How poorly does my forecast perform under a wrong ARIMA model? A decomposition-based simulation algorithm for assessment

Juan Manuel González Sopeña*, Tirthankar Dasgupta**, Vikram Pakrashi***, Han Xiao** and Bidisha Ghosh*

* Civil Structural & Environmental Engineering , Trinity College, Dubin, Ireland

**Department of Statistics, Rutgers University , New Jersey, USA

***School of Mechanical and Materials Engineering , University College Dublin, Ireland

Abstract

We investigate the problem of quantifying the inflation in the uncertainty associated with a forecast made from a possibly misspecified ARIMA model with respect to the uncertainty associated with a forecast made from the true model. We also decompose this inflation into two components associated with model misspecification and uncertainty of parameter estimation. The decomposition is based on a definition of "optimal" parameters of a misspecified model. We provide a simulation algorithm that would allow a researcher or practitioner to assess the consequences of using an incorrect ARIMA model with respect to an assumed true model in terms of the inflation in forecast error. We apply the algorithm to study the consequences of (i) using an autoregressive model of high order to approximate an ARMA model and (ii) making traffic volume forecasts on the basis of misspecified seasonal ARIMA models.

1 Introduction

In the age of data science, as time series forecasting finds an increasing number of widespread applications across various disciplines ranging from engineering to social sciences, interest in understanding and quantifying the uncertainty associated with such forecasts is also on the rise. Most applications of time series analysis consist of the following steps: (i) model specification or identification, (ii) model fitting, (iii) model checking or validation and (iv) using the fitted model to make an h-step ahead forecast, and estimate the associated forecast error.

One of the sources of forecast error is model misspecification (Chatfield, 1996). For example, suppose the data are actually generated by an auto regressive integrated moving average (ARIMA) model of the order (p, d, q), and is misspecified as an ARIMA(p^*, d^*, q^*) model. Although there are various tools available for model specification or identification that identify the true model almost surely from an asymptotic perspective, it is not uncommon for an analyst to identify an incorrect model from a sample of small size. Pukkila et al. (1990) suggested a method for determining the order of an ARIMA$(p, 0, q)$ or ARMA(p, q) model, that was shown to perform well with samples of size 100 or more for $p + q \leq 3$. However, for samples of size 50, the performance wasn't as good - e.g., an ARMA(1,2) model with the AR parameter $\phi_1 = 0.60$ and the MA parameters $\theta_1 = -0.50$ and $\theta_2 = -0.90$ was correctly identified only 34% of the time from samples of size 50. On the remaining occasions, the model was incorrectly identified as ARMA(2,0), ARMA(0,2), ARMA(1,0) and ARMA(3,0). An interesting question that immediately arises is, how much does the prediction suffer if one of these incorrect models is used.

It has sometimes been argued that incorrect model specification is sometimes a consequence of a nearly equivalent mathematical representation of the true model. For example, Kendall (1971) had argued that the time and effort spent in identifying the correct order of ARMA models can be saved by fitting moderately long autoregressive models. However, many authors, including Box and Jenkins (1973) have provided a counter-argument that such non-parsimonious models result in noisy forecasts due to uncertainty involved in estimation of a large number of model parameters. Thus, even if two models have almost similar mathematical representations, resulting in nearly equal forecast error, if the misspecified model is less parsimonious, the forecast error is likely to be inflated due to the uncertainty associated with estimation of model parameters. This aspect is often ignored in modern data science, where overfitting is considered benign by many.

This brings us to the question we want to investigate in this paper: if one makes a forecast using a time series model M, assumed different from the true data generating model, what are the sources of the forecast error and how is the forecast error distributed among these sources? Chatfield (1996) points out that in time series analysis, "there are primarily three sources of uncertainty: (1) Uncertainty about the structure of the model, (2) Uncertainty about estimates of the model parameters, assuming the model structure is known, and (3) Uncertainty about the data even when the model structure and the values of the model parameters are known." Therefore a more careful formulation of the above question is: can the forecast error associated with a possibly misspecified model be decomposed into the above three sources of uncertainty? If so, how?

Such a decomposition is useful in many real-life scenarios. One of the motivating examples we consider comes from traffic engineering. Traffic flow or traffic volume is referred to as the total number of vehicles crossing a point on a road-section over unit time (essentially, it is the traffic flow rate). Traffic engineers have extensively studied and modeled traffic data using simple ARIMA (Hamed et al., 1995) and seasonal ARIMA (SARIMA) models. However, the proposed models typically differ with respect to their autoregressive and moving average orders. Assuming that one of these models is true and the inherent error is the same, it is of practical interest to see how much the forecast error will be inflated if other models are chosen, and to which sources (model uncertainty and parameter estimation uncertainty) these errors will be attributed. This insight can be of relevance for other fields such as ecology (Mac Nally et al., 2018) or dynamical systems in general (Mangan et al., 2017), where there is an aim to understand how to better select models.

The goal of this paper is to propose a framework that would (a) allow decomposition of forecast error (measured by the mean squared error or MSE) from an arbitrary time series model into three components, and (b) outline a simulation procedure that would enable a researcher to decompose the forecast error of a proposed model, and compare it with that of a *benchmark model*. Surprisingly, in spite of the existence of a fairly large literature on time series model misspecification, a decomposition of forecast error like the one described above appears to have been scantily addressed and discussed. Our proposed framework is based on ideas similar to that developed by Davies and

Newbold (1980), who studied how forecast errors are inflated if misspecified ARIMA models are used in lieu of true data generating models. Our main contributions include: (i) providing a decomposition formula that helps to quantify the percentage contributions of the three aforementioned sources of forecast error, (ii) providing a new definition of "optimal parameters of a misspecified model" used in (i), and (iii) developing a simulation algorithm that can be used to estimate the percentage contributions of each source in the decomposition. We also discuss two applications of our proposed approach - one pertaining to a historical debate in statistics and the one from traffic engineering, briefly introduced in the previous paragraph. We restrict the discussion to the class of ARIMA and SARIMA models.

In the following section we introduce some notation, briefly describe the basics of forecasting with ARMA models and introduce the notion of parameter uncertainty as a component of forecast error. In Section 3, we explore the effect of model misspecification on forecast error, introduce the notion of "optimal misspecified model", and propose measures of inflation of the forecast error arising from parameter and model uncertainty. In Section 4, we lay out a comprehensive simulation framework to assess the source-wise inflation of error of forecast made from a misspecified model. In Section 5, we describe two applications of the proposed simulation framework. Some concluding remarks are presented in 6.

2 Notation, forecasting with ARMA models and contribution of parameter uncertainty in forecast error

We introduce some notation that can be found in most common and well-known time series textbooks like Box and Jenkins (1970) and Brockwell and Davis (2002). Suppose we have observed $Y_1, \ldots, Y_n$, i.e., n data points from a time series generated from a "true" model $\mathcal{M}$. Under a specified model M, the best or minimum mean squared error (MMSE) h-steps ahead forecast, i.e., forecast of Y_{n+h} from n data points under model M is given by

$$\hat{Y}^{(M)}_{n+h|n} = E_M(Y_{n+h}|Y_1, \ldots, Y_n), \tag{1}$$

where $E_M(\cdot)$ denotes expectation under model M, assuming the true model parameters are known. The forecast error is $e^{(M)}_{n+h|n} = \hat{Y}^{(M)}_{n+h|n} - Y_{n+h}$, and the MSE of this forecast is

$$MSE^{(M)}_{h|n} = E_{\mathcal{M}}\left[e^{(M)}_{n+h|n}\right]^2, \tag{2}$$

where the expectation is taken over the true model $\mathcal{M}$.

Suppose M represents a stationary ARMA(p,q) model with zero mean,

$$Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q},$$

where $\{\epsilon_t\}$ are assumed to be white noise (mutually independent) with zero mean and common variance σ^2_ϵ.

Let B denote the backward shift operator B such that $B^h Y_t = Y_{t-h}$. Then the above model can be written in the polynomial form

$$\phi(B)Y_t = \theta(B)\epsilon_t,$$

where $\phi(B) = (1 - \phi_1 B - \ldots - \phi_p B^p)$ and $\theta(B) = (1 + \theta_1 B + \ldots + \theta_q B^q)$ respectively represent the pth order AR and qth order MA polynomials. We assume that both $\phi(z) = 0$ and $\theta(z) = 0$ have no roots on or inside the unit circle, so that $\{Y_t\}$ is causal and strictly invertible, and has an infinite moving average representation:

$$Y_t = \psi_0 \epsilon_t + \psi_1 \epsilon_{t-1} + \ldots, \tag{3}$$

where for $j = 0, 1, 2, \ldots$, ψ_j is the coefficient of B^j in the infinite expansion $\phi^{-1}(B)\theta(B) = \sum_{j=0} \psi_j B^j$.

We have started with the general MMSE forecasting problem (1). From now on, we will focus on the Gaussian ARMA model, because the motivating traffic models and experiments we carry out will all be Gaussian. Under normality the MMSE predictor and the best linear predictor (BLP) coincide. Therefore, even without the normality assumption, the following discussion and results will remain the same if we consider the BLP instead of the MMSE predictor, which has been the convention of the forecasting based on ARMA models.

For a causal and invertible ARMA model, the temporal dependence decays geometrically fast, so the MMSE predictor based on $\{Y_1, \ldots, Y_n\}$ and on the infinite past $\{\ldots, Y_0, \ldots, Y_n\}$ are very close as long as n is reasonably large. As a result, we redefine the notation $\hat{Y}^{(M)}_{n+h|n}$ as

$$\hat{Y}^{(M)}_{n+h|n} = E_M(Y_{n+h}|\ldots, Y_0, \ldots, Y_n). \tag{4}$$

Using representation (3), one can write Y_{n+h} as:

$$Y_{n+h} = \psi_0\epsilon_{n+h} + \psi_1\epsilon_{n+h-1} + \ldots + \psi_{h-1}\epsilon_{n+1} + \psi_h\epsilon_n + \psi_{h+1}\epsilon_{n-1} + \ldots, \tag{5}$$

Then the MMSE predictor of Y_{n+h} is:

$$\hat{Y}^{(M)}_{n+h|n} = E_M(Y_{n+h}|Y_1, \ldots, Y_n) = E_M(Y_{n+h}|\epsilon_1, \ldots, \epsilon_n) = \psi_h\epsilon_n + \psi_{h+1}\epsilon_{n-1} + \ldots.$$

Now, let $\hat{Y}^{(\hat{M})}_{n+h|n}$ denote the minimum mean squared error (MMSE) predictor of Y_{n+h} when the parameters of model M are unknown and are estimated from observations $Y_1, \ldots, Y_n$. Then, this MMSE predictor is given by

$$\hat{Y}^{(\hat{M})}_{n+h|n} = \hat{\psi}_h\epsilon_n + \hat{\psi}_{h+1}\epsilon_{n-1} + \ldots,$$

where $\hat{\psi}_j$'s are estimated from the data $Y_1, \ldots, Y_n$. The corresponding forecast error and MSE are respectively given by

$$e_{n+h|n}^{(\hat{M})} = \hat{Y}_{n+h}^{(\hat{M})} - Y_{n+h}, \quad MSE_h^{(\hat{M})}|n = E_{\mathcal{M}}\left[e_{n+h|n}^{(\hat{M})}\right]^2, \tag{6}$$

2.1 Forecast errors from true model with known and unknown parameters

If the assumed model M is the true model $\mathcal{M}$ with known parameters, then the representation (5) is the correct expansion of Y_{n+h}, and consequently the forecast error

$$e_{n+h|n}^{(\mathcal{M})} = \hat{Y}_{n+h|n}^{(\mathcal{M})} - Y_{n+h} = \psi_0\epsilon_{n+h} + \psi_1\epsilon_{n+h-1} + \ldots + \psi_{h-1}\epsilon_{n+1}, \tag{7}$$

yields the MSE

$$MSE_{h|n}^{(\mathcal{M})} = \left(\psi_0^2 + \ldots + \psi_{h-1}^2\right)\sigma_\epsilon^2, \tag{8}$$

where, as assumed earlier, σ_ϵ^2 is the common variance of the residuals ϵ. This quantity $MSE_{h|n}^{(\mathcal{M})}$ can be referred to as the "inherent model error of forecast". It is a measure of forecast error in the *best possible scenario* where the true model specification (including the parameters) is known, and thus can be interpreted as the *unavoidable error*.

Now consider the situation where the parameters of the true model are unknown. Then the forecast error is given by $e_{n+h|n}^{(\widehat{\mathcal{M}})} = \hat{Y}_{n+h|n}^{(\widehat{\mathcal{M}})} - Y_{n+h}$. Thus the MSE can be decomposed as

$$\begin{aligned}
MSE_{h|n}^{(\widehat{\mathcal{M}})} &= E_{\mathcal{M}}\left[e_{n+h|n}^{(\widehat{\mathcal{M}})}\right]^2 = E\left[\hat{Y}_{n+h|n}^{(\widehat{\mathcal{M}})} - Y_{n+h}\right]^2 \\
&= E_{\mathcal{M}}\left[\hat{Y}_{n+h|n}^{(\widehat{\mathcal{M}})} - \hat{Y}_{n+h|n}^{(\mathcal{M})} + \hat{Y}_{n+h|n}^{(\mathcal{M})} - Y_{n+h}\right]^2 \\
&= E_{\mathcal{M}}\left[\hat{Y}_{n+h|n}^{(\widehat{\mathcal{M}})} - \hat{Y}_{n+h|n}^{(\mathcal{M})} + \epsilon_{n+h}^{(\mathcal{M})}\right]^2 \\
&= E_{\mathcal{M}}\left[\hat{Y}_{n+h|n}^{(\widehat{\mathcal{M}})} - \hat{Y}_{n+h|n}^{(\mathcal{M})}\right]^2 + E_{\mathcal{M}}\left[\epsilon_{n+h}^{(\mathcal{M})}\right]^2.
\end{aligned}$$

The last step follows from the following facts: (i) The difference $\hat{Y}_{n+h|n}^{(\widehat{\mathcal{M}})} - \hat{Y}_{n+h|n}^{(\mathcal{M})}$ (i.e, the difference between the predictor with the true model parameters and that with estimated model parameters) depends on the past residuals $\epsilon_n, \epsilon_{n-1}, \ldots, \ldots$ whereas, (ii) from (7) it is clear that $e^{(\mathcal{M})}_{n+h|n}$ depends only on the future residuals $\epsilon_{n+1}, \ldots, \epsilon_{n+h}$. Consequently, by mutual independence of residuals, $\hat{Y}^{(\widehat{\mathcal{M}})}_{n+h|n} - \hat{Y}^{(\mathcal{M})}_{n+h|n}$ is independent of $\epsilon^{(\mathcal{M})}_{n+h|n}$ and the expectation of the product term vanishes. We thus have the following well-known (e.g. Mazzeu et al., 2018) result:

Proposition 1. The MSE of forecast under the true model $\mathcal{M}$ with unknown coefficients, where the model parameters are estimated from observed data, can be decomposed as

$$MSE^{(\widehat{\mathcal{M}})}_{h|n} = \delta^{(\widehat{\mathcal{M}},\mathcal{M})}_{h|n} + MSE^{(\mathcal{M})}_{h|n}, \tag{9}$$

where $MSE^{(\mathcal{M})}_{h|n}$ is the inherent model error given by (8), and

$$\delta^{(\widehat{\mathcal{M}},\mathcal{M})}_{h|n} = E_{\mathcal{M}}\left[\hat{Y}^{(\widehat{\mathcal{M}})}_{n+h|n} - \hat{Y}^{(\mathcal{M})}_{n+h|n}\right]^2 \tag{10}$$

represents the contribution of "parameter uncertainty" to the overall forecast error.

Note that the quantity $\delta^{(\widehat{\mathcal{M}},\mathcal{M})}_{h|n}$ defined in Proposition 1 depends on the sample size n and the forecast horizon h, and loosely speaking, is expected to converge to zero for fixed h as n goes to infinity, under fairly mild conditions related to the convergence of the parameter estimators to the true model parameters (see Fuller (1996)). Proposition 1 also suggests that if it is possible to estimate $MSE^{(\widehat{\mathcal{M}})}_{h|n}$ and $MSE^{(\mathcal{M})}_{h|n}$, then their difference will provide an estimate of $\delta^{(\widehat{\mathcal{M}},\mathcal{M})}_{h|n}$.

3 Forecast errors from misspecified models

Now assume that M is a misspecified model with parameter θ_M (typically a vector), which is different from the true model $\mathcal{M}$ with parameter $\theta_{\mathcal{M}}$. To forecast Y_{n+h} from n data points using M, one has to estimate the parameters θ_M from observations $Y_1, \ldots, Y_n$ and follow the procedure described earlier to obtain the predictor $\hat{Y}^{(\hat{M})}_{n+h}$. The associated error will be denoted by $e^{(\hat{M})}_{n+h}$ and the MSE by $MSE^{(\hat{M})}_{h|n}$.

To decompose $MSE_{h|n}^{(\hat{M})}$, we visualize a *population version of the misspecified model* with some "true value" of θ_M, and define the MMSE predictor $\hat{Y}_{n+h}^{(M)}$ assuming that true value is known. While such a "true value" from a misspecified model does not make much sense, we can consider it to be the "best value" or "optimal value" (in some sense) that generates observations from the time series from the misspecified model similar to those generated by the true model. We will provide a more precise definition of such a best value in the context of ARMA models later. Now $MSE_{h|n}^{(\hat{M})}$ can be decomposed as follows:

$$\begin{aligned} MSE_{h|n}^{(\hat{M})} &= E_{\mathcal{M}}\left[\hat{Y}_{n+h|n}^{(\hat{M})} - Y_{n+h}\right]^2 \\ &= E_{\mathcal{M}}\left[\hat{Y}_{n+h|n}^{(\hat{M})} - \hat{Y}_{n+h|n}^{(M)} + \hat{Y}_{n+h|n}^{(M)} - Y_{n+h}\right]^2 \\ &= E_{\mathcal{M}}\left[\hat{Y}_{n+h|n}^{(\hat{M})} - \hat{Y}_{n+h|n}^{(M)} + e_{n+h|n}^{(M)}\right]^2 \\ &= E_{\mathcal{M}}\left[\hat{Y}_{n+h|n}^{(\hat{M})} - \hat{Y}_{n+h|n}^{(M)}\right]^2 + 2E_{\mathcal{M}}\left[\left(\hat{Y}_{n+h|n}^{(\hat{M})} - \hat{Y}_{n+h|n}^{(M)}\right) e_{n+h|n}^{(M)}\right] + E_{\mathcal{M}}\left[e_{n+h|n}^{(M)}\right]^2 \\ &= \delta_{n|h}^{(\hat{M},M)} + 2E_{\mathcal{M}}\left[\left(\hat{Y}_{n+h|n}^{(\hat{M})} - \hat{Y}_{n+h|n}^{(M)}\right) e_{n+h|n}^{(M)}\right] + MSE_{h|n}^{(M)}, \end{aligned} \tag{11}$$

which takes a form similar to (9) except for the fact that the product term does not vanish in this case. However, for a large sample size, both $\delta_{n|h}^{(\hat{M},M)}$ and the product term should be small. The last term can again be decomposed as

$$\begin{aligned} MSE_{h|n}^{(M)} &= E_{\mathcal{M}}\left[\hat{Y}_{n+h}^{(\hat{M})} - Y_{n+h}\right]^2 \\ &= E_{\mathcal{M}}\left[\hat{Y}_{n+h}^{(M)} - \hat{Y}_{n+h}^{(\mathcal{M})} + \hat{Y}_{n+h}^{(\mathcal{M})} - Y_{n+h}\right]^2 \\ &= \delta_{h|n}^{(M,\mathcal{M})} + MSE_{h|n}^{(\mathcal{M})}, \end{aligned} \tag{12}$$

where

$$\delta_{h|n}^{(M_1,M_2)} = E_{\mathcal{M}}\left[\hat{Y}_{n+h}^{(M_1)} - \hat{Y}_{n+h}^{(M_2)}\right]^2, \tag{13}$$

is a measure of the impact of *model uncertainty* on the forecast made from two models M_1 and M_2.

Remark 1. We note here that $\delta_{h|n}^{(M,\mathcal{M})} \geq 0$ for all h and n, with equality holding if and only if M is the true model $\mathcal{M}$. Consequently,

$$MSE_{h|n}^{(M)} \geq MSE_{h|n}^{(\mathcal{M})}, \tag{14}$$

with equality holding if and only if M is the true model $\mathcal{M}$.

Remark 2. Note that $\delta_{h|n}^{(M_1,M_2)}$ can be called the *Expected squared discrepancy* between forecasts of Y_{n+h} made from two models M_1 and M_2, that may or may not be completely specified (in terms of parameter values). Thus, $\delta_{h|n}^{(M,\mathcal{M})}$ in the RHS of (12), $\delta_{h|n}^{(\hat{M},M)}$ in the RHS of (11) and $\delta^{(\widehat{\mathcal{M}},\mathcal{M})}$ in (10) are all special cases of this discrepancy.

Davies and Newbold (1980) derived a closed form expression for $\delta_{h|n}^{(M,\mathcal{M})}$ when the true model $\mathcal{M}$ and the misspecified model M are respectively ARMA($p_{\mathcal{M}}, q_{\mathcal{M}}$) and ARMA($p_M, q_M$). We state the main result of Davies and Newbold (1980) below:

Proposition 2 (Davies and Newbold (1980))**.** Let the true model $\mathcal{M}$ be an ARMA($p_{\mathcal{M}}, q_{\mathcal{M}}$) model represented by $\phi_{\mathcal{M}}(B)Y_t = \theta_{\mathcal{M}}(B)\epsilon_t$ where $\phi_{\mathcal{M}}(B)$ and $\theta_{\mathcal{M}}(B)$ are polynomials of order $p_{\mathcal{M}}$ and $q_{\mathcal{M}}$ respectively and let the misspecified model M be an ARMA(p_M, q_M) model represented by $\phi_M(B)Y_t = \theta_M(B)\epsilon_t$ where $\phi_M(B)$ and $\theta_M(B)$ are polynomials of order p_M and q_M respectively. Also, let $\phi_{\mathcal{M}}(B)\theta_{\mathcal{M}}^{-1}(B)\epsilon_t = \sum_{j=0}^{\infty} \psi_{\mathcal{M},j}\epsilon_{t-j}$ and $\phi_M(B)\theta_M^{-1}(B)\epsilon_t = \sum_{j=0}^{\infty} \psi_{M,j}\epsilon_{t-j}$ be the infinite moving average representations of $\mathcal{M}$ and M respectively. Finally, let $\phi_M(B)\theta_M^{-1}(B)\phi_{\mathcal{M}}^{-1}(B)\theta_{\mathcal{M}}(B) = \sum_{j=0}^{\infty} \widetilde{\psi}_j B^j$. Then the contribution of the model uncertainty $\delta_{h|n}^{(M,\mathcal{M})}$ associated with the h-steps ahead forecast made from models M and $\mathcal{M}$ is given by

$$\delta_{h|n}^{(M,\mathcal{M})} = \sum_{j=0}^{\infty} \{\psi_{\mathcal{M},j+h} - a_j(h)\}^2, \tag{15}$$

where

$$a_j(h) = \sum_{k=0}^{j} \psi_{M,h+k}\widetilde{\psi}_{j-k}.$$

Proposition 2 assumes that the misspecified model is completely specified and the coefficients $\psi_{M,j}$'s are known. This is obviously not the case in practice. In the process of forecasting with a misspecified model, one would estimate the parameters of such a model from the data generated by the true model. We need to establish a connection of such an estimate with the specified value of the parameter in model M. We will revisit this topic in Section 3.1.

Finally, substituting (13) in (11) we arrive at the following proposition:

Proposition 3 (Decomposition of forecast MSE)**.** The MSE of prediction under the misspecified model M with unknown coefficients, where the model parameters are estimated from observed data, can be decomposed as

$$MSE_{h|n}^{(\hat{M})} = \delta_{h|n}^{(\hat{M},M)} + 2E_{\mathcal{M}}\left[\left(\hat{Y}_{n+h}^{(\hat{M})} - \hat{Y}_{n+h}^{(M)}\right)\epsilon_{n+h}^{(M)}\right] + \delta_{h|n}^{(M,\mathcal{M})} + MSE_{h|n}^{(\mathcal{M})}, \tag{16}$$

where $MSE_{h|n}^{(\mathcal{M})}$ is the inherent model error given by (8), $\delta_{h|n}^{(M,\mathcal{M})}$ is given by (13) and $\delta_{h|n}^{(\hat{M},M)}$ is given by (10).

Now, for any model M, we can write $\hat{Y}_{n+h}^{(\widehat{M})} = g_n(\widehat{\theta}_{M,n})$ and $\hat{Y}_{n+h}^{(M)} = g_n(\theta_M)$, where θ_M denotes the model parameter, $\widehat{\theta}_{M,n}$ its estimator based on observations $Y_1, \ldots, Y_n$, and $g_n(\cdot)$ is a continuous function which is finite for all n. If $\hat{\theta}_n$ is a consistent estimator of θ, as is the case for maximum likelihood estimators of ARMA model parameters, assuming normality of innovations. Then by the continuous mapping theorem, $g_n(\widehat{\theta}_{M,n}) - g_n(\theta_M)$ converges to zero in probability as $n \to \infty$. Consequently, the first two terms on the RHS of (16) $\delta_{h|n}^{(\hat{M},M)}$ and $2E_{\mathcal{M}}\left[\left(\hat{Y}_{n+h}^{(\hat{M})} - \hat{Y}_{n+h}^{(M)}\right)\right]$ both converge to zero. Thus we arrive at the following corollary.

Corollary 1. For large n, the MSE of prediction under the misspecified model M with unknown coefficients, where the model parameters are consistently estimated from observed data, can be approximated as

$$MSE_{h|n}^{(\widehat{M})} \approx MSE_{h|n}^{(M)} = \delta_{h|n}^{(\widehat{M},M)} + MSE_{h|n}^{(\mathcal{M})}.$$

3.1 "Optimal" parameters of a misspecified model

The foregoing discussion suggests the need to visualize and define "true" parameters of the misspecified model in terms of their estimated version. Davies and Newbold (1980) provided such a connection by assuming that the misspecified model ARMA(p_M, q_M) has an AR representation of order p_M, and estimated the parameters using least square estimation. The true parameters were then taken as the probability limits of the least square estimator. However, whereas this definition seems reasonable if the misspecified model is AR(p_M), it does not seem to incorporate additional model uncertainty if the misspecified model if ARMA(p_M, q_M) with $q_M \geq 1$, and consequently does not provide a way to define "true" parameters of the true misspecified model.

We now provide a more formal definition of an "optimal value of the parameter of a misspecified model" in the context of forecasting from ARMA models.

Definition 1 (Optimal value of misspecified model parameter)**.** Let M be any arbitrary ARMA model with parameter θ_M. The optimal value of θ_M is defined as the value θ_M^* that minimizes the MMSE of prediction of Y_{n+1} based on observations $\{\ldots, Y_{-1}, Y_0, Y_1, \ldots, Y_n\}$. Formally,

$$\theta_M^* = \arg\min_{\theta_M} E_{\mathcal{M}} \left[\hat{Y}_{n+1}^{\theta_M} - Y_{n+1}\right]^2,$$

where $\hat{Y}_{n+1}^{\theta_M}$ denotes the predictor of Y_{n+1} based on model M and parameter value θ_M.

Remark 3. If we require in addition that the ARMA model M is Gaussian, then this best misspecified model is equivalent to the one that minimizes the Kullback-Leibler divergence from θ_M to $\mathcal{M}$:

$$\theta_M^* = \arg\min_{\theta_M} \mathrm{KL}(\mathcal{M}||\theta_M),$$

where KL($\cdot||\cdot$) denotes the Kullback-Leibler divergence. We explain this equivalence in the Appendix.

Remark 4. Definition 1 is equivalent to the notion of the "true" misspecified model by Davies and Newbold (1980) if the misspecified model M is purely autoregressive. In fact, if the misspecified model is an autoregressive process, then according to Definition 1, θ_M^* is the Yule-Walker estimator based on the autocovariances of $\{Y_t\}$. Davies and Newbold (1980) suggest using the limit of the least squares estimators, which is exactly the Yule-Walker estimator. Therefore, the two definitions are equivalent when the misspecified model is autoregressive.

Remark 5. Definition 1 also provides us with a specific algorithm to obtain the value of θ_M using available data $Y_1, \ldots, Y_n$. Adopting the notations of Proposition 2, suppose the true model is a causal and invertible ARMA$(p_{\mathcal{M}}, q_{\mathcal{M}})$: $\phi_{\mathcal{M}}(B)Y_t = \theta_{\mathcal{M}}(B)\epsilon_t$, and the misspecified model M is a causal and invertible ARMA(p_M, q_M): $\phi_M(B)Y_t = \theta_M(B)\epsilon_t$. Here without loss of generality we assume there is no intercept in both models, and $\mathrm{Var}(\varepsilon_t) = 1$. We describe how to find the "optimal" parameters under the misspecified model. Instead of using AR and MA coefficients, the misspecified model can be equivalently parametrized by the factorizations $\phi_M(z) = (1-w_1 z)\cdots(1-w_{p_M} z)$ and $\theta^*(z) = (1-v_1 z)\cdots(1-v_{q_M} z)$. The one-step ahead prediction error (using the infinite past) under the model M is given by

$$\frac{\phi_M(B)}{\theta_M(B)} Y_t = \frac{\phi_M(B)}{\theta_M(B)} \times \frac{\theta_{\mathcal{M}}(B)}{\phi_{\mathcal{M}}(B)} \epsilon_t.$$

Let

$$\frac{\phi_M(z)}{\theta_M(z)} \times \frac{\theta_{\mathcal{M}}(z)}{\phi_{\mathcal{M}}(z)} = \sum_{k=0}^{\infty} \widetilde{\psi}_k z^k.$$

Note that each $\tilde{\psi}$ in the preceding equation depends on $\{w_1, \ldots, w_{p_M}, v_1, \ldots, v_{q_M}\}$ implicitly. According to Definition 1, the optimal values of the parameters $\{w_1^*, \ldots, w_{p_M}^*, v_1^*, \ldots, v_{q_M}^*\}$ are given by

$$\arg\min_{\{w_i, v_j\}} \sum_{k=0}^{n} \tilde{\psi}_k^2. \tag{17}$$

The optimization problem (17) is related to the maximum likelihood estimation (MLE) of the ARMA model. It will be convenient in practice to find θ_M^* through simulation: (i) simulate a long series (e.g. of length 100,000) from the true model, and (ii) find the MLE under the misspecified model. This MLE serves as an estimate of θ_M^*, whose accuracy can be controlled by the length of the simulated series.

Example 1. Suppose $\mathcal{M}$ is AR(1): $Y_t = \phi Y_{t-1} + \epsilon_t$, and let M: MA(1) be the misspecified model: $Y_t = \epsilon_t + \theta\epsilon_{t-1}$. Assume $|\phi| < 1$ so that the true model is invertible. It holds that

$$\sum_{k=0}^{n} \tilde{\psi}_k z^k := \frac{1}{(1-\phi z)(1+\theta z)} = \frac{1}{\phi+\theta}\left[\frac{\phi}{1-\phi z} + \frac{\theta}{1+\theta z}\right]$$

and

$$\sum_{k=1}^{n} \tilde{\psi}_k^2 = \frac{1}{(\phi+\theta)^2}\sum_{k=0}^{\infty}\left[\phi^{k+1} + (-1)^k\theta^{k+1}\right]^2.$$

Therefore, the optimal θ^* is the θ that minimizes the preceding infinite sum. We give the optimal θ^* corresponding to $\phi \in \{.1, .2, \ldots, .9\}$ in Table 1. The estimated $\hat{\theta}^*$ obtained from a simulated series of length 100,000 is also reported in the third row of the table.

ϕ	.1	.2	.3	.4	.5	.6	.7	.8	.9
θ^*	.099	.193	.279	.356	.428	.496	.565	.640	.735
$\hat{\theta}^*$	.098	.192	.280	.360	.429	.497	.562	.642	.735

Table 1: Optimal θ^* for the MA(1), when the true model is AR(1).

3.2 Inflation of prediction error by model misspecification and paramater uncertainty and its decomposition

Based on the definitions and results of the previous section, we now define measures of inflation of forecast MSE due to model misspecification and parameter uncertainty. First, note that in the problem of obtaining an h-step ahead forecast Y_{n+h} from observations $Y_1, \ldots, Y_n$ generated by a true ARMA model $\mathcal{M}$, the unavoidable or *intrinsic* uncertainty is $MSE_{h|n}^{(\mathcal{M})}$ given by (8). When an arbitrary model M is used to make the forecast, model parameters are estimated from the data and subsequently plugged into the forecast. This process inflates the forecast error by $MSE_{h|n}^{(\hat{M})} - MSE_{h|n}^{(\mathcal{M})}$, where $MSE_{h|n}^{(\hat{M})}$ is given by Proposition 3. We now formally define this measure of inflation and its components, relative to the intrinsic uncertainty $MSE_{h|n}^{\mathcal{M}}$.

Definition 2 (Total Percentage Inflation or TPI)**.** In the problem of obtaining an h-step ahead forecast Y_{n+h} from observations $Y_1, \ldots, Y_n$ generated by a true ARMA model $\mathcal{M}$, assuming model

M and estimating its parameters, the *total percentage inflation* (TPI) is given by

$$\begin{aligned}
\mathrm{TPI}_{h|n}^{(M)} &= \frac{MSE_{h|n}^{(\hat{M})} - MSE_{h|n}^{(\mathcal{M})}}{MSE_{h|n}^{(\mathcal{M})}} \times 100 \\
&= \begin{cases} \dfrac{\delta_{h|n}^{(\hat{M},M)}}{MSE^{(\mathcal{M})}} \times 100, & \text{if } M \equiv \mathcal{M}, \\ \\ \dfrac{\delta_{h|n}^{(\hat{M},M)} + 2E_{\mathcal{M}}\left[\left(\hat{Y}_{n+h}^{(\hat{M})} - \hat{Y}_{n+h}^{(M)}\right)\epsilon_{n+h}^{(M)}\right] + \delta_{h|n}^{(M,\mathcal{M})}}{MSE^{(\mathcal{M})}} \times 100, & \text{otherwise} \end{cases}
\end{aligned}$$

Definition 3 (Percentage Estimation Inflation or PEI)**.** In the problem of obtaining an h-step ahead forecast Y_{n+h} from observations $Y_1, \ldots, Y_n$ generated by a true ARMA model $\mathcal{M}$, assuming model M and estimating its parameters, the percentage *estimation inflation* (PEI) is given by

$$\begin{aligned}
\mathrm{PEI}_{h|n}^{(M)} &= \frac{MSE_{h|n}^{(\hat{M})} - MSE_{h|n}^{(M)}}{MSE_{h|n}^{(\mathcal{M})}} \times 100 \qquad (18) \\
&= \begin{cases} \dfrac{\delta_{h|n}^{(\hat{M},M)}}{MSE_{h|n}^{(\mathcal{M})}} \times 100, & \text{if } M \equiv \mathcal{M}, \\ \\ \dfrac{\delta_{h|n}^{(\hat{M},M)} + 2E_{\mathcal{M}}\left[\left(\hat{Y}_{n+h}^{(\hat{M})} - \hat{Y}_{n+h}^{(M)}\right)\epsilon_{n+h}^{(M)}\right]}{MSE_{h|n}^{(\mathcal{M})}} \times 100, & \text{otherwise} \end{cases}
\end{aligned}$$

where M represents the completely specified model with some underlying "best" parameter value s per Definition 1.

Definition 4 (Percentage Misspecification Inflation or PMI)**.** In the problem of obtaining an h-step ahead forecast Y_{n+h} from observations $Y_1, \ldots, Y_n$ generated by a true ARMA model $\mathcal{M}$, assuming model M and estimating its parameters, the percentage *misspecification inflation* (PMI) is given by

$$\mathrm{PMI}^{M}_{h|n} = \frac{MSE^{(M)}_{h|n} - MSE^{(\mathcal{M})}_{h|n}}{MSE^{(\mathcal{M})}_{h|n}} \times 100 \tag{19}$$
$$= \begin{cases} 0, & M \equiv \mathcal{M} \\ \\ \frac{\delta^{(M,\mathcal{M})}_{h|n}}{MSE^{(\mathcal{M})}_{h|n}} \times 100, & \text{otherwise} \end{cases}$$

where M represents the completely specified model with some underlying "best" parameter value as per Definition 1.

3.3 Extension to Seasonal ARMA (SARMA) models, ARIMA and SARIMA models

SARMA models allow users to incorporate seasonal effects into the ARMA framework. A typical multiplicative SARMA $(p, q) \times (P, Q)_s$ model (Brockwell and Davis, 2002) is of the form:

$$\phi_p(B)\Phi_P(B^s)Y_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t, \tag{20}$$

where as before, B^h denotes the h-step backshift operator for $h \geq 1$, $\phi_p(\cdot)$, $\Phi_P(\cdots)$, $\theta_q(\cdot)$ and $\Theta_Q(\cdot)$ are polynomials of orders p, P, q and Q respectively, and $\{\epsilon_t\}$ is a white noise term. Such models can include more than two seasonal polynomials and consequently allow for multiple sources of seasonality (e.g., daily, weekly etc.) to be incorporated into the model. Because model (20) can easily be expressed in the form (3), all the concepts, definitions and results discussed so far this section can be extended in a straightforward manner to SARMA models.

Similar decomposition of the mean squared prediction error can be given for the ARIMA process. For simplicity, we only consider the ARIMA models with integration order 1. Suppose the true model of $\{X_t\}$ is an ARIMA$(p_{\mathcal{M}}, 1, q_{\mathcal{M}})$

$$\phi(B)\Delta X_t = \theta(B)\epsilon_t.$$

Assume the (possibly) misspecified model M is ARIMA$(p_M, 1, q_M)$, *i.e.* it is still an ARIMA with integrated order 1, but with (possibly) misspecified p_M and q_M. Let $Y_t = \Delta X_t$, then $\{Y_t\} \sim$ ARMA$(p_{\mathcal{M}}, 0, q_{\mathcal{M}})$, and the decomposition discussed earlier holds for $\{Y_t\}$. We now describe how these results can be adapted for the process $\{X_t\}$. First of all, note that

$$\hat{X}_{h|n}^{(M)} = X_n + \sum_{k=1}^{h} \hat{Y}_{k|n}^{(M)}, \quad \text{and} \quad e_{h|n}^{(M)}(X) = X_{n+h} - \hat{X}_{h|n}^{(M)} = \sum_{k=1}^{h} e_{k|n}^{(M)}(Y).$$

for any model M, whether it is true or misspecified. Here we use (X) and (Y) to specify the prediction errors for $\{X_t\}$ or $\{Y_t\}$ respectively. Proposition 1 leads to

$$MSE_{h|n}^{(\widehat{\mathcal{M}})}(X) = \delta_{h|n}^{(\widehat{\mathcal{M}},\mathcal{M})}(X) + MSE_{h|n}^{(\mathcal{M})}(X),$$

where the inherent $MSE_{h|n}^{(\mathcal{M})}(X)$ is (comparing (8)):

$$MSE_{h|n}^{(\mathcal{M})}(X) = \sigma_\epsilon^2 \cdot \sum_{k=1}^{h} \left(\psi_0 + \ldots + \psi_{k-1}\right)^2.$$

Note that the inherent prediction MSE of $\{X_t\}$ goes to infinity as the forecast horizon h increases. Proposition 3 translates into

$$MSE_{h|n}^{(\hat{M})}(X) = \delta_{h|n}^{(\hat{M},M)}(X) + 2E_{\mathcal{M}}\left[\left(\hat{X}_{n+h}^{(\hat{M})} - \hat{X}_{n+h}^{(M)}\right) e_{n+h}^{(M)}(X)\right] + \delta_{h|n}^{(M,\mathcal{M})}(X) + MSE_{h|n}^{(\mathcal{M})}(X).$$

The Inflation measures TPI, PEI and PMI can then be defined similarly for the ARIMA process $\{X_t\}$. However, it is important to note that the above procedure can be generalized to any integration order only when it is the same for the ARIMA processes under comparison. Thus, our framework will allow us to compare an ARIMA(p_1,d_1, q_1) process with an ARIMA(p_2,d_2, q_2)

process if and only if $d_1 = d_2$.

4 Simulation algorithm and examples

We now present a simulation process that will help researchers estimate the different components of forecast inflation in a setting where model M is mistakenly used instead of a true model $\mathcal{M}$. It will also help compare different models by assessing their roles in inflating the inherent forecast error.

The input to the simulation code is (i) an ARIMA (p, d, q) model with specified values of known parameter $\theta_{\mathcal{M}}$ (assumed to be the true model $\mathcal{M}$), (ii) the order (p_M, d_M, q_M) of an ARIMA model which is the misspecified model M. A very long time series consisting of $N_{\max}$ observations from the known true model $\mathcal{M}$ is generated, and parameters θ^*_M of the misspecified model M are estimated from these data. Thus an ARIMA(p_M, d_M, q_M) model with parameters θ^*_M is assumed to be the optimal misspecified model M. The number of observations to estimate the parameters depends on the complexity of the true model $\mathcal{M}$, as a larger number of observations will be necessary for more complex processes.

Next, a sample size n and a forecast horizon h is fixed. For each iteration $i = 1, \dots, I$,

1. A total of $n + h$ observations are generated from the true model $\mathcal{M}$. Let $Y_{t,i}$ denote the tth observation, $t = 1, \dots, n + h$, $i = 1, \dots, I$.

2. The true model $\mathcal{M}$ and misspecified model M are now fitted to the first n observations, yielding estimated parameter $\hat{\theta}_{\mathcal{M},i}$ and $\hat{\theta}^*_{M,i}$ respectively.

3. Four forecasts: $\hat{Y}^{(\hat{M})}_{n+h|n,i}$, $\hat{Y}^{(M)}_{n+h|n,i}$, $\hat{Y}^{(\widehat{\mathcal{M}})}_{n+h|n,i}$ and $\hat{Y}^{(\mathcal{M})}_{n+h|n,i}$ of $Y_{n+h,i}$ are now obtained from the respective models using estimated parameter $\hat{\theta}^*$, optimal parameter θ^* of misspecified model M, estimated parameter $\hat{\theta}$ of the true model and known parameter θ of true model $\mathcal{M}$. The forecasts of any true model are estimated with the equations obtained using the backshift notation.

4. The corresponding forecast errors are estimated as:

$$\begin{aligned} e^{(\hat{M})}_{n+h|n,i} &= \hat{Y}^{(\hat{M})}_{n+h|n,i} - Y_{n+h,i}, \quad e^{(M)}_{n+h|n,i} = \hat{Y}^{(M)}_{n+h|n,i} - Y_{n+h,i} \\ e^{(\widehat{\mathcal{M}})}_{n+h|n,i} &= \hat{Y}^{(\widehat{\mathcal{M}})}_{n+h|n,i} - Y_{n+h,i}, \quad e^{(\mathcal{M})}_{n+h|n,i} = \hat{Y}^{(\mathcal{M})}_{n+h|n,i} - Y_{n+h,i} \end{aligned}$$

5. The mean squared errors of forecast $MSE^{(\hat{M})}_{n+h|n}$, $MSE^{(M)}_{n+h|n}$, $MSE^{(\widehat{\mathcal{M}})}_{n+h|n}$ and $MSE^{\mathcal{M}}_{n+h|n}$ are estimated by squaring and averaging the forecast errors $e^{\hat{M}}_{n+h|n,i}$, $e^{M}_{n+h|n,i}$, $e^{(\widehat{\mathcal{M}})}_{n+h|n,i}$ and $e^{\mathcal{M}}_{n+h|n,i}$ over $i = 1, \ldots, I$ respectively.

6. Estimation of inflation ratios:

   (i) $\mathrm{PEI}^{(\mathcal{M})}_{h|n}$ for the true model $\mathcal{M}$ is estimated by substituting the estimated mean squared errors in (18) for $M \equiv \mathcal{M}$. Recall that $\mathrm{PMI}^{(\mathcal{M})}_{h|n}$ for the true model is zero.

   (ii) $\mathrm{PEI}^{(M)}_{h|n}$ for the misspecified model M is estimated by substituting the estimated mean squared errors in (18).

   (iii) $\mathrm{PMI}^{(M)}_{h|n}$ for the misspecified model M is estimated by substituting the estimated mean squared errors in (19).

We now present an example of this simulation algorithm and presentation of the results.

Example: Misspecifying ARMA(1,1) as AR(1) or MA(1):

Assume that the true data generating process is ARMA(1,1) with parameters $\phi = 0.8$ and $\theta = -0.3$, and consider forecasting from such a process using two incorrect models: an AR(1) and an MA(1) process. The simulation was conducted using the process described above with $N_{\max} = 10^6$ data points used to estimate optimal parameters of the misspecified models. Various sample sizes n ranging from 50 to 500 in steps of 50 and forecast horizons $h = 1, 2, 3$ are considered for the simulations. For each (n, h) combination, the mean squared errors and inflation indices PEI and PMI were estimated from $I = 5000$ data sets.

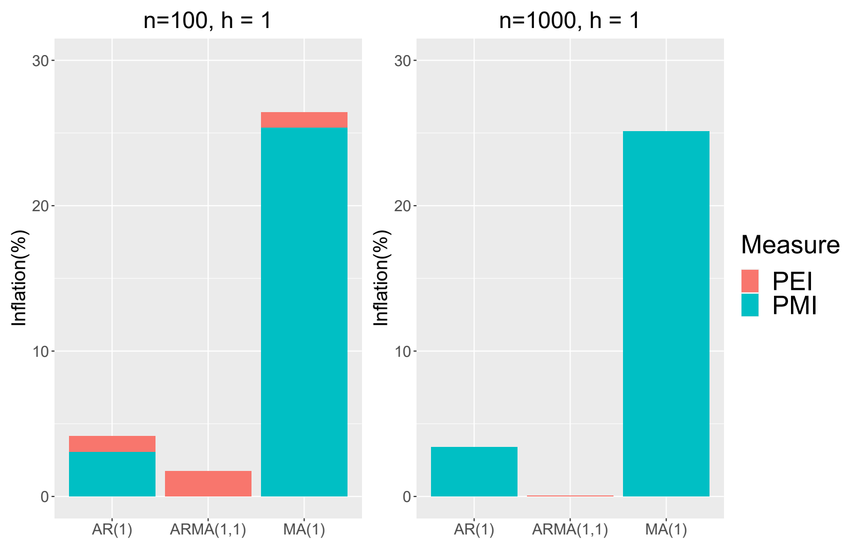

Figure 1: Decomposition of percentage inflation in forecast errors for a true ARMA(1,1) process with parameters $\phi = 0.8$, $\theta = -0.3$, and $\sigma^2 = 100$ under misspecified models AR(1) and MA(1). One-step ahead forecasting with $n = 100$ (left) and $n = 1000$ (right) is considered.

Figure 1 shows the break up of the inflation of forecast error associated with two sample sizes of $n = 100$ and $n = 1000$ under the three models. For $n = 100$, the contribution of the parameter uncertainty, i.e., $\mathrm{PEI}_{100|1}$ is 1.74 for the true model ARMA(1,1), and smaller (1.07 for AR(1) and 1.03 for MA(1)) for the two misspecified models. This is expected, because the number of parameters estimated under the true model is larger than that estimated under each of the misspecified models. On the contrary, there is no contribution of model uncertainty in the true model, whereas we have $\mathrm{PMI}_{100|1}$ as 3.06 and 25.38 for the AR(1) and the MA(1) models respectively. This is also expected, as MA(1) is a much poorer replacement of ARMA(1,1) than AR(1). For $n = 1000$, as expected, the PEI almost vanishes for all three models, but the PMI remains almost the same for all models.

Figure 2 presents a more comprehensive picture of this misspecification, in which indices PEI and PMI are plotted against the sample size n and represented by areas. Three different forecast horizons $h = 1, 2, 3$ are considered. Figure 2 shows that when the correct model is fitted, the inflation of forecast error is only affected by the estimation of parameters, which decreases as the sample size increases. Otherwise, when fitting misspecified models, the inflation is larger and comes mostly from the incorrect choice of model. The PEI becomes negligible as the sample size increases, but the PMI remains more or less unchanged. The PMI is much larger for MA(1), as compared to that of AR(1), and attains its maximum for $h = 2$. This behavior, of course, may change if the parameters of the true model change.

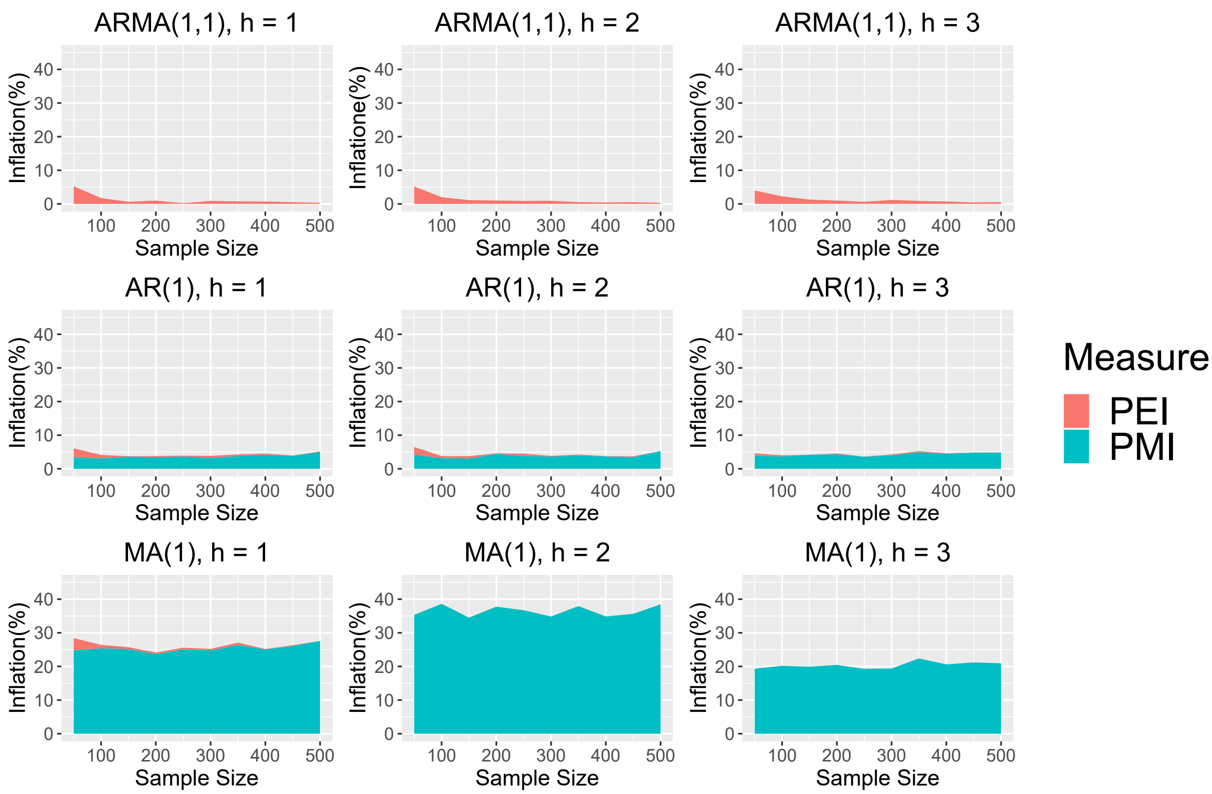

Figure 2: Decomposition of percentage inflation in forecast errors for a true ARMA(1,1) process with parameters $\phi = 0.8$, $\theta = -0.3$, and $\sigma^2 = 100$ under misspecified models AR(1) and MA(1).

Example: SARIMA models

We consider daily data produced by a seasonal ARMA $(2,2) \times (2,2)_7$ process with a weekly seasonal component, and consider forecasting using the following sequence of models, each of which is "weaker" (in terms of departure from the true model) than the preceding one, in the sense that:

- ARMA(2,2)(1,2)$_7$, where one autoregressive term of the season component is dropped.

- ARMA(2,2)(1,1)$_7$, where both the AR and MA orders of the seasonal component are one less than that of the true model.

- ARMA(2,2)(1,0)$_7$, where the moving average term of the weekly seasonal component is omitted.

- ARMA(2,2), completely dropping the seasonal component.

- ARMA(1,2), incorrectly modeling the main ARMA process.

- ARMA(1,1), the weakest model in the sequence.

Two additional models, an ARMA(2,2)(2,3)$_7$ and an ARMA(2,2)(3,3)$_7$, are also considered, with the aim of analyzing how the forecast is inflated for overfitted models.

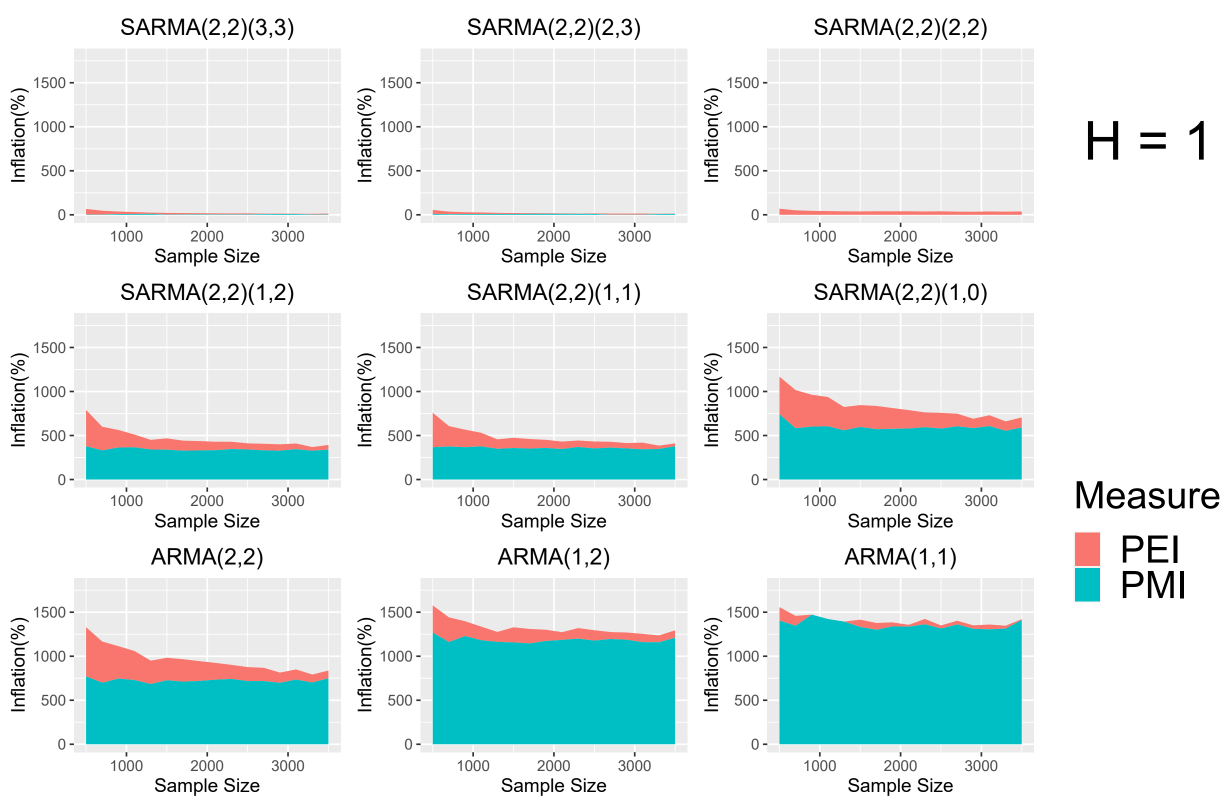

Figure 3: Decomposition of percentage inflation in forecast errors for a true SARMA(2,2)(2,2)$_7$ process for 1-step ahead predictions. 5000 simulations have been run for every scenario.

Figure 3 shows the decomposition of percentage inflation under each model for one-step ahead forecasts. The consequences of forecasting from misspecified models and how the inflation of forecast MSE increases as one drops components of the true model is consistent with the expectations. Additional simulations with 7 and 14-step ahead forecasts (not shown here) reveal that the consequences of misspecifying models becomes less severe as the forecast horizon increases.

5 Two Applications

In this Section, we demonstrate the proposed approach for assessment and decomposition of inflation of forecast error under possibly misspecified models with two examples.

5.1 Application 1: Re-visiting Kendall's suggested fitting of ARMA models

We now re-visit an interesting topic in time series that was first discussed by Kendall (1971), and subsequently debated (Box and Jenkins, 1973) and studied by researchers (e.g., Davies and Newbold, 1980). Kendall (1971) had suggested fitting moderately long autoregressive models as a quick and easy alternative of identifying orders of ARMA models. The counter-argument is the inflation of forecast error arising from increased parameter uncertainty that is a consequence of sacrificing parsimony.

As in our first simulation example, we assume that the true data generating process is a zero mean ARMA(1,1) process with parameters $\phi = 0.8$ and $\theta = -0.3$, and consider the misspecified model as a zero mean AR(p_M) process for different values of p_M. Let $\phi_{p_M,j}$ denote the jth parameter of the AR(p_M) process, where $j = 1, \ldots, p_M$, i.e., ϕ_{11} denotes the only parameter of the AR(1) process, ϕ_{21} and ϕ_{22} denote the two parameters of the AR(2) process, and so on. First, along the lines of Definition 1, we provide the optimal values of the AR parameters $\phi^*_{p_M,j}$ for $p_M \in \{1, \ldots, 8\}$ and $j = 1, \ldots, p_M$ in Table 2. It is interesting to note that for fixed j, the sequence $\{\phi^*_{p_M,j}\}_{j\geq 1}$ converges to some ϕ^*_j as p_M increases. That is, the autoregressive coefficients tend to stabilize as the AR order increases. For example, the optimal value of $\phi_{p_M 1}$ is 0.6281 for $p_M = 1$, but it converges to 0.5063 for $p_M \geq 4$.

Figure 4 shows the decomposition of percentage inflation in forecast errors when these autoregressive models are fitted to the ARMA(1,1) process. We note that for $h = 1, 2, 3$, an AR process of order 4 or greater appears to represent the ARMA(1,1) process fairly well, with the PMI index almost becoming negligible for $p_M \geq 4$. However, increasing the number of parameters has a clear adverse effect on the inflation of the forecast error through the increase of the PEI index, which is a consequence of sacrificing parsimony.

$\phi^*_{p_M,j}$	p_M							
	1	2	3	4	5	6	7	8
$\phi^*_{p_M,1}$	0.6281	0.5176	0.5071	0.5063	0.5063	0.5063	0.5063	0.5063
$\phi^*_{p_M,2}$		0.1759	0.1449	0.1431	0.1428	0.1428	0.1428	0.1428
$\phi^*_{p_M,3}$			0.0599	0.0536	0.0529	0.0527	0.0528	0.0528
$\phi^*_{p_M,4}$				0.0123	0.0097	0.0094	0.0096	0.0096
$\phi^*_{p_M,5}$					0.0052	0.0043	0.0047	0.0045
$\phi^*_{p_M,6}$						0.0018	0.0035	0.0029
$\phi^*_{p_M,7}$							-0.0032	-0.0053
$\phi^*_{p_M,8}$								0.0040

Table 2: Optimal $\phi^*_{p_M,j}$ for AR(p^*_M), when the true model is ARMA(1,1) with mean zero, $\phi = 0.8$ and $\theta = -0.3$.

5.2 Application 2: Traffic Engineering example

Some well-known daily seasonal ARIMA models considered appropriate for traffic data are: (i) $(1, 0, 1) \times (0, 1, 1)_{96}$ (Williams et al., 1998), (ii) $(2, 0, 1) \times (0, 1, 1)_{96}$ (Ghosh et al., 2005), and (iii) $(2, 0, 0) \times (0, 1, 1)_{144}$ (Kumar and Vanajakshi, 2015). The seasonal index in models (i) and (ii) is 96, whereas that for model (iii) is 144 is due to the slight difference in data collection intervals for these models. Whereas models (i) and (ii) were fitted using data collected at 15-minute intervals, model (iii) was fit using data collected at 10-minute intervals. However, because the patterns observed for 10-minute and 15-minute resolution traffic data are similar, model (iii) can be considered equivalent to a $(2, 0, 0) \times (0, 1, 1)_{96}$ model fitted with 15-minute resolution traffic data.

Among these three models, model (ii) has the highest number of parameters, and can be con-

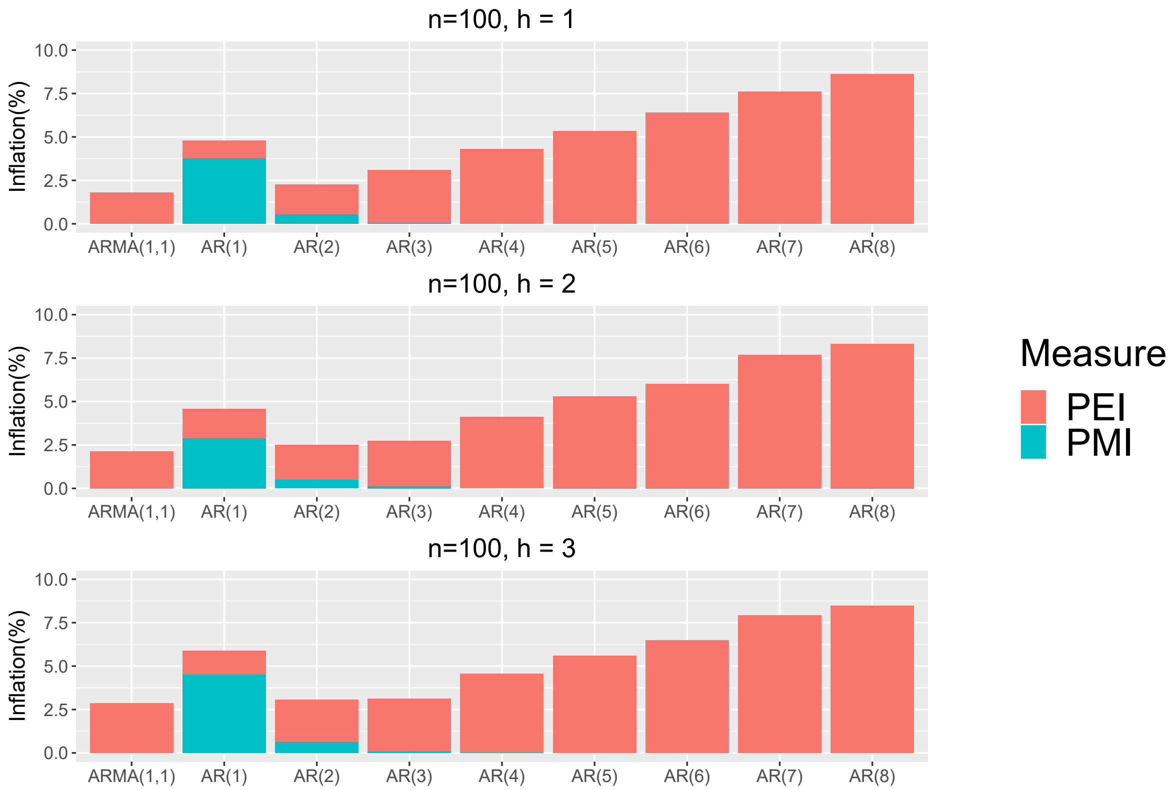

Figure 4: Decomposition of percentage inflation in forecast errors for a true ARMA(1,1) process with parameters $\phi = 0.8$, $\theta = -0.3$, and $\sigma^2 = 100$ fitting autoregressive models. 5000 simulations have been run for every scenario.

sidered to contain all the information that models (i) and (ii) have. Thus, assuming model (ii) to be the true model, we generate data and fit the three models to assess the inflation of forecast error for the remaining two. This assessment essentially boils down to the study of the influence of dropping a second order AR parameter (in model (i)) and a first order MA parameter (in model (ii)) when these two parameters exist in the true model (ii). For the sake of completeness, we also consider an additional model $(2, 0, 1) \times (1, 1, 1)_{96}$ to examine the influence of adding an unnecessary seasonal AR term. We will refer this model as model (iv) in the subsequent discussion.

Figure 5 shows the percentage inflation in forecast error when generating data from true $(2, 0, 1) \times (0, 1, 1)_{96}$ processes for 1- to 6-step ahead forecasts, which corresponds to 15-minute to 90-minute ahead forecasts for 15-minute resolution data. Models (ii) and (iv) exhibit a similar level of inflation in forecast error for every horizon, suggesting that the addition of a seasonal AR term does not hurt much. However, models (i) and (iii) both exhibit significant inflations in forecast error, with model (i) performing the worst in terms of PMI. This observation indicates the importance of the second AR term present in the true model, compared to the only MA term.

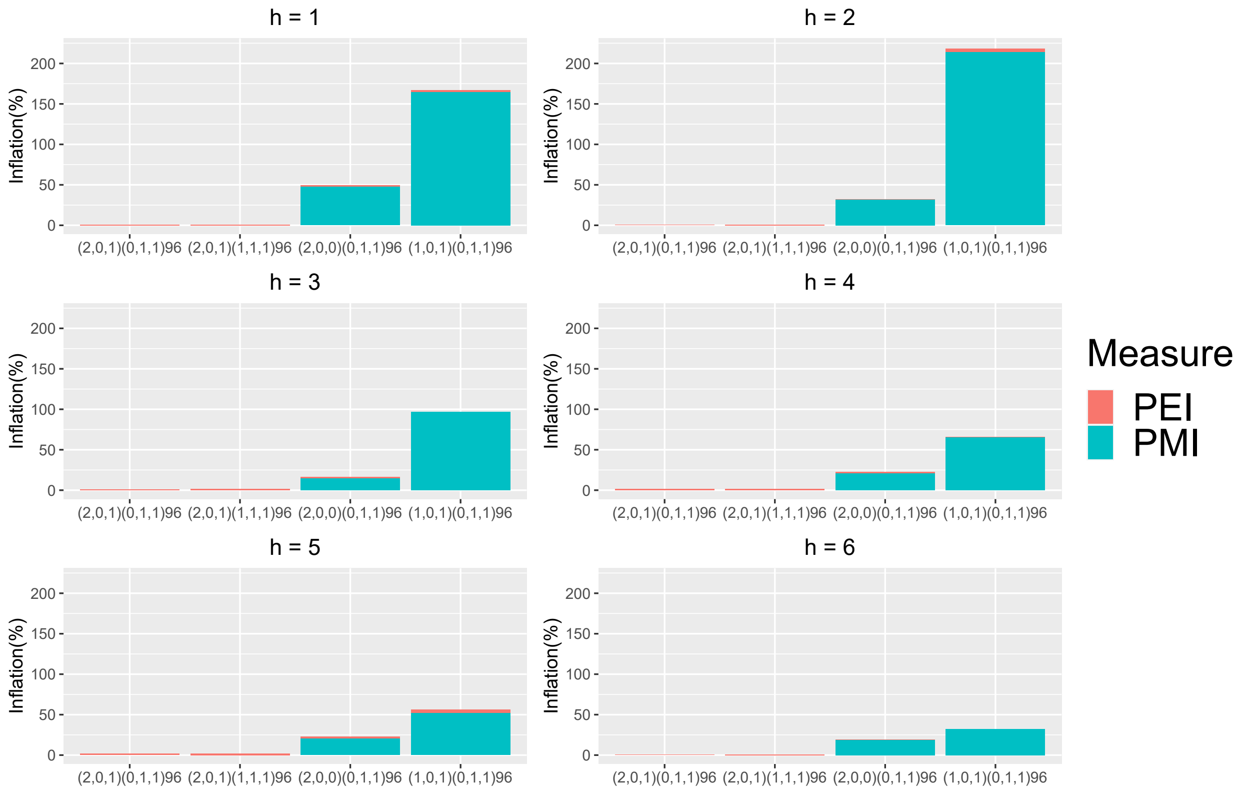

Figure 5: Decomposition of percentage inflation in forecast errors for true SARIMA$(2,0,1)(0,1,1)_{96}$ processes for a sample size n = 500. 1000 simulations have been run for every scenario.

6 Discussion

In this paper, we investigate the consequences of using a misspecified model for time series forecasting on the forecast error. On the basis of a decomposition of the MSE of the forecast obtained from the misspecified model, we define two indices associated with the inflation of the MSE compared to the true model. One quantifies the inflation associated with the incorrect use of model whereas the other measures the contribution of parameter estimation from the incorrect model. A simulation algorithm is proposed to perform this assessment for any ARIMA or SARIMA model, assuming that both the true model and the misspecified model can be converted to a stationary ARMA process by differencing the same number of times. The proposed framework helps to assess the consequences of sacrificing information by forecasting from models of lower order compared to true models of higher order. On the other hand, it also helps assess the consequences of using unnecessarily complex and larger models compared to the true model. We believe that the latter assessment is particularly important in a world where almost unlimited computing power is creating a natural tendency to overfit, without paying enough attention to the probable consequences of overfitting.

References

Boshnakov, G. and J. Halliday (2020). Package 'sarima'. `https://cran.r-project.org/web/packages/sarima/sarima.pdf`. [Online; last accessed 15-April-2022].

Box, G. and G. M. Jenkins (1973). Some comments on a paper by Chatfield and Prothero and on a review by Kendall. *Journal of the Royal Statistical Society, Series A 136*, 337–352.

Box, G. E. P. and G. M. Jenkins (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.

Brockwell, P. J. and R. A. Davis (2002). *Introduction to Time Series and Forecasting, 2nd. ed.* Springer-Verlang.

Chatfield, C. (1996). Model uncertainty and forecast accuracy. *Journal of Forecasting 15*, 495–508.

Davies, N. and P. Newbold (1980). Forecasting with misspecified models. *Journal of the Royal Statistical Society, Series C 29*, 87–92.

Fuller, W. A. (1996). *Introduction to Statistical Time Series Analysis*. San Francisco:John Wiley & Sons.

Ghosh, B., B. Basu, and M. O'Mahony (2005). Time-series modelling for forecasting vehicular traffic flow in Dublin. In *Proceedings of the 85th Transportation Research Board Annual Meeting, Washington, DC.*

Hamed, M. M., H. R. Al-Masaeid, and Z. M. B. Said (1995). Short-term prediction of traffic volume in urban arterials. *Journal of Transportation Engineering 121*(3), 249–254.

Hyndman, R. J., G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, and S. Razbash (2020). Package 'forecast'. *[Online, last accessed 17-Jan-2021] https://cran.r-project. org/web/packages/forecast/forecast.pdf*.

Kendall, M. G. (1971). Review of Box and Jenkins (1970). *Journal of the Royal Statistical Society, Series A 134*, 450–453.

Kumar, S. V. and L. Vanajakshi (2015). Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *European Transport Research Review 7*(3), 21.

Mac Nally, R., R. P. Duncan, J. R. Thomson, and J. D. Yen (2018). Model selection using information criteria, but is the "best" model any good? *Journal of Applied Ecology 55*(3), 1441–1444.

Mangan, N. M., J. N. Kutz, S. L. Brunton, and J. L. Proctor (2017). Model selection for dynamical systems via sparse regression and information criteria. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 473*(2204), 20170009.

Mazzeu, J. H. G., E. Ruiz, and H. Veiga (2018). Uncertainty and density forecasts of arma models: comparison of asymptotic, bayesian, and bootstrap procedures. *Journal of Economic Surveys 32*, 388–419.

Pukkila, T., S. Koreisha, and A. Kallinen (1990). The Identification of ARMA Models. *Biometrika 77*, 537–548.

Team, R. C. and C. Worldwide (2002). The R stats package. *R Foundation for Statistical Computing, Vienna, Austria: Available from: http://www.R-project.org*.

Williams, B. M., P. K. Durvasula, and D. E. Brown (1998). Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models. *Transportation Research Record 1644*(1), 132–141.

Appendix

Equivalence of Definition 1 and the minimizer of the KL distance

The precise meaning of this equivalence is elaborated as follows. Let $f_{\theta_M,n}(y_1,\dots,y_n)$ and $f_n(y_1,\dots,y_n)$ be the joint densities of $\{Y_1,\dots,Y_n\}$ under the model M with parameter θ_M and the true model $\mathcal{M}$ respectively. We shall consider the limit of the scaled KL-divergence

$$\frac{1}{n}\mathrm{KL}(f_{\theta_M,n}\|f_n) = \frac{1}{n}\int \log\left(\frac{f_n(y_1,\dots,y_n)}{f_{\theta_M,n}(y_1,\dots,y_n)}\right) f_n(y_1,\dots,y_n)\,dy_1\dots dy_n.$$

Let us use $\boldsymbol{y}_n := (y_1,\dots,y_n)'$ to simplify the notation. Also let $\hat{y}_k$ be the best linear prediction of y_k using $y_1,\dots,y_{k-1}$ under the model M with parameter θ_M, and denote by ν_j the corresponding prediction error variance. Under normality of M, it holds that

$$-\frac{1}{n}\log f_{\theta_M,n}(\boldsymbol{y}_n) = \frac{1}{2n}\sum_{k=1}^{n}\left[\log(2\pi\nu_k) + \frac{(y_k-\hat{y}_k)^2}{\nu_k}\right].$$

As $k\to\infty$, it holds that $\hat{y}_k \to \hat{y}_k^{\theta_M}$, and

$$\lim_{k\to\infty}\nu_k = \nu_\infty := E_{\theta_M}\left(y_k - \hat{y}_k^{\theta_M}\right)^2.$$

In the limit of the KL divergence,

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{KL}(f_{\theta_M,n}\|f_n) = \lim_{n\to\infty}\frac{1}{n}\int\log\left[f_n(\boldsymbol{y}_n)\right]f_n(\boldsymbol{y}_n)\,d\boldsymbol{y}_n - \lim_{n\to\infty}\frac{1}{n}\int\log\left[f_{\theta_M,n}(\boldsymbol{y}_n)\right]f_n(\boldsymbol{y}_n)\,d\boldsymbol{y}_n,$$

since the first term does not involve θ_M, we aim to minimize the second term, which in the limit becomes

$$-\lim_{n\to\infty}\frac{1}{n}\int\log\left[f_{\theta_M,n}(\boldsymbol{y}_n)\right]f_n(\boldsymbol{y}_n)\,d\boldsymbol{y}_n = \frac{1}{2}\log(2\pi\nu_\infty) + \frac{1}{2\nu_\infty}E_{\mathcal{M}}\left[\hat{Y}_{n+1}^{\theta_M} - Y_{n+1}\right]^2.$$

This leads to our definition of the best misspecified model.

Summary of R packages and simulations.

All the simulations have been run using R. The ARMA(1,1) series are obtained using the `arima.sim()` function available in the package stats (Team and Worldwide, 2002). As this function does not support seasonality, the seasonal ARIMA processes are modelled by the function `sim_sarima()` from the package sarima (Boshnakov and Halliday, 2020). The function `Arima()` from the package forecast (Hyndman et al., 2020) is used to fit both true and misspecified models. The predictions are finally estimated with the function `predict()`, also available in the package stats.