



# Testing heteroskedasticity for predictive regressions with nonstationary regressors<sup>☆</sup>



Shaoxin Hong<sup>a</sup>, Zhengyi Zhang<sup>b,\*</sup>, Zongwu Cai<sup>c</sup>

<sup>a</sup> Center for Economic Research, Shandong University, Jinan 250100, Shandong, China

<sup>b</sup> International School of Economics and Management, Capital University of Economics and Business, Beijing 100070, China

<sup>c</sup> Department of Economics, University of Kansas, Lawrence, KS 66045, USA

## ARTICLE INFO

### Article history:

Received 4 January 2021

Received in revised form 10 February 2021

Accepted 11 February 2021

Available online 13 February 2021

### JEL classification:

C12

C22

### Keywords:

Cramér–von Mises test statistic

Heteroskedasticity

Nonstationarity

Predictive regressions

Specification test

## ABSTRACT

This paper proposes the Cramér–von Mises type test statistic for testing heteroskedasticity in predictive regression when regressors are nonstationary. A Monte Carlo simulation study is conducted to illustrate the finite sample performance and a real empirical example is examined.

© 2021 Elsevier B.V. All rights reserved.

## 1. Introduction

In the recent three decades, testing predictability of asset returns has not only attracted vast amounts of attention from financial practitioners due to its key component to evaluate mutual fund managers' performance, examine the validity of asset pricing models, and improve asset allocation efficiency, but also has been an important research topic in financial economics in academia. A large literature has been devoted to testing whether asset returns are predictable or not. The typical econometric method used in literature is an ordinary least squares (OLS) regression of returns versus the lag of the financial variables, and conventional t-statistics are used to check the significance of coefficients.

For easy exposition, a structural simple predictive regression model is given by

$$y_t = \beta_0 + \beta_1 x_{t-1} + u_t, \quad 1 \leq t \leq T, \quad (1)$$

where  $x_{t-1}$  is a nonstationary time series. Here,  $\text{Var}(u_t | x_{t-1}) = \sigma_t^2$  might not be a constant. Then,  $u_t$  can be written as  $u_t = \sigma_t \epsilon_t$ ,

<sup>☆</sup> The authors acknowledge the financial supports, in part, from the National Natural Science Foundation of China with grant numbers 71631004 and 72033008.

\* Correspondence to: 121 Zhangjia Lukou, Fengtai District, Beijing 100070, China.

E-mail address: [zhangzhengyi@cueb.edu.cn](mailto:zhangzhengyi@cueb.edu.cn) (Z. Zhang).

where  $\{\epsilon_t\}$  is a sequence of iid random variable or a stationary time series with  $E(\epsilon_t) = 0$  and  $\text{Var}(\epsilon_t) = 1$ . Thus, (1) can be written as

$$y_t = \beta_0 + \beta_1 x_{t-1} + \sigma_t \epsilon_t, \quad (2)$$

where it is commonly assumed that  $\sigma_t$  is a constant and  $x_{t-1}$  satisfies the following autoregressive (AR) model

$$x_t = \rho x_{t-1} + v_t, \quad (3)$$

where  $\rho = 1 + c/T$  for some unknown constant  $c$ , and  $v_t$  is a weakly stationary time series, say, an  $\alpha$ -mixing process. Clearly, if  $c = 0$ ,  $x_t$  is a unit root process, denoted by  $I(1)$  and it is nearly integrated process, denoted by  $NI(1)$ , if  $c < 0$ . Under the assumption that  $\sigma_t$  is a constant and  $x_{t-1}$  satisfies (3), there is vast amounts of literature for testing the predictability as  $H_0 : \beta_1 = 0$ . The reader is referred to the recent survey paper by Liao et al. (2018) for details on the various methods, from which, one can see that when  $x_t$  is nonstationary, the assumption that  $\sigma_t$  is constant is key to derive the asymptotic theory for testing  $H_0 : \beta_1 = 0$ . Recently, by extending the work by Park (2002), Choi et al. (2016) reexamined stock return predictability by assuming that  $\sigma_t = \sigma_0(t/T)$  or  $\sigma_t = \sigma_0(z_t/\sqrt{T})$  for some known function  $\sigma_0(\cdot)$  to satisfy some conditions.

Therefore, by assuming that  $\sigma_t$  is possibly nonstationary and a function of  $x_{t-1}$  as  $\sigma_t = \sigma(x_{t-1})$ , it is of great importance to consider the following test problem; that is to test heteroskedasticity,

**Table 1**  
Size performance with nominal size 5%.

		$\delta = -0.95$		$\delta = -0.25$	
		$T = 500$	$T = 1000$	$T = 500$	$T = 1000$
$c = 0$	$h_1$	0.046	0.050	0.040	0.042
	$h_2$	0.049	0.046	0.044	0.048
	$h_3$	0.033	0.041	0.040	0.040
$c = -5$	$h_1$	0.036	0.041	0.032	0.044
	$h_2$	0.046	0.049	0.043	0.047
	$h_3$	0.026	0.037	0.022	0.033
$c = -30$	$h_1$	0.036	0.044	0.034	0.040
	$h_2$	0.041	0.045	0.049	0.045
	$h_3$	0.027	0.034	0.024	0.032

$$H_0 : \sigma(\cdot) = \sigma_0, \tag{4}$$

where  $\sigma_0$  is an unknown or known parameter. Of course, it would be interesting to consider other types of test.

The rest of the paper is organized as follows. The construction of the proposed test statistic is presented in Sections 2 and 3 is devoted to presenting studies on the Monte Carlo simulations and real applications. Concluding remarks conclude the paper in Section 4.

**2. Specification test**

Let  $\delta = \text{corr}(u_t, v_t)$ . As argued by Campbell and Yogo (2006),  $\delta$  might not be zero in many applications, so that the so-called *embedded endogeneity* causes the estimation bias for estimating  $\beta_0$  and  $\beta_1$  by running directly the model in (1). To overcome this difficulty, following Cai and Wang (2014), one can use a projection approach of  $u_t$  onto  $v_t$  as  $u_t = \gamma v_t + \eta_t$  and run the following regression with estimated regressor

$$y_t = \beta_0 + \beta_1 x_{t-1} + \gamma \hat{v}_t + \eta_t, \quad 1 \leq t \leq T, \tag{5}$$

to estimate  $\beta_0$  and  $\beta_1$ , where  $\hat{v}_t = x_t - \hat{\rho} x_{t-1}$  from (3), so that one can obtain the residual  $\hat{u}_t$ , from which  $\hat{u}_t = u_t + O_p(T^{-1/2})$  based on Theorem 1 in Cai and Wang (2014) under some mild conditions. Therefore, by letting  $r_t = u_t^2$  and  $\hat{r}_t = \hat{u}_t^2$ , we have

$$\hat{r}_t - r_t = 2 \left[ (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1) x_{t-1} \right] \sigma(x_{t-1}) \epsilon_t + O_p(T^{-1}). \tag{6}$$

It is intuitively clear that the biases of the squared residuals are of order  $O_p(T^{-1})$ , and this is the effect of estimated  $\beta_0$  and  $\beta_1$  in (5) on the estimated variance function  $\sigma^2(x)$ . This result also paves the way for employing a fully data-driven bandwidth procedure in the estimation given in (7) below.

Since  $\sigma^2(x_{t-1}) = E(r_t | x_{t-1})$ , then,  $r_t$  can be written as a mean regression form as follows

$$r_t = \sigma^2(x_{t-1}) + \xi_t,$$

where  $E(\xi_t | x_{t-1}) = 0$ . Thus, under  $H_0$  in (4), a consistent estimate of  $\sigma_0^2$  is

$$\tilde{\sigma}_0^2 = \frac{1}{T-1} \sum_{t=2}^T r_t$$

if  $r_t$  would be known. But,  $r_t$  is unknown in practice, it should be replaced by  $\hat{r}_t$  according to (6). Clearly, in view of (6),

$$\frac{1}{T-1} \sum_{t=2}^T \hat{r}_t - \tilde{\sigma}_0^2 = 2\sigma_0(\beta_0 - \hat{\beta}_0)\bar{\epsilon} + 2\sigma_0(\beta_1 - \hat{\beta}_1)\bar{x}\bar{\epsilon} + O_p(T^{-1}),$$

where  $\bar{\epsilon} = \sum_{t=2}^T \epsilon_t / (T-1)$  and  $\bar{x}\bar{\epsilon} = \sum_{t=2}^T x_{t-1} \epsilon_t / (T-1)$ , which can be easily shown that  $\bar{\epsilon} = o_p(1)$  and  $\bar{x}\bar{\epsilon} = O_p(1)$ . Therefore,

$$\tilde{\sigma}_0^2 = \frac{1}{T-1} \sum_{t=2}^T \hat{r}_t + o_p(1) \approx \frac{1}{T-1} \sum_{t=2}^T \hat{r}_t \equiv \hat{\sigma}_0^2,$$

so that  $\frac{1}{T-1} \sum_{t=2}^T \hat{r}_t$  is used to estimate  $\sigma_0^2$  consistently.

Under the alternative,  $\sigma^2(x_{t-1}) = E(r_t | x_{t-1})$  is a nonparametric function of  $x_{t-1}$ , so that based on nonparametric estimation procedure, one can estimate  $\sigma^2(x)$  for a given  $x$  by the following kernel method (local constant approach) as in Han and Zhang (2012),

$$\tilde{\sigma}^2(x) = \sum_{t=2}^T r_t K_h(x_{t-1} - x) / D_T(x), \tag{7}$$

where  $D_T(x) = \sum_{t=2}^T K_h(x_{t-1} - x)$ ,  $K_h(x) = K(x/h)/h$ ,  $K(x)$  is a kernel function, and  $h$  is the bandwidth satisfying  $h \rightarrow 0$  and  $Th \rightarrow \infty$ . Note that different from the local linear estimation of  $\sigma^2(x)$  for stationary case in Fan and Yao (1998), a local constant estimation is employed here since both have the exact same asymptotic behaviors for nonstationary regressors as argued in Cai (2011). Similar to Theorem 3 in Cai (2011) and Han and Zhang (2012), one can show easily that

$$\tilde{\sigma}^2(x) = \sigma^2(x) + B_T(x) + O_p(T^{-1/5}) \tag{8}$$

where  $B_T(x) = h^2 \mu_2(K) \ddot{\sigma}^2(x) / 2$  with  $\ddot{\sigma}^2(x)$  being the second derivative of  $\sigma^2(x)$  and  $\mu_2 = \int x^2 K(x) dx$ . Based on the above discussions in (6) and (8),  $r_t$  can be replaced by  $\hat{r}_t$ , so that

$$\hat{\sigma}^2(x) = \sum_{t=2}^T \hat{r}_t K_h(x_{t-1} - x) / D_T(x). \tag{9}$$

Similarly, one can show that (8) holds true for  $\hat{\sigma}^2(x)$ , which implies that  $\hat{\sigma}^2(x)$  is a consistent estimate of  $\sigma^2(x)$ .

To test  $H_0$ , similar to the test in Cai et al. (2015), we make use of the following weighted Cramér-von Mises test statistic ( $L_2$ -type) test statistic

$$J_T = \int [(\hat{\sigma}^2(x) - \hat{\sigma}_0^2) D_T(x)]^2 dx,$$

which can be simplified, by removing by the global center, i.e., the sum where  $t = s$  in the above equation, as follows

$$S_T = \sum_{t,s=2}^T \sum_{t \neq s} \hat{\xi}_t \hat{\xi}_s W((x_{t-1} - x_{s-1})/h), \tag{10}$$

where  $\hat{\xi}_t = \hat{r}_t - \hat{\sigma}_0^2$  and  $W(u) = \int K(v)K(u-v)dv$  is the convolution kernel. Note that the reason of employing the weighted  $L_2$ -type test statistic is to get ride of the random denominator, which has been commonly used in the literature.

Note that the test statistic in (10) is similar to the kernel smoothed test statistic, which has commonly been applied to test parametric specifications in the econometrics literature, see, for example, Fan and Li (1996) and Li (1999) for stationary time series mean regression and in Wang and Phillips (2012) for nonstationary regressor. Under some regularity assumptions, which are similar to Assumptions 1-5 in Wang and Phillips (2012), by following the proof in Wang and Phillips (2012), it is not difficult to show that

$$S_T / \sqrt{2 \hat{\Sigma}_T} \rightarrow_D N(0, 1),$$

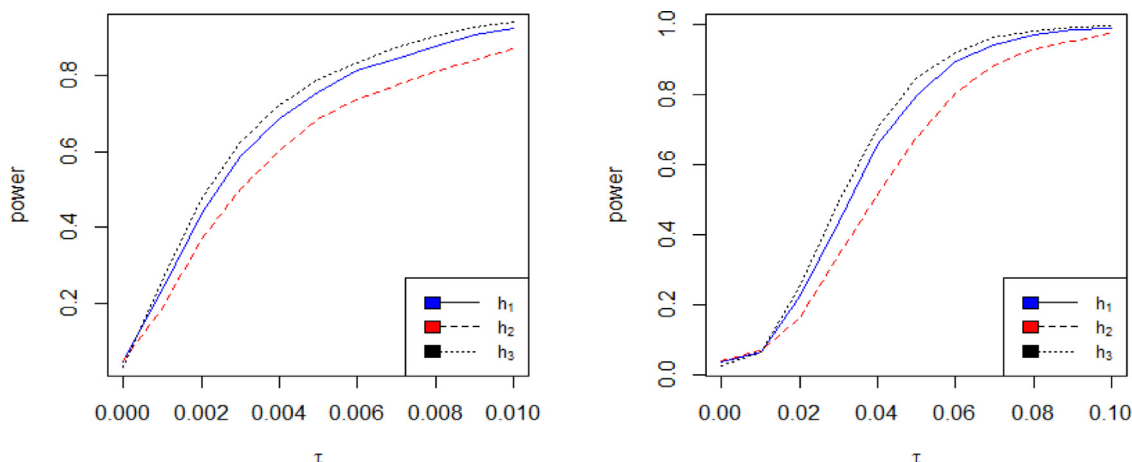
which is similar to (3.10) in Zheng (1996) for iid case, and (3.1) in Wang and Phillips (2012) for nonstationary situations, where

$$\hat{\Sigma}_T = \sum_{s,t=2}^T \sum_{s \neq t} \hat{\xi}_t^2 \hat{\xi}_s^2 W^2((x_t - x_s)/h),$$

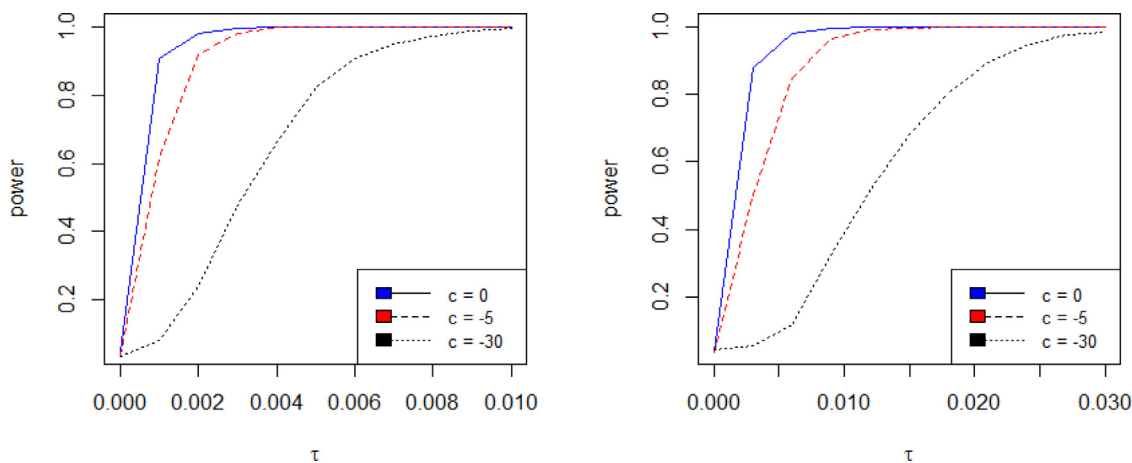
which is the exact same as that for the iid case in Zheng (1996) and for time series context in Fan and Li (1996) and Li (1999) as well as for nonstationary time series in Wang and Phillips (2012). Also, one can show that the test statistic  $S_T$  is a consistent test.

**Table 2**  
Testing results for both log D-P ratio and log E-P ratio.

	log D-P ratio			log E-P ratio		
	$\hat{\delta}$	Test statistics	p-value	$\hat{\delta}$	Test statistics	p-value
S&P 500	-0.8409	5.3865	<0.0001	-0.6589	2.6143	0.0089
NYSE/AMEX	-0.9473	5.6063	<0.0001	-0.6708	2.8815	0.0040



**Fig. 1.** Powers of the alternatives (11) with  $c = 0, \delta = -0.95$  and  $T = 500$  in the left figure and  $c = -30, \delta = -0.95$  and  $T = 500$  in the right figure.



**Fig. 2.** Powers of alternatives (11) with  $\delta = -0.95$  and  $T = 500$  in the left figures and  $\delta = -0.95$  and  $T = 1000$  in the right figure.

**3. Applications**

To investigate the finite sample performance of the proposed test statistics, we conduct a Monte Carlo simulation study with constant and time-varying  $\sigma^2(\cdot)$ . For simplicity, in following computations, the Gaussian kernel is used so that the convolution kernel is an  $N(0, 2)$  density function.

**3.1. Monte Carlo Simulations**

The data generating process from predictive regression model is

$$y_t = \beta_0 + \beta_1 x_{t-1} + \sigma(x_{t-1})\epsilon_t, \quad x_t = \rho x_{t-1} + v_t,$$

where  $\beta_0 = 0.5, \beta_1 = 0.75$  and  $\rho = 1 + c/T$  with  $c$  taking 0, -5, and -30. Clearly, the degree of persistence of  $x_t$  is controlled by nuisance parameter  $c$  and the three levels of persistence correspond to the cases when predictor  $x_t$  is a unit root process, a nearly unit root process and a stationary process, respectively.

The innovations  $\epsilon_t$  and  $v_t$  are jointly generated by standard bivariate normal distribution with correlation coefficient  $\delta$  taking -0.95 and -0.25.

To study the finite sample performance of proposed testing method, we consider the following case:

$$\sigma^2(x_{t-1}) = \sigma_0^2 + \tau x_{t-1}^2, \tag{11}$$

with  $\sigma_0^2 = 1$ . When  $\tau = 0$ , it collapses to the null hypothesis so that the empirical test size can be computed. When  $\tau \neq 0$ , it reduces to a series of alternatives, which delivers the empirical power indexed by  $\tau$ .

The Monte Carlo simulation is considered with  $T = 500$  and  $1000, \delta = -0.95$  and  $-0.25$ , and three values for nuisance parameter  $c$ , and it is repeated 1000 times for each setting. In order to demonstrate the finite sample behavior of the proposed test statistic, we examine the sizes and the powers under three choices of bandwidth  $(h_1, h_2, h_3) = (h, 0.5h, 1.5h)$  with  $h = T^{-1/10}$  for  $c = 0$  and  $-5$ , and  $h = T^{-1/5}$  for  $c = -30$ . The test sizes are reported in Table 1, from which one can observe that the test sizes are close to the nominal size under all settings.

Therefore, one can conclude that the test is not very sensitive to the choice of bandwidth as long as the bandwidth is in the reasonable range and performs very well. Now, we evaluate the power of the proposed test by two typical examples. Indeed, the power is computed by  $\tau$  in (11) in the range  $[0, 0.01]$  with an increment 0.001 when  $c = 0$  and  $\tau \in [0, 0.1]$  with an increment 0.01 when  $c = -30$  (see Fig. 1). The powers vary slightly with bandwidth but they are quite close in general, given such small increment in alternatives. Meanwhile, with smaller choice of  $\tau$  in the series of alternatives when  $c = 0$ , the performance of proposed test statistic is better for nonstationary scenario than for stationary case in terms of test power. Fig. 2 also demonstrates this property. In Fig. 2, we fix the bandwidth  $h = T^{-1/5}$  for all scenarios and compute powers by  $\tau \in [0, 0.1]$  with an increment 0.01 when  $T = 500$  and  $\tau \in [0, 0.03]$  with an increment 0.003 when  $T = 1000$ . We also compute powers with  $\delta = -0.25$ , and the results are almost the same and thus omitted for saving space.

### 3.2. An empirical example

As an illustration of our methodology, we test heteroskedasticity for predictive regression of stock return with dividend-price ratio and earning price ratio as the predictors. Two different series of stock returns are employed, the returns on NYSE/AMEX value-weighted index and S&P 500 index from the Center for Research in Security Prices (CRSP). We use monthly data from November 1926 to December 2019.

The dividend price (D-P) ratio is constructed as the ratio of average dividends during the last year over the current price level, we take the natural log on it in actual predictive regression. For both indexes, the log D-P ratios have been identified as an I(1) process by ADF test and we first consider the following predictive regression

$$r_t = \alpha + \beta x_{t-1} + u_t, \quad (12)$$

where  $r_t$  is the excess return and  $x_t$  represents the associated log D-P ratio. Denoted by  $\hat{v}_t$  the OLS residuals from the AR(1) model,  $x_t = \rho x_{t-1} + v_t$ , and  $\hat{u}_t$  the OLS residuals from model (12). The sample correlation coefficients between  $\hat{u}_t$  and  $\hat{v}_t$  are reported in Table 2, these nonzero values imply the existence of embedded endogeneity ( $x_{t-1}$  and  $u_t$  may be correlated) problem which leads to biased estimates. Following Cai and Wang (2014), we employ the projection approach to deal with endogeneity problem and the regression model is

$$r_t = \beta_0 + \beta_1 x_{t-1} + \gamma \hat{v}_t + \sigma(x_{t-1}) \epsilon_t.$$

Our interest is to test heteroskedasticity; that is, test  $H_0 : \sigma^2(x_{t-1}) = \sigma_0^2$ , where  $\sigma_0^2$  is unknown. The test statistics and p-values of the test are reported in Table 2. With 5% significance level, both null hypotheses are rejected. This means that  $\sigma(x_{t-1})$  depends on  $x_{t-1}$ .

In addition, we also consider the earning price (E-P) ratio as the predictor of stock return and the log E-P ratio is computed based on the cyclically adjusted price earnings ratio (CAPE), which is proposed by Shiller (2000). To be specific, we take the natural log on the inverse of CAPE and apply the same testing methodology as above. The results are reported in Table 2, from which one can conclude that  $\sigma(x_{t-1})$  depends on  $x_{t-1}$  in this application example as well.

## 4. Conclusion

In this paper, to test heteroskedasticity in predictive regression when regressors are nonstationary, a Cramér-von Mises type test statistic is proposed and the limiting distribution is derived. The simulation experiments show that our proposed test works very well even in a small sample size. We also apply our testing method to test heteroskedasticity for predictive regression of stock return with dividend-price ratio and earning price ratio as the predictors. We find the conditional variance  $\sigma^2(x_{t-1})$  is not a constant for both predictors.

Finally, we note several possible extensions of the present study. For example, the proposed test procedure might be applied to testing if the conditional variance  $\sigma^2(x_{t-1})$  might have a specific parametric form due to some particular application. We leave such extensions as possible future research topics.

## References

- Cai, Z., 2011. Nonparametric regression models with integrated covariates. In: Jiang, J., Roussas, G.G., Samaniego, F.J. (Eds.), *Nonparametric Statistical Methods and Related Topics*. In: *A Festschrift in Honor of Professor P.K. Bhattacharya on his 80th Birthday*, pp. 257–275.
- Cai, Z., Wang, Y., 2014. Testing predictive regression models with nonstationary regressors. *J. Econometrics* 178 (1), 4–14.
- Cai, Z., Wang, Y., Wang, Y., 2015. Testing instability in predictive regression model with nonstationary regressors. *Econ. Theory* 31 (5), 953–980.
- Campbell, J.Y., Yogo, M., 2006. Efficient tests of stock return predictability. *J. Financ. Econ.* 81 (1), 27–60.
- Choi, Y., Jacewitz, S., Park, J.Y., 2016. A reexamination of stock return predictability. *J. Econometrics* 192 (1), 168–189.
- Fan, Y., Li, Q., 1996. Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica* 64 (4), 865–890.
- Fan, J., Yao, Q., 1998. Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* 85 (3), 645–660.
- Han, H., Zhang, S., 2012. Non-stationary non-parametric volatility model. *Econom. J.* 15 (2), 204–225.
- Li, Q., 1999. Consistent model specification tests for time series econometric models. *J. Econometrics* 92 (1), 101–147.
- Liao, X., Cai, Z., Chen, H., 2018. A perspective on recent models for testing predictability of asset returns. *Appl. Math. B* 33 (2), 127–144.
- Park, J.Y., 2002. Nonstationary nonlinear heteroskedasticity. *J. Econometrics* 110 (1), 383–415.
- Shiller, R.J., 2000. *Irrational Exuberance*, first ed. Princeton University Press, Princeton, NJ.
- Wang, Q., Phillips, P.C.B., 2012. A specification test for nonlinear nonstationary models. *Annu. Statist.* 40 (2), 727–758.
- Zheng, J.X., 1996. A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* 75 (2), 263–289.