The Analysis of Variance Error Part I:Accurate Quantification

Brett Ninness* Håkan Hjalmarsson†

Abstract

This and a companion paper focus on the accurate quantification of the way noise induced estimation errors are dependent on model structure, underlying system frequency response, measurement noise and input excitation. This study exposes several new principles. In particular, it is shown that when employing Output–Error and Box–Jenkins model structures in a prediction-error framework, then the ensuing estimate variability in the frequency domain depends on the underlying system pole positions. As well, it is also established that the variability is affected by the choice of model structure. For example, with fixed noise model, it is twice as much when system poles are estimated as when they are a-priori known and fixed, even though the model order is the same in both cases. These results, together with others presented here and in a companion paper, are unexpected according to pre-existing theory. They depend on new techniques and results developed by the authors in the area of rational orthonormal bases, with the idea of a 'reproducing kernel' playing a key role.

Technical Report EE2002020, Department of Electrical and Computer Engineering, University of Newcastle, AUSTRALIA

1 Introduction

When identifying a system model on the basis of observed data, it is essential to quantify the likely error in that estimated model. Typically, this consists of two components. The first, a so-called "bias error", is the result of the model structure being less complex than the system being estimated. The second, called "variance error", is caused by corruption of the input-output data measurements.

Furthermore, when the corruption can be modelled as an additive stochastic process, and the underlying system is linear, then it is arguable that the total error in any identified model that passes a validation test is dominated by variance error [11].

In this common case, or when the model structure is rich enough to encompass the true underlying dynamics, the quantification of the total estimation error then becomes a question of assessing variance error. In relation to this, if the widely used prediction-error method with a quadratic criterion is employed [9, 22], then a seminal result is that under open-loop conditions the noise-induced

*This work was supported by the Australian Research Council. This author is with the School of Electrical Engineering & Computer Science, University of Newcastle, Australia and can be contacted at email:brett@ee.newcastle.edu.au or FAX: +61 49 21 69 93. Part of this work was completed while visiting S3-Automatic Control, The Royal Institute of Technology, Stockholm, Sweden.

†This author is with the Department of Sensors, Signals and Systems (Automatic Control), The Royal Institute of Technology, S-100 44 Stockholm, Sweden and may be contacted at email:hakan.hjalmarsson@s3.e.kth.se or FAX: +46 8 790 7329

error, as measured by the variability of the ensuing frequency response estimate $G(e^{j\omega}, \widehat{\theta}_N^n)$, may be approximated as [13, 16, 9, 15]

$$\text{Var}\{G(e^{j\omega}, \widehat{\theta}_N^n)\} \approx \frac{m}{N}\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)}. \tag{1}$$

Here Φ_ν and Φ_u are, respectively, the measurement noise and input excitation 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.

Apart from its simplicity, a key factor underlying the importance and popularity of the approximation (1) is that, according to its derivation [13, 16, 9, 15], it 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 [13]. Additionally, as shown in [9], it also applies when non-parametric (spectral based) estimation methods [1, 9] are employed provided that the m term in (1) is replaced by one dependent on the number of data points (and the windowing function) used.

Therefore, the only influence that the chosen model structure has on the right hand side of (1) is in terms of its order, and because of this the belief that $\text{Var}\{G(e^{j\omega}, \widehat{\theta}_N^n)\}$ is invariant to the particular choice of m'th order model structure has become a fundamental tenet of system identification.

Furthermore, it is also held as axiomatic that $\text{Var}\{G(e^{j\omega}, \widehat{\theta}_N^n)\}$ does not depend on the underlying true frequency response, again on account of the right hand side of (1) being independent of that quantity; see, for example, the work [3, 6, 27, 4, 26].

This and the companion paper [17] establishes these beliefs to be false, and that more accurate quantifications than (1) exist. In relation to this, a series of recent contributions [23, 24, 21, 19] has established a variance error quantification that is an extension of (1) and which is applicable to certain model structures which have poles or zeros fixed according to prior knowledge. For example, if an FIR structure is generalised so that its fixed poles $\{\xi_0, \cdots, \xi_{m-1}\}$ are not necessarily all at the origin, then [23, 24, 21, 19] has shown that in the interests of maximally accurate approximation, the quantification (1) should be modified to become

$$\text{Var}\{G(e^{j\omega}, \widehat{\theta}_N^n)\} \approx \frac{1}{N}\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)}\sum_{k=0}^{m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}. \tag{2}$$

Note that (2) reverts to (1) for the FIR case of $\xi_k = 0$. Furthermore, in [19] it has been shown that for ARX model structures with fixed noise model zeros, again not necessarily at points $\{\xi_0, \cdots, \xi_{m-1}\}$ which are at the origin, then again the expression (2) rather then (1) should be used in the interests of providing the most accurate approximation of $\text{Var}\{G(e^{j\omega}, \widehat{\theta}_N^n)\}$.

For both these generalised FIR and ARX cases where (2) is preferable, when actually computing $\widehat{\theta}_N^n$, the process of incorporating the fixed poles or zeros may be achieved by first pre-filtering the input data with an all-pole filter $F(q)$, and then using a conventional FIR or ARX structure [19].

The previous work [23, 24, 21, 19] has therefore illustrated that the effect of pre-filtering by $F(q)$ on $\text{Var}\{G(e^{j\omega}, \widehat{\theta}_N^n)\}$ cannot be accommodated by simply making the substitutions $\Phi_u \mapsto |F|^2\Phi_u$, $\Phi_\nu \mapsto |F|^2\Phi_\nu$. Instead the filter pole locations $\{\xi_k\}$ must be directly accounted for via (2). This has established that the quantification of $\text{Var}\{G(e^{j\omega}, \widehat{\theta}_N^n)\}$ cannot be expressed in a manner that is invariant to the nature (roughly speaking, smoothness) of the input and noise spectral densities.

The contribution of this paper and its companion [17] is to extend these results. In particular, in the case of an Output-Error structure and with $\{\xi_0, \cdots, \xi_{n-1}\}$ being the estimated poles of $G(q, \widehat{\theta}_N^n)$,

then a main result of this paper is to establish that the expression

$$\text{Var}\{G(e^{j\omega},\widehat{\theta}_N^n)\} \approx \frac{2}{N}\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)}\sum_{k=0}^{m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2} \tag{3}$$

is a significantly more accurate quantification than the widely held one (1).

As an example of the consequences of this expression, comparing (3) with (2) indicates that the variability $\text{Var}\{G(e^{j\omega},\widehat{\theta}_N^n)\}$ associated with estimating a model with known fixed poles $\{\xi_0,\cdots,\xi_{m-1}\}$, and hence only estimating a numerator, is only one half the variability associated with a model where the poles are estimated. Since, roughly speaking, twice as much information is being estimated, this result makes intuitive sense. However, it is completely at odds with pre-existing thought derived from (1) which, since the number of denominators polynomials $d = 1$ in both cases, would indicate that $\text{Var}\{G(e^{j\omega},\widehat{\theta}_N^n)\}$ is invariant to whether poles are estimated or not!

In dealing with this material, the work is organised into two papers. The one at hand essentially presents, proves and illustrates the main new variance error quantification result which is a generalisation of (3) to also accommodate the closed loop data case. It is organised to proceed in §2 with a simulation example designed to make this introductory discussion more concrete and to motivate the need for further study. The following §3 then precisely defines the problem scenario in terms of the available data and the estimation algorithm. With this in place, §4 then proceeds to directly state the main formal results of the paper, and this is followed in §5 with an extensive set of simulation studies illustrating the utility and accuracy of the new results of this paper, and in particular the approximation (3). The paper is then concluded in §6 and §7 with a discussion of how the results here are reconciled with pre-existing ones such as (1).

The companion paper [17] is concerned with the motivation and development of new analysis tools that are employed here in the proof of the main results. In particular, via the use of certain principles of 'Reproducing Kernels' and a parameterisation of the subspaces they specify via a particular rational orthonormal basis, new methods for understanding and quantifying certain Toeplitz matrix quadratic forms are derived. The associated work [17] also details how these methods and the ensuing main results of this paper are related to previous contributions [13, 16, 9], as well as establishing necessary and sufficient conditions for the accuracy of the new approximations here and in [17], while also deriving necessary conditions for the accuracy of the existing result (1), which seem very difficult to satisfy.

2 Motivation

In the interests of motivating the analysis and results to follow, it is preluded by an illustrative simulation example in which the following continuous time system

$$G(s) = \frac{0.0012(1-3.33s)^3}{(s+0.9163)^2(s+0.3567)^2(s+0.2231)^3}$$

is considered, and for which input-output samples are obtained at 1 second intervals with zero-order-held inputs. This implies a discrete time representation

$$G(q) = \frac{-0.0177(q^2-2.7192q+1.8489)(q-4.1377)(q-1.3298)(q+0.4466)(q+0.0463)}{(q-0.8)^3(q-0.7)^2(q-0.4)^2}$$

which is estimated using a 7'th order output error model structure and on the basis of observing a length $N = 10000$ sample input-output record for which the output is corrupted by white Gaussian

noise of variance $\sigma^2 = 0.01$, and with input which is a realisation of a stationary Gaussian process with spectral density

$$\Phi_u(\omega) = \frac{1}{1.25 - \cos\omega}.$$

The sample mean square error over 10000 estimation experiments with different input and noise realisations is used as an estimate of $\mathrm{Var}\{G(e^{j\omega}, \widehat{\theta}_N^n)\}$ and plotted as a solid line in figure 1. The 'classical' approximation (1) is shown as a dash-dot line in that same figure, and is clearly a poor approximation to (the estimate of) the true variability. For example, apart from quantitative errors, the approximation (1) is also qualitatively misleading by being of a 'high-pass' nature when the true variability appears to be 'low pass'.

By way of contrast, the modified approximation (3) (with the $\{\xi_k\}$ being the true poles in $G(q)$) which is shown as the dashed line in figure 1 appears to be quite an accurate approximation to the variability of the model estimate.

This provides clear empirical evidence that the true variability is in fact not invariant to either the model structure, or the true dynamics. Therefore, accurate approximation of the variance may need to take these into account, as the new expression (3) does. This and the companion paper [17] are now

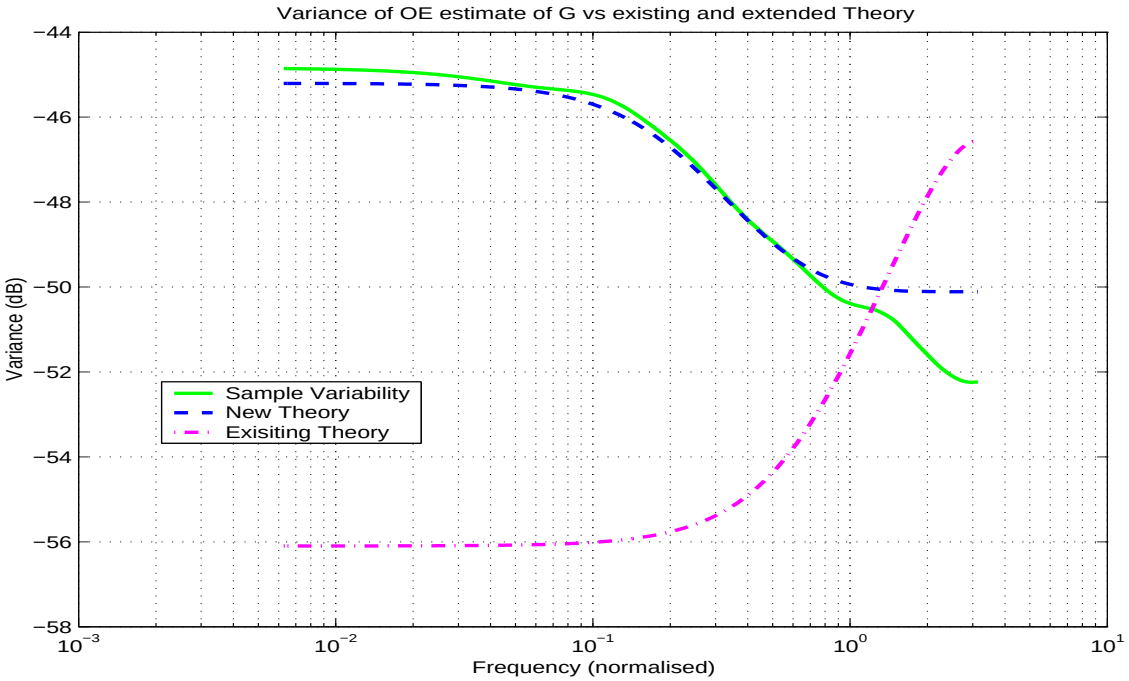


Figure 1: *Variability of Output–Error Estimate: True variability vs. theoretically derived approximations. Solid line is Monte–Carlo estimate of true variability, dash-dot line is the pre-existing approximation (3) which does not account for system poles or model structure. The dashed line is the new approximation presented in (3) whereby estimated system pole positions* $\{\xi_0, \cdots, \xi_{m-1}\}$ *and the fact that an Output–Error structure is employed are both accounted for.*

devoted to supplying a theoretical analysis of this issue, leading to the derivation of (3) and further new results.

3 Problem Formulation

In order to analyse the phenomenon illustrated in figure 1, it is necessary to precisely define the problem setting, which is one in which a model structure is used to describe the relationship between an observed input data record $\{u_t\}$ and output data record $\{y_t\}$ as

$$y_t = G(q, \theta^n)u_t + \nu_t, \qquad \nu_t = H(q, \theta^n)e_t. \tag{4}$$

Here $\{e_t\}$ is a zero-mean white noise sequence that satisfies $\mathbf{E}\{e_t^2\} = \sigma^2, \mathbf{E}\{|e_t|^{4+\epsilon}\} < \infty$ for some $\epsilon > 0$ and $G(q, \theta^n)$, $H(q, \theta^n)$ are transfer functions, rational in the forward shift operator q, and parameterised by a vector $\theta^n \in \mathbf{R}^n$ in such as way that $H(q, \theta^n)$ is monic; i.e. $\lim_{|q|\to\infty} H(q, \theta^n) = 1$.

In this case, the relationship (4) is commonly known [9, 2, 22] as a 'Box–Jenkins' model structure, or an 'Output–Error' model structure in the case where the noise model is fixed at $H(q, \theta^n) = 1$. The mean-square optimal one-step ahead predictor $\widehat{y}_t(\theta^n)$ based on the model structure (4) is [9]

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

with associated prediction error

$$\varepsilon_t(\theta^n) \triangleq y_t - \widehat{y}_t(\theta^n) = H^{-1}(q, \theta^n)\left[y_t - G(q, \theta_n)u_t\right]. \tag{5}$$

Using this, a quadratic estimation criterion may be defined as

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

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

$$\widehat{\theta}_N^n \triangleq \underset{\theta^n \in \mathbf{R}^n}{\arg\min}\, V_N(\theta^n). \tag{6}$$

Forming system estimates via the techniques (4)-(6) has become quite standard, in large part due to the availability of sophisticated software tools implementing the method [10], but also because of extensive theoretical understanding of the properties of such an approach.

For example, as has been established in [12, 9], under certain mild assumptions on the nature of the input $\{u_t\}$ (which will be discussed in detail later), the estimate $\widehat{\theta}_N^n$ converges with increasing N according to

$$\lim_{N\to\infty} \widehat{\theta}_N^n = \theta_\circ^n \triangleq \underset{\theta^n \in \mathbf{R}^n}{\arg\min} \lim_{N\to\infty} \mathbf{E}\left\{V_N(\theta^n)\right\} \qquad \text{w.p.1.} \tag{7}$$

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 [14, 2, 9]

$$\sqrt{N}(\widehat{\theta}_N^n - \theta_\circ^n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, P_n) \qquad \text{as } N \to \infty. \tag{8}$$

The $n \times n$ 'covariance matrix' P_n in (8) is defined in terms of two other matrices R_n and Q_n as

$$P_n \triangleq R_n^{-1} Q_n R_n^{-1} \tag{9}$$

which themselves are specified as

$$R_n \triangleq \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \left[\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_t^T(\theta_\circ^n)\right\} - \mathbf{E}\left\{\varepsilon_t(\theta_\circ^n)\left(\frac{\mathrm{d}\psi_t(\theta^n)}{\mathrm{d}\theta^n}\right)^T\right\}\right] \tag{10}$$

and

$$Q_n \triangleq \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}\sum_{\ell=1}^{N} \mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_\ell^T(\theta_\circ^n)\varepsilon_t(\theta_\circ^n)\varepsilon_\ell(\theta_\circ^n)\right\}. \tag{11}$$

The quantity $\psi_t(\theta^n)$ in the preceding expressions is the prediction error gradient given by

$$\psi_t(\theta^n) \triangleq \frac{\mathrm{d}\widehat{y}_t(\theta^n)}{\mathrm{d}\theta^n} = H^{-1}(q,\theta^n)\left[\frac{\mathrm{d}G(q,\theta^n)}{\mathrm{d}\theta^n}, \frac{\mathrm{d}H(q,\theta^n)}{\mathrm{d}\theta^n}\right]\zeta_t(\theta^n), \quad \zeta_t(\theta^n) \triangleq \begin{bmatrix} u_t \\ \varepsilon_t(\theta^n)\end{bmatrix}. \tag{12}$$

While an asymptotic distributional result like (8) is very satisfying theoretically, for practical applications it is rather less appealing, mainly due to the (just presented) intricate definition of P_n via Q_n, R_n and $\psi_t(\theta^n)$.

In response to this, the seminal work [7, 13, 16, 9, 15] has used an approach of investigating how (8) manifests itself in the variability of the frequency responses $G(e^{j\omega},\widehat{\theta}_N^n)$ and $H(e^{j\omega},\widehat{\theta}_N^n)$; the result being approximations such as (1).

The path towards achieving this involves noting that, with the definition

$$\Pi(q,\theta) \triangleq [G(q,\theta), H(q,\theta)] \tag{13}$$

then according to a first order Taylor expansion, the relationship between frequency domain and parameter space estimation errors is given as

$$\Pi^T(e^{j\omega},\widehat{\theta}_N^n) - \Pi^T(e^{j\omega},\theta_\circ^n) = \left[\left.\frac{\mathrm{d}\Pi(e^{j\omega},\theta^n)}{\mathrm{d}\theta^n}\right|_{\theta^n=\theta_\circ^n}\right]^T (\widehat{\theta}_N^n - \theta_\circ^n) + o(\|\widehat{\theta}_N^n - \theta_\circ^n\|^2). \tag{14}$$

Therefore, a consequence of (8) is that

$$\sqrt{N}\begin{bmatrix} G(e^{j\omega},\widehat{\theta}_N^n) - G(e^{j\omega},\theta_\circ^n) \\ H(e^{j\omega},\widehat{\theta}_N^n) - H(e^{j\omega},\theta_\circ^n)\end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Delta_n(\omega)), \qquad \text{as } N\to\infty \tag{15}$$

where

$$\Delta_n(\omega) \triangleq \left[\left.\frac{\mathrm{d}\Pi(e^{j\omega},\theta^n)}{\mathrm{d}\theta^n}\right|_{\theta^n=\theta_\circ^n}\right]^T P_n \left[\left.\frac{\mathrm{d}\Pi(e^{-j\omega},\theta^n)}{\mathrm{d}\theta^n}\right|_{\theta^n=\theta_\circ^n}\right]. \tag{16}$$

The main contribution of this, and the companion paper [17] is to rigorously establish that, as already illustrated in figure 1, this quantity $\Delta_n(\omega)$ may be accurately approximated (or in some cases exactly quantified) via simple expressions such as (3), and furthermore that the pre-existing quantification (1) can be an unreliable basis for quantifying $\Delta_n(\omega)$ for certain specific reasons.

4 Main Results

Before formally presenting the key contributions of this paper, it is first necessary to discuss two important technical points.

Firstly, and as indicated via (3) and the ensuing results in figure 1, the nature of the input sequence $\{u_t\}$ will contribute in a fundamental way to the accuracy of the parameter estimate $\widehat{\theta}_N^n$ and associated frequency response estimates $G(e^{j\omega},\widehat{\theta}_N^n)$ and $H(e^{j\omega},\widehat{\theta}_N^n)$. It is therefore appropriate to be explicit about any assumptions on $\{u_t\}$.

Input Assumption 4.1. *The input sequence $\{u_t\}$ is uniformly bounded (with probability one) as $|u_t| \leq C_u < \infty$ for all t, and is 'quasi-stationary' in the sense defined by Ljung (see, for example[9]) in that the following limit exists*

$$R_u(\tau) \triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{E}\{u_t u_{t-\tau}\}$$

where in addition $|\boldsymbol{E}\{u_t u_s\}| < \infty$ for all t and s (expectation is over the probability space that any random components of $\{u_t\}$ are defined on). Furthermore, it is assumed that $R_u(\tau)$ is absolutely summable so that the spectral density $\Phi_u(\omega)$ defined by

$$\Phi_u(\omega) \triangleq \sum_{\tau=-\infty}^{\infty} R_u(\tau) e^{-j\omega\tau}$$

exists and it is required that $0 < \Phi_u(\omega) < \infty$. □

Secondly, it is also important to emphasise that a crucial aspect of this and the companion paper [17] is the recognition of the need to carefully consider the relationship between the model order m for which a variance error quantification is required and any underlying 'true' system order. Indeed, given the usual complexity of real-world dynamics, any assumption of the existence of a true model order could be quite inappropriate.

In relation to this issue, the work here takes the perspective that, while on the one hand it is reasonable to assume that under-modelling-induced error decreases with increasing model order m, it is also reasonable to assume that the model order of interest has not surpassed any underlying true order, and hence does not imply pole-zero cancellations in the (asymptotic in N) estimated system.

This last premise is considered to be a realistic way of avoiding the supposition of a true model order, while still considering that some sort of model validation procedure, that checks for the appropriateness of the model structure (4) (eg. in terms of residual whiteness), and at the very least checks for pole-zero cancellation, is part of an overall estimation and error-quantification process.

Incorporating these ideas in a rigorous manner leads to certain technical difficulties in the proof of the following theorems and their corollaries. In particular, an essential point in the main theorem to follow is that it is not presumed that the asymptotic (in N) estimation residual $\varepsilon_t(\theta_\circ^n) = e_t =$ white noise.

Instead, a more general situation is considered whereby it is supposed that $\varepsilon_t(\theta_\circ^n)$ can be decomposed as

$$\varepsilon_t(\theta_\circ^n) = e_t + r_t^n \tag{17}$$

where r_t^n is independent of e_t (as in §3, $\{e_t\}$ is an i.i.d. zero mean white noise sequence for which $\mathbf{E}\{e_t^2\} = \sigma^2 < \infty$). Furthermore, it is supposed that for some $\rho, \beta > 1$ and finite C

$$\frac{1}{N}\sum_{t=1}^{N} \mathbf{E}\left\{|r_t^n r_{t-\tau}^n|\right\} \leq \frac{C}{n^\beta(1+|\tau|^\rho)}, \qquad \frac{1}{N}\sum_{t=1}^{N} \mathbf{E}\left\{|r_t^n e_{t-\tau}|\right\} \leq \frac{C}{n^\beta(1+|\tau|^\rho)}. \tag{18}$$

These requirements correspond to an assumption on the rate at which a finite dimensional model structure $G(q,\theta^n)$ is able to approximate an underlying true $G(q)$ that generates the observed input-output data via (4). More specifically, via (4) and (5), if it holds that $y_t = G(q)u_t + H(q)e_t$ for some true $G(q), H(q)$, then

$$r_t^n = H^{-1}(q,\theta_\circ^n)\left[G(q) - G(q,\theta_\circ^n), H(q) - H(q,\theta_\circ^n)\right]\zeta_t, \qquad \zeta_t \triangleq \begin{bmatrix} u_t \\ e_t \end{bmatrix}. \tag{19}$$

Hence, by Theorem 2.2 of [9], using (18), and assuming for simplicity, but without loss of generality, the case of open loop data collection

$$\frac{C}{n^\beta} \geq \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \mathbf{E}\left\{(r_t^n)^2\right\} \quad = \quad \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{|G(e^{j\omega}) - G(e^{j\omega},\theta_\circ^n)|^2}{|H(e^{j\omega},\theta_\circ^n)|^2}\Phi_u(\omega) + \sigma^2 \frac{|H(e^{j\omega}) - H(e^{j\omega},\theta_\circ^n)|^2}{|H(e^{j\omega},\theta_\circ^n)|^2}\,\mathrm{d}\omega \tag{20}$$

and therefore with μ denoting Lebesgue measure on $\mathbf{R}$, then via Chebychev's inequality

$$\mu\left\{\omega \in [-\pi,\pi] : \left|\frac{G(e^{j\omega}) - G(e^{j\omega},\theta_\circ^n)}{H(e^{j\omega},\theta_\circ^n)}\right|^2 \Phi_u(\omega) + \sigma^2\left|\frac{H(e^{j\omega}) - H(e^{j\omega},\theta_\circ^n)}{H(e^{j\omega},\theta_\circ^n)}\right|^2 > \epsilon\right\} \leq \frac{C}{\epsilon n^\beta}. \tag{21}$$

Therefore, the bounds (18) imply an assumption on the worst-case frequency domain estimation error of $\|G(e^{j\omega}) - G(e^{j\omega},\theta_\circ^n)\|_\infty, \|H(e^{j\omega}) - H(e^{j\omega},\theta_\circ^n)\|_\infty = o(1/n^\beta)$ as $n \to \infty$.

It should also be recognised that the assumptions (18) on the $\{r_t^n\}$ process defined in (17) are different to those employed in the original work [13]. There $N^{-1}\sum_{t=1}^{N}\mathbf{E}\{|r_t^n|^2\} = o(1/n^2)$ as $n \to \infty$ was assumed, while the work here, via (18), only requires $N^{-1}\sum_{t=1}^{N}\mathbf{E}\{|r_t^n|^2\} = o(1/n)$. This is achieved at the expense of assumptions about how $N^{-1}\sum_{t=1}^{N}\mathbf{E}\{|r_t^n r_{t+\tau}^n|\}$ behaves as a function of τ, and such concerns were avoided in [13].

With these preliminary comments in mind, the following Theorem 4.1 and Corollaries 4.1–4.3 provide formal statements of the main technical results of this paper.

Theorem 4.1. *Suppose that $\widehat{\theta}_N^n$ is calculated via (6) using the model structure (4). Then*

$$\sqrt{N}\begin{pmatrix} G(e^{j\omega},\widehat{\theta}_N^n) - G(e^{j\omega},\theta_\circ^n) \\ H(e^{j\omega},\widehat{\theta}_N^n) - H(e^{j\omega},\theta_\circ^n) \\ G(e^{j\lambda},\widehat{\theta}_N^n) - G(e^{j\lambda},\theta_\circ^n) \\ H(e^{j\lambda},\widehat{\theta}_N^n) - H(e^{j\lambda},\theta_\circ^n) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma_m(\omega,\lambda))$$

as $N \to \infty$, where provided that

1. *$\{u_t\}$ satisfies input-assumption 4.1 with $\Phi_u(\omega) \in Lip(\alpha)$ for some $\alpha > 0$;*
2. *Either $G(z,\theta_\circ^n)$ or $H(z,\theta_\circ^n)$ (or both) contains no pole-zero cancellation for any model order m;*
3. *It holds that $y_t = G(q)u_t + H(q)e_t$ for some asymptotically stable true system $G(q)$ and such that the bounds (17), (18) are satisfied;*
4. *Under the factorisation*

$$\frac{G(z,\theta_\circ^n)}{A(z,\theta_\circ^n)}H^{-1}(z,\theta_\circ^n) = \frac{B_\dagger(z)}{A(z,\theta_\circ^n)A_\dagger(z)}H_\dagger^{-1}(z) \tag{22}$$

where $B_\dagger(z)$ is a polynomial, $H_\dagger(z)$ is rational and bi-proper and

$$A(z,\theta_\circ^n)A_\dagger(z) = (z-\xi_0)(z-\xi_1)\cdots(z-\xi_{2m-1}), \quad D(z,\theta_\circ^n)C(z,\theta_\circ^n) = (z-\eta_0)(z-\eta_1)\cdots(z-\eta_{2m-1}) \tag{23}$$

then the zeros $\{\xi_k\}$ and $\{\eta_k\}$ are contained in the open unit disk $\mathbf{D}$ for any m,

then for $\omega \neq \lambda$

$$\lim_{m\to\infty} \Sigma_m(\omega,\lambda) \left[\begin{array}{cc} K_m(\omega) & \varnothing \\ \varnothing & K_m(\lambda) \end{array} \right]^{-1} = \sigma^2 \left[\begin{array}{cc} |H_\circ(e^{j\omega})|^2 \Phi_\zeta^{-1}(\omega) & 0 \\ 0 & |H_\circ(e^{j\lambda})|^2 \Phi_\zeta^{-1}(\lambda) \end{array} \right] \quad (24)$$

where

$$K_m(\omega) \triangleq \left[\begin{array}{cc} \kappa_m(\omega) & 0 \\ 0 & \widetilde{\kappa}_m(\omega) \end{array} \right], \qquad \kappa_m(\omega) \triangleq \sum_{k=0}^{2m-1} \frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}, \qquad \widetilde{\kappa}_m(\omega) \triangleq \sum_{k=0}^{2m-1} \frac{1-|\eta_k|^2}{|e^{j\omega}-\eta_k|^2} \quad (25)$$

and

$$\Phi_\zeta(\omega) \triangleq \left[\begin{array}{cc} \Phi_u(\omega) & \Phi_{ue}(\omega) \\ \Phi_{ue}(\omega) & \sigma^2 \end{array} \right], \qquad H_\circ(e^{j\omega}) \triangleq \lim_{n\to\infty} H(e^{j\omega}, \theta_\circ^n). \quad (26)$$

Proof. The proof is somewhat lengthy and confined to appendix A. It depends on new technical tools and results developed in the companion paper [17]. □

The implication of this result is that since it asserts that (recall $n = 4m$)

$$\lim_{m\to\infty} \lim_{N\to\infty} N \cdot \mathrm{Cov} \left\{ \left[\begin{array}{c} G(e^{j\omega}, \widehat{\theta}_N^n) - G(e^{j\omega}, \theta_\circ^n) \\ H(e^{j\omega}, \widehat{\theta}_N^n) - H(e^{j\omega}, \theta_\circ^n) \end{array} \right] \right\} \cdot K_m^{-1}(\omega) = \sigma^2 |H_\circ(e^{j\omega})|^2 \Phi_\zeta^{-1}(\omega) \quad (27)$$

then one could expect that the above equality should nearly hold for finite m and N so that, taking the case of open loop data collection in which $\Phi_{ue}(\omega) = 0$ as an example, it should occur that

$$\mathbf{E} \left\{ \left| G(e^{j\omega}, \widehat{\theta}_N^n) - G(e^{j\omega}, \theta_\circ^n) \right|^2 \right\} \approx \frac{1}{N} \frac{\sigma^2 |H(e^{j\omega}, \widehat{\theta}_N^n)|^2}{\Phi_u(\omega)} \sum_{k=0}^{2m-1} \frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}, \quad (28)$$

$$\mathbf{E} \left\{ \left| H(e^{j\omega}, \widehat{\theta}_N^n) - H(e^{j\omega}, \theta_\circ^n) \right|^2 \right\} \approx \frac{1}{N} |H(e^{j\omega}, \widehat{\theta}_N^n)|^2 \sum_{k=0}^{2m-1} \frac{1-|\eta_k|^2}{|e^{j\omega}-\eta_k|^2} \quad (29)$$

are good approximations. The first leads to the new quantification (3) given in the introduction. Nevertheless, since convergence in distribution guarantees nothing about mean-square convergence, then it is not strictly correct to progress to a variance approximation like (28) on the basis of the distributional convergence illustrated in Theorem 4.1. This gap can be closed provided that assumptions on the noise process $\{e_t\}$ are strengthened.

Corollary 4.1. *Under the same conditions Theorem 4.1, but with a strengthened requirement that* $\mathbf{E}\left\{e_t^8\right\} < \infty$ *then with the definition*

$$\widetilde{\Pi}_m(e^{j\omega}) \triangleq \Pi^T(e^{j\omega}, \widehat{\theta}_N^n) - \Pi^T(e^{j\omega}, \widehat{\theta}_\circ^n) \quad (30)$$

with $\Pi(z,\theta)$ *defined in (13) and with* $K_m(\omega)$ *given by (25)*

$$\lim_{m\to\infty} \lim_{N\to\infty} N \cdot \mathbf{E} \left\{ \widetilde{\Pi}_m(e^{j\omega}) \widetilde{\Pi}_m(e^{j\omega})^\star \cdot K_m^{-1}(\omega) \right\} = \sigma^2 |H_\circ(e^{j\omega})|^2 \Phi_\zeta^{-1}(\omega). \quad (31)$$

Proof. Follows from Theorem 4.1 using the methods in Appendix 9B of [9]. □

As is made clear in the companion paper [17], the analysis leading to (27) has been specifically designed to maximise the precision of the approximation (28), and in fact using Theorem 7.1 of the companion work [17], the convergence rate of (27), and hence the accuracy of (28),(29) can be quantified. The following §5 will also establish the precision of (28) and (29) empirically.

In relation to this issue of accuracy, note that the factorisation in (22) is not necessarily unique in its definition of $A_\dagger(z)$, and hence some of the zeros $\{\xi_m, \cdots, \xi_{2m-1}\}$ governing the value of $\kappa_m(\omega)$ in (25) are also not always uniquely defined. In the strictly asymptotic sense of Theorem 4.1, this will be immaterial in that, assuming the number of non-unique zeros is fixed, the effect of these components in the formation of $\kappa_m(\omega)$ will become increasingly negligible compared to the size of $\kappa_m(\omega)$ as m grows.

However, as will be detailed further in the companion paper [17] and illustrated in §5, for a given finite m of interest, the accuracy of ensuing approximations like (28) are maximised by choosing the non-unique zeros of $A_\dagger(z)$ in order that $\Phi_\zeta(\omega)|H_\dagger^{-1}(e^{j\omega})|^2$ is as Lipschitz smooth as possible.

In certain cases though, the factorisation (22),(23) is unique. For example, when employing an Output-Error model structure in which the noise model in (4) is fixed at $H(q,\theta^n) = 1$, then $A_\dagger(z)$ is clearly uniquely given as $A_\dagger(z) = A(q,\theta_\circ^n)$. Therefore, all the zeros $\{\xi_k\}$ occur twice as $\xi_{m+k} = \xi_k$ so that

$$\kappa_m(\omega) = \sum_{k=0}^{2m-1} \frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2} = 2\sum_{k=0}^{m-1} \frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}$$

which leads to the quantification

$$\mathbf{E}\left\{\left|G(e^{j\omega},\widehat{\theta}_N^n) - G(e^{j\omega},\theta_\circ^n)\right|^2\right\} \approx \frac{2}{N}\frac{\sigma^2}{\Phi_u(\omega)}\sum_{k=0}^{m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}. \tag{32}$$

In fact, this case of fixed noise model warrants special mention since, coupled with an open loop measurement scenario, it is a situation in which variance error quantification can be achieved without requiring that the model class be able to asymptotically encapsulate the true underlying noise properties.

Corollary 4.2. *Under the same conditions as Theorem 4.1 save for the modifications that*

1. *In the model structure (4) the noise model is fixed at $H(q,\theta) = H_*(q)$ which is not necessarily equal to any true underlying one;*
2. *It holds that $y_t = G(q)u_t + H(q)e_t$ for some asymptotically stable true system $G(q)$ such that with $\theta_\circ^n$ defined by (7) and for some $C < \infty$, $\beta > 1$*

$$|G(e^{j\omega}) - G(e^{j\omega},\theta_\circ^n)| \le Cn^{-\beta};$$

3. *The cross spectrum $\Phi_{ue}(\omega) = 0$,*

then

$$\sqrt{N}\begin{pmatrix} G(e^{j\omega},\widehat{\theta}_N^n) - G(e^{j\omega},\theta_\circ^n) \\ G(e^{j\lambda},\widehat{\theta}_N^n) - G(e^{j\lambda},\theta_\circ^n)\end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma_m(\omega,\lambda)) \qquad \text{as } N\to\infty \tag{33}$$

where with $\kappa_m(\omega)$ being defined as

$$\kappa_m(\omega) \triangleq \sum_{k=0}^{m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}$$

where the $\{\xi_k\}$ *are the poles of* $G(q,\theta^n_\circ)$ *and for* $\omega \neq \lambda$

$$\lim_{m\to\infty} \Sigma_m(\omega,\lambda) \left[\begin{array}{cc} \kappa_m^{-1}(\omega) & 0 \\ 0 & \kappa_m^{-1}(\lambda) \end{array} \right] = 2 \left[\begin{array}{cc} \Phi_\nu(\omega)|H_*(e^{j\omega})|^{-2}\Phi_u^{-1}(\omega) & 0 \\ 0 & \Phi_\nu(\lambda)|H_*(e^{j\lambda})|^{-2}\Phi_u^{-1}(\lambda) \end{array} \right]. \tag{34}$$

Furthermore, under the strengthened requirement of $\boldsymbol{E}\left\{|e_t|^8\right\} < \infty$

$$\lim_{m\to\infty}\lim_{N\to\infty} \frac{N}{\kappa_m(\omega)} \boldsymbol{E}\left\{|G(e^{j\omega},\widehat{\theta}^n_N) - G(e^{j\omega},\theta^n_\circ)|^2\right\} = \frac{2\Phi_\nu(\omega)}{|H_*(e^{j\omega})|^2\Phi_u(\omega)}. \tag{35}$$

Proof. See Appendix B. □

The ensuing quantification of

$$\mathbf{E}\left\{\left|G(e^{j\omega},\widehat{\theta}^n_N) - G(e^{j\omega},\theta^n_\circ)\right|^2\right\} \approx \frac{2}{N}\frac{\Phi_\nu(\omega)}{|H_*(e^{j\omega})|^2\Phi_u(\omega)} \sum_{k=0}^{m-1} \frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}. \tag{36}$$

then encapsulates the Output-Error model structure as a special case of $H_*(q) = 1$, although the result also has application to data pre-filtering situations, which are equivalent to employing a fixed noise model [9].

Finally, in many applications, it may be unappealing that quantifications like (28) and (36) depend on asymptotic in model order m arguments. In fact, depending on the experimental conditions, it is possible to supply quantifications that avoid this reservation by being asymptotic only in the data length N, as illustrated by the final main result of this section.

Corollary 4.3. *Under the conditions imposed in Theorem 4.1, together with further assumptions that*

1. $\Phi_u(\omega)/|H_\dagger(e^{j\omega})|^2$ *is a constant;*
2. $\varepsilon_t(\theta^n_\circ) = e_t$ *where* $\{e_t\}$ *is 'white noise' satisfying the assumptions of Corollary 4.1;*
3. *The process* $\{e_t\}$ *satisfies* $\boldsymbol{E}\left\{|e_t|^8\right\} < \infty$*;*
4. *The cross spectrum* $\Phi_{ue}(\omega) = 0$*,*

then

$$\lim_{N\to\infty} N \cdot \mathrm{Cov}\left\{ \left[\begin{array}{c} G(e^{j\omega},\widehat{\theta}^n_N) - G(e^{j\omega},\theta^n_\circ) \\ H(e^{j\omega},\widehat{\theta}^n_N) - H(e^{j\omega},\theta^n_\circ) \end{array} \right] \right\} = \Phi_\nu(\omega)\Phi_\zeta^{-1}(\omega)K_m(\omega). \tag{37}$$

If the condition number 1 is dropped, then the above equality still holds except for the top left element of the above 2×2 *matrix valued functions, which will be an approximation.*

Proof. This is established in Appendix C. See also the discussion in §6 of the companion paper [17]. □

Again using the Output-Error case as an example in which the unique factorisation $H_\dagger(z) = 1$ exists, then for white input excitation $\Phi_u = \gamma$ the preceding result asserts that the quantification

$$\mathbf{E}\left\{|G(e^{j\omega},\widehat{\theta}^n_N) - G(e^{j\omega},\theta^n_\circ)|^2\right\} \approx \frac{2\sigma^2}{N\gamma}\sum_{k=0}^{m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}$$

has an accuracy that depends only on the amount N of data, and hence is applicable for arbitrarily low model order, as will be illustrated in the following simulation section. The only other results of this non-asymptotic in m nature have appeared in [25, 18], where less general cases that do not encapsulate the general Box–Jenkins model structure allowed in Corollary 4.3 were addressed.

Note that in relation to Corollaries 4.1 and 4.3, the condition that $\Phi_{ue}(\omega) = 0$ is satisfied by open loop data collection as well as some closed loop scenarios, such as that of a controller which is strictly causal.

5 Further Simulation Example

As has already been emphasised, the work in this paper leading to the new quantification (3), while being derived via asymptotic analysis, is designed to be as accurate as possible for finite model orders m at which it is likely to be used. The companion paper [17] is devoted to explaining and establishing this point.

Certainly, the example shown in figure 1 indicates that this is the case, but it is still what might be considered a relatively high order (m=7) situation. In consideration of this and other aspects surrounding (28),(29) this section presents several additional simulation examples designed to provide comprehensive empirical evidence substantiating the utility of the new approximation (28), (29).

These studies are organised according to the type of system simulated, the colouring of the input spectra, the amount N of observed data, the model structure type, the experimental conditions, and the input-output dynamics which are one of the following.

System1: Low-Order

$$G(q) = \frac{0.1}{(q - 0.9)}, \tag{38}$$

System2: Mid-Order

$$G(q) = \frac{0.06(q - 0.8)(q - 0.9)}{(q - 0.99)(q - 0.7)(q - 0.6)}, \tag{39}$$

System3: Low-Order Resonant

$$G(q) = \frac{0.0342q + 0.0330}{(q - 0.95e^{j\pi/12})(q - 0.95e^{-j\pi/12})}, \tag{40}$$

System4: Mid-Order Resonant

$$G(q) = \frac{0.1176(q + 8.0722)(q + 0.8672)(q + 0.0948)}{(q - 0.75e^{j\pi/3})(q - 0.75e^{-j\pi/3})(q - 0.95e^{j\pi/12})(q - 0.95e^{-j\pi/12})}. \tag{41}$$

For each of these systems, two possible input spectrum are considered

$$\Phi_u(\omega) = \frac{1}{1.25 - \cos\omega} \qquad \text{and} \qquad \Phi_u(\omega) = 1$$

as well as both long ($N = 10,000$) and short ($N = 200$) data lengths.

Firstly, in order to most directly illustrate quantification accuracy, attention is initially restricted to the simplest case of white Gaussian measurement noise of variance $\sigma^2 = 0.0001$ and the employment of an Output–Error model of order equal to the true system.

System	Input Spectrum			
	Coloured		White	
	$\mathbf{N = 10000}$	$\mathbf{N = 200}$	$\mathbf{N = 10000}$	$\mathbf{N = 200}$
1: Low-Order	Fig 2(a)	Fig 2(b)	Fig 3(a)	Fig 3(b)
2: Mid-Order	Fig 4(a)	Fig 4(b)	Fig 5(a)	Fig 5(b)
3: Low-Order Resonant	Fig 6(a)	Fig 6(b)	Fig 7(a)	Fig 7(b)
4: High-Order Resonant	Fig 8(a)	Fig 8(b)	Fig 9(a)	Fig 9(b)

Table 1: *Organisation of Simulation Examples*

This structure is fitted over 10000 different input and measurement noise realisations to allow for the computation of the true estimate variability via sample average over these Monte–Carlo simulations, which is then compared to the new expression (3), (36) as well as the pre-existing one (1) in figures 2–9 and according to the organisation given in table 1.

In each of these figures, (the estimate of) the true variability is shown as a solid line, the new variance expression (3), (36) of this paper is shown as a dashed line, and the pre-existing approximation (1) is illustrated via a dash-dot line. The consideration of all these examples reveals some important points.

Firstly, the new approximation (3) is clearly quite robust. It provides an informative quantification across the full range of scenarios, even for the case of very low model order $m = 1$ and very low data length $N = 200$ as shown in figure 2(b). This accuracy for low model order may seem surprising, but as will be exposed in the companion paper [17], the reason that the factor of 2 is present in the new quantification (3), (36) is that the a key approximating step of Fourier convergence is with respect to an expansion *twice* as long as the model order m, and variance quantification accuracy is proportional to this expansion length.

Secondly, as shown in the cases of white input, the new approximation (3) is essentially exact in these cases regardless of model order, save for small errors at very low data lengths. This, of course, is consistent with Corollary 4.3.

Thirdly, as illustrated in the case of resonant systems, even when the true variability has a quite complicated nature, the new approximation (3) is able to provide an informative and accurate quantification.

Finally, as suggested by examination of the dash-dot line representing (1) in each of figures 1–9, the pre-existing and widely used quantification (1) can be unreliable, which leads to the suggestion of this paper that in fact it should be replaced by (3) and its generalisations (28), (29).

Turning now to the case of Box–Jenkins model structures, consider the situation of the input-output dynamics being that of the low-order system (38), with output measurements now subject to noise coloured as

$$H(q) = \frac{q}{q - 0.9}.$$

In this case, the factorisation (22) of Theorem 4.1 is uniquely given as (using $G(q, \theta_\circ^n) = G(q)$, $H(q, \theta_\circ^n) = H(q)$)

$$\frac{G(q)}{A(q)H(q)} = \frac{0.1}{(q - 0.9)(q - 0.1)} = \frac{B_\dagger(q)}{A(q)A_\dagger(q)} H_\dagger^{-1}(q)$$

so that $H_\dagger(q) = 1$ and $A_\dagger(q) = (q - 0.1)$. Furthermore, in the case of white input, then $\Phi_u/|H_\dagger|^2$ is a constant, so that according to Corollary 4.3, the variance quantification (28) should be 'exact', in

the sense of its accuracy being preserved for arbitrarily low model order m. Reference to figure 10(a) shows that indeed this is the case, with the approximation (28) shown as a dashed line being identical to the Monte–Carlo estimated variability shown as a solid line. Again, for comparison the dash-dot line there is the pre-existing quantification (1), which is clearly far less accurate.

However, consider a modification of this simulation experiment in which the noise colouring is changed to

$$H(q) = \frac{q}{q - 0.85}.$$

In this case, the factorisation (22) is not unique, and could be performed as either of

$$\frac{G(q)}{A(q)H(q)} = \frac{0.1}{(q-0.9)\underbrace{(q-0.9)}_{A_\dagger}}\underbrace{\left(\frac{q-0.85}{q}\right)}_{H_\dagger^{-1}} \quad \text{or} \quad \frac{G(q)}{A(q)H(q)} = \frac{0.1}{(q-0.9)\underbrace{q}_{A_\dagger}}\underbrace{\left(\frac{q-0.85}{q-0.9}\right)}_{H_\dagger^{-1}}.$$

However, as mentioned in §4, and as proved in the companion paper [17], the convergence rate of the asymptotic expression (27), and hence the accuracy of the attendant approximation (28) shown in figure 10(b) increases with increasing Lipschitz smoothness of $\Phi_u/|H_\dagger|^2$. Since in this example, Φ_u is a constant, then the smoothness is clearly greatest for the second factorisation choice of $A_\dagger = q$, $H_\dagger = (q-0.9)(q-0.85)^{-1}$, and this leads to the approximation shown as the dashed line in figure 10(b). The alternative choice of $A_\dagger = (q-0.9)$, $H_\dagger = q(q-0.85)^{-1}$ is shown as the dash-dot line and while still a reasonably accurate quantification, especially considering the very low model orders involved, is inferior to the alternative. The variability $\mathrm{Var}\{H(e^{j\omega}, \widehat{\theta}_N^n)\}$ versus the quantification (29) is shown in figure 11(a), and as predicted by Corollary 4.3, there is exact agreement even though the model order is only $m = 1$.

Finally, since this is an open-loop example, then Corollary 4.2 indicates that the approximation (32) with fixed noise model set to $H_* = 1$ should quantify the variability of an Output-Error structure fitted to this coloured noise case. The validity of this is illustrated in figure 11(b) where the upper solid line of Monte–Carlo estimated Output-Error variability can be compared with (32) shown as a dash-dot line. These true and quantified variabilities can be contrasted with the Box–Jenkins estimates for the same data sets shown below these curves. Although the difference between these cases is not great, it is clearly substantial enough to indicate that, despite pre-existing thought, the variability $\mathrm{Var}\{G(e^{j\omega}, \widehat{\theta}_N^n)\}$ is not invariant to the model structure choice, and this was predicted by the variance quantifications (28) and (32) since the former involves a $\kappa_m(\omega)$ defined via $[\xi_0, \xi_1] = [0.9, 0.0]$ while the latter involves a $\kappa_m(\omega)$ defined via $[\xi_0, \xi_1] = [0.9, 0.9]$.

6 Rapprochement with Pre-existing work

It is important to reconcile these new results with the pre-existing work [13, 9] which inspired the paper here. Superficially, Theorem 4.1 might appear to be in contradiction with Theorem 3.1 of [13], since the latter also addresses the same estimation scenarios considered in Theorem 4.1, but presents a different result in that the $K_m(\omega)$ term arising in Theorem 4.1 is replaced by the frequency independent factor m.

The key point in resolving this is apparent discrepancy is to recognise that this paper takes a different perspective to [13] on the relationship between model order m and any underlying true system order. In [13], it is accepted that as m increases, eventually it must exceed an underlying true model order. On the other hand, this paper assumes that regardless of how large m is, pole-zero cancellation in the asymptotic estimated model never occurs. The motivation for this is to most

accurately represent the common finite model order m application in which an estimated model order is chosen so as to avoid pole-zero cancellations.

Recognising this shows that, in a purely asymptotic sense, there is no contradiction between the convergence results in Theorems 4.1- 4.3 and pre-existing variance convergence results such as those of Theorem 3.1 in [13].

Specifically, if one assumes that a true finite model order $m = m_\star$ exists, then the main theorems of the companion paper [17] which underpin the results of this paper cannot be used since they depend upon R_n defined in (10) being non-singular, and if $m > m_\star$ then pole zero cancellations in $G(z, \theta_\circ^n)$ will lead to rank deficiency in R_n.

Indeed, since this is the Hessian of the asymptotic value of the cost $V_N(\theta)$, then the asymptotic value $\theta_\circ$ specified by (7) is not even well defined, which makes any analysis of variance error ill-posed. To address this, consider the strategy employed in [13] and other works of introducing the *regularised* cost which, for some regularising parameter $\delta > 0$, is defined as

$$V_N(\theta) = \frac{1}{2N}\sum_{t=1}^{N} \varepsilon_t^2(\theta) + \frac{\delta}{2}\|\theta - \theta_\circ\|^2. \tag{42}$$

The Hessian of this modified cost is always positive definite, and hence $\widehat{\theta}_N$ is always well defined. Furthermore, in the limit as $N \to \infty$, the criterion (42) has a unique minimum at $\theta_\circ$ for an arbitrarily small δ.

Asymptotic variance results can therefore be derived with respect to this modified cost in the over-modelled case of $m > m_\star$. Furthermore, as shown in [**?**] for the ARMA time-series case, the reproducing kernel approach used here, and developed in the companion work [17] can also be employed, albeit in a different fashion.

Now, although the work [**?**] addressed only the time series case, the methods used there are directly extendible to the situations considered here. Taking the Output-Error case for the sake of example, the methods of [**?**] then provide the conclusion that if $G(z, \theta_\circ)$ can be factored as

$$G(z, \theta_\circ) = \frac{B_{m_\star}(z, \theta_\circ)}{A_{m_\star}(z, \theta_\circ)} \cdot \frac{T(z)}{T(z)} \tag{43}$$

where $B_{m_\star}(z, \theta_\circ)/A_{m_\star}(z, \theta_\circ)$ is of minimal order $m_\star$ and $T(z)$ is a polynomial of order $m - m_\star$ containing the pole-zero cancellations in $G(z, \theta_\circ)$, then in the limit at $\delta \to 0$, the results of Theorem 4.1 still hold with the zeros $\{\xi_0, \cdots, \xi_{m-1}\}$ being defined by

$$A_{m_\star}(z, \theta_\circ)T(z) = (z - \xi_0)(z - \xi_1)\cdots(z - \xi_{m-1}). \tag{44}$$

Now, if the regularisation point $\theta_\circ$ is chosen such that the pole-zero cancellation points $\{\xi_{m_\star}, \cdots, \xi_{m-1}\}$ are all at the origin, then this implies that the factor $\kappa_m(\omega)$ defined in Theorem 4.1 and appearing in (32) is given as

$$\kappa_m(\omega) = (m - m_\star) + \sum_{k=0}^{m_\star - 1} \frac{1 - |\xi_k|^2}{|e^{j\omega} - \xi_k|^2}$$

and hence

$$\frac{\kappa_m(\omega)}{m} = 1 + O(1/m), \qquad \text{as } m \to \infty. \tag{45}$$

Therefore, in a strictly asymptotic sense, normalisation by $\kappa_m(\omega)$ as in Theorem 4.1 or by m as in Theorem 3.1 of [13] will be equivalent, but *only* if $\theta_\circ$ is chosen so that excess pole-zero cancellations occur at the origin; see [**?**] for more on this last point.

However, *in practice*, for this (possible) asymptotic equivalence to imply variance approximation equivalence, the model order m employed would need to be much larger than the true underlying model order $m_\star$ so that, via (45), $\kappa_m(\omega) \approx m$. In contrast to this, practical interest is usually focused on the situation of roughly correct model order $m \approx m_\star$, in which case the *interpretations* of Theorem 4.1 and Theorem 3.1 of [13] are different.

Specifically, by virtue of the particular adapted orthonormal basis employed in developing Theorem 4.1, the convergence rate involved there is likely to be significantly faster than that associated with Theorem 3.1 of [13] where the standard trigonometric basis is used. As a result and as illustrated in figures 1–9, assuming that (24) has approximately converged for some finite $m \approx m_\star$ can provide a more reliable quantification (3) of variance error than if the convergence in Theorem 3.1 of [13] is assumed to have occurred for finite m to lead to (1).

Indeed, the companion paper [17] shows that when normalisation by $\kappa_m(\omega)$ is involved, then a convergence rate bound can be established that guarantees improving accuracy of $\mathrm{Var}\{G(e^{j\omega}, \widehat{\theta}_N^n)\}$ approximation with increasing model order m, and such results are unavailable via the trigonometric basis methods used in the previous works [13, 16, 9] which led to (1).

7 Conclusion

The main theme of this paper was to highlight that the variability of quadratic-cost prediction-error estimates is not invariant to the choice of model structure, nor is it necessarily invariant to the dynamics of the actual system being estimated.

This is counter to previous thought that has argued that since only shift invariance is required of the model structure for (1) to hold, and since (1) depends only on model order, data length and signal-to-noise ratio, then in fact the precise choice of model structure, and the dynamics of the estimated system are irrelevant to the variance error.

The following companion paper [17] will detail how this latter argument can fail since convergence of a certain Fourier series is central to the approximation (1) being accurate, and this can easily be upset. A strategy developed in [17], and employed in the proof of the results here involves reparameterisation with a certain rational orthonormal basis, which lead to an extension (3) of (1) which can offer improved accuracy by explicitly accounting for factors (such as estimated pole positions) that may otherwise destroy convergence, and hence approximation.

A Proof of Theorem 4.1

Proof. In what follows, C will be used generically to denote a finite constant that may be different in different parts of the same expression. Furthermore, the developments presented below depend on the definition of certain quantities and the employment of certain new results which are given in the companion paper [17] and not repeated here. With this in mind, note that via (8)

$$\sqrt{N}(\widehat{\theta}_N^n - \theta_\circ^n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, P_n), \qquad \text{as } N \to \infty \tag{A.1}$$

where $P_n = R_n^{-1} Q_n R_n^{-1}$, with Q_n being defined via by (11) and R_n being defined in (10). Considering R_n first, then using the developments (22)-(25) of [17]

$$R_n = T_n\left(\Omega_\circ \frac{Z_\circ \Phi_\zeta Z_\circ^\star}{|H_\circ^n|^2} \Omega_\circ^\star\right) - W_n, \qquad W_n \triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mathbf{E}\left\{\varepsilon_t(\theta_\circ^n) \frac{\mathrm{d}\psi_t(\theta^n)}{(\mathrm{d}\theta^n)^T}\right\}. \tag{A.2}$$

However, via the expression (12) and using equation (22) of the companion paper [17]

$$\frac{\mathrm{d}\psi_t(\theta^n)}{(\mathrm{d}\theta^n)^T} = (\Lambda_m(q) \otimes I_4) L(q, \theta_\circ^n)[\lambda_t(\theta_\circ^n) \otimes I_2](\Lambda_m^T(q) \otimes I_4)$$

where

$$\lambda_t(\theta_\circ^n) \triangleq \left[\begin{array}{cc} u_t & 0 \\ 0 & \varepsilon_t(\theta_\circ^n) \end{array} \right],$$

$$L(q, \theta_\circ^n) \triangleq \left[\begin{array}{cccc} 2G(q,\theta_\circ^n)/A^2(q,\theta_\circ^n) & -1/A^2(q,\theta_\circ^n) & & \\ -1/A^2(q,\theta_\circ^n) & 0 & \varnothing & \\ & & 2H(q,\theta_\circ^n)/D^2(q,\theta_\circ^n) & -1/D^2(q,\theta_\circ^n) \\ \varnothing & & -1/D^2(q,\theta_\circ^n) & 0 \end{array} \right].$$

Therefore, after recognising that $\mathbf{E}\{e_t\psi_t(\theta_\circ^n)\} = 0$ because $\psi_t(\theta_\circ^n)$ only depends on $[u_k, e_k + r_k^n]^T$ for $k \leq t-1$, then the application of Lemma A.1 implies

$$W_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} (\Lambda_m(e^{j\omega}) \otimes I_2) L(e^{j\omega}, \theta_\circ^n)[\Phi(\omega) \otimes I_2](\Lambda_m^T(e^{j\omega}) \otimes I_2)\, \mathrm{d}\omega$$

where $\Phi(\omega)$ is formed from the cross spectrum Φ_{ru} between $\{r_t^n\}$ and $\{u_t\}$ as well as the auto-spectrum $\Phi_r(\omega)$ of $\{r_t^n\}$ as

$$\Phi(\omega) = \left[\begin{array}{cc} \Phi_{ru}(\omega) & 0 \\ 0 & \Phi_r(\omega) \end{array} \right].$$

Now, as already argued via equation (19), $\|\Phi(\omega) \otimes I_2\| \leq \max\{|\Phi_{ru}(\omega)|, Cn^{-\beta}\} < C < \infty$. Therefore, for arbitrary $\epsilon > 0$ define the regions

$$\Omega_1 = \{\omega \in [-\pi, \pi] : \|\Phi(\omega) \otimes I_2\| > \epsilon\}, \qquad \Omega_2 = [-\pi, \pi] \setminus \Omega_1$$

and also take $x \in \mathbf{R}^n$ arbitrary but such that $x^T x = 1$. Then by the above formulation and with $C < \infty$

$$\begin{aligned} \left|x^T W_n x\right| \;\leq\; & \epsilon \times \sup_{\omega\in[-\pi,\pi]} \left\|L(e^{j\omega}, \theta_\circ^n)\right\| \cdot \frac{1}{2\pi}\int_{\Omega_2} \left|x^T(\Lambda_m(e^{j\omega}) \otimes I_4)(\Lambda_m^T(e^{j\omega}) \otimes I_4)x\right| \,\mathrm{d}\omega + \\ & \sup_{\omega\in[-\pi,\pi]} \|\Phi(\omega) \otimes I_2\| \cdot \|L(e^{j\omega}, \theta_\circ^n)\| \cdot \frac{1}{2\pi}\int_{\Omega_1} \left|x^T(\Lambda_m(e^{j\omega}) \otimes I_4)(\Lambda_m^T(e^{j\omega}) \otimes I_4)x\right| \,\mathrm{d}\omega. \end{aligned}$$

Now, by the assumptions on the poles $\{\xi_k\}$ and $\{\eta_k\}$, all the elements of $L(e^{j\omega}, \theta_\circ^n)$ are bounded so that $\|L(e^{j\omega}, \theta_\circ^n)\| < \infty$ and hence, for some $C < \infty$ (recall, $m = n/4$) and with $\lfloor\cdot\rfloor$ denoting the integer part of a number

$$\begin{aligned} & \epsilon \sup_{\omega\in[-\pi,\pi]} \|L(e^{j\omega}, \theta_\circ^n)\| \cdot \frac{1}{2\pi}\int_{\Omega_2} \left|x^T(\Lambda_m(e^{j\omega}) \otimes I_4)(\Lambda_m^T(e^{j\omega}) \otimes I_4)x\right| \,\mathrm{d}\omega \\ \leq\; & \frac{C\epsilon}{2\pi}\int_{-\pi}^{\pi} \left|\sum_{k=0}^{n-1}\sum_{\ell=0}^{n-1} x_k x_\ell e^{j\omega(\lfloor k/4\rfloor)} e^{j\omega(\lfloor \ell/4\rfloor)}\right| \,\mathrm{d}\omega \\ =\; & \frac{C\epsilon}{2\pi}\int_{-\pi}^{\pi} \left|\sum_{k=0}^{n-1} x_k e^{j\omega(\lfloor k/4\rfloor)}\right|^2 \,\mathrm{d}\omega \end{aligned}$$

$$
\begin{aligned}
&= C\epsilon \sum_{k=0}^{n-1}\sum_{\ell=0}^{n-1} x_k \overline{x_\ell} \int_{-\pi}^{\pi} e^{j\omega(\lfloor k/4 \rfloor)} e^{-j\omega(\lfloor \ell/4 \rfloor)}\, \mathrm{d}\omega \\
&= C\epsilon \sum_{k=0}^{n-1}\sum_{\ell=0}^{n-1} x_k \overline{x_\ell} \delta\left(\lfloor k/4 \rfloor - \lfloor \ell/4 \rfloor\right) \\
&= C\epsilon \left[\sum_{k=0}^{n-1} |x_k|^2 + 2\sum_{k=0}^{n-2} x_k \overline{x_{k+1}} + 2\sum_{k=0}^{n-3} x_k \overline{x_{k+2}} + 2\sum_{k=0}^{n-4} x_k \overline{x_{k+3}} \right] \leq C\epsilon
\end{aligned}
$$

where the Cauchy-Schwartz inequality and the fact that $x^T x = 1$ was used in the last line.

Now, for the second term over-bounding $|x^T W_n x|$, first note that by assumption 3 of the theorem and with the definition $\widetilde{\Pi}_n(\omega) \triangleq \Pi(e^{j\omega}, \theta_\circ^n) - [G(e^{j\omega}), H(e^{j\omega})]$

$$
\Phi_r(\omega) = |H(e^{j\omega}, \theta_\circ^n)|^{-2} \widetilde{\Pi}_n(\omega) \Phi_\zeta(\omega) \widetilde{\Pi}_n(\omega)^\star \tag{A.3}
$$

$$
|\Phi_{ru}(\omega)|^2 = \Phi_r(\omega)\Phi_u(\omega) + \left| \frac{H(e^{j\omega}) - H(e^{j\omega}, \theta_\circ^n)}{H(e^{j\omega}, \theta_\circ^n)} \right|^2 \left(|\Phi_{ue}(\omega)|^2 - \sigma^2 \Phi_u(\omega) \right). \tag{A.4}
$$

Now by assumption, $\Phi_\zeta(\omega)$ is invertible for all ω so that from (A.3) and assumption 3

$$
\widetilde{\Pi}(\omega)\widetilde{\Pi}(\omega)^\star \leq C\Phi_r(\omega) \leq \frac{C}{n^\beta} \tag{A.5}
$$

Combining this with (A.4) then implies that

$$
|\Phi_{ru}(\omega)| < Cn^{-\beta} \tag{A.6}
$$

Furthermore, by the dyadic construction of $(\Lambda_m(e^{j\omega}) \otimes I_4)(\Lambda_m^T(e^{j\omega}) \otimes I_4)$ it has only four non-zero eigenvalues corresponding to eigenvectors which are the columns of $\Lambda_m(e^{j\omega}) \otimes I_4$. Hence, the eigenvalues are equal and equal to the sum of the squares of the elements of these eigenvectors which itself easily evaluates to $e^{j(m-1)\omega} \sin(m\omega)/\sin(\omega)$. Therefore,

$$
\left| x^T (\Lambda_m(e^{j\omega}) \otimes I_4)(\Lambda_m^T(e^{j\omega}) \otimes I_4) x \right| \leq \left(\frac{\sin \omega m}{\sin \omega} \right) x^T x \leq m
$$

and hence using (A.6) and with μ being Lebesgue measure on $\mathbf{R}$, then for large enough n

$$
\sup_{\omega \in [-\pi,\pi]} \|\Phi(\omega) \otimes I_2\| \cdot \|L(e^{j\omega}, \theta_\circ^n)\| \cdot \frac{1}{2\pi} \int_{\Omega_1} \left| x^T (\Lambda_m(e^{j\omega}) \otimes I_4)(\Lambda_m^T(e^{j\omega}) \otimes I_4) x \right| \, \mathrm{d}\omega
$$

$$
\leq \quad Cn \int_{\Omega_1} \mathrm{d}\omega = Cn\, \mu\left\{\omega \in [-\pi, \pi] : |\Phi_{ru}(\omega)| > \epsilon\right\} \leq \frac{C}{\epsilon n^{1-\beta}}
$$

where Chebychev's inequality and the bound (A.6) was used. Therefore

$$
\left| x^T W_n x \right| \leq C\epsilon + \frac{C}{\epsilon\, n^{1-\beta}}
$$

so that since ϵ and x are arbitrary and since W_n is symmetric then $\|W_n\| \to 0$ as $n \to \infty$. Therefore, employing the $\sim$ notation denoting asymptotic matrix equivalence which that is introduced via

equation (10) of the companion paper [17] (In using this, the choice $\xi_k = 0$ should be made so that $\Gamma_m = \Lambda_m$)

$$R_n \sim T_n\left(\Omega_\circ \frac{Z_\circ \Phi_\zeta Z_\circ^\star}{|H_\circ^n|^2}\Omega_\circ^\star\right) \qquad \text{as } n \to \infty. \tag{A.7}$$

Turning now to Q_n, by recognising that since $\psi_t(\theta_\circ^n)$ depends only on past data it is independent of e_t

$$\begin{aligned}\frac{1}{N}\sum_{t=1}^{N}\sum_{\ell=1}^{N}\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_\ell^T(\theta_\circ^n)\varepsilon_t(\theta_\circ^n)\varepsilon_\ell(\theta_\circ^n)\right\} &= \sigma^2\frac{1}{N}\sum_{t=1}^{N}\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_t^T(\theta_\circ^n)\right\} + \\ &\frac{1}{N}\sum_{t=1}^{N}\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_t^T(\theta_\circ^n)(r_t^n)^2\right\} + \\ &\frac{2}{N}\sum_{\tau=1}^{N-1}\sum_{t=1}^{N-\tau}\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_{t+\tau}^T(\theta_\circ^n)r_t^n r_{t+\tau}^n\right\} + \\ &\frac{1}{N}\sum_{\tau=1}^{N-1}\sum_{t=1}^{N-\tau}\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_{t+\tau}^T(\theta_\circ^n)e_t r_{t+\tau}^n\right\}.\end{aligned}$$

Considering the k,ℓ'th element of the third term in the above expression, by the assumptions on $\{r_t^n\}$ and the boundedness assumption on $\{u_t\}$ combined with the stability of the filters involved in $\psi_t(\theta_\circ^n)$

$$\left|\frac{1}{N}\sum_{t=1}^{N-\tau}\left[\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_{t+\tau}^T r_t^n r_{t+\tau}^n\right\}\right]_{k,\ell}\right| \le \frac{C}{N}\sum_{t=1}^{N}\mathbf{E}\left\{|r_t^n r_{t+\tau}^n|\right\} \le \frac{C}{n^\beta(1+|\tau|^\rho)}$$

so that since the 2-norm $\|\cdot\|$ of a matrix is bounded by the 1-norm which is the maximum absolute row sum

$$\left\|\frac{2}{N}\sum_{\tau=1}^{N-1}\sum_{t=1}^{N-\tau}\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_{t+\tau}^T r_t^n r_{t+\tau}^n\right\}\right\| \le \frac{C}{n^{\beta-1}}\sum_{\tau=1}^{N-1}\frac{1}{1+|\tau|^\rho} \le \frac{C}{n^{\beta-1}} \to 0, \qquad \text{as } n \to \infty.$$

An identical argument applies to the terms involving $\mathbf{E}\{\psi_t(\theta_\circ^n)\psi_{t+\tau}^T e_t r_{t+\tau}^n\}$ and $\mathbf{E}\{\psi_t(\theta_\circ^n)\psi_t^T(r_t^n)^2\}$ so that asymptotically in n the only component of Q_n to contribute to quadratic forms involving it will be

$$Q_n \sim \lim_{N\to\infty}\frac{\sigma^2}{N}\sum_{t=1}^{N}\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_t^T(\theta_\circ^n)\right\} = \sigma^2 T_n\left(\Omega_\circ\frac{Z_\circ\Phi_\zeta Z_\circ^\star}{|H_\circ^n|^2}\Omega_\circ^\star\right).$$

Therefore

$$P_n \sim R_n^{-1}Q_nR_n^{-1} = \sigma^2 T_n\left(\Omega_\circ\frac{Z_\circ\Phi_\zeta Z_\circ^\star}{|H_\circ^n|^2}\Omega_\circ^\star\right). \tag{A.8}$$

Consequently, using (A.1) together with this expression for P_n and the Taylor expansion in (14) leads to

$$\sqrt{N}\left[\Pi^T(e^{j\omega},\widehat{\theta}_N^n) - \Pi^T(e^{j\omega},\theta_\circ^n)\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Delta_m(\omega)) \tag{A.9}$$

where, according to the developments leading to equation (29) of [17]

$$\lim_{m\to\infty}\Delta_m(\omega) = \sigma^2\lim_{m\to\infty}|H(e^{j\omega},\theta_\circ^n)|^2\varphi_m(\omega,\omega)$$

where

$$\varphi_m(\lambda,\omega) \triangleq \frac{Z_\circ^\star(e^{j\omega})}{H(e^{-j\omega},\theta_\circ^n)}\Omega_\circ^\star(e^{j\omega})[\Lambda_m^\star(e^{j\omega})\otimes I_4]T_n^{-1}\left(\Omega_\circ\frac{Z_\circ\Phi_\zeta Z_\circ^\star}{|H_\circ^n|^2}\Omega_\circ^\star\right)[\Lambda_m(e^{j\lambda})\otimes I_4]\Omega_\circ(e^{j\lambda})\frac{Z_\circ(e^{j\lambda})}{H(e^{j\lambda},\theta_\circ^n)}.$$

Therefore, by employing Theorem 6.1 of [17]

$$\lim_{m\to\infty}\Delta_m(\omega)K_m^{-1}(\omega) = \sigma^2|H_\circ(e^{j\omega})|^2\Phi_\zeta^{-1}(\omega). \tag{A.10}$$

Extending this result to consider neighbouring frequencies ω and λ may then be achieved by using (A.9) to conclude that

$$\sqrt{N}\left[\begin{array}{c}\Pi^T(e^{j\omega},\widehat{\theta}_N^n)-\Pi^T(e^{j\omega},\theta_\circ^n)\\ \Pi^T(e^{j\lambda},\widehat{\theta}_N^n)-\Pi^T(e^{j\lambda},\theta_\circ^n)\end{array}\right]\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Sigma_m(\omega,\lambda))$$

where

$$\Sigma_m(\omega,\lambda)\triangleq\sigma^2\left[\begin{array}{cc}\varphi_m(\omega,\omega) & \varphi_m(\omega,\lambda)\\ \varphi_m(\lambda,\omega) & \varphi_m(\lambda,\lambda)\end{array}\right].$$

Furthermore, since

$$\Sigma_m(\omega,\lambda)\left[\begin{array}{cc}K_m^{-1}(\omega) & \varnothing\\ \varnothing & K_m^{-1}(\lambda)\end{array}\right]=\sigma^2\left[\begin{array}{cc}\varphi_m(\omega,\omega)K_m^{-1}(\omega) & \varphi_m(\omega,\lambda)K_m^{-1}(\lambda)\\ \varphi_m(\lambda,\omega)K_m^{-1}(\omega) & \varphi_m(\lambda,\lambda)K_m^{-1}(\lambda)\end{array}\right],$$

then employment of (56) of Theorem 6.1 of [17] establishes that the asymptotic (as $m\to\infty$) values of the diagonal entries of the above quantity are (respectively) $\sigma^2|H_\circ(e^{j\omega})|^2\Phi_\zeta^{-1}(\omega)$ and $\sigma^2|H_\circ(e^{j\lambda})|^2\Phi_\zeta^{-1}(\lambda)$, while those on the off-diagonal tend asymptotically to zero. □

Lemma A.1. *Suppose that $\{u_t\}$ and $\{y_t\}$ are jointly quasi-stationary processes with bounded spectral densities and which are filtered as*

$$x_t = H(q)u_t, \qquad z_t = G(q)y_t$$

where both filters have absolutely summable and real-valued impulse responses. Then

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{E}\{x_tz_t\} = \frac{1}{2\pi}\int_{-\pi}^{\pi}H(e^{j\omega})\overline{G(e^{j\omega})}\Phi_{uy}(\omega)\,\mathrm{d}\omega.$$

Proof. By the assumptions of absolute summability and bounded-ness of spectral densities, ensuing re-orderings of summations and exchanges of integrals and summations are justified so as to calculate:

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\mathbf{E}\{x_tz_t\} &= \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sum_{\ell=0}^{\infty}\sum_{m=0}^{\infty}h_\ell g_m\mathbf{E}\{u_{t-\ell}y_{t-m}\}\\ &= \sum_{\ell=0}^{\infty}\sum_{m=0}^{\infty}h_\ell g_m\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\mathbf{E}\{u_{t-\ell}y_{t-m}\}\\ &= \sum_{\ell=0}^{\infty}\sum_{m=0}^{\infty}h_\ell g_m\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_{uy}(\omega)e^{j\omega(m-\ell)}\,\mathrm{d}\omega\\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}H(e^{j\omega})\overline{G(e^{j\omega})}\Phi_{uy}(\omega)\,\mathrm{d}\omega.\end{aligned}$$

□

B Proof of Corollary 4.2

Proof. As before, C denotes a finite constant that may be different in different parts of the same expression, and the ensuing developments below depend on the definition of certain quantities and the employment of certain new results which are given in the companion paper [17] and not repeated here. Since $H(q,\theta)$ is fixed at $H(q,\theta) = H_*(q)$, then (A.2) becomes

$$R_n = T_n\left(\frac{Z_\circ \Phi_u Z_\circ^\star}{|A_\circ|^2}\right) - W_n, \qquad W_n \triangleq \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \mathbf{E}\left\{\varepsilon_t(\theta_\circ^n)\frac{\mathrm{d}\psi_t(\theta^n)}{(\mathrm{d}\theta^n)^T}\right\} \tag{B.1}$$

where now

$$Z(q,\theta) \triangleq \begin{bmatrix} -G(q,\theta) \\ 1 \end{bmatrix} \tag{B.2}$$

and

$$\frac{\mathrm{d}\psi_t(\theta^n)}{(\mathrm{d}\theta^n)^T} = (\Lambda_m(q)\otimes I_2)L(q,\theta_\circ^n)(\Lambda_m^T(q)\otimes I_2)A^{-2}(q,\theta_\circ^n)u_t, \qquad L(q,\theta_\circ^n) \triangleq \begin{bmatrix} 2G(q,\theta_\circ^n) & -1 \\ -1 & 0 \end{bmatrix}$$

so that according to Lemma A.1

$$W_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} (\Lambda_m(e^{j\omega})\otimes I_2)L(e^{j\omega},\theta_\circ^n)(\Lambda_m^T(e^{j\omega})\otimes I_2)\frac{\Phi_{\varepsilon u}(\omega)}{A^2(e^{j\omega},\theta_\circ^n)}\,\mathrm{d}\omega.$$

Furthermore, under assumption 3 of Theorem 4.1 and using the condition that $\Phi_{ue}(\omega) = 0$ then

$$\Phi_{\varepsilon u}(\omega) = H_*(e^{j\omega})^{-1}\left[G(e^{j\omega}) - G(e^{j\omega},\theta_\circ^n)\right]\Phi_u(\omega) \tag{B.3}$$

so that according to the assumptions of the corollary

$$|\Phi_{\varepsilon u}(\omega)| \le C\left|G(e^{j\omega}) - G(e^{j\omega},\theta_\circ^n)\right| \le \frac{C}{n^\beta}. \tag{B.4}$$

Therefore, by the same argument as employed between (A.2) and (A.7)

$$R_n \sim T_n\left(\frac{Z_\circ \Phi_u Z_\circ^\star}{|A_\circ|^2}\right) \qquad \text{as } n\to\infty. \tag{B.5}$$

Turning now to Q_n, defining $\varepsilon_t = \delta_t^n + \epsilon_t$ where $\delta_t^n \triangleq H_*^{-1}(q)[G(q) - G(q,\theta_\circ^n)]u_t$, $\epsilon_t \triangleq H_*^{-1}(q)H(q)e_t$ leads to

$$\begin{aligned} Q_n = \frac{1}{N}\sum_{t=1}^{N}\sum_{\ell=1}^{N}\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_\ell^T(\theta_\circ^n)\varepsilon_t(\theta_\circ^n)\varepsilon_\ell(\theta_\circ^n)\right\} &= \frac{1}{N}\sum_{t=1}^{N}\sum_{\ell=1}^{N}\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_\ell^T(\theta_\circ^n)\epsilon_t\epsilon_\ell\right\} + \\ &\quad \frac{1}{N}\sum_{t=1}^{N}\sum_{\ell=1}^{N}\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_\ell^T(\theta_\circ^n)\delta_t^n\delta_\ell^n\right\} \end{aligned}$$

where the assumption of open loop data collection implying $\mathbf{E}\{u_t\epsilon_t\} = 0$ has been employed. Furthermore, by the assumptions of the Corollary

$$\mathbf{E}\left\{(\delta_t^n)^2\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{G(e^{j\omega}) - G(e^{j\omega},\theta_\circ^n)}{H_*(e^{j\omega})}\right|^2 \mathrm{d}\omega \le \frac{C}{n^{2\beta}}.$$

Therefore, application of the Cauchy–Schwartz inequality and Lemma A.1 of [19] indicate that

$$Q_n \sim T_n\left(\frac{Z_\circ \Phi_u \Phi_\nu Z_\circ^\star}{|H_*|^2|A_\circ|^2}\right) \qquad \text{as } n \to \infty.$$

Consequently, using (A.1) and the Taylor expansion in (14) leads to

$$\sqrt{N}\begin{bmatrix} G(e^{j\omega},\widehat{\theta}_N^n) - G(e^{j\omega},\theta_\circ^n) \\ G(e^{j\lambda},\widehat{\theta}_N^n) - G(e^{j\lambda},\theta_\circ^n) \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Delta_m(\omega,\lambda)) \tag{B.6}$$

where, according to the developments leading to equation (29) of [17]

$$\begin{aligned}\Delta_m(\omega,\lambda) &= \frac{\sigma^2}{|A(e^{j\omega},\theta_\circ^n)|^2} Z_\circ^\star(e^{j\omega})[\Lambda_m^\star(e^{j\omega})\otimes I_2]T_n^{-1}\left(\frac{Z_\circ\Phi_u Z_\circ^\star}{|A_\circ|^2}\right)\times \\ &\quad T_n\left(\frac{Z_\circ\Phi_u\Phi_\nu Z_\circ^\star}{|H_*|^2|A_\circ|^2}\right)T_n^{-1}\left(\frac{Z_\circ\Phi_u Z_\circ^\star}{|A_\circ|^2}\right)[\Lambda_m(e^{j\lambda})\otimes I_2]Z_\circ(e^{j\lambda}) \qquad \text{(B.7)} \\ &= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\frac{Z_\circ^\star(e^{j\omega})}{A_\circ^\star(e^{j\omega})}[\Lambda_m^\star(e^{j\omega})\otimes I_2]T_n^{-1}\left(\frac{Z_\circ\Phi_u Z_\circ^\star}{|A_\circ|^2}\right)[\Lambda_m(e^{j\mu})\otimes I_2]\frac{Z_\circ(e^{j\mu})}{A_\circ(e^{j\mu})}\times \\ &\quad \frac{\Phi_u(\mu)\Phi_\nu(\mu)}{|H_*(\mu)|^2}\frac{Z_\circ^\star(e^{j\mu})}{A_\circ(e^{j\mu})}[\Lambda_m^\star(e^{j\mu})\otimes I_2]T_n^{-1}\left(\frac{Z_\circ\Phi_u Z_\circ^\star}{|A_\circ|^2}\right)[\Lambda_m(e^{j\lambda})\otimes I_2]\frac{Z_\circ(e^{j\lambda})}{A_\circ(e^{j\lambda})}\,\mathrm{d}\mu.\end{aligned}$$

However, according to the reproducing kernel arguments in §6 of [17] which employs a rational orthonormal basis re-parameterisation detailed in §5,6 of [17]

$$\frac{Z_\circ^\star(e^{j\omega})}{A_\circ^\star(e^{j\omega})}[\Lambda_m^\star(e^{j\omega})\otimes I_2]T_n^{-1}\left(\frac{Z_\circ\Phi_u Z_\circ^\star}{|A_\circ|^2}\right)[\Lambda_m(e^{j\mu})\otimes I_2]\frac{Z_\circ(e^{j\mu})}{A_\circ(e^{j\mu})} = \Gamma_{2m}^\star(\omega)M_n^{-1}(\Phi_u)\Gamma_m(\mu)$$

which when substituted into the above expression gives

$$\begin{aligned}\Delta_m(\omega,\lambda) &= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\Gamma_{2m}^\star(\omega)M_n^{-1}(\Phi_u)\Gamma_{2m}(\mu)\frac{\Phi_u(\mu)\Phi_\nu(\mu)}{|H_*(\mu)|^2}\Gamma_{2m}^\star(\mu)M_n^{-1}(\Phi_u)\Gamma_{2m}(\lambda)\,\mathrm{d}\mu. \\ &= \Gamma_{2m}^\star(\omega)M_n^{-1}(\Phi_u)M_n\left(\frac{\Phi_u\Phi_\nu}{|H_*|^2}\right)M_n^{-1}(\Phi_u)\Gamma_{2m}(\lambda).\end{aligned}$$

However, according to Theorem 5.2 of [20]

$$M_n^{-1}(\Phi_u)M_n\left(\frac{\Phi_u\Phi_\nu}{|H_*|^2}\right)M_n^{-1}(\Phi_u) \sim M_n\left(\frac{\Phi_\nu}{\Phi_u|H_*|^2}\right) \qquad \text{as } n\to\infty$$

so that by Theorem 2.2 of [17] (see also Theorem 4.3 of [20])

$$\lim_{m\to\infty}\frac{\Delta_m(\omega,\lambda)}{2\kappa_m(\omega)} = \lim_{m\to\infty}\frac{1}{2\kappa_m(\omega)}\Gamma_{2m}^\star(\omega)M_n\left(\frac{\Phi_\nu}{\Phi_u|H_*|^2}\right)\Gamma_{2m}(\lambda) = \begin{cases}\dfrac{\Phi_\nu(\omega)}{\Phi_u(\omega)|H_*(e^{j\omega})|^2} & ;\omega=\lambda \\ 0 & ;\omega\neq\lambda.\end{cases}$$

This establishes (34), and (35) then follows from this via the same argument as used in the proof of Corollary 4.1. □

C Proof of Corollary 4.3

Proof. Referring to the proof of Theorem 4.1, under the assumption of $\varepsilon_t(\theta_\circ^n) = e_t$, then $Q_n = \sigma^2 R_n$ and hence $P_n = \sigma^2 R_n^{-1}$ with, again since $\varepsilon_t(\theta_\circ^n) = e_t$

$$R_n = T_n^{-1}\left(\Omega_\circ \frac{Z_\circ \Phi_\zeta Z_\circ^\star}{|H_\circ^n|^2}\Omega_\circ^\star\right).$$

Therefore, using the Taylor expansion ideas leading to (A.9) and then the argument of Appendix 9B of [9] employing the assumption of $\mathbf{E}\left\{|e_t|^8\right\} < \infty$ to allow passage from distributional convergence to mean-square convergence then implies that

$$\lim_{N\to\infty} N \cdot \mathrm{Cov}\left\{\begin{bmatrix} G(e^{j\omega}, \widehat{\theta}_N^n) \\ H(e^{j\omega}, \widehat{\theta}_N^n) \end{bmatrix}\right\} = \Delta_m(\omega) \tag{C.1}$$

where, according to the developments leading to equation (29) of [17]

$$\Delta_m(\omega) = \sigma^2 |H(e^{j\omega}, \theta_\circ^n)|^2 \varphi_m(\omega, \omega)$$

with

$$\varphi_m(\lambda,\omega) \triangleq \frac{Z_\circ^\star(e^{j\omega})}{H(e^{-j\omega},\theta_\circ^n)}\Omega_\circ^\star(e^{j\omega})[\Lambda_m^\star(e^{j\omega})\otimes I_4]T_n^{-1}\left(\Omega_\circ \frac{Z_\circ \Phi_\zeta Z_\circ^\star}{|H_\circ^n|^2}\Omega_\circ^\star\right)[\Lambda_m(e^{j\lambda})\otimes I_4]\Omega_\circ(e^{j\lambda})\frac{Z_\circ(e^{j\lambda})}{H(e^{j\lambda},\theta_\circ^n)}$$

and then, by Theorem 6.1 of [17]

$$\Delta_m(\omega) = \sigma^2 |H_\circ(e^{j\omega})|^2 \Phi_\zeta^{-1}(\omega) K_m(\omega). \tag{C.2}$$

□

References

[1] D. R. BRILLINGER, *Time Series: Data Analysis and Theory*, Holden-Day, 1981.

[2] P. CAINES, *Linear Stochastic Systems*, John Wiley and Sons, New York, 1988.

[3] U. FORSELL AND L. LJUNG, *Closed-loop identification revisited*, Automatica, 53 (1999), pp. 1215–1241.

[4] U. FORSSELL AND L. LJUNG, *Some results on optimal experiment design*, Automatica, 36 (2000), pp. 749–756.

[5] M. GEVERS, *Essays on control: Perspectives in the theory and its applications*, Birkhäuser, Boston, 1993, ch. Towards a joint design of identification and Control?, pp. 111–151.

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

[7] E. HANNAN AND D. NICHOLLS, *The estimation of the prediction error variance*, Journal of the American Statistical Association, 72 (1977), pp. 834–840.

[8] H. HJALMARSSON, M. GEVERS, AND F. DE BRUYNE, *For some model based control design criteria, closed loop identification gives better performance*, Automatica, 32 (1996), pp. 1659–1673.

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

[10] ——, *MATLAB System Identification Toolbox Users Guide, Version 5*, The Mathworks, 2000.

[11] L. LJUNG AND L. GUO, *The role of model validation for assessing the size of the unmodeled dynamics*, IEEE Trans.Automatic Control, AC-42 (1997), pp. 1230–1240.

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

[13] ——, *Asymptotic variance expressions for identified black-box transfer function models*, IEEE Transactions on Automatic Control, AC-30 (1985), pp. 834–844.

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

[15] 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.

[16] L.LJUNG AND Z.D.YUAN, *Asymptotic properties of black-box identification of transfer functions*, IEEE Transactions on Automatic Control, 30 (1985), pp. 514–530.

[17] B. NINNESS AND H. HJALMARSSON, *The analysis of variance error part II: Fundamental principles*, Submitted to IEEE Transactions on Automatic Control, (2001).

[18] B. NINNESS AND H. HJALMARSSON, *Exact quantification of variance error*, in Proceedings of the IFAC World Congress, 2002.

[19] 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.

[20] ——, *Generalised Fourier and Toeplitz results for rational orthonormal bases*, SIAM Journal on Control and Optimization, 37 (1999), pp. 429–460.

[21] P.M.J. VAN DEN HOF, P.S.C. HEUBERGER, AND J. BOKOR, *System identification with generalized orthonormal basis functions*, Automatica, 31 (1995), pp. 1821–1834.

[22] T.SÖDERSTRÖM AND P.STOICA, *System Identification*, Prentice Hall, New York, 1989.

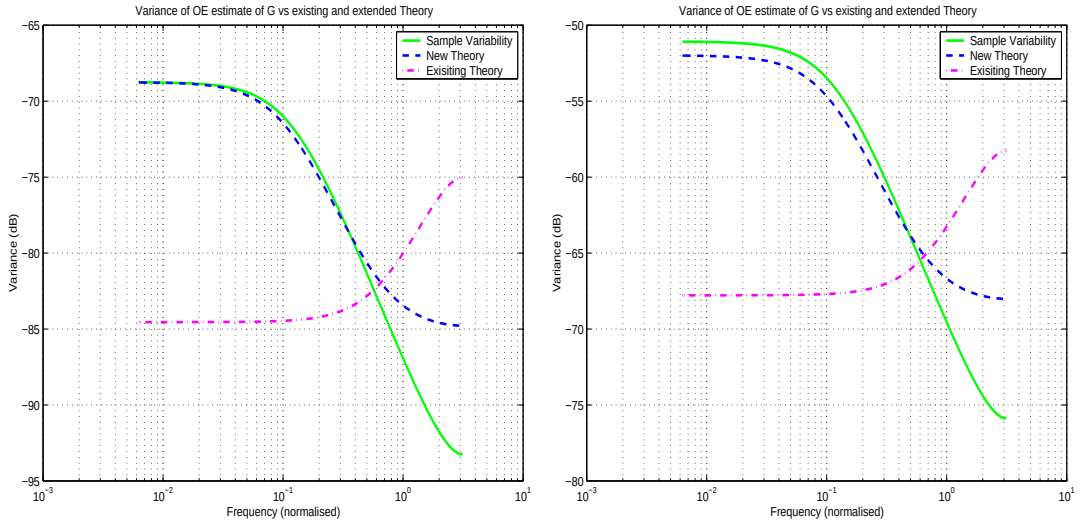
[23] B. WAHLBERG, *System identification using Laguerre models*, IEEE Transactions on Automatic Control, AC-36 (1991), pp. 551–562.

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

[25] L.-L. XIE AND L. LJUNG, *Asymptotic variance expressions for estimated frequency functions*, IEEE Transactions on Automatic Control, 46 (2001), pp. 1887–1899.

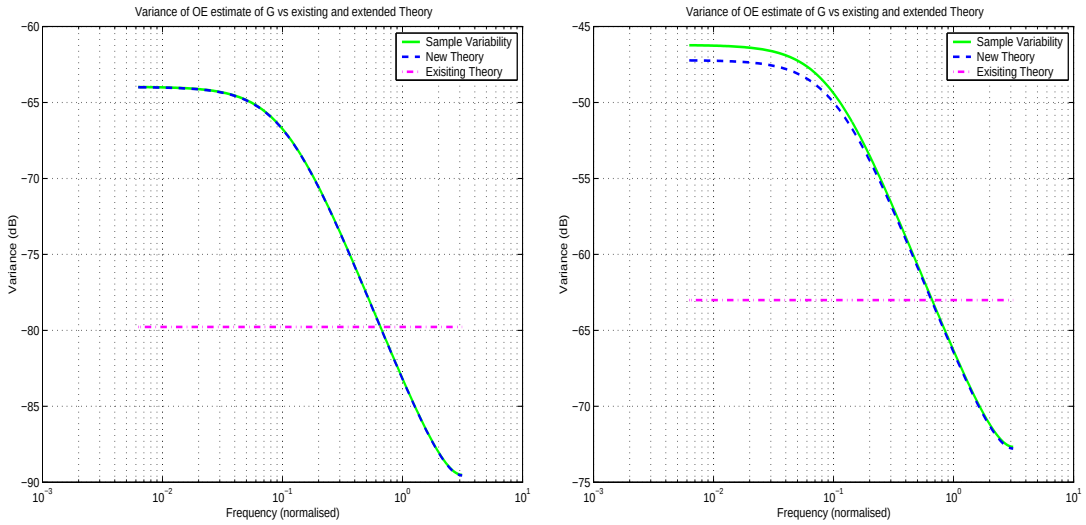
[26] Y. ZHU, *Multivariable process identification for MPC: the asymptotic method and its applications*, Journal of Process Control, 8 (1998), pp. 101–115.

[27] Y. ZHU AND P. VAN DEN BOSCH, *Optimal closed–loop identification test design for internal model control*, Automatica, 36 (2000), pp. 1237–1241.



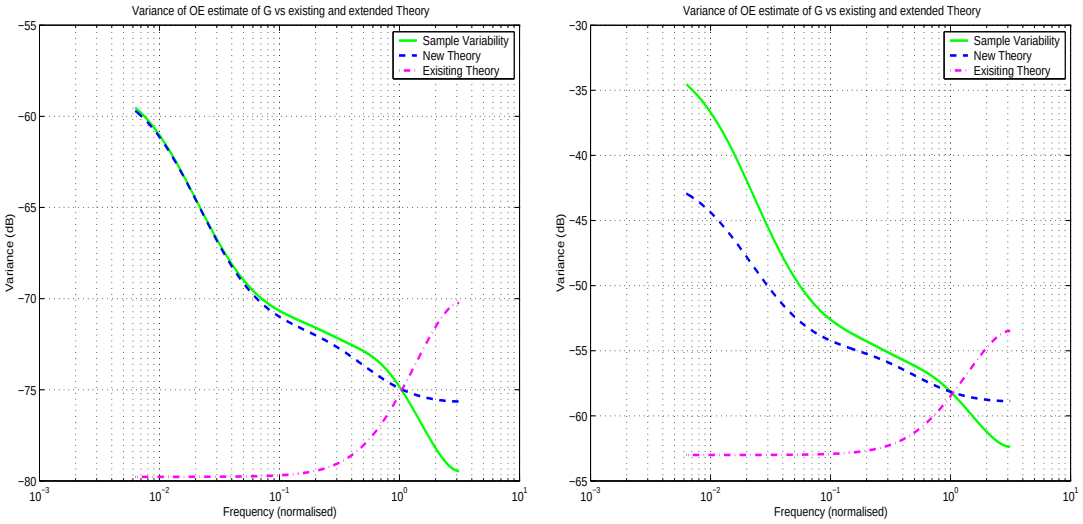
(a) Low-Order, Coloured Φ_u, $N = 10,000$ (b) Low-Order, Coloured Φ_u, $N = 200$

Figure 2: *System 1, very low-order. True variability is solid line, new quantification (3) is the dashed line, and the existing quantification (1) is the dash-dot line.*



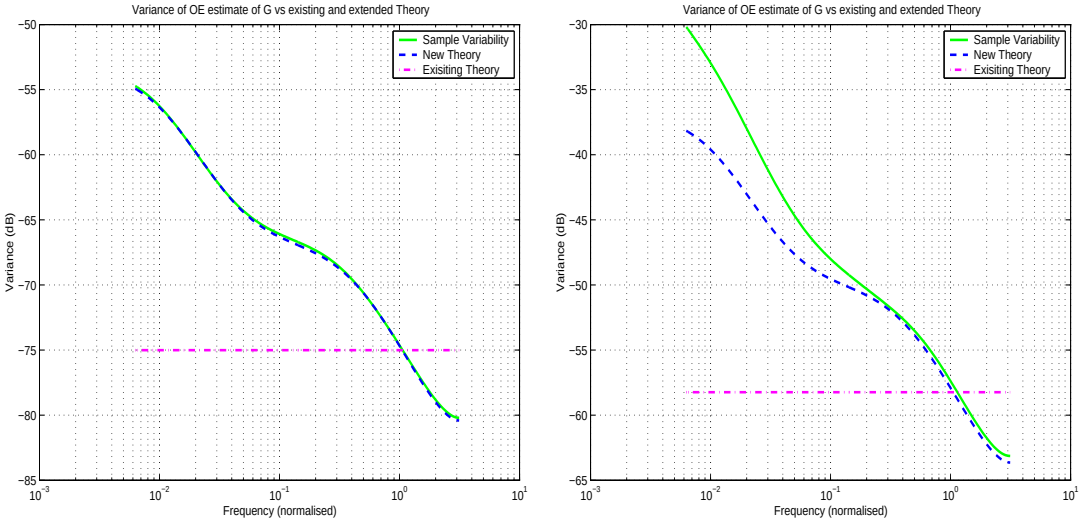
(a) Low-Order, White Φ_u, $N = 10,000$ (b) Low-Order, White Φ_u, $N = 200$

Figure 3: *System 1, very low-order. True variability is solid line, new quantification (3) is the dashed line, and the existing quantification (1) is the dash-dot line.*



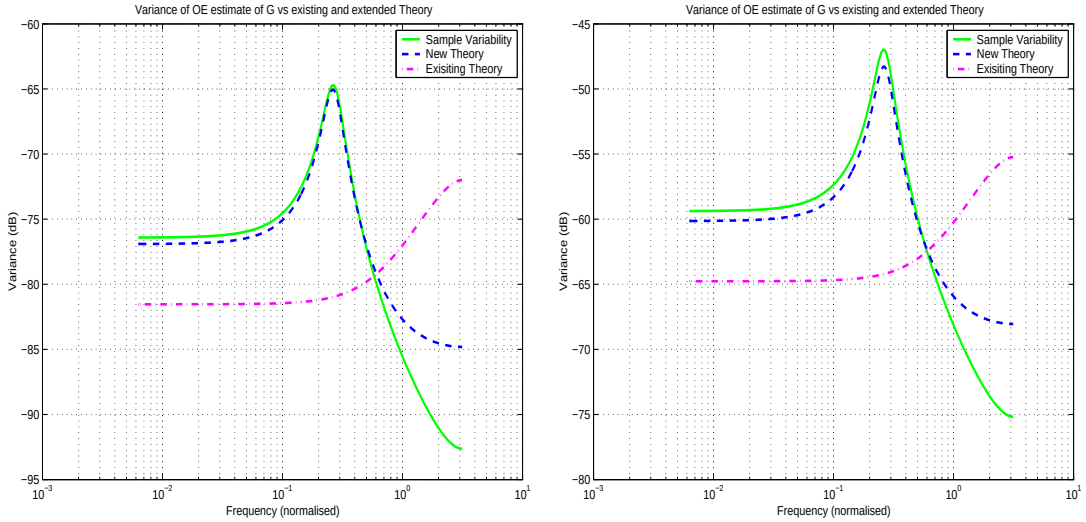
(a) Mid-Order, Coloured Φ_u, $N = 10,000$ (b) Mid-Order, Coloured Φ_u, $N = 200$

Figure 4: *System 2, mid-order. True variability is solid line, new quantification (3) is the dashed line, and the existing quantification (1) is the dash-dot line.*



(a) Mid-Order, White Φ_u, $N = 10,000$ (b) Mid-Order, White Φ_u, $N = 200$

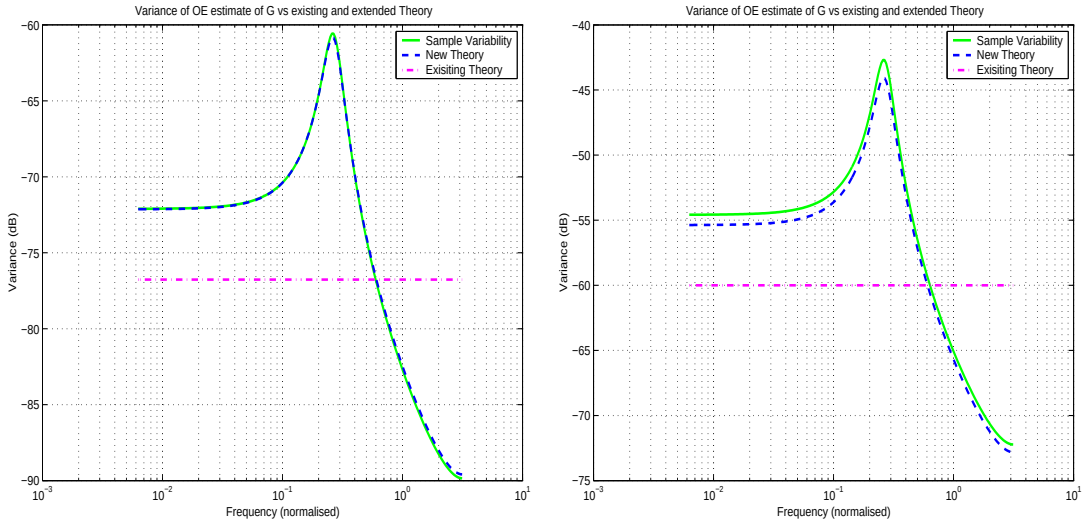
Figure 5: *System 2, mid-order. True variability is solid line, new quantification (3) is the dashed line, and the existing quantification (1) is the dash-dot line.*



(a) Low-Order Resonant, Coloured Φ_u, $N = 10,000$

(b) Low-Order Resonant, Coloured Φ_u, $N = 200$

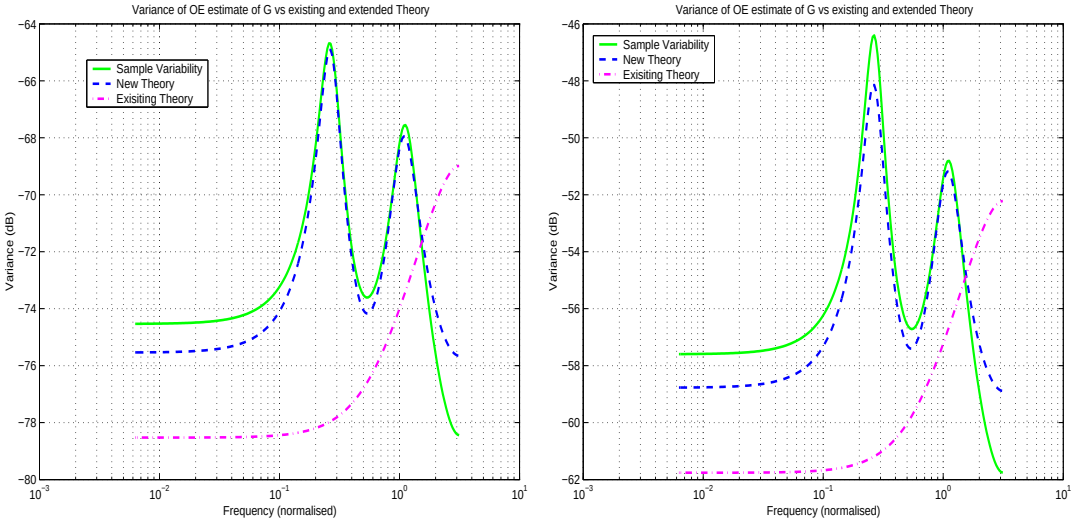
Figure 6: *System 3, low-order resonant. True variability is solid line, new quantification (3) is the dashed line, and the existing quantification (1) is the dash-dot line.*



(a) Low-Order Resonant, White Φ_u, $N = 10,000$

(b) Low-Order Resonant, White Φ_u, $N = 200$

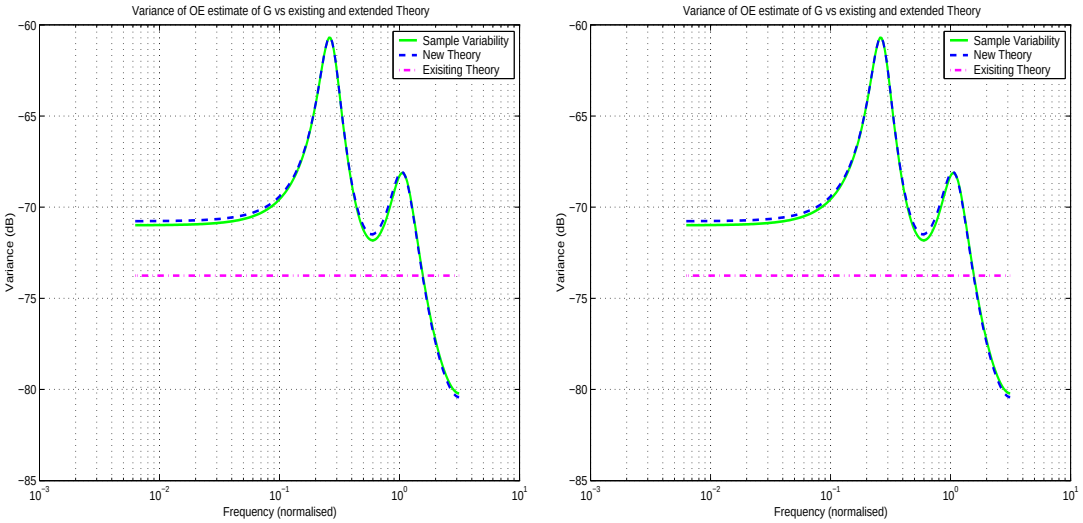
Figure 7: *System 3, low-order resonant. True variability is solid line, new quantification (3) is the dashed line, and the existing quantification (1) is the dash-dot line.*



(a) Mid-Order Resonant, Coloured Φ_u, $N = 10,000$

(b) Mid-Order Resonant, Coloured Φ_u, $N = 200$

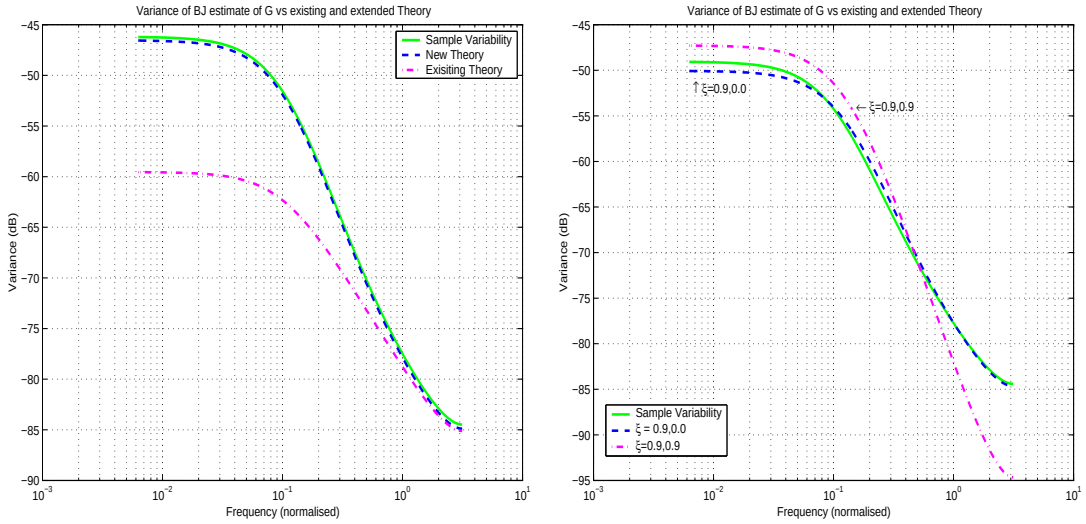
Figure 8: *System 4, mid-order resonant. True variability is solid line, new quantification (3) is the dashed line, and the existing quantification (1) is the dash-dot line.*



(a) Mid-Order Resonant, White Φ_u, $N = 10,000$

(b) Mid-Order Resonant, White Φ_u, $N = 200$

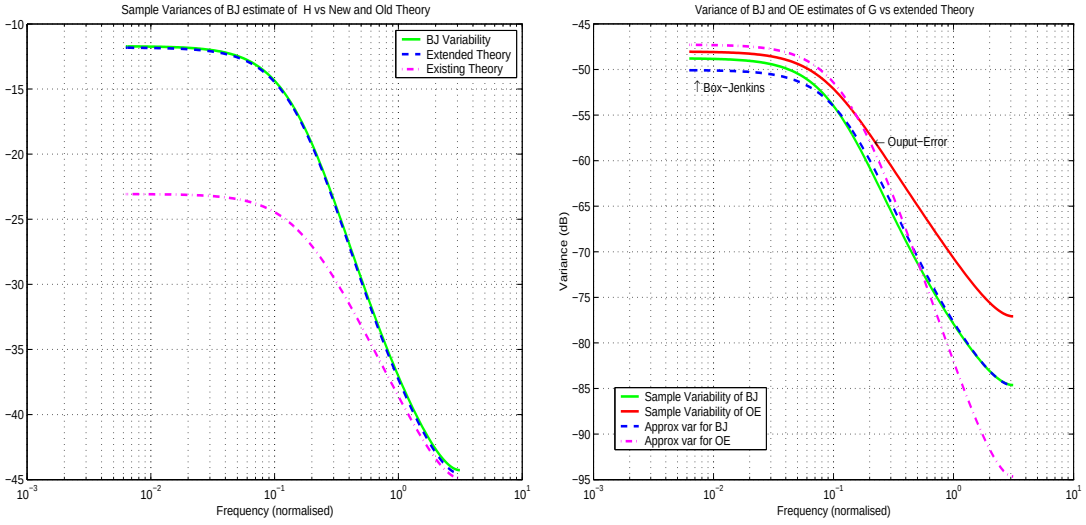
Figure 9: *System 4, mid-order resonant. True variability is solid line, new quantification (3) is the dashed line, and the existing quantification (1) is the dash-dot line.*



(a) Low Order, $H(q) = q(q - 0.9)^{-1}$

(b) Low Order, $H(q) = q(q - 0.85)^{-1}$

Figure 10: *System 1, low order, with coloured measurement noise and estimation using Box–Jenkins model structure. In both figures, the true variability is solid line and the new quantification (28) is the dashed line. On the left the existing quantification (1) is the dash-dot line, while on the right it is the new quantification (28) but with a different choice for $A_\dagger$ as used for the dashed line approximation.*



(a) Low Order, $H(q) = q(q - 0.85)^{-1}$, Noise model variability

(b) Low Order, $H(q) = q(q - 0.85)^{-1}$, Box–Jenkins and Output-Error

Figure 11: *System 1, low order, with coloured measurement noise and estimation using both Box–Jenkins and Output–Error model structures. In both figures, the true variability is solid line and the new quantification (28) is the dashed line for the Box–Jenkins case, and the dash-dot line for the Output-Error case.*