Testing for predictability in a noninvertible ARMA model

Lanne, Markku and Meitz, Mika and Saikkonen, Pentti

2012

1 Introduction

Testing for autocorrelation or predictability in time series data has attracted considerable interest in various fields of application. A prominent example is testing the predictability of asset returns which has rather a long history in empirical finance (see Campbell, Lo, and MacKinlay (1997, Ch. 2)). As serial correlation in asset returns is weak at best, tests based directly on the sample autocorrelation function tend to lack power, and testing has, in addition, been based on time series processes implied by structural models. Amongst these, the first-order autoregressive moving average (ARMA(1,1)) process implied by the price-trend model of Taylor (1982) and the mean-reversion model of Poterba and Summers (1988) has been influential.

Testing for autocorrelation within the conventional (stationary and invertible) ARMA(1,1) model is a nonstandard problem, and appropriate tests have been proposed by Andrews and Ploberger (1996) and Nankervis and Savin (2010), whose work is closely related to ours. However, these authors, as well as all others we are aware of, base their tests on invertible ARMA models. In this paper, we take a different approach and develop tests for autocorrelation in the context of a noninvertible ARMA(1,1) model. We argue that, in addition to the convenience of leading to standard tests, the employed noninvertible ARMA model may also provide a more powerful framework for testing autocorrelation or predictability than its invertible counterpart. One reason for this is that, unlike its invertible counterpart, the noninvertible ARMA model is capable of capturing conditional heteroskedasticity likely to be involved in many empirical applications. Furthermore, while the tests of Andrews and Ploberger (1996) and Nankervis and Savin (2010) are only designed for testing for autocorrelation, our noninvertible ARMA model also makes possible to discriminate between autocorrelation and nonlinear predictability. These convenient features require that the data generation process is non-Gaussian which can be seen as a necessary identification condition. However, from the practical point of view this may not be a serious limitation in that Gaussianity is frequently rejected in applications, especially when testing the predictability of asset returns.

As a starting point both Andrews and Ploberger (1996) and Nankervis and Savin (2010) have the stationary and invertible ARMA(1,1) model

$$y_t = \phi_0 y_{t-1} + \varepsilon_t - \vartheta_0 \varepsilon_{t-1}, \tag{1}$$

where the parameters ϕ_0 and ϑ_0 satisfy $|\phi_0| < 1$ and $|\vartheta_0| < 1$, and ε_t is an uncorrelated zero mean error term, that is, white noise. For testing purposes the model is reparameterized as

$$y_t = (\vartheta_0 + \beta_0)\, y_{t-1} + \varepsilon_t - \vartheta_0 \varepsilon_{t-1},$$

where $\beta_0 = \phi_0 - \vartheta_0$. The null hypothesis of interest is that y_t is white noise or that $\beta_0 = 0$. As the null hypothesis implies that $y_t = \varepsilon_t$, the nuisance parameter ϑ_0 is present only under the alternative, explaining why the testing problem is nonstandard (cf. Davies (1977)).

Based on a Gaussian likelihood Andrews and Ploberger (1996) show that their likelihood ratio (LR) test and so-called supremum Lagrange multiplier (LM) tests and average exponential LM tests have desirable asymptotic properties, and they justify the asymptotic distributions of these tests without invoking Gaussianity and independence that were initially used to motivate the tests. Nankervis and Savin (2010) modify (some of) the tests of Andrews and Ploberger (1996) to make them applicable to a wider range of data generation processes. Using a general near epoch dependence assumption, they develop several test statistics with the same asymptotic distributions as the corresponding tests statistics of Andrews and Ploberger (1996). Unlike the tests of Andrews and Ploberger (1996), those of Nankervis and Savin (2010) are valid for data that are uncorrelated but dependent exhibiting, for instance, ARCH type conditional heteroskedasticity.

A noninvertible version of the ARMA(1,1) model (1) is obtained by assuming $|\vartheta_0| > 1$ instead of $|\vartheta_0| < 1$. Asymptotic estimation theory of this kind of ARMA models has been studied in several papers (see, e.g., Lii and Rosenblatt (1996) and Wu and Davis (2010)). This theory is not limited to the first order case, and it also allows for the possibility that the autoregressive part is noncausal, which in the special case (1) means that $|\phi_0| > 1$. We shall not consider noncausal ARMA models in this paper. Some of the recent work on the estimation of noninvertible ARMA models has been focused on so-called all-pass models which in the first order (causal) case are obtained from (1) by imposing the restriction $\phi_0 = 1/\vartheta_0$ ($|\vartheta_0| > 1$) (see Breidt, Davis, and Trindade (2001) and Andrews, Davis, and Breidt (2006, 2007)). From the viewpoint of testing for serial correlation, all-pass processes are interesting in that they can generate time series that are uncorrelated but dependent. This is a useful feature in testing for predictability as it facilitates detecting nonlinear dependence in case no autocorrelation is found. However, as already indicated the dependence of an all-pass process requires non-Gaussian data which is also required in the estimation theory discussed above as well as in the tests developed in this paper (see the aforementioned references and the discussion in the next section).

The LR and Wald tests to be derived in this paper are not directly based on the formulation of the noninvertible ARMA(1,1) model discussed in the preceding paragraph and the references therein. Instead, we use a formulation similar to that in Meitz and Saikkonen (2011) where parameter estimation in a noninvertible ARMA model with autoregressive conditionally heteroskedastic errors is studied. The reason for this is the fact that while assuming $|\vartheta_0| > 1$ in (1) makes the model noninvertible, it also complicates the derivation of the tests. This can be seen by noticing that in (1) $|\phi_0| < 1$ is assumed and the null hypothesis $\beta_0 = 0$ can equivalently be stated as $\phi_0 = \vartheta_0$, provided $|\vartheta_0| > 1$ is not assumed. In addition to a general noninvertible ARMA model we also consider tests within the corresponding all-pass model. Due to the aforementioned reasons of identification, our theoretical results assume a general non-Gaussian error distribution. In the empirical application of the paper, Student's t–distribution is employed.

We illustrate the usefulness of our tests by an application to quarterly U.S. stock returns, for which the noninvertible ARMA(1,1) model is found to provide an adequate description. In particular, and in contrast to the invertible ARMA(1,1) model, this model is able to capture the conditional heteroskedasticity present in the return series. The tests indicate that while there is little evidence in favor of autocorrelation, the returns are still dependent and, hence, predictable. The tests of Nankervis and Savin (2010) agree on the absence of autocorrelation, but they have little to say about predictability in general.

The remainder of the paper is organized as follows. Section 2 introduces our formulation of the noninvertible ARMA(1,1) model and discusses its properties. The test procedures are derived in Section 3 and studied by means of Monte Carlo simulation experiments in Section 4. An empirical application to testing the predictability of U.S. stock returns is presented in Section 5. Finally, Section 6 concludes. Some technical details are deferred to Appendices.

2 Noninvertible ARMA(1,1) and all-pass models

In this section, we discuss our formulation of the noninvertible ARMA(1,1) model, and its special case the all-pass model, in some detail. We define our noninvertible ARMA(1,1) model by the equation

$$y_t = \phi_0 y_{t-1} + \epsilon_{t-1} - \theta_0 \epsilon_t, \tag{2}$$

where ϕ_0 and θ_0 are parameters satisfying $|\phi_0| < 1$ and $|\theta_0| < 1$, and ϵ_t is an uncorrelated error term with zero mean and finite variance σ_0^2. (Throughout the paper, a subscript zero signifies a true (but unknown) parameter value.) Letting B denote the backward shift operator ($B^k \epsilon_t = \epsilon_{t-k}$, $k = 0, \pm 1, \pm 2, \ldots$) we can write equation (2) as

$$(1 - \phi_0 B)\, y_t = (1 - \theta_0 B^{-1}) B \epsilon_t. \tag{3}$$

When $\theta_0 \neq 0$, the connection between the specifications (1) and (2) is given by $\theta_0 = 1/\vartheta_0$ and $\epsilon_t = -\vartheta_0 \varepsilon_t$; when $\theta_0 = 0$, the moving average part in (2) reduces to the uncorrelated error term ϵ_{t-1}, and the same is achieved in (1) by setting $\vartheta_0 = 0$ and $\varepsilon_t = \epsilon_{t-1}$.

As already discussed in the Introduction, meaningful application of a noninvertible ARMA model requires a non-Gaussian data generation process. To illustrate this, using well-known results on linear filters (see, e.g., Brockwell and Davis (1991, Sec. 4.4)), the spectral density function of the noninvertible ARMA(1,1) process y_t in (2) can be seen to equal

$$\frac{\sigma_0^2}{2\pi}\frac{|(1-\theta_0 e^{i\omega})e^{-i\omega}|^2}{|1-\phi_0 e^{-i\omega}|^2}=\frac{\sigma_0^2}{2\pi}\frac{|1-\theta_0 e^{-i\omega}|^2}{|1-\phi_0 e^{-i\omega}|^2}. \tag{4}$$

The right-hand side of (4) coincides with the spectral density function of a conventional invertible ARMA(1,1) process with the same parameter values as in (2). This means that invertible and noninvertible ARMA processes cannot be distinguished by the spectral density function and, hence, by the autocovariance function. As the Gaussian likelihood function of an ARMA model is determined by the autocovariance function of the observed process, it also becomes understandable why estimation and statistical testing in noninvertible ARMA models assumes a non-Gaussian data generation process (in our case, a specific reason is that the information matrix based on a Gaussian likelihood becomes singular under the considered null hypotheses, as demonstrated in Appendix C). Known results on maximum likelihood (ML) and quasi ML estimation of noninvertible ARMA models also require that the error term is independent and identically distributed (IID). Unless otherwise stated we shall therefore assume that the error term ϵ_t in our model (2) is non-Gaussian and IID.

An interesting special case of the noninvertible ARMA(1,1) model (2) is obtained when $\phi_0 = \theta_0$. In this case, the process is called an all-pass process and defined as

$$y_t = \phi_0 y_{t-1} + \epsilon_{t-1} - \phi_0 \epsilon_t, \tag{5}$$

where $|\phi_0| < 1$ and ϵ_t is non-Gaussian and IID, as in (2). All-pass processes are uncorrelated (the spectral density in (4) reduces to $\sigma_0^2/2\pi$ when $\phi_0 = \theta_0$), but dependent (note that in (3) the operators $1-\phi_0 B$ and $1-\theta_0 B^{-1}$ do not cancel out even when $\phi_0 = \theta_0$). It may be worth noting, however, that, even though uncorrelated, all-pass processes are in general predictable. A formal justification for this fact is given in Appendix A, where it is demonstrated that the best (in mean square error sense) predictor of a (non-Gaussian) all-pass process is nonzero.

As discussed by Breidt, Davis, and Trindade (2001), all-pass models can, to some extent, allow for nonlinear behavior, especially of the kind typically modeled by ARCH type models.

The same holds true for noninvertible ARMA models in general. In Appendix A we derive the autocorrelation function of squared observations from a noninvertible ARMA(1,1) process, $\mathrm{Cor}(y_t^2, y_{t+k}^2)$, and from the expression therein it can be seen that for non-Gaussian errors one may expect to observe autocorrelation in squared observations generated by a noninvertible ARMA(1,1) process or an all-pass process. For certain values of the parameters ϕ_0 and θ_0, the autocorrelation in squared observations can be quite strong. This happens especially when the signs of ϕ_0 and θ_0 are different in which case the first order autocorrelation can even be close to unity. In such cases the series itself is also autocorrelated so that the situation is different from that in (pure) GARCH processes. In the all-pass process, the autocorrelation in squared observations is always rather mild, indicating that it may not be appropriate, say, for frequently observed financial time series which are typically (nearly) uncorrelated and exhibit strong conditional heteroskedasticity.

As a remark, we also note that if an invertible ARMA model is fitted to a time series generated by a (non-Gaussian) noninvertible ARMA process, the resulting squared residuals tend to be autocorrelated. To see this, write equation (3) as

$$(1-\phi_0 B)\, y_t = (1-\theta_0 B)\xi_t, \quad \xi_t = \frac{(1-\theta_0 B^{-1})}{(1-\theta_0 B)} B\epsilon_t,$$

where ξ_t is an all-pass process. Thus, when $\theta_0 \neq 0$, the errors ξ_t are uncorrelated but, as discussed in the previous paragraph, their squares are generally correlated.

We close this section by introducing the hypotheses we are interested in testing within the noninvertible ARMA(1,1) model (2) or the all-pass model (5). As there are more than one hypothesis, some kind of a sequential approach may be employed. One possibility is to start from the totally unrestricted model and test whether it reduces to an all-pass model. The hypothesis of interest is then

$$\mathsf{H}_{\text{AP}} : \phi_0 = \theta_0 \text{ in model (2).}$$

The alternative is $\phi_0 \neq \theta_0$. If this hypothesis is rejected the conclusion is that the process is autocorrelated. In case of nonrejection it is still possible that the common value of the parameters ϕ_0 and θ_0 is zero in which case the observed process is IID. For studying this, the relevant hypothesis is

$$\mathsf{H}_{\mathrm{IID}}^{(\mathrm{AP})} : \phi_0 = 0 \text{ in model (5)}$$

with the alternative being $\phi_0 \neq 0$. In this case, a rejection means that the process is uncorrelated but dependent and (nonlinearly) predictable, whereas a nonrejection supports the IID hypothesis.

If the IID process is a priori highly plausible, it may be a good idea to test for the IID hypothesis

$$\mathsf{H}_{\mathrm{IID}} : \phi_0 = \theta_0 = 0 \text{ in model (2)}$$

directly within the unrestricted noninvertible ARMA(1,1) model. If a rejection results, the relevant hypothesis to test next is the all-pass hypothesis H_{AP}. However, according to our simulation experiments in Section 4 the LR and Wald tests of the IID hypothesis H_{IID} may have relatively low power against close alternatives, suggesting that slight deviations from independence may not be detected. Therefore, if on a priori grounds the IID hypothesis is implausible, it may be advisable to start out with the all-pass hypothesis H_{AP} so as not to dismiss potential weak nonlinear dependence. For instance, in testing the predictability of asset returns, the general wisdom seems to be that the IID hypothesis is very unlikely to hold so that the all-pass hypothesis should be more interesting than the IID hypothesis.

As already discussed, all-pass processes exhibit ARCH-type dependence in the form of correlation in the squared observations. In this respect, our assumption of the error term ϵ_t in (2) (and (5)) being IID may not be so restrictive after all, as mild forms of conditional heteroskedasticity are permitted under the null hypothesis of most interest. As a comparison, the tests of Andrews and Ploberger (1996) do not allow for this kind of dependence but those of Nankervis and Savin (2010) do. The assumptions of Nankervis and Savin (2010) are in fact very general allowing for nonlinear dependences not covered by our noninvertible ARMA model. On the other hand, their tests do not facilitate discrimination between nonlinear dependence and independence although this may not be such a serious shortcoming, if the IID assumption can be precluded as incredible as seems to be the case in certain applications.

3 Test procedures

We now formulate (approximate) Wald and LR tests for the hypotheses introduced in the previous section. We start with a brief discussion of the assumptions needed.

We have already assumed the error term ϵ_t to be non-Gaussian and IID. Assume further that ϵ_t has a continuous distribution with a density function $f_{\sigma_0}(x;\lambda_0) = \sigma_0^{-1} f\left(\sigma_0^{-1}x;\lambda_0\right)$ that may also depend on the parameter vector λ_0 $(d \times 1)$ in addition to the scale parameter σ_0. For our theoretical developments, the function $f(x;\lambda)$ has to satisfy a number of regularity conditions similar to those used in related previous work on noninvertible and noncausal ARMA models (see Breidt et al. (1991), Lii and Rosenblatt (1996), Andrews, Davis, and Breidt (2006), and Lanne and Saikkonen (2011)). These conditions are technical in nature, and we relegate their precise formulation to Appendix C, where further discussion is also provided. Here we only note that the required conditions are satisfied by several conventional distributions such as the (rescaled) Student's t–distribution and weighted averages of Gaussian distributions. Their exact formulation is adopted from a recent paper by Meitz and Saikkonen (2011) where an asymptotic estimation theory for noninvertible ARMA models is extended to allow for ARCH type conditional heteroskedasticity. In the present context, the assumptions used in Meitz and Saikkonen (2011) are convenient because, unlike in other related previous work, the formulation of the employed noninvertible ARMA model is similar to (2). On the other hand, because some of these assumptions were originally designed to deal with conditional heteroskedasticity they may be unnecessarily strong in our context. However, as the main focus of our paper is to present a new approach of testing for autocorrelation and predictability, no attempt is made to find the weakest possible assumptions.

To set notation, collect the parameters of the model (2) in the vector $\delta = (\phi, \theta, \sigma, \lambda)$. The permissible parameter space is given by $|\phi| < 1$, $|\theta| < 1$, $\sigma > 0$, and $\lambda \in \Lambda$ where $\Lambda \subset \mathbb{R}^d$. Suppose that observations $y_0, \ldots, y_T$ are available. We estimate the parameters by maximizing the approximate log-likelihood function (divided by T)

$$\tilde{L}_T(\delta) = T^{-1} \sum_{t=1}^{T} \log f\left(\frac{\tilde{\epsilon}_{t-1}(\delta)}{\sigma}; \lambda\right) - \frac{1}{2}\log\sigma^2$$

the derivation of which is discussed in some detail in Appendix B. Here we only note that the quantities $\tilde{\epsilon}_{T-1}(\delta), \ldots, \tilde{\epsilon}_0(\delta)$ are solved by using the backward recursion

$$\tilde{\epsilon}_{t-1}(\delta) = y_t - \phi y_{t-1} + \theta\tilde{\epsilon}_t(\delta), \qquad t = T, \ldots, 1,$$

with end condition $\tilde{\epsilon}_T(\delta) = 0$. It is demonstrated in Appendix C that, under the regularity conditions stated therein, the approximate log-likelihood function $\tilde{L}_T(\delta)$ has a (local) maximizer $\tilde{\delta}$ which is consistent and asymptotically normally distributed. Specifically, we have

$$\sqrt{T}(\tilde{\delta} - \delta_0) \xrightarrow{d} N(0, \mathcal{I}(\delta_0)^{-1}) \ \text{ as } \ T \to \infty,$$

where the positive definite matrix $\mathcal{I}(\delta_0)$ is defined in Appendix C. Here it suffices to note that a consistent estimator of $\mathcal{I}(\delta_0)$ is obtained in the usual way from the Hessian of the log-likelihood function (but not from the outer product of the score matrix). Thus, denoting $\mathcal{J}_T(\delta) = -\partial^2\tilde{L}_T(\delta)/\partial\delta\partial\delta'$ we have $\mathcal{J}_T(\tilde{\delta}) \xrightarrow{p} \mathcal{I}(\delta_0)$.

The preceding discussion can readily be modified to concern estimation of parameters of the all-pass model (5). It suffices to redefine $\delta = (\phi, \sigma, \lambda)$ and compute $\tilde{\epsilon}_{t-1}(\delta)$ in the approximate log-likelihood function $\tilde{L}_T(\delta)$ with the restriction $\phi = \theta$. The asymptotic normality result of the obtained estimator then applies with a consistent estimator of the limiting covariance matrix defined in terms of the Hessian of the employed counterpart of $\tilde{L}_T(\delta)$.

Based on the preceding results we can derive Wald and LR tests for the hypotheses introduced in the preceding section. First consider the all-pass hypothesis H_{AP} and partition the parameter vector δ as $\delta = (\delta_1, \delta_2)$, where $\delta_1 = (\phi, \theta)$ and $\delta_2 = (\sigma, \lambda)$. Let $\tilde{\delta} = (\tilde{\delta}_1, \tilde{\delta}_2)$ and $\mathcal{J}_T(\tilde{\delta}) = [\mathcal{J}_{\delta_i\delta_j,T}(\tilde{\delta})]$ $(i, j = 1, 2)$ be the corresponding partitions of the estimator $\tilde{\delta}$ and the matrix $\mathcal{J}_T(\tilde{\delta})$. To simplify notation, we denote $\tilde{\mathcal{J}}_{ij} = \mathcal{J}_{\delta_i\delta_j,T}(\tilde{\delta})$ $(i, j = 1, 2)$. Then, defining the vector $a = (1, -1)$ we can write the Wald test statistic for H_{AP} as

$$\mathsf{W}_{\text{AP}} = T\tilde{\delta}_1' a\big[a'(\tilde{\mathcal{J}}_{11} - \tilde{\mathcal{J}}_{12}\tilde{\mathcal{J}}_{22}^{-1}\tilde{\mathcal{J}}_{21})^{-1}a\big]^{-1}a'\tilde{\delta}_1 \xrightarrow{d} \chi_1^2 \quad \text{under } \mathsf{H}_{\text{AP}}.$$

Of course, one can alternatively obtain a Wald test with asymptotic standard normal distribution. For the corresponding likelihood ratio test, let $\tilde{\delta}_{\text{AP}} = (\tilde{\phi}_{\text{AP}}, \tilde{\phi}_{\text{AP}}, \tilde{\sigma}_{\text{AP}}, \tilde{\lambda}_{\text{AP}})$ signify the ML estimator of δ_0 constrained by H_{AP}. Then, the LR test statistic for testing H_{AP} reads as

$$\mathsf{LR}_{\text{AP}} = 2T\big[\tilde{L}_T(\tilde{\delta}) - \tilde{L}_T(\tilde{\delta}_{\text{AP}})\big] \xrightarrow{d} \chi_1^2 \quad \text{under } \mathsf{H}_{\text{AP}}.$$

Obtaining a Wald test for the IID hypothesis $\mathsf{H}_{\text{IID}}^{(\text{AP})}$ with a standard normal limiting distribution is simple. The test statistic is just the estimator $\tilde{\phi}_{\text{AP}}$ divided by its approximate standard error obtained from the square root of the first diagonal element of the inverse of the relevant Hessian discussed above. For the corresponding LR test one needs to estimate the nuisance parameters σ_0 and λ_0, that is, maximize the likelihood function defined by assuming that the observed series y_t, $t = 0, \ldots, T$, is IID with marginal distribution characterized by the density $f_{\sigma_0}(x; \lambda_0)$. Denoting the resulting restricted estimator of the parameter vector δ_0 by $\tilde{\delta}_{\text{IID}} = (0, 0, \tilde{\sigma}_{\text{IID}}, \tilde{\lambda}_{\text{IID}})$ we get the LR test statistic

$$\mathsf{LR}_{\text{IID}}^{(\text{AP})} = 2T\big[\tilde{L}_T(\tilde{\delta}_{\text{AP}}) - \tilde{L}_T(\tilde{\delta}_{\text{IID}})\big] \xrightarrow{d} \chi_1^2 \quad \text{under } \mathsf{H}_{\text{IID}}^{(\text{AP})}.$$

Along the same lines one can also construct a test for the IID hypothesis H_{IID}. The Wald test statistic can be formed by replacing the vector a in test statistic W_{AP} by a 2×2 identity matrix whereas the corresponding LR test statistic can be formed by replacing the estimator $\tilde{\delta}_{\text{AP}}$ in test statistic LR_{AP} by the estimator $\tilde{\delta}_{\text{IID}}$. Both of these test statistics have an asymptotic χ_2^2 distribution under the null hypothesis.

It may be noted that the preceding tests obtained in the considered noninvertible ARMA(1,1) model are "standard" leading to an asymptotic chi-squared or standard normal limiting distribution. This is not the case when serial correlation is tested in the conventional invertible

ARMA(1,1) model. Then a non-standard testing problem with complicated limiting distributions results because a nuisance parameter is present in the model only under the alternative hypothesis; see Andrews and Ploberger (1996) and Nankervis and Savin (2010).

4 Simulation study

In this section, we explore the finite-sample properties of the proposed tests by means of Monte Carlo simulation experiments. In addition to reporting the results of a number of size and power simulations of the Wald and LR tests, we also simulated the tests of Nankervis and Savin (2010) for comparison. The new tests are shown to be superior in testing the null hypothesis of no autocorrelation against the noninvertible ARMA(1,1) process. Throughout the results are based on 10,000 replications, and two sample sizes, 200 and 500, are considered.

Table 1 presents the rejection rates of the 5% level Wald and LR tests, when the data are generated from the noninvertible ARMA(1,1) model with the error term having Student's t–distribution with 5 degrees of freedom (this error distribution is also used in other simulation experiments discussed in this section). As far as the all-pass hypothesis H_{AP} in the noninvertible ARMA(1,1) model is concerned, the size of the Wald test seems to be closer to the nominal size than that of the LR test. In order to study the power of the tests, we consider alternative data generation processes with ϕ_0 fixed at 0.8 and and θ_0 taking values between 0.85 and 0.65. Comparable parameter values are likely to be encountered in typical empirical applications of these tests. The rejection rates of both tests increase as a function of the distance of θ_0 from 0.8, with an equal distance resulting in greater empirical power when θ_0 exceeds 0.8. In general, both tests have good power compared to the tests of Nankervis and Savin (2010) (to be discussed in more detail below), and the differences between them are minor.

The rejection rates of the Wald and LR tests for the IID hypothesis $\mathsf{H}_{\mathrm{IID}}^{(\mathrm{AP})}$ in the all-pass model are reported for different values of the parameters in the middle panel of Table 1. With 200 observations, both tests slightly overreject, but with the greater sample size, the rejection rates lie close to the nominal size. The empirical power is reasonable already for relatively small deviations from the null hypothesis, and it increases steadily with the parameter values. The differences between the Wald and LR tests are minor.

The lower panel of Table 1 presents the rejection rates of the Wald and LR tests for the IID hypothesis H_{IID} in the noninvertible ARMA(1,1) model. In accordance with the tests of the IID hypothesis $\mathsf{H}_{\text{IID}}^{(\text{AP})}$ in the all-pass model, both tests tend to overreject with only 200 observations, with the overrejection problem relieved as the sample size increases. With 200 observations, the size of the Wald test is more seriously distorted, but at the greater sample size, the difference between the tests is negligible. The power properties of the tests developed for the hypotheses H_{IID} and $H_{\text{IID}}^{(\text{AP})}$ are similar but the latter seem to be superior, especially at alternatives close to the null hypothesis. This suggests that it might be preferable to test for independence in the all-pass model instead of the unrestricted noninvertible ARMA(1,1) model, provided the all-pass restriction is not rejected.

For comparison, in Table 2, we report the rejection rates of the Exp-LM$_\infty$ test of Nankervis and Savin (2010), which is the one of their tests that they recommend for nonseasonal applications in economics and finance. As the null hypothesis of these tests is that of no autocorrelation, they should be compared to our tests of the all-pass hypothesis H_{AP}. The finite-sample behavior of their sup LM and Exp-LM$_0$ tests is similar in our setup (to save space the results are not reported, but they are available upon request). Application of these tests requires choice of a weight matrix, i.e., an estimator of the asymptotic covariance matrix of the sample autocorrelations. We consider three different weight matrices: the identity matrix and the VARHAC estimator with the lag length selected by the Akaike and Bayesian information criteria (AIC and BIC, respectively). The program used for VARHAC was downloaded from Wouter den Haan's web page (http://www.wouterdenhaan.com/varhac.html). In accordance with the simulation results of Nankervis and Savin (2010), we find the test based on the Akaike information criterion clearly inferior in small samples. In particular, in this case the test tends to overreject heavily, while there is only slight overrejection with the other choices of the weight matrix, especially with as many as 500 observations. As far as power is concerned, qualitatively the findings are similar to those for the tests of the all-pass hypothesis H_{AP} in the upper panel of Table 1, but in each case the Exp-LM$_\infty$ test is beaten by our Wald and LR tests by a considerable margin. For instance, when $\theta_0 = 0.9$ and $T = 200$, the Wald and Exp-LM$_\infty$ tests reject approximately 75% and 30% of the time, respectively, with both tests suffering from small size distortions of similar magnitude. This indicates that for testing for serial correlation in the noninvertible ARMA model, the new tests are clearly superior to the those of Nankervis and Savin (2010), especially when autocorrelation is relatively weak.

As discussed in Section 2, when the values of ϕ_0 and θ_0 are very different, the noninvertible ARMA(1,1) process can exhibit even strong ARCH-type dependence. Hence, it is plausible that our tests have some power against serially uncorrelated GARCH processes, i.e., they may reject the all-pass hypothesis H_{AP} although the series is not autocorrelated but strongly conditionally heteroskedastic. It is worth noting, however, that there are no statistical reasons why our tests should maintain their size in cases like these which are outside the model class assumed to derive the tests.

To check how sensitive our tests are to ARCH-type dependence we ran simulations with the GARCH(1,1) model as the data generation process. We used the same parameter values as Nankervis and Savin (2010) did in their simulations, but instead of Gaussian errors, we generated errors from Student's t–distribution with 5 degrees of freedom. The parameter values of Nankervis and Savin (2010) imply rather strong ARCH effects, and, not surprisingly, our tests have nonnegligible, but relatively low power against this data generation process. The rejection rates of the Wald test of the all-pass hypothesis H_{AP} equal 0.13 and 0.15 with 200 and 500 observations, respectively. With the LR test the corresponding figures are higher, 0.19 and 0.25, respectively, indicating a somewhat stronger sensitivity to the considered GARCH process. Further experiments with a number of other parameter values resulted in similar outcomes. Thus, although the rejection rates for pure GARCH models are not very high, test results must be interpreted with care to avoid mixing autocorrelation up with conditional heteroskedasticity. Therefore, the tests cannot without reservation be recommended for frequently sampled financial time series exhibiting strong conditional heteroskedasticity.

5 Empirical application

As pointed out in the Introduction, testing the predictability of asset returns continues to be an active area of research in finance. In the early literature, the absence of predictability was considered an indication of market efficiency, i.e., it was argued that because rational investors use information efficiently, returns should be unpredictable. However, the later developments in dynamic asset pricing theory have demonstrated that this is the case only under very special conditions, including the assumption of risk neutral investors. Hence, the results of predictability tests, in general, yield no direct conclusions concerning market efficiency, but they are interesting from the viewpoint of studying asset pricing models and investment strategies.

Campbell, Lo, and MacKinlay (1997, Ch. 2) discuss three different types of dependencies in asset returns that have been explored in the empirical finance literature. Under the strictest assumption considered, returns are independently and identically distributed, while the majority of the empirical literature concentrates on testing the presence of autocorrelation. Our two-stage testing procedure outlined in Section 3 provides a unified framework that encompasses both of these rather separate literatures. In addition, it facilitates testing nonlinear predictability whose presence is in line with modern asset pricing theory (cf. Singleton (2006, Ch. 9)). The third assumption made in part of the previous literature, independence with heterogeneity, is largely unexplored due to lack of suitable methods.

In our testing approach, rejection of the all-pass hypothesis H_{AP} indicates the presence of autocorrelation. On the other hand, if it is not rejected, the conditional IID hypothesis $H_{\text{IID}}^{(\text{AP})}$ can next be tested, and its rejection indicates that the returns follow an all-pass process, and are thus predictable. Conversely if $H_{\text{IID}}^{(\text{AP})}$ is not rejected, the returns are deemed unpredictable. The underlying assumption of the procedure is that the returns are generated by the noninvertible

ARMA(1,1) process. Previous analyses have often been based on its invertible counterpart implied by the price-trend model of Taylor (1982) and the mean-reversion model of Poterba and Summers (1988). Both models produce the same autocorrelation function, but the noninvertible ARMA model has a number of benefits discussed above. Of course, it is necessary to confirm the fit of this model by diagnostic checks before proceeding with the tests.

In our empirical analysis, we test the predictability of quarterly U.S. returns on three value-weighted size-ordered stock portfolios and the market portfolio. The data were obtained from Kenneth French's web page (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html), and the portfolios include all NYSE, AMEX, and NASDAQ stocks with data for June of each year. The monthly simple returns from 1947:1 to 2007:12 were converted to continuously compounded quarterly return series. We consider only quarterly returns because more frequently sampled return series presumably exhibit too strong autoregressive conditional heteroskedasticity coupled with too weak autocorrelation for the noninvertible ARMA process to capture.

As a preliminary analysis, we estimate the invertible Gaussian ARMA(1,1) model for each (demeaned) series. These models are adequate in the sense of capturing all autocorrelation, but the squared residuals are strongly autocorrelated, with the fourth-order McLeod-Li test rejecting at the 1% level in each case. Furthermore, as expected, the residuals exhibit considerable excess kurtosis, which shows up as rejections in the Jarque-Bera test at any reasonable significance level. Hence, we proceed with noninvertible ARMA(1,1) models with Student's t–distributed errors.

The estimation results are presented in the upper panel of Table 3. For all portfolios, the estimated degrees-of-freedom parameter is small, reconfirming the need for a leptokurtic error distribution, and also the Q-Q plots of the residuals (not shown) indicate good fit. The estimates of ϕ_0 and θ_0 are always large and lie rather close to each other. All parameters are quite accurately estimated. Neither the residuals nor their squares are autocorrelated, indicating that the noninvertible ARMA model successfully captures conditional heteroskedasticity. For comparison, we also estimated the invertible ARMA(1,1) model with Student's t–distributed errors for each series. Although this model produces serially uncorrelated residuals as well, some conditional heteroskedasticity seems to be remaining. For the two portfolios with the smallest firms, the fourth-order McLeod-Li test rejects even at the 5% level. Thus, in terms of fit, the noninvertible models seem superior.

The result of autocorrelation and independence tests are reported in the lower panel of Table 3. We could first test for independence, i.e., hypothesis H_{IID} in the noninvertible ARMA(1,1) model, but because stock returns are highly unlikely to be IID, we follow the two-stage testing procedure outlined in Section 3. As far as the all-pass hypothesis H_{AP} is concerned, the tests lend little support to autocorrelation in the returns. It is only for the returns on the stocks of the smallest firms that the Wald test rejects at the 10% level. Although virtually no evidence in favor of autocorrelation is found, the returns may still exhibit nonlinear predictability, and we proceed with the tests of the IID hypothesis $\mathsf{H}_{\text{IID}}^{(\text{AP})}$ in the all-pass model. The Wald test rejects with very small p-values in each case, while the results of the LR test are more varied. For the market return, $\mathsf{H}_{\text{IID}}^{(\text{AP})}$ is rejected at the 1% level, and for the two portfolios consisting of the stocks of the largest firms at the 10% level. However, for the smallest firms, the LR test does not reject at any reasonable level of significance, suggesting independence. With the potential exception of the smallest firms, we can thus conclude that the returns are neither independently and identically distributed nor autocorrelated, but they are still predictable in the sense of being generated by the all-pass model.

The discrepancies in the results of the Wald and LR tests for the smallest firms most likely follow from the fact that the likelihood surface is very flat with two local maxima, the global one given in Table 3 and another one in the vicinity of the point $\phi_0 = \theta_0 = 0$. Under the constraint $\phi_0 = \theta_0$, the global optimum (–0.072) lies close to the latter, which explains the nonrejection of the IID hypothesis in the LR test even though the ML estimates of ϕ_0 and θ_0 in Table 3 are quite different from zero. Given these considerations, we are inclined to conclude that also the returns on the stocks of the smallest firms are not IID but predictable albeit not autocorrelated.

The result of the Exp-LM$_\infty$ test reported on the bottom row of Table 3 indicate no rejections at reasonable significance levels. The other two tests of Nankervis and Savin (2010) lead to similar conclusions. As far as autocorrelation is concerned, this test yields the same conclusion as our tests of the hypothesis H_{AP}. However, our two-stage testing procedure goes beyond this in finding (nonlinear) predictability, while the Nankervis-Savin tests are only designed for testing autocorrelation.

6 Conclusion

The test procedures for autocorrelation and predictability developed in this paper within the noninvertible ARMA(1,1) model add to the available tests previously obtained within the conventional invertible ARMA(1,1) model. A convenient feature of the procedures, not shared by their previous counterparts, is that in addition to testing for autocorrelation, they also facilitate testing for nonlinear predictability. The noninvertible ARMA model also differs from its invertible counterpart in that, to some extent, it can allow for conditional heteroskedasticity in the data. These features require a non-Gaussian data generation process which, however, need not be a serious limitation in that Gaussianity is quite often found inappropriate, for instance, in modeling economic and financial time series. This also turned out to be the case in our empirical application to testing the predictability of U.S. stock returns.

Although the noninvertible ARMA(1,1) model is theoretically well motivated and probably empirically adequate in typical financial applications, our test procedures can be extended to higher-order ARMA models in a straightforward manner. In contrast, the corresponding testing problem in the conventional invertible ARMA model is nonstandard, making such extensions tedious. In addition to extensions to higher-order models, our results open a number of other avenues for future research. First, instead of ML estimation, the least absolute deviation estimation developed in Breidt, Davis, and Trindade (2001) and Wu and Davis (2010) could be employed. Although this approach tends to result in less powerful tests, it has the benefit of not having to specify the error distribution. This feature may be useful when modeling highly leptokurtic financial data. Second, it may also be of interest to consider similar tests within a noninvertible ARMA model involving an ARCH or GARCH component. Some work in this direction was recently done by Meitz and Saikkonen (2011) who studied ML estimation by allowing for conventional ARCH errors in the noninvertible ARMA model. Such extensions are likely to be needed in applications to frequently observed financial data. Third, while we have considered methods of assessing predictability by means of tests on the parameters of the model, quantifying the accuracy of out-of-sample forecasts might also be of interest. This calls for the development of a forecasting method for the noninvertible ARMA model that lies outside the scope of this paper. Finally, in addition to quarterly U.S. stock returns studied in this paper, also returns from other markets and more frequently sampled returns, including those from the foreign exchange market, would be interesting.

Appendix A: Additional technical details

Predictability of an all-pass process. Consider an all-pass process $y_t = \phi_0 y_{t-1} + \epsilon_{t-1} - \phi_0 \epsilon_t$ with $|\phi_0| < 1$, $\phi_0 \neq 0$, and ϵ_t non-Gaussian and IID. Denote $u_t = y_t - \phi_0 y_{t-1} = \epsilon_{t-1} - \phi_0 \epsilon_t$. Because u_t is a noninvertible MA process, according to Rosenblatt (2000, Corollary 5.4.3) and subject to mild moment conditions on ϵ_t (see op. cit.), the best one-step predictor of u_t, $E[u_t \mid u_{t-s},\ s \geq 1]$, must be nonlinear. On the other hand, as the AR-polynomial $1 - \phi_0 B$ is causal, $y_t = \sum_{j=0}^{\infty} \phi_0^j u_{t-j}$, and the σ–algebras $\sigma\left(u_{t-s},\ s \geq 1\right)$ and $\sigma\left(y_{t-s},\ s \geq 1\right)$ coincide, so that

$$E[u_t \mid u_{t-s},\ s \geq 1] = E[u_t \mid y_{t-s},\ s \geq 1] = E[y_t - \phi_0 y_{t-1} \mid y_{t-s},\ s \geq 1] = E[y_t \mid y_{t-s},\ s \geq 1] - \phi_0 y_{t-1}.$$

If y_t is not predictable, $E[y_t \mid y_{t-s},\ s \geq 1] = 0$, in which case $E[u_t \mid u_{t-s},\ s \geq 1] = -\phi_0 y_{t-1} = -\phi_0 \sum_{j=0}^{\infty} \phi_0^j u_{t-1-j}$, an expression linear in u_{t-s}, $s \geq 1$, a contradiction. Therefore y_t must be predictable, with the best predictor being nonlinear.

Autocorrelation function of squared observations from a noninvertible ARMA(1,1) process. First conclude from (2) that y_t has the linear representation

$$y_t = \sum_{j=-1}^{\infty} \psi_{0,j} \epsilon_{t-1-j}, \tag{6}$$

where $\psi_{0,j}$ is the coefficient of z^j in the Laurent series expansion of $(1 - \phi_0 z)^{-1}(1 - \theta_0 z^{-1}) \overset{def}{=} \psi_0\left(z\right)$. Now assume that ϵ_t has finite fourth moments and consider the autocorrelation function of y_t^2. As in Brockwell and Davis (1991, the proof of Proposition 7.3.1) one obtains

$$\operatorname{Cor}\left(y_t^2, y_{t+k}^2\right) = \frac{(\kappa_0 - 3)\sum_{j=-1}^{\infty} \psi_{0,j}^2 \psi_{0,j+k}^2 + 2\left(\sum_{j=-1}^{\infty} \psi_{0,j}\psi_{0,j+k}\right)^2}{(\kappa_0 - 3)\sum_{j=-1}^{\infty} \psi_{0,j}^4 + 2\left(\sum_{j=-1}^{\infty} \psi_{0,j}^2\right)^2}, \quad k \geq 0, \tag{7}$$

where $\kappa_0 = E\left(\epsilon_t^4\right)/\sigma_0^4$. This shows that the squared process y_t^2 is autocorrelated when $\kappa_0 \neq 3$ or, equivalently, when the (excess) kurtosis of ϵ_t is nonzero. Thus, for non-Gaussian errors one may expect to observe autocorrelation in squared observations obtained from a noninvertible ARMA(1,1) processes. For all-pass processes the right hand side of (7) simplifies because, due to lack of autocorrelation, the second term in the numerator vanishes. However, the first term is generally nonzero, illustrating the aforementioned dependence of all-pass processes.

An explicit expression for the right hand side of (7) as a function of the parameters ϕ_0, θ_0, and κ_0 is derived next. First note that the coefficients $\psi_{0,j}$ in the linear representation (6) are given by $\psi_{0,-1} = -\theta_0$ and $\psi_{0,j} = (1 - \phi_0\theta_0)\,\phi_0^j$, $j = 0, 1, \ldots$. With straightforward computation one obtains

$$\begin{aligned}\sum_{j=-1}^{\infty} \psi_{0,j}^2 &= \theta_0^2 + \frac{(1-\phi_0\theta_0)^2}{1-\phi_0^2} \\ \sum_{j=-1}^{\infty} \psi_{0,j}\psi_{0,j+k} &= (1-\phi_0\theta_0)\,\phi_0^{k-1}\left[\frac{(1-\phi_0\theta_0)\,\phi_0}{1-\phi_0^2} - \theta_0\right], \quad k > 0.\end{aligned}$$

Note that the expression in the brackets vanishes when the all-pass restriction $\phi_0 = \theta_0$ holds. Further computations give

$$\begin{aligned}\sum_{j=-1}^{\infty} \psi_{0,j}^4 &= \theta_0^4 + \frac{(1-\phi_0\theta_0)^4}{1-\phi_0^4} \\ \sum_{j=-1}^{\infty} \psi_{0,j}^2\psi_{0,j+k}^2 &= (1-\phi_0\theta_0)^2\,\phi_0^{2k-2}\left[\frac{(1-\phi_0\theta_0)^2\,\phi_0^2}{1-\phi_0^4} + \theta_0^2\right], \quad k > 0.\end{aligned}$$

Substituting the preceding expressions on the right hand side of (7) and simplifying yields

$$\mathrm{Cor}\left(y_t^2, y_{t+k}^2\right) = (1-\phi_0\theta_0)^2\,\phi_0^{2k-2}\frac{(\kappa_0-3)\left[\frac{(1-\phi_0\theta_0)^2\phi_0^2}{1-\phi_0^4} + \theta_0^2\right] + 2\left[\frac{(1-\phi_0\theta_0)\phi_0}{1-\phi_0^2} - \theta_0\right]^2}{(\kappa_0-3)\left[\theta_0^4 + \frac{(1-\phi_0\theta_0)^4}{1-\phi_0^4}\right] + 2\left[\theta_0^2 + \frac{(1-\phi_0\theta_0)^2}{1-\phi_0^2}\right]^2}$$

for $k > 0$.

Appendix B: The approximate likelihood function of the noninvertible ARMA(1,1) model

Following Lii and Rosenblatt (1996), Andrews, Davis, and Breidt (2006), and Meitz and Saikkonen (2011) we estimate the parameters of the model by an approximate ML procedure. As in these papers it can be shown that, conditional on the initial value y_0, the log-likelihood of the parameter vector $\delta = (\phi, \theta, \sigma, \lambda)$ based on the observed data $(y_1, \ldots, y_T)$ (divided by T) can be approximated by

$$L_T(\delta) = T^{-1} \sum_{t=1}^{T} \log f\left(\frac{\epsilon_{t-1}(\delta)}{\sigma}; \lambda\right) - \frac{1}{2} \log \sigma^2,$$

where

$$\epsilon_{t-1}(\delta) = \theta(B^{-1})^{-1} \phi\left(B\right) y_t = \sum_{j=-1}^{\infty} \varpi_j y_{t+j}$$

with ϖ_j the coefficient of z^j in the Laurent series expansion of $\theta(z^{-1})^{-1}\phi\left(z\right) \overset{def}{=} \varpi\left(z\right)$. However, as computing $\epsilon_t(\delta)$ for $t = 1, \ldots, T$ is not feasible in terms of the available data, a further approximation is needed. To obtain a likelihood feasible in practice we need an approximation for $\epsilon_{t-1}(\delta)$, $t = 1, \ldots, T$, expressible in terms of the observations $y_0, y_1, \ldots, y_T$ and the parameters. To this end, set $\tilde{\epsilon}_T(\delta) = 0$ and recursively solve for $\tilde{\epsilon}_{T-1}(\delta), \ldots, \tilde{\epsilon}_0(\delta)$ by using the backward recursion

$$\tilde{\epsilon}_{t-1}\left(\delta\right) = y_t - \phi_1 y_{t-1} + \theta_1 \tilde{\epsilon}_t(\delta), \qquad t = T, \ldots, 1.$$

As in the aforementioned papers, the resulting approximate log-likelihood then takes the form

$$\tilde{L}_T(\delta) = T^{-1} \sum_{t=1}^{T} \log f\left(\frac{\tilde{\epsilon}_{t-1}(\delta)}{\sigma}; \lambda\right) - \frac{1}{2} \log \sigma^2.$$

In practice, estimation is carried out by maximizing $\tilde{L}_T(\delta)$ over the permissible parameter space (the infeasible counterpart $L_T(\delta)$ can be used in theoretical derivations).

Appendix C: Asymptotic properties of the approximate ML estimator

In this appendix, we discuss the asymptotic properties of the the approximate ML estimator introduced in Appendix B, thereby justifying the test procedures presented in Secion 3. We use results in Meitz and Saikkonen (2011). The noninvertible ARMA model considered in that paper differs in one minor respect of the one used in this paper. The difference only concerns the time index in the error term ϵ_t. The formulation employed in Meitz and Saikkonen (2011) is obtained from that in (1) by replacing ϵ_t by ϵ_{t+1}. From the viewpoint of parameter estimation and deriving asymptotic properties of the estimators this difference is of no importance. However, in order to facilitate comparison with the arguments in Meitz and Saikkonen (2011) we explicitly present the formulation of their model which is

$$y_t = \phi_0 y_{t-1} + \epsilon_t - \theta_0 \epsilon_{t+1},$$

where the notation is exactly as in (2) except for the fact that ϵ_t has been replaced by ϵ_{t+1}. As in Meitz and Saikkonen (2011), we first introduce the infeasible log-likelihood function

$$L_T(\delta) = T^{-1} \sum_{t=1}^{T} \log f\left(\frac{\epsilon_t(\delta)}{\sigma}; \lambda\right) - \frac{1}{2} \log \sigma^2,$$

where

$$\epsilon_t(\delta) = \theta(B^{-1})^{-1} \phi(B)\, y_t = \sum_{j=-1}^{\infty} \varpi_j y_{t+j},$$

with ϖ_j the coefficient of z^j in the Laurent series expansion of $\theta(z^{-1})^{-1}\phi(z) \overset{def}{=} \varpi(z)$. The feasible log-likelihood function $\tilde{L}_T(\delta)$ is obtained from this by replacing $\epsilon_t(\delta)$ by $\tilde{\epsilon}_t(\delta)$, $t = 1, \ldots, T$, by setting $\tilde{\epsilon}_{T+1}(\delta) = 0$ and recursively solving for $\tilde{\epsilon}_T(\delta), \ldots, \tilde{\epsilon}_1(\delta)$ with the backward recursion

$$\tilde{\epsilon}_t(\delta) = y_t - \phi y_{t-1} + \theta \tilde{\epsilon}_{t+1}(\delta), \qquad t = T, \ldots, 1.$$

To present the assumptions required for the asymptotic distribution of the approximate ML estimator we need some notation. As we here use standardized innovations, we write ϵ_t as $\epsilon_t = \sigma_0 \eta_t$ and we also use a subscript to signify a partial derivative indicated by the subscript, for instance $f_x(x; \lambda) = \frac{\partial}{\partial x} f(x; \lambda)$, $f_\lambda(x; \lambda) = \frac{\partial}{\partial \lambda} f(x; \lambda)$, and $f_{xx}(x; \lambda) = \frac{\partial^2}{\partial x^2} f(x; \lambda)$. For brevity, we set $e_{x,t} = \frac{f_x(\eta_t;\lambda_0)}{f(\eta_t;\lambda_0)}$ and $e_{\lambda,t} = \frac{f_\lambda(\eta_t;\lambda_0)}{f(\eta_t;\lambda_0)}$, and let $|\cdot|$ signify the Euclidean norm. The following assumptions are sufficient to obtain the desired results.

Assumption C.1.

(i) *The innovation process* η_t *is a sequence of IID random variables with* $E[\eta_t] = 0$, $E[\eta_t^2] = 1$, *and* $E[\eta_t^4] < \infty$. *The distribution of* η_t *is non-Gaussian, and has a (Lebesgue) density* $f(x; \lambda_0)$ *which (possibly) depends on a parameter vector* λ_0 *taking values in an open subset of* $\mathbb{R}^d$.

(ii) *For all* $x \in \mathbb{R}$ *and* λ *in some neighborhood of* λ_0, $f(x;\lambda) > 0$ *and* $f(x;\lambda)$ *is twice continuously differentiable with respect to* $(x;\lambda)$.

(iii) *For all* λ *in some neighborhood of* λ_0, $\int x f(x;\lambda)\,dx = 0$ *and* $\int x^2 f(x;\lambda)\,dx = 1$.

(iv) *The matrix* $E[e_{\lambda,t}e'_{\lambda,t}]$ *is positive definite.*

(v) $\int f_{xx}(x;\lambda_0)\,dx = 0$ *and* $\int x^2 f_{xx}(x;\lambda_0)\,dx = 2$.

(vi) *For all* $x \in \mathbb{R}$, *all* λ *in some neighborhood of* λ_0, *and every* λ_i, $i = 1, \ldots, d$, *the functions*

$$x^4 \frac{f_x^4(x;\lambda)}{f^4(x;\lambda)}, \quad \frac{f_{\lambda_i}^4(x;\lambda)}{f^4(x;\lambda)}, \quad x^4 \frac{f_{xx}^2(x;\lambda)}{f^2(x;\lambda)}, \quad \frac{f_{\lambda_i x}^2(x;\lambda)}{f^2(x;\lambda)}, \quad \text{and} \quad \left|\frac{f_{\lambda\lambda}(x;\lambda)}{f(x;\lambda)}\right|$$

are dominated by $d_1(1 + |x|^{d_2})$ *with* $d_1, d_2 \geq 0$ *and* $\int |x|^{d_2} f(x;\lambda_0)\,dx < \infty$.

(vii) *For all* $x \in \mathbb{R}$ *and* λ *in some neighborhood of* λ_0, *the functions* $|x^2 f_\lambda(x;\lambda)|$ *and* $|f_{\lambda\lambda}(x;\lambda)|$ *are dominated by a function* $\overline{f}(x)$ *such that* $\int \overline{f}(x)\,dx < \infty$.

(viii) *For all* $x \in \mathbb{R}$, $\triangle x \in \mathbb{R}$, *and* λ *in some neighborhood of* λ_0, *and for some* $C < \infty$ *and* $d_1, d_2 > 0$,

$$|v(x + \triangle x;\lambda) - v(x;\lambda)| \leq C\big((1 + |x|^{d_1})\,|\triangle x| + |\triangle x|^{d_2}\big)$$

for the following choices of the function $v(x;\lambda)$:

(a) (i) $v(x;\lambda) = \frac{f_x(x;\lambda)}{f(x;\lambda)}$, *(ii)* $v(x;\lambda) = \frac{f_\lambda(x;\lambda)}{f(x;\lambda)}$.

(b) (i) $v(x;\lambda) = \frac{f_{xx}(x;\lambda)}{f(x;\lambda)}$, *(ii)* $v(x;\lambda) = \frac{f_{\lambda x}(x;\lambda)}{f(x;\lambda)}$, *(iii)* $v(x;\lambda) = \frac{f_{\lambda\lambda}(x;\lambda)}{f(x;\lambda)}$.

Assumption C.1 consists of conditions modified from Assumptions 1–7 of Meitz and Saikkonen (2011). These authors consider maximum likelihood estimation of a noninvertible ARMA(P,Q) model in which the error terms ϵ_t are conditionally heteroskedastic and follow a standard ARCH(R)–model. The noninvertible ARMA(1,1) model considered here is obtained as a special case by setting $P = Q = 1$ and assuming the ϵ_t to be IID with constant variance σ_0^2.

Besides minor differences in presentation, there are two essential differences between the conditions above and Assumptions 1–7 of Meitz and Saikkonen (2011). First, we assume the errors to be non-Gaussian. As was discussed in Section 2, in the present context it is necessary to rule out Gaussian innovations. In Meitz and Saikkonen (2011) the situation is different as there Gaussian innovations can be allowed due to the assumed ARCH-structure. Second, in Meitz and Saikkonen (2011) the innovations were assumed to have a symmetric distribution. The only reason for this was to simplify the otherwise complex derivations, and in the present context this is not necessary.

As mentioned in Section 3, some of the employed assumptions in Meitz and Saikkonen (2011) were imposed to deal with the assumed ARCH-structure and may, therefore, be unnecessarily strong. For instance, it seems possible that the assumption of a finite fourth moment could be replaced by a milder alternative. On the other hand, this moment condition is already marginally weaker than what is assumed by Andrews and Ploberger (1996) and Nankervis and Savin (2010) who, however, allowed for a considerably more general data generation process than we do. Regarding other conditions in Assumption C.1, most of them have analogs in Lii and Rosenblatt (1996) and Andrews, Davis, and Breidt (2006).

We can now state a result summarizing the asymptotic properties of the (feasible) ML estimator $\tilde{\delta}_T$.

Theorem C.1. *If Assumption C.1 holds, there exists a sequence of solutions $\tilde{\delta}_T$ to the (feasible) likelihood equations $\partial \tilde{L}_T(\delta)/\partial\delta = 0$ such that $T^{1/2}(\tilde{\delta}_T - \delta_0) \xrightarrow{d} N(0, \mathcal{I}(\delta_0)^{-1})$ as $T \to \infty$, where*

$\mathcal{I}(\delta_0)$ *takes the form*

$$\mathcal{I}(\delta_0) = \begin{bmatrix} E[e_{x,t}^2](1-\phi_0^2)^{-1} & -(1-\phi_0\theta_0)^{-1} & 0 & 0 \\ -(1-\phi_0\theta_0)^{-1} & E[e_{x,t}^2](1-\theta_0^2)^{-1} & 0 & 0 \\ 0 & 0 & \frac{1}{4\sigma_0^4}E[(e_{x,t}\eta_t+1)^2] & -\frac{1}{2\sigma_0^2}E[e_{x,t}\eta_t e_{\lambda,t}'] \\ 0 & 0 & -\frac{1}{2\sigma_0^2}E[e_{x,t}\eta_t e_{\lambda,t}] & E[e_{\lambda,t}e_{\lambda,t}'] \end{bmatrix}.$$

Moreover, a consistent estimator for the limiting covariance matrix is given by $-(\partial^2\tilde{L}_T(\tilde{\delta}_T)/\partial\delta\partial\delta')^{-1}$*, that is,* $-(\partial^2\tilde{L}_T(\tilde{\delta}_T)/\partial\delta\partial\delta')^{-1} \to \mathcal{I}(\delta_0)^{-1}$ *a.s. as* $T \to \infty$*.*

Theorem C.1 gives the conventional results concerning the asymptotic properties of a (local) ML estimator. Theorem C.1 and the arguments used to prove it imply the validity of conventional Wald and Likelihood Ratio test procedures, justifying the asymptotic distributions of the test statistics presented in the text. Note that these results do not hold if η_t is Gaussian because then $E[e_{x,t}^2] = E[\eta_t^2] = 1$, showing that, when $\phi_0 = \theta_0$, the upper left hand corner of the matrix $\mathcal{I}(\delta_0)$ in Theorem C.1 is singular.

Proof of Theorem C.1 (outline). Theorem 1 of Meitz and Saikkonen (2011) gives the result of Theorem C.1 in the more general context of a noninvertible ARMA(P,Q) model with ARCH-errors. Given Assumption C.1, the result of Theorem C.1 can be established as in Meitz and Saikkonen (2011). In broad terms, the line of proof is standard, mainly consisting of establishing the following three facts: (i) the rescaled score vector evaluated at the true parameter value is asymptotically normally distributed with zero mean and positive definite covariance matrix $\mathcal{I}(\delta_0)$, (ii) the expectation of the Hessian evaluated at the true parameter value coincides with $-\mathcal{I}(\delta_0)$, and (iii) the rescaled Hessian matrix converges uniformly in some neighborhood of the true parameter value. These and the other necessary facts required to establish Theorem C.1 can be proven by following the steps in the proof Theorem 1 in Meitz and Saikkonen (2011). In general, the required derivations are now considerably shorter than in Meitz and Saikkonen (2011) because therein ARCH-errors were allowed for. We omit the details of the proof, but in what follows briefly discuss two substantial differences in the proofs.

One issue requiring additional explanation is the proof of positive definiteness of $\mathcal{I}(\delta_0)$. As seen above, we now have to assume the errors to be non-Gaussian which differs from Meitz and Saikkonen (2011). Despite this, one can still follow the general line of proof in that paper although the argument can be made considerably simpler. In the considered first order case the positive definiteness of the upper left hand corner of $\mathcal{I}(\delta_0)$ is readily seen by computing the determinant of this matrix and making use of the fact that in the non-Gaussian case $E[e_{x,t}^2] > 1$ holds (see Andrews, Davis, and Breidt (2006), Remark 2). The positive definiteness of the lower right hand corner of $\mathcal{I}(\delta_0)$ can be established by using (a simplified version of) the argument in the proof of Lemma 2 of Meitz and Saikkonen (2011) (see the beginning of Step 4 in their proof).

The second main difference in the proofs comes from the fact that we allow the distribution of η_t to be asymmetric. In Meitz and Saikkonen (2011) this was ruled out in order to simplify the otherwise extremely complex derivations. One reason why symmetricity is not required here is that the derivations are much simpler than in Meitz and Saikkonen (2011), with a majority of the terms present in their derivations now dropping out. Second reason is that those terms that remain also simplify (often because therein the error variance depends on the observations, and here it is a constant). Third reason why symmetricity is not required is that some derivations in Meitz and Saikkonen (2011) could have been justified without symmetricity, but this was made use of because it lead to substantially shorter arguments. ■

References

Andrews, B., Davis, R. A., and Breidt, F. J. (2006). Maximum likelihood estimation for all-pass time series models. Journal of Multivariate Analysis 97, 1638–1659.

Andrews, B., Davis, R. A., and Breidt, F. J. (2007). Rank-based estimation for all-pass time series models. Annals of Statistics 35, 844–869.

Andrews, D. W. K., and Ploberger, W. (1996). Testing for serial correlation against an ARMA(1,1) process. Journal of the American Statistical Association 91, 1331–1342.

Breidt, J., R.A. Davis, K.S. Lii, and M. Rosenblatt (1991). Maximum likelihood estimation for noncausal autoregressive processes. Journal of Multivariate Analysis 36, 175–198.

Breidt, F. J., Davis, R. A., and Trindade, A. A. (2001). Least absolute deviation estimation for all-pass time series models. Annals of Statistics 29, 919–946.

Brockwell, P. J., and Davis, R. A. (1991). Time Series: Theory and Methods, 2nd edn. Springer-Verlag. New York.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 64, 247–254

Lanne, M., and Saikkonen, P. (2011). Noncausal autoregressions for economic time series. Journal of Time Series Econometrics, Vol. 3, Iss. 3, Article 2.

Lii, K.-S., and Rosenblatt, M. (1996). Maximum likelihood estimation for nonGaussian nonminimum phase ARMA sequences. Statistica Sinica 6, 1–22.

Meitz, M., and Saikkonen, P. (2011). Maximum likelihood estimation of a noninvertible ARMA model with autoregressive conditional heteroskedasticity. Unpublished manuscript.

Nankervis, J. C., and Savin, N. E. (2010). Testing for serial correlation: Generalized Andrews-Ploberger tests. Jounal of Business and Economic Statistics 28, 246–255.

Poterba, J. M., and Summers, L. H. (1988). Mean reversion in stock prices: evidence and implications. Journal of Financial Economics 22, 27–59.

Rosenblatt, M. (2000). Gaussian and Non-Gaussian Linear Time Series and Random Fields. Springer-Verlag, New York.

Singleton, K. J. (2006). Empirical Dynamic Asset Pricing. Princeton University Press, Princeton, NJ.

Taylor, S. J. (1982). Tests of the random walk hypothesis against a price-trend hypothesis. Journal of Financial and Quantitative Analysis 17, 37–61.

Wu, R., and Davis, R. A. (2010). Least absolute deviation estimation for general autoregressive moving average time-series models. Journal of Time Series Analysis 31, 98–112.

Table 1: Rejection rates of nominal 5% level Wald and LR tests: ARMA(1,1) models.

ϕ_0	θ_0	Wald Test		LR Test	
		$T = 200$	$T = 500$	$T = 200$	$T = 500$
		$\mathsf{H}_{\text{AP}} : \phi_0 = \theta_0$			
0.00	0.00	0.063	0.054	0.074	0.074
0.80	0.80	0.063	0.053	0.081	0.071
0.80	0.85	0.277	0.544	0.274	0.541
0.80	0.90	0.746	0.988	0.722	0.987
0.80	0.95	0.945	1.000	0.942	1.000
0.80	0.75	0.177	0.430	0.261	0.507
0.80	0.70	0.549	0.927	0.635	0.949
0.80	0.65	0.843	0.998	0.889	0.999
		$\mathsf{H}_{\text{IID}}^{(\text{AP})} : \phi_0 = 0$ in All-Pass Model			
0.00	0.00	0.089	0.047	0.077	0.066
0.10	0.10	0.222	0.327	0.213	0.411
0.20	0.20	0.525	0.847	0.466	0.819
0.40	0.40	0.890	0.991	0.836	0.988
0.60	0.60	0.967	0.997	0.917	0.997
		$\mathsf{H}_{\text{IID}} : \phi_0 = \theta_0 = 0$			
0.00	0.00	0.101	0.066	0.071	0.065
0.10	0.10	0.208	0.303	0.155	0.284
0.20	0.20	0.469	0.756	0.393	0.741
0.40	0.40	0.859	0.984	0.797	0.986
0.60	0.60	0.955	0.997	0.895	0.996

The entries are rejection rates of the null hypotheses $\phi_0 = \theta_0 = 0$ (upper panel) and $\phi_0 = \theta_0$ (lower panel). The DGP is the ARMA(1,1) process with values of the ϕ_0 and θ_0 parameters given in the first and second column, respectively. The errors are generated from Student's t–distribution with 5 degrees of freedom. The number of replications is 10,000.

Table 2: Rejection rates of nominal 5% level Exp-LM$_\infty$ test: ARMA(1,1) models.

ϕ_0	θ_0	Weight Matrix					
		I		AIC		BIC	
		$T=200$	$T=500$	$T=200$	$T=500$	$T=200$	$T=500$
0.00	0.00	0.067	0.051	0.131	0.077	0.080	0.055
0.80	0.80	0.062	0.041	0.163	0.088	0.082	0.060
0.80	0.85	0.184	0.205	0.184	0.205	0.126	0.208
0.80	0.90	0.295	0.787	0.323	0.709	0.311	0.750
0.80	0.95	0.540	0.967	0.522	0.954	0.553	0.964
0.80	0.75	0.139	0.246	0.220	0.239	0.162	0.265
0.80	0.70	0.344	0.755	0.312	0.489	0.352	0.716
0.80	0.65	0.628	0.975	0.397	0.667	0.596	0.916

The entries are rejection rates of the Exp-LM$_\infty$ test with the weight matrix being the identity matrix (I) or selected by the Akaike (AIC) or the Bayesian (BIC) information criterion. The number of autocorrelation coefficients, T_r, included in the test statistic equals 20. The DGP is the ARMA(1,1) process with values of the ϕ_0 and θ_0 parameters given in the first and second column, respectively. The errors are generated from Student's t–distribution with 5 degrees of freedom. The number of replications is 10,000.

Table 3: Estimation and test results for quarterly returns on the value-weighted market return and returns on portfolios formed on size.

	Portfolio			
	Market	Bottom 30%	Middle 40%	Top 30%
ϕ_0	0.793	0.867	0.723	0.791
	(0.070)	(0.034)	(0.081)	(0.070)
θ_0	0.790	0.934	0.789	0.760
	(0.078)	(0.033)	(0.078)	(0.081)
σ_0	8.156	11.809	9.852	7.727
	(0.803)	(1.106)	(0.859)	(0.688)
λ_0	4.146	4.518	4.638	4.486
	(1.241)	(1.688)	(1.575)	(1.415)
	$\mathsf{H}_{\text{AP}} : \phi_0 = \theta_0$			
Wald	0.931	0.053	0.136	0.482
LR	0.930	0.162	0.157	0.460
	$\mathsf{H}_{\text{IID}}^{(\text{AP})} : \phi_0 = 0$ in All-Pass Model			
Wald	6.20e–30	8.96e–9	1.16e–20	4.48e–27
LR	0.006	0.391	0.066	0.085
Exp-LM_∞	0.246	0.085	1.076	0.041

The ML estimates of the noninvertible ARMA(1,1) model are based on the assumption of Student's t–distributed errors with λ_0 degrees of freedom. The figures in parenteheses are standard errors computed from the Hessian of the log-likelihood function. For the Wald and LR tests, p-values are reported. The 10% and 5% critical values of the Exp-LM_∞ test equal 1.418 and 1.973, respectively. The weight matrix in the Exp-LM_∞ test is selected by the BIC with a maximum of three lags, and the number of autocorrelation coefficients, T_r, included in the test statistic equals 20.