathbf{C}^{p\times p}$ for this space X_n of $\mathbf{C}^p$ valued functions is an entity such that for any $\alpha\in\mathbf{C}^p$ [1],

$$f_\omega(\cdot) \triangleq \varphi_n(\cdot,\omega)\alpha \in X_n \qquad \forall\omega\in[-\pi,\pi] \quad (6)$$

and for any $f\in X_n$

$$\langle f(\cdot),\varphi_n(\cdot,\omega)\alpha\rangle = \alpha^\star f(\omega). \quad (7)$$

Although the reproducing kernel φ for a space is unique, there may be multiple ways of expressing it, one of which is essential to this paper.

Lemma 3.1 (Expressions for the reproducing kernel): Consider the subspace $X_n \subseteq L_2$ defined via (5) with alternative basis $\{\mathcal{B}_k\}$

$$X_n = \mathit{Span}\{\mathcal{B}_1(\lambda),\cdots,\mathcal{B}_n(\lambda)\} \quad (8)$$

which is orthonormal in that $\langle\mathcal{B}_k,\mathcal{B}_\ell\rangle = \delta(k-\ell)$ where the latter is the Kronecker delta function. Then the reproducing kernel $\varphi_n(\lambda,\omega)$ on X_n may be expressed as

$$\varphi_n(\lambda,\omega) = \sum_{k=1}^{n}\mathcal{B}_k(\lambda)\mathcal{B}_k^\star(\omega). \quad (9)$$

■

In what follows, we will mainly be concerned with functions $f(\omega) : [-\pi,\pi]\to\mathbf{C}^p$ that arise as the restriction of $f(z) : \mathbf{C}\to\mathbf{C}^p$ to the domain $z = e^{j\omega}$, $\omega\in[-\pi,\pi]$. As such, the paper will alternate between notation $f(\omega)$, $f(e^{j\omega})$ and $f(z)$ as convenient.

The relevance of these reproducing kernel ideas to the problem of quantification of variance error for frequency function estimates stems from the fact that when the prediction errors are white, the associated variance error in the frequency domain can be recognised as being the reproducing kernel for a particular space that depends on the model structure and the spectral density of the input.

Therefore, since the preceding lemma also establishes that the reproducing kernel, which is unique, can be expressed as (9), then this provides a means for *exact* quantification of variance error provided that an explicit expression for the orthonormal basis $\mathcal{B}_k(z)$ spanned by the elements of $\Psi(z)$ can be found.

With this is in mind, in the scalar case $p=1$, the following lemma details an important situation in which an explicit formulation of the necessary orthonormal basis is available.

Lemma 3.2 (Orthonormal basis: Fixed denominator Space): Consider the space

$$X_n \triangleq \mathit{Span}\left\{\frac{z^{-1}}{L_n(z)},\frac{z^{-2}}{L_n(z)},\cdots,\frac{z^{-m}}{L_n(z)}\right\} \quad (10)$$

where

$$L_n(z) = \prod_{k=1}^{n}(1-\xi_k z^{-1}), \qquad |\xi_k|<1 \tag{11}$$

for some set of specified poles $\{\xi_1,\cdots,\xi_n\}$ *and where* $m \geq n$*. Then it holds that*

$$X_n = \mathit{Span}\,\{\mathcal{B}_1(z),\cdots,\mathcal{B}_n(z)\} \tag{12}$$

where

$$\mathcal{B}_k(z) \triangleq \frac{\sqrt{1-|\xi_k|^2}}{z-\xi_k}\cdot\phi_{k-1}(z),\ \phi_k(z)\triangleq\prod_{\ell=1}^{k}\frac{1-\overline{\xi_\ell}z}{z-\xi_\ell} \tag{13}$$

where $\phi_0(z)\triangleq 1$ *and and with* $\xi_k = 0$ *for* $k = n+1,\ldots,m$*. Furthermore, the functions* $\{\mathcal{B}_k(z)\}$ *defined in (13) satisfy* $\langle\mathcal{B}_k,\mathcal{B}_\ell\rangle = \delta(k-\ell)$ *and hence form an orthonormal basis.* ■

IV. Problem Formulation

With the necessary technical tools established, this paper now proceeds to precisely define the format of the estimation problem considered here and to provide the links between them and the preceding material.

Here, it is supposed that the relationship between an observed input data record $\{u_t\}$ and output data record $\{y_t\}$ is modelled according to

$$\mathcal{M}: \quad y_t = G(q,\theta^n)u_t + H(q,\theta^n)e_t \tag{14}$$

where the 'dynamics model' $G(q,\theta^n)$ and the 'noise model' $H(q,\theta^n)$ are jointly parameterised by a vector $\theta^n\in\mathbf{R}^n$ and are of the rational forms ($A(q,\theta^n) - D(q,\theta^n)$ below are all polynomials in the backward shift operator q^{-1})

$$G(q,\theta^n) = \frac{B(q,\theta^n)}{A(q,\theta^n)}, \qquad H(q,\theta^n) = \frac{C(q,\theta^n)}{D(q,\theta^n)} \tag{15}$$

while $\{e_t\}$ in (14) is a zero-mean white noise sequence that satisfies $\mathbf{E}\{e_t^2\} = \sigma^2, \mathbf{E}\{|e_t|^8\}<\infty$.

The postulated relationship (14) can encompass a range of model structures such as FIR, ARMAX, 'Output–Error' and 'Box–Jenkins' [4]. For all these cases, since $H(q,\theta^n)$ is also constrained to be monic (i.e. $\lim_{|q|\to\infty} H(q,\theta^n) = 1$), for all θ, then the mean-square optimal one-step ahead predictor $\widehat{y}_t(\theta^n)$ based on the model structure (14) is [4]

$$\widehat{y}_t(\theta^n) = H^{-1}(q,\theta^n)G(q,\theta^n)u_t + \left[1 - H^{-1}(q,\theta^n)\right]y_t \tag{16}$$

with associated prediction error

$$\varepsilon_t(\theta^n) \triangleq y_t - \widehat{y}_t(\theta^n) = H^{-1}(q,\theta^n)\left[y_t - G(q,\theta_n)u_t\right]. \tag{17}$$

Using this, a quadratic estimation criterion may be defined as

$$V_N(\theta^n) = \frac{1}{2N}\sum_{t=1}^{N}\varepsilon_t^2(\theta^n) \tag{18}$$

and then used to construct the prediction error estimate $\widehat{\theta}_N^n$ of θ^n as

$$\widehat{\theta}_N^n \triangleq \arg\min_{\theta^n\in\mathbf{R}^n} V_N(\theta^n). \tag{19}$$

Forming system estimates via the formulation (14)-(19) has become quite standard, in large part due to the availability of sophisticated software tools (MATLAB ID Toolbox), but also because of its statistical efficiency and further extensive theoretical understanding of the properties of such an approach [4].

For example, as has been established in [5], [4], under certain mild assumptions on the nature of the input $\{u_t\}$, the estimate $\widehat{\theta}_N^n$ converges with increasing N according to

$$\lim_{N\to\infty}\widehat{\theta}_N^n = \theta_\circ^n \triangleq \arg\min_{\theta^n\in\mathbf{R}^n}\lim_{N\to\infty}\mathbf{E}\{V_N(\theta^n)\} \qquad \text{w.p.1.} \tag{20}$$

As well, it also holds that as N increases, the estimate $\widehat{\theta}_N^n$ converges in law to a Normally distributed random variable with mean value $\theta_\circ^n$ according to [7], [4]

$$\sqrt{N}(\widehat{\theta}_N^n - \theta_\circ^n) \xrightarrow{\mathcal{D}} \mathcal{N}(0,P_n) \qquad \text{as } N\to\infty \tag{21}$$

and furthermore, under the added assumption of $\mathbf{E}\left\{|e_t|^8\right\}<\infty$ then as established in [4, Appendix 9B]

$$\lim_{N\to\infty}\text{Var}\{\widehat{\theta}_N^n - \theta_\circ^n\} = P_n. \tag{22}$$

The matrix P_n, which gives a measure of parameter space estimation error, is of central importance to this paper. Its formulation is, in general, problem specific, but in the particular case of the model structure (14) being rich enough to encompass any true underlying dynamics [4]

$$P_n^{-1} = \frac{1}{\sigma^2}\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_t^T(\theta_\circ^n)\right\} \tag{23}$$

where for some matrix of transfer functions $\Pi(q,\theta^n)$, and some quasi-stationary (possibly vector valued) signal $\zeta_t(\theta^n)$

$$\psi_t(\theta^n) \triangleq -\frac{\mathrm{d}}{\mathrm{d}\theta^n}\widehat{y}_t(\theta^n) = -H^{-1}(q,\theta^n)\frac{\mathrm{d}\Pi(q,\theta^n)}{\mathrm{d}\theta^n}\zeta_t(\theta^n). \tag{24}$$

Unfortunately, while this explicit formulation of P_n exists, in general it does not provide significant insight into how various design variables affect the accuracy of the estimated frequency functions $G(e^{j\omega},\widehat{\theta}_N^n)$ and $H(e^{j\omega},\widehat{\theta}_N^n)$. In response to this, the seminal work [3], [6], [9], [4], [8] has used an approach of investigating how (21) manifests itself in the variability $\Delta_n(\omega)$ of $G(e^{j\omega},\widehat{\theta}_N^n)$ and $H(e^{j\omega},\widehat{\theta}_N^n)$; the result being approximations such as (2).

Central to the contribution of this paper is the novel approach of recognising that the problem of quantifying $\Delta_n(\omega)$ is closely related to the problem of quantifying the reproducing kernel for a certain space X_n which is defined via the rows of the matrix (θ^n assumed to be a column vector)

$$\Psi(z,\theta_\circ^n) \triangleq H^{-1}(z,\theta_\circ^n)\left.\frac{\mathrm{d}\Pi(z,\theta^n)}{\mathrm{d}\theta^n}\right|_{\theta^n=\theta_\circ^n} S_{\zeta_\circ}(z) \tag{25}$$

according to

$$X_n \triangleq \text{Span}\left\{[\Psi(z,\theta_\circ^n)]_1^T,\cdots,[\Psi(z,\theta_\circ^n)]_n^T\right\} \tag{26}$$

and where, in (25), the term $S_{\zeta_\circ}(z)$ is the spectral factor associated with the process $\{\zeta_t(\theta_\circ^n)\}$.

To make this more concrete, and for future use in the following sections, the space X_n for certain important model structures is provided in the following lemmas.

Lemma 4.1 (Characterization of space – Box-Jenkins): Suppose that the model structure (14) is parameterised with polynomials of the form

$$A(q,\theta^n) = 1 + a_1 q^{-1} + a_2 q^{-1} + \cdots + a_{m_a} q^{-m_a}, \quad (27)$$
$$B(q,\theta^n) = b_1 q^{-1} + b_2 q^{-1} + \cdots + b_{m_b} q^{-m_b}, \quad (28)$$
$$D(q,\theta^n) = 1 + d_1 q^{-1} + d_2 q^{-1} + \cdots + d_{m_d} q^{-m_d}, \quad (29)$$
$$C(q,\theta^n) = 1 + c_1 q^{-1} + c_2 q^{-1} + \cdots + c_{m_c} q^{-m_c}, \quad (30)$$

for some integers m_a, m_b, m_c, m_d. Then (24) holds with

$$\Pi(q,\theta^n) = [G(q,\theta^n), H(q,\theta^n)], \qquad \zeta_t(\theta^n) = \begin{bmatrix} u_t \\ \varepsilon_t(\theta^n) \end{bmatrix} \quad (31)$$

and therefore, with $F(z)$ being the spectral factor of the input spectrum of $\Phi_u(\omega)$, and in the case where $\Phi_{ue} \equiv 0$, the space X_n defined in (26),(25) may be expressed as

$$X_n = \mathit{Span}\left\{f_1(z), \cdots, f_{m_a+m_b}(z), g_1(z), \cdots, g_{m_c+m_d}(z)\right\} \quad (32)$$

where

$$f_k(z) \triangleq \begin{bmatrix} \dfrac{z^{-k}F(z)}{A^2(z,\theta^n_\circ)H(z,\theta^n_\circ)}, & 0 \end{bmatrix}^T \quad (33)$$

$$g_k(z) \triangleq \begin{bmatrix} 0, & \dfrac{z^{-k}}{C(z,\theta^n_\circ)D(z,\theta^n_\circ)} \end{bmatrix}^T. \quad (34)$$

■

Lemma 4.2 (Characterization of space – Output-Error): Suppose that the model structure (14) is parameterised with the numerator and denominator polynomials of the form (27), (28) and

$$C(q,\theta^n) = D(q,\theta^n) = 1 \quad (35)$$

for some integers m_a, m_b. Then (24) holds with

$$H(q,\theta^n_\circ) = 1, \qquad \Pi(q,\theta^n) = G(q,\theta^n), \qquad \zeta_t(\theta^n) = u_t \quad (36)$$

and therefore, with $F(z)$ being the spectral factor of the input spectrum of $\Phi_u(\omega)$, the space X_n defined in (26),(25) may be expressed as

$$X_n = \mathit{Span}\left\{\frac{z^{-1}F(z)}{A^2(z,\theta^n_\circ)}, \frac{z^{-2}F(z)}{A^2(z,\theta^n_\circ)}, \ldots, \frac{z^{-(m_a+m_b)}F(z)}{A^2(z,\theta^n_\circ)}\right\}. \quad (37)$$

■

Finally, in what follows it will also be important to consider the relationship between the true underlying system $\mathcal{S}$ given as

$$\mathcal{S}: \quad y_t = G(q)u_t + \nu_t, \qquad \nu_t = H(q)e_t \quad (38)$$

and the model structure $\mathcal{M}$ defined by (14).

In this paper, it is assumed that $\mathcal{S} \in \mathcal{M}$, hence $\varepsilon_t(\theta^n_\circ) = e_t$ and $G(q,\theta^n_\circ) = G(q), H(q,\theta^n_\circ) = H(q)$, in which case the latter shorter notation will often be used together with

$$\Phi_\nu(\omega) = \sigma^2 |H(e^{j\omega})|^2 \quad (39)$$

representing the power spectral density of the process $\{\nu_t\}$. The case of $\mathcal{S} \notin \mathcal{M}$ is treated in the extended companion paper [10].

V. Main Results

With the necessary technical tools and problem formulation established, the paper now proceeds to present the main results which are new quantifications that are 'exact' for finite model order and permit analysis of the phenomenon illustrated in Figure 1.

The central result to be employed here finally makes completely explicit the link between variance error and reproducing kernels, as hinted at in the previous sections. It applies for all cases encompassed by the model structure (14), including those of FIR, ARMAX, OE and Box–Jenkins type.

Theorem 5.1 (Frequency Domain Variability – $\mathcal{S} \in \mathcal{M}$): Suppose that $\widehat{\theta}^n_N$ is calculated via (19) using the model structure (14) and that the following assumptions are satisfied

1) *$\varepsilon_t(\theta^n_\circ) = e_t$ where $\{e_t\}$ is a zero mean i.i.d. process that satisfies $\mathbf{E}\left\{|e_t|^8\right\} < \infty$;*
2) *The relationship (24) holds for some $\Pi(q,\theta^n)$, and some quasi-stationary (possibly vector valued) signal $\{\zeta_t(\theta^n)\}$ and for which the power spectral density $\Phi_{\zeta_\circ}(\omega)$ of $\{\zeta_t(\theta^n_\circ)\}$ satisfies $\Phi_{\zeta_\circ}(\omega) > 0$;*
3) *Neither of $G(z,\theta^n_\circ)$ or $H(z,\theta^n_\circ)$ contain any pole-zero cancellations.*

Then denoting $S_{\zeta_\circ}(z)$ as the spectral factor of the power spectral density $\Phi_{\zeta_\circ}(\omega)$ of $\{\zeta_t(\theta^n_\circ)\}$

$$\lim_{N\to\infty} N \cdot \mathit{Cov}\left\{\begin{bmatrix} G(e^{j\omega},\widehat{\theta}^n_N) \\ H(e^{j\omega},\widehat{\theta}^n_N) \end{bmatrix}\right\} = \Delta_n(\omega) \quad (40)$$

where

$$\Delta_n(\omega) = \Phi_\nu(\omega)\, S^{-\star}_{\zeta_\circ}(e^{j\omega})\, \varphi_n(\omega,\omega)\, S^{-1}_{\zeta_\circ}(e^{j\omega}) \quad (41)$$

with $\varphi_n(\lambda,\omega)$ being the reproducing kernel for the space X_n defined via (25), (26).

■

The problem of deriving an explicit expression $\Delta_n(\omega)$ for the estimate covariance in the frequency domain is therefore established as being equivalent to the problem of finding an explicit expression for the reproducing kernel for a certain space X_n which is that spanned by the rows of $\Psi(z,\theta^n_\circ)$ defined in (25). This space, and hence the reproducing kernel $\varphi_m(\lambda,\omega)$, depends on the model structure employed. Therefore the covariance $\Delta_n(\omega)$ of the dynamic system estimate, will also depend on the model structure as will now be made explicit via the following corollaries to this main theorem.

Corollary 5.1 (Variability of Box–Jenkins model $\mathcal{S} \in \mathcal{M}$): Suppose that all the conditions of Lemma 4.1 are satisfied and hence that a Box–Jenkins model structure is employed. Suppose further that the conditions of Theorem 5.1 are satisfied and that

$$A_\dagger(z) = A^2(z,\theta^n_\circ)\frac{H(z,\theta^n_\circ)}{F(z)} \quad (42)$$

is a polynomial in z^{-1} of degree at most $m_a + m_b$. Define the zeros $\{\xi_k\}$ and $\{\eta_k\}$ according to

$$z^{m_a+m_b} A_\dagger(z) = \prod_{k=1}^{m_a+m_b} (z - \xi_k),$$

$$z^{m_c+m_d} C(z, \theta_\circ^n) D(z, \theta_\circ^n) = \prod_{k=1}^{m_c+m_d} (z - \eta_k) \tag{43}$$

and use these to define the functions $\kappa(\omega)$ and $\widetilde{\kappa}(\omega)$ according to

$$\kappa(\omega) \triangleq \sum_{k=1}^{m_a+m_b} \frac{1 - |\xi_k|^2}{|e^{j\omega} - \xi_k|^2}, \qquad \widetilde{\kappa}(\omega) \triangleq \sum_{k=1}^{m_c+m_d} \frac{1 - |\eta_k|^2}{|e^{j\omega} - \eta_k|^2}. \tag{44}$$

Then

$$\lim_{N\to\infty} N \,\mathrm{Cov}\left\{ \left[\begin{array}{c} G(\widehat{\theta}_N^n) \\ H(\widehat{\theta}_N^n) \end{array} \right] \right\} = \Phi_\nu(\omega) \left[\begin{array}{cc} \dfrac{\kappa(\omega)}{\Phi_u} & 0 \\ 0 & \dfrac{\widetilde{\kappa}(\omega)}{\sigma^2} \end{array} \right]. \tag{45}$$

Furthermore, regardless of whether (42) is satisfied for some polynomial $A_\dagger(z)$

$$\lim_{N\to\infty} N \,\mathit{Var}\left\{ H(e^{j\omega}, \widehat{\theta}_N^n) \right\} = \widetilde{\kappa}(\omega) |H(e^{j\omega})|^2. \tag{46}$$

■

Corollary 5.2 (Variability of Output-Error model $\mathcal{S} \in \mathcal{M}$): Suppose that all the conditions of Lemma 4.2 are satisfied and hence that an Output-Error model structure is employed. Suppose further that the conditions of Theorem 5.1 are satisfied and that

$$A_\dagger(z) = \frac{A^2(z, \theta_\circ^n)}{F(z)} \tag{47}$$

is a polynomial in z^{-1} of degree at most $m_a + m_b$. Define the zeros $\{\xi_k\}$ and the function $\kappa(\omega)$ according to

$$z^{m_a+m_b} A_\dagger(z) = \prod_{k=1}^{m_a+m_b} (z - \xi_k), \;\; \kappa(\omega) \triangleq \sum_{k=1}^{m_a+m_b} \frac{1 - |\xi_k|^2}{|e^{j\omega} - \xi_k|^2}. \tag{48}$$

Then

$$\lim_{N\to\infty} N \cdot \mathrm{Var}\left\{ G(e^{j\omega}, \widehat{\theta}_N^n) \right\} = \sigma^2 \cdot \frac{\kappa(\omega)}{\Phi_u(\omega)}. \tag{49}$$

■

The essential implications of these corollaries are that although the equality in the variance expressions (45), (49) depend on N being infinitely large, it could be expected that for finite N the convergence results (45), (49) should still hold approximately, so that the following quantifications are useful

$$\mathbf{E}\left\{ \left| G(\widehat{\theta}_N^n) - G \right|^2 \right\} \approx \frac{1}{N} \, \frac{\widehat{\sigma}^2 |H(\widehat{\theta}_N^n)|^2}{\Phi_u(\omega)} \sum_{k=1}^{m_a+m_b} \frac{1 - |\xi_k|^2}{|e^{j\omega} - \xi_k|^2}, \tag{50}$$

$$\mathbf{E}\left\{ \left| H(\widehat{\theta}_N^n) - H \right|^2 \right\} \approx \frac{1}{N} |H(\widehat{\theta}_N^n)|^2 \sum_{k=1}^{m_c+m_d} \frac{1 - |\eta_k|^2}{|e^{j\omega} - \eta_k|^2}. \tag{51}$$

where

$$\widehat{\sigma}^2 \triangleq \frac{1}{N} \sum_{t=1}^{N} \varepsilon_t^2(\widehat{\theta}_N^n). \tag{52}$$

There are some important facets of, and conclusions to be drawn from these quantifications. Firstly, and most importantly, the expressions (50), (51) are 'exact' for finite model order in in the sense that, unlike most pre-existing results such as (2), they are not derived from an asymptotic in model order argument. As such, they are likely to be far more accurate for practical cases of finite, and indeed low model order as has already been illustrated in Figure 1.

Secondly, these results in Theorem 5.1 and Corollary 5.1 represent an extension of those given in [16]. There, model structures with poles fixed at those of the true underlying system, and fixed noise model were considered, where here we consider the more general case where the poles and the noise model are unknown, and hence estimated.

Thirdly, note that a key point is that the quantification (50) resolves an outstanding paradox in the theory of system identification. Namely, a consequence of the existing quantification (2) is that, since it applies for any shift invariant structure which includes the FIR, and Box–Jenkins cases then (as explained in [16]) it suggests that there is no variance penalty to be paid for estimating pole locations in the Box–Jenkins model structure (14), as opposed to fixing those pole locations and estimating just a numerator term as an FIR structure.

This is counter-intuitive, and indeed the expression (50) indicates that it is in fact untrue. Specifically, consider the case of white input and white noise, i.e. $\Phi_u(\omega) = \Phi_\nu(\omega) \equiv 1$. Then, in the case of the denominator order m_a equalling the numerator order m_b, the results of [16] applying for the case of fixed denominator modelling with poles at $\{\xi_k\}$ (as well as Theorem 5.1) give an exact quantification for finite numerator order $m = m_b = m_a$ of

$$\lim_{N\to\infty} N \,\mathbf{E}\left\{ \left| G(\widehat{\theta}_N^n) - G \right|^2 \right\} = \frac{\sigma^2}{\Phi_u(\omega)} \sum_{k=1}^{m} \frac{1 - |\xi_k|^2}{|e^{j\omega} - \xi_k|^2}. \tag{53}$$

whereas when the denominator is estimated, $A^2(z, \theta_\circ^n) H(z, \theta_\circ^n)/F(z) = A^2(z, \theta_\circ^n)$ and Corollary 5.1 then asserts that

$$\begin{aligned} \lim_{N\to\infty} N \,\mathbf{E}\left\{ \left| G(\widehat{\theta}_N^n) - G \right|^2 \right\} &= \frac{\sigma^2}{\Phi_u(\omega)} \sum_{k=1}^{2m} \frac{1 - |\xi_k|^2}{|e^{j\omega} - \xi_k|^2} \\ &= 2 \cdot \frac{\sigma^2}{\Phi_u(\omega)} \sum_{k=0}^{m} \frac{1 - |\xi_k|^2}{|e^{j\omega} - \xi_k|^2}. \end{aligned} \tag{54}$$

Therefore, the new quantification (50) of this paper is an expression that, as well as being exact for finite model orders m_a, m_b, is exactly *double* that pertaining to the case examined in [16] were poles are fixed. Since, roughly speaking, twice as much information (i.e. a numerator *and* a denominator) is being estimated, this new result now corroborates intuitive belief and resolves the paradox.

Finally, note that if the underlying system is, in fact, of FIR type so that all the poles $\{\xi_k\}$ are at the origin, then (54) implies

$$\mathbf{E}\left\{\left|G(e^{j\omega},\widehat{\theta}^n_N) - G(e^{j\omega})\right|^2\right\} \approx 2\cdot\frac{m\sigma^2}{N\Phi_u(\omega)}.$$

Therefore, modulo a factor of 2, there is a rapprochement in the FIR modelling case between the new quantifications derived here, and the pre-existing one (2).

In relation to this last point where Φ_u was assumed white, there are a range of non-trivial possibilities of input spectrum for which exact variance quantification is possible, as illustrated by the further corollary to Theorem 5.1.

Corollary 5.3 (Box–Jenkins, coloured input): Suppose that all the conditions of Lemma 4.1 and Theorem 5.1 are satisfied, and that the spectral factor $F(z)$ of $\Phi_u(\omega)$ and the limiting noise model $H(z,\theta^n_\circ)$ are of the form ($\mu \in \mathbf{C}$)

$$F(z) = \frac{\mu}{L(z)}, \quad L(z) = \prod_{k=1}^{m_f}(1 - f_k z^{-1}),$$

$$H(z,\theta^n_\circ) = \prod_{k=1}^{m_h}(1 - h_k z^{-1}). \tag{55}$$

Then if $m_b = m_a + m_f + m_h + m_\star$, $m_\star \geq 0$

$$\lim_{N\to\infty} N\cdot\mathrm{Cov}\left\{\begin{bmatrix} G(\widehat{\theta}^n_N) \\ H(\widehat{\theta}^n_N)\end{bmatrix}\right\} = \Phi_\nu(\omega)\begin{bmatrix} \dfrac{\rho(\omega)}{\Phi_u(\omega)} & 0 \\ 0 & \dfrac{\widetilde{\kappa}(\omega)}{\sigma^2}\end{bmatrix} \tag{56}$$

where, with the $\{\xi_k\}$ being chosen as the zeros of $A(z,\theta^n_\circ)$,

$$\rho(\omega) \triangleq m_\star + 2\cdot\sum_{k=1}^{m_a}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2} + \sum_{k=1}^{m_f}\frac{1-|f_k|^2}{|e^{j\omega}-f_k|^2} + \sum_{k=1}^{m_h}\frac{1-|h_k|^2}{|e^{j\omega}-h_k|^2}$$

and $\widetilde{\kappa}(\omega)$ is defined in (44) of Theorem 5.1.

■

The preceding corollary shows that when $\mathcal{S} \in \mathcal{M}$ and the input is of AR(p) type with noise of MA(r) type, then provided that $p + r$ more 'lags' are modelled in the numerator $B(q,\theta^n)$ than the denominator $A(q,\theta^n)$, then the quantification

$$\mathbf{E}\left\{\left|G(\widehat{\theta}^n_N) - G\right|^2\right\} \approx \frac{\sigma^2}{N}\frac{|H(e^{j\omega},\widehat{\theta}^n_N)|^2}{\Phi_u(\omega)}\times$$

$$\left[m_\star + 2\sum_{k=1}^{m_a}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2} + \sum_{k=1}^{p}\frac{1-|\ell_k|^2}{|e^{j\omega}-\ell_k|^2} + \sum_{k=1}^{r}\frac{1-|c_k|^2}{|e^{j\omega}-c_k|^2}\right]$$

is 'exact' for finite model order. This is a direct extension of the results of [16] from the fixed-and-correct denominator and noise model case considered there, to the more general Box–Jenkins modelling situation in which the denominator and noise model are both estimated.

Furthermore, comparing the above to the quantification presented in [16] establishes that the effect of modelling the denominator is to double the size of the component $\sum_{k=1}^{m_a}(1-|\xi_k|^2)/|e^{j\omega}-\xi_k|^{-2}$ (which depends on the poles $\{\xi_k\}$ of $G(q,\theta^n_\circ)$) in the variance of the dynamics estimate, and as mentioned in the introduction and previous discussion, this makes intuitive sense.

VI. CONCLUSIONS

This paper has established the new principle that the problem of quantifying variance error is equivalent to that of quantifying the reproducing kernel for a certain subspace of L_2. Since this subspace was shown to depend on the model structure being employed, this exposes the important result that the variance error is also dependent on the model structure, and this fact is counter to what is suggested by pre-existing quantifications such as (2) which depend on an asymptotic in model order argument.

Furthermore, for certain important special cases, it was shown how the reproducing kernel, and hence the variance error, could be expressed in closed form, and for a finite model order of interest. These results, since they apply for the very general Box–Jenkins and Output-Error model structures, are an extension and generalisation of previous results that have applied only for FIR-type model structures with fixed denominators.

VII. REFERENCES

[1] D. ALPAY, *The Schur Algorithm, Reproducing Kernel Spaces and System Theory*, American Mathematical Society, 1998.

[2] M. GEVERS, L. LJUNG, AND P. VAN DEN HOF, *Asymptotic variance expressions for closed loop id.}, Automatica, 37 (2001), pp. 781–786.*

[3] E. HANNAN AND D. NICHOLLS, The estimation of the prediction error variance, *J. Amer. Stat. Assoc, 72 (1977), pp. 834–840.*

[4] L. LJUNG, System Identification: Theory for the User, (2nd edition), *Prentice-Hall, Inc., New Jersey, 1999.*

[5] L.LJUNG, Convergence analysis of parametric identification methods, *IEEE Transactions on Automatic Control, AC-23 (1978), pp. 770–783.*

[6] ———, Asymptotic variance for identified black-box transfer function models, *IEEE Trans. on Auto. Cont, AC-30 (1985), pp. 834–844.*

[7] L.LJUNG AND P.E.CAINES, Asymp. Normality of pred. error estimators for approximate system models, *Stochastics, 3 (1979), pp. 29–46.*

[8] L.LJUNG AND B. WAHLBERG, Asymptotic properties of the least squares method for estimating transfer functions and disturbance spectra, *Advances in Applied Probability, 24 (1992), pp. 412–440.*

[9] L.LJUNG AND Z.D.YUAN, Asymp. prop. of black-box identification of transfer functions, *IEEE Trans. Auto. Cont, 30 (1985), pp. 514–530.*

[10] B. NINNESS AND H. HJALMARSSON, Variance Error Quantifications that are Exact for Finite Model Order, *Submitted to IEEE Trans. on Auto. Cont, (2003). Available at* www.ee.newcastle.edu.au

[11] B. NINNESS, H. HJALMARSSON, AND F. GUSTAFSSON, The fundamental role of general orthonormal bases in system identification, *IEEE Transactions on Automatic Control, 44 (1999), pp. 1384–1406.*

[12] P.M.J. VAN DEN HOF, P.S.C. HEUBERGER, AND J. BOKOR, Sys. id. with gen. ortho. basis func., *Automatica, 31 (1995), pp. 1821–1834.*

[13] P. VAN DEN HOF, Closed-loop issues in system identification, *Annual Reviews in Control, 22 (1998), pp. 173–186.*

[14] B. WAHLBERG, System identification using Laguerre models, *IEEE Transactions on Automatic Control, AC-36 (1991), pp. 551–562.*

[15] B. WAHLBERG, System identification using Kautz models, *IEEE Transactions on Automatic Control, AC-39 (1994), pp. 1276–1282.*

[16] L.-L. XIE AND L. LJUNG, Asymp. var. expr. for estimated frequency functions, *IEEE Transa. Auto. Cont, 46 (2001), pp. 1887–1899.*