TEXTO PARA DISCUSSÃO

No. 637

Adaptive LASSO estimation for ARDL models with GARCH innovations

Marcelo C. Medeiros
Eduardo F. Mendes

DEPARTAMENTO DE ECONOMIA
www.econ.puc-rio.br

ADAPTIVE LASSO ESTIMATION FOR ARDL MODELS WITH GARCH INNOVATIONS

Marcelo C. Medeiros
Department of Economics
Pontifical Catholic University of Rio de Janeiro
Rua Marquês de São Vicente 225, Gávea
Rio de Janeiro, RJ, 22451-900, BRAZIL
E-mail: `mcm@econ.puc-rio.br`

Eduardo F. Mendes
School of Applied Mathematics
Fundação Getulio Vargas
Praia de Botafogo, 190, Botafogo
Rio de Janeiro, RJ, 22250-900, BRAZIL
E-mail: `eduardo.mendes@fgv.br`

JEL: C22.

Keywords: ARDL, GARCH, sparse models, shrinkage, LASSO, adaLASSO, time series.

Acknowledgements: M. C. Medeiros acknowledges partial support from CNPq/Brazil. E. F. Mendes was partially supported by the Australian Center of Excellence Grant CE140100049. We are indebt to Gabriel Vasconcelos for superb research assistance.

Abstract: In this paper we show the validity of the adaptive LASSO procedure in estimating stationary ARDL(p,q) models with innovations in a broad class of conditionally heteroskedastic models. We show that the adaptive Lasso selects the relevant variables with probability converging to one and that the estimator is oracle efficient, meaning that its distribution converges to the same distribution of the oracle assisted least squares, i.e., the least squares estimator calculated as if we knew the set of relevant variables beforehand. Finally, we show that the LASSO estimator can be used to construct the initial weights. The performance of the method in finite samples is illustrated using Monte Carlo simulation.

1. Introduction

We consider the problem of estimating linear autoregressive distributed lag (ARDL) models with non-Gaussian GARCH errors when the number of regressors is possibly larger than the sample size, but only a finite and small number of regressors is relevant (sparsity). We focus on the adaptive Least Absolute Shrinkage and Selection Operator (adaLASSO) proposed by Zou (2006) as a generalization of the LASSO procedure pioneered by Tibshirani (1996).

Medeiros and Mendes (2016) put forward a number of sufficient high-level conditions for the adaLASSO to be model selection consistent and oracle efficient in linear time-series models with heteroskedastic and non-Gaussian errors. In this paper, we adapt the Medeiros and Mendes'(2015) conditions to the case of ARDL models with non-Gaussian GARCH errors. By model selection consistency we mean that the correct set of regressors are selected as the number of observations diverges to infinity. The oracle property means that the adaLASSO estimator has the same asymptotic distribution as the ordinary least squares (OLS) estimator under the knowledge of the correct set of relevant regressors (Fan and Li, 2001). Since our results are asymptotic, the high-dimension is understood as the number of candidate covariates increases polynomially with the sample size.

Estimation procedures that rely on shrinking the parameters towards zero have been receiving increasing attention in the times-series literature; see, for example, Nardi and Rinaldo (2011), Kock and Callot (2015), Liao and Phillips (2015), Basu and Michailidis (2015), among many others. The reason for the recent popularity of such methods is mainly due their ability to handle situations where the number of parameters to be estimated is larger than the available sample size.

The paper is organized as follows. In Section 2 we present the model and the assumptions. The key results of the paper are shown in Section 3. A Monte Carlo simulation is presented in Section 4. Finally, Section 5 concludes the paper. All mathematical proofs are relegated to the appendix.

2. Model and Main Assumptions

We consider a stochastic process $\{y_t\}$ generated from

$$y_t = \sum_{i=1}^{p} \phi_{0i} y_{t-i} + \sum_{i=0}^{q} \boldsymbol{\theta}'_{0i} \boldsymbol{z}_{t-i} + \varepsilon_t = \boldsymbol{\beta}'_0 \boldsymbol{x}_t + \varepsilon_t, \tag{1}$$

where only a fraction of the elements of $\boldsymbol{\beta}$ is non-zero, i.e., $\boldsymbol{\beta}$ is sparse. Consider the following assumption about the data generating process (DGP).

Assumption (Data Generating Process)**.** *The DGP is such that:*

(A1) The roots of the polynomial $1 - \sum_{i=1}^{p} \phi_{0i} L^i$ *lie outside the unity circle.*

(A2) $\boldsymbol{z}_t$ *is an infinite-order vector moving average process, VMA(∞),* $\boldsymbol{z}_t = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\eta}_{t-j}$*, where*

- *a.* $\{\boldsymbol{\eta}_t\}_{t=-\infty}^{\infty}$ *is a zero mean, strictly stationary, uncorrelated, strong mixing process taking values on* $\mathbb{R}^{d_z}$*,* $d_z \in \mathbb{N}$*. The strong mixing coefficients* $\{\alpha_m\}_{m=-\infty}^{\infty}$ *decrease geometrically with* m*, i.e.* $\alpha_m = O(\alpha^m)$ *for some* $|\alpha| < 1$*.*
- *b. For some* $d \geq 1$*,* $\mathbb{E}(\boldsymbol{\eta}_{0,i}^{2d}) < \infty$*,* $i = 1, \ldots, d_z$*. There exists a positive constant* $\sigma_{\eta,\max}$*, such that* $\max_{\boldsymbol{b}'\boldsymbol{b}=1} \|\boldsymbol{b}'\boldsymbol{\eta}_0\|_{2d} \leq \sigma_{\eta,\max}$*.*[1]
- *c.* $\sum_{j=k}^{\infty} \|\boldsymbol{\Psi}_j\| \leq \kappa_k$*, where* κ_k *is some non-negative, decreasing sequence satisfying* $\kappa_k = O(\kappa^k)$ *for some* $0 \leq \kappa < 1$*, and* $\|\boldsymbol{\Psi}_j\|$ *denotes the operator norm of* $\boldsymbol{\Psi}_j$*.*

(A3) The strictly stationary, strong mixing innovation process $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ *is a difference martingale sequence with respect to the* σ*-algebra generated by* $\{\boldsymbol{\eta}_{t-i}, \varepsilon_{t-1-i}\}_{i=0}^{\infty}$*. The strong mixing coefficients* $\{\alpha_m\}_{m=-\infty}^{\infty}$ *decrease geometrically with* m*, i.e.,* $\alpha_m = O(\alpha^m)$ *for some* $|\alpha| < 1$*. For some* $d \in [1, \infty)$*,* $\mathbb{E}(\varepsilon_0^{2d}) < \infty$*, and* $\mathbb{E}(\varepsilon_0^2) = \sigma_\varepsilon^2$*.*

(A4) $\{\boldsymbol{\eta}_t\}_{t=-\infty}^{\infty}$ *and* $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ *are such that* $\mathbb{E}(\boldsymbol{\eta}_t \varepsilon_{t-i})$ *depends only on* i*,* $i = 1, 2, \ldots$*, for any* t*.*

(A5) $\boldsymbol{\beta}_0 = [\boldsymbol{\beta}_0(1)', \boldsymbol{\beta}_0(2)']' \in \mathbb{R}^{p+qd_z}$ *is sparse, in a sense that it* $\boldsymbol{\beta}_0(1) \neq \boldsymbol{0}$*,* $\boldsymbol{\beta}_0(2) = \boldsymbol{0}$ *and* $\boldsymbol{\beta}_0(1) \in \mathbb{R}^s$*,* $s \in \mathbb{N}$*. Moreover,* $\min_{1 \leq i \leq s} |\beta_{0i}| \geq \beta_{\min} > 0$ *and* $\sum_{i=0}^{\infty} \|\boldsymbol{\theta}_{0i}\| \leq \tilde{\theta} < \infty$*.*

(A6) The smallest eigenvalue of $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \mathbb{E}\left(\boldsymbol{x}_0 \boldsymbol{x}'_0\right)$ *is bounded away from zero.*

[1] We write $\|\boldsymbol{b}'\boldsymbol{v}\|_{2d} = (\mathbb{E}|\boldsymbol{b}'\boldsymbol{v}|^{2d})^{1/2d} = (\mathbb{E}|\boldsymbol{b}'\boldsymbol{v}\boldsymbol{v}'\boldsymbol{b}|^d)^{1/2d}$, and $\|\cdot\| = \|\cdot\|_2$.

Under Assumptions (A1)–(A4), y_t admits an MA(∞) expansion such that $y_t = \sum_{j=0}^{\infty} \gamma_j s_{t-j}$, where $\sum_{j=k+1}^{\infty} |\gamma_j| \leq \zeta_k$, ζ_k is a geometrically decreasing sequence, and $s_t = \sum_{i=0}^{q} \boldsymbol{\theta}_{0i}' \boldsymbol{z}_{t-i} + \varepsilon_t$ is weakly stationary. The conditions also imply that y_t and z_t have $2d$ moments. Assumption (A5) requires that the smallest coefficient that enters the model is lower bounded by $\beta_{\min} > 0$ that may decrease as the number of variables, s, increases. It also requires that the parameters $\boldsymbol{\theta}_{0j}$, $j = 0, 1, \ldots$, are not too large, which is required to show that $\mathbb{E}|y_0|^{2d}$ does not increase as q and the dimension of z_t increase. Finally, Assumption (A6) restricts our result to processes where $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ is positive definite.

Assumption (A2) covers a large range of cases. For instance, if we assume that $\boldsymbol{A}(L)\boldsymbol{z}_t = \boldsymbol{B}(L)\boldsymbol{\eta}_t$ where $\boldsymbol{A}(L)$ and $\boldsymbol{B}(L)$ are polynomials with finite degree, L is the lag operator, and the innovations $\boldsymbol{\eta}_i \overset{\text{iid}}{\sim} \mathsf{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta)$, Assumption (A2) part b is satisfied. If the largest eigenvalue of the companion matrix of the system is less than one in absolute value, Assumption (A2) parts a and c are also satisfied. Assumption (A4) is valid if $\boldsymbol{\eta}_t$ is a martingale difference process with respect to the σ-algebra generated by $\{\boldsymbol{\eta}_{t-i}, \varepsilon_{t-i} : i = 1, 2, ...\}$.

Assumption (A3) is satisfied by a number of conditionally heteroskedastic error specifications. In the following examples we consider two cases of interest.

2.1. **ARDL(p,q)-GARCH(l,m) model.** Consider model (1) with the error defined as

$$\varepsilon_t = \sqrt{h_t} v_t, \quad v_t \overset{\text{iid}}{\sim} (0,1), \quad h_t = \alpha_0 + \sum_{i=1}^{l} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{m} \pi_i h_{t-i}. \tag{2}$$

Under this GARCH specification we may replace Assumption (A3) by:

Assumption ((A3) for GARCH(l,m))**.** *The coefficients of the GARCH(l,m) innovations satisfy, for some* $d \in [1, \infty)$,

a. $\sum_{i=1}^{l} \alpha_i + \sum_{i=1}^{m} \pi_i < 1$
b. $\mathbb{E}|v_0|^{2d} < \infty$ *and* $\left(\frac{\sum_{i=1}^{l} \alpha_i}{1 - \sum_{i=1}^{m} \pi_i} \right)^d \mathbb{E}|v_0|^{2d} < 1$
c. The distribution of v_0 *is absolutely continuous with respect to the Lebesgue measure, being strictly positive in a neighbourhood of zero.*

These conditions ensure that the GARCH innovations are strictly stationary, strong mixing with geometrically decreasing rate, and have $2d$ moments (Lindner, 2009).

2.2. **ARDL(p,q) with Meitz and Saikkonen's (2008) GARCH family.** Consider model (1) with the error defined as

$$\varepsilon_t = f(h_t) v_t, \quad v_t \overset{\text{iid}}{\sim} (0,1), \quad h_t = g_1(h_{t-1}) + g_2(\varepsilon_{t-1}, h_{t-1}). \tag{3}$$

This specification is discussed in Meitz and Saikkonen (2008). It nests a number of first order GARCH specifications such as the Hentschel's family of GARCH models (Hentschel, 1995). The authors derived general conditions for stationarity, strong-mixingness and the existence of moments for this family of models. We replace (A3) by the following set of assumptions.

Assumption ((A3) for Meitz and Saikkonen GARCH family)**.** *a. v_0 has a probability density function $\phi_v(\cdot)$ supported on $\mathbb{R}$ and bounded away from zero on compact subsets of $\mathbb{R}$.*

b. The functions $g_1 : \mathbb{R}_+ \to \mathbb{R}_+$ and $g_2 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are bounded on bounded subsets of their domains and, for some $\underline{g} > 0$, $\inf_{h \in \mathbb{R}_+, u \in \mathbb{R}} [g_1(h) + g_2(u, h)] = \underline{g}$.

c. There exists a real number $a \in [0, \infty)$ such that $g_1(h) \leq ah + o(h)$ as $h \to \infty$.

d. The function g_2 satisfies the following three conditions.

d_1. There exists an unbounded interval of $\mathbb{R}_+$ that is, for all $h > 0$, contained in the image set $g_2[(-\infty, \infty), h]$.

d_2. For all $h > 0$, the function $g_2(\cdot, h)$ is continuous from the right (or alternatively, continuous from the left).

d_3. There exists a real number $R > 0$ such that, for $u > R$ and all $h > 0$, $g_2(u, h)$ is continuous and monotonically increasing and the inverse function $g_2^{-1}(v, h)$ is such that $\partial g_2^{-1}(v, h)/\partial v$ is bounded away from zero on compact subsets of its domain.

e. The exist a Borel measurable function $b : \mathbb{R} \to \mathbb{R}_+$, nonconstant and continuous on some open set, and a real number $c \in [0, \infty)$ such that $g_2(f(h)v, h) \leq hb(v) + c$ for all $h \in \mathbb{R}_+$. Furthermore, $\mathbb{E}[b(v_0)^d] < \infty$ for some $d \in \mathbb{R}_+$.

The reader is referred to the original paper for a detailed discussion of the assumptions along with examples of conditional variance specifications.

3. Results

The adaLASSO estimator of $\boldsymbol{\beta}$ used in this paper is given by

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in B} \sum_{t=\max(p,q)+1}^{T} \left(y_t - \boldsymbol{\beta}' \boldsymbol{x}_t\right)^2 + \lambda \sum_{i=j}^{p+qd_z} w_j |\beta_j|, \tag{4}$$

where $w_j = 1/|\widehat{\beta}_{I,j}|^{-1}$ are the initial weights constructed from the LASSO initial estimates, $\widehat{\boldsymbol{\beta}}_I$ of $\boldsymbol{\beta}_0$, and $B \subset \mathbb{R}^{p+qd_z}$ is an open ball around $\{\mathbf{0}\}$ that satisfy the constraints in Assumptions (A1) and (A5). When $\hat{\beta}_{I,j} = 0$, we remove the predictor from the equation.

Recently, Fan et al. (2014) showed that the adaLASSO is just an one-step implementation of the family of folded concave penalized least-squares of Fan and Li (2001). Furthermore, using the LASSO as initial estimator can be regarded as the two-step implementation of the local linear approximation in Fan et al. (2014) with a zero initial estimate.

We show that, under Assumptions (A1) – (A6), and further conditions on λ and $\beta_{\min}$, the LASSO can indeed be used as an initial estimator, the adaLASSO consistently selects the correct set of relevant variables, and that the asymptotic distribution of the estimator is the same as the oracle assisted least squares. The number of candidate variables can be larger than the number of observations, but it depends on the moment Assumptions (A2) and (A3).

The next theorem summarises the main results of this paper. We use the concept of sign consistency that implies model selection consistency. We define $\operatorname{sign}(x) = I(x > 0) - I(x < 0)$.

Theorem 1. *Assume (A1)–(A6) hold with $\beta_{\min} > cs^{-\frac{1}{2}}$, for all T sufficiently large and some constant $c > 0$. For some $0 < \xi < 1$, $p = q = o\left(T^{\frac{\xi}{4}}\right)$, $s = o\left(T^{\frac{\xi}{4}}\right)$, $s^{1/2}pd_z = o\left[T^{\frac{d-1}{4}(1-\xi)}\right]$, and the regularization parameter λ_I of the initial LASSO estimator satisfies $\frac{(pd_z)^{\frac{1}{d}}T^{\frac{1}{2}}}{\lambda_I} \to 0$ and $\frac{\lambda_I s^{\frac{3}{2}}}{T^{1-\xi/2}} \to 0$. Furthermore, if the adaLASSO regularization parameter λ satisfies $\frac{(pd_z)^{\frac{1}{d}}T^{\frac{1-\xi}{2}}}{\lambda} \to 0$ and $\frac{\lambda s}{T^{1/2}} \to 0$, then $\Pr\left[\operatorname{sign}(\widehat{\boldsymbol{\beta}}) = \operatorname{sign}(\boldsymbol{\beta}_0)\right] \to 1$, as $T \to \infty$, where the equality should be taken elementwise. Denote $\widehat{\boldsymbol{\beta}}_{ols}(1)$ the OLS estimator of $\boldsymbol{\beta}_0(1)$. For any s-dimensional vector $\boldsymbol{\delta}$ with Euclidean norm 1, $\sqrt{T}\boldsymbol{\delta}'\left[\widehat{\boldsymbol{\beta}}(1) - \boldsymbol{\beta}_0(1)\right] = \sqrt{T}\boldsymbol{\delta}'\left[\widehat{\boldsymbol{\beta}}_{ols}(1) - \boldsymbol{\beta}_0(1)\right] + o_p(1)$.*

The rate in which n and s increase is controlled by ξ and d, in Assumptions (A2) and (A3). The constant ξ is related to the number of candidate variables and variables in the model. Larger values of ξ means that s can be larger at a cost of less candidate variables in the pool, on the other hand, a small ξ allows for a larger number of candidate variables at the cost of stronger conditions on s. We derive the rates of increase in $(p+1)d_z$ and s under two scenarios: $s = O(1)$ and $s = O\left[(pd_z)^{2/(d-1)}\right]$. In the first scenario, since s is constant, the number of candidate variables is $(p+1)d_z = o\left[T^{\frac{d-1}{4}(1-\xi)}\right]$. In the second scenario $pd_z = o\left[T^{\frac{(d-1)^2}{4d}(1-\xi)}\right]$, and $s = o\left\{T^{\min\left[\frac{d-1}{2d}(1-\xi), \frac{\xi}{4}\right]}\right\}$. In all cases, increasing the number of moments is essential to increase the number of candidate and relevant variables. Different specifications of s may also be considered and the respective rates derived. Both regularization parameters λ and λ_I are constrained by the number of candidate variables, the number of variables in the model, and number of moments d.

The selection of the regularization parameter λ is critical. Traditionally, one employs cross-validation and selects λ within a grid that maximizes some predictive measure. In a time-dependent framework cross-validation is more complicated. An alternative approach that has received more attention in recent years is to choose λ using information criteria, such as the BIC; see Zou et al. (2007). We adopt the BIC to select both λ_I and λ.

4. Simulation

Consider the following DGP:

$$
\begin{aligned}
y_t &= \phi y_{t-1} + \boldsymbol{\theta}'\boldsymbol{z}_{t-1}(1) + u_t, \quad u_t = h_t^{1/2}\varepsilon_t, \quad \varepsilon_t \overset{\text{iid}}{\sim} \mathsf{t}^*(5), \quad h_t = 5\times 10^{-4} + \pi_1 h_{t-1} + \alpha_1 u_{t-1}^2 \\
\boldsymbol{z}_t &= \begin{bmatrix} \boldsymbol{z}_t(1) \\ \boldsymbol{z}_t(2) \end{bmatrix} = \boldsymbol{A}_1 \begin{bmatrix} \boldsymbol{z}_{t-1}(1) \\ \boldsymbol{z}_{t-1}(2) \end{bmatrix} + \boldsymbol{A}_4 \begin{bmatrix} \boldsymbol{z}_{t-4}(1) \\ \boldsymbol{z}_{t-4}(2) \end{bmatrix} + v_t, \; v_t \overset{\text{iid}}{\sim} \mathsf{N}(0,1),
\end{aligned}
\tag{5}
$$

where the typical element of $\boldsymbol{\theta}$ is given by $\theta_i = \frac{1}{\sqrt{s}}(-1)^i$. $\boldsymbol{z}_t(1)$ is a $(s-1)\times 1$ vector of included (relevant) variables. The vector $\boldsymbol{z}_t = [\boldsymbol{z}_t(1)', \boldsymbol{z}_t(2)']' \in \mathbb{R}^{(n-1)}$, has $n-s$ irrelevant variables and follows a fourth-order VAR model with Gaussian errors as in Kock and Callot (2015). The matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_4$ are block diagonals with each block of dimension 5×5 and typical element 0.15 and -0.1, respectively. The errors of the VAR have a covariance matrix equal to the identity. The error term of the ARDL model is t-distributed with 5 degrees of freedom. $\mathsf{t}^*(5)$ denotes an standardized t-distribution with 5 degrees of freedom, such that all the errors have zero mean and unit variance. The vector of candidate variables is $\boldsymbol{x}_t = (y_{t-1}, \boldsymbol{z}_{t-1}')'$. Finally, $\phi = 0.5$, 0.8, or 0.9 and the GARCH parameters (α_1, π_1) can be either $(0.1, 0.8)$ or $(0.5, 0.9)$. The two GARCH processes induce different moment structure for the error term in the model.

We simulate $T = 100, 500, 1000$ observations of DGP (5) for different combinations of n, s, and values for ϕ, π_1, and α_1 . We consider $n = 101, 301, 501$ and $s = 5, 10, 15$.

We start by analyzing the properties of the estimators for ϕ. Figures 1–3 illustrate the distribution of oracle and the adaLASSO estimators. For $T = 100$ the adaLASSO estimator is biased downwards as expected but the bias reduces as T increases. The bias is also more severe for $\phi = 0.8$ or $\phi = 0.9$ as compared to case where $\phi = 0.5$. The differences in performance among the two GARCH specifications are minor. It is also clear that when $\phi = 0.5$ or $\phi = 0.8$ the adaLASSO distribution gets closer to the oracle as the sample size increases. For $\phi = 0.9$ there is still a small bias even when $T = 1000$.

Table 1 shows the average bias and the average mean squared error (MSE) for the adaLASSO estimator over the Monte Carlo simulations and the candidate variables, i.e.,

$$\mathsf{Bias} = \frac{1}{1000n}\sum_{j=1}^{1000}\left[\widehat{\phi} - \phi + \sum_{i=1}^{n-1}\left(\widehat{\beta}_i - \beta_i\right)\right] \text{ and } \mathsf{MSE} = \frac{1}{1000n}\sum_{j=1}^{1000}\left[\left(\widehat{\phi} - \phi\right)^2 + \sum_{i=1}^{n-1}\left(\widehat{\beta}_i - \beta_i\right)^2\right].$$

It is clear that both variance and bias are very low. This is explained, as expected, by the large number of zero estimates. There are not much difference between the two GARCH specifications. On the other hand, the higher is the persistence of the model, the higher is the bias.

Table 2 presents model selection results. Panel (a) shows the fraction of replications where the correct model has been selected, i.e., all the relevant variables included and all the irrelevant regressors excluded from the final model (correct sparsity pattern). It is clear the performance of the adaLASSO improves as $T \to \infty$ and gets worse as q increases. Furthermore, there is a slightly deterioration as n increases. Finally, the higher the persistence the worse the results, specially in small samples. Panel (b) shows the fraction of replications where the relevant variables are all included. For $T = 300$ and $T = 1000$, the true model is included almost every time. For smaller sample sizes the performance decreases dramatically as s and ϕ increase. Panel (c) presents the fraction of relevant variables included and Panel (d) shows the fraction of irrelevant variables excluded. It is clear that the fraction of included relevant variables is extremely high, as well as the fraction of excluded irrelevant regressors.

Table 3 shows the MSE ratio (adaLASSO over the oracle) for one-step-ahead out-of-sample forecasts. We consider a total of 100 out-of-sample observations. As expected, for low values of q, the adaLASSO has a similar performance than the oracle. For $q = 10$ or $q = 15$, the results are reasonable only for $T = 500$ or $T = 1000$. The performance of the adaLASSO also improves as T increases.

5. Conclusions

We consider the estimation of ARDL model with non-Gaussian and GARCH errors in high-dimensions. We advocate the use of the adaLASSO procedure to estimate the parameters and we show that the adaLASSO is model selection consistent and possesses the oracle property under some set of conditions.

Appendix A. Proof of Theorem 1

We show that the conditions in Theorem 1 satisfies conditions DGP, DESIGN, WEIGHTS, and REG in Medeiros and Mendes (2016).

Assumption DGP(1) follows because $\mathbb{E}(\boldsymbol{z}_t) = 0$, $\mathbb{E}(\varepsilon_t) = 0$ and the results in Lemmas 1 and 2. DGP(2) follows from Assumption (A3). DGP(3) is satisfied by Lemma 8 along with the constraints on pd_z and s. It follows from the Cauchy-Schwarz inequality that $\|\varepsilon_t x_{it}\|_d \leq \|\varepsilon_t\|_{2d}\|x_{it}\|_{2d}$. Lemma 1 ensures that $\|y_{t-j}\|_{2d}$ and $\|z_{i,t-j}\|_{2d}$ are bounded independently of T. Assumptions DESIGN(1) and DESIGN(2) are satisfied by Assumption (A5).

In Section 5 of Medeiros and Mendes (2016), the authors show that, under regularity conditions, the LASSO initial estimator satisfies conditions WEIGHTS(1) and WEIGHTS(2) in their paper. In summary, we require that, with probability converging to one: (1) $\max_{1\leq i,j\leq p+qd_z} |[\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} - \boldsymbol{\Sigma}_{\boldsymbol{x}}]_{ij}| \leq \rho_0/16s$, where $\rho_0 > 0$ is the smallest eigenvalue of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$; (2) The LASSO parameter estimates satisfy $\|\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0\|_1 \leq 4\frac{\lambda_I s}{T\rho_0}$, where λ_I is the regularization parameter of the initial estimator; (3) $\frac{(p+qd_z)^{1/d}T^{1/2}}{\lambda_I} \to 0$ and $\frac{\lambda_I s^{3/2}}{T^{1-\xi/2}} \to 0$; and (4) $\beta_{\min} > 2w_{\max}^{-1}\max\left(1, 4\frac{\lambda_I s\, w_{\max}}{T\rho_0}, 8\frac{\lambda\sqrt{s}w_{\max}}{T^{1-\xi/2}\rho_0}\right)$, where $\sum_{i=1}^{s}|\theta_{I,i}|^{-1} \leq sw_{\max}^2$.

We assume that $p = q$ and $s^{1/2}pd_z = o\left[T^{\frac{d-1}{4}(1-\xi)}\right]$, making $\nu = \xi/4$. It then follows from Lemma 8 that

$$\Pr\left(\max_{1\leq i,j\leq p+qd_z} |[\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} - \boldsymbol{\Sigma}_{\boldsymbol{x}}]_{ij}| \leq \frac{\rho_0}{16s}\right) \geq 1 - c_1\frac{sp^2d_z^2}{T^{(d-1)(1-\xi)/2}} \to 1.$$

Hence, it follows from Medeiros and Mendes (2016, Lemma 1) that the *restricted eigenvalue condition* holds with bound ρ_0, and DGP(3) is satisfied with bound $\rho_0/16$.

Suppose that $\{u_t x_{it}\}$ is a martingale difference sequence with $\max_t \mathbb{E}|u_t x_{it}|^d < \infty$. Then, applying the Burkhölder-Davis-Gundy and C_r inequalities, $\mathbb{E}|T^{-1/2}\sum_{t=1}^T u_t x_{it}|^d \leq c \max_t \mathbb{E}|u_t x_{it}|^d < \infty$ and

$$
\begin{aligned}
\Pr\left(2 \max_{i=1,\ldots,p+qd_z}\left|\sum_{t=1}^T u_t x_{it}\right| > \sqrt{T}\lambda_0\right) &\geq 1 - \frac{2}{\lambda_0}\mathbb{E}\left(\max_{i=1,\ldots,p+qd_z}\left|T^{-1/2}\sum_{t=1}^T u_t x_{it}\right|\right) \\
&\geq 1 - \frac{2(p+qd_z)^{1/d}}{\lambda_0}\max_{i=1,\ldots,p+qd_z}\left\|\frac{1}{T}\sum_{t=1}^T u_t x_{it}\right\|_d \\
&\geq 1 - 4c_2\frac{(p+qd_z)^{1/d}T^{1/2}}{\lambda_I} \to 1,
\end{aligned}
$$

where we take $2\lambda_0 = T^{-1/2}\lambda_I$ and use DGP(4) to show that $\|u_t x_{i,t}\|_d < \infty$. It follows from Bülhmann and van der Geer (2011, Theorem 6.1) that $\|\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0\|_1 \leq 4\frac{\lambda_I s}{T\rho_0}$ with probability converging to one. The conditions on the regularization parameter for the LASSO initial estimator, λ_I, are assumed.

Finally, we take $w_{\max} = 2s^{\frac{1}{2}}/c_3$, for some constant $c_3 > 0$. It means that $\frac{\lambda_I s\, w_{\max}}{T\rho_0} \propto \frac{\lambda_I s^{3/2}}{T} \to 0$ and $\frac{\lambda\sqrt{s}w_{\max}}{T^{1-\xi/2}\rho_0} \propto \frac{\lambda s}{T^{1-\xi/2}} \to 0$, from assumptions on λ_I and λ. Therefore, for all T sufficiently large it is enough to assume $\beta_{\min} > c_3 s^{-\frac{1}{2}}$.

It follows from Medeiros and Mendes (2016, Proposition 1) that there exists $0 \leq \xi < 1$ satisfying condition WEIGHTS(1) and (2), as far as $2s^{\frac{1}{2}}/c_3 < T^{\xi/2}$, for sufficiently large T. Hence, the LASSO estimates can be used to construct the initial weights. Therefore, the result follows from Theorem 1 and 2 in Medeiros and Mendes (2016).

Appendix B. Properties of y_t and s_t

Lemma 1. *Under Assumptions (A1)–(A5), (1) $\|\boldsymbol{\delta}'\boldsymbol{z}_t\|_{2d} \leq \sigma_{\eta,\max}\kappa_0\|\boldsymbol{\delta}\|$, for any $\delta \in \mathbb{R}^{d_z}\setminus\{\mathbf{0}\}$; (2) $\|\boldsymbol{s}_t\|_{2d} \leq c_d$, where $c_d = 2(\sigma_{\eta,\max}\kappa_0\tilde{\theta} \vee 1)(\|\varepsilon_0\|_{2d} \vee 1)$; and (3) $\|y_t\|_{2d} = \zeta_0 c_d$.*

Proof. For any $\boldsymbol{\delta} \in \mathbb{R}^{d_z}$, the triangle inequality and sub-multiplicative property of the operator norm yields

$$
\begin{aligned}
\|\boldsymbol{\delta}'\boldsymbol{z}_t\|_{2d} &\leq \sum_{j=0}^{\infty}\|\boldsymbol{\delta}'\boldsymbol{\Psi}_j\boldsymbol{\eta}_{t-j}\|_{2d} \leq \sum_{j=0}^{\infty}\|\boldsymbol{\delta}'\boldsymbol{\Psi}_j\|\max_{\boldsymbol{b}'\boldsymbol{b}=1}\|\boldsymbol{b}'\boldsymbol{\eta}_{t-j}\|_{2d} \\
&\leq \left[\max_{\boldsymbol{b}'\boldsymbol{b}=1}\mathbb{E}(\boldsymbol{b}'\boldsymbol{\eta}_0\boldsymbol{\eta}_0'\boldsymbol{b})^d\right]^{1/2d}\|\boldsymbol{\delta}\|\sum_{j=0}^{\infty}\|\boldsymbol{\Psi}_j\| \leq \sigma_{\eta,\max}\|\boldsymbol{\delta}\|\kappa_0.
\end{aligned}
$$

(2) is proved in two steps. Let $A_t = \sum_{i=0}^q \boldsymbol{\theta}_{0i}'\boldsymbol{z}_{t-i}$ and $B_t = \epsilon_t$. Hence,

$$
\|A_t\|_{2d} \leq \sum_{i=0}^q \|\boldsymbol{\theta}_{0i}'\boldsymbol{z}_{t-i}\|_{2d} \leq \kappa_0\sigma_{\eta,\max}\,\tilde{\theta}, \quad \|B_t\|_{2d} = \|\varepsilon_0\|_{2d}, \quad \text{and}
$$

$$
\begin{aligned}
\mathbb{E}|s_t|^{2d} &= \sum_{j=0}^{2d}\binom{2d}{j}\mathbb{E}|A_t^j B_t^{2d-j}| \leq \sum_{j=0}^{2d}\binom{2d}{j}\|A_t\|_{2d}^j\|B_t\|_{2d}^{2d-j} \\
&\leq 2^{2d}\max_{1\leq j\leq 2d}\|A_t\|_{2d}^j\|B_t\|_{2d}^{2d-j} \leq \left[2(\kappa_0\sigma_{\eta,\max}\tilde{\theta} \vee 1)(\|\varepsilon_0\|_{2d} \vee 1)\right]^{2d}.
\end{aligned}
$$

(3) is proved by induction by applying Hölder's inequality d times:

$$\mathbb{E}|y_t|^{2d} = \mathbb{E}\left(\sum_{i=0}^{\infty}\gamma_i s_{t-i}\right)^{2d} \leq \sum_{i_1=0}^{\infty}\cdots\sum_{i_{2d}=0}^{\infty}\prod_{j=1}^{2d}|\gamma_{i_j}|\,\mathbb{E}\left|\prod_{j=1}^{2d}s_{t-i_j}\right| \leq \max_t \|s_t\|_{2d}^{2d}\left(\sum_{i=0}^{\infty}|\gamma_j|\right)^{2d} \leq c_d^{2d}\zeta_0^{2d},$$

where we use the fact that $\mathbb{E}|\prod_{j=1}^{d}x_{i_j}| \leq \prod_{j=1}^{d}\|x_{i_j}\|_d$. □

Lemma 2. *Under Assumptions (A1)–(A5), (1) $\{\boldsymbol{z}_t\}$ is weakly stationary; (2) $\{s_t\}$ is weakly stationary; and (3) the vector process $\{y_t, \boldsymbol{z}_t, \varepsilon_t\}$ is weakly stationary.*

Proof. The existence of second moments is established on Lemma 1. Hence, we have to show that the covariances and autocorrelations of y_t, $\boldsymbol{z}_t$ and ε_t do not depend on t.

Under Assumption (A2), $\boldsymbol{z}_t = \sum_{i=0}^{\infty}\boldsymbol{\Psi}_i\boldsymbol{\eta}_{t-i}$, where $\mathbb{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_j') = \boldsymbol{0}$. Therefore,

$$\mathbb{E}\left(\boldsymbol{z}_t\boldsymbol{z}_{t-i}'\right) = \mathbb{E}\left[\left(\sum_{j=0}^{\infty}\boldsymbol{\Psi}_j\boldsymbol{\eta}_{t-j}\right)\left(\sum_{j=0}^{\infty}\boldsymbol{\Psi}_j\boldsymbol{\eta}_{t-i-j}\right)'\right] = \sum_{j=0}^{\infty}\boldsymbol{\Psi}_{i+j}\boldsymbol{\Sigma}_\eta\boldsymbol{\Psi}_j' = \boldsymbol{F}_z(i), \tag{6}$$

where $\boldsymbol{F}_z(i)$ is a matrix that depends on i. Hence, since $\mathbb{E}(\boldsymbol{z}_t) = \boldsymbol{0}$, $\{\boldsymbol{z}_t\}$ is weakly stationary.

Assumption (A3) requires that ε_t is strictly stationary and is a martingale difference process with respect to the information $\mathcal{F}_t = \sigma\{\boldsymbol{\eta}_{t-i}, \varepsilon_{t-i-1} : i = 0, 1, ...\}$. Hence,

$$\mathbb{E}(\varepsilon_t s_{t-i}) = \mathbb{E}\left[\left(\sum_{j=0}^{q}\boldsymbol{\theta}_{0j}'\boldsymbol{z}_{t-i-j} + \varepsilon_{t-i}\right)\mathbb{E}(\varepsilon_t|\mathcal{F}_t)\right] = 0. \tag{7}$$

Define $\mathbb{E}(\varepsilon_{t-i}\boldsymbol{z}_{t-j}) = \boldsymbol{F}_{z\varepsilon}(i-j)$. Under Assumption (A3) $\mathbb{E}(\varepsilon_t\boldsymbol{z}_{t-i}) = \boldsymbol{F}_{z\varepsilon}(-i) = \boldsymbol{0}$, for $i = 0, 1, ...$ Assumption (A4) together with $\mathbb{E}(\varepsilon_t\boldsymbol{\eta}_{t-i}) = \boldsymbol{0}$, $i = 0, 1, \ldots$, implies that

$$\mathbb{E}(\boldsymbol{z}_t\varepsilon_{t-i}) = \sum_{l=0}^{\infty}\boldsymbol{\Psi}_l\mathbb{E}(\boldsymbol{\eta}_{t-l}\varepsilon_{t-i}) = \sum_{l=0}^{i-1}\boldsymbol{\Psi}_l\mathbb{E}(\boldsymbol{\eta}_{t-l}\varepsilon_{t-i}) = \sum_{l=0}^{i-1}\boldsymbol{\Psi}_l\boldsymbol{F}_{\eta\varepsilon}(i-l) = \boldsymbol{F}_{z\varepsilon}(i). \tag{8}$$

Since ε_ts are uncorrelated, it follows that

$$\mathbb{E}(s_t\varepsilon_{t-i}) = \mathbb{E}\left[\left(\sum_{j=0}^{q}\boldsymbol{\theta}_{0j}'\boldsymbol{z}_{t-j} + \varepsilon_t\right)\varepsilon_{t-i}\right] = \sum_{j=0}^{q}\boldsymbol{\theta}_{0j}'\mathbb{E}\left(\boldsymbol{z}_{t-j}\varepsilon_{t-i}\right) = \sum_{j=0}^{q}\boldsymbol{\theta}_{0j}'\boldsymbol{F}_{z\varepsilon}(i-j) = F_{s\varepsilon}(i,q), \tag{9}$$

that is a function of $\boldsymbol{F}_{z\varepsilon}(i)$ and q.

Finally, since $\boldsymbol{z}_t$ is weakly stationary, $r_t(q) = \sum_{i=0}^{q}\boldsymbol{\theta}_{0i}'\boldsymbol{z}_{t-i}$ is also weakly stationary, with autocorrelation $\mathbb{E}[r_t(q)r_{t-i}(q)] = F_r(i,q)$. Combining the previous results we show that s_t is weakly stationary, i.e., $\mathbb{E}(s_t s_{t-i}) = F_{s\varepsilon}(i,q) + F_r(i,q) = F_s(i,q)$. Since $\mathbb{E}(s_t) = 0$ for all t, we conclude that s_t is weakly stationary. Furthermore y_t is also weakly stationary.

To prove the last claim we have to show that the individual processes y_t, $\boldsymbol{z}_t$ and ε_t are weakly stationary, and show that expected values of their lagged cross-products do not depend on t. We have already shown that the individual processes are weakly stationary. The expected value $\mathbb{E}(\boldsymbol{z}_{t-j}\varepsilon_{t-i}) = \boldsymbol{F}_{z\varepsilon}(i-j)$ does not depend on t for any i, j. Since $y_{t-i} = \sum_{j=0}^{\infty} \gamma_j s_{t-i-j}$, it follows from (7) and (9) that $\mathbb{E}(\varepsilon_{t-j} y_{t-i})$ is a function of i and j, but not t. Finally, it follows from (6), (8), and the fact that ε_t is a martingale difference process, that $\mathbb{E}(\boldsymbol{z}_{t-j} y_{t-i})$ is a function of only i and j. $\square$

Lemma 3. *For some $k \in \mathbb{N}$, let $y_{t,k} = \sum_{j=0}^{k} \gamma_j s_{t,k}$, where $s_{t,k} = \sum_{i=0}^{q} \boldsymbol{\theta}_{0i}' \boldsymbol{z}_{t,k} + \varepsilon_t$ and $\boldsymbol{z}_{t,k} = \sum_{j=0}^{k} \boldsymbol{\Psi}_j \boldsymbol{\eta}_{t-j}$. Under Assumptions (A1)-(A5), for $i \in \mathbb{N}$, (1) $y_{t-i,k}$ is a measurable function of $(\boldsymbol{\eta}_{t-i}', \ldots, \boldsymbol{\eta}_{t-i-q-k}, \varepsilon_{t-i}, \ldots, \varepsilon_{t-i-k})$; (2) $\boldsymbol{z}_{t-j,k}$ is a measurable function of $(\boldsymbol{\eta}_{t-j}', \ldots, \boldsymbol{\eta}_{t-j-k})$; (3) $\|y_{t-i,k}\|_{2d} \leq \zeta_0 c_d$; (4) $\|\boldsymbol{\delta}' \boldsymbol{z}_{t-j,k}\|_{2d} \leq \sigma_{\eta,\max} \kappa_0 \|\boldsymbol{\delta}\|$, for $\delta \in \mathbb{R}^{dz} \backslash \{0\}$; (5) $\|y_{t-i} - y_{t-i,k}\|_{2d} \leq c_d \, \zeta_{k+1}$; and (6) $\|\boldsymbol{\delta}'(\boldsymbol{z}_{t-i} - \boldsymbol{z}_{t-i,k})\|_{2d} \leq \sigma_{\eta,\max} \|\boldsymbol{\delta}\| \, \kappa_{k+1}$, for $\delta \in \mathbb{R}^{dz} \setminus \{0\}$.*

Proof. Parts (1) and (2) follow from the definition of $y_{t,k}$ and $\boldsymbol{z}_{t,k}$; (3) and (4) follow as in Lemma 1 and the derivation will be omitted. It suffices to prove (5) for $i = 0$. Write $\|y_t - y_{t,k}\| = \left\| \sum_{j=k+1}^{\infty} \gamma_j s_{t-j} \right\|_{2d} \leq \max_t \|s_{t,k}\|_{2d} \sum_{j=k+1}^{\infty} |\gamma_j| \leq c_d \zeta_{k+1}$, where $\|s_{t,k}\|_{2d} < c_d \tilde{r}$ follows after the same arguments used in Lemma 1. Similarly, for (6), we have $\|\boldsymbol{\delta}'(\boldsymbol{z}_{t-i} - \boldsymbol{z}_{t-i,k})\| \leq \sigma_{\eta,\max} \|\boldsymbol{\delta}\| \kappa_{k+1}$. $\square$

Appendix C. A probability Bound

Proposition 1 (Triplex Inequality, Jiang (2009, Theorem 1))**.** *Let $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ be an increasing sequence of σ-fields, and ρ_t be a random variable that is $\mathcal{F}_t$-measurable for each t. Then, for each ε_T, $C_T > 0$ and positive integers m and T, we have*

$$
\begin{aligned}
\Pr\left\{ \left| \sum_{t=1}^{T} [\rho_t - \mathbb{E}(\rho_t)] \right| > T\varepsilon_T \right\} \leq\; & 2m \exp\left[-T\varepsilon_T^2 / (288 m^2 C_T^2) \right] \\
& + (6/\varepsilon_T) T^{-1} \sum_{t=1}^{T} \mathbb{E} \left| \mathbb{E}(\rho_t | \mathcal{F}_{t-m}) - \mathbb{E}(\rho_t) \right| \\
& + (15/\varepsilon_T) T^{-1} \sum_{t=1}^{T} \mathbb{E} \left[|\rho_t| \, I\left(|\rho_t| > C_T \right) \right],
\end{aligned} \tag{10}
$$

as long as the RHS exists and is smaller than one.

The first term in the RHS depends on the dependence window m, the upper bound C_T, and ε_T. The second term on the RHS is called the *dependence term* and is described in the framework of ℓ_1-mixingales (see, e.g., Chapter 16, Davidson, 1994). The third term on the RHS captures the tail behaviour of ρ_t.

Lemma 4. *Assume there exist $c_d > 0$ and $d > 1$ such that $\mathbb{E}|\rho_t|^d < c_d$ for all t. Then, $\mathbb{E}[|\rho_t| I(|\rho_t| > C_t)] \leq c_d C_T^{-(d-1)}$ and the tail condition in (10) is satisfied with $\frac{15}{\varepsilon_T} T^{-1} \sum_{t=1}^{T} \mathbb{E}\left[|\rho_t| \, I\left(|\rho_t| > C_T \right) \right] \leq \frac{15 c_d}{\varepsilon_T C_T^{d-1}}$.*

Proof. It follows after application of Hölder and Markov inequalities that $\mathbb{E}[|\rho_t| I(|\rho_t| > C_t)] \leq \mathbb{E}\left(|\rho_t|^d \right)^{1/d}$ $\Pr(|\rho_t| > C_t)^{(d-1)/d} \leq \mathbb{E}(|\rho_t|^d)^{1/d} \mathbb{E}(|\rho_t|^d)^{(d-1)/p} / C_t^{d(d-1)/d} = \mathbb{E}(|\rho_t|^d) C_t^{-(d-1)}$. $\square$

Bounds on the *dependence term* are more involved and we shall focus on processes that can be arbitrarily approximated by a strong mixing sequence.

Lemma 5. *Let $\{\varepsilon_t\}_{-\infty}^{\infty}$ denote a strong mixing sequence with mixing coefficients $\{\alpha_m\}_{m=-\infty}^{\infty}$ and let $\mathcal{F}_t = \sigma\{\varepsilon_t, \varepsilon_{t-1}, \dots\}$. If $\rho_{t,k} = \rho(\varepsilon_{t-1}, \dots, \varepsilon_{t-k})$ is $\mathcal{F}_t$-measurable, for finite k, then $\rho_{t,k}$ is also strong mixing with coefficients $\{\alpha_{m-k}\}_{m=-\infty}^{\infty}$. Furthermore, any positive $d \geq 1$, $\mathbb{E}\left|\mathbb{E}(\rho_{t,k}|\mathcal{F}_{t-m}) - \mathbb{E}(\rho_{t,k})\right| \leq 6\alpha_{m-k}^{1-1/d}\|\rho_{t,k}\|_d$, as far as the RHS exists.*

Proof. The proof follows directly from Theorems 14.1 and 14.2 in Davidson (1994). □

Lemma 6. *Let $\{\rho_t\}_{t=-\infty}^{\infty}$ denote a $\mathcal{F}_t$ measurable process that can be arbitrarily approximated by a strong mixing process $\{\rho_{t,k}\}_{t=-\infty}^{\infty}$, with mixing coefficients $\{\alpha_{m-k}\}_{m=-\infty}^{\infty}$, in a sense that there exist finite constants $c_t > 0$ and a decreasing sequence $\{\nu_k\}_{k=0}^{\infty}$ such that $E|\rho_t - \rho_{k,t}| \leq c_t\nu_k$ (approximation assumption). Then,*

$$\mathbb{E}\left|\mathbb{E}(\rho_t|\mathcal{F}_{t-m}) - \mathbb{E}(\rho_t)\right| \leq 6\alpha_{m-k}^{1-1/d}\|\rho_{t,k}\|_d + 2c_t\nu_k. \tag{11}$$

Proof. It follows from the triangle inequality, Lemma 5, and the approximation assumption that

$$\begin{aligned}\mathbb{E}\left|\mathbb{E}(\rho_t|\mathcal{F}_{t-m}) - \mathbb{E}(\rho_t)\right| &\leq \mathbb{E}\left|\mathbb{E}(\rho_{t,k}|\mathcal{F}_{t-m}) - \mathbb{E}(\rho_{t,k})\right| + \mathbb{E}\left|\mathbb{E}(\rho_t|\mathcal{F}_{t-m}) - \mathbb{E}(\rho_{t,k}|\mathcal{F}_{t-m})\right| + \left|\mathbb{E}(\rho_t) - \mathbb{E}(\rho_{t,k})\right| \\ &\leq \mathbb{E}\left|\mathbb{E}(\rho_{t,k}|\mathcal{F}_{t-m}) - \mathbb{E}(\rho_{t,k})\right| + 2\mathbb{E}\left|\rho_t - \rho_{t,k}\right| \leq 6\alpha_{m-k}^{1-1/d}\|\rho_{t,k}\|_d + 2c_t\nu_k.\end{aligned}$$

□

Our goal is to use the triplex inequality to bound $\sum_{i=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t' - \mathbb{E}(\boldsymbol{x}_0\boldsymbol{x}_0')$ element-wise, where $\boldsymbol{x}_t$ is defined in (1). We consider $\rho_t^{(1)}(i) = y_t y_{t-i}$, $\rho_t^{(2)}(j) = \boldsymbol{\delta}_1'\boldsymbol{z}_t\boldsymbol{\delta}_2'\boldsymbol{z}_{t-j}$ and $\rho_t^{(3)}(i,j) = y_{t-i}\boldsymbol{\delta}_1'\boldsymbol{z}_{t-j}$, where $1 \leq i \leq p$, $0 \leq j \leq q$ and $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ are r-dimensional vectors of the form $\boldsymbol{e}_s = (0, \dots, 1, \dots, 0)'$ that have zeros in all positions except for $1 \leq s \leq d_z$.

Lemma 7. *Under Assumptions (A1)–(A5), for $m > p + q + k$: (1) $\mathbb{E}\left|\mathbb{E}\left[\rho_t^{(1)}(i)|\mathcal{F}_m\right] - \mathbb{E}\left[\rho_t^{(1)}(i)\right]\right| \leq 6(\zeta_0 c_d)^2\alpha_{m-i-q-k}^{1-1/d} + 2\zeta_0 c_2^2\zeta_{k+1}$; (2) $\mathbb{E}\left|\mathbb{E}\left[\rho_t^{(2)}(j)|\mathcal{F}_m\right] - \mathbb{E}\left[\rho_t^{(2)}(j)\right]\right| \leq 6\left(\sigma_{\eta,\max}\kappa_0\right)^2\alpha_{m-j-k}^{1-1/d} + 2\sigma_{\eta,\max}^2\kappa_0\kappa_{k+1}$; (3) $\mathbb{E}\left|\mathbb{E}\left[\rho_t^{(3)}(i,j)|\mathcal{F}_m\right] - \mathbb{E}\left[\rho_t^{(3)}(i,j)\right]\right| \leq \sigma_{\eta,\max}\left[6\zeta_0 c_d\kappa_0\alpha_{m-i-q-k}^{1-1/d} + (c_2\kappa_0 + c_d\zeta_0)(\kappa_{k+1} \vee \zeta_{k+1})\right]$.*

Proof. The proof consists on applying Lemma 6 to each $\rho_t^{(i)}$, using the approximation $\rho_{t,k}^{(i)}$ that has y_t replaced by $y_{t,k}$ and $\boldsymbol{z}_t$ replaced by $\boldsymbol{z}_{t,k}$, e.g., $\rho_{t,k}^{(1)}(i) = y_{t,k}y_{t-i,k}$.

We bound the first term on the RHS of equation (11). It follows from Lemmas 3 (1)–(2) and 5 that $\rho_{t,k}^{(1)}(i)$ is strong mixing with coefficients $\alpha_{m-i-q-k}$, $\rho_{t,k}^{(2)}(j)$ is strong mixing with coefficients α_{m-j-k} and $\rho_{t,k}^{(3)}(i,j)$ is strong mixing with coefficients $\alpha_{m-i-q-k}$. It also follows from Lemma 3 (3)–(4) and the Cauchy-Schwarz inequality that $\|\rho_{t,k}^{(1)}(i)\|_d \leq \|y_{t,k}\|_{2d}^2 \leq (\zeta_0 c_d)^2$, $\|\rho_{t,k}^{(2)}(j)\|_d \leq (\sigma_{\eta,\max}\kappa_0)^2$, and $\|\rho_{t,k}^{(3)}(i,j)\|_d \leq \zeta_0 c_d\sigma_{\eta,\max}\kappa_0$.

We use the following basic inequality to bound the second term on the RHS of equation (11). For any random variables X,Y,X' and Y', $\|XY - X'Y'\|_1 \leq \|X\|_2\|Y - Y'\|_2 + \|Y'\|_2\|X - X'\|_2$. Applying this inequality to $\|\rho_t^{(1)}(i) - \rho_{t,k}^{(1)}(i)\|_1$, and using Lemmas 1 and 3 we have

$$\|\rho_t^{(1)}(i) - \rho_{t,k}^{(1)}(i)\|_1 \leq \|y_t\|\|y_{t-i} - y_{t-i,k}\| + \|y_{t-i,k}\|\|y_t - y_{t,k}\| \leq 2\zeta_0 c_2 \times c_d \zeta_{k+1} \leq 2\zeta_0 c_2^2 \zeta_{k+1}.$$

Similarly, $\|\rho_t^{(2)}(j) - \rho_{t,k}^{(2)}(j)\|_1 \leq 2\sigma_{\eta,\max}^2 \kappa_0 \kappa_{k+1}$ and $\|\rho_t^{(3)}(i,j) - \rho_{t,k}^{(3)}(i,j)\|_1 \leq \sigma_{\eta,\max}(c_2\kappa_0 + c_d\zeta_0)(\kappa_{k+1} \vee \zeta_{k+1})$. The result follows by combining the bounds. $\square$

Lemma 8. *Let (A1)–(A4) hold and T be sufficiently large. Suppose that $\alpha_m = \alpha^m$ $(0 < \alpha < 1)$, $\kappa_m = \kappa^m$ $(0 < \kappa < 1)$, and $\zeta_m = \zeta^m$ $(0 < \zeta < 1)$. Then, for constants $a_1, a_2 > 0$, $p = q$, and s the number of non-zero coefficients in model* (1),

$$\Pr\left\{\max_{1 \leq i,j \leq p+qd_z} \left| \left[\sum_{t=p}^{T} \boldsymbol{x}_t \boldsymbol{x}_t' - \mathbb{E}(\boldsymbol{x}_t \boldsymbol{x}_t')\right]_{i,j} \right| \geq T\frac{a_1}{s}\right\} \leq a_2 \frac{sp^2 d_z^2}{T^{(d-1)(1-4\nu)/2}}, \tag{12}$$

where $0 < \nu < 1/4$, $s \vee p = o(T^\nu)$, and d_z is at most polynomial in T.[2]

Proof. We use $c^{(i)}$, $i = 1, 2, \ldots$, as arbitrary constants that do not depend on p, q, d_z, or T. Despite the bound being derived for each T, it is of asymptotic nature in a sense that it is not our goal to provide optimal constants, and that some steps will hold only "for all $T \geq T_0$ sufficiently large".

The proof consists in applying the triplex inequality to the processes $\left\{\left[\sum_{i=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t' - \mathbb{E}(\boldsymbol{x}_0 \boldsymbol{x}_0')\right]_{i,j}\right\}_T$ for $1 \leq i, j \leq p + q \times d_z$. We shall split the terms into three blocks. The first one involves the terms in $\tilde{\boldsymbol{y}}_{t-1}\tilde{\boldsymbol{y}}_{t-1}'$ where $\tilde{\boldsymbol{y}}_{t-1} = (y_{t-1}, \ldots, y_{t-p})'$, the second block has terms $\tilde{\boldsymbol{z}}_t \tilde{\boldsymbol{z}}_t'$, where $\tilde{\boldsymbol{z}}_t = (\boldsymbol{z}_t', \ldots, \boldsymbol{z}_{t-q}')'$, and the third block involves terms $\tilde{\boldsymbol{y}}_{t-1}\tilde{\boldsymbol{z}}_t'$.

Using the union bound on the first block, we have

$$\begin{aligned}\Pr&\left[\max_{1 \leq i,j \leq p} \left| \left[\sum_{t=p}^{T} \tilde{\boldsymbol{y}}_{t-1}\tilde{\boldsymbol{y}}_{t-1}' - \mathbb{E}(\tilde{\boldsymbol{y}}_{t-1}\tilde{\boldsymbol{y}}_{t-1}')\right]_{i,j} \right| \geq T\frac{a_1}{4s}\right] \\ &\leq p^2 \max_{1 \leq i,j \leq p} \Pr\left[\left|\sum_{t=p}^{T} y_{t-i}y_{t-j} - \mathbb{E}(y_0 y_{i-j})\right| > T\frac{a_1}{4s}\right].\end{aligned} \tag{13}$$

The *dependence term* is bounded using Lemma 7 (1). Let $\tau = -(\log\alpha \vee \log\zeta \vee \log\kappa)$ and $m = T^\nu$, for some $0 < \nu < 1/4$ and $p + q = o(T^\nu)$. Then, the dependence term is bounded by

$$p^2 s \max_{1 \leq i,j \leq p} T^{-1} \sum_{t=p}^{T} \mathbb{E}|\mathbb{E}(y_{t-i}y_{t-j}|\mathcal{F}_{t-m}) - \mathbb{E}(y_{t-i}y_{t-j})| \leq c^{(1)} p^2 s e^{-c^{(2)}\tau T^\nu/2}.$$

First we choose $k = [(d-1)/(2d-1)](m-p-q)$ and set $e^{(1-1/d)(m-p-q-k)\log\alpha} \leq e^{-[(d-1)/(2d-1)]\tau(m-p-q)}$, then use the fact that, for T sufficiently large, $p + q < m/2$, to reach the result.

[2]We use the notation $[A]_{i,j}$ to denote the $(i,j)^{th}$ element of matrix A.

In the *tail term*, we use lemmas 1 and 4. Let $C_T = T^{(1-4\nu)/2}$,

$$p^2 s \max_{1\le i,j\le p} T^{-1}\sum_{t=p}^{T}\mathbb{E}\left[|y_{t-i}y_{t-j}|I(|y_{t-i}y_{t-j}|>C_T)\right] \le c^{(3)}p^2\left[\frac{s^{1/(d-1)}}{T^{(1-4\nu)/2}}\right]^{d-1}.$$

First we use the Cauchy-Schwarz inequality to find $\mathbb{E}|y_{t-i}y_{t-j}|^d \le \mathbb{E}|y_0|^{2d} < (\zeta_0 c_d)^{2d}$, then we plug this value in Lemma 4 and reach the result.

The first term in the triplex inequality involves $T/C_T^2 m^2 = T^{1-2\nu-(1-4\nu)} = T^{2\nu}$. Then,

$$(13) \le 2T^{\nu}p^2 e^{-c^{(4)}\left(\frac{T^{\nu}}{s}\right)^2} + c^{(1)}p^2 s e^{-c^{(2)}\tau T^{\nu}/2} + c^{(3)}\left[\frac{p^{2/(d-1)}s^{1/(d-1)}}{T^{(1-4\nu)/2}}\right]^{d-1}.$$

The second block is treated similarly. We first apply the union bound

$$\Pr\left[\max_{1\le i,j\le qd_z}\left|\sum_{t=q}^{T}\tilde{\boldsymbol{z}}_t\tilde{\boldsymbol{z}}_t' - \mathbb{E}(\tilde{\boldsymbol{z}}_t\tilde{\boldsymbol{z}}_t')\right|_{i,j} \ge T\frac{a_1}{4s}\right]$$
$$\le q^2 d_z^2 \max_{1\le i,j\le qd_z}\Pr\left[\left|\sum_{t=q}^{T}\tilde{\boldsymbol{z}}_{t,i}\tilde{\boldsymbol{z}}_{t,j} - \mathbb{E}(\tilde{\boldsymbol{z}}_{t,i}\tilde{\boldsymbol{z}}_{t,j})\right| > T\frac{a_1}{4s}\right]. \quad (14)$$

Using the same arguments as before, we find

$$(14) \le 2T^{\nu}(qd_z)^2 e^{-c^{(5)}\left(\frac{T^{\nu}}{s}\right)^2} + c^{(6)}(qd_z)^2 s e^{-c^{(7)}\tau T^{\nu}/2} + c^{(8)}\left[\frac{(qd_z)^{2/(d-1)}s^{1/(d-1)}}{T^{(1-4\nu)/2}}\right]^{d-1}.$$

The third block is treated in the same way of the previous two. We first apply the union bound

$$\Pr\left[\max_{1\le i\le p,1\le j\le qd_z}\left|\sum_{t=q}^{T}\tilde{\boldsymbol{y}}_{t-1}\tilde{\boldsymbol{z}}_t' - \mathbb{E}(\tilde{\boldsymbol{y}}_{t-1}\tilde{\boldsymbol{z}}_t')\right|_{i,j} \ge T\frac{a_1}{4s}\right]$$
$$\le pqd_z \max_{1\le i\le p,1\le j\le qd_z}\Pr\left[\left|\sum_{t=q}^{T}y_{t-i}\tilde{\boldsymbol{z}}_{t,j} - \mathbb{E}(y_{t-i}\tilde{\boldsymbol{z}}_{t,j})\right| > T\frac{a_1}{4s}\right]. \quad (15)$$

Using the same arguments as before, we find

$$(15) \le 2T^{\nu}pqd_z e^{-c^{(9)}\left(\frac{T^{\nu}}{s}\right)^2} + c^{(10)}pqd_z s e^{-c^{(11)}\tau T^{\nu}/2} + 2c^{(12)}\left[\frac{(pqd_z)^{1/(d-1)}s^{1/(d-1)}}{T^{(1-4\nu)/2}}\right]^{d-1}.$$

The last term is the dominant one in each inequality. Using the fact that $p = q$, that all terms are positive, and choosing $a_2 = c^{(12)}$ sufficiently large: $(13) + (14) + (15) \le c^{(12)}\frac{sp^2 d_z^2}{T^{(d-1)(1-4\nu)/2}}$.

□

The right-hand side of this probability inequality converges to zero if $s = O\left[pd_z^{2/(d-1)}\right]$ and $pd_z = o\left\{T^{[(d-1)^2/4d](1-4\nu)}\right\}$. In this case, pd_z is superlinear whenever $d > 1 + 2(1+\sqrt{2-4\nu})/(1-4\nu)$. If $s = O(1)$ and $pd_z = o(T^{[(d-1)/4](1-4\nu)})$, the right-hand side also converges to zero and pd_z is superlinear whenever $d > 1 + 4/(1-4\nu)$.

References

Basu, S. and G. Michailidis: 2015, 'Regularized estimation in sparse high-dimensional time series models'. *The Annals of Statistics* **43**, 1535–1567.

Bülhmann, P. and S. van der Geer: 2011, *Statistics for High Dimensional Data*. Springer.

Davidson, J.: 1994, *Stochastic Limit Theory*. Oxford: Oxford University Press.

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

Fan, Y., J. Fan, and E. Barut: 2014, 'Adaptive Robust Variable Selection'. *Annals of Statistics* **42**, 324–351.

Hentschel, L.: 1995, 'All in the family: Nesting Symmetric and AQsymmetric GARCH Models'. *Journal of Financial Economics* **39**, 71–104.

Jiang, W.: 2009, 'On uniform deviations of general empirical risks with unboundedness, dependence and high dimensionality'. *Journal of Machine Learning Research* **10**, 977–996.

Kock, A. and L. Callot: 2015, 'Oracle Inequalities for High Dimensional Vector Autoregressions'. *Journal of Econometrics* **186**, 325–344.

Liao, Z. and P. Phillips: 2015, 'Automated Estimation of Vector Error Correction Models'. *Econometric Theory* **31**, 581–646.

Lindner, A. M.: 2009, 'Stationarity, mixing, distributional properties and moments of GARCH (p, q)–processes'. In: *Handbook of financial time series*. Springer, pp. 43–69.

Medeiros, M. and E. Mendes: 2016, 'ℓ_1-Regularization of High-dimensional Time-Series Models with non-Gaussian and Heteroskedastic Innovations'. *Journal of Econometrics* **191**(1).

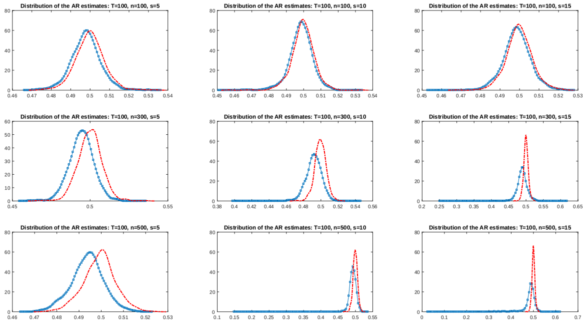
Meitz, M. and P. Saikkonen: 2008, 'Ergodicity, Mixing, and Existence of Moments of a Class of Markov models with applications to GARCH and ACD models'. *Econometric Theory* **24**, 1291–1320.

Nardi, Y. and A. Rinaldo: 2011, 'Autoregressive process modeling via the Lasso procedure'. *Journal of Multivariate Analysis* **102**, 528–549.

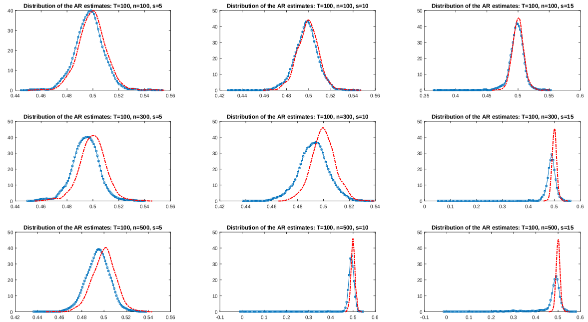
Tibshirani, R.: 1996, 'Regression shrinkage and selection via the Lasso'. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(1), 267–288.

Zou, H.: 2006, 'The adaptive lasso and its oracle properies'. *Journal of the American Statistical Association* **101**, 1418–1429.

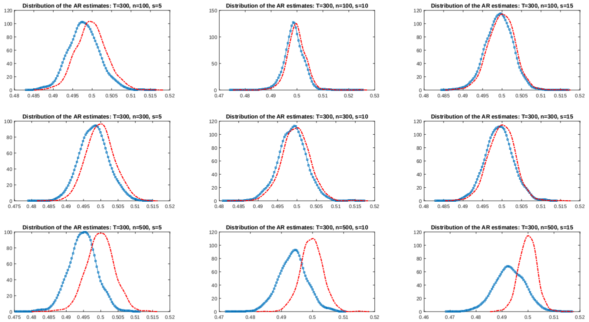
Zou, H., T. Hastie, and R. Tibshirani: 2007, 'On the degrees of freedom of the lasso'. *Annals of Statistics* **35**, 2173–2192.



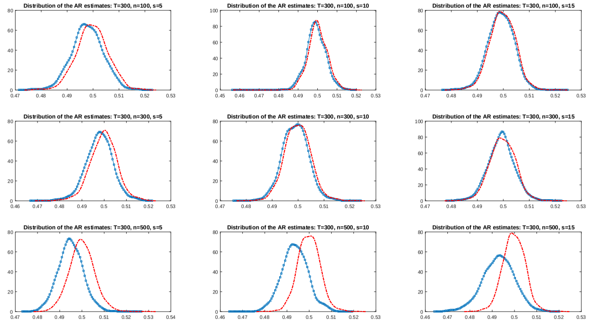
(a) $T = 100$, $\alpha_1 = 0.01$ and $\pi_1 = 0.8$



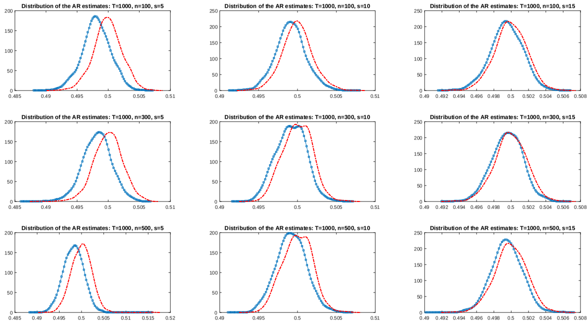
(b) $T = 100$, $\alpha_1 = 0.05$ and $\pi_1 = 0.9$



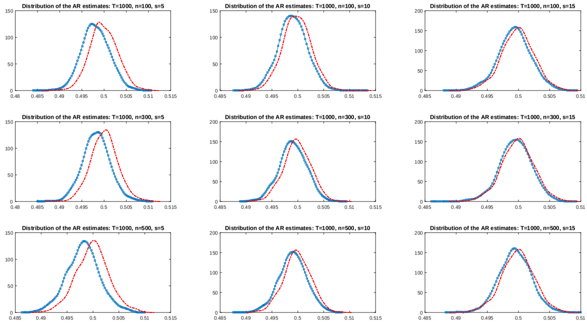
(c) $T = 300$, $\alpha_1 = 0.01$ and $\pi_1 = 0.8$



(d) $T = 300$, $\alpha_1 = 0.05$ and $\pi_1 = 0.9$

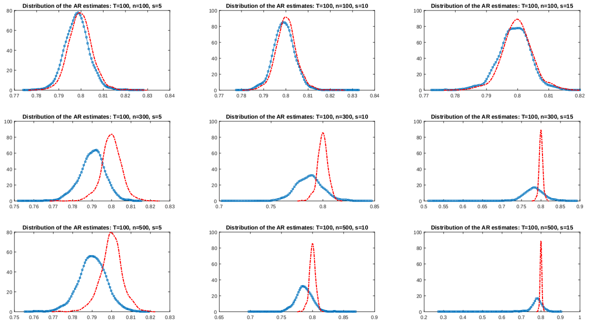


(e) $T = 1000$, $\alpha_1 = 0.01$ and $\pi_1 = 0.8$

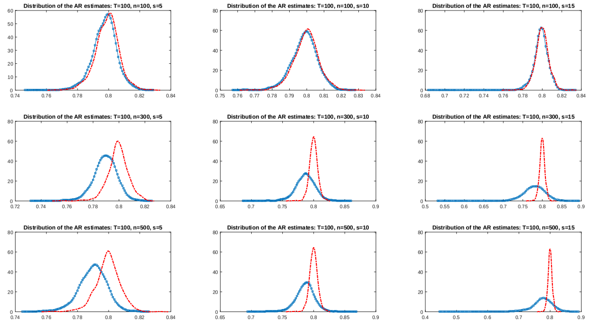


(f) $T = 1000$, $\alpha_1 = 0.05$ and $\pi_1 = 0.9$

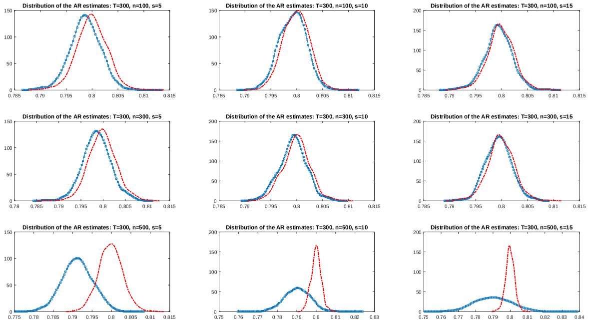
FIGURE 1. Distribution of the adaLASSO (blue) and Oracle estimators (red) for the parameter $\phi = 0.5$.



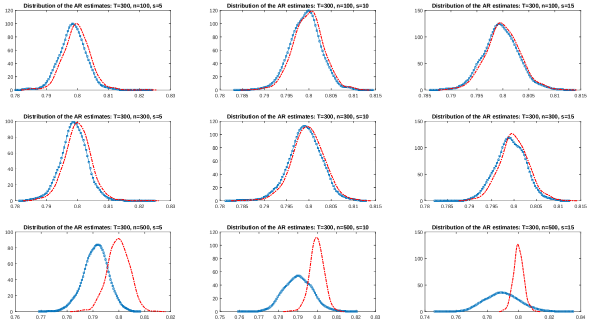
(a) $T = 100$, $\alpha_1 = 0.01$ and $\pi_1 = 0.8$



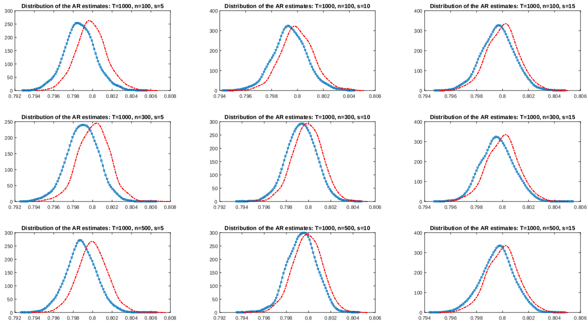
(b) $T = 100$, $\alpha_1 = 0.05$ and $\pi_1 = 0.9$



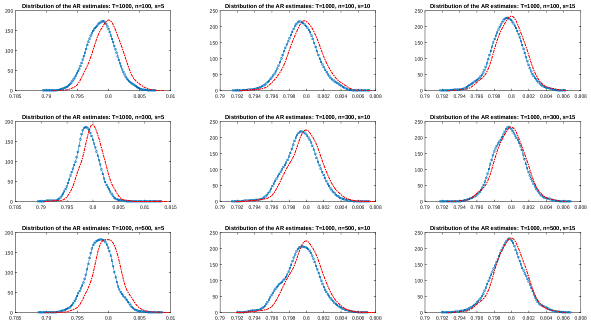
(c) $T = 300$, $\alpha_1 = 0.01$ and $\pi_1 = 0.8$



(d) $T = 300$, $\alpha_1 = 0.05$ and $\pi_1 = 0.9$

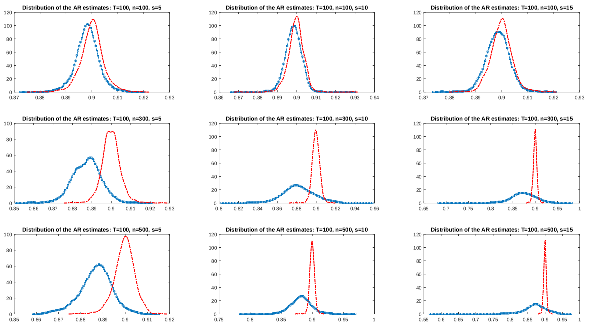


(e) $T = 1000$, $\alpha_1 = 0.01$ and $\pi_1 = 0.8$

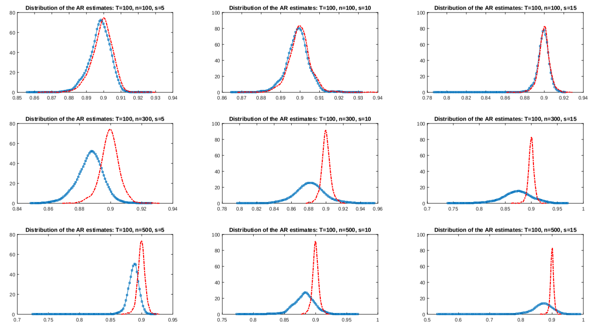


(f) $T = 1000$, $\alpha_1 = 0.05$ and $\pi_1 = 0.9$

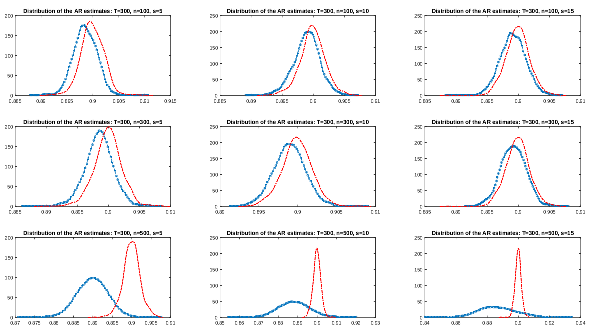
FIGURE 2. Distribution of the adaLASSO (blue) and Oracle estimators (red) for the parameter $\phi = 0.8$.



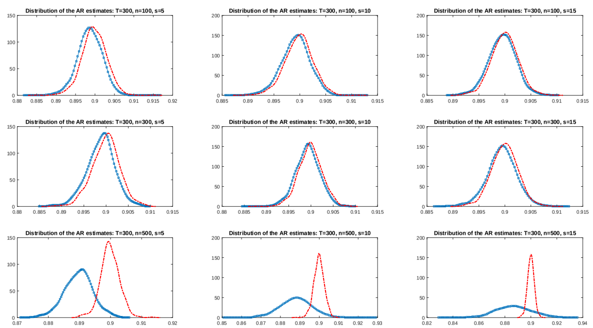
(a) $T = 100$, $\alpha_1 = 0.01$ and $\pi_1 = 0.8$



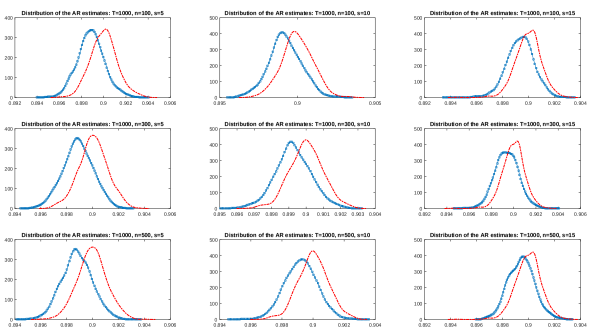
(b) $T = 100$, $\alpha_1 = 0.05$ and $\pi_1 = 0.9$



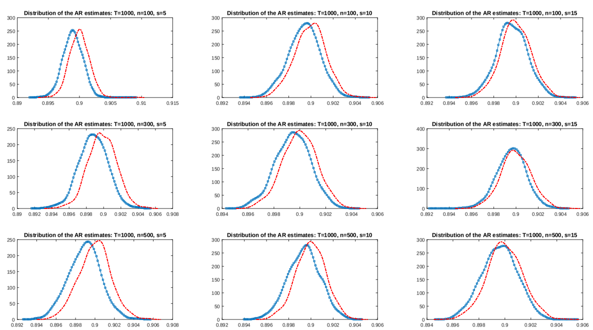
(c) $T = 300$, $\alpha_1 = 0.01$ and $\pi_1 = 0.8$



(d) $T = 300$, $\alpha_1 = 0.05$ and $\pi_1 = 0.9$



(e) $T = 1000$, $\alpha_1 = 0.01$ and $\pi_1 = 0.8$



(f) $T = 1000$, $\alpha_1 = 0.05$ and $\pi_1 = 0.9$

FIGURE 3. Distribution of the adaLASSO (blue) and Oracle estimators (red) for the parameter $\phi = 0.9$.

Table 1. Parameter Estimates: Descriptive Statistics for Model (5) with GARCH parameters.

The table reports the average absolute bias, Panels (a) and (c), and the average mean squared error (MSE), Panels (b) and (d), over all parameter estimates and Monte Carlo simulations. n is the number of candidate variables whereas q is the number of relevant regressors.

		$T=100$			$T=300$			$T=1000$		
ϕ	$q\backslash n$	100	300	500	100	300	500	100	300	500
					$\alpha_0 = 5\times10^{-4}$, $\alpha_1 = 0.1$ and $\pi_1 = 0.8$					
					Panel (a): Bias ($\times10^{-3}$)					
	5	-1.0207	-0.3498	-0.2125	-1.0219	-0.3400	-0.2104	-1.0192	-0.3391	-0.2036
0.5	10	-1.0224	-0.3612	-0.2170	-1.0125	-0.3374	-0.2113	-1.0091	-0.3362	-0.2017
	15	-1.0228	-0.3711	-0.2442	-1.0080	-0.3352	-0.2132	-1.0048	-0.3345	-0.2011
	5	1.9840	0.6324	0.3802	1.9862	0.6615	0.3824	1.9881	0.6634	0.3977
0.8	10	1.9993	0.6439	0.3873	1.9958	0.6655	0.3944	1.9972	0.6656	0.3991
	15	1.9889	0.6000	0.3572	1.9991	0.6647	0.3777	1.9963	0.6661	0.3990
	5	2.9817	0.9584	0.5741	2.9867	0.9956	0.5797	2.9889	0.9960	0.5976
0.9	10	2.9919	0.9837	0.5951	2.9980	0.9998	0.6018	3.0001	1.0005	0.5999
	15	2.9816	0.9096	0.5366	2.9936	0.9962	0.5708	2.9952	0.9978	0.5991
					Panel (b): MSE ($\times10^{-3}$)					
	5	0.1073	0.0394	0.0239	0.1054	0.0351	0.0230	0.1045	0.0348	0.0209
0.5	10	0.1097	0.0603	0.0423	0.1046	0.0349	0.0306	0.1029	0.0343	0.0206
	15	0.1154	0.1261	0.1872	0.1047	0.0349	0.0456	0.1026	0.0341	0.0205
	5	0.3963	0.1383	0.0838	0.3966	0.1320	0.0835	0.3960	0.1324	0.0793
0.8	10	0.4053	0.2871	0.1798	0.4002	0.1335	0.1459	0.3996	0.1331	0.0799
	15	0.4102	0.7195	0.5328	0.4026	0.1342	0.2865	0.4011	0.1336	0.0802
	5	0.8976	0.3264	0.1953	0.8955	0.2986	0.1998	0.8959	0.2987	0.1791
0.9	10	0.9079	0.6892	0.4367	0.9045	0.3016	0.3710	0.9034	0.3011	0.1806
	15	0.9174	1.4300	0.9744	0.9097	0.3032	0.7424	0.9071	0.3023	0.1814
					$\alpha_0 = 5\times10^{-4}$, $\alpha_1 = 0.05$ and $\pi_1 = 0.9$					
					Panel (c): Bias ($\times10^{-3}$)					
	5	-1.0244	-0.3522	-0.2107	-1.0192	-0.3410	-0.2100	-1.0193	-0.3391	-0.2037
0.5	10	-1.0173	-0.3572	-0.2195	-1.0074	-0.3364	-0.2112	-1.0091	-0.3367	-0.2019
	15	-1.0227	-0.3749	-0.2634	-1.0097	-0.3366	-0.2144	-1.0044	-0.3350	-0.2012
	5	1.9749	0.6321	0.3802	1.9855	0.6628	0.3820	1.9893	0.6632	0.3973
0.8	10	1.9868	0.6451	0.3876	1.9977	0.6641	0.3921	1.9969	0.6656	0.3993
	15	2.0192	0.5991	0.3402	1.9953	0.6639	0.3757	1.9968	0.6655	0.3991
	5	2.9719	0.9558	0.5744	2.9876	0.9953	0.5793	2.9884	0.9959	0.5979
0.9	10	2.9980	0.9858	0.6002	2.9966	0.9995	0.5985	2.9984	0.9988	0.6002
	15	2.9912	0.8977	0.5383	2.9931	0.9983	0.5761	2.9978	0.9989	0.5992
					Panel (d): MSE ($\times10^{-3}$)					
	5	0.1132	0.0403	0.0240	0.1063	0.0354	0.0231	0.1047	0.0348	0.0209
0.5	10	0.1196	0.0620	0.0500	0.1056	0.0353	0.0311	0.1036	0.0345	0.0207
	15	0.1642	0.1435	0.2494	0.1072	0.0358	0.0472	0.1034	0.0344	0.0206
	5	0.3957	0.1387	0.0847	0.3965	0.1327	0.0834	0.3964	0.1323	0.0792
0.8	10	0.4093	0.2973	0.1955	0.4012	0.1337	0.1493	0.3998	0.1333	0.0800
	15	0.4580	0.7473	0.5945	0.4049	0.1345	0.2923	0.4008	0.1339	0.0803
	5	0.8927	0.3280	0.1994	0.8953	0.2987	0.2006	0.8959	0.2983	0.1792
0.9	10	0.9077	0.7055	0.4330	0.9010	0.3004	0.3778	0.8999	0.3001	0.1799
	15	0.9351	1.4828	0.9913	0.9050	0.3016	0.7543	0.9020	0.3006	0.1804

Table 2. Model Selection: Descriptive Statistics for Model (5) with GARCH parameters $\alpha_0 = 5 \times 10^{-4}$, $\alpha_1 = 0.1$ and $\pi_1 = 0.8$.

The table reports several statistics concerning model selection. Panel (a) presents the fraction of replications where the correct model has been selected, i.e., all the relevant variables included and all the irrelevant regressors excluded. Panel (b) shows the fraction of replications where the relevant variables are all included. Panel (c) presents the fraction of relevant variables included. Panel (d) shows the fraction of irrelevant variables excluded.

		$\alpha_0 = 5 \times 10^{-4}$, $\alpha_1 = 0.1$ and $\pi_1 = 0.8$									$\alpha_0 = 5 \times 10^{-4}$, $\alpha_1 = 0.05$ and $\pi_1 = 0.9$								
		$T = 100$			$T = 300$			$T = 1000$			$T = 100$			$T = 300$			$T = 1000$		
ϕ	$q \backslash n$	100	300	500	100	300	500	100	300	500	100	300	500	100	300	500	100	300	500
		Panel (a): Correct Sparsity Pattern									Panel (a): Correct Sparsity Pattern								
	5	0.994	1.000	0.999	0.994	1.000	1.000	0.994	1.000	1.000	0.972	0.997	0.999	0.972	1.000	1.000	0.972	1.000	1.000
0.5	10	0.958	0.997	0.986	1.000	1.000	1.000	1.000	1.000	1.000	0.800	0.991	0.977	0.997	0.999	1.000	1.000	1.000	1.000
	15	0.864	0.965	0.830	0.999	0.999	1.000	1.000	1.000	1.000	0.590	0.944	0.741	0.995	0.988	1.000	1.000	1.000	1.000
	5	0.999	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000	0.984	1.000	0.999	0.984	1.000	1.000	0.984	1.000	1.000
0.8	10	0.977	0.997	0.979	1.000	0.999	1.000	1.000	1.000	1.000	0.902	0.994	0.980	1.000	1.000	1.000	1.000	1.000	1.000
	15	0.918	0.818	0.702	1.000	0.999	0.999	1.000	1.000	1.000	0.704	0.793	0.611	0.999	0.996	0.999	1.000	1.000	1.000
	5	0.998	1.000	1.000	0.998	1.000	1.000	0.998	1.000	1.000	0.988	1.000	0.999	0.988	1.000	1.000	0.988	1.000	1.000
0.9	10	0.983	0.961	0.931	1.000	1.000	1.000	1.000	1.000	1.000	0.934	0.943	0.936	1.000	0.999	0.999	1.000	1.000	1.000
	15	0.952	0.449	0.348	0.999	1.000	0.836	1.000	1.000	1.000	0.782	0.389	0.305	0.999	1.000	0.813	1.000	1.000	1.000
		Panel (b): True Model Included									Panel (b): True Model Included								
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5	10	1.000	0.998	0.994	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.989	1.000	1.000	1.000	1.000	1.000	1.000
	15	1.000	0.983	0.875	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.979	0.836	1.000	1.000	1.000	1.000	1.000	1.000
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.8	10	1.000	0.997	0.981	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.981	1.000	1.000	1.000	1.000	1.000	1.000
	15	1.000	0.818	0.708	1.000	1.000	0.999	1.000	1.000	1.000	0.998	0.800	0.619	1.000	1.000	0.999	1.000	1.000	1.000
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.9	10	1.000	0.961	0.931	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.943	0.938	1.000	1.000	0.999	1.000	1.000	1.000
	15	1.000	0.449	0.348	1.000	1.000	0.836	1.000	1.000	1.000	0.999	0.390	0.306	1.000	1.000	0.813	1.000	1.000	1.000
		Panel (c): Fraction of Relevant Variables Included									Panel (c): Fraction of Relevant Variables Included								
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5	10	1.000	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	1.000	1.000	1.000	1.000	1.000	1.000
	15	1.000	0.997	0.969	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.994	0.954	1.000	1.000	1.000	1.000	1.000	1.000
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.8	10	1.000	0.999	0.994	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.995	1.000	1.000	1.000	1.000	1.000	1.000
	15	1.000	0.969	0.919	1.000	1.000	0.999	1.000	1.000	1.000	0.999	0.964	0.892	1.000	1.000	0.999	1.000	1.000	1.000
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.9	10	1.000	0.994	0.984	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.990	0.989	1.000	1.000	0.999	1.000	1.000	1.000
	15	1.000	0.874	0.802	1.000	1.000	0.982	1.000	1.000	1.000	0.999	0.862	0.789	1.000	1.000	0.979	1.000	1.000	1.000
		Panel (d): Fraction of Irrelevant Excluded									Panel (d): Fraction of Irrelevant Excluded								
	5	0.999	1.000	1.000	1.0000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5	10	0.998	0.999	0.999	1.0000	1.000	1.000	1.000	1.000	1.000	0.991	1.000	0.999	0.999	1.000	1.000	1.000	1.000	1.000
	15	0.992	0.999	0.996	1.0000	1.000	1.000	1.000	1.000	1.000	0.968	0.999	0.994	0.999	0.999	1.000	1.000	1.000	1.000
	5	1.000	1.000	1.000	1.0000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.8	10	0.999	1.000	1.000	1.0000	1.000	1.000	1.000	1.000	1.000	0.996	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	15	0.998	0.999	0.999	1.0000	1.000	1.000	1.000	1.000	1.000	0.986	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	5	0.999	1.000	1.000	1.0000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.9	10	0.999	1.000	1.000	1.0000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	15	0.998	0.999	0.999	1.0000	1.000	1.000	1.000	1.000	1.000	0.992	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000

Table 3. Forecasting: Ratio of the mean squared errors (MSE) for Model (5).

The table reports the ratio of the one-step-ahead mean squared error (MSE) for the adaLASSO over the oracle. n is the number of candidate variables whereas q is the number of relevant regressors.

		$T = 100$			$T = 300$			$T = 1000$		
ϕ	$q\backslash n$	100	300	500	100	300	500	100	300	500
				$\alpha_0 = 5 \times 10^{-4}$, $\alpha_1 = 0.1$ and $\pi_1 = 0.8$						
	5	1.0116	1.0810	1.0808	1.0101	1.0087	1.0726	1.0093	1.0087	1.0104
0.5	10	1.0443	2.0360	2.6073	1.0160	1.0156	1.7135	1.0164	1.0153	1.0169
	15	1.1458	5.0896	16.1101	1.0209	1.0205	3.0673	1.0187	1.0180	1.0199
	5	1.0219	1.9611	2.0077	1.0190	1.0181	1.9998	1.0198	1.0186	1.0203
0.8	10	1.0784	9.3232	10.3846	1.0349	1.0344	7.6378	1.0287	1.0326	1.0276
	15	1.1216	30.8723	41.3265	1.0497	1.0472	21.4417	1.0350	1.0374	1.0359
	5	1.1147	3.4576	3.6502	1.0673	1.0724	3.9863	1.0513	1.0522	1.0534
0.9	10	1.2201	22.3899	23.5280	1.1768	1.1818	19.8754	1.1448	1.1427	1.1407
	15	1.3263	59.1769	71.7225	1.2300	1.2218	53.5847	1.2080	1.1848	1.2017
				$\alpha_0 = 5 \times 10^{-4}$, $\alpha_1 = 0.05$ and $\pi_1 = 0.9$						
	5	1.0337	1.0413	1.0379	1.0074	1.0046	1.0384	1.0047	1.0046	1.0039
0.5	10	1.0733	1.5114	2.0495	1.0082	1.0072	1.3803	1.0078	1.0081	1.0069
	15	1.7215	3.3597	11.6598	1.0110	1.0110	2.0726	1.0096	1.0103	1.0081
	5	1.0176	1.4773	1.5535	1.0095	1.0099	1.4911	1.0082	1.0096	1.0102
0.8	10	1.0613	5.5934	6.2592	1.0162	1.0143	4.5985	1.0152	1.0160	1.0143
	15	1.5607	16.7974	23.0624	1.0177	1.0181	10.6860	1.0154	1.0178	1.0168
	5	1.0256	2.4542	2.5230	1.0177	1.0189	2.6059	1.0175	1.0164	1.0173
0.9	10	1.0579	12.1933	12.4650	1.0302	1.0280	10.9742	1.0241	1.0252	1.0245
	15	1.2584	32.2624	35.9492	1.0339	1.0339	27.9254	1.0260	1.0303	1.0294