Testing for Time-Varying Properties Under Misspecified Conditional Mean and Variance

Daiki Maki1 · Yasushi Ota2

Accepted: 3 July 2020 / Published online: 16 July 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
This study examines statistical performance of tests for time-varying properties under misspecified conditional mean and variance. When time-varying properties of the conditional mean are tested in the case in which data have no time-varying mean but have time-varying variance, asymptotic tests have size distortions. The distortions are minimized by the use of a bootstrap method. Similarly, when time-varying properties of the conditional variance are tested in the case in which data have time-varying mean but no time-varying variance, asymptotic tests have large size distortions. The distortions are not improved even by the use of bootstrap methods. We show that tests for time-varying properties of the conditional mean by the bootstrap are robust regardless of the time-varying variance model, whereas tests for time-varying properties of the conditional variance do not perform well in the presence of misspecified time-varying mean.

Keywords Time-varying properties · Mean · Variance · Misspecified models · Size distortions

1 Introduction

Many economic and financial time-series data have time-varying properties. Their properties are roughly classified into two types. One is a property about the conditional mean. Constant and/or autoregressive parameters change with time as time-varying properties for the mean. For example, Lundbergh et al. (2003) developed a time-varying smooth transition autoregressive model and illustrated

---

1 Faculty of Commerce, Doshisha University, Karasuma-higashi-iru, Imadegawa-dori, Kamigyo-ku, Kyoto 602-8580, Japan
2 Faculty of Management, Okayama University of Science, Okayama, Japan
empirical applications to U.S macroeconomic data. In addition, Holt and Balagtas (2009) explored the possibility of time-varying properties for the mean of U.S. meat demand. Another time-varying property is the conditional variance of the error term for a regression model. Variance of the error term (i.e., volatility) is frequently modeled by autoregressive conditional heteroskedasticity (ARCH). As introduced by Dahlhaus and Rao (2006), a property of the time-varying model for volatility is that the parameters of volatility models change with time. Amado and Teräsvirta (2013) and Kim and Kim (2016) recently provided a time-varying volatility model and its applications.

Linearity tests are usually used to investigate time-varying properties. When we test for time-varying properties of the conditional mean, we assume a homoskedastic variance or a correctly specified linear variance process for the error term of the conditional mean. When we test for time-varying properties of the conditional variance, we assume a correctly specified linear process for a mean. However, it is difficult to know a priori which time-varying property for the mean or variance is present. This challenge implies that researchers may erroneously test for time-varying properties of the mean, although volatility actually has time-varying properties. Similarly, researchers may erroneously test for time-varying properties of the variance, although the mean actually has time-varying properties. In fact, Pitarakis (2004) and Perron and Yamamoto (2019) showed that a test for a change in regression coefficients (variance) that ignores or misspecifies the presence of change in variance (regression coefficients) causes poor statistical properties. For the influence of ARCH on inference of misspecified models, Lumsdaine and Ng (1999) showed overrejection of ARCH tests in the presence of misspecified conditional mean models. Van Dijk et al. (1999b) demonstrated size distortions of ARCH tests when a process has additive outliers. Balke and Kapetanios (2007) pointed out that spurious ARCH effects appear when nonlinearity of the mean is ignored. If we were to perform erroneous tests and fail to obtain reliability from the derived results, it would be difficult to construct and evaluate the correct model. Accordingly, it is important to clarify the influence of tests on the misspecified models.

This study examines the influence of time-varying tests on misspecified conditional mean and variance. In particular, we clarify the statistical performance of tests for time-varying variance when a process has time-varying mean with homoskedastic variance. We also investigate statistical performance of tests for time-varying mean when a process has time-varying variance with a linear mean model. Some studies, including Lumsdaine and Ng (1999), Van Dijk et al. (1999a), and Balke and Kapetanios (2007), have investigated problems of misspecified models in ARCH tests. However, previous studies have not clarified the influence of time-varying tests on misspecified conditional mean and variance. As mentioned above, it is important to clarify time-varying properties correctly. This study uses a logistic smooth transition function to model time-varying mean and variance. This model smooths threshold or structural break models in time. The time-varying tests used in this study are based on the method introduced by Luukkonen et al. (1988a). They developed linearity tests using a Taylor series approximation to overcome the identification problem pointed out by Davies (1977, 1987).
The simulation results in this study provide evidence that the asymptotic test for time-varying mean has size distortions when the conditional variance model is misspecified. However, the wild bootstrap method introduced by Liu (1988) improves the size distortions. In fact, Becker and Hurn (2009) provided evidence that the wild bootstrap improves size properties when the conditional mean is being tested. We can test for time-varying mean by using the wild bootstrap without depending on the form of volatility. When we test for time-varying variance in the presence of a misspecified conditional mean, asymptotic tests have large size distortions. The properties are not improved by the use of the bootstrap ARCH test proposed by Gel and Chen (2012) and the wild bootstrap. The results show that the wild bootstrap tests for time-varying mean are robust regardless of the misspecified conditional variance, whereas tests for time-varying variance do not perform well in the presence of misspecified conditional mean.

The remainder of this paper is organized as follows. Section 2 presents the tests for time-varying mean and variance. Section 3 provides statistical properties of the tests for time-varying mean and variance. Section 4 illustrates empirical applications. Finally, Sect. 5 summarizes and concludes.

2 Tests for Time-Varying Mean and Variance

We assume the following regression model to test for time-varying mean.

\[ y_t = \alpha_0 + \sum_{i=1}^{p} \beta_{0i} y_{t-i} + (\alpha_1 + \sum_{i=1}^{p} \beta_{1i} y_{t-i}) F(t, \gamma, c) + u_t, \tag{1} \]

where \( u_t \sim N(0, \sigma^2) \) and \( F(t, \gamma, c) \) is a transition function to model time-varying properties of \( (1) \). \( (1) \) has a time-varying constant and slope for \( y_{t-i} \) depending on the value of \( F(t, \gamma, c) \). Lin and Teräsvirta (1994) provided the asymptotic theory for time-varying regression models. \( (1) \) reduces to a linear autoregressive model for \( F(t, \gamma, c) = 0 \) and/or \( \alpha_1 = 0 \) and \( \beta_1 = 0 \). \( F(t, \gamma, c) \) is denoted as

\[ F(t, \gamma, c) = (1 + \exp\{-\gamma(t - c)\})^{-1} - \frac{1}{2}, \tag{2} \]

where \( \gamma \) is a parameter of smoothness and \( c \) is a threshold parameter. We assume \( \gamma > 0 \) and \( c > 0 \). \( F(t, \gamma, c) \) is bounded between \(-1/2\) and \(1/2\). \( F(t, \gamma, c) \) moves toward \(-1/2\) when \( t < c \) and small \( \gamma(t - c) \), and moves toward \(1/2\) when \( t > c \) and large \( \gamma(t - c) \). \( F(t, \gamma, c) \) takes 0 for \( t = c \). \( F(t, \gamma, c) \) with \( \gamma = \infty \) equals a structural break model, because \( F(t, \gamma, c) \) with \( \gamma = \infty \) becomes an indicator function that takes the value \(-1/2\) or \(1/2\). Figures 1 and 2 demonstrate the values of \( F(t, \gamma, c) \), for example. Figure 1 depicts the graph of \( F(t, \gamma, c) \) with \( T = 200, \gamma = (0.01, 0.1) \), and \( c = T/2 \). Figure 2 depicts it with \( T = 1,000, \gamma = (0.01, 0.1) \), and \( c = T/2 \). Figure 1 shows that \( F(t, \gamma, c) \) with \( \gamma = 0.01 \) is almost linear, whereas \( F(t, \gamma, c) \) with \( \gamma = 0.1 \) has a smooth change around \( T = 100 \). For Fig. 2, \( F(t, \gamma, c) \) with \( \gamma = 0.01 \) has a smooth change around \( T = 500 \). By contrast, \( F(t, \gamma, c) \) with \( \gamma = 0.1 \) has a rapid change around \( T = 500 \). The time-varying properties of \( (1) \) depend on the values of \( t, \gamma \), and \( c \).
The null and alternative hypotheses to test for time-varying mean in (1) are denoted as follows:

\[ H_0 : \gamma = 0, \quad H_1 : \gamma > 0. \] (3)

(1) becomes a simple autoregressive (AR) model under \( H_0 \) and has a time-varying constant and slope under \( H_1 \) with \( \alpha_1 \) and/or \( \beta_{1i} \). Note that we cannot test for \( \gamma = 0 \) directly, because the null hypothesis has an identification problem about \( \alpha_1 \) and \( \beta_{1i} \), which are identified only under the alternative hypothesis with \( \gamma > 0 \). Davies (1977,

---

**Fig. 1** Value of the transition function with \( T = 200 \)

---

**Fig. 2** Value of the transition function with \( T = 1,000 \). Solid line has \( \gamma = 0.01 \). Dashed line has \( \gamma = 0.1 \)
(1987) discussed the identification problem in which a nuisance parameter is present only under the alternative hypothesis. A solution for the identification problem is to use the Taylor series approximation introduced by Luukkonen et al. (1988a).

When we use a Taylor series approximation around \( \gamma = 0 \), (1) with \( F(t, \gamma, c) \) is given by

\[
y_i = \phi_{10} + \sum_{i=1}^{\nu} \phi_{1i}y_{t-i} + \phi_{20}t + \sum_{i=1}^{\nu} \phi_{2i}ty_{t-i} + e_t,
\]

where \( e_t \) is an error term for the new regression model. We can formulate the null and alternative hypotheses to test for time-varying mean in (4) as follows.

\[
H_0 : \phi_{20} = \phi_{21} = \cdots = \phi_{2p} = 0, \quad H_1 : H_0 \text{ is not true.}
\]

We rewrite (4) as

\[
y_t = \Phi'Y_t + e_t,
\]

where \( \Phi = (\phi_{10}, \phi_{11}, \ldots, \phi_{1p}, \phi_{20}, \phi_{21}, \ldots, \phi_{2p})' \) and \( Y_t = (1, y_{t-1}, \ldots, y_{t-p}, t, ty_{t-1}, \ldots, ty_{t-p})' \). The usual test for (5) uses the following Wald statistic.

\[
M_a = \frac{1}{\hat{\sigma}^2} \hat{\Phi}' R \left( \sum_{i=1}^{T} Y_tY_t' \right)^{-1} R' \hat{\Phi}_2,
\]

where \( \hat{\Phi}_2 \) is the estimate of \( \Phi_2 = (\phi_{20}, \phi_{21}, \ldots, \phi_{2p})' \), \( \hat{\sigma}^2 \) is the estimate of the residual variance obtained from (4), and \( R \) is a matrix such that \( R\hat{\Phi} = \hat{\Phi}_2 \). \( M_a \) has an \( F \) distribution with \( (p+1, T - (2 + 2p)) \) degrees of freedom under the null hypothesis.

Standard asymptotic tests for linearity often cause spurious nonlinearity when errors have heteroskedastic variance, including ARCH, generalized ARCH (GARCH), stochastic volatility, and structural breaks. A method to relieve the influence is to use the heteroskedasticity-consistent covariance matrix estimator (HCCME) introduced by White (1980). This is a popular method to test for linearity in the presence of heteroskedasticity, since test statistics based on the HCCME asymptotically have the same distribution as the original test statistics under the null hypothesis. However, linearity tests using the HCCME do not perform well under the null hypothesis of linearity with ARCH- and GARCH-type errors, particularly for finite samples. The HCCME cannot sufficiently improve the influence of heteroskedastic variance on inference of the conditional mean. As pointed out by Pavlidis et al. (2013), the presence of ARCH errors complicates tests for linearity, because ARCH errors are like nonlinear processes for conditional mean. Some studies, including Pavlidis et al. (2010) and Maki (2014), have demonstrated this problem. A better method is the wild bootstrap proposed by Liu (1988). This method can resample data with unknown heteroskedastic variance and yield better statistical performance than the asymptotic or HCCME tests can. The wild bootstrap test for time-varying mean adopts the following procedure.
Step 1. Estimate regression model (4) and compute test statistic (7) by ordinary least squares (OLS).

Step 2. Estimate the regression model under the null hypothesis with $\phi_{20} = \phi_{21} = \cdots = \phi_{2p} = 0$ and obtain the parameter estimates and residuals denoted as $\hat{e}_{0t}$ by OLS.

Step 3. Generate the bootstrapped sample as follows:

$$y_t^* = \hat{\phi}_{10} + \sum_{i=1}^{p} \hat{\phi}_{1i} y_{t-i} + e_t^*,$$

where $e_t^* = e_t \hat{e}_{0t}$. We set $e_t$ to independent and identically distributed $N(0, 1)$.

Step 4. Estimate regression model (4) using the generated bootstrap sample (8) and compute test statistic (7) denoted as $M_{wb}$ by OLS.

Step 5. Repeat steps 3 and 4 many times.

Step 6. Compute the bootstrap $p$-value as follows:

$$P(M_{wb}) = \frac{1}{M} \sum_{m=1}^{M} I(M_{wb}(m) > M_a),$$

where $M$ is the number of bootstrap iterations and $I(\cdot)$ is an indicator function such that $I(\cdot)$ is 1 if (\cdot) is true and 0 otherwise. Usually, more than 1,000 times are enough for the number of bootstrap iterations. Andrews and Buchinsky (2000) and Davidson and MacKinnon (2000) discussed the problem of the number of bootstrap iterations. The null hypothesis is rejected if (9) is smaller than a significant level.

We next explain the test for time-varying smooth transition variance. Consider the following data generating process (DGP).

$$y_t = a_0 + \sum_{i=1}^{p} \hat{\beta}_0 y_{t-i} + u_t,$$

$$u_t = h_t \varepsilon_t,$$

$$h_t^2 = a_0 + \sum_{j=1}^{q} b_{0j} u_{t-j}^2 + (a_1 + \sum_{j=1}^{q} b_{1j} u_{t-j}^2) F(t, \gamma, c),$$

$$F(t, \gamma, c) = (1 + \exp\{-\gamma(t - c)\})^{-1} - \frac{1}{2}.$$

(10) has a linear AR model for mean and time-varying smooth transition ARCH process (12) for variance. While Hagerud (1996) proposed the smooth transition variance

---

1 Davidson and Flachaire (2008) discussed the method of the wild bootstrap, including the two-point distribution and Rademacher distribution in addition to $N(0, 1)$ for $\varepsilon_t$. Although it is necessary for $\varepsilon_t$ to satisfy $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = 1$, the results among distributions for $\varepsilon_t$ are quite similar.
GARCH model with the transition variable $u_{t-1}$, (13) has the transition variable $t$. The transition function for time-varying ARCH is similar to that of time-varying mean. Since (12) also has an identification problem, we use a Taylor series approximation around $\gamma = 0$ to test for the time-varying smooth transition ARCH. (12) is rewritten as

$$h_t^2 = \rho_{10} + \sum_{j=1}^{q} \rho_{1j} u_{t-j}^2 + \rho_{20} t + \sum_{j=1}^{q} \rho_{2j} u_{t-j}^2 + v_t,$$  \hspace{1cm} (14)

where $v_t$ is an error term, including white noise and the remainder term from the Taylor approximation. The null hypothesis and alternative hypothesis of the test for time-varying ARCH are given by

$$H_0 : \rho_{11} = \cdots = \rho_{1q} = \rho_{20} = \rho_{21} = \cdots = \rho_{2q} = 0, \hspace{1cm} H_1 : H_0 \text{ is not true}. \hspace{1cm} (15)$$

(15) nests ARCH test. To test for (15), we first estimate (10) and obtain residuals by OLS. We next estimate (14) by OLS. The test statistic for (15) is

$$V_a = \frac{(SSR(h)_0 - SSR(h)_1)/(2q + 1)}{SSR(h)_1/(T - (2q + 2))}, \hspace{1cm} (16)$$

where $SSR(h)_0$ is the sum of the squared residuals obtained from the estimation of (14) under the null hypothesis and $SSR(h)_1$ is the sum of the squared residuals obtained from the estimation of (14). $V_a$ has an $F$ distribution with $(q + 1, T - (2q + 2))$ degrees of freedom under the null hypothesis.

Gel and Chen (2012) introduced a new bootstrap test for ARCH to improve the size and power of the asymptotic test. We use their approach to test for time-varying ARCH. The procedure is denoted as follows.

\textit{Step 1}. Estimate regression model (14) and compute test statistic (16) by OLS.

\textit{Step 2}. Estimate the regression model under the null hypothesis with $\rho_{11} = \cdots = \rho_{1q} = \rho_{20} = \rho_{21} = \cdots = \rho_{2q} = 0$ and obtain the parameter estimates $\hat{\rho}_{10}$ and residuals denoted as $\hat{\epsilon}_t$ by OLS.

\textit{Step 3}. Generate the bootstrapped sample as follows:

\begin{itemize}
  \item [(12)] includes conditional and unconditional heteroskedasticity. Meanwhile, conditional heteroskedasticity has $b_{0j} > 0$ and/or $b_{1j} > 0$ with $F(\cdot) \neq 0$, unconditional heteroskedasticity has $a_0 > 0$ or $a_1 > 0$ with $F(\cdot) \neq 0$ in addition to $b_{0j} = \cdots = b_{0q} = b_{1j} = \cdots = b_{1q} = 0$.
  \item [(3)] It is possible to test for the null hypothesis of $\rho_{20} = \rho_{21} = \cdots = \rho_{2q} = 0$. This means that in the null hypothesis, there is an ARCH error and in the alternative hypothesis, there is a time-varying ARCH error. However, it is difficult to know that the variance process has ARCH properties a priori. When we have the null hypothesis $\rho_{20} = \rho_{21} = \cdots = \rho_{2q} = 0$, we have to test for the hypothesis of whether variance has ARCH error before testing for $\rho_{20} = \rho_{21} = \cdots = \rho_{2q} = 0$. This requires a two-step approach and cannot test for time-varying ARCH directly. This may make comparisons among the tests difficult and lead to misleading results. Therefore, we adopt hypotheses in (15) to avoid these problems and simply test for time-varying ARCH.
  \item [(4)] The test statistic is also expressed by $TR^2$, where $R^2$ is the coefficient of the determination of (12). $TR^2$ asymptotically has $\chi^2(2q + 1)$ distribution. See Gel and Chen (2012).
\end{itemize}
where $v^*_{t}$ is randomly selected from $(\hat{\vartheta}_{01} \cdots \hat{\vartheta}_{0T})$.

**Step 4.** Estimate regression model (14) using the generated bootstrap sample (17) and compute test statistic (16) denoted as $V_b$ by OLS.

**Step 5.** Repeat steps 3 and 4 many times.

**Step 6.** Compute the bootstrap $p$-value as follows:

$$P(V_b) = \frac{1}{M} \sum_{m=1}^{M} I(V_{b(m)} > V_a).$$  \hspace{1cm} (18)

The null hypothesis is rejected if the $p$-value (18) is smaller than a significant level. We also use the wild bootstrap test for (15). This is similar to the bootstrap test shown above. The main difference is Step 3. The wild bootstrap approach replaces $v^*_{t}$ as $\tilde{v}_t = \hat{\vartheta}_t \eta_t$, where $\eta_t \sim N(0,1)$. We denote the wild bootstrap test for time-varying ARCH as $V_{wb}$. The $p$-value is given by

$$P(V_{wb}) = \frac{1}{M} \sum_{m=1}^{M} I(V_{wb(m)} > V_a).$$  \hspace{1cm} (19)

### 3 Statistical Properties of Time-Varying Tests

This section examines the statistical properties of the time-varying tests reviewed in Sect. 2. We conduct Monte Carlo simulations to compare the size and power of the test statistics. The simulations are based on 10,000 replications, the nominal level at 5%, and sample sizes with $T = 100, 200, 400$, and 1000. We show only nominal power properties and do not present size-corrected power, because the purpose of this study is to compare the tests and applied practitioners do not usually use size-corrected tests. Bootstrap tests have 1,000 replications. To avoid the effect of initial conditions, data with $T + 100$ are generated. The initial 100 samples are discarded and we use the data with sample size $T$. Tests compared in this section are denoted as $M_a, M_{wb}, V_a, V_b$, and $V_{wb}$.

First, as a benchmark, autoregressive processes with homoskedastic error are generated.

$$y_t = \alpha_0 + \beta_0 y_{t-1} + u_t,$$  \hspace{1cm} (20)

where $u_t \sim$ i.i.d.$N(0,1)$. $\alpha_0$ is set to $\alpha_0 = 1$. The persistent parameter $\beta_0$ is set to $\beta_0 = 0.3$ and $0.9$ in order to investigate the influence of persistence of the process on time-varying tests. Since (20) has no time-varying properties in mean and variance, the rejection frequencies of each test demonstrate the empirical size for (20). Table 1 presents the rejection frequencies of each test. When the persistent parameter is $\beta_0 = 0.3$, $V_a$ has slight underrejection for $T = 100, 200$, and 400. Other tests
have empirical size near to the 5% nominal level and reasonable size properties regardless of sample size. Although time-varying tests for mean, $M_a$ and $M_{wb}$, have overrejection in small samples with $\beta_0 = 0.9$, the overrejection decreases with the increased sample size. Unlike time-varying tests for the mean, time-varying tests for the variance do not have size distortions, except for $V_a$ with $T = 100$, which has slight underrejection.

Next, we consider the following DGP to examine the power properties of $M_a$ and $M_{wb}$ and the size properties of $V_a$, $V_b$, and $V_{wb}$ under the following time-varying mean model.

$$y_t = 1 + 0.3y_{t-1} + (\alpha_1 + \beta_1 y_{t-1})F(\cdot) + u_t,$$

$$F(\cdot) = (1 + \exp\{-\gamma(t - c)\})^{-1} - \frac{1}{2}.$$  \hfill (21)

Parameters for the time-varying mean are set to $(\alpha_1, \beta_1) = (0, 0.3), (0, 0.6), (0.5, 0.3), \text{ and } (1, 0.3)$. $F(\cdot)$ moves from $-1/2$ to $1/2$. For $(\alpha_1, \beta_1) = (1, 0.3)$, (21) has time-varying properties from $y_t = 0.5 + 0.15y_{t-1} + u_t$ to $y_t = 1.5 + 0.45y_{t-1} + u_t$. We examine the effect of the magnitude of change for constant and AR parameters on performance of the tests. Smoothness parameter $\gamma$ in $F(\cdot)$ is set to $\gamma = (0.01, 0.1)$. $\gamma = 0.01$ means more smooth and slight change in $F(\cdot)$ than $\gamma = 0.1$. In addition, the threshold parameter $c$ is set to $c = T/2$. The simulation results are shown in Table 2. The power of $M_a$ and $M_{wb}$ is low for parameter sets $(\alpha_1, \beta_1, \gamma) = (0, 0.3, 0.01)$ with $T = 100$ and 200. It increases rapidly with large sample size. We observe that the magnitude of $\alpha_1$ and $\beta_1$ affects the power of $M_a$ and $M_{wb}$. For example, the power of

| $\beta_0 = 0.3$ | $M_a$ | $M_{wb}$ | $V_a$ | $V_b$ | $V_{wb}$ |
|-----------------|-------|---------|-------|-------|---------|
| $T = 100$       | 0.049 | 0.051  | 0.034 | 0.046 | 0.051  |
| $T = 200$       | 0.048 | 0.049  | 0.041 | 0.045 | 0.054  |
| $T = 400$       | 0.053 | 0.054  | 0.040 | 0.050 | 0.051  |
| $T = 1000$      | 0.046 | 0.054  | 0.046 | 0.050 | 0.046  |

| $\beta_0 = 0.9$ | $M_a$ | $M_{wb}$ | $V_a$ | $V_b$ | $V_{wb}$ |
|-----------------|-------|---------|-------|-------|---------|
| $T = 100$       | 0.106 | 0.104  | 0.035 | 0.047 | 0.048  |
| $T = 200$       | 0.080 | 0.074  | 0.043 | 0.049 | 0.044  |
| $T = 400$       | 0.054 | 0.061  | 0.045 | 0.049 | 0.051  |
| $T = 1000$      | 0.056 | 0.057  | 0.045 | 0.050 | 0.045  |

We conduct Monte Carlo experiments under various other situations. For example, while tests have larger rejection frequencies when $c$ is smaller than $T/2$, they have smaller rejection frequencies when $c$ is larger than $T/2$. However, differences among the tests are similar to those of $c = T/2$. The results for other parameter sets $(\alpha_1, \beta_1, \gamma, \text{ and } c)$ are available from the author on request.
Ma and Mwb for \((\alpha_1, \beta_1) = (0, 0.3)\) with \(T = 200\) are 0.134 and 0.132, respectively, whereas those for \((\alpha_1, \beta_1) = (1, 0.3)\) with \(T = 200\) are 0.694 and 0.703, respectively. When the smoothness parameter is \(\gamma = 0.1\), \(Ma\) and \(Mwb\) have sufficient power. \(\gamma\) has clear impact on the performance of \(Ma\) and \(Mwb\). Generally, we observe that \(Ma\) and \(Mwb\) are able to find time-varying mean properties in the presence of homoskedastic variance.

For \(Va\), \(Vb\), and \(Vwb\), the results in Table 2 show the empirical size properties of the tests, because (21) has homoskedastic variance. \(Va\) performs well in small samples. It has overrejection for \(\gamma = 0.01\) and \(T = 1, 000\). In particular, \(Va\) has a rejection frequency of 0.325 and large size distortions when \((\alpha_1, \beta_1) = (0, 0.6)\) and \(T = 1, 000\). We find that the increase in the time-varying AR parameter causes high rejection frequencies for the time-varying variance test under the null hypothesis of homoskedastic variance. Although the change of a constant of (21) also increases the rejection frequencies, the effect is not larger than the increase in the AR parameter. The rejection frequencies of \(Va\) for \(T = 1, 000\) with \((\alpha_1, \beta_1) = (0, 0.3)\) and \((\alpha_1, \alpha_1) = (1, 0.3)\) are 0.068 and 0.119, respectively. Similar properties are observed for \(Vb\). The results imply that the bootstrap time-varying ARCH test does not perform well under the null hypothesis of homoskedastic variance with a misspecified conditional mean. By contrast, the size properties of \(Vwb\) outperform those of \(Va\) and \(Vb\). \(Vwb\) does not have overrejection even for \(T = 1, 000\). However, \(Vwb\) has size distortions for \(\gamma = 0.1\).

### Table 2 Rejection frequencies under time-varying AR with homoskedastic error

| \(\gamma = 0.01\) | \(\gamma = 0.1\) |
|---------------------|---------------------|
| \(M_a\) | \(M_{wb}\) | \(V_a\) | \(V_b\) | \(V_{wb}\) | \(M_a\) | \(M_{wb}\) | \(V_a\) | \(V_b\) | \(V_{wb}\) |
| \(T = 100\) | 0.056 | 0.061 | 0.036 | 0.043 | 0.048 | 0.380 | 0.382 | 0.036 | 0.046 | 0.042 |
| \(T = 200\) | 0.134 | 0.132 | 0.040 | 0.048 | 0.049 | 0.789 | 0.794 | 0.036 | 0.051 | 0.035 |
| \(T = 400\) | 0.625 | 0.630 | 0.049 | 0.053 | 0.041 | 0.986 | 0.988 | 0.042 | 0.067 | 0.035 |
| \(T = 1000\) | 1 | 0.999 | 0.068 | 0.072 | 0.031 | 1 | 1 | 0.119 | 0.086 | 0.036 |
| \(\alpha_1 = 0, \beta_1 = 0.3\) | \(T = 100\) | 0.091 | 0.090 | 0.035 | 0.043 | 0.048 | 0.943 | 0.941 | 0.055 | 0.071 | 0.029 |
| \(T = 200\) | 0.428 | 0.424 | 0.041 | 0.053 | 0.045 | 1 | 0.999 | 0.126 | 0.133 | 0.040 |
| \(T = 400\) | 0.996 | 0.997 | 0.064 | 0.076 | 0.027 | 1 | 1 | 0.286 | 0.296 | 0.083 |
| \(T = 1000\) | 1 | 1 | 0.325 | 0.332 | 0.057 | 1 | 1 | 0.704 | 0.703 | 0.253 |
| \(\alpha_1 = 0, \beta_1 = 0.6\) | \(T = 100\) | 0.085 | 0.091 | 0.036 | 0.042 | 0.044 | 0.880 | 0.875 | 0.038 | 0.046 | 0.034 |
| \(T = 200\) | 0.357 | 0.365 | 0.040 | 0.047 | 0.048 | 0.998 | 0.999 | 0.049 | 0.054 | 0.045 |
| \(T = 400\) | 0.990 | 0.992 | 0.046 | 0.055 | 0.038 | 1 | 1 | 0.068 | 0.077 | 0.044 |
| \(T = 1000\) | 1 | 1 | 0.074 | 0.078 | 0.035 | 1 | 1 | 0.139 | 0.144 | 0.070 |
| \(\alpha_1 = 0.5, \beta_1 = 0.3\) | \(T = 100\) | 0.121 | 0.129 | 0.036 | 0.042 | 0.042 | 0.994 | 0.994 | 0.040 | 0.052 | 0.031 |
| \(T = 200\) | 0.694 | 0.703 | 0.036 | 0.043 | 0.044 | 1 | 1 | 0.070 | 0.080 | 0.044 |
| \(T = 400\) | 1 | 1 | 0.042 | 0.052 | 0.034 | 1 | 1 | 0.127 | 0.132 | 0.058 |
| \(T = 1000\) | 1 | 1 | 0.119 | 0.125 | 0.040 | 1 | 1 | 0.300 | 0.311 | 0.113 |

\(M_a\) and \(M_{wb}\) for \((\alpha_1, \beta_1) = (0, 0.3)\) with \(T = 200\) are 0.134 and 0.132, respectively, whereas those for \((\alpha_1, \beta_1) = (1, 0.3)\) with \(T = 200\) are 0.694 and 0.703, respectively. When the smoothness parameter is \(\gamma = 0.1\), \(Ma\) and \(Mwb\) have sufficient power. \(\gamma\) has clear impact on the performance of \(Ma\) and \(Mwb\). Generally, we observe that \(Ma\) and \(Mwb\) are able to find time-varying mean properties in the presence of homoskedastic variance.
Actually, the rejection frequencies of $V_{wb}$ for $T = 1,000$ with $(\alpha_1, \beta_1) = (0, 0.6)$ and $(\alpha_1, \beta_1) = (1, 0.3)$ are 0.253 and 0.113, respectively. $V_{wb}$ does not improve the size properties when the DGP has time-varying mean with homoskedastic error. $V_a$ and $V_b$ for $\gamma = 0.1$ have larger size distortions than those for $\gamma = 0.01$.

When we test for time-varying ARCH in the presence of a misspecified conditional mean, the tests cannot lead to reliable results, particularly for a large change in time-varying parameters with a large smoothness parameter $\gamma$. As shown by Balke and Kapetanios (2007), neglected nonlinearity in the conditional mean causes spurious heteroskedasticity, because the conditional variance includes extra nonlinearity and lags. This is clearer when $T$ increases\(^6\). If the mean is misspecified, the residual includes nonlinearity of the mean. Large parameter shift and sample size lead to stronger nonlinearity. This is the reason for poor performance of the time-varying variance test. Therefore, time-varying ARCH tests in the presence of a misspecified conditional mean have size distortions.

Table 3 reports the rejection frequencies of the tests for the AR process with ARCH error. The DGP is as follows.

$$y_t = 1 + 0.3y_{t-1} + u_t,$$

$$u_t = h_t \epsilon_t,$$

$$h_t^2 = a_0 + b_0 u_{t-1}^2$$

\(^6\) See Luukkonen et al. (1988b) for similar results.
We set a constant parameter for ARCH to $a_0 = 1$. Persistence parameter $b_0$ for ARCH is set to $b_0 = 0.3, 0.6,$ and 0.9 in order to examine the effect of persistence of variance on the performance of the tests. $b_0 < 1$ is the necessary and sufficient condition for weak stationarity of a semi-strong process. For (23), (24), and (25), the rejection frequencies of $M_a$ and $M_{wb}$ show the empirical size, because the process does not have time-varying mean. The performances of $V_a, V_b,$ and $V_{wb}$ show frequencies of finding ARCH properties. $M_a$ overrejects the null hypothesis at 5% significance and its overrejection increases when the persistence parameter and/or sample size increases. $M_a$ is sensitive to persistence in the presence of the ARCH effect. $M_{wb}$ has slight overrejection only for $b_0 = 0.9$ with large samples. However, $M_{wb}$ clearly outperforms $M_a$ and has reasonable and acceptable rejection frequencies in most cases. This is due to the property that the wild bootstrap can deal with heteroskedasticity of unknown form when the conditional mean is correctly specified. We observe the difference of power properties among $V_a, V_b,$ and $V_{wb}$. $V_a$ and $V_b$ have sufficient power to find the ARCH effect. Their abilities to detect the ARCH effect increases for large $b_0$ and $T$. The ability of $V_{wb}$ is lower than that of $V_a$ and $V_b$. For example, the power of $V_a, V_b,$ and $V_{wb}$ for $b_1 = 0.3$ and $T = 400$ is 0.937, 0.937, and 0.260, respectively. While the size properties of $V_{wb}$ are superior to $V_a$ and $V_b$ under a misspecified conditional mean reported in Table 2, the power of $V_{wb}$ is clearly inferior to $V_a$ and $V_b$. The higher power of $V_a$ and $V_b$ is due to size distortions presented in Table 2. The comparison indicates that the use of these tests is not effective from the viewpoint of size and power, because size-corrected tests are needed.

We finally investigate the statistical properties of the tests for the AR process with time-varying ARCH error. The DGP is given by

$$y_t = 1 + 0.3y_{t-1} + u_t,$$  \hspace{1cm} (26)

$$u_t = h_t \epsilon_t,$$ \hspace{1cm} (27)

$$h_t^2 = 1 + 0.3u_{t-1}^2 + (a_1 + b_1 u_{t-1}^2)F(\cdot),$$ \hspace{1cm} (28)

$$F(t, \gamma, c) = \left(1 + \exp \left\{-\gamma \left(t - \frac{T}{2}\right)\right\}\right)^{-1} - \frac{1}{2}.$$ \hspace{1cm} (29)

Time-varying parameters $(a_1, b_1)$ are set to $(a_1, b_1) = (0, 0.3), (0, 0.6), (0.5, 0.3),$ and $(1, 0.3)$, respectively. We consider DGP with $\gamma = 0.01$ and 0.1. These settings are similar to those in Table 2, in which the DGP has time-varying mean. The results are presented in Table 4. $M_{wb}$ does not have overrejection for all cases and is close to the nominal size at 5%. This means that $M_{wb}$ is a reliable test for time-varying mean and does not lead to spurious time-varying mean. Although $M_a$ has size distortions under time-varying ARCH error, the rejection frequencies do not overly depend on the size of parameters $a_1$, $b_1$, and $\gamma$. The rejection frequencies mainly depend on sample size and slowly increase when the sample size increases. The results are similar to those of Zhou (2013) and Boldea et al. (2019). They show that the wild bootstrap is asymptotically valid for a change in mean or regression coefficients in the
presence of conditional and unconditional heteroskedasticity\textsuperscript{7}. The wild bootstrap can replicate unknown heteroskedasticity of the errors. This property yields valid tests for linearity of the mean.

The power of $V_b$ is a little better than that of $V_a$. These tests have sufficient power to find time-varying ARCH effects. Furthermore, the increase in the smoothness parameter $\gamma$ brings higher power of time-varying ARCH tests. In particular, the effect is clear for $(a_1, b_1) = (1, 0.3)$. When the sample size is $T = 100$, the power of $V_a$ and $V_b$ is 0.387 and 0.429 for $\gamma = 0.01$, respectively, and 0.774 and 0.813 for $\gamma = 0.1$, respectively. Note that the results of $V_{wb}$ are clearly different from those of $V_a$ and $V_b$. The power of $V_{wb}$ is lower than that of $V_a$ and $V_b$ in all cases. For example, when the process has $(a_1, b_1) = (0.5, 0.3)$ and $T = 200$, the power of $V_a$, $V_b$, and $V_{wb}$ are 0.724, 0.744, and 0.154 for $\gamma = 0.01$, respectively. $V_a$ and $V_b$ cause large overrejection under the null hypothesis with time-varying mean, whereas they have higher power under the alternative hypothesis. By contrast, $V_{wb}$ has better size properties than $V_a$ and $V_b$ do, whereas $V_{wb}$ has clearly lower power than $V_a$ and $V_b$ do. The Monte Carlo simulation results provide evidence that $M_{wb}$ tests yield more reliable

\textsuperscript{7} Zhang and Wu (2019) proposed a nonparametric test for a change in regression coefficients in the presence of time-varying variance.
results than do $M_a$ under ARCH and time-varying ARCH, and time-varying ARCH tests do not perform well in the presence of time-varying mean.

4 Empirical Applications

In this section, we apply the tests to economic and financial variables. We use five types of economic and financial data, namely, the industrial production index (IPI), consumer price index (CPI), S&P500 Index, 10-year treasury rate (10YTR), and Real M2. While the first difference of logarithm is used for IPI, CPI, S&P500, and Real M2, the level variable is employed for 10YTR. The data are obtained from the Federal Reserve Bank of St. Louis. Monthly IPI, CPI, 10YTR, and Real M2 data consist of 1214 (January 1919–February 2020), 878 (January 1947–February 2020), 700 (January 1962–March 2020), and 734 (January 1959–February 2020) observations, respectively. Daily S&P500 data consist of 1261 (April 8, 2015–April 8 2020) observations. The lag lengths of each tests are selected by the Akaike information criterion (AIC).

The empirical results are presented in Table 5, which tabulates the $p$ values of each test. The results support simulation evidence demonstrated in Sect. 3. For IPI, the $p$ values of $M_a$ and $M_{wb}$ are 0.016 and 0.031, indicating that IPI has the time-varying properties of the mean. As shown in Sect. 3, $M_a$ and $M_{wb}$ have enough power to find time-varying properties. The $p$ values of $V_a$, $V_b$, and $V_{wb}$ are also small. However, it is possible that the results of $V_a$, $V_b$, and $V_{wb}$ do not imply time-varying ARCH effects, because Monte Carlo simulation in Sect. 3 demonstrates that when time-series processes have time-varying mean properties, $V_a$, $V_b$, and $V_{wb}$ overreject the null hypothesis and tend to lead to spurious time-varying ARCH effects. Similar results are observed for 10YTR and Real M2. While the $p$ values of $M_a$ and $M_{wb}$ are small enough to reject the null hypothesis of linearity of the mean for 10YTR and Real M2, $V_a$ and $V_b$ also have small $p$ values. The $p$ values of $V_{wb}$ are larger than those of $V_a$ and $V_b$. The reason is that $V_{wb}$ does not have overrejection whereas $V_a$ and $V_b$ do, as presented in Table 2.

For CPI and S&P500, $V_a$ and $V_b$ can reject the null hypothesis of linearity of the variance, whereas $M_{wb}$ does not reject the null hypothesis of linearity of the mean at the 5% significance level. The results indicate that CPI and S&P500 have linear mean properties with time-varying ARCH effects, because $M_{wb}$ has superior size properties in the presence of heteroskedasticity and $V_a$ and $V_b$ have enough power to find time-varying ARCH effects. Although the $p$ value of $M_a$ for S&P500 is less

| Table 5 | Empirical results |
|---------|-------------------|
|         | $M_a$ | $M_{wb}$ | $V_a$ | $V_b$ | $V_{wb}$ |
| IPI     | 0.016 | 0.031   | 0.000 | 0.000 | 0.057    |
| CPI     | 0.465 | 0.798   | 0.052 | 0.010 | 0.356    |
| S&P500  | 0.000 | 0.141   | 0.000 | 0.001 | 0.267    |
| 10YTR   | 0.072 | 0.009   | 0.037 | 0.012 | 0.072    |
| Real M2 | 0.000 | 0.030   | 0.065 | 0.004 | 0.146    |
than 1%, it is possible that the results are due to the overrejection of the null hypothesis in the presence of time-varying ARCH effects, as demonstrated in Tables 3 and 4. We observe that $V_{wb}$ does not provide evidence of time-varying ARCH effects for CPI and S&P500. This is because $V_{wb}$ has lower power than $V_a$ and $V_b$. The empirical applications show that $M_{wb}$ can appropriately find time-varying properties of the mean, whereas there are cases in which $M_a, V_a, V_b,$ and $V_{wb}$ lead to unreliable results.

5 Summary and Conclusion

This study examined the statistical performance of time-varying tests under misspecified conditional mean and variance. Although time-varying properties are frequently observed in various economic and financial data, it is difficult to know a priori which time-varying property for mean or variance is present. When employing misspecified conditional models, researchers may obtain unreliable results. Therefore, it is important to clarify the statistical properties of time-varying tests in misspecified conditional mean or variance models. Monte Carlo simulation results reveal that asymptotic tests for time-varying mean have size distortions when the variance model is misspecified, whereas the wild bootstrap method improves the size distortions. Wild bootstrap tests for time-varying mean are robust regardless of the misspecified variance, and can lead to reliable results of tests for time-varying mean. However, bootstrap tests in addition to asymptotic tests for time-varying variance have size distortions in the presence of the misspecified conditional mean. Tests for time-varying variance do not perform well or provide reliable results in the presence of the misspecified conditional mean. Robust time-varying variance tests in the presence of the misspecified conditional mean remain for further study.

Acknowledgement This research was supported by KAKENHI (Grant Number: 16K03604).

References

Amado, C., & Teräsvirta, T. (2013). Modelling volatility by variance decomposition. *Journal of Econometrics, 175*, 142–153.

Andrews, D. W. K., & Buchinsky, M. (2000). A three-step method for choosing the number of bootstrap repetitions. *Econometrica, 68*, 23–51.

Balke, A. P., & Kapetanios, G. (2007). Testing for ARCH in the presence of nonlinearity of unknown form in the conditional mean. *Journal of Econometrics, 137*, 472–488.

Becker, R., & Hurn, S. (2009). Testing for nonlinearity in mean in the presence of heteroskedasticity. *Economic Analysis & Policy, 39*, 313–326.

Boldea, O., Cornea-Madeira, A., & Hall, A. R. (2019). Bootstrapping structural change tests. *Journal of Econometrics, 213*, 359–397.

Catani, P. S., & Ahlgren, N. J. C. (2017). Combined Lagrange multiplier test for ARCH in vector autoregressive models. *Econometrics and Statistics, 1*, 62–84.

Dahlhaus, R., & Rao, S. S. (2006). Statistical inference for time-varying ARCH processes. *The Annals of Statistics, 34*, 1075–1114.

Davidson, R., & Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics, 146*, 162–169.
Davidson, R., & MacKinnon, J. G. (2000). Bootstrap tests: How many bootstraps? *Econometric Reviews, 19*, 55–68.

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

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika, 74*, 33–43.

Gel, Y. R., & Chen, B. (2012). Robust Lagrange multiplier test for detecting ARCH/GARCH effect using permutation and bootstrap. *The Canadian Journal of Statistics, 40*, 405–426.

Hagerud, G. (1996). *A new non-linear GARCH model*. Stockholm: Stockholm School of Economics: EFI Economic Research Institute.

Holt, M. T., & Balagtas, J. (2009). Estimating structural change with smooth transition regressions: An application to meat demand. *American Journal of Agricultural Economics, 91*, 1424–1431.

Kim, K. H., & Kim, T. (2016). Capital asset pricing model: A time-varying volatility approach. *Journal of Empirical Finance, 37*, 268–281.

Lin, C.-F. J., & Teräsvirta, T. (1994). Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics, 62*, 211–228.

Liu, R. Y. (1988). Bootstrap procedure under some non-i.i.d. models. *Annals of Statistics, 16*, 1696–1708.

Lumsdaine, R. L., & Ng, S. (1999). Testing for ARCH in the presence of a possibly misspecified conditional mean. *Journal of Econometrics, 93*, 257–279.

Lundbergh, S., Teräsvirta, T., & Van Dijk, V. (2003). Time-varying smooth transition autoregressive models. *Journal of Business & Economic Statistics, 21*, 104–121.

Luukkonen, R., Saikkonen, P., & Teräsvirta, T. (1988a). Testing linearity against smooth transition autoregressive models. *Biometrika, 70*, 491–499.

Luukkonen, R., Saikkonen, P., & Teräsvirta, T. (1988b). Testing linearity in univariate time series models. *Scandinavian Journal of Statistics, 15*, 161–175.

Maki, D. (2014). A comparison of linearity tests based on wild bootstrap. *Advances and Applications in Statistics, 4*(3), 93–107.

Pavlidis, E., Paya, I., & Peel, D. (2010). Specifying smooth transition regression models in the presence of conditional heteroskedasticity of unknown form. *Studies in Nonlinear Dynamics and Econometrics, 14*(3), 1–40.

Pavlidis, E., Paya, I., & Peel, D. (2013). Nonlinear causality tests and multivariate conditional heteroskedasticity: A simulation study. *Studies in Nonlinear Dynamics and Econometrics, 17*(3), 297–312.

Perron, P., & Yamamoto, Y. (2019). Pitfalls of two-step testing for changes in the error variance and coefficients of a linear regression model. *Econometrics, 7*(2), 22, 1–11.

Pitarakis, J. Y. (2004). Least squares estimation and tests of breaks in mean and variance under misspecification. *Econometrics Journal, 7*, 32–54.

Van Dijk, D., Franses, P. H., & Lucas, A. (1999a). Testing for ARCH in the presence of additive outliers. *Journal of Applied Econometrics, 14*, 539–562.

Van Dijk, D., Franses, P. H., & Lucas, A. (1999b). Testing for smooth transition nonlinearity in the presence of outliers. *Journal of Business & Economic Statistics, 17*, 217–235.

Van Dijk, D., Teräsvirta, T., & Franses, P. H. (2002). Smooth transition autoregressive models—A survey of recent developments. *Econometric Reviews, 21*, 1–47.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 14*, 1261–1295.

Zhang, E., & Wu, J. (2019). Testing for structural changes in linear regressions with time-varying variance. *Communications in Statistics-Theory and Methods* (in press).

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association, 108*, 726–740.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.