Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

A MISSPECIFICATION TEST FOR SIMULATION METAMODELS

Shiyu Wang
Guangwu Liu
Kun Zhang

Department of Management Sciences, College of Business
City University of Hong Kong
83 Tat Chee Avenue, Kowloon, HONG KONG

ABSTRACT

In this paper we propose a novel misspecification test for simulation metamodels. It is a consistent test that helps to assess the adequacy of simulation metamodels. The test statistic we construct is shown to be asymptotically normally distributed under the null hypothesis that the metamodel is correct, while diverging to infinity at a rate of $\sqrt{n}$, where n is the test sample size if the given metamodel is inadequate. Furthermore, as a by-product, we construct confidence intervals for mean squared errors of the metamodels. Preliminary numerical studies show that the test works quite well and has good finite-sample properties.

1 INTRODUCTION

Simulation modeling often serves as an effective tool to study complex real systems with noises. Simulation models can provide useful insights on the behavior of the real systems. For instance, they can be used to study how the system performance changes with respect to a set of design parameters. This information can then be used to find better design parameters to improve or optimize the system performance. Let θ denote the vector-valued design parameter of the system being simulated. The simulated output of the system is usually a function of θ with noise, denoted by $Y(\theta)$, and the performance of the system is often measured by the expected value of $Y(\theta)$. In other words, the performance of the system of interest is a function of the design parameter vector θ:

$$f(\theta) = \mathbb{E}[Y(\theta)], \quad \theta \in \Theta,$$

where Θ denotes a set within which θ may take value.

The functional form of f is typically unknown, and simulation experiments may be conducted to learn this unknown form. In principle, for any given value of θ, one may simulate a number of independent and identically distributed (i.i.d.) copies of $Y(\theta)$ and take their sample average to estimate $f(\theta)$. However, applying this estimation procedure for a large number of possible values of θ is often time consuming. To circumvent this difficulty, one may attempt to find some approximating functions of f that are inexpensive to compute. Such approximating functions are usually referred to as simulation metamodels that are essentially models of the simulation model; see, e.g., Kleijnen (1986) and Barton and Meckesheimer (2006).

When θ is considered as the design parameter vector, it takes deterministic values within Θ. This setting can be extended to a more general case by allowing θ to take random values according to a certain probability distribution. For instance, in portfolio risk measurement, θ may represent a set of risk factors valued at a future time horizon, while $Y(\theta)$ represents the (discounted) random loss of a portfolio at maturity. Then $f(\theta) \triangleq \mathbb{E}[Y(\theta)|\theta]$ represents the loss over the considered time horizon for a given scenario of the risk factors θ, i.e., the loss over the considered time horizon is a function of θ. In later presentation of the paper, we abuse the notation a bit and let

$$f(\mathbf{x}) \triangleq \mathbb{E}[Y|\mathbf{X} = \mathbf{x}]$$

denote the unknown function f, where the dependence of Y on $\mathbf{X}$ is suppressed for ease of notation. In our setting, $\mathbf{X}$ is allowed to take deterministic values or random values, analogous to fixed design and random design in regression analysis. Here $\mathbf{X}$ and $\mathbf{x}$ take values in $\mathbb{R}^d$, while the random variable Y is one-dimensional.

When a metamodel is used to approximate the unknown function f, one of the major issues is on the adequacy of the metamodel, i.e., how well the metamodel approximates f. Let $g(\mathbf{x})$ denote the metamodel. Then a natural idea is to construct a statistical test based on certain measures of distance between $g(\mathbf{x})$ and $f(\mathbf{x})$ for the $\mathbf{x}$ values of interest. However, how to construct such a test is far from trivial because the functional form of f is unknown. In this paper, we attempt to tackle this issue by developing a hypothesis test to assess the adequacy of a simulation metamodel. More specifically, we construct a misspecification test based on an estimator of the mean squared error (MSE) of the metamodel, which is defined as

$$\mathbb{E}\left[(g(\mathbf{X})-f(\mathbf{X}))^2\right].$$

While the meaning of the expectation is straightforward when $\mathbf{X}$ is a random vector, it should be pointed out that when $\mathbf{X}$ takes deterministic values, definition of the above expectation needs to be defined appropriately. For instance, if $\{\mathbf{x}_1,\ldots,\mathbf{x}_m\}$ is a set of values of $\mathbf{x}$ that are important in assessing the metamodel adequacy, then the expectation can be defined as weighted average of $\left\{(g(\mathbf{x}_i)-f(\mathbf{x}_i))^2, i=1,\ldots,m\right\}$ where the weights can be specified by the simulation user based on relative importance of the $\mathbf{x}_i$'s.

An attractive feature of the proposed misspecification test is that estimating the MSE of the metamodel $g(\cdot)$ does not require knowing the functional form of f, which is inspired by a recent work of Goda (2017) on estimating the variance of a conditional expectation. This feature allows for substantial flexibility in conducting misspecification test. For instance, $\mathbf{X}$ is allowed to take discrete values, continuous values, or even mixed.

This work is related to a vast literature in statistics and econometrics on misspecification tests (sometimes also called goodness-of-fit tests) of regression models. A main stream of this literature is based on smoothing ideas that estimate the unknown functional form of f in a nonparametric manner and then compare the approximating function g to the nonparametric estimate of f. In particular, many tests for a parametric/semiparameteric model contrast the parametric model with a completely nonparametric one which can guarantee the consistency of the tests. The noteworthy works along this line include Lee (1988), Yatchew (1992), Eubank and Spiegelman (1990), Hardle and Mammen (1993), Wooldridge (1992), Zheng (1996), Dette (1999), Guerre and Lavergne (2005), and Hsiao, Li, and Racine (2007). Both Lee's and Yatchew's tests are based on comparing the parametric sum of squared residuals with the nonparametric sum of squared residuals. Eubank and Spiegelman construct their test by fitting a spline smoothing to the residuals from a linear regression. Wooldridge proposes a test for neglected nonlinearities. It is based on the residual test of Davidson and MacKinnon (1981) and estimation of an infinite-dimensional alternative. Hardle and Mammen's test is based on the integrated squared difference between the parametric fit and the nonparametric fit. Zheng proposes a quadratic form of test that combines the idea of the conditional moment of the residuals via a nonparametric kernel method. Dette's test is based on a comparison of a nonparametric and a parametric estimator of the integrated variance function. Guerre and Lavergne construct new data-driven smoothing tests which select the smoothing parameter through a new criterion and are proved to be adaptive rate-optimal and consistent. Hsiao, Li, and Racine's test is an extension of Zheng's to the case where covariates $\mathbf{X}$ may contain a discrete component.

However, all of the aforementioned misspecification tests suffer from the "curse of dimensionality" that refers to the poor performances of the methods as the dimension of the covariates $\mathbf{X}$ increases, see, e.g., Stone et al. (1980). The curse of dimensionality stems from the nature of nonparametric estimation of the unknown functional form of f. This difficulty motivates different modifications of the misspecification tests. To address this issue, the works by Lavergne and Patilea (2008), Xia (2009) and Guo, Wang, and Zhu (2016) should be noticed. The main idea of their studies, inspired by the projection pursuit, is to use a single linear index $\mathbf{X}^T\beta$ as a conditioning variable instead of $\mathbf{X}$, where $\|\beta\|=1$. In this way, the test behaves almost as if the dimension of the covariates was one. However, such tests rely on that f depends on $\mathbf{X}$ through the index $\mathbf{X}^T\beta$ for some beta, a restrictive assumption that is typically not verifiable for practical problems. In addition, tests based on empirical regression processes to avoid the selection of a smoothing parameter can be considered as refinements of misspecification tests in the literature. An omnibus literature review on misspecification tests for regression models is provided by González-Manteiga and Crujeiras (2013).

In short, a major contribution of this paper is on the development of a new misspecification test that is tailor-made to simulation metamodeling. Compared to existing misspecification tests in the vast literature in econometrics, the test we propose in this paper has the following advantages.

- It does not suffer from the curse of dimensionality.
- It is easy to implement and does not involve any smoothing parameters.
- It can be applied when the covariates $\mathbf{X}$ has both discrete and continuous components.

It shall be seen in later presentation that the proposed test requires two independent simulation outputs for a scenario of $\mathbf{X}$. While this is indeed a limitation when real data is used to conduct the test, requirement of two independent copies of simulated data is usually not an issue in the context of simulation metamodeling. Throughout the paper, we assume that two independent simulation outputs for any scenario of $\mathbf{X}$ can be simulated.

The rest of this paper is organized as follows. In Section 2, we introduce the problem and construct an estimator of the MSE of any given metamodel . Based on the estimator of the MSE, we construct a test statistic and establish the asymptotic distributions of it under the null and the fixed alternative in Section 3. We further construct the confidence intervals for the MSE of the metamodel. In Section 4, we analyze the power of the test against local misspecification approaching the true simulation model at rates slower or equal to $n^{-1/4}$. Two numerical experiments are presented in Section 5, followed by conclusions in Section 6.

2 ESTIMATION OF MEAN SQUARED ERROR

In this paper, we assume that the the system output Y is one-dimensional. When the system performance of interest is multidimensional, one may construct a metamodel for each of the dimensions separately (cf. Kleijnen 1986). We further assume that $\mathbf{X}$ is d-dimensional, taking values in $\mathscr{X} \subset \mathbb{R}^d$, where its individual coordinates may take discrete or continuous values.

We let g denote a given metamodel. For any given covariates $\mathbf{x} \in \mathscr{X}$, the metamodel provides a quick prediction $g(\mathbf{x})$ that is used to approximate $f(\mathbf{x})$. Our objective is to assess the adequacy of the given metamodel g by testing whether the metamodel g exactly describes the true functional form of f. To do so, we define the null and the alternative hypotheses as follows:

$$H_0: \ \mathbb{P}(g(\mathbf{X}) = f(\mathbf{X})) = 1,$$
$$H_1: \ \mathbb{P}(g(\mathbf{X}) = f(\mathbf{X})) < 1.$$

Denote the MSE of the metamodel g as

$$M_g = \mathbb{E}\left[(f(\mathbf{X}) - g(\mathbf{X}))^2\right]. \tag{1}$$

MSE measures the deviation of g from the true functional form of f, and can be used to characterize the fidelity of the metamodel.

Then under H_0, since $g(\mathbf{X}) = f(\mathbf{X})$ almost surely, we have

$$M_g = 0,$$

while under H_1, we have

$$M_g > 0.$$

Therefore, we may construct an estimator of M_g to form a test. Inspired by Goda (2017), we construct an estimator of M_g as follows. We simply rewrite M_g as

$$\begin{aligned}
M_g &= \mathbb{E}\left[f(\mathbf{X}) - g(\mathbf{X})\right]^2 = \mathbb{E}[\mathbb{E}[Y - g(\mathbf{X})|\mathbf{X}]]^2 \\
&= \mathbb{E}\left[\mathbb{E}[Y - g(\mathbf{X})|\mathbf{X}] \cdot \mathbb{E}[Y - g(\mathbf{X})|\mathbf{X}]\right] \\
&= \mathbb{E}\left[\mathbb{E}[Y - g(\mathbf{X})|\mathbf{X}] \cdot \mathbb{E}[Y' - g(\mathbf{X})|\mathbf{X}]\right] \\
&= \mathbb{E}\left[\mathbb{E}[(Y - g(\mathbf{X})) \cdot (Y' - g(\mathbf{X}))|\mathbf{X}]\right] \\
&= \mathbb{E}\left[(Y - g(\mathbf{X})) \cdot (Y' - g(\mathbf{X}))\right],
\end{aligned}$$

where Y and Y' are i.i.d. conditional on $\mathbf{X}$. In this way, given the samples $\{\mathbf{X}_i, Y_i, Y_i'\}_{1 \le i \le n}$, where $\mathbf{X}_i$ is sampled randomly from the marginal distribution of $\mathbf{X}$, and Y_i and Y_i' are sampled independently from the conditional distribution of Y given $\mathbf{X}_i$, we obtain an estimator of M_g as follows:

$$V_n = \frac{1}{n} \sum_{i=1}^{n} (Y_i - g(\mathbf{X}_i))(Y_i' - g(\mathbf{X}_i)). \tag{2}$$

3 TEST STATISTIC AND ITS LIMITING DISTRIBUTION

In this section, we construct a test statistic and analyze its limiting distribution. More specifically, a test statistic is constructed based on V_n. To this end, a natural starting point is to analyze the asymptotic distribution of V_n. To facilitate analysis, we make the following assumption.

Assumption 1 The metamodel $g(\mathbf{x})$ is a Borel measurable function on $\mathscr{X}$ and

$$\mathbb{E}\left[(Y - g(\mathbf{X}))^4\right] < \infty.$$

Assumption 1 is a mild moment condition on $Y - g(\mathbf{X})$. It is satisfied when the fourth moments of Y and $g(\mathbf{X})$ are finite. Based on this assumption, we can derive the asymptotic distributions of V_n under the null and the alternative hypotheses. In particular, the result on asymptotic distribution of V_n under the null hypothesis is summarized in the following proposition, whose proof is provided in the appendix.

Proposition 1 If Assumption 1 holds, then under the null hypothesis,

$$\sqrt{n} V_n \xrightarrow{d} N(0, \sigma_0^2),$$

as $n \to \infty$, where "$\xrightarrow{d}$" denotes convergence in distribution, and $N(0, \sigma_0^2)$ is a normal distribution with mean zero and a variance σ_0^2 defined by

$$\sigma_0^2 \triangleq \mathbb{E}\left[\mathrm{Var}(Y|\mathbf{X})\right]^2. \tag{3}$$

Moreover, σ_0^2 can be consistently estimated by $\hat{\sigma}_n^2$, where

$$\hat{\sigma}_n^2 \triangleq \frac{1}{n} \sum_{i=1}^{n} (Y_i - g(\mathbf{X}_i))^2 (Y_i' - g(\mathbf{X}_i))^2. \tag{4}$$

Proposition 1 shows that under the null hypothesis, $\sqrt{n} V_n$ is asymptotically normally distribution, and its asymptotic variance can be estimated by $\hat{\sigma}_n^2$. Therefore, we construct a test statistic T_n defined as

$$T_n = \sqrt{n} V_n / \hat{\sigma}_n. \tag{5}$$

It can be easily seen that T_n converges in distribution to a standard normal distribution under the null hypothesis when the sample size n goes to infinity. To summarize, we have the following theorem on limiting distribution of T_n.

Theorem 1 If Assumption 1 holds, then under the null hypothesis,

$$T_n \xrightarrow{d} N(0,1),$$

as $n \to \infty$.

Theorem 1 could be used to compute the asymptotic critical value for our misspecification test. Denote z_α as the α-quantile of the standard normal distribution. If the test statistic T_n is falling beyond the range of $[z_{\alpha/2}, -z_{\alpha/2}]$, we reject the null hypothesis with a significance level of α, implying that the given metamodel is inadequate. Otherwise, we fail to reject the null which means that we accept the statement "the metamodel is an adequate approximation of the true model" with a significance level of α.

A test that has asymptotic power equal to 1 is said to be consistent. To understand the power and show the consistency of the test, we establish the asymptotic distribution of V_n under the alternative hypothesis, which is summarized in the following proposition. Proof of the proposition is provided in the appendix.

Proposition 2 If Assumption 1 holds, then under the alternative hypothesis H_1,

$$V_n \xrightarrow{p} M_g > 0,$$

and

$$\hat{\sigma}_n^2 \xrightarrow{p} \mathbb{E}\left\{\mathbb{E}[(Y-g(\mathbf{X}))^2|\mathbf{X}]\right\}^2 > 0,$$

where "$\xrightarrow{p}$" denotes convergence in probability.

Applying Proposition 2, the asymptotic distribution of the test statistic T_n under the alternative hypothesis then follows.

Theorem 2 If Assumption 1 holds, then under the alternative hypothesis,

$$T_n/\sqrt{n} \xrightarrow{p} \frac{M_g}{\mathbb{E}\left\{\mathbb{E}[(Y-g(\mathbf{X}))^2|\mathbf{X}]\right\}^2} > 0,$$

as $n \to \infty$.

Theorem 2 shows that $T_n \to \infty$ in probability under the alternative meaning that if the metamodel is inadequate. The probability of rejecting the null hypothesis converges to 1 as $n \to \infty$. Namely, the asymptotic power of the test is 1. In addition, T_n goes to infinity at a rate of $\sqrt{n}$.

Furthermore, by the central limit theorem, we have the following result.

Theorem 3 If Assumption 1 holds, then under the alternative hypothesis,

$$\sqrt{n}(V_n - M_g) \xrightarrow{d} N(0, \sigma_1^2),$$

as $n \to \infty$, where

$$\sigma_1^2 \triangleq \mathbb{E}\left\{\mathbb{E}[(Y-g(\mathbf{X}))^2|\mathbf{X}]\right\}^2. \tag{6}$$

Theorem 3 allows the construction of tests for precise hypotheses (cf. Berger and Delampady 1987)

$$H_0:\ M_g > \Delta, \tag{7}$$
$$H_1:\ M_g \le \Delta.$$

The null hypothesis is rejected if $\sqrt{n}(V_n - \Delta) \le \hat{\sigma}_n \cdot z_\alpha$, where $\hat{\sigma}_n^2$ is defined in (4) that is a consistent estimator of σ_1^2.

It should also be mentioned that Theorem 3 allows the construction of confidence intervals (CIs) for the MSE M_g, i.e., a CI with a confidence level β for M_g is given by

$$\left[V_n - z_{1-(1-\beta)/2} \cdot \frac{\hat{\sigma}_n}{\sqrt{n}},\ V_n + z_{1-(1-\beta)/2} \cdot \frac{\hat{\sigma}_n}{\sqrt{n}}\right]. \tag{8}$$

4 TESTS OF LOCAL ALTERNATIVES

In the previous section, we have shown that our test is consistent against all fixed alternatives. To further understand the power of the test, it is helpful to investigate how the test behaves against local misspecifications. When a sequence of local alternatives are allowed to get closer and closer to the null hypothesis as sample size increases, the power of the test may not converge to unity. In this section we focus on local power analysis which is defined as the evaluation of the behavior of the power function of our test in a neighborhood of the null hypothesis. Local power analysis has become an important technique in econometrics. For a more detailed account of local alternatives and related literature, interested readers are referred to McManus (1991).

Fundamental to this analysis is a study on how the test behaves at a sequence of alternatives that converge to the null as sample size grows. That is, we consider a sequence of local alternatives as follows:

$$H_{1n} : f(\mathbf{X}) = g(\mathbf{X}) + \delta_n \cdot e(\mathbf{X}), \tag{9}$$

where the known function $e(\cdot)$ has finite second moment and $\delta_n \to 0$ as $n \to \infty$.

The following theorem gives the power of our test against local alternatives, whose proof is provided in the appendix.

Theorem 4 Suppose that Assumption 1 holds. Under the local alternative hypothesis (9), if $\delta_n = n^{-1/4}$, then,

$$T_n \xrightarrow{d} N(\mu, 1),$$

as $n \to \infty$, where

$$\mu = \mathbb{E}[e^2(\mathbf{X})]/\sigma_1,$$

with σ_1^2 defined in (6).

Theorem 4 shows that our proposed test statistic T_n will converge to a normal with nonzero mean under the local alternatives approaching to the null at a rate of $n^{-1/4}$. In addition, from the proof of Theorem 4 it can be shown that T_n converges to infinity when δ_n converges to zero slower than $n^{-1/4}$. Therefore, it should be clear that our test has an asymptotic power equal to 1 against local alternatives that approach to the null at rates slower than $n^{-1/4}$. Note that we only require finite second moment of $e(\mathbf{X})$ without imposing any smoothness restriction on $e(\mathbf{X})$, as is frequent in local alternative analysis, see e.g. Zheng (1996).

5 SIMULATION STUDY

In this section, we consider two examples to examine the performances of the proposed misspecification test and to validate the constructed confidence intervals for mean squared errors of the metamodels. The first example is on portfolio risk measurement where $\mathbf{X}$ denotes the risk factors on which the discounted loss at maturity Y depends, while the second example is an M/G/2 queue in which $\mathbf{X}$ is taken to be the arrival rate and Y is an average waiting time.

5.1 Loss Function of A Portfolio

Consider a portfolio that consists of a European vanilla call option and an up-and-out barrier call option, written on two different underlying assets, respectively. Price dynamics of the underlying assets, denoted as $S_t^1, S_t^2, t \geq 0$, are governed by a two-dimensional geometric Brownian motion, i.e.,

$$dS_t^1 = \mu_1 S_t^1 dt + \sigma_1 S_t^1 dB_t^1, \qquad dS_t^2 = \mu_2 S_t^2 dt + \sigma_2 S_t^2 dB_t^2,$$

where μ_i, σ_i $(i = 1, 2)$ represent the return rates and the volatilities of ith asset, respectively, and B_t^1 and B_t^2 are two independent standard Brownian motions.

Suppose that options in the portfolio have the same maturity T. Denote the portfolio value at time t by V_t. At a future time $\tau(\tau < T)$, the loss of the portfolio is a function of $S_\tau^1, S_\tau^2, \max_{0\le t\le\tau} S_t^2$, and can be written as

$$\begin{aligned} f\left(S_\tau^1, S_\tau^2, \max_{0\le t\le\tau} S_t^2\right) &= V_0 - V_\tau \\ &= \mathbb{E}\left[V_0 - e^{-r(T-\tau)}(S_T^1 - K_1)^+ - e^{-r(T-\tau)}(S_T^2 - K_2)^+ \mathbf{1}\{\max_{0\le t\le T} S_t^2 < H\} \,\middle|\, S_\tau^1, S_\tau^2, \max_{0\le t\le\tau} S_t^2 \right], \end{aligned}$$

where r is the risk-free interest rate, K_1 and K_2 are strike prices of these two options respectively, H is the barrier level of the barrier option, and the current portfolio value V_0 is a known constant .

To approximate the unknown functional form of f, we consider the following four metamodels, where the coefficients are obtained by least-squares method and serve as inputs to the misspecification tests.

1. A metamodel which is a linear functional of the true f:

$$g_1\left(S_\tau^1, S_\tau^2, \max_{0\le t\le\tau} S_t^2\right) = a_0 + a_1 f\left(S_\tau^1, S_\tau^2, \max_{0\le t\le\tau} S_t^2\right).$$

2. A quadratic polynomial metamodel in the following form:

$$g_2\left(S_\tau^1, S_\tau^2, \max_{0\le t\le\tau} S_t^2\right) = b_0 + b_1(S_\tau^1 + S_\tau^2) + b_2(S_\tau^1 + S_\tau^2)^2.$$

3. A quadratic polynomial metamodel in the following form:

$$g_3\left(S_\tau^1, S_\tau^2, \max_{0\le t\le\tau} S_t^2\right) = c_0 + c_1(S_\tau^1 + S_\tau^2) + c_2(S_\tau^1 + S_\tau^2)^2 + c_3 \max_{0\le t\le\tau} S_t^2.$$

4. A quadratic polynomial metamodel in the following form:

$$g_4\left(S_\tau^1, S_\tau^2, \max_{0\le t\le\tau} S_t^2\right) = d_0 + d_1 S_\tau^1 + d_2 S_\tau^2 + d_3 \max_{0\le t\le\tau} S_t^2 + d_4\left(S_\tau^1\right)^2 + d_5(S_\tau^2)^2 + d_6\left(\max_{0\le t\le\tau} S_t^2\right)^2.$$

We vary the sample sizes n from 10^3 to 10^5, and set $\mu_1 = 15\%, \mu_2 = 8\%$, $\sigma_1 = 15\%, \sigma_2 = 8\%$, $K_1 = K_2 = 90$, $r = 5\%$, $H = 120$, $S_0 = 100$, $T = 1$, and $\tau = 1/25$. To demonstrate the performance of the test, 250 independent testing experiments are conducted to show the performances of the test statistics defined in (5) and the confidence intervals for the MSE of metamodels according to (8).

The proportions of rejections for four metamodels are shown in Figure 1. Results in this figure are interpreted as follows. For Metamodel 1 which represents the true model, it can be seen that the rejection rates in three plots become closer to the corresponding significance levels (1%, 5%, and 10% respectively) when n becomes larger. It indicates that our test gives an appropriate rejection rate for the correctly specified metamodel. For Metamodels 2 and 3 which represent inadequate metamodels, the proportions of rejections are higher and converge to 1 as n becomes larger. It is consistent with the theoretical result on the power of the test. For Metamodel 4, its proportion of rejections is close to the significance level although it doesn't describe the exact form of the portfolio loss. The failure to reject Metamodel 4 can be regarded as a recognition that it is an adequate approximation of the functional form of f.

Performances of the confidence intervals for the MSEs (M_g) of metamodels are reported in Figure 2. In Figure 2, we plot the observed coverage probabilities of the confidence intervals with nominal confidence levels 90%, 95% and 99% with respect to sample sizes n from 10^3 to 10^5. The observed coverage probability of a confidence interval is the proportion of replications in which the interval contains the true value of MSE. It can be seen from Figure 2 that the coverage probabilities in each plot are close to the corresponding nominal confidence levels when n becomes larger, which implies that the confidence intervals are asymptotically valid ones.

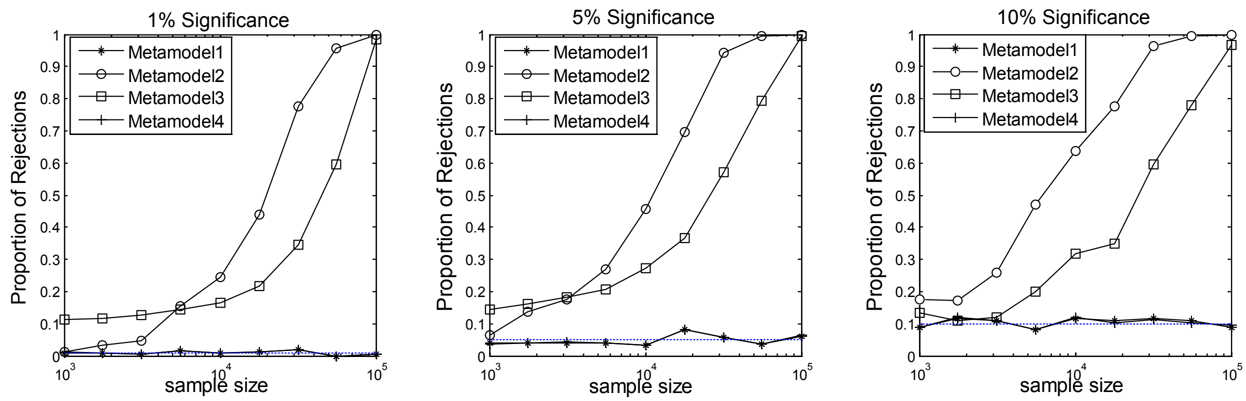

Figure 1: Proportion of rejections for metamodels with respect to different sample sizes for the portfolio example.

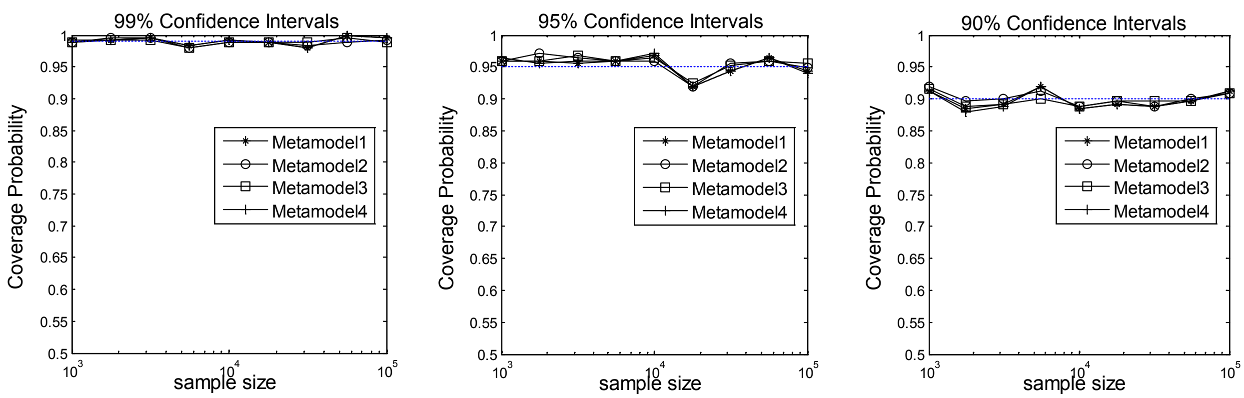

Figure 2: Observed coverage probabilities with respect to different sample sizes for MSEs of the metamodels for the portfolio example.

5.2 An M/G/2 Queue

Considering an M/G/2 queue where the arrivals follow a Poisson process with rate X, and service time distributions of both servers are independent and follow a Gamma distribution with shape parameter 2 and scale parameter 0.5. Denote the average waiting time of the first 100 customers by Y. To avoid explosion of the queue length, we let X take value in $(0,2)$. In the experiment, we sample X from a uniform $(0,2)$ distribution. We are interested in the functional form of $\mathbb{E}[Y|X]$ as a function of X, denoted by

$$f(x) = \mathbb{E}[Y|X = x].$$

To approximate the unknown functional form of f, we consider three metamodels, where the coefficients are obtained by least-squares method.

1. A linear metamodel:
$$g_1(x) = a_0 + a_1 x.$$
2. A quadratic polynomial metamodel:
$$g_2(x) = b_0 + b_1 x + b_2 x^2.$$
3. A cubic polynomial metamodel:
$$g_3(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3.$$

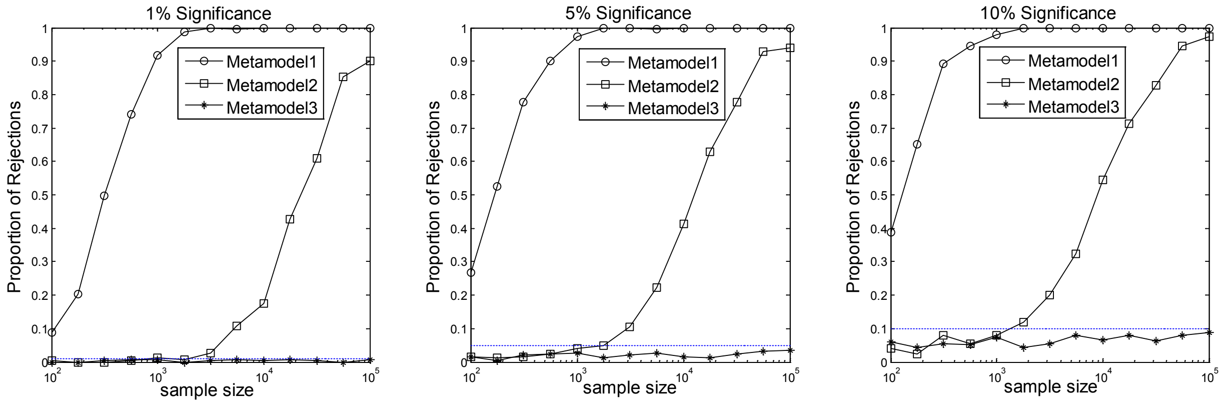

Figure 3: Proportion of rejections for metamodels with respect to different sample sizes for the M/G/2 Queue.

We conduct 250 independent testing experiments to measure the proportions of rejections for the three metamodels, and vary the sample size n from 10^2 to 10^5. The numerical results are summarized in Figure 3. For Metamodel 1, the proportions of rejections are higher and converge to 1 as n becomes larger. We reject Metamodel 1, suggesting that it is not a good approximation to the true function f. The proportions of rejections for Metamodel 2 is also becoming closer to 1 as sample sizes increases, implying a failure to pass the test. We reject Metamodel 2 according to the test results. The plot of Metamodel 3 shows that it has the best performance among the three metamodels since its proportion of rejections is close to the significance level as n becomes larger. This suggests that Metamodel 3 is a good approximation of the unknown function f.

6 CONCLUSIONS

In this paper we have proposed a misspecification test to assess the adequacy of simulation metamodels. The test is shown to be consistent against general inadequate metamodels and local inadequate metamodels that approach the true model at rates slower or equal to $n^{-1/4}$. The test is easy to implement and does not suffer from the curse of dimensionality. Moreover, it is suitable for both continuous and discrete covariates. Preliminary simulation studies of two examples show that the proposed test works quite well. One of the possible extensions is to test whether a class of metamodels parametrized by unknown parameters is adequate for the unknown functional form of f, which is part of an ongoing work of the authors.

ACKNOWLEDGMENTS

This research is partially supported by the Hong Kong Research Grants Council [GRF 192313 11504014].

A PROOF OF PROPOSITION 1

Since $\{\mathbf{X}_i, Y_i, Y_i'\}_{1\le i\le n}$ are i.i.d., applying the central limit theorem, it follows that

$$\sqrt{n}(V_n - \mathbb{E}[V_n]) \xrightarrow{d} N(0, \sigma_0^2),$$

where

$$\sigma_0^2 = \mathrm{Var}\left((Y_i - g(\mathbf{X}_i))(Y_i' - g(\mathbf{X}_i))\right).$$

Under the null hypothesis, it holds that

$$\begin{aligned}
\mathbb{E}[V_n] &= \mathbb{E}\left[(Y_i - g(\mathbf{X}_i))(Y_i' - g(\mathbf{X}_i))\right] \\
&= \mathbb{E}\left[\mathbb{E}[(Y - g(\mathbf{X}_i))(Y_i' - g(\mathbf{X}_i))|\mathbf{X}_i]\right] \\
&= \mathbb{E}\left[\mathbb{E}[Y_i - g(\mathbf{X}_i)|\mathbf{X}_i] \cdot \mathbb{E}[Y_i' - g(\mathbf{X}_i)|\mathbf{X}_i]\right] \\
&= \mathbb{E}\left[\mathbb{E}[Y - g(\mathbf{X})|\mathbf{X}] \cdot \mathbb{E}[Y - g(\mathbf{X})|\mathbf{X}]\right] \\
&= \mathbb{E}[\mathbb{E}[Y - g(\mathbf{X})|\mathbf{X}]]^2 = M_g = 0,
\end{aligned}$$

and

$$\begin{aligned}
\sigma_0^2 &= \mathrm{Var}\left((Y_i - g(\mathbf{X}_i))(Y_i' - g(\mathbf{X}_i))\right) \\
&= \mathbb{E}\left[(Y_i - g(\mathbf{X}_i))^2(Y_i' - g(\mathbf{X}_i))^2\right] - \left(\mathbb{E}\left[(Y_i - g(\mathbf{X}_i))(Y_i' - g(\mathbf{X}_i))\right]\right)^2 \\
&= \mathbb{E}\left[\mathbb{E}[(Y_i - g(\mathbf{X}_i))^2(Y_i' - g(\mathbf{X}_i))^2|\mathbf{X}_i]\right] - 0 \\
&= \mathbb{E}\left[\mathbb{E}[(Y_i - g(\mathbf{X}_i))^2|\mathbf{X}_i] \cdot \mathbb{E}[(Y_i' - g(\mathbf{X}_i))^2|\mathbf{X}_i]\right] \\
&= \mathbb{E}\left[\mathbb{E}[(Y - g(\mathbf{X}))^2|\mathbf{X}] \cdot \mathbb{E}[(Y - g(\mathbf{X}))^2|\mathbf{X}]\right] \\
&= \mathbb{E}\left[\mathbb{E}[(Y - g(\mathbf{X}))^2|\mathbf{X}]\right]^2 = \mathbb{E}\left[\mathrm{Var}(Y - g(\mathbf{X})|\mathbf{X}) + (\mathbb{E}[Y - g(\mathbf{X})|\mathbf{X}])^2\right]^2 \\
&= \mathbb{E}\left[\mathrm{Var}(Y - g(\mathbf{X})|\mathbf{X})\right]^2 = \mathbb{E}\left[\mathrm{Var}(Y|\mathbf{X})\right]^2.
\end{aligned}$$

Furthermore, by applying the weak law of large number, it holds that

$$\hat{\sigma}_n^2 \xrightarrow{p} \mathbb{E}[\hat{\sigma}_n^2],$$

where

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}_n^2] &= \mathbb{E}\left[(Y_i - g(\mathbf{X}_i))^2(Y_i' - g(\mathbf{X}_i))^2\right] \\
&= \mathbb{E}\left[\mathbb{E}[(Y_i - g(\mathbf{X}_i))^2(Y_i' - g(\mathbf{X}_i))^2|\mathbf{X}_i]\right] \\
&= \mathbb{E}\left[\mathbb{E}[(Y_i - g(\mathbf{X}_i))^2|\mathbf{X}_i]\mathbb{E}[(Y_i' - g(\mathbf{X}_i))^2|\mathbf{X}_i]\right] \\
&= \mathbb{E}\left[\mathbb{E}[(Y - g(\mathbf{X}))^2|\mathbf{X}]\mathbb{E}[(Y' - g(\mathbf{X}))^2|\mathbf{X}]\right] \\
&= \mathbb{E}\left[\mathbb{E}[(Y - g(\mathbf{X}))^2|\mathbf{X}]\right]^2 = \sigma_0^2.
\end{aligned}$$

Therefore, the proof is completed.

B PROOF OF PROPOSITION 2

Similar to the proof of Proposition 1, we apply the weak law of large number and obtain

$$V_n \xrightarrow{p} \mathbb{E}\left[(Y_i - g(\mathbf{X}_i))(Y_i' - g(\mathbf{X}_i))\right],$$

and

$$\hat{\sigma}_n^2 \xrightarrow{p} \mathbb{E}\left[(Y_i - g(\mathbf{X}_i))^2(Y_i' - g(\mathbf{X}_i))^2\right].$$

Under the alternative,

$$\mathbb{E}\left[(Y_i - g(\mathbf{X}_i))(Y_i' - g(\mathbf{X}_i))\right] = \mathbb{E}\left[\mathbb{E}[(Y - g(\mathbf{X}))|\mathbf{X}]\right]^2 = M_g > 0,$$

and

$$\mathbb{E}\left[(Y_i - g(\mathbf{X}_i))^2(Y_i' - g(\mathbf{X}_i))^2\right] = \mathbb{E}\left[\mathbb{E}[(Y - g(\mathbf{X}))^2|\mathbf{X}]\right]^2 \geq \mathbb{E}\left\{\mathbb{E}[Y - g(\mathbf{X})|\mathbf{X}]\right\}^4$$

$$\geq \left(\mathbb{E}\left[\mathbb{E}[Y-g(\mathbf{X})|\mathbf{X}]\right]^2\right)^2 = M_g^2 > 0,$$

where the first inequality holds because $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$ and the second inequality holds due to Jensen's inequality. This completes the proof.

C PROOF OF THEOREM 4

Under the local alternatives, since we have $Y_i - g(\mathbf{X}_i) = Y_i - f(\mathbf{X}_i) + \delta_n \cdot e(\mathbf{X}_i)$, $\sqrt{n}V_n$ can be decomposed as

$$\begin{aligned}
\sqrt{n}V_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_i - f(\mathbf{X}_i))(Y_i' - f(\mathbf{X}_i)) + \delta_n \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e(\mathbf{X}_i)(Y_i - f(\mathbf{X}_i)) \\
&+ \delta_n \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e(\mathbf{X}_i)(Y_i' - f(\mathbf{X}_i)) + \sqrt{n}\delta_n^2 \frac{1}{n}\sum_{i=1}^{n} e^2(\mathbf{X}_i) \\
&= Q_{1n} + \delta_n \cdot Q_{2n} + \delta_n \cdot Q_{3n} + \sqrt{n}\delta_n^2 \cdot Q_{4n}.
\end{aligned}$$

For Q_{1n}, since

$$\mathbb{E}\left[(Y_i - f(\mathbf{X}_i))(Y_i' - f(\mathbf{X}_i))\right] = \mathbb{E}[\mathbb{E}[Y - f(\mathbf{X})|\mathbf{X}]]^2 = 0,$$

applying the central limit theorem it holds that Q_{1n} converges in distribution to a normal random variable with mean zero and variance

$$\operatorname{Var}\left[(Y_i - f(\mathbf{X}_i))(Y_i' - f(\mathbf{X}_i))\right] = \mathbb{E}\left[(Y_i - f(\mathbf{X}_i))(Y_i' - f(\mathbf{X}_i))\right]^2 = \sigma_1^2,$$

For Q_{2n} and Q_{3n}, since

$$\mathbb{E}\left[e(\mathbf{X}_i)(Y_i' - f(\mathbf{X}_i))\right] = \mathbb{E}\left[e(\mathbf{X}_i)(Y_i - f(\mathbf{X}_i))\right] = \mathbb{E}\left[e(\mathbf{X})(\mathbb{E}[Y|\mathbf{X}] - f(\mathbf{X}))\right] = 0,$$

applying the central limit theorem it holds that

$$Q_{2n} \xrightarrow{d} N\left(0, \mathbb{E}[e^2(\mathbf{X})\operatorname{Var}(Y|\mathbf{X})]\right) \qquad \text{and} \qquad Q_{3n} \xrightarrow{d} N\left(0, \mathbb{E}[e^2(\mathbf{X})\operatorname{Var}(Y|\mathbf{X})]\right).$$

Furthermore, by the weak law of large numbers, $Q_{4n} \xrightarrow{p} \mathbb{E}[e^2(\mathbf{X})]$. If $\delta_n = n^{-1/4}$, then

$$\delta_n \cdot Q_{2n} = n^{-1/4} \cdot Q_{2n} \xrightarrow{p} 0, \quad \delta_n \cdot Q_{3n} = n^{-1/4} \cdot Q_{3n} \xrightarrow{p} 0, \quad \sqrt{n}\delta_n^2 \cdot Q_{4n} = \sqrt{n} \cdot n^{-1/2} Q_{4n} \xrightarrow{p} \mathbb{E}[e^2(\mathbf{X})].$$

Hence we have

$$\sqrt{n}V_n \xrightarrow{p} N\left(\mathbb{E}[e^2(\mathbf{X})], \sigma_1^2\right),$$

and

$$T_n = \sqrt{n}V_n/\sigma_1 \xrightarrow{p} N\left(\mathbb{E}[e^2(\mathbf{X})]/\sigma_1, 1\right),$$

which completes the proof.

REFERENCES

Barton, R. R., and M. Meckesheimer. 2006. "Metamodel-based simulation optimization". *Handbooks in Operations Research and Management Science* 13:535–574.

Berger, J. O., and M. Delampady. 1987. "Testing Precise Hypotheses". *Statistical Science* 2 (3): 317–335.

Davidson, R., and J. G. MacKinnon. 1981. "Several tests for model specification in the presence of alternative hypotheses". *Econometrica* 49 (3): 781–93.

Dette, H. 1999. "A consistent test for the functional form of a regression based on a difference of variance estimators". *The Annals of Statistics* 27 (3): 1012–1040.
Eubank, R. L., and C. H. Spiegelman. 1990. "Testing the goodness of fit of a linear model via nonparametric regression techniques". *Journal of the American Statistical Association* 85 (410): 387–392.
Goda, T. 2017. "Computing the variance of a conditional expectation via non-nested Monte Carlo". *Operations Research Letters* 45 (1): 63–67.
González-Manteiga, W., and R. M. Crujeiras. 2013. "An updated review of Goodness-of-Fit tests for regression models". *Test* 22 (3): 361–411.
Guerre, E., and P. Lavergne. 2005. "Data-driven rate-optimal specification testing in regression models". *Annals of Statistics* 33 (2): 840–870.
Guo, X., T. Wang, and L. Zhu. 2016. "Model checking for parametric single-index models: a dimension reduction model-adaptive approach". *Journal of the Royal Statistical Society Series B* 78 (5): 1013–1035.
Hardle, W., and E. Mammen. 1993. "Comparing nonparametric versus parametric regression fits". *Annals of statistics* 21 (4): 1926–1947.
Hsiao, C., Q. Li, and J. S. Racine. 2007. "A consistent model specification test with mixed discrete and continuous data". *Journal of Econometrics* 140 (2): 802–826.
Kleijnen, J. P. 1986. *Statistical Tools for Simulation Practitioners*. Marcel Dekker, Inc.
Lavergne, P., and V. Patilea. 2008. "Breaking the curse of dimensionality in nonparametric testing". *Journal of Econometrics* 143 (1): 103–122.
Lee, B. 1988. *A nonparametric model specification test using a kernel regression method*. Ph. D. thesis, University of Wisconsin, Madison.
McManus, D. A. 1991. "Who invented local power analysis?". *Econometric Theory* 7 (2): 265–268.
Stone, C. J. et al. 1980. "Optimal Rates of Convergence for Nonparametric Estimators". *The Annals of Statistics* 8 (6): 1348–1360.
Wooldridge, J. M. 1992. "A test for functional form against nonparametric alternatives". *Econometric Theory* 8 (04): 452–475.
Xia, Y. 2009. "Model checking in regression via dimension reduction". *Biometrika* 96 (1): 133–148.
Yatchew, A. J. 1992. "Nonparametric regression tests based on least squares". *Econometric Theory* 8 (04): 435–451.
Zheng, J. X. 1996. "A consistent test of functional form via nonparametric estimation techniques". *Journal of Econometrics* 75 (2): 263–289.

AUTHOR BIOGRAPHIES

SHIYU WANG is a PhD candidate in the Department of Management Sciences, College of Business at City University of Hong Kong. Her research focuses on stochastic simulation and optimization, and their applications. Her e-mail address is shiyuwang.work@gmail.com.

GUANGWU LIU is an associate professor in the Department of Management Sciences, College of Business at City University of Hong Kong. His research interests include stochastic simulation, financial engineering, and risk management. His e-mail address is msgw.liu@cityu.edu.hk.

KUN ZHANG is a PhD candidate in the Department of Management Sciences, College of Business at City University of Hong Kong. His research focuses on stochastic simulation and risk management. His e-mail address is kkunzhang2013@gmail.com.