Departamento de Economía

Predictive Accuracy of Impulse Responses Estimated Using Local Projections and Vector Autoregressions

Autorías ditellianas: Sola, Martín; Yunis, Patricio
Otras autorías: Psadarakis, Zacharias (*University of London); Spagnol, Nicola (Brunel University)*

Predictive Accuracy of Impulse Responses Estimated Using Local Projections and Vector Autoregressions*

Zacharias Psaradakis†
Birkbeck, University of London, U.K.
z.psaradakis@bbk.ac.uk

Martin Sola
Universidad Torcuato di Tella, Argentina
msola@utdt.edu

Nicola Spagnolo
Brunel University London, U.K.
nicola.spagnolo@brunel.ac.uk

Patricio Yunis
Universidad Torcuato di Tella, Argentina
patriciojyunis@gmail.com

May 20, 2024

Abstract

We examine the small-sample accuracy of impulse responses obtained using local projections (LP) and vector autoregressive (VAR) models. In view of the fact that impulse responses are differences between multistep predictors, we propose to assess the relative performance of impulse-response estimators using tests for equal predictive accuracy. In our Monte Carlo experiments, LP-based and VAR-based estimators are found to be equally accurate in large samples under a mean squared error risk function. VAR-based estimators tend to have an advantage over LP-based ones in small and moderately sized samples, particularly at long horizons.

Keywords: Local projections; Predictive accuracy; VAR models.
JEL Classification: C32; C53.

*The authors thank Martin Gonzalez-Rozada, Constantino Hevia and Fabio Spagnolo for helpful comments and discussions.

†Corresponding author. E-mail: z.psaradakis@bbk.ac.uk; Address: Birkbeck Business School, Birkbeck, University of London, Malet Street, London WC1E 7HX, United Kingdom.

1 Introduction

Estimating impulse responses by the method of local projections (LP), as proposed by Jordà (2005), has become an increasingly popular alternative to the conventional approach based on vector autoregressive (VAR) models. For a wide range of empirically relevant data-generating processes (DGPs), the two approaches are known to agree approximately, both in population and in large samples, especially at impulse-response horizons of short length (see Plagborg-Møller and Wolf (2021)). Nevertheless, there is ongoing debate as to whether LP-based estimators fare better or worse than VAR-based estimators, in terms of bias and efficiency, in samples of relatively small or intermediate sizes like those that are often used in empirical work (see Brugnolini (2018), Herbst and Johannsen (2021) and Li et al. (2024), inter alia).

This paper contributes to the expanding literature on the relative merits of LP-based and VAR-based approaches to impulse-response estimation by considering finite-sample accuracy from a new perspective. Specifically, motivated by the observation that impulse responses are forecasts conditional on specific innovations, we propose to assess the relative performance of LP-based and VAR-based estimators by means of tests for equal predictive accuracy (cf. Diebold and Mariano (1995), Harvey et al. (1997)). This way of evaluating the accuracy of the two approaches, based on their relative performance with respect to a mean squared error risk function, allows us to draw conclusions about the statistical significance, or otherwise, of the differences between the estimation errors associated with the two competing methods of impulse-response estimation. The approach is, to our knowledge, novel and a useful addition to the more familiar bias/efficiency comparisons that are found in the literature.

By means of Monte Carlo experiments, we demonstrate that, under a finite-order VAR DGP, there is not much to choose between the two estimators of impulse responses in large samples in terms of predictive ability (evaluated with a quadratic loss function). By contrast, in small and moderately sized samples, VAR-based estimators outperform LP-based ones, particularly at long horizons where LP-based estimators of impulse responses tend to be significantly biased.

The organization of the paper is as follows. Section 2 introduces the impulse-response estimators of interest. Section 3 presents and discusses the results of a simulation study that compares LP-based and VAR-based estimators of impulse responses in terms of finite-sample bias and predictive accuracy. Section 4 summarizes and concludes.

2 VAR-Based and LP-Based Impulse-Response Estimators

Consider the first-order VAR model for the K-dimensional time series $y_t = (Y_{1t}, \ldots, Y_{Kt})'$ given by

$$y_t = \mu + \Phi y_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, T, \tag{1}$$

where μ is a K-dimensional vector of constants, Φ is a $K \times K$ parameter matrix, and $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Kt})'$ are K-dimensional random vectors such that $\mathbf{E}[\varepsilon_t] = 0$, $\mathbf{E}[\varepsilon_t \varepsilon_r'] = 0$ for $t \neq r$ and $\mathbf{E}[\varepsilon_t \varepsilon_t'] = \Sigma$, with Σ being a finite and positive definite matrix. Higher-order VAR models can be written in this form with appropriate redefinition of y_t, μ, Φ and ε_t.

For any nonnegative integer τ, the impulse response of y_t at horizon τ to a shock of size $\delta = (\delta_1, \ldots, \delta_K)'$ at date t may be defined in general as the difference between two conditional expectations of $y_{t+\tau}$, one which is conditional on the information set $\mathbb{Y}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$ and on a shock of size δ to ε_t and one which is conditional on $\mathbb{Y}_{t-1}$ alone (Koop et al. (1996)). More formally, the impulse response function of y_t to a shock of size δ is

$$R(\tau, \delta) = \mathbf{E}[y_{t+\tau} | \varepsilon_t = \delta, \mathbb{Y}_{t-1}] - \mathbf{E}[y_{t+\tau} | \varepsilon_t = 0, \mathbb{Y}_{t-1}], \quad \tau = 0, 1, 2, \ldots. \tag{2}$$

Hence, when the DGP is the model in (1) with parameter matrix Φ of spectral radius less than 1, it can be easily verified that

$$R(\tau, \delta) = \Phi^{\tau} \delta. \tag{3}$$

Evidently, $R(\tau, \delta)$ is invariant with respect to $\mathbb{Y}_{t-1}$ in this case and tends to the zero vector as τ tends to infinity.

Given any horizon τ, the conventional approach to estimating impulse responses (under a maintained first-order VAR model) amounts to replacing Φ in (3) with its ordinary least-squares estimator $\widehat{\Phi}$ to obtain the VAR-based estimator

$$\widehat{R}_{\mathrm{VAR}}(\tau, \delta) = \widehat{\Phi}^{\tau}\delta.$$

An alternative approach, put forward by Jordà (2005), is to rely on the linear projection model

$$y_{t+\tau} = c_\tau + B_\tau y_t + u_{\tau,t+\tau}, \quad t = 1, 2, \ldots, T - \tau, \tag{4}$$

where c_τ is a K-dimensional vector of constants, B_τ is a $K \times K$ parameter matrix, and $u_{\tau,t+\tau}$ is a K-dimensional projection error that is autocorrelated in general (for instance, $u_{\tau,t+\tau} = \sum_{k=1}^{\tau} \Phi^{\tau-k}\varepsilon_{t+k}$ when y_t satisfies (1) and Φ has spectral radius less than 1). The standard LP estimator of $R(\tau, \delta)$ is

$$\widehat{R}_{\mathrm{LP}}(\tau, \delta) = \widehat{B}_\tau \delta,$$

where $\widehat{B}_\tau$ is the ordinary least-squares estimator of B_τ in (4) (with $\widehat{B}_0$ equal to the identity matrix of order K). Note that, unless it is maintained that the DGP is the VAR model in (1), LP impulse response estimates may be obtained from linear projections of $y_{t+\tau}$ onto a constant and $(y_t, y_{t-1}, \ldots, y_{t-p})$ for some positive integer p that need not be the same for different horizons τ (in practice, p may be determined by means of conventional information-based criteria).

3 Bias and Predictive Accuracy of Impulse-Response Estimators

This section presents and discusses the results of Monte Carlo experiments designed to shed light on the comparative performance of VAR-based and LP-based impulse responses in terms of bias and predictive accuracy. In order to focus on these two aspects of impulse responses, we abstract from issues associated with structural identification of impulse responses (by assuming that the covariance matrix of the Wold innovations is diagonal), as well as from issues related to model misspecification (by assuming that VAR-based impulse responses are obtained from a correctly specified model).

The DGP used in the experiments is a bivariate version ($K = 2$) of the VAR model in (1), with $\mu = (0.4, 0.4)'$ and ε_t being normally distributed with identity covariance matrix. We consider two stable parameter configurations:

$$\text{DGP-1: } \Phi = \begin{bmatrix} 0.49 & 0.36 \\ 0.36 & 0.49 \end{bmatrix}; \qquad \text{DGP-2: } \Phi = \begin{bmatrix} 0.80 & 0.10 \\ 0.75 & 0.40 \end{bmatrix}.$$

Letting $e_1 = (1, 0)'$ and $e_2 = (0, 1)'$, the objects of interest are the horizon-τ impulse responses $R_i(\tau, e_j) = e_i' R(\tau, e_j) = e_i' \Phi^\tau e_j$ $(i, j = 1, 2)$ and their VAR-based and LP-based estimators $\widehat{R}_{\text{VAR},i}(\tau, e_j) = e_i' \widehat{R}_{\text{VAR}}(\tau, e_j)$ and $\widehat{R}_{\text{LP},i}(\tau, e_j) = e_i' \widehat{R}_{\text{LP}}(\tau, e_j)$, respectively. All reported simulation results are obtained from 1000 artificial samples.

3.1 Bias

Monte Carlo estimates of the bias of VAR-based and LP-based estimators of impulse responses at each horizon $\tau = 1, 2, \ldots, 15$ are shown in Figure 1 for sample sizes $T = 100, 400, 1600$. The clear pattern that emerges from these plots is that the finite-sample bias of both estimators is a decreasing function of the sample size and that the VAR-based estimator has an advantage at all horizons. A notable difference between the two estimators is that the bias of the VAR-based estimators of impulse responses tends to decrease in general (albeit not necessarily monotonically) as τ increases. By contrast, the bias of LP-based estimators has a tendency to increase as the horizon length τ becomes longer. In the next sub-section we consider whether the differences in the observed estimation errors committed by the two approaches are significant enough for one procedure to be considered to outperform the other.

3.2 Predictive Accuracy

Since horizon-τ impulse responses are in essence differences between τ-step-ahead predictors, the relative accuracy of VAR-based and LP-based estimators may be assessed by comparing their predictive accuracy using the methodology of Diebold and Mariano (1995), or its small-sample modification discussed in Harvey et al. (1997). Specifically, let

$$d_{ij}(\tau) = \{R_i(\tau, e_j) - \widehat{R}_{\text{LP},i}(\tau, e_j)\}^2 - \{R_i(\tau, e_j) - \widehat{R}_{\text{VAR},i}(\tau, e_j)\}^2, \quad i, j = 1, 2,$$

be the difference between the squared errors associated with the LP-based and VAR-based estimators of the horizon-τ impulse response of Y_{it} to ε_{jt}. The hypothesis of equal accuracy, with respect to the mean squared error risk function, may be formulated as $\mathbf{E}[d_{ij}(\tau)] = 0$, whereas $\mathbf{E}[d_{ij}(\tau)] > 0$ implies that VAR-based impulse responses are more accurate than LP-based ones.

Following Harvey et al. (1997), when N horizon-τ differentials $d_{ij,s}(\tau)$, $s = 1, 2, \ldots, N$ ($N \geq \tau$), are available, with arithmetic mean $\bar{d}_{ij}(\tau) = N^{-1} \sum_{s=1}^{N} d_{ij,s}(\tau)$, a test for equal accuracy may be based on the statistic

$$\mathcal{D}_{ij}(\tau) = \left\{ \frac{N + 1 - 2\tau + N^{-1}\tau(\tau - 1)}{\widehat{\omega}_{ij}(\tau)} \right\}^{1/2} \bar{d}_{ij}(\tau), \tag{5}$$

where $\widehat{\omega}_{ij}(\tau)$ is the estimator of the asymptotic variance of $N^{1/2}\bar{d}_{ij}(\tau)$ given by

$$\widehat{\omega}_{ij}(\tau) = N^{-1} \sum_{s=1}^{N} \{d_{ij,s}(\tau) - \bar{d}_{ij}(\tau)\}^2 + 2N^{-1} \sum_{k=1}^{\tau-1} \sum_{s=k+1}^{N} \{d_{ij,s}(\tau) - \bar{d}_{ij}(\tau)\}\{d_{ij,s-k}(\tau) - \bar{d}_{ij}(\tau)\}.$$

The null hypothesis of equal accuracy is rejected at level α, in favour of the one-sided alternative that VAR-based responses are superior, if $\mathcal{D}_{ij}(\tau)$ exceeds the $(1-\alpha)$th quantile of the Student-t distribution with $N - 1$ degrees of freedom.

The differentials $d_{ij,s}(\tau)$, $s = 1, \ldots, N$, are constructed in our analysis using a recursive scheme. More specifically, for each artificial time series of length $T = 100, 200, 400, 800, 1600$ from DGP-1 or DGP-2 and each horizon $\tau = 1, 2, \ldots, 15$, impulse responses $\widehat{R}_{\mathrm{VAR},i}(\tau, e_j)$ and $\widehat{R}_{\mathrm{LP},i}(\tau, e_j)$ are computed using the first $T_0 = 60$ data points to obtain $\widehat{\Phi}$ and $\widehat{B}_{\tau}$. This is then repeated for the first $T_0+1, T_0+2, \ldots, T$ data points to obtain $N = T - T_0 + 1$ differentials $d_{ij,s}(\tau)$ in total.

Tables 1 and 2 show the percentage of artificial samples in which VAR-based impulse responses are found to be more accurate than LP-based responses according to a 5%-level test based on the statistic in (5). The results suggest that the use of VAR-based impulse responses is preferable in samples comprising 200 observations or less, especially so at long horizons; at such horizons, VAR-based responses have a clear advantage over LP-based responses under both DGPs. For medium-sized samples of 400 observations, there is not much to choose between the two competing estimators, although VAR-based impulse responses tend to have a slight advantage the longer the impulse horizon is. Finally, for samples of 800 observations or more, the hypothesis of equal accuracy is very rarely rejected.

Simulation results (not shown) for a 5%-level test of the hypothesis of equal accuracy ($\mathbf{E}[d_{ij}(\tau)] = 0$) against the one-sided alternative that VAR-based impulse responses are less accurate than LP-based responses ($\mathbf{E}[d_{ij}(\tau)] < 0$) reveal that the null hypothesis is very rarely rejected. The highest rejection rates (ranging between 8% and 13%) were obtained for $R_1(\tau, e_1)$ and $R_2(\tau, e_1)$ under DGP-2, when $\tau \geq 2$ and $T = 100$.

We note that these results based on predictive accuracy tests are generally in agreement with the findings in the simulation exercises of Brugnolini (2018) and Herbst and Johannsen (2021), who also report that VAR-based estimators of impulse responses have superior performance when using well-specified models that contain the DGP. For DGPs that do not admit a finite-order VAR representation, the results in Li et al. (2024) suggest that while LP-based and VAR-based estimators tend to perform similarly at short horizons, they behave differently at intermediate and long horizons, with VAR-based estimators being a more attractive choice in terms of mean squared error loss.

4 Conclusion

This paper has investigated the comparative performance of VAR-based and LP-based estimators of impulse responses in terms of finite-sample predictive accuracy. Motivated by the observation that impulse responses are differences between optimal predictors, we have examined whether VAR-based and LP-based estimators of impulse responses have equal predictive ability (with respect to the mean squared error risk function) using a modification of the Diebold–Mariano approach. This is a novel way of assessing the statistical significance of the differences in the estimation errors associated with the two competing approaches. The results of a simulation study have shown that there is not much to choose between the two estimators of impulse responses in large samples. However, in small and moderately sizes samples similar to those commonly used in applied macroeconomics, VAR-based estimators have an advantage over LP-based ones, particularly at long horizons where LP-based estimators tend to be significantly biased.

Obvious potential extensions of the approach put forward in this paper involve exploration of the properties of VAR-based and LP-based estimators, and of bias-corrected variants of them, under more general DGPs, different structural identification schemes, and potential VAR model misspecification.

References

Brugnolini, L. (2018). About local projection impulse response function reliability. CEIS Working Paper No. 440, University of Rome Tor Vergata.

Diebold, F. X. and R. S. Mariano (1995). Testing the equality of prediction mean squared errors. *Journal of Business & Economic Statistics 13*(3), 253–263.

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting 13*(2), 281–291.

Herbst, E. and B. K. Johannsen (2021). Bias in local projections. FEDS Working Paper No. 2020-010R1, Board of Governors of the Federal Reserve System.

Jordà, O. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review 95*(1), 161–182.

Koop, G., H. M. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics 74*(1), 119–147.

Li, D., M. Plagborg-Møller, and C. K. Wolf (2024). Local projections vs. VARs: Lessons from thousands of DGPs. *Journal of Econometrics*.

Plagborg-Møller, M. and C. K. Wolf (2021). Local projections and VARs estimate the same impulse responses. *Econometrica 89*(2), 955–980.

Figure 1: Simulation-Estimated Bias of LP-Based and VAR-Based Estimators

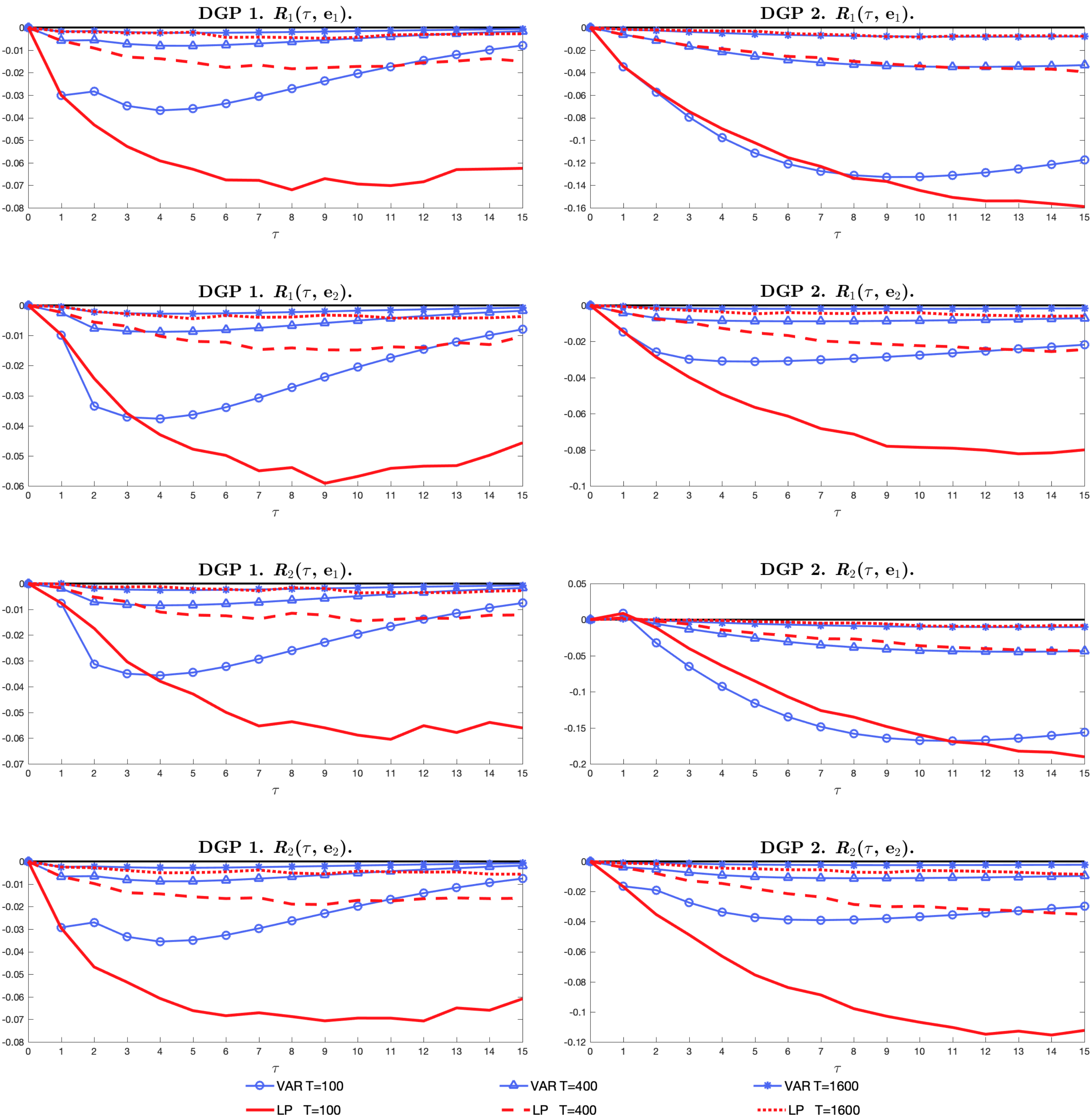

Table 1: Percentage Rejections of the Hypothesis of Equal Accuracy Under DGP-1

$R_1(\tau, e_1)$

T \ τ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
100	0.0	24.5	26.5	30.1	32.5	34.9	38.3	39.2	42.5	46.7	51.4	54.2	60.5	66.1	69.6
200	0.0	12.1	12.4	15.0	14.6	16.3	17.8	17.9	19.9	22.7	24.6	26.9	29.3	32.6	34.5
400	0.0	5.7	5.2	6.8	6.8	6.5	7.0	6.6	7.9	9.4	10.9	13.9	13.6	14.0	17.4
800	0.0	2.0	2.4	3.3	3.3	1.9	3.7	4.0	5.7	4.9	6.0	6.7	6.2	6.7	7.5
1600	0.0	1.0	1.0	0.7	0.8	1.0	1.7	2.2	2.2	2.6	2.6	2.2	2.1	2.9	3.1

$R_1(\tau, e_2)$

T \ τ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
100	0.0	20.1	22.7	24.5	27.8	29.5	36.0	38.9	42.4	45.4	48.4	55.1	61.8	65.4	70.9
200	0.0	8.9	9.6	11.4	13.4	14.5	16.1	18.2	19.8	23.1	26.3	30.2	32.0	31.7	36.9
400	0.0	4.1	5.2	5.0	4.9	6.1	7.2	7.8	9.5	9.2	11.3	11.3	12.8	14.0	16.2
800	0.0	1.9	2.0	2.3	2.3	2.7	3.1	3.4	3.0	4.5	4.2	4.3	5.4	5.6	6.5
1600	0.0	0.7	1.2	0.8	1.3	1.5	2.2	1.0	1.5	1.6	2.2	2.2	3.2	3.2	3.3

$R_2(\tau, e_1)$

T \ τ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
100	0.0	22.0	25.3	25.5	27.6	33.5	36.2	38.9	40.4	47.5	52.4	57.2	60.9	67.3	70.0
200	0.0	8.4	11.4	12.4	13.8	16.2	16.7	18.6	18.2	21.6	24.5	27.6	29.1	32.4	34.9
400	0.0	4.5	6.4	5.2	5.3	6.1	5.6	7.1	7.9	10.3	12.4	11.8	14.5	14.2	16.1
800	0.0	2.2	2.8	2.4	2.8	3.0	3.8	3.8	5.0	4.8	6.1	6.1	5.7	6.2	6.9
1600	0.0	0.7	0.8	1.2	0.8	1.4	1.6	1.6	2.3	1.9	2.0	2.2	2.8	2.3	3.4

$R_2(\tau, e_2)$

T \ τ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
100	0.0	25.1	25.0	25.3	30.8	34.3	36.5	39.9	43.2	46.0	49.6	54.3	61.6	64.5	69.4
200	0.0	12.3	11.4	13.6	14.6	17.0	18.3	17.9	18.8	23.1	27.7	29.0	31.7	32.2	35.8
400	0.0	4.9	5.1	5.2	6.1	6.1	6.7	7.9	8.9	10.3	12.3	12.6	13.3	13.5	15.4
800	0.0	2.3	2.3	2.3	2.7	2.8	3.8	3.5	3.7	4.1	4.8	4.2	5.9	6.0	6.7
1600	0.0	0.5	0.4	1.6	0.9	1.4	1.1	2.0	1.6	2.1	2.8	2.5	2.7	2.8	2.1

.

Table 2: Percentage Rejections of the Hypothesis of Equal Accuracy Under DGP-2

$R_1(\tau, e_1)$

T \ τ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
100	0.0	21.0	22.1	21.9	25.0	27.4	29.1	30.8	32.2	35.6	39.4	42.3	45.6	49.6	51.5
200	0.0	7.8	9.4	10.7	10.2	10.6	10.8	11.8	11.4	13.4	15.2	17.3	19.2	21.5	23.7
400	0.0	3.1	3.9	4.2	4.3	3.4	3.7	4.4	4.4	5.6	5.2	6.6	7.2	8.3	8.9
800	0.0	1.7	1.1	1.8	1.2	1.1	1.4	1.4	1.4	1.6	1.4	1.5	2.2	2.0	2.7
1600	0.0	0.6	0.7	0.5	0.5	0.4	0.4	0.4	0.5	0.4	0.4	0.3	0.2	0.4	0.5

$R_1(\tau, e_2)$

T \ τ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
100	0.0	19.4	22.3	24.1	26.7	29.8	34.4	38.5	42.5	48.0	50.4	55.4	61.1	63.2	68.9
200	0.0	8.7	10.7	12.2	12.1	14.0	15.3	16.7	18.6	21.4	23.3	27.3	28.9	30.7	34.3
400	0.0	3.5	3.7	4.9	4.6	4.9	6.5	6.4	7.4	7.0	8.4	10.0	10.5	12.0	13.3
800	0.0	2.5	1.3	1.9	2.0	2.3	3.3	2.8	3.2	3.2	3.3	3.4	3.4	3.9	4.3
1600	0.0	1.2	0.6	0.9	1.5	1.5	1.8	1.6	1.8	2.1	2.0	1.8	1.4	2.0	2.5

$R_2(\tau, e_1)$

T \ τ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
100	0.0	17.9	17.0	18.4	21.1	22.1	24.6	26.5	29.3	32.8	33.9	39.9	42.9	43.9	49.5
200	0.0	8.1	7.5	7.9	9.0	8.2	9.4	8.9	10.5	12.4	13.2	15.7	17.6	16.9	20.3
400	0.0	3.4	3.5	3.8	3.6	2.6	2.4	3.0	3.9	5.0	5.9	6.2	7.2	7.9	7.3
800	0.0	1.8	2.1	1.4	1.8	1.3	1.3	1.3	1.8	1.3	1.1	1.6	2.1	2.3	2.2
1600	0.0	0.3	0.6	0.7	0.6	0.6	0.4	0.4	0.9	0.7	0.6	0.6	0.6	0.7	0.2

$R_2(\tau, e_2)$

T \ τ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
100	0.0	23.7	23.0	25.2	27.7	31.2	34.4	38.3	39.6	47.2	49.9	55.2	59.9	63.9	66.8
200	0.0	10.4	11.4	12.5	13.9	13.9	16.2	16.5	17.0	20.1	23.8	26.7	30.4	31.2	34.6
400	0.0	3.6	5.5	5.2	5.1	5.7	6.6	6.9	6.4	8.3	7.6	10.0	10.7	13.4	13.6
800	0.0	2.0	2.3	2.0	2.3	2.0	2.7	2.6	3.5	3.9	3.5	4.1	3.5	3.9	4.5
1600	0.0	0.5	0.8	1.7	1.4	1.3	1.5	1.3	1.5	1.7	1.4	2.1	1.0	1.7	1.7