# Oracle Properties and Finite Sample Inference of the Adaptive Lasso for Time Series Regression Models

Francesco Audrino[*] Lorenzo Camponovo[†]
University of St.Gallen University of St.Gallen

October 2013

**Abstract**

We derive new theoretical results on the properties of the adaptive least absolute shrinkage and selection operator (adaptive lasso) for time series regression models. In particular we investigate the question of how to conduct finite sample inference on the parameters given an adaptive lasso model for some fixed value of the shrinkage parameter. Central in this study is the test of the hypothesis that a given adaptive lasso parameter equals zero, which therefore tests for a false positive. To this end we construct a simple (conservative) testing procedure and show, theoretically and empirically through extensive Monte Carlo simulations, that the adaptive lasso combines efficient parameter estimation, variable selection, and valid finite sample inference in one step. Moreover, we analytically derive a bias correction factor that is able to significantly improve the empirical coverage of the test on the active variables. Finally, we apply the introduced testing procedure to investigate the relation between the short rate dynamics and the economy, thereby providing a statistical foundation (from a model choice perspective) to the classic Taylor rule monetary policy model.

**JEL classifications**: C12; C22; E43.

**Keywords**: Adaptive lasso; Time series; Oracle properties; Finite sample inference; Taylor rule monetary policy model.

[*]E-mail: francesco.audrino@unisg.ch. Address: Bodanstrasse 6, 9000 St.Gallen, Switzerland.
[†]E-mail: lorenzo.camponovo@unisg.ch. Address: Bodanstrasse 6, 9000 St.Gallen, Switzerland.

# 1 Introduction

Recent years have seen a steady increase in the availability of large amounts of economic data. This raises the question of how best to exploit this information to refine benchmark techniques used by the private industry and the academic community. In this study we focus on the question of how to perform consistent variable selection and (finite sample) inference in classical time series regression models with a fixed number of candidate variables and a general error distribution.

To this end we consider a technique recently introduced in the machine learning community belonging to the group of shrinkage methodologies (that is, following the idea of shrinking to zero the coefficients of the irrelevant variables) that has proven its worth and is becoming increasingly popular: the Least Absolute Shrinkage and Selection Operator (lasso), introduced by Tibshirani (1996), and its refined version known as the adaptive lasso, proposed by Zou (2006).[1] The main problem with the lasso is that it requires a condition denoted as the *irrepresentable condition*, which is essentially a necessary condition for exact recovery of the non-zero coefficients that is much too restrictive in many cases.[2] In fact, irrepresentable conditions show that the lasso typically selects too many variables and that so-called false positives are unavoidable. Zou (2006) proposed the adaptive lasso to alleviate this problem and to try to reduce the number of false positives. Moreover, the adaptive lasso estimator also fulfills the oracle property in the sense introduced by Fan and Li (2001).

The interest in using lasso-type techniques in general applications as well, such as those in economics and finance, raises the question of how to extend the most advanced theoretical results derived for the iid cross-sectional setting to the time-series setting. Recent papers investigating this question in settings requiring different assumptions and conditions on the number of active variables and on the error distribution include Wang et al. (2007), Hsu et al. (2008), Nardi and Rinaldo (2011), Song and Bickel (2011), Kock and Callot (2012), Park and Sakaori (2012), and Audrino and Knaus (2012).

Two studies recently proposed are closely related to our work: Kock (2012) and Medeiros and Mendes (2012). Each investigates the asymptotic properties of the adaptive lasso estimator in single-equation linear time series models: While Kock (2012) focuses more on the conditions needed to perform consistent variable selection in stationary and non-stationary autoregressive models using the adaptive lasso with a fixed number of variables, Medeiros and Mendes (2012) extend the basic linear model investigated in the studies cited above to include exogenous vari-

[1] We refer to Bühlmann and van de Geer (2011) for a detailed discussion of the lasso estimator and its generalizations in the cross-sectional setting for independent and identically distributed (iid) variables.

[2] See, for example, Zhao and Yu (2006) for more details.

ables, non-Gaussian, conditionally heteroscedastic, and possibly time-dependent errors, and a number of variables (candidate and relevant ones) that is allowed to increase as a function of the sample size. Although some of their results are identical to those of the present study, conditions on the model and proofs are substantially (if not completely) different.

Moreover, in this paper we take the discussion a step further and contribute to the literature on the adaptive lasso along two main dimensions. First, we quantify the bias in finite samples that is incurred when making inference on the active variables in the time series regression by introducing an explicit formula. Second, we show how one can make (conservative) inference in the case of the non-active variables in the predictive regression. In particular, we introduce a very simple procedure based on the least squares estimator on the predictive regression with all candidate variables to test the hypothesis that a given parameter is zero, i.e. that the corresponding variable does not belong to the active set. Our theoretical results show that the adaptive lasso may combine efficient parameter estimation, variable selection, and valid finite sample inference in one step.

To the best of our knowledge, these are new results. They extend the usefulness of the (adaptive) lasso beyond variable selection for performing statistical inference (such as tests of hypotheses and the construction of confidence intervals) in a broad spectrum of applications in all fields dealing with a large amount of iid and time series data.

Some related research on the significance of the active variables in a lasso model has been recently proposed by Lockhart et al. (2013). Nevertheless, their proposed covariance statistic for testing the significance of predictor variables as they enter the active set, along the lasso solution path, differs considerably from the approach we propose in this study. As Lockhart et al. (2013) maintain in their discussion section at the end of the paper, the question of how to carry out a significance test for any predictor in the active set given a lasso model at some fixed value of $\lambda_n$ (i.e. the tuning parameter of the lasso that controls the amount of shrinkage) was left for future research. Finally, a similar approach for the lasso in the Gaussian iid setting has been proposed by Javanmard and Montanari (2013).

Using an extensive simulation study based on data generating processes with a different number of variables, error distributions, and number of observations at disposal, we investigate the relevance of the bias correction factor and the effectiveness of the introduced testing procedure. First, results show the importance in finite samples of the bias correction factor for the active variables: the empirical coverages are significantly improved, in particular for variables with small coefficients. Second, although conservative in their construction, tests of the null hypothesis that coefficients are equal to zero in particular for non-active variables give accurate results, yielding reasonably small proportions of false rejections. Third, if enough data is available, tests on the active variables with small coefficients also produce satisfactory power results.

Finally, we apply the adaptive lasso to a classic problem in financial economics that has been investigated in the academic community for the last twenty years: the relation between interest rates and the state of the economy.[3] In particular, we analyze which variables, from a group of macroeconomic and financial indicators, are relevant explaining the short-term interest rate in a simple Taylor rule-type monetary policy model (see Taylor, 1993, page 202). Not surprisingly, we find that the only predictors belonging to the set of active variables identified by the adaptive lasso are the following three: one-lag past short rate values (which take into account the well-known persistence of the short rate dynamics and act as a proxy for additional macroeconomic, monetary policy, or even financial variables), an inflation indicator, and the unemployment rate. This result adds a purely statistical foundation to the classic economic intuition driving the Taylor rule, suggesting that the Federal Reserve System (Fed) increases interest rates in times of high inflation, or when employment is above the full employment levels, and decreases interest rates in the opposite situations.

The content of this paper can be summarized as follows: Section 2 introduces the model we are going to consider. Oracle properties of the adaptive lasso estimated on time series regressions are discussed in Section 3. In Section 4 we introduce the statistical testing procedure that can be used to make finite samples inference on both active and non-active variables. Monte Carlo simulation results are shown in Section 5 and the application on the prediction of the short-term interest rate is performed in Section 6. Finally, Section 7 concludes. All proofs of the theorems in the paper are provided in the appendix.

# 2 Model and Notation

Consider the stationary linear regression model

$$Y_t = \sum_{i=1}^{p_1} \rho_i^* Y_{t-i} + \sum_{i=1}^{p_2} \gamma_i^* W_{i,t} + \sum_{i=1}^{p_3} \beta_i^* X_{i,t-1} + \epsilon_t, \tag{1}$$

where $p_1 + p_2 + p_3 = p < \infty$, $W_t = (W_{1,t}, \dots, W_{p_2,t})'$ is a vector of covariates at time $t$, $X_{t-1} = (X_{1,t-1}, \dots, X_{p_3,t-1})'$ is a vector of regressors at time $t-1$ assumed to predict $Y_t$, $\epsilon_t$ is a zero-mean error term, and $\theta^* = (\rho_1^*, \dots, \rho_{p_1}^*, \gamma_1^*, \dots, \gamma_{p_2}^*, \beta_1^*, \dots, \beta_{p_3}^*)'$ is the unknown parameter of interest. Important examples of linear regression models (1) include: autoregressive models (when $\gamma_i^* = 0$ and $\beta_j^* = 0$, for $i = 1, \dots, p_2$ and $j = 1, \dots, p_3$); iid linear regression models (when

[3] See, among others, Clarida et al. (2000), Ang and Piazzesi (2003), Dewachter and Lyrio (2006), Moench (2008), Ang et al. (2008), Rudebusch (2010), and Filipova et al. (2013).

$\rho_i^* = 0$ and $\beta_j^* = 0$, for $i = 1, \ldots, p_1$ and $j = 1, \ldots, p_3$); and predictive regression models (when $\rho_i^* = 0$ and $\gamma_j^* = 0$, for $i = 1, \ldots, p_1$ and $j = 1, \ldots, p_2$).

To simplify our notation, we just write $\theta^* = (\theta_1^*, \ldots, \theta_p^*)$, i.e., we set $\theta_i^* = \rho_i^*$, for $i = 1, \ldots, p_1$, $\theta_i^* = \gamma_{i-p_1}^*$, for $i = p_1 + 1, \ldots, p_1 + p_2$, and $\theta_i^* = \beta_{i-p_1-p_2}^*$, for $i = p_1 + p_2 + 1, \ldots, p_1 + p_2 + p_3$. A common way to estimate the unknown parameter $\theta^*$ relies on the least squares approach. More precisely, we can introduce the least squares estimator $\hat{\theta}_{LS} = (\hat{\theta}_{LS,1}, \ldots, \hat{\theta}_{LS,p})'$ of $\theta^*$ defined as

$$\hat{\theta}_{LS} = arg \min_{\theta} \frac{1}{n} \sum_{t=1}^{n} (Y_t - \theta' Z_t)^2,$$

where $Z_t = (Y_{t-1}, \ldots, Y_{t-p_1}, W_t', X_{t-1}')'$ and $n$ denotes the sample size. Furthermore, under some regularity conditions and using standard techniques, we can show that

$$\sqrt{n}(\hat{\theta}_{LS} - \theta^*) \to N(0, V),$$

i.e, $\hat{\theta}_{LS}$ is a consistent estimator of $\theta^*$ with normal limit distribution and covariance matrix $V$.

Let $A = \{i : \theta_i^* \neq 0\}$ denote the set of active variables, and assume that $|A| = q < p$. Similarly, Let $\hat{A}_{LS} = \{i : \hat{\theta}_{LS,i}^* \neq 0\}$. Then, in general $|\hat{A}_{LS}| = p \neq q$. Thus, in spite of an efficient estimate of the unknown parameter, the least squares approach does not perform variable selection. In the iid context, Zou (2006) introduces a lasso procedure which combines both efficient parameter estimation and variable selection in one step. To achieve this objective in predictive regression models as well, we extend the lasso method to our setting. More precisely, we introduce the adaptive lasso estimator $\hat{\theta}_{AL} = (\hat{\theta}_{AL,1}, \ldots, \hat{\theta}_{AL,p})'$ of $\theta^*$ defined as

$$\hat{\theta}_{AL} = arg \min_{\theta} \frac{1}{n} \sum_{t=1}^{n} (Y_t - \theta' Z_t)^2 + \frac{\lambda_n}{n} \sum_{i=1}^{p} \lambda_{n,i} |\theta_i|, \tag{2}$$

where $\lambda_n$ is a tuning parameter and $\lambda_{n,i} = 1/|\hat{\theta}_{LS,i}|$. To simplify the presentation of our results, we consider only the weights $\lambda_{n,i} = 1/|\hat{\theta}_{LS,i}|$ instead of the more general penalizations $\lambda_{n,i}^{(\gamma)} = 1/|\hat{\theta}_i|^{\gamma}$ proposed in Zou (2006), where $\gamma > 0$ and $\hat{\theta}_i$ is a root-$n$ consistent estimator of $\theta_i^*$. However, with some slight modifications we can extend the results in Sections 3 and 4 to this more general framework.

It is important to note that the penalization of the variables in (2) also depends on the least squares estimate $\hat{\theta}_{LS}$. In particular, variables with least squares estimates close to zero are more penalized. This property represents a key condition for ensuring valid variable selection, as highlighted in Zou (2006) in the iid context. In the next section, we derive the asymptotic properties of the adaptive lasso for time series regression models.

## 3 Oracle Properties of the Adaptive Lasso

In the iid context, the adaptive lasso procedure introduced in Zou (2006) possesses so-called oracle properties. More precisely, the adaptive lasso simultaneously performs correct variable selection and provides an efficient estimate of the non-zero coefficients as if only the relevant variables had been included in the model. In the next theorem we show that the adaptive lasso enjoys these properties in time series regression models as well. Before presenting the main results, first we introduce some notation. Let $\theta^{*A} = (\theta_1^{*A}, \ldots, \theta_q^{*A})'$ denotes the sub-vector of the non-zero coefficients of $\theta^*$. Similarly, let $\hat{\theta}_{LS}^A = (\hat{\theta}_{LS,1}^A, \ldots, \hat{\theta}_{LS,q}^A)'$ and $\hat{\theta}_{AL}^A = (\hat{\theta}_{AL,1}^A, \ldots, \hat{\theta}_{AL,q}^A)'$ denote the least squares and adaptive lasso estimates of $\theta^{*A}$. Furthermore, let $V^A$ be the asymptotic covariance matrix of $\hat{\theta}_{LS}^A$. Finally, let $\hat{A}_{AL} = \{i : \hat{\theta}_{AL,i} \neq 0\}$. The oracle properties of the adaptive lasso are derived in the next theorem.

**Theorem 3.1.** *Let $p_1 + p_2 + p_3 = p < \infty$. Assume that $\{Y_t\}$ and $\{Z_t\}$ are stationary processes such that $\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^n Z_t Z_t' = C$ exists and is of full rank, and $\frac{1}{\sqrt{n}}\sum_{t=1}^n \epsilon_t Z_t$ converges in distribution to a random variable with $N(0,\Omega)$ for some covariance matrix $\Omega$. If (i) $\lambda_n \to +\infty$ and (ii) $\frac{\lambda_n}{\sqrt{n}} \to 0$, then:*

*(I) Variable Selection: $\lim_{n\to\infty} P(\hat{A}_{AL} = A) = 1$.*

*(II) Limit Distribution:*

$$\sqrt{n}\left(\hat{\theta}_{AL}^A - \theta^{*A}\right) + \hat{b}_{AL}^A \to N(0, V^A),$$

*where the bias term is given by*

$$\hat{b}_{AL}^A = \left(\frac{1}{n}\sum_{i=1}^n Z_t^A Z_t'^A\right)^{-1} \cdot \left(\frac{\lambda_n}{2\sqrt{n}}\lambda_{n,1}^A sign(\hat{\theta}_{AL,1}^A), \ldots, \frac{\lambda_n}{2\sqrt{n}}\lambda_{n,q}^A sign(\hat{\theta}_{AL,q}^A)\right)',$$

*$Z_t^A$ is the sub-vector of $Z_t$ for the non-zero coefficients, and $\lambda_{n,i}^A = 1/|\hat{\theta}_{LS,i}^A|$, $i = 1, \ldots, q$.*

The assumptions in Theorem 3.1 are mild conditions which are also required for proving the consistency and deriving the limit distribution of the least squares estimator $\hat{\theta}_{LS}$. Statement (I) of Theorem 3.1 establishes that the adaptive lasso also performs correct variable selection in time series settings, i.e. the adaptive lasso asymptotically identifies the sub-vector of the non-zero coefficients of $\theta^*$. Furthermore, in statement (II) we derive the limit distribution. In particular, we note that the adaptive lasso has the same limit distribution of the least squares estimator. Therefore, the adaptive lasso performs variable selection and efficient parameter estimation in one step.

The oracle properties (I) and (II) discussed above are in line with the results shown in Kock (2012) and Medeiros and Mendes (2012) that were derived using substantially different arguments and proofs. Moreover, moving beyond those studies, in statement (II) we also provide an explicit formula for the bias term $\hat{b}^A_{AL}$ that is incurred when making inference on the active variables. This term is asymptotically negligible but provides important refinements for finite sample inference as highlighted in Section 5.

# 4 Finite Sample Inference with the Adaptive Lasso

In the previous section, we proved the asymptotic properties of the adaptive lasso for time series regression models. In particular, we showed that the limit distribution of the estimates of the non-zero coefficients is normal, while those of the zero coefficients collapse to zero. This allowed us to use these results to introduce inference on the non-zero coefficients. However, since a priori we do not know the non-zero coefficients $\theta^{*A}$, the practical implementation of testing procedures in this context remains unclear. To deepen our understanding of this issue, suppose that the estimate of the first component of $\theta^*$ is different from zero, i.e., $\hat{\theta}_{AL,1} \neq 0$. Then, we have two cases: (i) $\theta^*_1 \neq 0$ or (ii) $\theta^*_1 = 0$. If $\theta^*_1 \neq 0$, then by Theorem 3.1 the limit distribution of $\hat{\theta}_{AL,1}$ is normal. Thus, we can construct Gaussian confidence intervals for the parameter of interest. On the other hand, if $\theta^*_1 = 0$, then by Theorem 3.1 we can only conclude that $\hat{\theta}_{AL,1}$ must collapse to zero asymptotically. Since a priori we do not know whether (i) or (ii) is satisfied, it turns out that we also do not know how to make inference on the parameter $\theta^*_1$.

The aim of this section is to clarify how to introduce valid finite sample inference with the adaptive lasso. In particular, we show that the adaptive lasso may combine efficient parameter estimation, variable selection, and valid finite sample inference in one step. To achieve this objective, we introduce some notation and terminology in line with Andrews and Guggenberger (2010). In particular, we first show that the limit distribution of the adaptive lasso is discontinuous in the tuning parameter $\lambda_n$. Then, we prove that with an appropriate selection of the critical values, adaptive lasso tests have the correct asymptotic size, where the asymptotic size is the limit of the exact size of the test, as defined in (5) below.

To this end, we slightly change our notation. In particular, we write $\hat{\theta}_{AL,\lambda_n}$ instead of $\hat{\theta}_{AL}$ to point out that $\hat{\theta}_{AL,\lambda_n}$ depends on the tuning parameter $\lambda_n$. Let $0 \leq \lambda_n < \infty$ and for $i = 1, \dots, p$, let $0 \leq \lambda_{0,i} < \infty$ denote the limit of $\lambda_n|\hat{\lambda}_{n,i}|/\sqrt{n}$, i.e.,

$$\frac{\lambda_n|\hat{\lambda}_{n,i}|}{\sqrt{n}} \to \lambda_{0,i}.$$

Note that for the non-zero coefficients, $\lambda^A_{0,i} = 0$. Then, in the next theorem we derive the limit distribution of $\sqrt{n}(\hat{\theta}_{AL,\lambda_n} - \theta^*)$.

**Theorem 4.1.** *Let $p_1 + p_2 + p_3 = p < \infty$. Assume that $\{Y_t\}$ and $\{Z_t\}$ are stationary processes such that $\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' = C$ exists and is of full rank, and $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \epsilon_t Z_t$ converges in distribution to a random variable with $N(0,\Omega)$ for some covariance matrix $\Omega$. Let $0 \leq \lambda_n = \lambda < \infty$. Then,*

$$\sqrt{n}(\hat{\theta}_{AL,\lambda_n} - \theta^*) \rightarrow arg\min(R),$$

*where*

$$R(u) = -2u'W + u'Cu + \sum_{i=1}^{p} \lambda_{0,i}|u_i|, \tag{3}$$

*and $W \sim N(0,\Omega)$.*

Note that since for the non-zero coefficients $\lambda_{0,i}^A = 0$, it turns out that in (3) only the zero coefficients are penalized. Furthermore, when $\lambda_n = \lambda = 0$, then for $i = 1,\ldots,p$, $\lambda_{0,i} = 0$, and therefore $arg\min(R) = C^{-1}W \sim N(0,V)$. Finally, when $\lambda_n \rightarrow \infty$, the estimates of the zero coefficients collapse to zero.

Using the result in Theorem 4.1 we can show that the adaptive lasso may provide valid finite sample inference. To this end, for $i = 1,\ldots,p$, let

$$T_{\lambda_n,i} = \sqrt{n}|\hat{\theta}_{AL,\lambda_n,i} - \theta_i^*|.$$

Furthermore, let $c_{\lambda_n,i,1-\alpha}$ denote the $1-\alpha$ quantile of the limit distribution of the statistic $T_{\lambda_n,i}$, where $\alpha \in (0,1)$. For instance, when $\lambda_n = 0$, then $c_{0,i,1-\alpha}$ is simply the $1-\alpha$ quantile of the random variable $|S|$, where $S \sim N(0,V_i)$, and $V_i$ is the $i$-th diagonal term of $V$. Using the result in Theorem 4.1, we can easily verify that for all $0 \leq \lambda_n < \infty$,

$$c_{0,i,1-\alpha} \geq c_{\lambda_n,i,1-\alpha}, \tag{4}$$

i.e., the $1-\alpha$ quantile $c_{\lambda_n,i,1-\alpha}$ is maximized at $\lambda_n = 0$.

To better understand this point, in Figure 1 below we plot the 0.95-quantiles of the distribution of the random variable $|u|$, where $u$ minimizes the function $R$ defined in (3).

[Figure 1 about here.]

Given the illustrative goal of the figure, we consider the simple case where $u \in \mathbb{R}$ and $C = \Omega = \mathbb{I}$, where $\mathbb{I}$ denotes the identity matrix. The horizontal solid line in Figure 1 represents the 0.95-quantile of the distribution of $|u|$ with $u \sim N(0,1)$, i.e, $\lambda_{0,1} = 0$ and $c_{0,1,0.95} = 1.96$. The dashed line represents instead the 0.95-quantiles of the random variable $|u|$ for different values of $\lambda_{0,1} \in [0,4]$ in equation (3). We can easily verify that the quantiles are indeed maximized by

$c_{0,1,0.95} = 1.96$. Then, they decrease almost linearly as $\lambda_{0,1}$ increases. Finally, when $\lambda_{0,1} = 4$ the 0.95-quantile is practically zero.

The result in (4) is the key condition for introducing valid inference with the adaptive lasso. Indeed, assume that we want to test the null hypothesis $H_{0,i} : \theta_i^* = \theta_{0i}^*$ versus the alternative $H_{1,i} : \theta_i^* \neq \theta_{0i}^*$, for some $\theta_{0i}^* \in \mathbb{R}$. Consider the test statistic $T_{\lambda_n,i}(\theta_{0i}^*) = \sqrt{n}|\hat{\theta}_{AL,\lambda_n,i} - \theta_{0i}^*|$, and the critical value $c_{0,i,1-\alpha}$. Following Andrews and Guggenberger (2010), we define the exact size and asymptotic size of the test of the null hypothesis $H_{0,i}$ as

$$\begin{aligned} ExSz_n(\theta_{0i}^*) &= \sup_{\lambda_n \in \mathbb{R}^+} P_{H_{0,i}}(T_{n,\lambda_n,i}(\theta_{0i}^*) > c_{0,i,1-\alpha}), \\ AsySz(\theta_{0i}^*) &= \limsup_{n\to\infty} ExSz_n(\theta_{0i}^*). \end{aligned} \tag{5}$$

Then, by (4) it follows trivially that $AsySz(\theta_{0i}^*) = \alpha$, i.e., the adaptive lasso test implies a correct asymptotic size. This result is summarized in the following corollary:

**Corollary 4.1.** *Let $p_1 + p_2 + p_3 = p < \infty$. Assume that $\{Y_t\}$ and $\{Z_t\}$ are stationary processes such that $\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n Z_t Z_t' = C$ exists and is of full rank, and $\frac{1}{\sqrt{n}} \sum_{t=1}^n \epsilon_t Z_t \to N(0, \Omega)$ for some covariance matrix $\Omega$. Let $0 \leq \lambda_n = \lambda < \infty$, and let $\hat{\theta}_{AL,\lambda_n}$ be the adaptive lasso estimate of $\theta^*$. Consider the test statistic $T_{\lambda_n,i}(\theta_{0i}^*) = \sqrt{n}|\hat{\theta}_{AL,\lambda_n,i} - \theta_{0i}^*|$, with the critical value $c_{0,i,1-\alpha}$. Then, the asymptotic size of the test of the null hypothesis $H_{0,i} : \theta_i^* = \theta_{0i}^*$ versus the alternative $H_{1,i} : \theta_i^* \neq \theta_{0i}^*$ satisfies*

$$AsySz(\theta_{0i}^*) = \alpha. \tag{6}$$

Using the result in Corollary 4.1 inference based on the adaptive lasso is straightforward. Indeed, to test the null hypothesis $H_{0,i}$ we can simply use the adaptive lasso statistic $T_{\lambda_n,i}(\theta_{0i}^*) = \sqrt{n}|\hat{\theta}_{AL,\lambda_n,i} - \theta_{0i}^*|$ with the normal critical value $c_{0,i,1-\alpha}$. This results shows that the adaptive lasso can combine efficient parameter estimation, variable selection, as well as valid finite sample inference in one step.

# 5 Monte Carlo

In this section we use Monte Carlo simulations to study the accuracy of inference based on the adaptive lasso. In particular, we consider three different settings.

**Setting 1:** $p_1 = p_2 = p_3 = 5$ and $\epsilon_t \sim_{\text{iid}} N(0, 1)$.

We generate $N = 5000$ samples according to model (1) with $p_1 = p_2 = p_3 = 5$, $\rho_1^* = \gamma_1^* = \beta_1^* = 0.3$, $\rho_2^* = \gamma_2^* = \beta_2^* = 0.1$, and $\rho_i^* = \gamma_i^* = \beta_i^* = 0$, for $i = 3, 4, 5$. We consider Gaussian error terms $\epsilon_t \sim_{\text{iid}} N(0,1)$. Furthermore, for $i = 1,\dots,5$ and $t = 1,\dots,n$, let $W_{i,t} \sim_{\text{iid}} N(0,1)$ and $X_{i,t-1} \sim_{\text{iid}} N(0,1)$. The simulated sample sizes are $n = 800$ and $n = 1600$.

In a first exercise, we study the accuracy of the inference for the active parameters. More precisely, using the results in Theorem 3.1, we construct 0.95-confidence intervals for the active parameters $\rho_1^* = \gamma_1^* = \beta_1^* = 0.3$, and $\rho_2^* = \gamma_2^* = \beta_2^* = 0.1$, where the tuning parameter $\lambda_n$ is selected according to the Bayesian Schwartz Information Criterion (BIC). The empirical coverages are summarized in Table 1, Panel A. In the first part of Table 1 Panel A we apply the results in Theorem 3.1 without the bias term $\hat{b}^A_{AL}$. In contrast, in the bottom part we use instead the bias-corrected limit distribution.

[Table 1 about here.]

In the top panel we note that the adaptive lasso without a correction term provides valid inference for the active parameters $\rho_1^* = \gamma_1^* = \beta_1^* = 0.3$ (see columns 2, 4, and 6). The empirical coverages are quite close to the nominal coverage probability 0.95. For instance, when $n = 1600$ the empirical coverages for $\rho_1^*$, $\gamma_1^*$ and $\beta_1^*$ are 0.9228, 0.9218 and 0.9222, respectively. In contrast, the adaptive lasso without correction term does not provide accurate inference for the active parameters $\rho_2^* = \gamma_2^* = \beta_2^* = 0.1$. In fact, the empirical coverages are quite far from the nominal coverage probability 0.95 (see columns 3, 5 and 7). For instance, when $n = 800$ the empirical coverages for $\rho_2^*$, $\gamma_2^*$ and $\beta_2^*$ are 0.8638, 0.8468 and 0.8536, respectively. Thus, the difference between empirical coverages and nominal coverage probability is approximately 0.1.

In the bottom panel of Table 1, Panel A, we note that the bias-corrected limit distribution substantially improves the accuracy of the adaptive lasso inference. The empirical coverages using the bias-corrected limit distribution are always closer to the nominal coverage probability than those computed without the bias term $\hat{b}^A_{AL}$. In particular, it is interesting to note that using the bias-corrected distribution the empirical coverages for the active parameters $\rho_2^* = \gamma_2^* = \beta_2^* = 0.1$ are very close to 0.95 as well. For instance, when $n = 800$ the empirical coverages for $\rho_2^*$, $\gamma_2^*$ and $\beta_2^*$ are 0.9346, 0.9322 and 0.9302, respectively. Thus, in this case the difference between empirical coverages and nominal coverage probability is always smaller than 0.02. These results show that the bias-corrected limit distribution is particularly important and useful for improving inference on the active parameters with small coefficients close to zero.

In a second exercise we study the finite sample power of the introduced (conservative) adaptive lasso test for the null hypothesis $H_{0,i} : \theta_i^* = 0$ versus the alternative $H_{1,i} : \theta_i^* \neq 0$, $i = 1,\dots,p$. Empirical frequencies of rejection of $H_{0,i}$ using the results in Corollary 4.1 with significance level $\alpha = 0.05$ are reported in Table 1, Panel B.

First, the table shows that for the active parameters $\rho_1^* = \gamma_1^* = \beta_1^* = 0.3$ (second column) we always reject the null hypothesis, for both $n = 800$ and $n = 1600$ sample sizes. Second, we note that for the active parameters $\rho_2^* = \gamma_2^* = \beta_2^* = 0.1$ (third column), the power of the adaptive lasso test significantly increases as $n$ increases. Thus, in case the coefficients of the active variables are small and close to zero, sufficient data is needed in order for the test to reach high power values. For instance, for $\rho_2^* = 0.1$ the empirical frequencies of rejection of the null hypothesis are 0.7014 and 0.9548 for $n = 800$ and $n = 1600$, respectively.

Finally, the empirical frequencies of rejection for the inactive parameters are in a range between 0.025 and 0.035 (see columns 4, 5 and 6). Ideally, the correct value in those cases should equal $\alpha = 0.05$, the size of the test. Nevertheless, our results are not surprising and there are two main factors that help explain the values (lower than $\alpha$) we find. First, it is important to remember that the adaptive lasso shrinks some of the coefficients (in particular those of the inactive variables) exactly to zero. In those cases no true asymptotic distribution exists, and therefore the number of non-rejections of the null hypothesis becomes larger. Second, the test we introduced is conservative per construction: this means that we expect to get fewer rejections of the null hypothesis than tests performed under ideal conditions.

**Setting 2:** $p_1 = 1$, $p_2 = p_3 = 10$ and $\epsilon_t \sim_{\text{iid}} N(0, 1)$.

In this second setting, we generate $N = 5000$ samples according to model (1) with $p_1 = 1$, $p_2 = p_3 = 10$, $\rho_1^* = 0.5$, $\gamma_1^* = \beta_1^* = 0.3$, $\gamma_2^* = \beta_2^* = 0.1$, and $\gamma_i^* = \beta_i^* = 0$, for $i = 3, \ldots, 10$. We consider Gaussian error terms $\epsilon_t \sim_{\text{iid}} N(0, 1)$. Furthermore, for $i = 1, \ldots, 10$ and $t = 1, \ldots, n$, let $W_{i,t} \sim_{\text{iid}} N(0, 1)$ and $X_{i,t-1} \sim_{\text{iid}} N(0, 1)$. The simulated sample sizes are $n = 800$ and $n = 1,600$. Note that this setting is very similar to the empirical analysis discussed in Section 6.

As in the previous setting, we perform the same two exercises. We focus first on the empirical coverages and we construct 0.95-confidence intervals for the active parameters $\rho_1^* = 0.5$, $\gamma_1^* = \beta_1^* = 0.3$, and $\gamma_2^* = \beta_2^* = 0.1$, where the tuning parameter $\lambda_n$ is selected according to BIC. In the second exercise we investigate the finite sample power of the adaptive lasso test for the null hypothesis $H_{0,i} : \theta_i^* = 0$ versus the alternative $H_{1,i} : \theta_i^* \neq 0$, $i = 1, \ldots, p$. The significance level is $\alpha = 0.05$. Table 2 summarize the results.

[Table 2 about here.]

Table 2, Panel A, summarizes the empirical coverages for the active parameters. Similarly to the previous setting, in this case as well we can observe that the adaptive lasso provides valid inference for the parameters of interest. The inference based on the bias-corrected limit distribution always outperforms one based on the limit distribution without correcting for the finite sample bias. For example, when $n = 1600$ the empirical coverages for $\gamma_2^* = 0.1$ and $\beta_2^* = 0.1$

using the bias-corrected limit distributions are 0.9288 and 0.9310, respectively. In contrast, the empirical coverages for $\gamma_2^* = 0.1$ and $\beta_2^* = 0.1$ without the bias term $\hat{b}_{AL}^A$ are 0.9128 and 0.9094, respectively.

Furthermore, again in this setting we show in Table 2, Panel B, that the adaptive lasso provides a valid statistical tool for testing the null hypothesis $H_{0,i} : \theta_i^* = 0$ versus the alternative $H_{1,i} : \theta_i^* \neq 0$, $i = 1, \ldots, p$. In particular, the adaptive lasso test always rejects the null hypothesis for $\gamma_1^* = \beta_1^* = 0.3$ when $n = 800$ and $n = 1600$ (see column 2); relatively high values of the power can also be seen for the other active parameters with small coefficients (in particular for $n = 1600$, see column 3); and the proportion of false rejections for the inactive parameters is in the range between 0.038 and 0.05, close to the level of the test $\alpha$.

**Setting 3:** $p_1 = p_2 = p_3 = 5$, $\epsilon_t \sim_{\text{iid}} t_5$ or/and GARCH error terms.

In this final setting we want to test the accuracy of the introduced procedure when dealing with a different error distribution with heavier tails or/and allowing for heteroscedasticity. For this purpose, we generate $N = 5000$ samples according to model (1) with $p_1 = p_2 = p_3 = 5$, $\rho_1^* = \gamma_1^* = \beta_1^* = 0.3$, $\rho_2^* = \gamma_2^* = \beta_2^* = 0.1$, and $\rho_i^* = \gamma_i^* = \beta_i^* = 0$, for $i = 3, 4, 5$. Furthermore, for $i = 1, \ldots, 5$ and $t = 1, \ldots, n$, let $W_{i,t} \sim_{\text{iid}} N(0,1)$ and $X_{i,t-1} \sim_{\text{iid}} N(0,1)$. The simulated sample sizes are $n = 800$ and $n = 1600$.

We consider two different distributions for the error terms. In the first case, we assume that $\epsilon_t \sim_{\text{iid}} t_5$. In the second case we assume instead the following GARCH representation

$$\begin{aligned} \epsilon_t &= \sqrt{h_t} e_t, \\ h_t &= 0.1 + 0.7 h_{t-1} + 0.1 h_{t-1} e_{t-1}^2, \end{aligned}$$

where $e_t \sim_{\text{iid}} t_5$. As in the previous settings, we analyze both the empirical coverages of 0.95-confidence intervals for the active parameters and the finite sample power of the adaptive lasso test at level $\alpha = 0.05$ for the null hypothesis $H_{0,i} : \theta_i^* = 0$ versus the alternative $H_{1,i} : \theta_i^* \neq 0$, $i = 1, \ldots, p$. The tuning parameter $\lambda_n$ is selected according to BIC. Results are reported in Tables 3 and 4.

[Table 3 about here.]

[Table 4 about here.]

The results in Tables 3 and 4, Panel A, confirm that inference with the adaptive lasso based on the bias-corrected limit distribution substantially outperforms inference based on the limit distribution without finite sample bias correction. For instance, in Table 3 when $\epsilon_t \sim_{\text{iid}} t_5$ and

$n = 800$ the empirical coverages for $\rho_2^*$, $\gamma_2^*$ and $\beta_2^*$ using the bias-corrected limit distribution are 0.9388, 0.9212 and 0.9230, respectively. In contrast, the empirical coverages for $\rho_2^*$, $\gamma_2^*$ and $\beta_2^*$ without the bias term $\hat{b}_{AL}^A$ are 0.8704, 0.8532 and 0.8556, respectively. The adaptive lasso provides a valid statistical tool for testing the null hypothesis $H_{0,i} : \theta_i^* = 0$ in this final, more complex setting as well. Indeed, in Tables 3 and 4, Panel B, we show that the adaptive lasso test always rejects the null hypothesis for $\rho_1^* = \gamma_1^* = \beta_1^* = 0.3$ both when $\epsilon_t \sim_{\text{iid}} t_5$ and when allowing for GARCH error terms. As shown in these tables, the power of the test for small coefficients close to zero can be moderate, in particular when not enough data is available. Finally, the proportion of false rejections for the inactive parameters is once again in most cases around 0.03, close to the level $\alpha$ of the test.

In sum, the results we get from the Monte Carlo simulations in the different settings confirm that the adaptive lasso provides a valid approach for testing the null hypothesis $H_{0,i} : \theta_i^* = 0$, $i = 1, \ldots, p$.

# 6 Empirical Illustration

We consider the relation between the short-term interest rate and the state of the economy in a Taylor-type monetary policy model, i.e. a linear regression model for the short rate that has as possible regressor candidates all macroeconomic and financial variables such as inflation, unemployment, industrial production, or monetary variables. The macroeconomic data was downloaded from the Federal Reserve Bank of Philadelphia and is part of the database called *Real-Time Data Set for Macroeconomists*, which consists of vintages of the most relevant macroeconomic variables. In our study we use the vintage available at the end of 2013. The time period under consideration goes from January 1959 to December 2012, for a total of 648 monthly observations. We collected the data for 19 macroeconomic variables, including:

- *Price level indices*: Consumer Price Index (PCPI), Core Consumer Price Index (PCPIX), and Produced Price Index (PPPI);

- *Monetary and financial*: M1 Money Stock (M1), M2 Money Stock (M2), Monetary Base (BASEBASA), Total Reserves (TRBASA), Nonborrowed Reserves (NBRBASA), and Nonborrowed Reserves Plus Extended Credit (NBRECBASA);

- *Industrial production and capacity utilization*: Capacity Utilization Rate Manufacturing (CUM), Industrial Production Index Total (IPT), Industrial Production Index Manufacturing (IPM);

- *Housing*: Housing Starts (HSTARTS);

- *Labor market*: Nonfarm Payroll Employment (EMPLOY), Aggregate Weekly Hours Goods-Producing (HG), Civilian Labor Force (LFC), Participation Rate, Constructed (LFPART), Civilian Noninstitutional Population (POP), Unemployment Rate (RUC).

The dependent variable is the US 3-month short rate. Data was downloaded from different sources: from 1959 to the end of 1969 from MacCulloch, from 1970 to the end of 1981 from Fama-Bliss (CRSP), and the final period until the end of 2013 from the Board of Governors of the Federal Reserve System (3-Month Treasury constant maturity rate). Where needed, all variables were seasonally adjusted. To take into account the time series dynamics of the short rate, we included the first lagged short rate value as predictor in the regression. Thus, we have $p_1 = 1, p_2 = 19, p_3 = 0$ in model (1).

In Table 5, we report the adaptive lasso point estimates. The tuning parameter $\lambda_n$ is selected according to BIC.

[Table 5 about here.]

Applying Corollary 4.1 we test the null hypothesis $H_{0,i} : \theta_i^* = 0$, $i = 1, \ldots, 20$, (19 regressors and the first lagged short rate as predictors). As shown in Table 5, column 2, the only variables that are significantly different from zero (at all relevant significance levels) using the testing procedure for the adaptive lasso are the lagged short rate, the Producer Price Index, and the Unemployment Rate. It is interesting to see that there are other variables with adaptive lasso coefficients different from zero. Without the use of the testing procedure introduced in this study we would not have been able to classify them as false positives.

This result is not surprising. In fact, the predictors we found to be statistically significant and to belong to the active set of variables identified by the adaptive lasso procedure are those also commonly thought to be economically relevant in the Taylor rule monetary policy model for the short rate. According to this rule, the central bank sets the nominal short-term interest rate, $r_t$, based on the following equation

$$r_t = \gamma_0 + \rho r_{t-1} + \gamma_\pi \pi_t + \gamma_g g_t + \varepsilon_t^r,$$

where $\pi_t$ denotes inflation, $g_t$ is the output gap, and $\varepsilon_t^r$ is a sequence of independent and normally distributed innovations with mean zero and variance $\sigma_r^2$. Thus, our result adds a purely statistical foundation (from the viewpoint of variable choice in the regression) to this economically intuitive rule. Moreover, the sign and (partially) the magnitude of the coefficients of the active variables are in line with the literature, that is a positive relation between inflation and the short rate, a negative relation between unemployment and the short rate, and a high persistence of the short rate dynamics.[4]

It is important to highlight that this result would not have been possible without the theory for testing null hypotheses of the type $H_{0,i} : \theta_i^* = \theta_{0i}^*$ versus the alternative $H_{1,i} : \theta_i^* \neq \theta_{0i}^*$, for some $\theta_{0i}^* \in \mathbb{R}$, developed in this study. In fact we would have found more active predictors using both the adaptive lasso and the classical full least squares estimates (whose results are summarized in column 4 of Table 5), completely losing the economic intuition behind the Taylor rule and rendering the interpretation of the results very difficult.

# 7 Conclusions

We presented new theoretical and empirical results on the finite sample and asymptotic properties of the adaptive lasso in time series regression models. We extended previous results presented in the literature along two main lines: (i) computing analytically a bias correction term for doing finite sample inference on the active variables in the adaptive lasso, and (ii) introducing a simple, conservative, but effective testing procedure for the null hypothesis that a parameter is equal to zero in the adaptive lasso model with a fixed amount of shrinkage.

Through extensive Monte Carlo simulations with a changing number of candidate variables, two different error distributions, and different sample sizes, we showed the accuracy of the testing procedure in finite sample. Testing our procedure in a more involved simulation experiment where we relaxed the assumption of iid errors, we also empirically confirmed the theoretical results and showed that the methodology is robust against this kind of deviation from the standard setting. This result is not surprising and confirms the recent findings and discussions in Medeiros and Mendes (2012) and Kock (2012).

Finally, we investigated the implications of the new testing procedure in an empirical application concerning the relation between the short-term interest rate dynamics and the (macro)economy. To this end we considered a Taylor rule monetary policy model, where we let the adaptive lasso choose from a number of macroeconomic and financial predictors the relevant ones to put in the active set. We then tested using the new procedure to see whether all remaining active variables had a corresponding coefficient significantly different from zero. In contrast with the full OLS on all variables, the only variables with a coefficient different than zero identified by our testing procedure were an inflation indicator, the unemployment rate, and the one-lagged past short rate. We interpreted this result as a statistical confirmation of the Taylor rule.

Our theoretical results are general and can be applied to a broad spectrum of iid and time

[4] See, for example, Filipova et al. (2013) and the references therein.

series applications, in particular when the researcher has to do variable selection and inference among many candidate variables such as realized volatility modeling, excess returns or inflation prediction. Moreover, in light of the theoretical results proved in this study, an alternative way of conducting finite sample inference can be envisaged. In the spirit of the recent works proposed by Chatterjee and Lahiri (2011) and Chatterjee and Lahiri (2013), we plan to develop bootstrap simulation techniques that can be applied to the (adaptive) lasso to perform finite sample testing of the resulting parameters. This is left for future research.

# References


Andrews, W.K., and P. Guggenberger (2010). Asymptotic size and a problem with subsampling and with the $m$ out of $n$ bootstrap. *Econometric Theory*, **26**, 426-468.

Ang, A., Beckaert, G., and Wei, W. (2008). The term structure of real rates and expected inflation. *Journal of Finance* **64**, 797-849.

Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* **50**, 745-787.

Audrino, F. and Knaus, S. (2012). Lassoing the HAR model: A model selection perspective on realized volatility dynamics. University of St. Gallen, SEPS Discussion paper series.

Bühlmann, P. and van de Geer, S. (2011). *Statistics for high-dimensional data: Methods, theory and applications*. Springer Series in Statistics, Springer, Heidelberg.

Chatterjee, A. and Lahiri, S.N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association* **106**, 608-625.

Chatterjee, A. and Lahiri, S.N. (2013). Rates of convergence of the Adaptive LASSO estimators to the Oracle distribution and higher order refinements by the bootstrap. *Annals of Statistics* **41**, 1232-1259.

Clarida, R., Gali, J., and Gertler, M. (2000). Monetary policy rules and macroeconomic stability: evidence and some theory. *The Quarterly Journal of Economics* **115**, 147-180.

Dewachter, H., and Lyrio, M. (2006). Macro factors and term structure of interest rates. *Journal of Money, Credit and Banking* **38**, 119-140.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348-1360.

Fan, J., and H. Peng (2004). On nonconcave penalized likelihood with diverging number of parameters. *The Annals of Statistics*, **32**, 928-961.

Filipova, K., Audrino, F., and De Giorgi, E. (2013). Monetary Policy Regimes: Implications for the Yield Curve and Bond Pricing. Working paper, University of St. Gallen.

Geyer, C. (1994). On the asymptotics of constrained M-estimation. *The Annals of Statistics*, **22**, 1993-2010.

Hsu, N., Hung, H. and Chang, Y. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis* **52**, 3645-3657.

Knight, K., and W. Fu (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics*, **28**, 1356-1378.

Kock, A.B. (2012). On the oracle property of the adaptive lasso in stationary and nonstationary autoregressions. CREATES Research Papers 2012-05, Aarhus University.

Kock, A.B., and Callot, L.A.F. (2012). Oracle inequalities for high dimensional vector autoregressions. CREATES Research Paper 2012-12, Aarhus University.

Javanmard, A. and Montanari, A. (2013). Hypothesis Testing in High-Dimensional Regression under the Gaussian Random Design Model: Asymptotic Theory. ArXiv e-prints.

Lockhart, T. Taylor, J., Tibshirani R.J., and Tibshirani R. (2013). A significance test for the lasso. ArXiv e-prints.

Medeiros, M.C., and E.F. Mendes (2012). Estimating high-dimensional time series models. CREATES Research Paper 2012-37, Aarhus University.

Moench, E. (2008). Forecasting the yield curve in a data-rich environment: a no-arbitrage factor-augmented VAR approach. *Journal of Econometrics* **146**, 26-43.

Nardi, Y., and Rinaldo, A. (2011). Autoregressive process modeling via the lasso procedure. *Journal of Multivariate Analysis* **102**, 528-549.

Park, H., and Sakaori, F. (2012). Lag weighted lasso for time series model. *Computational Statistics*, forthcoming.

Rudebusch, G. (2010). Macro-finance models of interest rates and the economy. *The Manchester School* **7**, 25-52.

Song, S., and Bickel, P.J. (2011). Large vector autoregressions (2011). ArXiv e-prints.

Taylor, J.B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy* **39**, 195-214.

Tibshirani, R. (1996). Regression analysis and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, **58**, 267-288.

Wang, H., Li, G. and Tsai, C. (2007). Regression coefficient and autoregressive order shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* **69**, 63-78.

Zhao, P. and Yu, B. (2006). On model consistency of lasso. *Journal of Machine Learning Research* **7**, 2541-2563.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101**, 1418-1429.

# Appendix: Proofs of the Theorems

**Proof of Theorem 3.1:** First we derive the limit distribution of the adaptive lasso. In particular, we adopt the same argument as in the proof of Theorem 2 in Zou (2006). Finally, we use this result to prove (I) Variable Selection and to compute the bias term in (II) Limit Distribution.

Let

$$R_n(u) = \sum_{t=1}^{n} \left[ (\epsilon_t - u' Z_t/\sqrt{n})^2 - \epsilon_t^2 \right] + \lambda_n \sum_{i=1}^{p} \lambda_{n,i} \left[ |\theta_i^* + u_i/\sqrt{n}| - |\theta_i^*| \right].$$

Note that $R_n$ is minimized at $\sqrt{n}(\hat{\theta}_{AL} - \theta^*)$. Furthermore, we know that

$$\sum_{t=1}^{n} \left[ (\epsilon_t - u' Z_t/\sqrt{n})^2 - \epsilon_t^2 \right] \to -2u'W + u'Cu,$$

where $W \sim N(0, \Omega)$. Now, consider the limit of the second term $\lambda_n \sum_{i=1}^{p} \lambda_{n,i} \left[ |\theta_i^* + u_i/\sqrt{n}| - |\theta_i^*| \right]$. If $\theta_i^* \neq 0$, then $\lambda_{n,i} \to |\theta_i^*|^{-1}$, and consequently $\lambda_n \lambda_{n,i} \left[ |\theta_i^* + u_i/\sqrt{n}| - |\theta_i^*| \right] \to 0$. If $\theta_i^* = 0$, then $|\theta_i^* + u_i/\sqrt{n}| - |\theta_i^*| = u_i/\sqrt{n}$, and furthermore $\lambda_n \lambda_{n,i} = \lambda_n C$, where $C/\sqrt{n} = O_p(1)$. Let $R(u)$ denote the limit of $R_n(u)$. Then, we can conclude that

$$R(u) = \begin{cases} -2u_A' W^A + u_A' C^A u_A & \text{if } u_i = 0, \text{ for } i \notin A, \\ \infty & \text{otherwise,} \end{cases}$$

where $W^A \sim N(0, \Omega^A)$ and $\Omega^A$ is the sub-matrix of $\Omega$ for the non-zero coefficients. Note that $R_n$ is convex, and the unique minimum of $R$ is $((C^A)^{-1} W^A, 0)'$. Therefore, by Geyer (1994) it follows that

$$\begin{aligned} \sqrt{n} \left( \hat{\theta}_{AL}^A - \theta^{*A} \right) &\to N(0, V^A), \\ \sqrt{n} \hat{\theta}_{AL}^{A^c} &\to 0, \end{aligned}$$

where $\hat{\theta}_{AL}^{A^c}$ denotes the adaptive lasso estimate of the zero coefficients $\theta^{*A^c}$ of $\theta^*$.

Using this result, we can prove (I) Variable Selection. We adopt the same argument as in the proof of Lemma 5 in Fan and Peng (2004). Let

$$Q(\theta) = \frac{1}{n} \sum_{t=1}^{n} (Y_t - \theta' Z_t)^2 + \frac{\lambda_n}{n} \sum_{i=1}^{p} \lambda_{n,i} |\theta_i|.$$

With some abuse of notation, we write $\theta^* = (\theta'^{*A}, \theta'^{*A^c})'$. We show that with probability tending to 1, for any $\hat{\theta}^A$ satisfying $\|\hat{\theta}^A - \theta^{*A}\| = O_p(1/\sqrt{n})$ and any constant $C$,

$$Q((\hat{\theta}'^A, 0')') = min_{\|\hat{\theta}^{A^c}\| \leq C/\sqrt{n}} Q((\hat{\theta}'^A, \hat{\theta}'^{A^c})').$$

To this end, for $j \notin A$ consider

$$\begin{aligned}\frac{\partial Q(\theta)}{\partial \theta_j} &= -\frac{2}{n}\sum_{i=1}^{n}(Y_t - \theta' Z_t)Z_t^{(j)} + \frac{\lambda_n}{n}\lambda_{n,j} sign(\theta_j) \\ &= J_1 + J_2,\end{aligned}$$

where $Z_t^{(j)}$ denotes the $j$-component of the vector $Z_t$. Note that $J_1 = O_p(1/\sqrt{n})$, while the dominant term is $J_2$, since (i) $\lambda_n \to +\infty$ and (ii) $\frac{\lambda_n}{\sqrt{n}} \to 0$. Thus, the sign of $\theta_j$ determines the sign of $\frac{\partial Q(\theta)}{\partial \theta_j}$. More precisely, we have

$$\begin{aligned}&\tfrac{\partial Q(\theta)}{\partial \theta_j} < 0, \quad \text{when } -\epsilon_n < \theta_j < 0. \\ &\tfrac{\partial Q(\theta)}{\partial \theta_j} > 0, \quad \text{when } 0 < \theta_j < \epsilon_n.\end{aligned}$$

This concludes the proof of (I) Variable Selection.

Finally, using these results, we can focus on (II) Limit Distribution and also derive the bias term. Note that for $n$ large enough, for $j \in A$ we have

$$\frac{\partial Q(\theta)}{\partial \theta_j} = -\frac{2}{n}\sum_{i=1}^{n}(Y_t - \hat{\theta}'_{AL} Z_t)Z_t^{(j)} + \frac{\lambda_n}{n}\lambda_{n,j} sign(\hat{\theta}_{AL,j}) = 0. \tag{7}$$

Furthermore, for $n$ large enough $\hat{\theta}_{AL,j} = 0$ for $j \notin A$. Thus, we can rewrite the $q$ equations (7) in matrix form

$$0 = \frac{2}{n}\sum_{i=1}^{n}(Y_t - \hat{\theta}'^{A}_{AL} Z_t^A)Z_t^A - \Lambda_{AL}^A, \tag{8}$$

where $\Lambda_{AL}^A = (\frac{\lambda_n}{n}\lambda_{n,1} sign(\hat{\theta}^A_{AL,1}), \ldots, \frac{\lambda_n}{n}\lambda_{n,q} sign(\hat{\theta}^A_{AL,q}))'$. Now consider the term $\frac{1}{n}\sum_{i=1}^{n}(Y_t - \hat{\theta}'_{AL} Z_t)Z_t$. A Taylor expansion around $\theta^*$ yields

$$\frac{1}{n}\sum_{i=1}^{n}(Y_t - \hat{\theta}'_{AL} Z_t)Z_t = \frac{1}{n}\sum_{i=1}^{n}(Y_t - \theta'^{*} Z_t)Z_t - \frac{1}{n}\sum_{i=1}^{n} Z_t Z_t'(\hat{\theta}_{AL} - \theta^*). \tag{9}$$

Again, since $\hat{\theta}_{AL,j} = \theta_j^* = 0$ for $j \notin A$ and $n$ large enough, from (9) it turns out that

$$\frac{1}{n}\sum_{i=1}^{n}(Y_t - \hat{\theta}'^{A}_{AL} Z_t^A)Z_t^A = \frac{1}{n}\sum_{i=1}^{n}(Y_t - \theta'^{*A} Z_t^A)Z_t^A - \frac{1}{n}\sum_{i=1}^{n} Z_t^A Z_t'^{A}(\hat{\theta}^A_{AL} - \theta^{*A}). \tag{10}$$

Therefore, by combining (8) and (10) we have

$$0 = \frac{2}{n}\sum_{i=1}^{n}(Y_t - \theta'^{*A} Z_t^A)Z_t^A - \frac{2}{n}\sum_{i=1}^{n} Z_t^A Z_t'^{A}(\hat{\theta}^A_{AL} - \theta^{*A}) - \Lambda_{AL}^A,$$

i.e.,

$$\sqrt{n}(\hat{\theta}^A_{AL} - \theta^{*A}) = \left(\frac{1}{n}\sum_{i=1}^{n} Z_t^A Z_t'^{A}\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_t - \theta'^{*A} Z_t^A)Z_t^A - \frac{\sqrt{n}}{2}\Lambda_{AL}^A\right).$$

Since $\frac{\lambda_n}{\sqrt{n}} \to 0$, it turns out that for $j \in A$, $\frac{\lambda_n}{\sqrt{n}}\lambda_{n,j} sign(\hat{\theta}_{AL,j}) \to 0$. Therefore,

$$\sqrt{n}\left(\hat{\theta}^A_{AL} - \theta^{*A}\right) + \hat{b}^A_{AL} \to N(0, V^A),$$

where the bias term is given by

$$\hat{b}^A_{AL} = \left(\frac{1}{n}\sum_{i=1}^{n} Z^A_t Z'^A_t\right)^{-1} \left(\frac{\lambda_n}{2\sqrt{n}}\lambda^A_{n,1} sign(\hat{\theta}^A_{AL,1}), \ldots, \frac{\lambda_n}{2\sqrt{n}}\lambda^A_{n,q} sign(\hat{\theta}^A_{AL,q})\right)',$$

This concludes the proof.

**Proof of Theorem 4.1:** To prove Theorem 4.1, we use the same arguments as in the proof of Theorem 2 in Knight and Fu (2000). More precisely, let

$$R_n(u) = \sum_{t=1}^{n}\left[(\epsilon_t - u'Z_t/\sqrt{n})^2 - \epsilon_t^2\right] + \lambda_n \sum_{i=1}^{p} \lambda_{n,i}\left[|\theta^*_i + u_i/\sqrt{n}| - |\theta^*_i|\right].$$

Note that $R_n$ is minimized at $\sqrt{n}(\hat{\theta}_{AL,\lambda_n} - \theta^*)$. Furthermore,

$$\begin{aligned}
\sum_{t=1}^{n}\left[(\epsilon_t - u'Z_t/\sqrt{n})^2 - \epsilon_t^2\right] &\to -2u'W + u'Cu, \\
\lambda_n \sum_{i=1}^{p+r} \lambda_{n,i}\left[|\theta^*_i + u_i/\sqrt{n}| - |\theta^*_i|\right] &\to \sum_{i=1}^{p} \lambda_{0,i}|u_i|.
\end{aligned}$$

Thus, $R_n(u) \to R(u)$ as $n \to \infty$. Since $R_n$ is convex and $R$ has a unique minimum, it follows from Geyer (1994) that

$$arg\min(R_n) = \sqrt{n}(\hat{\theta}_{AL,\lambda_n} - \theta^*) \to arg\min(R).$$

This concludes the proof.

**Monte Carlo Simulations: Setting 1**

Panel A: Empirical coverages for the active parameters

| | $\rho_1^* = 0.3$ | $\rho_2^* = 0.1$ | $\gamma_1^* = 0.3$ | $\gamma_2^* = 0.1$ | $\beta_1^* = 0.3$ | $\beta_2^* = 0.1$ |
|---|---|---|---|---|---|---|
| $n = 800$ | 0.9134 | 0.8638 | 0.9100 | 0.8468 | 0.9146 | 0.8536 |
| $n = 1600$ | 0.9228 | 0.8832 | 0.9218 | 0.8890 | 0.9222 | 0.8924 |

| | $\rho_1^* = 0.3$ | $\rho_2^* = 0.1$ | $\gamma_1^* = 0.3$ | $\gamma_2^* = 0.1$ | $\beta_1^* = 0.3$ | $\beta_2^* = 0.1$ |
|---|---|---|---|---|---|---|
| $n = 800$ | 0.9224 | 0.9346 | 0.9172 | 0.9322 | 0.9210 | 0.9302 |
| $n = 1600$ | 0.9288 | 0.9358 | 0.9280 | 0.9332 | 0.9254 | 0.9342 |

Panel B: Empirical frequencies of rejection of the null hypothesis $H_0$

| | $\rho_1^* = 0.3$ | $\rho_2^* = 0.1$ | $\rho_3^* = 0$ | $\rho_4^* = 0$ | $\rho_5^* = 0$ |
|---|---|---|---|---|---|
| $n = 800$ | 1.0000 | 0.7014 | 0.0272 | 0.0246 | 0.0292 |
| $n = 1600$ | 1.0000 | 0.9548 | 0.0252 | 0.0234 | 0.0294 |

| | $\gamma_1^* = 0.3$ | $\gamma_2^* = 0.1$ | $\gamma_3^* = 0$ | $\gamma_4^* = 0$ | $\gamma_5^* = 0$ |
|---|---|---|---|---|---|
| $n = 800$ | 1.0000 | 0.6476 | 0.0326 | 0.0310 | 0.0254 |
| $n = 1600$ | 1.0000 | 0.9168 | 0.0312 | 0.0260 | 0.0242 |

| | $\beta_1^* = 0.3$ | $\beta_2^* = 0.1$ | $\beta_3^* = 0$ | $\beta_4^* = 0$ | $\beta_5^* = 0$ |
|---|---|---|---|---|---|
| $n = 800$ | 1.0000 | 0.6390 | 0.0356 | 0.0280 | 0.0296 |
| $n = 1600$ | 1.0000 | 0.9236 | 0.0260 | 0.0290 | 0.0224 |

Table 1: The data is generated according to model (1) with Gaussian innovations. The simulated sample sizes are $n = 800$ and $n = 1600$ and the results are based on $N = 5000$ simulations. In Panel A we report the empirical coverages of 0.95-confidence intervals for the active parameters $\rho_1^* = \gamma_1^* = \beta_1^* = 0.3$ and $\rho_2^* = \gamma_2^* = \beta_2^* = 0.1$. In the top panel we apply the results in Theorem 3.1 without the bias term $\hat{b}_{AL}^A$. In the bottom panel, we use instead the bias-corrected limit distribution. In Panel B we summarize the empirical frequencies of rejection of the null hypothesis $H_{0,i} : \theta_i^* = 0$ versus the alternative $H_{1,i} : \theta_i^* \neq 0$, $i = 1, \ldots, p$, using the results in Corollary 4.1 with significance level $\alpha = 0.05$. In the top, second, and bottom panels, we consider $\rho_i^*$, $\gamma_i^*$ and $\beta_i^*$, $i = 1, \ldots, 5$, respectively.

**Monte Carlo Simulations: Setting 2**

Panel A: Empirical coverages for the active parameters

| | $\rho_1^* = 0.5$ | $\gamma_1^* = 0.3$ | $\gamma_2^* = 0.1$ | $\beta_1^* = 0.3$ | $\beta_2^* = 0.1$ |
|---|---|---|---|---|---|
| $n = 800$ | 0.9174 | 0.9034 | 0.8920 | 0.9086 | 0.8876 |
| $n = 1600$ | 0.9192 | 0.9158 | 0.9128 | 0.9144 | 0.9094 |

| | $\rho_1^* = 0.5$ | $\gamma_1^* = 0.3$ | $\gamma_2^* = 0.1$ | $\beta_1^* = 0.3$ | $\beta_2^* = 0.1$ |
|---|---|---|---|---|---|
| $n = 800$ | 0.9260 | 0.9172 | 0.9298 | 0.9156 | 0.9254 |
| $n = 1600$ | 0.9294 | 0.9324 | 0.9288 | 0.9262 | 0.9310 |

Panel B: Empirical frequencies of rejection of the null hypothesis $H_0$

| | $\gamma_1^* = 0.3$ | $\gamma_2^* = 0.1$ | $\gamma_3^* = 0$ | $\gamma_4^* = 0$ | $\gamma_5^* = 0$ |
|---|---|---|---|---|---|
| $n = 800$ | 1.0000 | 0.7226 | 0.0432 | 0.0382 | 0.0456 |
| $n = 1600$ | 1.0000 | 0.9474 | 0.0390 | 0.0404 | 0.0422 |
| | $\gamma_6^* = 0$ | $\gamma_7^* = 0$ | $\gamma_8^* = 0$ | $\gamma_9^* = 0$ | $\gamma_{10}^* = 0$ |
| $n = 800$ | 0.0482 | 0.0448 | 0.0432 | 0.0470 | 0.0426 |
| $n = 1600$ | 0.0454 | 0.0432 | 0.0380 | 0.0454 | 0.0412 |

| | $\beta_1^* = 0.3$ | $\beta_2^* = 0.1$ | $\beta_3^* = 0$ | $\beta_4^* = 0$ | $\beta_5^* = 0$ |
|---|---|---|---|---|---|
| $n = 800$ | 1.0000 | 0.7124 | 0.0492 | 0.0468 | 0.0422 |
| $n = 1600$ | 1.0000 | 0.9492 | 0.0468 | 0.0382 | 0.0404 |
| | $\beta_6^* = 0$ | $\beta_7^* = 0$ | $\beta_8^* = 0$ | $\beta_9^* = 0$ | $\beta_{10}^* = 0$ |
| $n = 800$ | 0.0388 | 0.0438 | 0.0482 | 0.0462 | 0.0428 |
| $n = 1600$ | 0.0342 | 0.0412 | 0.0422 | 0.0410 | 0.0380 |

Table 2: The data is generated according to model (1) with Gaussian innovations. The simulated sample sizes are $n = 800$ and $n = 1600$ and the results are based on $N = 5000$ simulations. In Panel A we report the empirical coverages of 0.95-confidence intervals for the active parameters $\rho_1^* = 0.5$, $\gamma_1^* = \beta_1^* = 0.3$ and $\gamma_2^* = \beta_2^* = 0.1$. In the top panel we apply the results in Theorem 3.1 without the bias term $\hat{b}_{AL}^A$. In the bottom panel, we use instead the bias-corrected limit distribution. In Panel B we summarize the empirical frequencies of rejection of the null hypothesis $H_{0,i} : \theta_i^* = 0$ versus the alternative $H_{1,i} : \theta_i^* \neq 0$, $i = 1, \ldots, p$, using the results in Corollary 4.1 with significance level $\alpha = 0.05$. In the top and bottom panels, we consider $\gamma_i^*$ and $\beta_i^*$, $i = 1, \ldots, 10$, respectively.

**Monte Carlo Simulations: Setting 3**

Panel A: Empirical coverages for the active parameters

| | $\rho_1^* = 0.3$ | $\rho_2^* = 0.1$ | $\gamma_1^* = 0.3$ | $\gamma_2^* = 0.1$ | $\beta_1^* = 0.3$ | $\beta_2^* = 0.1$ |
|---|---|---|---|---|---|---|
| $n = 800$ | 0.9340 | 0.8704 | 0.9322 | 0.8532 | 0.9268 | 0.8556 |
| $n = 1600$ | 0.9370 | 0.9164 | 0.9340 | 0.8974 | 0.9290 | 0.8996 |

| | $\rho_1^* = 0.3$ | $\rho_2^* = 0.1$ | $\gamma_1^* = 0.3$ | $\gamma_2^* = 0.1$ | $\beta_1^* = 0.3$ | $\beta_2^* = 0.1$ |
|---|---|---|---|---|---|---|
| $n = 800$ | 0.9366 | 0.9388 | 0.9340 | 0.9212 | 0.9284 | 0.9230 |
| $n = 1600$ | 0.9422 | 0.9412 | 0.9348 | 0.9462 | 0.9334 | 0.9478 |

Panel B: Empirical frequencies of rejection of the null hypothesis $H_0$

| | $\rho_1^* = 0.3$ | $\rho_2^* = 0.1$ | $\rho_3^* = 0$ | $\rho_4^* = 0$ | $\rho_5^* = 0$ |
|---|---|---|---|---|---|
| $n = 800$ | 1.0000 | 0.6864 | 0.0256 | 0.0148 | 0.0272 |
| $n = 1600$ | 1.0000 | 0.9554 | 0.0248 | 0.0204 | 0.0222 |

| | $\gamma_1^* = 0.3$ | $\gamma_2^* = 0.1$ | $\gamma_3^* = 0$ | $\gamma_4^* = 0$ | $\gamma_5^* = 0$ |
|---|---|---|---|---|---|
| $n = 800$ | 1.0000 | 0.4612 | 0.0190 | 0.0272 | 0.0264 |
| $n = 1600$ | 1.0000 | 0.7914 | 0.0212 | 0.0238 | 0.0260 |

| | $\beta_1^* = 0.3$ | $\beta_2^* = 0.1$ | $\beta_3^* = 0$ | $\beta_4^* = 0$ | $\beta_5^* = 0$ |
|---|---|---|---|---|---|
| $n = 800$ | 1.0000 | 0.4554 | 0.0314 | 0.0238 | 0.0272 |
| $n = 1600$ | 1.0000 | 0.7894 | 0.0286 | 0.0264 | 0.0246 |

Table 3: The data is generated according to model (1) with $\epsilon_t \sim_{\text{iid}} t_5$. The simulated sample sizes are $n = 800$ and $n = 1600$ and the results are based on $N = 5000$ simulations. In Panel A we report the empirical coverages of 0.95-confidence intervals for the active parameters $\rho_1^* = \gamma_1^* = \beta_1^* = 0.3$ and $\rho_2^* = \gamma_2^* = \beta_2^* = 0.1$. In the top panel we apply the results in Theorem 3.1 without the bias term $\hat{b}_{AL}^A$. In the bottom panel, we use instead the bias-corrected limit distribution. In Panel B we summarize the empirical frequencies of rejection of the null hypothesis $H_{0,i} : \theta_i^* = 0$ versus the alternative $H_{1,i} : \theta_i^* \neq 0$, $i = 1, \ldots, p$, using the results in Corollary 4.1 with significance level $\alpha = 0.05$. In the top, second, and bottom panels, we consider $\rho_i^*$, $\gamma_i^*$ and $\beta_i^*$, $i = 1, \ldots, 5$, respectively.

## Monte Carlo Simulations: Setting 3

Panel A: Empirical coverages for the active parameters

| | $\rho_1^* = 0.3$ | $\rho_2^* = 0.1$ | $\gamma_1^* = 0.3$ | $\gamma_2^* = 0.1$ | $\beta_1^* = 0.3$ | $\beta_2^* = 0.1$ |
|---|---|---|---|---|---|---|
| $n = 800$ | 0.9152 | 0.8892 | 0.9092 | 0.8476 | 0.9136 | 0.8520 |
| $n = 1600$ | 0.9240 | 0.9114 | 0.9168 | 0.8834 | 0.9210 | 0.8858 |

| | $\rho_1^* = 0.3$ | $\rho_2^* = 0.1$ | $\gamma_1^* = 0.3$ | $\gamma_2^* = 0.1$ | $\beta_1^* = 0.3$ | $\beta_2^* = 0.1$ |
|---|---|---|---|---|---|---|
| $n = 800$ | 0.9184 | 0.9148 | 0.9168 | 0.9124 | 0.9212 | 0.9168 |
| $n = 1600$ | 0.9280 | 0.9410 | 0.9242 | 0.9358 | 0.9260 | 0.9388 |

Panel B: Empirical frequencies of rejection of the null hypothesis $H_0$

| | $\rho_1^* = 0.3$ | $\rho_2^* = 0.1$ | $\rho_3^* = 0$ | $\rho_4^* = 0$ | $\rho_5^* = 0$ |
|---|---|---|---|---|---|
| $n = 800$ | 1.0000 | 0.5008 | 0.0318 | 0.0380 | 0.0364 |
| $n = 1600$ | 1.0000 | 0.7896 | 0.0344 | 0.0278 | 0.0324 |

| | $\gamma_1^* = 0.3$ | $\gamma_2^* = 0.1$ | $\gamma_3^* = 0$ | $\gamma_4^* = 0$ | $\gamma_5^* = 0$ |
|---|---|---|---|---|---|
| $n = 800$ | 1.0000 | 0.5854 | 0.0330 | 0.0326 | 0.0312 |
| $n = 1600$ | 1.0000 | 0.8524 | 0.0248 | 0.0286 | 0.0224 |

| | $\beta_1^* = 0.3$ | $\beta_2^* = 0.1$ | $\beta_3^* = 0$ | $\beta_4^* = 0$ | $\beta_5^* = 0$ |
|---|---|---|---|---|---|
| $n = 800$ | 1.0000 | 0.5782 | 0.0328 | 0.0298 | 0.0270 |
| $n = 1600$ | 1.0000 | 0.8602 | 0.0296 | 0.0254 | 0.0248 |

Table 4: The data is generated according to model (1) with the error term following a GARCH process with a $t_5$ distributed innovation. The simulated sample sizes are $n = 800$ and $n = 1600$ and the results are based on $N = 5000$ simulations. In Panel A we report the empirical coverages of 0.95-confidence intervals for the active parameters $\rho_1^* = \gamma_1^* = \beta_1^* = 0.3$ and $\rho_2^* = \gamma_2^* = \beta_2^* = 0.1$. In the top panel we apply the results in Theorem 3.1 without the bias term $\hat{b}_{AL}^A$. In the bottom panel, we use instead the bias-corrected limit distribution. In Panel B we summarize the empirical frequencies of rejection of the null hypothesis $H_{0,i} : \theta_i^* = 0$ versus the alternative $H_{1,i} : \theta_i^* \neq 0$, $i = 1, \ldots, p$, using the results in Corollary 4.1 with significance level $\alpha = 0.05$. In the top, second and bottom panels, we consider $\rho_i^*$, $\gamma_i^*$ and $\beta_i^*$, $i = 1, \ldots, 5$, respectively.

| Variables | AL Estimate | Standard Errors | LS Estimate |
|---|---|---|---|
| US 3-month one-lag | 0.8937*** | 0.0319 | 0.9102*** |
| Consumer Price Index | −0.0604 | 0.0443 | −0.0618 |
| Core Consumer Price Index | 0 | 0.0246 | 0.0185 |
| Produced Price Index | 0.0615*** | 0.0164 | 0.0657*** |
| M1 Money Stock | 0 | 0.0003 | −0.0007** |
| M2 Money Stock | 0 | 0.0002 | −0.0007*** |
| Monetary Base | 0.0017 | 0.0020 | 0.0062*** |
| Total Reserves | −0.0025 | 0.0022 | −0.0057*** |
| Nonborrowed Reserves | 0.0154 | 0.0429 | 0.0209 |
| Nonborrowed Reserves Plus Extended Credit | −0.0155 | 0.0429 | −0.0210 |
| Capacity Utilization Rate Manufacturing | 0 | 0.0167 | −0.0108 |
| Industrial Production Index Total | 0.0550 | 0.0678 | 0.0771 |
| Industrial Production Index Manufacturing | −0.0558 | 0.0568 | −0.0845 |
| Housing Starts | 0 | 0.0001 | 0.0002** |
| Nonfarm Payroll Employment | 0 | 0.0001 | −0.0001 |
| Aggregate Weekly Hours Goods- Producing | 0 | 0.0166 | 0.0108 |
| Civilian Labor Force | 0 | 0.0001 | 0.0001 |
| Participation Rate, Constructed | −0.0277 | 0.0280 | −0.0311 |
| Civilian Noninstitutional Population | 0 | 0.0001 | −0.0001 |
| Unemployment Rate | −0.2626*** | 0.0892 | −0.2812*** |

Table 5: **Taylor rule monetary policy model for the short rate: Adaptive lasso point estimation and inference.** We report the adaptive lasso estimates (column 2), the standard errors (column 3) and the full least squares estimates (column 4) for the empirical analysis introduced in Section 6. The dependent variable is the US 3-month short rate. The period under investigation ranges from January 1959 to December 2012, for a total of 648 monthly observations. Asterisks * ,** ,*** denote significance at the 10%, 5% and 1% level, respectively.

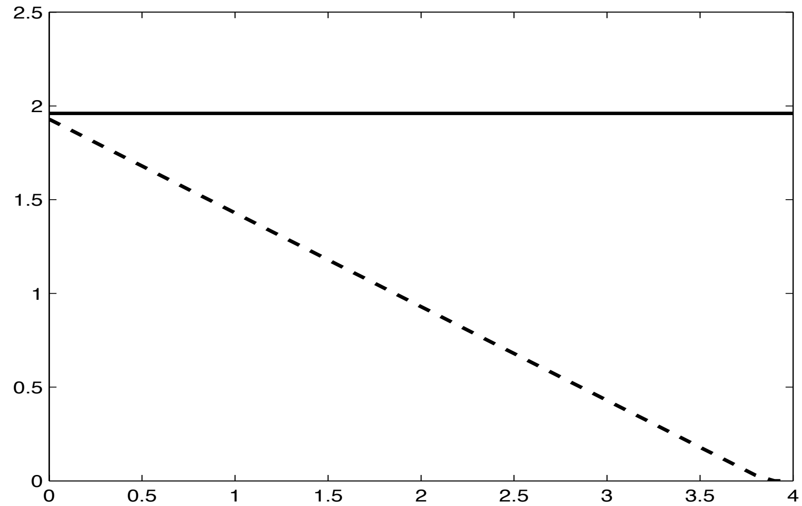


Figure 1: **Quantiles Adaptive Lasso Estimates.** We plot the 0.95-quantiles of the distribution of the random variable $|u|$, where $u$ minimizes the function $R$ defined in (3). The horizontal solid line represents the 0.95-quantile of the distribution of $|u|$ with $u \sim N(0,1)$. The dashed line represents instead the 0.95-quantiles of the random variable $|u|$ for different values of $\lambda_{0,1} \in [0,4]$.