


Research Article

Comparison of Weighted Lag Adaptive LASSO with Autometrics for Covariate Selection and Forecasting Using Time-Series Data

Sara Muhammadullah ¹, Amena Urooj,¹ Faridoon Khan,¹ Mohammed N Alshahrani,² Mohammed Alqawba,³ and Sanaa Al-Marzouki⁴

¹Department of Economics and Econometrics, Pakistan Institute of Development Economics, Islamabad, Pakistan

²Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam University, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia

³Department of Mathematics, College of Science and Arts, Qassim University, Ar Rass, Saudi Arabia

⁴Statistics Department, Faculty of Sciences, King AbdulAziz University, Jeddah 21551, Saudi Arabia

Correspondence should be addressed to Sara Muhammadullah; saramuhammadullah@gmail.com

Received 12 November 2021; Revised 11 December 2021; Accepted 23 December 2021; Published 11 January 2022

Academic Editor: Peican Zhu

Copyright © 2022 Sara Muhammadullah et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to reduce the dimensionality of parameter space and enhance out-of-sample forecasting performance, this research compares regularization techniques with Autometrics in time-series modeling. We mainly focus on comparing weighted lag adaptive LASSO (WLAdaLASSO) with Autometrics, but as a benchmark, we estimate other popular regularization methods LASSO, AdaLASSO, SCAD, and MCP. For analytical comparison, we implement Monte Carlo simulation and assess the performance of these techniques in terms of out-of-sample Root Mean Square Error, Gauge, and Potency. The comparison is assessed with varying autocorrelation coefficients and sample sizes. The simulation experiment indicates that, compared to Autometrics and other regularization approaches, the WLAdaLASSO outperforms the others in covariate selection and forecasting, especially when there is a greater linear dependency between predictors. In contrast, the computational efficiency of Autometrics decreases with a strong linear dependency between predictors. However, under the large sample and weak linear dependency between predictors, the Autometrics potency $\rightarrow 1$ and gauge $\rightarrow \alpha$. In contrast, LASSO, AdaLASSO, SCAD, and MCP select more covariates and possess higher RMSE than Autometrics and WLAdaLASSO. To compare the considered techniques, we made the Generalized Unidentified Model for covariate selection and out-of-sample forecasting for the trade balance of Pakistan. We train the model on 1985–2015 observations and 2016–2020 observations as test data for the out-of-sample forecast.

1. Introduction

Since the beginning of time-series analysis, modeling and forecasting have been the center of attraction. The accuracy of the model in time-series analysis is always unknown. Only one in a million models can be accurate; “essentially, all models are wrong, but some are useful” [1]. However, the massive availability of data in the current era leads us to a new phase of time-series analysis for model selection and forecasting. Including many financial and economic

covariates in the time-series model for superior prediction may yield considerable benefits. However, parsimonious models in time-series analysis are superior in forecasting. Failure to decrease dimensionality may lead to poor performance due to cumulative estimation losses from redundant or insignificant variables.

On the other hand, the traditional time-series modeling for covariates and lag selection in Autoregressive Distributed Lag (ARDL) modeling uses Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) [2, 3]. This

technique is limited to the number of covariates, and their lag must not be greater than the number of observations. The traditional Ordinary Least Square method fails to estimate the forthcoming models with huge regressors and limited observations due to inadequate degrees of freedom. Several statistical techniques exist in the literature for model selection and forecasting when covariates and their lags are more than the number of observations. Meanwhile, classical approach (Autometrics, general-to-specific) and regularization techniques (Machine Learning) are frequently used in time-series modeling when covariates exceed the number of observations. Besides these techniques, complex network theories provide an efficient and reliable solution for handling time-series issues. In recent years, the complex network has been extensively used in socioeconomic phenomena [4–6]. However, this study aimed to identify the true covariate and evaluate the model’s forecasting performance, and we only concentrated on regularization techniques and the classical approach.

The use of sparse modeling has grown widely in time-series analysis as it can efficiently handle big macroeconomic data sets and substitute the factor models [7–13]. For the time being, Medeiros & Mendes [14] find that the adaptive LASSO (AdaLASSO) consistently chooses the essential covariates as the number of observations grows (model selection consistency) even when the errors are non-Gaussian and conditionally heteroscedastic. Audrino and Camponovo [15] illustrate the theoretical and empirical efficiency of AdaLASSO, as it asymptotically selects covariate with finite-sample in time-series regression models. Covariates and their lag selections are challenging in time-series modeling, mainly when there is a mixture of serial correlation [16]. To probe this gap, Konzen and Ziegelmann [17] introduce Weighted Lag adaptive LASSO (WLAdaLASSO), which applies various weights to each coefficient and penalizes coefficients of higher-lagged variables. The WLAdaLASSO outperforms LASSO and AdaLASSO in forecasting and covariate selection, even in a greater linear dependency between predictors with many candidate lags, whereas Uematsu and Tanaka [18] use folded concave penalties for ultra-high-dimensional time-series forecasting and covariate selection. They verify the oracle inequalities of folded concave penalties (SCAD and MCP) for macroeconomic time series under appropriate conditions with the theoretical and empirical contribution.

In the meantime, very few studies exist that utilize the classical technique (Autometrics) in the context of macroeconomic forecasting [19–22]. In cross-sectional modeling, Epprecht et al. [23] compare the LASSO and AdaLASSO estimate with classical technique (Autometrics) in forecasting and covariate selection. The result indicates that LASSO and AdaLASSO estimates outperform Autometrics in prediction. Conversely, for time-series modeling with dynamic structure, WLAdaLASSO outperforms forecasting and covariate selection than LASSO and AdaLASSO. However, we have not come across a work that has compared the computational efficiency of regularization techniques, particularly WLAdaLASSO, SCAD, and MCP, with classical technique (Autometrics) in dynamic time-series

modeling. For this purpose, we implement an updated regularization technique for dynamic time-series modeling to assess their performance with the classical approach (Autometrics) for covariate selection and forecasting theoretically and empirically. Furthermore, we assess the efficiency of these techniques in simulation experiments where the real data generating process (DGP) has a dynamic structure. To summarize the entire discussion, our main contribution is a comparison of WLAdaLASSO and Autometrics for covariate selection and forecasting under different scenarios with various autocorrelation coefficients (0.1, 0.5, and 0.8) of regressors and T sample sizes (50, 100, and 500), as well as application to macroeconomic data to provide a conclusive solution to predictability. The computational efficiency of these techniques is assessed in terms of gauge, potency, and out-of-sample Root Mean Square Error (RMSE). We constructed a Generalized Unidentified Model (GUM) and considered all the possible macroeconomic determinants of the trade balance for real data analysis. The techniques are not restricted to the balance of trade but valid for any time series.

The rest of the paper is organized as follows: in Section 2, we will briefly illustrate the model selection techniques. Section 3 is based on the simulation experiment and results. Section 4 discusses real data analysis. Finally, Section 5 presents the conclusion and remarks on the efficacy of considered techniques.

2. Model Selection Techniques

Technically, two broad spectrums of model selection exist in the literature: regularization techniques and classical approach (Autometrics, general-to-specific), whenever P regressors are greater than N number of observations. The classical approach (Autometrics, general-to-specific) starts with a fully saturated model and uses a backward elimination with the multipath search process, and the selection of the model mainly depends on the predefined significance level. However, the regularization technique applies the sparsity on the p -dimensional parameter vector, which forces many of its components to be zero. This technique combats the issues posed by high-dimensionality. We describe each of these techniques in more detail, but we only considered orthogonal regularization techniques.

2.1. Autometrics Algorithms for Covariate and Lag Selection. Autometrics is a third-generation algorithm created on similar concepts of PcGets. Hoover et al. [24] proposed the general-to-specific model selection technique that aggregates many elements of the “Hendry” methodologies and “London School of Economics (LSE).” Doornik [25] proposed PcGets as a second-generation method extended by Krolzig and Hendry [26], prolonging and enlightening Hoover and Perez’s algorithm [26, 27]. The concept of general-to-specific (gets) modeling is the cornerstone of the Autometrics approach:

- (i) Initially, the GUM includes the overall covariates and estimates it by the OLS method with expelling

statistically irrelevant covariates; the reliability of the reduced model is confirmed at each stage to prove the congruence with diagnostic tests.

- (ii) Autometrics uses a tree-path search with multistep simplifications along numerous paths. Final models are calculated using a tree-path search and confirmed using diagnostic tests; if the coefficient estimates are statistically insignificant, the model is discarded. When a large number of terminal models are identified, Autometrics retests their union. A new GUM is created when the “surviving” terminal models are combined, allowing for one more tree-path search repeat. The entire exploration process is repeated, with the terminal models and their combinations being examined once again. If many models pass the encompassing tests, the final choice is based on predetermined information criteria.

Diagnostic tests are used to double-check the simplified models, while comprehensive tests resolve numerous terminal models. For diagnostic tests, Autometrics uses Jarque and Bera [28] residual normality test, Breusch and Pagan [29], and Godfrey [30] second-order residual autocorrelation, autocorrelated conditional heteroscedasticity (ARCH) to second-order [31], and in-sample stability [32]. In some aspects, Autometrics is a partially black box [23]. However, it allows the user to choose between “nominal significance level” and “1-cut and tight significance level” when establishing modeling approaches. The multipath approach avoids path dependency by using a tree structure and a similar stepwise backward elimination, a built-in function of the gets package in R environments [33].

2.2. Regularization Techniques. Regularization techniques handle saturated models with irrelevant regressors even if regressors are more than the number of observations and shrink the irrelevant coefficients equal to zero with some bias. Several regularization techniques exist in the literature; we only opt orthogonal regularization techniques for dynamic covariate selection and forecasting: LASSO, AdaLASSO, WLAdaLASSO, SCAD, and MCP.

2.2.1. LASSO and AdaLASSO Estimate. Due to lower computation cost, the Least Absolute Shrinkage and Selection Operator (LASSO) is a popular estimation method in a linear regression framework, introduced by Tibshirani (1996). The LASSO method is like ridge regression; however, it set some coefficients precisely equal to zero with a substantial bias. The resulting model is easy to interpret and possesses the least forecast error. Consider a linear regression model, where $y = (y_{1t}, y_{2t}, \dots, y_{nt})$ are the continuous response regressors, and $x_{it} = (x_{1t}, x_{1t-1}, \dots, x_{pt-1})$ are the covariates with its lag, and γ_j are the estimated coefficients. The equation can be defined as

$$\hat{\gamma}_j = \operatorname{argmin}_{\gamma} \left\| y_{it} - \sum_{i=1}^p \gamma_{it} x_{it} \right\|_2^2 + \sum_{i=1}^p p_{\lambda_j}(|\gamma_j|), \quad (1)$$

where $p_{\lambda_j}(\cdot)$ is a penalty function and λ_j is the hyperparameter.

The second term in the above equation is defined as “L1 penalty,” and λ leads to a sparse solution with a shrinking specific set the coefficients precisely equal to zero with a certain amount of bias. The amount of shrinkage depends upon the selection of λ , whereas its range is $0 < \lambda < \infty$.

Zou [34] demonstrated that the LASSO estimator lacks the oracle characteristic and introduced the adaptive LASSO, a simple and effective solution. In contrast, the coefficients in LASSO are all penalized equally in the L_1 penalty. However, in AdaLASSO, each coefficient is given a distinct weight. Zou [34] showed that the AdaLASSO could have the best results if the weights are data-dependent and carefully chosen; then, the AdaLASSO can possess the oracle property.

$$\hat{\gamma}^{\text{adaLASSO}} = \operatorname{argmin}_{\gamma} \left\| y_{it} - \sum_{i=1}^j \gamma_j x_{it} \right\|_2^2 + \lambda \sum_{i=1}^j \hat{w}_j |\gamma_j|. \quad (2)$$

$\hat{w}_j = 1/\hat{\gamma}_j^{*\tau}$, $\tau > 0$, and $\hat{\gamma}_j^*$ is an initial parameter estimate. The weights for zero coefficients diverge (to infinity) as the sample size expands, and nonzero coefficients converge to a finite constant. To estimate the $\hat{\gamma}_j^*$, Zou [34] recommended the OLS method. However, when the number of candidate variables exceeds the number of observations, the OLS method does not work. A ridge estimate can be employed as an initial estimator in this case.

2.2.2. Weighted Lag Adaptive LASSO (WLAdaLASSO). The Weighted Lag Adaptive LASSO (WLAdaLASSO) was introduced by Konzen and Ziegelmann [17] and established on the concept of Park and Sakaori [35] work. It is defined as another type of LASSO estimate specifically for time-series modeling with lag structure. The idea is similar to AdaLASSO and built for the time-series ARDL framework, as the more distant lags have a more negligible effect in predicting the dependent variable, imposing more enormous penalties on them.

$$\hat{\gamma}^{\text{adaLASSO}} = \operatorname{argmin}_{\gamma} \left\| y_{it} - \sum_{i=1}^j \gamma_j x_{it} \right\|_2^2 + \lambda \sum_{i=1}^j \hat{w}_j |\gamma_j|. \quad (3)$$

Here, $\hat{w}_j = (\hat{\gamma}_j^{\text{bridge}} |e^{-al}|)^{-\tau}$, l is the lag length, $\tau > 0$, and $\alpha \geq 0$ are tuning parameters. Moreover, $\hat{\gamma}_j^*$ is an initial parameter estimate. $\tau=1$ like in AdaLASSO. To pick α , Konzen and Ziegelmann (2016) suggest estimating the model for a given λ using a grid (0; 0.5; 1; :::; 10) and choose the one with the lowest BIC and the λ parameter selected on the same criteria of the lowest BIC.

2.2.3. SCAD and MCP Estimate. Smoothly Clipped Absolute Deviation is unbiased and sparse (i.e., small estimated coefficients automatically set to zero) and fulfills the condition of continuity proposed by Fan and Li [36]. The smoothly clipped absolute deviation (SCAD) for covariate selection and its lags are defined as

$$\hat{\gamma}^{SCAD} = \operatorname{argmax}_{\gamma} \left\| y_{it} - \sum_{i=1}^j \gamma_j x_{it} \right\|_2^2 + \sum_{j=1}^d p_j(|\gamma_j|; \lambda; \alpha),$$

$$(\text{SCAD}) p_{\lambda}(|\gamma_j|) = \lambda \begin{cases} |\gamma|, & \text{if } |\gamma| \leq \lambda, \\ -\frac{(\gamma^2 - 2a\lambda|\gamma| + \lambda^2)}{2(a+1)\lambda}, & \text{if } \lambda < |\gamma| \leq a\lambda \text{ and,} \\ \frac{1}{2}(a+1)\lambda, & \text{if } |\gamma| \geq a\lambda. \end{cases} \quad (4)$$

where x is the matrix of covariates and its lag, and the second term in the above equation is $\sum_{j=1}^d p_j(|\gamma_j|; \lambda; \alpha)$, that is, a penalized term designed to meet all three requirements (unbiasedness, sparsity, and continuity). The SCAD has proven to be effective in many statistical circumstances, such as cross-sectional regression and time-series modeling [18]. $P(\gamma|\lambda, \alpha)$ is a folded concave penalty; unlike LASSO, it depends on two tuning parameters, and penalties depend on λ in a nonmultiplicative way, so that $P(\alpha\lambda) = \lambda P(\alpha)$. Additionally, the tuning parameter α controls the concavity of the penalty. The maximization of the objective function depends on α and λ , whereas α equals 3.7 and λ is selected via cross-validation [36].

Zhang [37] introduced the Minimax Concave Penalty (MCP), a nonconvex penalization strategy that employs sparse area up to a particular variable selection threshold, resulting in an unbiased estimate.

$$(\text{MCP}) p_{\lambda}(|\gamma_j|) = \lambda \begin{cases} \left(\lambda - \frac{|\gamma|}{\alpha} \operatorname{sign}(\gamma) \right), & \text{if } |\gamma| \leq \alpha\lambda, \\ 0, & \text{if } |\gamma| > \alpha\lambda. \end{cases} \quad (5)$$

MCP uses $\sum_{j=1}^d p_j(|\gamma_j|; \lambda; \alpha)$ regularization path based on the family of nonconvex penalty function with two tuning parameters α and λ , where α is fixed, and λ is selected via cross-validation. The tuning parameter λ controls the amount of shrinkage and α concavity of penalty. MCP prevents the spares convexity to a greater extent due to minimizing the maximum concavity [37]. The regularization parameter tends to larger α coefficient affords less unbiased and more convexity [37]. SCAD and MCP estimates belong to a family of folded concave penalties, as $P(\cdot)$ penalty function is neither convex nor concave.

2.3. Selection of Tuning Parameters for Regularization Techniques. The selection of λ tuning parameter is crucial as it governs the complexity of the selected model. The choice of the optimal tuning parameter provides a parsimonious model with a precise prediction performance. In practice, the tuning parameter is frequently selected using a cross-validation approach to achieve prediction optimality. Such prediction optimality is frequently at odds with covariates

selection; however, the objective is to recover the underlying set of sparse variables: frequently, a bigger penalty parameter is required for covariate selection than the optimal prediction [38]. However, the BIC criterion is superior to cross-validation for covariate selection, but it has no theoretical explanation. Meanwhile, WLAdaLASSO with BIC-based tuning parameter outperforms others in covariate selection and out-of-sample forecast [17]. Hence, we use BIC-based tuning parameters for covariate selection and out-of-sample forecast in simulation exercises and real data analysis.

2.4. Theoretical Comparison. To compare these techniques, we use Gauge, Potency, and out-of-sample RMSE. Gauge is the empirical null retention frequency of how irrelevant covariates are retained, whereas potency is known as correct covariate identifications. The comparison of regularization techniques and Autometrics assessed via a correct zero identification interpreted as potency, and incorrect zero identification is referred to as Gauge [39]. We use RMSE for out-of-sample forecasting to evaluate the performance of concerned techniques in a simulation study and real data analysis. If the approaches correctly identify the accurate model, the estimations of the following parameters should be expected:

- (1) Gauge approaches to nominal significance level α or tight significance level (0.01 or 0.001).

$$E\left(\frac{\widehat{k}_{\text{rel}}}{k_{\text{rel}}}\right) \longrightarrow \alpha. \quad (6)$$

- (2) Potency approaches 1 if considered estimation techniques efficiently estimate the accurate model.

$$E\left(\frac{\widehat{k}_{\text{rel}}}{k_{\text{rel}}}\right) \longrightarrow 1. \quad (7)$$

3. Simulation Experiments and Results

The simulation study has been performed in R -free statistical software; for Autometrics, we used the gets package of R , which is freely available, and for regularization techniques, we use the glmnet for LASSO, AdaLASSO, and WLAdaLASSO as for SCAD and MCP, the ncvmreg package. The performance of Autometrics for covariate selection and forecasting is assessed with two levels of significance 0.05 and 0.01.

3.1. Data Generating Process (DGP). We use Konzen and Ziegelmann [17] DGP for statistical comparison. Regarding covariate and lag selection performance, our goal is to compare the gauge (“size”) and potency (“power”) of the model. We also emphasize the out-of-sample forecasting performance of considered techniques. To illustrate our purpose, we chose Konzen and Ziegelmann [17] DGP with 10 independent time-series covariates that follow AR(1) as $x_{i,t} = \phi x_{i,t-1} + \mu_{i,t}$ where $\mu_{i,t} \sim N(0, 1)$ and $i = 1, 2, \dots, 10$. We assess the performance of considered techniques under different scenarios based on the same linear model with varying autocorrelation coefficients AR(1) ϕ equal to 0.1, 0.5, and 0.8 and T number of observations equal to 50, 100, and 500.

The considered DGP is as follows:

$$\begin{aligned} y_t &= 0.8y_{t-1} + 0.6x_{1,t-1} + 0.3x_{1,t-2} - 0.5x_{2,t-1} - 0.2x_{2,t-2} \\ &\quad + 0.4x_{3,t-1} + 0.3x_{3,t-2} + 0.4x_{4,t-1} - 0.3x_{5,t-1} \\ &\quad + 0.2x_{6,t-1} + \varepsilon_t, \\ \varepsilon_t &\sim N(0, 1) \quad t = 1, 2, \dots, T. \end{aligned} \quad (8)$$

We employ WLAdaLASSO, Autometrics, and other regularization techniques to estimate the model. The lag length of dependent and independent regressors is equal to 5 throughout the simulation study with varying T observations and ϕ parameter of independent regressors. We eliminate the last ten observations of the simulated series to implement the out-of-sample RMSE. The RMSE of the out-of-sample forecast is reported in the below figures, and simulation is repeated 1000 times.

3.2. Simulation Results. Tables 1–3 illustrate the simulation findings of considered techniques in terms of average gauge and potency. The simulated result of out-of-sample RMSE is presented in the figures. The above indicates that, among all concerned techniques, WLAdaLASSO outperforms others in potency equal to 63.6%, with T being equal to 50, whereas Autometrics with a 0.01 level of significance retains the least potency of 16.1% on average. As the sample size increases, the performance of considered techniques is improved in average potency (increases) and average gauge (decreases). Table 1 indicates that Autometrics retains its average gauge around its nominal significance level (0.05) and 0.01 at the cost of least average potency. However, with an increase in sample T equal to 500, the Autometrics with 0.05 significance performs near WLAdaLASSO both in potency and gauge. Among regularization techniques, LASSO, AdaLASSO, SCAD, and MCP perform inferior to WLAdaLASSO in gauge and potency.

Tables 2 and 3 illustrate the simulated result with ϕ (autocorrelation coefficients) of regressors equal to 0.5 and 0.8. The WLAdaLASSO estimate outperforms others with an average potency of 64.5%, ϕ equal to 0.5, and T equal to 50. As the T sample of WLAdaLASSO increases, the average gauge approaches nominal significance level, and average potency approaches 1. The simulation result indicates that

the WLAdaLASSO estimate is not sensitive to autocorrelation coefficients as with ϕ equal to 0.1 and T equal to 50; the average retains potency equal to 63.6% and 64.5% with ϕ being equal to 0.5. However, the Autometrics performs poorly as the autocorrelation coefficient increases from 0.1 to 0.5.

Meanwhile, Autometrics with ϕ equal to 0.8 and T equal to 50 possess 11.5% gauge, which is higher than the 5% significance level. The performance of Autometrics does not get enhanced (gauge $\rightarrow \alpha$, and potency $\rightarrow 1$) as the sample size increases with ϕ being equal to 0.8. However, WLAdaLASSO performs superior in average potency and gauge compared to all other techniques. The simulation experiment indicates that WLAdaLASSO performs robust even with a stronger linear dependence between predictors. With increasing sample, the performance of Autometrics, LASSO, AdaLASSO, SCAD, and MCP does not get enhanced as that of WLAdaLASSO. The WLAdaLASSO performs superior to other considered regularization techniques and to Autometrics as well in average gauge and potency even with higher and weak linear dependency between predictors and small sample size.

Figures 1–3 illustrate the RMSE of considered techniques with T samples being equal to 50, 100, and 500 and ϕ equal to 0.1, 0.5, and 0.8. The result shows that the WLAdaLASSO outperforms other techniques in out-of-sample forecasting. The WLAdaLASSO estimate is insensitive to autocorrelation coefficients, as the forecast performance and average potency have not decreased because ϕ equals 0.8 even with a small sample. However, with ϕ being equal to 0.8, all other techniques perform poorly in out-of-sample forecasting, whereas WLAdaLASSO possesses the least RMSE. Autometrics with autocorrelation coefficient equals 0.1, and T equal to 50 performs poorly in RMSE compared to WLAdaLASSO, but with sample size increment, the RMSE decreases because the average potency increases. However, Autometrics with a ϕ being equal to 0.8 and T equal to 50 performs the worst among all other techniques. The overall simulation result indicates that WLAdaLASSO outperforms Autometrics and other regularization techniques in potency and out-of-sample forecasting.

4. Real Data Analysis

For the real data analysis, we aim to probe the determinants of the trade balance for Pakistan and implement the considered techniques and assess their performance. Trade has played an important role in developing countries as a growth engine in various eras. The trade deficit or surplus is a term used to describe trade imbalances. Since independence, Pakistan has been in a trade deficit, except for three years: 1947–1948, 1950–1951, and 1972–1973 [40]. According to economic literature, a variety of factors are thought to be responsible for long-term trade deficits in various economies, including ineffective public policies, shocks in major trading countries, oil price hikes if the economy is heavily reliant on oil imports, residents’ socioeconomic conditions, and increased urbanization [41, 42]. The existing studies in the case of Pakistan considered only a few macroeconomic

TABLE 1: Simulated result with ϕ equal to 0.1.

		$T = 50$	$T = 100$	$T = 500$
WLAdaLASSO	Gauge	0.268	0.055	0.014
	Potency	0.636	0.708	0.954
Autometrics (0.05)	Gauge	0.069	0.037	0.009
	Potency	0.256	0.375	0.960
Autometrics (0.01)	Gauge	0.028	0.016	0.001
	Potency	0.161	0.257	0.892
LASSO	Gauge	0.465	0.186	0.134
	Potency	0.619	0.612	0.713
AdaLASSO	Gauge	0.197	0.092	0.080
	Potency	0.389	0.523	0.681
SCAD	Gauge	0.176	0.174	0.098
	Potency	0.380	0.584	0.700
MCP	Gauge	0.178	0.155	0.083
	Potency	0.360	0.553	0.694

TABLE 2: Simulated result with ϕ equal to 0.5.

		$T = 50$	$T = 100$	$T = 500$
WLAdaLASSO	Gauge	0.210	0.055	0.013
	Potency	0.649	0.830	0.997
Autometrics (0.05)	Gauge	0.065	0.042	0.033
	Potency	0.280	0.407	0.807
Autometrics (0.01)	Gauge	0.034	0.018	0.025
	Potency	0.243	0.306	0.704
LASSO	Gauge	0.510	0.199	0.125
	Potency	0.692	0.660	0.717
AdaLASSO	Gauge	0.251	0.097	0.076
	Potency	0.460	0.520	0.691
SCAD	Gauge	0.210	0.175	0.093
	Potency	0.351	0.581	0.699
MCP	Gauge	0.221	0.148	0.079
	Potency	0.350	0.556	0.691

TABLE 3: Simulated result with ϕ equal to 0.8.

		$T = 50$	$T = 100$	$T = 500$
WLAdaLASSO	Gauge	0.220	0.056	0.011
	Potency	0.696	0.707	0.992
Autometrics (0.05)	Gauge	0.115	0.050	0.075
	Potency	0.282	0.471	0.601
Autometrics (0.01)	Gauge	0.083	0.027	0.066
	Potency	0.206	0.437	0.522
LASSO	Gauge	0.298	0.231	0.157
	Potency	0.587	0.683	0.719
AdaLASSO	Gauge	0.112	0.108	0.078
	Potency	0.327	0.516	0.683
SCAD	Gauge	0.078	0.177	0.091
	Potency	0.339	0.537	0.675
MCP	Gauge	0.071	0.150	0.083
	Potency	0.291	0.528	0.680

variables as like GDP, exchange rate, broad money supply, inflation, and Foreign Direct Investment [40, 43–47]. This study intakes the Generalized Unidentified Model (GUM) that includes each and every possible determinant of trade balance with 11 regressors, namely, Domestic Investment (log), Domestic Consumption (log), FDI (log), GDP (log), Inflation (log), Budget Deficit (log), Remittances (log),

Exchange Rate (log), Population (log), Urban population (log), and Government expenditure (log).

We use annual frequency data from 1980 to 2020. The data has been compiled from World Data Indicator. The model contains 11 regressors (with a difference) and includes 5 lags of each covariates and the lags of the dependent variable. The GUM includes 71 covariates; due to

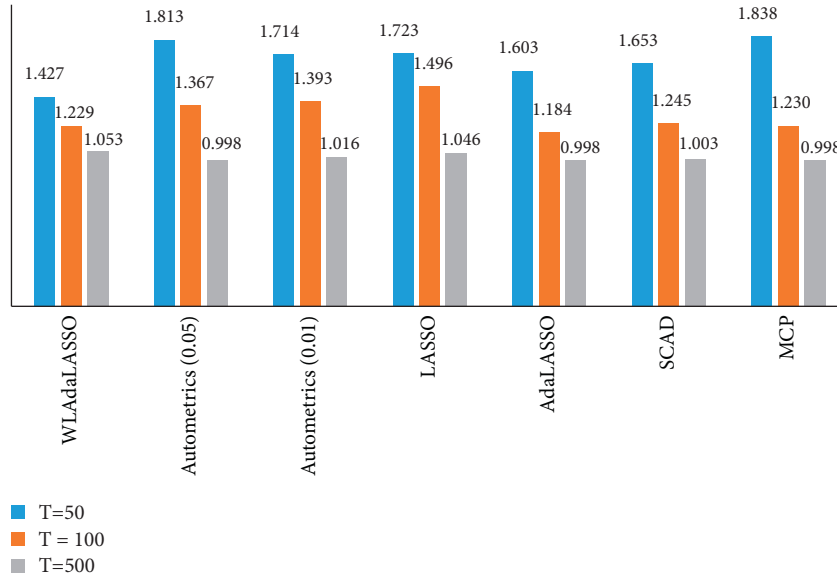


FIGURE 1: Simulated result of RMSE with ϕ equal to 0.1.

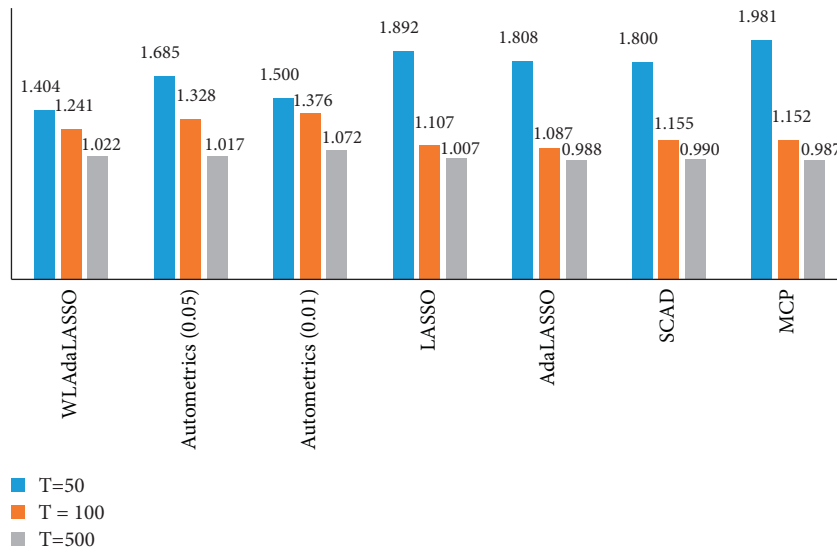


FIGURE 2: Simulated result of RMSE with ϕ equal to 0.5.

differencing data and including 5 lags of covariates, we have a total of 35 observations. We train the model on 30 observations from 1985 to 2015 as the last 5 observations 2016–2020 have been discarded for test data. Throughout simulation experiment and real data analysis, we use BIC-based tuning parameter for regularization techniques, while, for Autometrics, we select the model with 0.01 and 0.05 significance levels.

The real data analysis illustrated in Figure 4 verifies our simulation findings as WLAdaLASSO outperforms all other techniques with the least out-of-sample RMSE being equal to 0.0069 followed by Autometrics (0.01) with RMSE of 0.018. Autometrics with 0.05 possesses a higher RMSE equal to 0.111 than Autometrics with a 0.01 significance level. The finding is aligned with the simulation experiment as

Autometrics with a 0.05 level of significance possesses a slightly higher average gauge with higher RMSE than a 0.01 level of significance. SCAD, MCP, LASSO, and AdaLASSO estimate higher RMSE as the model selects more irrelevant covariates and lag than WLAdaLASSO and Autometrics. WLAdaLASSO selects three covariates, namely, difference of urban population (dupop), difference of log GDP lag 1 (dlnGDP(-1)), and difference of log population lag 4 (dpop(-4)). Autometrics with a 0.05 significance level selects five covariates and their lag, and with 0.01 significance level, it selects three covariates. dlnGDP(-1) is a common covariate between WLAdaLASSO and Autometrics with 0.05 and 0.01 significance levels. SCAD, MCP, and LASSO select too many covariates and their lag, due to which these techniques possess higher RMSE compared to

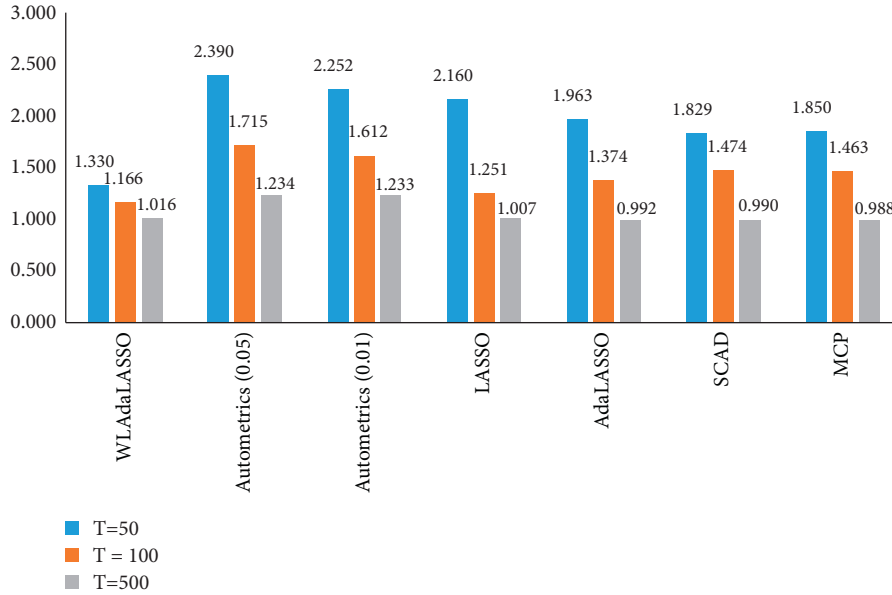
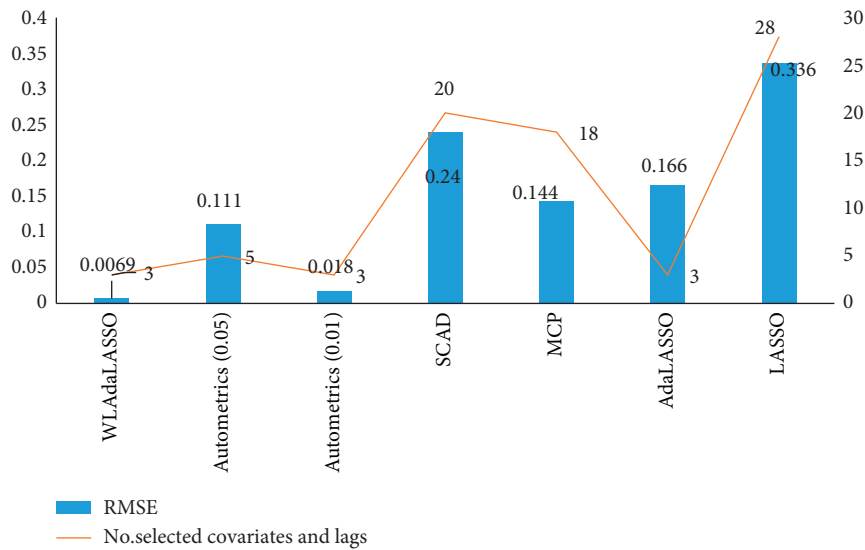
FIGURE 3: Simulated result of RMSE with ϕ equal to 0.8.

FIGURE 4: Real data analysis.

WLAdaLASSO. However, AdaLASSO selects three covariates and their lag with RMSE being equal to 0.16, which is higher than that of WLAdaLASSO and Autometrics.

5. Conclusion

Regularization techniques have become extremely popular in time-series modeling in recent years due to availability of massive data. This study aims to analyze the performance of the WLAdaLASSO with Autometrics for covariate selection and forecasting. The simulation study illustrates that the WLAdaLASSO with the stronger linear dependency between predictors outperforms Autometrics and other regularization techniques. However, Autometrics with ϕ being equal 0.1, the performance of gauge $\rightarrow \alpha$ (0.05 or 0.01 level of

significance), potency $\rightarrow 1$, and the Average RMSE also decrease, with sample size increment. On the contrary, the situation is limited to ϕ equal to 0.1; however, ϕ equal to 0.8 and increasing sample size does not significantly enhance the performance of Autometrics compared to WLAdaLASSO. Autometrics with 0.05 significances level includes irrelevant covariates that increase the RMSE compared to 0.01 significance, and the finding is aligned with real data analysis. However, other than the WLAdaLASSO, all considered regularization techniques perform poorly in covariate selection and forecasting even with ϕ being equal to 0.1 and T equal to 50, whereas the performance of considered techniques is improved with an increase in sample size; still, WLAdaLASSO outperformed others among all simulation experiments. The simulation experiment and real data

analysis are evidence that the WLAdaLASSO is a more robust technique than all other considered regularization techniques and Autometrics as well in out-of-sample forecasting and covariate selection even with the stronger linear dependence between predictors and small sample size.

5.1. Limitations. One of the study's constraints is that it only considers linear models and annual data. These simulation experiments are limited to Gaussian distributed errors.

Data Availability

Coding of the simulation study, coding of the real data analysis, and data used for analysis can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] G. E. P. Box, "Robustness in the strategy of scientific model building," *Robustness in Statistics*, pp. 201–236, 1979.
- [2] M. H. Pesaran, Y. Shin, and R. J. Smith, "Bounds testing approaches to the analysis of level relationships," *Journal of Applied Econometrics*, vol. 16, no. 3, pp. 289–326, 2001.
- [3] M. H. Pesaran and R. Smith, "Estimating long-run relationships from dynamic heterogeneous panels," *Journal of Econometrics*, vol. 68, no. 1, pp. 79–113, 1995.
- [4] X. Cui, J. Hu, Y. Ma, P. Wu, P. Zhu, and H.-J. Li, "Investigation of stock price network based on time series analysis and complex network," *International Journal of Modern Physics B*, vol. 35, no. 13, Article ID 2150171, 2021.
- [5] J. Hu, C. Xia, H. Li, P. Zhu, and W. Xiong, "Properties and structural analyses of USA's regional electricity market: a visibility graph network approach," *Applied Mathematics and Computation*, vol. 385, Article ID 125434, 2020.
- [6] H.-H. Qiao, Z.-H. Deng, H.-J. Li, J. Hu, Q. Song, and L. Gao, "Research on historical phase division of terrorism: an analysis method by time series complex network," *Neurocomputing*, vol. 420, pp. 246–265, 2021.
- [7] J. Bai and S. Ng, "Determining the number of factors in approximate factor models," *Econometrica*, vol. 70, no. 1, pp. 191–221, 2002.
- [8] C. De Mol, D. Giannone, L. Reichlin, C. De Mol, D. Giannone, and L. Reichlin, "Forecasting using a large number of predictors: is Bayesian shrinkage a valid alternative to principal components?" *Journal of Econometrics*, vol. 146, no. 2, pp. 318–328, 2008.
- [9] Y. Hua, *Macroeconomic Forecasting Using Large Vector Auto Regressive Model*, Doctoral dissertation, Humboldt-Universität zu Berlin, Berlin, Germany, 2011.
- [10] J. Li, "Monetary policy analysis based on lasso-assisted vector autoregression (lavar)," *SSRN Electronic Journal* 30 pages, 2012.
- [11] J. Li and W. Chen, "Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models," *International Journal of Forecasting*, vol. 30, no. 4, pp. 996–1015, 2014.
- [12] C. Marsilli, "Variable selection in predictive MIDAS models," *SSRN Electronic Journal*, vol. 39, 2014.
- [13] W. B. Nicholson, D. S. Matteson, and J. Bien, "VARX-L: structured regularization for large vector autoregressions with exogenous variables," *International Journal of Forecasting*, vol. 33, no. 3, pp. 627–651, 2017.
- [14] M. C. Medeiros and E. F. Mendes, "Estimating High-Dimensional Time Series Models," *Estimating High Dimensional Time Series*, <https://www.econstor.eu/bitstream/10419/176085/1/td602.pdf>, 2012.
- [15] F. Audrino and L. Camponovo, "Oracle properties and finite sample inference of the adaptive lasso for time series regression models," *SSRN Electronic Journal*, 2013.
- [16] S. Song and P. J. Bickel, "Large Vector Auto Regressions," 2011, <https://arxiv.org/abs/1106.3915v1>.
- [17] E. Konzen and F. A. Ziegelmann, "LASSO-type penalties for covariate selection and forecasting in time series," *Journal of Forecasting*, vol. 35, no. 7, pp. 592–612, 2016.
- [18] Y. Uematsu and S. Tanaka, "High dimensional macroeconomic forecasting and variable selection via penalized regression," *The Econometrics Journal*, vol. 22, no. 1, pp. 34–56, 2019.
- [19] J. L. Castle, M. P. Clements, and D. F. Hendry, "Forecasting by factors, by variables, by both or neither?" *Journal of Econometrics*, vol. 177, no. 2, pp. 305–319, 2013.
- [20] J. L. Castle, J. A. Doornik, and D. F. Hendry, "Modelling non-stationary 'big data,'" *International Journal of Forecasting*, vol. 37, no. 4, pp. 1556–1575, 2021.
- [21] F. Khan, A. Urooj, S. A. Khan, A. Alsubie, Z. Almaspoor, and S. Muhammadullah, "Comparing the forecast performance of advanced statistical and machine learning techniques using huge big data: evidence from Monte Carlo experiments," *Complexity*, vol. 2021, Article ID 6117513, 11 pages, 2021.
- [22] F. Khan, A. Urooj, K. Ullah, B. Alnssyan, and Z. Almaspoor, "A comparison of Autometrics and penalization techniques under various error distributions: evidence from Monte Carlo simulation," *Complexity*, vol. 2021, Article ID 9223763, 8 pages, 2021.
- [23] C. Epprecht, D. Guégan, Á. Veiga, and J. Correa da Rosa, "Variable selection and forecasting via automated methods for linear models: LASSO/adaLASSO and Autometrics," *Communications in Statistics - Simulation and Computation*, vol. 50, no. 1, pp. 103–122, 2021.
- [24] K. D. Hoover, S. J. Perez, and / . ~kdhoover, "Data mining reconsidered: encompassing and the general to specific approach to specification search," *The Econometrics Journal*, vol. 2, no. 2, pp. 167–191, 1999.
- [25] J. A. Doornik, *Econometric Model Selection with More Variables than Observations*, Citeseer, Princeton, NJ, USA, 2009.
- [26] H. M. Krolzig and D. F. Hendry, "Computer automation of general-to-specific model selection procedures," *Journal of Economic Dynamics and Control*, vol. 25, no. 6–7, pp. 831–866, 2001.
- [27] D. F. Hendry and H. M. Krolzig, "We ran one regression. In oxford bulletin of economics and statistics," *John Wiley & Sons, Ltd*, vol. 66, no. 5, pp. 799–810, 2004.
- [28] C. M. Jarque and A. K. Bera, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Economics Letters*, vol. 6, no. 3, pp. 255–259, 1980.
- [29] T. S. Breusch and A. R. Pagan, "The Lagrange multiplier test and its applications to model specification in econometrics," *The Review of Economic Studies*, vol. 47, no. 1, p. 239, 1980.
- [30] L. G. Godfrey, "Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables," *Econometrica*, vol. 46, no. 6, p. 1303, 1978.

- [31] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, p. 987, 1982.
- [32] G. C. Chow, "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, vol. 28, no. 3, p. 591, 1960.
- [33] F. Pretis, J. J. Reade, and G. Sucarrat, "Automated general-to-specific (GETS) regression modeling and indicator saturation for outliers and structural breaks," *Journal of Statistical Software*, vol. 86, 2018.
- [34] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, 2006.
- [35] H. Park and F. Sakaori, "Lag weighted lasso for time series model," *Computational Statistics*, vol. 28, no. 2, pp. 493–504, 2013.
- [36] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [37] C. H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Annals of Statistics*, vol. 38, no. 2, pp. 894–942, 2010.
- [38] P. Bühlmann and S. Van De Geer, *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer Science & Business Media, Berlin, Germany, 2011.
- [39] J. A. Doornik and D. F. Hendry, "Statistical model selection with "big data"," *Cogent Economics & Finance*, vol. 3, no. 1, Article ID 1045216, 2015.
- [40] K. Asif, "Determinants of trade balance: a comparison between Pakistan and India," *IBA Business Review*, vol. 9, no. 1, 2014.
- [41] C. Grupe and A. Rose, "China, India, and the socioeconomic determinants of their competitiveness," *Economics Research International*, vol. 2010, Article ID 860425, 14 pages, 2010.
- [42] V. Manual and W. San, "Dynamic relationship between trade balance and macroeconomic elements: empirical evidence from emerging economies in Malaysia," *International Journal of Recent Technology and Engineering*, vol. 7, no. 5, pp. 47–54, 2019.
- [43] A. Awan, N. Asghar, and H. U. Rehman, "The impact of exchange rate, fiscal deficit and terms of trade on external debt of Pakistan," *Australian Journal of Business and Management Research*, vol. 1, no. 3, p. 10, 2011.
- [44] A. Hussain and S. D. Muhammad, "The role of exchange rate on balance of trade: empirical from Pakistan," *European Journal of Social Sciences*, vol. 14, no. 1, p. 150, 2010.
- [45] M. K. Kakar, R. Kakar, and W. Khan, "The determinants of Pakistan's trade balance: an ARDL cointegration approach," *Lahore Journal of Economics*, vol. 15, no. 1, 2010.
- [46] S. D. Muhammad, "Determinant of balance of trade: case study of Pakistan," *European Journal of Scientific Research*, vol. 41, no. 1, pp. 13–20, 2010.
- [47] M. Shahbaz, A. Jalil, and F. Islam, *Real Exchange Rate Changes and Trade Balance in Pakistan: A Revisit*, 2010.