*Econometrica*, Vol. 80, No. 3 (May, 2012), 1157–1183

# TESTING FOR SMOOTH STRUCTURAL CHANGES IN TIME SERIES MODELS VIA NONPARAMETRIC REGRESSION

BY BIN CHEN AND YONGMIAO HONG[1]

Checking parameter stability of econometric models is a long-standing problem. Almost all existing structural change tests in econometrics are designed to detect abrupt breaks. Little attention has been paid to smooth structural changes, which may be more realistic in economics. We propose a consistent test for smooth structural changes as well as abrupt structural breaks with known or unknown change points. The idea is to estimate smooth time-varying parameters by local smoothing and compare the fitted values of the restricted constant parameter model and the unrestricted time-varying parameter model. The test is asymptotically pivotal and does not require prior information about the alternative. A simulation study highlights the merits of the proposed test relative to a variety of popular tests for structural changes. In an application, we strongly reject the stability of univariate and multivariate stock return prediction models in the postwar and post-oil-shocks periods.

KEYWORDS: Kernel, model stability, nonparametric regression, parameter constancy, smooth structural change.

## 1. INTRODUCTION

DETECTION OF STRUCTURAL CHANGES in economic relationships is a long-standing problem in econometrics. However, most existing tests are designed for abrupt structural breaks. As Hansen (2001) pointed out, "it may seem unlikely that a structural break could be immediate and might seem more reasonable to allow a structural change to take a period of time to take effect." Indeed, technological progress, preference change, and policy switch are some leading driving forces of structural changes that usually exhibit evolutionary changes in the long term.

During the past two decades, time-varying time series models have appeared as a novel tool to capture the evolutionary behavior of economic time series. An example is the smooth transition regression (STR) model developed by Lin and Teräsvirta (1994). By the use of a transition function, the STR model

[1]This paper is based on Chen's job market paper titled "Testing for Smooth Structural Changes in Time Series Econometric Models via Local Linear Regression." We are grateful to a coeditor and two anonymous referees for exceptionally helpful comments and suggestions. We also thank Mark Bils, Zongwu Cai, Yongsung Chang, Ron Gallant, George Jakubson, Robert Jarrow, Nick Kiefer, Chung-Ming Kuan, Haitao Li, Alan Stockman, George Tauchen, Chunchi Wu, and seminar participants at Cornell University, Duke University, National University of Singapore, Texas A&M University, University of Pittsburgh, University of Rochester, and the Sino-Korean Econometrics Workshop in Xiamen, China for their comments. All remaining errors are solely ours. Chen acknowledges financial support from the department of economics, University of Rochester. Hong acknowledges financial support from the outstanding oversea youth fund of National Science Foundation of China.

 DOI: 10.3982/ECTA7990

allows both the intercept and the slope to change smoothly over time. Parametric models for time-varying parameters lead to more efficient estimation if the coefficient functions are correctly specified. However, economic theories usually do not suggest any concrete functional form for time-varying parameters; the choice of a functional form is somewhat arbitrary. A nonparametric time-varying parameter model was introduced by Robinson (1989, 1991) and further studied by Orbe, Ferreira, and Rodriguez-Poo (2000, 2005) and Cai (2007). One advantage of this nonparametric model is that little restriction is imposed on the functional forms of the time-varying intercept and slope, except for the condition that they evolve over time smoothly. Motivated by its flexibility, we use this model as the alternative to test smooth structural changes for a linear regression model.

To our knowledge, there are only two tests designed explicitly for smooth structural changes in the literature. Farley, Hinich, and McGuire (1975) constructed an $F$ test by comparing a linear time series model with a parametric alternative whose slope is a linear function of time. Lin and Teräsvirta (1994) developed Lagrange multiplier- (LM) type tests against a STR alternative. These tests use a specific parametric time-varying parameter model. While these tests have the best power against the assumed alternative, no prior information about the true alternative is usually available for practitioners. In such scenarios, it is highly desirable to develop consistent tests that have good power against all-round alternatives of structural changes.

This paper proposes a new consistent Wald-type test for smooth structural changes as well as abrupt structural breaks. The test complements the existing tests for abrupt structural breaks and avoids the difficulty associated with whether there are multiple breaks and whether breakpoints are unknown. We estimate smooth-changing parameters by local linear regression and compare them with the ordinary least squares (OLS) parameter estimator. The proposed Wald-type test can be viewed as a generalization of Hausman's (1978) test from the parametric framework to the nonparametric framework. A generalized Chow (1960) $F$-type test could also be constructed by comparing the sums of squared residuals (SSRs) between the restricted constant parameter model and the unrestricted time-varying parameter model (see Chen and Hong (2008) for details). Interestingly, unlike Chow's (1960) test, which is optimal in the context of the classical linear regression model with independent and identically distributed (i.i.d.) normal errors, the generalized Chow test is no longer optimal. We show that the generalized Hausman test is asymptotically more powerful than the generalized Chow test. For this reason, this paper focuses on the generalized Hausman test.

Compared with the existing tests for structural breaks in the literature, the proposed test has a number of appealing features. First, it is consistent against a large class of smooth time-varying parameter alternatives as well as multiple sudden structural breaks with unknown breakpoints. Second, no prior information on a structural change alternative is needed. In particular, we do not need to know whether the structural changes are smooth or abrupt, and in the cases of abrupt structural breaks, we do not need to know the dates or the number of breaks. Third, different from many tests for structural breaks in the literature, our test is asymptotically pivotal. The only inputs required are the OLS and local linear time-varying parameter estimators. The latter is a locally weighted least squares estimator. Hence, any standard econometric software can be used to implement the test. Fourth, because only local information is employed in estimating parameters at each time point, our test has symmetric power against structural breaks that occur either in the first or second half of the sample period. In contrast, some existing tests (e.g., Brown, Durbin, and Evans' (1975) cumulative sum (CUSUM) test) have different powers against structural breaks that have the same magnitudes but occur at different time points. Fifth, unlike Andrews' (1993) supremum test and Bai and Perron's (1998) double maximum test, no trimming of the boundary region near the end points of the sample period is needed for our test. Finally, as a by-product, the nonparametric local linear estimators of the time-varying parameters can provide insight into the stability of the economic relationship.

In Section 2, we introduce the framework and state the hypotheses of interest. Section 3 describes our approach and the test statistic. Section 4 derives the asymptotic null distribution and Section 5 investigates the asymptotic power. In Section 6, a simulation study examines the reliability of the asymptotic theory in finite samples. Section 7 applies our test to stock return predictability models and documents strong evidence against model stability. Section 8 concludes. All mathematical proofs are collected in the Supplemental Material (Chen and Hong (2012)).

## 2. HYPOTHESES OF INTEREST

Consider the data generating process (DGP)

$$Y_t = \mathbf{X}_t'\alpha_t + \varepsilon_t, \quad t = 1, \dots, T, \tag{2.1}$$

where $Y_t$ is a dependent variable, $\mathbf{X}_t$ is a $d \times 1$ vector of explanatory variables, $\alpha_t$ is a $d \times 1$ possibly time-varying parameter vector, $\varepsilon_t$ is an unobservable disturbance with $E(\varepsilon_t|\mathbf{X}_t) = 0$ almost surely (a.s.), $d$ is a fixed positive integer, and $T$ is the sample size. The regressor vector $\mathbf{X}_t$ can contain exogenous explanatory variables and lagged dependent variables. Thus, both static and dynamic regression models are covered.

Like the bulk of the literature on structure changes, we are interested in testing the constancy of the regression parameter in (2.1). The null hypothesis of interest is

$$\mathbb{H}_0 : \alpha_t = \alpha \text{ for some constant vector } \alpha \in \mathbb{R}^d \text{ and for all } t.$$

The alternative hypothesis $\mathbb{H}_A$ is that $\mathbb{H}_0$ is false. Under $\mathbb{H}_0$, the unknown constant parameter vector $\alpha$ can be consistently estimated by (e.g.) the OLS estimator

$$\hat{\alpha} = \arg\min_{\alpha \in \mathbb{R}^d} \sum_{t=1}^{T} (Y_t - \mathbf{X}_t' \alpha)^2. \tag{2.2}$$

Under the alternative $\mathbb{H}_A$, $\alpha_t$ is a time-varying parameter vector. Examples include Chow's (1960) single break model, Hall and Hart's (1990) deterministic trend model, and Lin and Teräsvirta's (1994) STR model. Tests for parametric structural change alternatives (e.g., Lin and Teräsvirta's (1994) LM tests) have the best power against the assumed alternative. Unfortunately, usually no prior information about the structural change alternative is available in practice. To cover a wide range of alternatives, we consider the smooth time-varying parameter model

$$Y_t = \mathbf{X}_t' \alpha(t/T) + \varepsilon_t, \quad t = 1, \ldots, T, \tag{2.3}$$

where $\alpha : [0, 1] \to \mathbb{R}^d$ is an unknown smooth function except for a finite number of points on [0, 1]. Discontinuities of $\alpha(\cdot)$ at a finite number of points in [0, 1] allow abrupt changes.

This model was introduced by Robinson (1989, 1991) and its nonparametric estimation was further considered in Robinson (1989, 1991), Orbe, Ferreira, and Rodriguez-Póo (2000, 2005) and Cai (2007).[2] It avoids restrictive parameterization of $\alpha(\cdot)$. The specification that $\alpha(\cdot)$ is a function of ratio $t/T$ rather than time $t$ only is a common scaling scheme in the literature (e.g., Phillips and Hansen (1990)). The reason for this requirement is that a nonparametric estimator for $\alpha_t$ is not consistent unless the amount of data on which it depends increases, and merely increasing the sample size does not necessarily improve estimation of $\alpha_t$ at some fixed point $t$, even if some smoothness condition is imposed on $\alpha_t$. The amount of local information must increase suitably if the variance and the bias of a nonparametric estimator of $\alpha_t$ are to decrease. A convenient way to achieve this is to regard $\alpha_t$ as ordinates of smooth function $\alpha(\cdot)$ on an equally spaced grid over [0, 1], which becomes finer as $T \to \infty$, and to consider estimation of $\alpha(u)$ at fixed points $u \in [0, 1]$. Consistent estimation of model (2.3) was considered in Robinson (1991) and Cai (2007) using local constant and local linear smoothing, respectively.

The specification of $\alpha_t = \alpha(t/T)$ does not regard the sampling of $(Y_t, \mathbf{X}_t')'$ as taking place on a grid on [0, 1], which would make the preservation of independence or weak dependence properties as $T$ increases implausible. We note

[2]These authors considered pointwise consistent nonparametric estimation of time-varying parameters $\alpha(t/T)$ and $\sigma^2(t/T)$, where $\text{var}(\varepsilon_t) = \sigma^2(t/T)$. They did not consider testing parameter constancy.

that the device of taking $(Y_t, \mathbf{X}_t')'$ to be observations at intervals $1/T$ on a continuous process on [0, 1] that itself is independent of $T$ does not work because it does not achieve the accumulation of new information as $T$ increases, which is needed for consistency. Making parameter $\alpha_t$ depend on $T$ is common in econometrics. A well known example is local power analysis, where local alternatives are specified as a function of $T$.

Model (2.3) includes the locally stationary autoregressive model in Dahlhaus (1996),

$$Y_t = \alpha_0(t/T) + \sum_{j=1}^{p} \alpha_j(t/T) Y_{t-j} + \varepsilon_t,$$

where $\varepsilon_t = \sigma(t/T)v_t$ and $v_t \sim$ i.i.d. $N(0, 1)$. A locally stationary process is a nonstationary time series whose behavior can be locally approximated by a stationary process. In time series analysis, it is often assumed that nonstationary economic time series can be transformed—by removing time trends and/or taking differences—into a stationary process. In fact, the transformed series may still not be stationary, even after trending components are removed. Locally stationary time series models nicely fill this gap and provide new insight into modelling economic time series.

We assume that $\alpha(\cdot)$ is continuous except for a finite number of points on $[0, 1]$. In other words, we permit $\alpha(\cdot)$ to have finitely many discontinuities. Hence, a single structural break or multiple breaks with known or unknown breakpoints, as often considered in this literature, are special cases of model (2.3). For example, suppose $\alpha(u) = \alpha_0$ if $u \leq u_0$, and $\alpha(u) = \alpha_1$ otherwise. Then we obtain the single break model originally considered in Chow (1960).

## 3. NONPARAMETRIC TESTING

We now propose a consistent test for smooth structural changes. Recall that under $\mathbb{H}_0$, we have a constant parameter regression model $Y_t = \mathbf{X}_t'\alpha + \varepsilon_t$, where $\alpha$ can be consistently estimated by the OLS estimator $\hat{\alpha}$ in (2.2). Under the alternative $\mathbb{H}_A$, $\alpha_t = \alpha(t/T)$ is changing over time. The OLS estimator $\hat{\alpha}$ is no longer suitable because there exists no parameter $\alpha$ such that $E(Y_t|\mathbf{X}_t) = \mathbf{X}_t'\alpha$ a.s. under $\mathbb{H}_A$. However, a nonparametric estimator can consistently estimate the time-varying parameter $\alpha_t$.

Various nonparametric methods could be used to estimate $\alpha_t$. Robinson (1991) and Cai (2007) studied the pointwise consistency and asymptotic normality of the kernel and local linear estimators, respectively. Here, we use local linear smoothing, which includes the kernel method as a special case. Cai (2007) showed that although the kernel and the local linear estimators share the same asymptotic properties at the interior points, the latter converge faster than the former in the boundary regions near the end points of the sample period. The use of local linear smoothing is quite suitable here.[3] In particular, structural changes near the boundary regions are notoriously difficult to detect, as shown by many previous works in the literature (e.g., Chu, Hornik, and Kuan (1995)). Unlike many existing tests, no trimming is needed for the local linear smoother, which can estimate structural changes near boundary regions for sufficiently large samples. Thus, it is expected to give better power in such cases.

Put $\mathbf{Z}_{st} = (1, \frac{s-t}{T})'$ and $k_{st} = k(\frac{s-t}{Th})$, where the kernel $k(\cdot) : [-1, 1] \to \mathbb{R}^+$ is a prespecified symmetric probability density, and $h \equiv h(T)$ is a bandwidth with $h \to 0$ and $Th \to \infty$ as $T \to \infty$. For notational simplicity, we have suppressed the dependence of $\mathbf{Z}_{st}$ and $k_{st}$ on $T$ and $h$. Examples of $k(\cdot)$ include the uniform, Epanechnikov, and quartic kernels.

We note that although local linear smoothing can enhance the convergence rate of the asymptotic bias in the boundary regions $[1, Th] \cup [T - Th, T]$ from $h$ to $h^2$, the scale differs from that of an interior point. As shown in Cai (2007), the asymptotic bias at an interior point is proportional to $h^2 \int_{-1}^{1} u^2 k(u)\, du$ while that at a boundary point is proportional to $h^2 b(c)$, where $b(c) = (\mu_{2c}^2 - \mu_{1c}\mu_{3c})/(\mu_{0c}\mu_{2c} - \mu_{1c}^2)$, $\mu_{ic} = \int_{-c}^{1} u^i k(u)\, du$, $i = 0, 1, 2, 3$, and $c \in [0, 1]$. Although the convergence rate is the same as in the interior region, the asymptotic variance at a boundary point tends to be larger, because fewer observations are available to the estimators in the boundary regions. These differences would complicate the form of our test statistics. Moreover, the observations contained in the boundary regions $[1, Th] \cup [T - Th, T]$ are not trivial. For example, if $h = (1/\sqrt{12})T^{-1/5}$, about 23%, 17%, and 10% of the observations fall into the boundary regions when the sample size $T = 100, 500$, and 5000, respectively.

To make the behavior of the local linear estimator at boundary points similar to that at interior points, we follow Hall and Wehrly (1991) to reflect the data in the boundary regions, obtaining pseudodata $(Y_t, \mathbf{X}_t') = (Y_{-t}, \mathbf{X}_{-t}')$ for $-\lfloor Th \rfloor \le t \le -2$, where $\lfloor Th \rfloor$ denotes the integer part of $Th$ and $(Y_t, \mathbf{X}_t') = (Y_{2T-t}, \mathbf{X}_{2T-t}')$ for $T + 1 \le t \le T + \lfloor Th \rfloor$. We use the synthesized data (the union of the original data and the pseudodata) to estimate $\alpha_t$. By construction, symmetric data points are available in the boundary regions $[1, Th] \cup [T - Th, T]$. In addition, in nonparametric regression estimation, this method has also been described as "reflection about the boundaries" by Cline and Hart (1991) with regard to nonparametric density estimation. It has not been used to estimate time-varying coefficients in the previous literature.

[3]Both local linear smoothing and the conventional kernel method are local smoothing. Global smoothing (e.g., series approximation) is another class of nonparametric method. The coefficient function $\alpha(\cdot)$ may not have a nice shape and many terms are needed when using a serial approximation, which complicates the estimation. On the other hand, structural change is the local behavior of parameters and hence local smoothing is expected to have better approximation in many cases.

The local linear parameter estimator is obtained by minimizing the local SSRs,

$$(3.1)\qquad \min_{\beta\in\mathbb{R}^{2d}} \sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor} k_{st}\left[Y_s-\alpha_0'\mathbf{X}_s-\alpha_1'\left(\frac{s-t}{T}\right)\mathbf{X}_s\right]^2 = \sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor} k_{st}(Y_s-\beta'\mathbf{Q}_{st})^2,$$

where $\beta=(\alpha_0',\alpha_1')'$ is a $2d\times 1$ vector, $\alpha_j$ is a $d\times 1$ coefficient vector for $(\frac{s-t}{T})^j\mathbf{X}_s$, $j=0,1$, $\mathbf{Q}_{st}=\mathbf{Z}_{st}\otimes\mathbf{X}_s$ is a $2d\times 1$ vector, and $\otimes$ is the Kronecker product. Note that the device of using pseudodata does not affect the estimation at interior points $[Th, T-Th]$. By solving the optimization problem in (3.1), we obtain the solution

$$(3.2)\qquad \hat{\beta}_t=\left(\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor} k_{st}\mathbf{Q}_{st}\mathbf{Q}_{st}'\right)^{-1}\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor} k_{st}\mathbf{Q}_{st}Y_s,\quad t=1,\ldots,T.$$

This is a locally weighted least squares estimator. As pointed out by Cai (2007), the local linear estimator could be regarded as the OLS estimator of the transformed model

$$k_{st}^{1/2}Y_s=k_{st}^{1/2}\mathbf{X}_s'\alpha_0+k_{st}^{1/2}\left(\frac{s-t}{T}\right)\mathbf{X}_s'\alpha_1+\varepsilon_s,\quad s=1,\ldots,T.$$

Hence the estimation can be implemented by standard econometric software.

Put $\mathbf{e}_1=(1,0)'$. Then the local linear estimator for $\alpha_t$ is given by

$$(3.3)\qquad \hat{\alpha}_t=(\mathbf{e}_1'\otimes\mathbf{I}_d)\hat{\beta}_t,\quad t=1,\ldots,T.$$

Note that with the reflection method for the boundary regions, one can also use the kernel method, which is equivalent to a local linear estimation with the restriction $\alpha_1=\mathbf{0}$. The test statistic has the same asymptotic distribution for both local constant and linear estimators.

With $\hat{\alpha}_t$, we can construct a Wald-type test by comparing the OLS and nonparametric regression estimators. This can be interpreted as a generalized Hausman (1978) test. Hausman's (1978) test compares two parameter estimators, where one is efficient but inconsistent under the alternative and the other is inefficient but consistent under the alternative. We extend Hausman's (1978) idea from a parametric regression to a nonparametric regression, where the OLS regression estimator $\mathbf{X}_t'\hat{\alpha}$ can be viewed as an efficient estimator for $E(Y_t|\mathbf{X}_t)$ under $\mathbb{H}_0$, and the nonparametric time-varying parameter regression estimator $\mathbf{X}_t'\hat{\alpha}_t$ can be viewed as an inefficient but consistent estimator for $E(Y_t|\mathbf{X}_t)$ under $\mathbb{H}_A$. We compare these parametric and nonparametric fitted values via a sample quadratic form,

$$\hat{Q}=\frac{1}{T}\sum_{t=1}^{T}(\mathbf{X}_t'\hat{\alpha}_t-\mathbf{X}_t'\hat{\alpha})^2.$$

The statistic $\hat{Q}$ converges to 0 under $\mathbb{H}_0$, but to a strictly positive constant under $\mathbb{H}_A$, giving our one-sided test asymptotic unit power. Any significant departure of $\hat{Q}$ from 0 is evidence of structural changes.[4] Formally, our generalized Hausman test is a standardized version of $\hat{Q}$,

$$\hat{H} = (T\sqrt{h}\hat{Q} - \hat{A}_H)/\sqrt{\hat{B}_H}, \tag{3.4}$$

where $\hat{A}_H = h^{-1/2}C_A \operatorname{trace}(\hat{\Omega}\hat{M}^{-1})$, $\hat{B}_H = 4C_B \operatorname{trace}(\hat{M}^{-1}\hat{\Omega}\hat{M}^{-1}\hat{\Omega})$, $C_A = T^{-1}h^{-1}\sum_{j=-\lfloor Th\rfloor}^{\lfloor Th\rfloor}(1-\frac{|j|}{T})\, k(\frac{j}{Th})[k(\frac{j}{Th}) + h\int_{-1}^{1} k(\frac{j}{Th}+2u)\,du] = \int_{-1}^{1} k^2(u)\,du + o(1)$, $C_B = T^{-1}h^{-1}\sum_{j=1}^{T-1}(1-\frac{j}{T})[\int_{-1}^{1} k(u)k(u+\frac{j}{Th})\,du]^2 = \int_0^1[\int_{-1}^{1} k(u)k(u+v)\,du]^2\,dv + o(1)$, $\hat{\mathbf{M}} = T^{-1}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{X}_t'$, and $\hat{\Omega} = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2\mathbf{X}_t\mathbf{X}_t'$. Note that $C_A$ and $C_B$ are independent of the random sample $\{Y_t, \mathbf{X}_t'\}_{t=1}^{T}$. The factors $\hat{A}_H$ and $\hat{B}_H$ are approximately the mean and the variance of $T\sqrt{h}\hat{Q}$. They take into account the impact of conditional heteroscedasticity and higher order serial dependence in $\{\varepsilon_t\}$. Consequently, the test statistic is robust to conditional heteroscedasticity and time-varying higher order conditional moments of unknown form. If $\varepsilon_t$ is conditionally homoscedastic, then $\hat{A}_H$ and $\hat{B}_H$ can be simplified to $h^{-1/2}dC_A\hat{\sigma}^2$ and $4dC_B\hat{\sigma}^4$, respectively, where $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}(Y_t - \mathbf{X}_t'\hat{\alpha}_t)^2$ is the estimated variance.

## 4. ASYMPTOTIC DISTRIBUTION

To derive the asymptotic distribution of $\hat{H}$, we impose the following regularity conditions.

ASSUMPTION 1: *$\{\mathbf{X}_t', \varepsilon_t\}'$ is a $(d+1)\times 1$ stationary $\beta$-mixing process with mixing coefficients $\{\beta(j)\}$ satisfying $\sum_{j=1}^{\infty} j^2\beta(j)^{\delta/(1+\delta)} < C$ for some $0<\delta<1$.*

ASSUMPTION 2: *$\{\varepsilon_t\}$ is a martingale difference sequence (m.d.s.) such that $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ and $E(\varepsilon_t^2) = \sigma^2$, where $\mathcal{F}_{t-1} = \{\mathbf{X}_t', \mathbf{X}_{t-1}', \ldots, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$.*

ASSUMPTION 3: (i) *The $d\times d$ matrix $\mathbf{M} = E(\mathbf{X}_t\mathbf{X}_t')$ is finite and positive definite*; (ii) $E(\mathbf{X}_{ti}^8) < \infty$ *for $i = 1,\ldots,d$*; (iii) $E(Y_t^8) < \infty$.

ASSUMPTION 4: *$\hat{\alpha}$ is a parameter estimator such that $\sqrt{T}(\hat{\alpha} - \alpha^*) = O_P(1)$, where $\alpha^* = p\lim_{T\to\infty}\hat{\alpha}$ and $\alpha^* = \alpha$ under $\mathbb{H}_0$, where $\alpha$ is given in $\mathbb{H}_0$.*

[4]Alternatively, we could compare $\hat{\alpha}_t$ and $\hat{\alpha}$ directly, and the asymptotic derivation is similar. However, multiplying the coefficients by $\mathbf{X}_t$ gives a comparison between fitted values of the restricted and unrestriced models, and makes our test asymptotically pivotal.

ASSUMPTION 5: $k : [-1, 1] \to \mathbb{R}^+$ *is a symmetric bounded probability density function*.

ASSUMPTION 6: *The bandwidth* $h = cT^{-\lambda}$ *for* $0 < \lambda < 1$ *and* $0 < c < \infty$.

The $\beta$-mixing condition in Assumption 1 imposes a restriction on the temporal dependence in $\{\mathbf{X}_t', \varepsilon_t\}$. Assumption 2 allows dynamic regression models when $\mathbf{X}_t$ contains both exogenous and lagged dependent variables, and conditional heteroscedasticity of unknown form.[5] We note that our m.d.s. assumption is weaker than Lin and Teräsvirta's (1994), who assumed that $\{\varepsilon_t\}$ is an m.d.s. with $\lim_{t\to\infty} E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$. Assumption 2 requires that the linear regression model be correctly specified under $\mathbb{H}_0$: violation of correct model specification may lead to spurious rejection of model stability. Assumption 3 imposes moment conditions on $\mathbf{X}_t$ and $Y_t$, commonly assumed in the regression literature. It could be relaxed to allow time-varying moments (i.e., $\mathbf{M}(t/T) = E(\mathbf{X}_t\mathbf{X}_t')$ is a function of standardized time $t/T$) at the cost of more tedious proof and test statistics. Assumption 4 holds for any $\sqrt{T}$-consistent estimator for $\alpha$ under $\mathbb{H}_0$. We allow but are not restricted to the OLS estimator $\hat{\alpha}$ in (2.2).

Assumption 5 implies $\int_{-1}^{1} k(u)\,du = 1$, $\int_{-1}^{1} uk(u)\,du = 0$, and $\int_{-1}^{1} u^2 \times k(u)\,du < \infty$. All examples noted in Section 3 satisfy this assumption. Assumption 6 implies $h \to 0$ and $Th \to \infty$. This is the standard condition for the bandwidth and covers the optimal bandwidth $h \propto T^{-1/5}$ that minimizes the integrated mean squared error (MSE) of the nonparametric estimation for $\alpha(u)$, $u \in [0, 1]$. In practice, $h$ can be chosen via a simple rule-of-thumb approach, namely $h = (1/\sqrt{12})T^{-1/5}$, where $1/\sqrt{12}$ is the standard deviation of $U(0, 1)$, which could be viewed as the limiting distribution of the grid points $\{\frac{t}{T},\ t = 1, \ldots, T\}$ as $T \to \infty$. Alternatively, as suggested by Robinson (1989), an automatic method such as cross-validation (CV) may be used. Define a "leave-one-out" estimator $\hat{\alpha}_{-t} = (\mathbf{e}_1' \otimes \mathbf{I}_d)\hat{\beta}_{-t}$, where $\hat{\beta}_{-t} = (\sum_{s=t-\lfloor Th\rfloor, s\neq t}^{t+\lfloor Th\rfloor} k_{st}\mathbf{Q}_{st}\mathbf{Q}_{st}')^{-1} \sum_{s=t-\lfloor Th\rfloor, s\neq t}^{t+\lfloor Th\rfloor} k_{st}\mathbf{Q}_{st}Y_s$. Then a data-driven choice of $h$ is $\hat{h}_{\text{CV}} = \arg\min_{c_1T^{-1/5}\leq h\leq c_2T^{-1/5}} \text{CV}(h)$, where $\text{CV}(h) = \sum_{t=1}^{T}(Y_t - \mathbf{X}_t'\hat{\alpha}_{-t})^2$, and $c_1$ and $c_2$ are two prespecified constants. We conjecture that under $\mathbb{H}_0$, $\hat{h}_{\text{CV}}$ may approach the upper bound $c_2T^{-1/5}$ with probability approaching 1, and under $\mathbb{H}_A$, it minimizes the integrated MSE of $\hat{\alpha}_t$ asymptotically.[6] CV does not affect

[5]Assumption 2 rules out linear regression models with endogeneity. For such cases, we could compare a two-stage least square (2SLS) estimator and a local 2SLS estimator, and construct a test statistic accordingly. This is left for future research.

[6]The theoretical property of $\hat{\alpha}_t$ is an open issue to be investigated, but this is beyond the scope of this paper. Wong (1983) showed consistency of CV for estimating a time-varying intercept. Fryzlewicz, Sapatinas, and Subba Rao (2008) also applied a CV method in the estimation of a time-varying autoregressive conditional heteroscedasticity (ARCH) model.

the limiting distribution of our test statistic, as long as $\hat{h}_{\text{CV}}/h \to 1$ sufficiently fast,[7] but it may affect the size of the test in finite samples because the use of CV induces additional sampling noise. The main objective of using CV is to increase power. We note that the CV-based bandwidth does not maximize power, but we expect reasonable and robust power in finite samples from using CV, especially when compared to the use of rule-of-thumb or other ad hoc bandwidth selection methods.[8] We investigate this method in the simulation study below.

We now state the asymptotic distribution of $\hat{H}$ under $\mathbb{H}_0$.

THEOREM 1: *Suppose Assumptions 1–6 and $\mathbb{H}_0$ hold.* (i) *Then $\hat{H} \xrightarrow{d} N(0,1)$ as $T \to \infty$.* (ii) *Suppose in addition $\operatorname{var}(\varepsilon_t|\mathbf{X}_t) = \sigma^2$ a.s. Then $\hat{A}_H = h^{-1/2} d C_A \hat{\sigma}^2$ and $\hat{B}_H = 4 d C_B \hat{\sigma}^4$ in the definition of $\hat{H}$.*

As an important feature of $\hat{H}$, the use of the restricted parametric estimator $\hat{\alpha}$ in place of the regression parameter $\alpha$ under $\mathbb{H}_0$ has no impact on the limit distribution of $\hat{H}$. Intuitively, $\hat{\alpha}$ converges to $\alpha$ faster than the nonparametric estimator $\hat{\alpha}_t$. Consequently, the asymptotic distribution of $\hat{H}$ is solely determined by the nonparametric estimator $\hat{\alpha}_t$. In small samples, the distribution of $\hat{H}$ may not be well approximated by $N(0,1)$. Accurate finite sample critical values can be obtained via bootstrap; see Section 6 for more discussion.

## 5. ASYMPTOTIC POWER

To study the asymptotic power of $\hat{H}$ under $\mathbb{H}_A$, we impose the following assumption:

ASSUMPTION 7: *The coefficient function $\alpha : [0,1] \to \mathbb{R}^d$ is continuous except for a finite number of discontinuity points on $[0,1]$ and $\sup_{v \in (0,1)} \| \lim_{u \to v^+} \alpha(u) - \lim_{u \to v^-} \alpha(u) \| \leq C$.*

This allows both smooth structural changes and abrupt structural breaks with known or unknown breakpoints. For abrupt structural breaks, the break size is bounded.

[7] Using an argument similar to Chen and Hong (2010, Theorem 2), we can show that $\hat{H}_{\hat{h}_{\text{CV}}}$ also has an asymptotic $N(0,1)$ distribution, where $\hat{H}_{\hat{h}_{\text{CV}}}$ is computed in the same way as $\hat{H}$, with $\hat{h}_{\text{CV}}$ replacing $h$.

[8] Following Sun, Phillips, and Jin (2008), we may consider a data-driven bandwidth selection by minimizing a weighted average loss function, namely $\hat{h} = \arg\min_h \frac{w_T}{1+w_T} e_T^{\text{I}} + \frac{1}{1+w_T} e_T^{\text{II}}$, where $w_T$ is a weighting function, and $e_T^{\text{I}}$ and $e_T^{\text{II}}$ are the type I and type II errors of $\hat{H}$, respectively. However, how to find the analytical expressions for the leading terms of $e_T^{\text{I}}$ and $e_T^{\text{II}}$ is rather complicated and hence will be pursued in a subsequent study.

THEOREM 2: *Suppose Assumptions 1–7 hold. Then for any sequence of nonstochastic constants* $\{C_T = o(T\sqrt{h})\}$, $P(\hat{H} > C_T) \to 1$ *under* $\mathbb{H}_A$ *as* $T \to \infty$.

Our test statistic is based on a sample quadratic form, which is nonnegative. Hence, negative values of $\hat{H}$ can only occur under $\mathbb{H}_0$ with probability approaching 1. Therefore, our generalized Hausman test is a one-sided test. Theorem 2 implies that $\hat{H}$ is consistent against all alternatives to $\mathbb{H}_0$ at any given significance level, subject to Assumption 7. Thus, for $T$ sufficiently large, $\hat{H}$ can detect any structural changes, including those that occur close to the starting and ending points of the sample period, because no trimming has to be used. This is rather appealing because no prior information about the alternative is available in practice. It avoids the blindness of searching for possible alternatives of structural changes. We note that for (and merely for) simplicity, stationarity for $\mathbf{X}_t$ is assumed under $\mathbb{H}_A$. One could allow $\mathbf{X}_t$ to be a locally stationary process.

To gain more insight into the power property of $\hat{H}$, we consider two classes of local alternatives.

CASE 1—Local Smooth Structural Change: We have

$$\mathbb{H}_{1A}(j_T): \alpha(u) = \alpha + j_T g(u), \quad u \in [0, 1],$$

where $g : [0, 1] \to \mathbb{R}^d$ is a twice continuously differentiable vector function. The term $j_T g(u)$ characterizes the departure of the smooth-changing coefficient $\alpha(u)$ from the constant $\alpha$ at each point $u \in [0, 1]$ and $j_T$ is the speed at which the departure vanishes to 0 as $T \to \infty$. For notational simplicity, we have suppressed the dependence of $\alpha(u)$ on $T$.

CASE 2—Local Sharp Structural Change at Some Point $u_0$: We have

$$\mathbb{H}_{2A}(b_T, r_T): \alpha(u) = \alpha + b_T f[(u - u_0)/r_T], \quad u \in [0, 1],$$

where $u_0$ is a given point in $[0, 1]$, $f : \mathbb{R} \to \mathbb{R}^d$ is a twice continuously differentiable vector function with $\sup_{z\in\mathbb{R}} \|f(z)\| \le C$ and $\sup_{z\in\mathbb{R}} \|d^2 f(z)/dz^2\| \le C$, $b_T = b(T) \to 0$, and $r_T = r(T) \to 0$ as $T \to \infty$. This type of alternatives was first studied by Rosenblatt (1975) in a different context. Under $\mathbb{H}_{2A}(b_T, r_T)$, the coefficient function $\alpha(u)$ becomes a nonsmooth spike at location $u_0$ (i.e., a temporary change around $u_0$) as $T \to \infty$, due to the existence of the shrinking width parameter $r_T$. Here, $r_T$ controls the sharpness of the structural change around $u_0$, and $b_T$ is the speed at which the departure of $\alpha(u)$ from $\alpha$ at each point $u \in [0, 1]$ vanishes to 0 as $T \to \infty$. For concreteness, we use OLS estimation under $\mathbb{H}_0$ in Theorem 3 below.

THEOREM 3: *Suppose Assumptions* 1–6 *hold and let* $\hat{\alpha}$ *be the OLS estimator*. (i) *Under* $\mathbb{H}_{1A}(j_T)$ *with* $j_T = T^{-1/2}h^{-1/4}$, $\hat{H} \xrightarrow{d} N(\delta_1, 1)$ *as* $T \to \infty$, *where* $\delta_1 = [\int_0^1 g(u)'\mathbf{M}g(u)\,du - \int_0^1 g(u)'\,du\,\mathbf{M}\int_0^1 g(u)\,du]/\sqrt{B_H}$, $B_H = 4\,\mathrm{trace}(\mathbf{M}^{-1}\Omega\mathbf{M}^{-1}\Omega)\int_0^1[\int_{-1}^1 k(u)k(u+v)\,du]^2\,dv$, $\mathbf{M} = E(\mathbf{X}_t\mathbf{X}_t')$ *and* $\Omega = E(\varepsilon_t^2\mathbf{X}_t\mathbf{X}_t')$. (ii) *Under* $\mathbb{H}_{2A}(b_T, r_T)$ *with* $b_T \to 0$, $r_T \to 0$, $b_T^2 r_T = T^{-1}h^{-1/2}$, *and* $h = o(r_T)$, $\hat{H} \xrightarrow{d} N(\delta_2, 1)$ *as* $T \to \infty$, *where* $\delta_2 = [\int_{-\infty}^{\infty} f(z)'\mathbf{M}f(z)\,dz]/\sqrt{B_H}$.[9]

Our test has nontrivial power against the class of smooth alternatives $\mathbb{H}_{1A}(j_T)$ with rate $j_T = T^{-1/2}h^{-1/4}$, which is slightly slower than the parametric rate $T^{-1/2}$ as $h \to 0$.[10] For example, $j_T = T^{-9/20}$ if $h = T^{-1/5}$. In contrast, popular tests such as Andrews' (1993) supremum and Bai and Perron's (1998) double maximum tests have nontrivial power against $\mathbb{H}_{1A}(j_T)$ with $j_T = T^{-1/2}$ and with nonconstant $g(u)$ on $\Pi$ (e.g., $\Pi = [0.15, 0.85]$), where $\Pi$ is a strict subset of [0, 1]. Thus, these tests could be more powerful than ours against the smooth alternatives $\mathbb{H}_{1A}(j_T)$ when $g(u)$ is not a constant function on $\Pi$. However, these tests are not consistent against $\mathbb{H}_{1A}(j_T)$ with $j_T = T^{-1/2}$ for many $g(u)$ functions which are constant on $\Pi$ but not constant outside $\Pi$, as can be verified.[11]

Moreover, our test can have better power than the aforementioned tests against the class of temporary sharp alternatives $\mathbb{H}_{2A}(b_T, r_T)$ for suitable sequences of $b_T$ and $r_T$. For example, suppose $r_T = T^{-1/5}(\ln\ln T)^{\epsilon}$ and $b_T = T^{-7/20}(\ln\ln T)^{-\epsilon/2}$ for small $\epsilon > 0$, and we choose $h = T^{-1/5}$. Since $b_T r_T = o(T^{-1/2})$, it can be shown that the noncentrality parameters of Andrews' (1993) and Bai and Perron's (1998) tests converge to 0 as $T \to \infty$, and so they have no power.[12] Intuitively, the local sharp structural changes are rather spiky, which

[9]It may first seem odd that the noncentrality parameter $\delta_2$ does not depend on the given location point $u_0 \in [0, 1]$. This is due to the fact that $\mathbf{M}$ is not time-varying. If we allow for the time-varying second moment of $\mathbf{X}_t$, $\delta_2$ would depend on $u_0$.

[10]We note that no "curse of dimensionality" problem exists here, as the nonparametric regression is implemented with respect to the scalar $t/T$. Unlike hypotheses concerning the regression function at a single point, our test is a global measure, namely $\alpha_t = \alpha$ for all $t$, rather than at some point only. Specifically, our generalized Hausman test is based on an integrated $L_2$ norm. The local rate $T^{-1/2}h^{-1/4}$ has been studied in the literature with univariate nonparametric regression (see, e.g., Härdle and Mammen (1993)).

[11]A simple example can be constructed as $\alpha(u) = \alpha + j_T g_1(u)$ if $0 \le u < c$, as $\alpha(u) = \alpha$ if $c \le u \le 1 - c$, and as $\alpha(u) = \alpha + j_T g_2(u)$ otherwise, where $\int_0^c g_1(u)\,du = \int_{1-c}^1 g_2(u)\,du = 0$ and $\Pi$ is a subset of $[c, 1-c]$.

[12]Following the notations and proof of Theorem 4 in Andrews (1993), we have $\bar{m}_T(\pi) = T^{-1}\sum_{t=1}^{T\pi}\{\mathbf{X}_t\varepsilon_t + \mathbf{X}_t\mathbf{X}_t' b_T f[(u-u_0)/r_T]\ \ \mathbf{0}\}' + T^{-1}\sum_{t=T\pi+1}^{T}\{\mathbf{0}\ \ \mathbf{X}_t\varepsilon_t + \mathbf{X}_t\mathbf{X}_t' b_T f[(u-u_0)/r_T]\}' = \{T^{-1}\sum_{t=1}^{T\pi}[\mathbf{X}_t\varepsilon_t\ \ \mathbf{0}]' + T^{-1}\sum_{t=T\pi+1}^{T}[\mathbf{0}\ \ \mathbf{X}_t\varepsilon]'\} + T^{-1}\sum_{t=1}^{T\pi}\{\mathbf{X}_t\mathbf{X}_t' b_T f[(u-u_0)/r_T]\ \ \mathbf{0}\}' + T^{-1}\sum_{t=T\pi+1}^{T}\{\mathbf{0}\ \ \mathbf{X}_t\mathbf{X}_t' b_T f[(u-u_0)/r_T]\}' \equiv A_1(\pi) + A_2(\pi) + A_3(\pi)$. By Theorem 1 of Andrews (1993), $\sqrt{T}A_1(\cdot) \Rightarrow G(\cdot)$, where $G(\cdot)$ is defined in equation (3.4) of Andrews (1993) and $A_j(\pi) = O_P(b_T r_T)$ for $j = 2, 3$. Therefore, if $b_T r_T\sqrt{T} \to 0$, we have $W_T(\cdot) \Rightarrow Q(\cdot)$ and $\sup_{\pi\in\Pi} W_T(\pi) \to^d \sup_{\pi\in\Pi} Q(\pi)$, where $Q(\pi) = [B(\pi) - \pi B(1)]'[B(\pi) - \pi B(1)]/[\pi(1-\pi)]$.

are similar to jumps or nonpersistent temporal structural breaks considered in the simulation. Andrews' (1993) and Bai and Perron's (1998) tests have good power against structural changes which last forever but are ineffective in detecting local sharp changes. In contrast, since local smoothing can capture local structural changes, our test can be asymptotically more powerful than those tests under $\mathbb{H}_{2A}(b_T, r_T)$.

In the classical normal linear regression model, Chow's (1960) $F$ test has the optimal power against a single structural break. We can also construct a generalized Chow $F$-type test by comparing the SSRs between the restricted constant parameter model and the unrestricted time-varying parameter model, namely

$$\hat{C} = [\sqrt{h}(\mathrm{SSR}_0 - \mathrm{SSR}_1) - \hat{A}_C]/\sqrt{\hat{B}_C},$$

where $\mathrm{SSR}_0 = \sum_{t=1}^{T}(Y_t - \mathbf{X}_t'\hat{\alpha})^2$, $\mathrm{SSR}_1 = \sum_{t=1}^{T}(Y_t - \mathbf{X}_t'\hat{\alpha}_t)^2$, and $\hat{A}_C$ and $\hat{B}_C$ are some suitable centering and scaling factors (see Chen and Hong (2008)). Interestingly, this optimality property disappears for $\hat{C}$ in the present setup, because it is asymptotically less powerful than $\hat{H}$ under the same local or global alternative. This is established in Theorem 4 below.

THEOREM 4: (i) *Suppose the conditions of Theorem* 3 *hold*, *and the same kernel* $k(\cdot)$ *and bandwidth* $h$ *are used for the* $\hat{H}$ *and* $\hat{C}$ *tests*. *Then* $\hat{H}$ *is asymptotically more efficient than* $\hat{C}$ *under* $\mathbb{H}_{1A}(j_T)$ *and* $\mathbb{H}_{2A}(b_T, r_T)$, *respectively*. (ii) *Suppose Assumptions* 1–7 *hold*. *Then* $\hat{H}$ *is asymptotically more efficient than* $\hat{C}$ *in terms of the Bahadur asymptotic efficiency criterion*.

Theorem 4(i) suggests that for all functions $g(\cdot)$ and $f(\cdot)$, $\hat{H}$ is more efficient than $\hat{C}$ in terms of the Pitman asymptotic efficiency criterion, which is suitable for local power analysis. Theorem 4(ii) shows that the relative efficiency of $\hat{H}$ over $\hat{C}$ carries over to the global alternative. Pitman's (1979) and Bahadur's (1960) asymptotic relative efficiency criteria are the limit ratios of the sample sizes required by the two tests to achieve the same asymptotic $p$-value under the same *local* alternative or *global* alternative, respectively. Theorem 4 implies that under the same set of conditions, including the same local or global alternative, the same bandwidth, and the same kernel, the generalized Hausman test is more efficient than the generalized Chow test.

Intuitively, the relative efficiency of $\hat{H}$ over $\hat{C}$ follows because the comparison of fitted values between the restricted and unrestricted models has a smaller sampling variation than the comparison of the SSRs between two models. To see this, we decompose

$$\text{SSR}_0 - \text{SSR}_1 = 2\sum_{t=1}^{T}(\hat{\alpha}_t - \hat{\alpha})'\mathbf{X}_t\varepsilon_t + \sum_{t=1}^{T}(\mathbf{X}_t'\hat{\alpha}_t - \mathbf{X}_t'\hat{\alpha})^2 + \text{remainder term.} \tag{5.1}$$

The asymptotic distribution of $\hat{C}$ is jointly determined by the first two terms in (5.1), whose total variance is greater than the variance of the second term, thus causing lower power than $\hat{H}$. The asymptotic distribution of $\hat{H}$ is determined by the second term of (5.1) only.

The relative efficiency of $\hat{H}$ over $\hat{C}$ is sizable. We show in the Supplemental Material that with the choice of bandwidth $h = cT^{-\lambda}$, Pitman and Bahadur relative efficiencies of $\hat{H}$ to $\hat{C}$ are

$$\text{RE}(\hat{H}:\hat{C}) = \left\{\int_{-1}^{1}\left[2k(v) - \int_{-1}^{1}k(u)k(u+v)\,du\right]^2 dv \Big/ \int_{-1}^{1}\left[\int_{-1}^{1}k(u)k(u+v)\,du\right]^2 dv\right\}^{1/(2-\lambda)},$$

which is always greater than 1 for any kernel satisfying Assumption 5. Suppose the bandwidth rate parameter $\lambda = 1/5$, which gives the optimal bandwidth minimizing an integrated MSE for estimating $\alpha(\cdot)$ on $[0, 1]$. Then for most commonly used kernels, such as the uniform, Epanechnikov, and quartic kernels, we have $\text{RE}(\hat{H} : \hat{C}) = 2.80$, 2.04, and 1.99, respectively.

Caution must be taken when the generalized Hausman and Chow tests reject $\mathbb{H}_0$. It is possible that the rejection is due to a nonlinear relationship or other model misspecifications rather than structural changes. For example, the choice of an inappropriate functional form and omitted variables can result in spurious structural changes. Of course, this is not particular to the proposed tests, but relevant to all existing tests for structural breaks.

## 6. FINITE SAMPLE PERFORMANCE

We now compare the finite sample performance of the proposed tests and those of Lin and Teräsvirta (1994), Andrews (1993), Bai and Perron (1998), and Elliott and Müller (2006).

To examine the size of all tests under $\mathbb{H}_0$, we consider the following DGP:

DGP S.1—No Structural Change:

$$Y_t = 1 + 0.5X_t + \varepsilon_t,$$
$$X_t = 0.5X_{t-1} + \nu_t, \quad \nu_t \sim \text{i.i.d. } N(0, 1).$$

To examine robustness of tests, we consider three cases for $\{\varepsilon_t\}$: (i) $\varepsilon_t \sim$ i.i.d. $N(0,1)$; (ii) $\varepsilon_t = \sqrt{h_t}u_t$, $h_t = 0.2 + 0.5\varepsilon_{t-1}^2$, $u_t \sim$ i.i.d. $N(0,1)$; (iii) $\varepsilon_t = \sqrt{h_t}u_t$, $h_t = 0.2 + 0.5X_t^2$, $u_t \sim$ i.i.d. $N(0,1)$. Note that $\text{var}(\varepsilon_t|X_t) \neq \sigma^2$ under case (iii). We generate 5000 data sets of the random sample $\{X_t, Y_t\}_{t=1}^T$ for each $T = 100, 250$, and 500.

We use the uniform kernel for $\hat{H}$ and $\hat{C}$ tests. Our experience suggests that the choice of $k(\cdot)$ has little impact on the performance of our tests. To conserve space, we report results based on the simple rule-of-thumb bandwidth $h = (1/\sqrt{12})T^{-1/5}$, which attains the optimal rate for local smoothing.[13] We compare $\hat{H}$ and $\hat{C}$ with a variety of popular tests, namely, Lin and Teräsvirta's (1994) LM test based on the first order Taylor expansion, Andrews' (1993) supremum LM test, Bai and Perron's (1998) UD max test, and Elliott and Müller's (2006) quasilocal level (qLL) test.[14] Following Andrews (1993), we choose the trimming region $\Pi = [0.15, 0.85]$ for the tests of Andrews (1993) and Bai and Perron (1998). For Bai and Perron's (1998) test, we set the upper bound of the number of breaks at 5. We consider both heteroscedasticity-robust and homoscedasticity-specific versions of all tests (the former are all denoted as -het), following Elliott and Müller (2006).

Table I reports the rejection rates of all tests under DGP S.1 at the 5% significance level, using asymptotic theory. Under i.i.d. and ARCH errors, both $\hat{H}$ and $\hat{C}$ overreject $\mathbb{H}_0$ when $T = 100$, but not excessively, and improve as $T$ increases; the $\hat{H}$ and $\hat{C}$ tests derived under conditional homoscedasticity and i.i.d. have better sizes than $\hat{H}$-het and $\hat{C}$-het, respectively. Under conditional heteroscedastic errors, the $\hat{H}$ and $\hat{C}$ tests derived under conditional homoscedasticity display strong overrejection, as is expected. For other tests, Andrews' supremum LM is quite conservative, especially its heteroscedasticity version. In contrast, Bai and Perron's UD max test shows quite a bit of overrejection. Overall, Lin and Teräsvirta's LM and Elliott and Müller's qLL tests have the best sizes in small samples, but our tests also have reasonable sizes.

Because the sizes of our tests using asymptotic theory differ from the nominal level in small samples and are a bit sensitive to bandwidth selection, we consider a wild bootstrap:

[13]We also try the rule-of-thumb bandwidth with different scaling parameters and the CV-based bandwidth described in Section 4. Simulation results, reported in the Supplemental Material, show that empirical sizes and powers are a bit sensitive to the bandwidth selection without bootstrap. However, the wild bootstrap described below alleviates the sensitivity to the choice of the bandwidth.

[14]To save space, simulation results for Brown, Durbin, and Evans' (1975) CUSUM test, Hackl's (1980) moving sum (MOSUM) test, Lin and Teräsvirta's (1994) LM2, LM3, Andrews and Ploberger's (1994) exponential and average LM tests, and Bai and Perron's (1998) WD max test can be found in the Supplemental Material.

TABLE I

EMPIRICAL LEVELS OF TESTS[a]

| Test[b] | $\varepsilon_t \sim$ i.i.d. $N(0,1)$ 100 | 250 | 500 | $\varepsilon_t \sim$ ARCH(1) 100 | 250 | 500 | $\varepsilon_t \mid X_t \sim N(0, f(X_t))$ 100 | 250 | 500 |
|---|---|---|---|---|---|---|---|---|---|
| | Rejection Rates Based on Bootstrap Critical Values | | | | | | | | |
| $\hat{H}$-het | 0.065 | 0.059 | 0.050 | 0.071 | 0.059 | 0.055 | 0.083 | 0.061 | 0.052 |
| $\hat{H}$ | 0.070 | 0.066 | 0.053 | 0.079 | 0.063 | 0.058 | 0.117 | 0.084 | 0.063 |
| $\hat{C}$-het | 0.069 | 0.060 | 0.052 | 0.071 | 0.063 | 0.052 | 0.081 | 0.056 | 0.050 |
| $\hat{C}$ | 0.073 | 0.066 | 0.055 | 0.079 | 0.069 | 0.056 | 0.118 | 0.080 | 0.063 |
| | Rejection Rates Based on Asymptotic Critical Values | | | | | | | | |
| $\hat{H}$-het | 0.095 | 0.078 | 0.053 | 0.116 | 0.087 | 0.066 | 0.120 | 0.092 | 0.071 |
| $\hat{H}$ | 0.079 | 0.071 | 0.047 | 0.097 | 0.080 | 0.062 | 0.350 | 0.428 | 0.471 |
| $\hat{C}$-het | 0.066 | 0.053 | 0.042 | 0.077 | 0.060 | 0.051 | 0.081 | 0.061 | 0.048 |
| $\hat{C}$ | 0.052 | 0.047 | 0.035 | 0.067 | 0.050 | 0.048 | 0.243 | 0.288 | 0.328 |
| LM-het | 0.043 | 0.054 | 0.048 | 0.044 | 0.047 | 0.053 | 0.045 | 0.045 | 0.050 |
| LM | 0.046 | 0.054 | 0.049 | 0.052 | 0.048 | 0.052 | 0.150 | 0.168 | 0.177 |
| Sup-LM-het | 0.018 | 0.035 | 0.043 | 0.013 | 0.029 | 0.037 | 0.010 | 0.022 | 0.034 |
| Sup-LM | 0.029 | 0.043 | 0.045 | 0.048 | 0.050 | 0.055 | 0.205 | 0.282 | 0.330 |
| UDMax-het | 0.138 | 0.085 | 0.067 | 0.153 | 0.082 | 0.066 | 0.260 | 0.132 | 0.096 |
| UDMax | 0.051 | 0.052 | 0.050 | 0.095 | 0.072 | 0.068 | 0.338 | 0.393 | 0.431 |
| qLL-het | 0.060 | 0.050 | 0.051 | 0.081 | 0.067 | 0.055 | 0.060 | 0.058 | 0.054 |
| qLL | 0.065 | 0.055 | 0.052 | 0.088 | 0.074 | 0.061 | 0.454 | 0.503 | 0.514 |

[a] 5% significance level; 5000 iterations.

[b] $\hat{H}$ and $\hat{C}$ are the generalized Hausman and Chow tests, LM is Lin and Teräsvirta's (1994) LM test based on the first-order Taylor expansion; Sup-LM is Andrews' (1993) supremum LM test; UDMax is Bai and Perron's (1998) double maximum test; qLL is Elliott and Müller's (2006) efficient test based on a "quasilocal level" model. *-het denotes the heteroscedasticity-robust version of the corresponding * test.

*Step* (i). Use the sample $\{Y_t, \mathbf{X}_t'\}_{t=1}^T$ to estimate the model via OLS and nonparametric regression, respectively, and compute the $\hat{H}$ statistic and the nonparametric residual $\hat{\varepsilon}_t = Y_t - \mathbf{X}_t'\hat{\alpha}_t$.

*Step* (ii). Obtain a wild bootstrap residual $\hat{\varepsilon}_t^*$ from the centered nonparametric residual $\bar{\varepsilon}_t = \hat{\varepsilon}_t - T^{-1}\sum_{t=1}^T \hat{\varepsilon}_t$ and construct a bootstrap sample $\{Y_t^*, \mathbf{X}_t'\}_{t=1}^T$, where $Y_t^* = \mathbf{X}_t'\hat{\alpha} + \hat{\varepsilon}_t^*$.[15]

*Step* (iii). Compute the bootstrap statistic $\hat{H}^*$ in the same way as $\hat{H}$, with $\{Y_t^*, \mathbf{X}_t'\}_{t=1}^T$ replacing the original sample $\{Y_t, \mathbf{X}_t'\}_{t=1}^T$.

*Step* (iv). Repeat Steps (ii) and (iii) $B$ times to obtain $B$ bootstrap test statistics $\{\hat{H}_l^*\}_{l=1}^B$, where $B$ is sufficiently large.

[15] We generate a wild bootstrap residual according to the formula that $\hat{\varepsilon}_t^* = a\bar{\varepsilon}_t$ with probability $1 - a/\sqrt{5}$ and $\hat{\varepsilon}_t^* = (1-a)\bar{\varepsilon}_t$ with probability $a/\sqrt{5}$, where $a = (1+\sqrt{5})/2$.

*Step* (v). Compute the bootstrap $p$-value $p^* \equiv B^{-1}\sum_{l=1}^{B}\mathbf{1}(\hat{H}_l^* > \hat{H})$, where $\mathbf{1}(\cdot)$ is the indicator function.

We generate 5000 data sets of random sample $\{Y_t, \mathbf{X}_t'\}_{t=1}^{T}$ and use $B = 499$ bootstrap iterations for each simulated data set. Table I shows that the bootstrap indeed approximates the finite sample distribution of test statistics more accurately. Table A.VII in the Supplemental Material shows that the bootstrap $p$-values are not sensitive to the choice of bandwidth $h$.

To investigate the power of all tests in detecting structural changes, we consider five alternatives:

DGP P.1—Single Structural Break[16]: We have

$$Y_t = \begin{cases} 1 + 0.5X_t + \varepsilon_t, & \text{if } t \leq 0.3T, \\ 1.2 + X_t + \varepsilon_t, & \text{otherwise.} \end{cases}$$

DGP P.2—Multiple Structural Breaks: We have

$$Y_t = \begin{cases} 0.6 + 0.3X_t + \varepsilon_t, & \text{if } 0.1T \leq t \leq 0.2T \text{ or } 0.7T \leq t \leq 0.8T, \\ 1.5 + X_t + \varepsilon_t, & \text{if } 0.4T \leq t \leq 0.5T, \\ 1 + 0.5X_t + \varepsilon_t, & \text{otherwise.} \end{cases}$$

DGP P.3—Nonpersistent Temporal Structural Breaks: We have

$$Y_t = \begin{cases} 1 + 0.5X_t + \varepsilon_t, & \text{if } t \leq 0.4T \text{ or } t \geq 0.6T, \\ 1.5 + X_t + \varepsilon_t, & \text{otherwise.} \end{cases}$$

DGP P.4—Smooth Structural Changes: We have

$$Y_t = F(\tau)(1 + 0.5X_t) + \varepsilon_t,$$

where $\tau = \frac{t}{T}$ and $F(\tau) = 1.5 - 1.5\exp[-3(\tau - 0.5)^2]$.

DGP P.5—Unit Root in Parameters: We have

$$Y_t = \rho_{1t} + \rho_{2t}X_t + \varepsilon_t,$$

where $\rho_{jt} = \rho_{jt-1} + u_{jt}$, $u_{jt} \sim$ i.i.d. $N(0, 1/15)$, and $j = 1, 2$.

For each of DGPs P.1–P.5, we generate 1000 data sets of the random sample $\{X_t, Y_t\}_{t=1}^{T}$ for each $T = 100$, 250, and 500. Table II reports the rejection rates of all tests with empirical critical values (ECVs), which are size-adjusted critical

[16]For robustness, we consider different locations of structural changes. To save space, results are reported in the Supplemental Material.

TABLE II

EMPIRICAL POWERS OF TESTS[a]

| | DGP P.1 Single Break | | | DGP P.2 Multiple Breaks | | | DGP P.3 Nonpersistent Temporal Breaks | | | DGP P.4 Smooth Changes | | | DGP P.5 Unit Root in Parameters | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Test[b] | 100 | 250 | 500 | 100 | 250 | 500 | 100 | 250 | 500 | 100 | 250 | 500 | 100 | 250 | 500 |
| | | | | | | | Rejection Rates Based on Bootstrap Critical Values | | | | | | | | |
| $\hat{H}$-het | 0.399 | 0.847 | 0.998 | 0.293 | 0.673 | 0.951 | 0.324 | 0.770 | 0.977 | 0.484 | 0.889 | 0.997 | 0.636 | 0.988 | 1.00 |
| $\hat{H}$ | 0.418 | 0.855 | 0.998 | 0.319 | 0.694 | 0.953 | 0.343 | 0.788 | 0.977 | 0.512 | 0.900 | 0.997 | 0.654 | 0.989 | 1.00 |
| $\hat{C}$-het | 0.320 | 0.751 | 0.975 | 0.269 | 0.676 | 0.955 | 0.304 | 0.729 | 0.968 | 0.406 | 0.799 | 0.984 | 0.560 | 0.983 | 1.00 |
| $\hat{C}$ | 0.337 | 0.761 | 0.979 | 0.286 | 0.690 | 0.961 | 0.321 | 0.746 | 0.970 | 0.428 | 0.809 | 0.985 | 0.589 | 0.985 | 1.00 |
| | | | | | | | Rejection Rates Based on Empirical Critical Values | | | | | | | | |
| $\hat{H}$-het | 0.416 | 0.857 | 0.990 | 0.308 | 0.646 | 0.954 | 0.319 | 0.748 | 0.990 | 0.443 | 0.884 | 0.998 | 0.617 | 0.983 | 1.00 |
| $\hat{H}$ | 0.416 | 0.857 | 0.990 | 0.306 | 0.665 | 0.957 | 0.325 | 0.765 | 0.987 | 0.452 | 0.886 | 0.998 | 0.626 | 0.987 | 1.00 |
| $\hat{C}$-het | 0.314 | 0.728 | 0.972 | 0.276 | 0.615 | 0.942 | 0.306 | 0.693 | 0.971 | 0.378 | 0.786 | 0.992 | 0.555 | 0.982 | 1.00 |
| $\hat{C}$ | 0.352 | 0.731 | 0.972 | 0.296 | 0.621 | 0.949 | 0.330 | 0.700 | 0.977 | 0.401 | 0.793 | 0.992 | 0.564 | 0.985 | 1.00 |
| LM-het | 0.404 | 0.814 | 0.993 | 0.052 | 0.054 | 0.061 | 0.059 | 0.058 | 0.059 | 0.065 | 0.055 | 0.059 | 0.557 | 0.889 | 0.971 |
| LM | 0.458 | 0.853 | 0.993 | 0.053 | 0.045 | 0.061 | 0.046 | 0.050 | 0.046 | 0.074 | 0.072 | 0.063 | 0.585 | 0.896 | 0.972 |
| Sup-LM-het | 0.427 | 0.888 | 0.999 | 0.115 | 0.231 | 0.478 | 0.129 | 0.330 | 0.665 | 0.215 | 0.598 | 0.942 | 0.563 | 0.945 | 0.999 |
| Sup-LM | 0.501 | 0.923 | 1.00 | 0.145 | 0.286 | 0.554 | 0.143 | 0.348 | 0.708 | 0.281 | 0.675 | 0.964 | 0.623 | 0.964 | 0.999 |
| UDMax-het | 0.393 | 0.871 | 0.996 | 0.183 | 0.400 | 0.773 | 0.236 | 0.634 | 0.973 | 0.258 | 0.683 | 0.978 | 0.524 | 0.785 | 0.966 |
| UDMax | 0.494 | 0.922 | 1.00 | 0.240 | 0.556 | 0.914 | 0.262 | 0.736 | 0.993 | 0.345 | 0.781 | 0.989 | 0.631 | 0.815 | 0.978 |
| qLL-het | 0.376 | 0.865 | 0.996 | 0.217 | 0.637 | 0.928 | 0.222 | 0.709 | 0.968 | 0.418 | 0.892 | 0.997 | 0.613 | 0.979 | 1.00 |
| qLL | 0.428 | 0.873 | 0.996 | 0.297 | 0.657 | 0.942 | 0.324 | 0.734 | 0.976 | 0.440 | 0.897 | 0.997 | 0.645 | 0.983 | 1.00 |

[a]5% significance level; 1000 iterations.

[b]$\hat{H}$ and $\hat{C}$ are the generalized Hausman and Chow tests, LM is Lin and Teräsvirta's (1994) LM test based on the first-order Taylor expansion; Sup-LM is Andrews' (1993) supremum LM test; UDMax is Bai and Perron's (1998) double maximum test; qLL is Elliott and Müller's (2006) efficient test based on a quasilocal level model. *-het denotes the heteroscedasticity-robust version of the corresponding * test.

values,[17] under DGPs P.1–P.5 at the 5% level. We first consider the deterministic single break (DGP P.1), namely, a single break with a given breakpoint and size. For the interior breakpoint, all tests have power against DGP P.1, although Andrews' sup-LM-het is most powerful among all heteroscedasticity-robust tests. The $\hat{H}$-het test performs slightly better than LM-het, UD max-het, and qLL-het tests when $T$ is small. Results for the single break at $t = 0.1T$ (near the left boundary of the sample period) are report in the Supplemental Material. Here, $\hat{H}$-het outperforms LM-het, UD max-het, sup-LM-het, and qLL-het tests. We note that all homoscedasticity-specific tests are more powerful than their heteroscedasticity-robust counterparts, and $\hat{H}$ is more powerful than $\hat{C}$ under DGP P.1, confirming our theory.

Next, we consider multiple breaks. Under DGP P.2, the $\hat{H}$ and $\hat{C}$ tests dominate all other tests. Lin and Teräsvirta's LM test has no power even when $T = 500$. Bai and Perron's UD max test improves a lot upon Andrews' single break test, which confirms Perron's (2006) observation that "while the test for one break is consistent against alternatives involving multiple changes, its power in finite samples can be rather poor."

Under DGP P.3, the break lasts only for some period of time. The $\hat{H}$ test outperforms other tests for all sample sizes. Lin and Teräsvirta's LM test has low or little power. UD max and qLL tests perform slightly worse than ours but better than Andrews' sup-LM test.

DGP P.4 is an alternative with nonmonotonic smooth structural changes. This is a STR model considered in Lin and Teräsvirta (1994), where the transition function is a second-order logistic function. Not surprisingly, Lin and Teräsvirta's LM test, which is based on a first-order Taylor expansion, has no power. Our tests and the qLL test outperform other tests.

Finally, we consider the alternative with unit root in parameters (DGP P.5). Again, the $\hat{H}$ test outperforms all other tests. The qLL test is slightly less powerful than $\hat{H}$, but more powerful than LM and sup-LM tests. We note that in most cases, the power results for tests using bootstrap critical values are similar to those using ECVs.

To sum up, (i) the empirical sizes of the $\hat{H}$ and $\hat{C}$ tests are larger than the nominal levels, but they improve as the sample size increases. Under conditional homoscedasticity, the homoscedasticity-specific tests, $\hat{H}$ and $\hat{C}$, have better sizes than heteroscedasticity-robust tests $\hat{H}$-het and $\hat{C}$-het, respectively. Under conditional heteroscedasticity, $\hat{H}$-het and $\hat{C}$-het continue to have reasonable levels, but homoscedasticity-specific tests strongly overreject the correct model. Other homoscedasticity-specific and heteroscedasticity-robust tests have similar patterns. (ii) Our tests have reasonable all-around power

[17]For example, the 5%-level ECV is the 95% quantile of the empirical distribution of the statistic obtained in 5000 replications under DGP S.1.

against both smooth and abrupt structural changes. They outperform all other tests in detecting smooth structural changes; they have good power against various multiple structural breaks, including the alternatives where the break occurs near the boundary of the sample period. (iii) Our tests are not always the most powerful in detecting each of the alternatives considered. However, they have relatively omnibus power against all five DGPs, provided the sample size is sufficiently large. Tests for parametric smooth structural changes are powerful against the specified alternatives, but they may have low power if the polynomial order is not high enough; tests with the trimmed range also have a danger of omitting breaks occurring near the boundary of the sample period; and tests for single break may have rather poor power against alternatives involving multiple breaks. (iv) The generalized Hausman test $\hat{H}$ is more powerful than the generalized Chow test $\hat{C}$ in most cases, confirming our asymptotic theory. (v) The heteroscedasticity-robust generalized Hausman and Chow tests have power similar to their respective homoscedasticity-specific counterparts in most cases. This feature is not shared by other tests.

## 7. STABILITY OF RETURN PREDICTION MODELS

Stock return predictability is an important yet controversial issue in empirical finance. Numerous studies document the predictability of stock returns using various lagged financial and macroeconomic variables, such as the dividend price ratio, earning price ratio, book-to-market ratio, term spread, default premium, interest rates, and inflation rate as well as corporate payout and financing activity. Most existing works focus on in-sample tests.

A recent critique that challenges the conventional wisdom of return predictability emphasizes that predictive regressions have poor out-of-sample performance. Welch and Goyal (2008) showed that all aforementioned financial and macroeconomic variables fail to yield better out-of-sample forecasts of the U.S. equity premium than the simple historical mean equity returns. This striking finding triggers vigorous debates in the literature. One possible reason that significant in-sample evidence of predictability is often accompanied by weak out-of-sample evidence of predictability is the existence of structural changes. Indeed, Clark and McCracken (2005) presented analytical evidence on the effects of structural breaks on the tests for equal forecast accuracy and encompassing. They showed that out-of-sample predictive evidence can be harder to detect because the results of out-of-sample tests are highly dependent on the timing of the predictability.

We now use our tests to check whether the predictive regression of stock returns is stable over time. Some existing studies have considered structural breaks in the equity premium, but results are mixed. For example, Kim, Morley, and Nelson (2005) found a one-time structural break in the equity premium in the 1940s, but no additional breaks in the postwar period. Paye and Timmermann (2006) examined the stability of return prediction models for

10 Organization for Economic Cooperation and Development (OECD) countries. They found strong evidence against stability in a multivariate regression with the dividend yield, short rate, term spread, and default spread, but in the univariate regressions, they found fairly weak evidence on instability in the dividend yield regression or default premium regression. Using Elliott and Müller's (2006) test, Rapach and Wohar (2006) cannot reject structural stability in three of the eight predictive regressions (the price earning ratio, term spread, and short rate) for S&P 500 returns. As emphasized by Paye and Timmermann (2006), all existing tests focus on occasional, large shifts in coefficients rather than a gradual evolution. We avoid this restriction by using our tests, which have power against both smooth structural changes and sudden breaks.

We consider a standard predictive regression $Y_{t+1} = \alpha + \beta' X_t + \varepsilon_{t+1}$, where $Y_{t+1} = \log[(P_{t+1} + D_{t+1})/P_t] - r_t$, $P_t$ is the S&P 500 index, $D_t$ is the dividend paid on the S&P 500 index, $r_t$ is the 3-month Treasury bill rate, and $X_t$ is a predetermined predictor. Following Welch and Goyal (2008) and Rapach, Strauss, and Zhou (2010), we consider 14 financial and economic variables:

(i) *Log dividend price ratio* (D/P): The log difference between dividends and the S&P 500 index, where dividends are computed via a 1-year moving sum.

(ii) *Log dividend yield* (D/Y): The log difference between dividends and the lagged S&P 500 index.

(iii) *Log earnings price ratio* (E/P): The log difference between earnings and the lagged S&P 500 index, where earnings are computed via a 1-year moving sum.

(iv) *Log dividend payout ratio* (D/E): The log difference between dividends and earnings.

(v) *Stock variance* (SVAR): The sum of squared daily returns on the S&P 500 index.

(vi) *Book-to-market ratio* (B/M): The ratio of book value to market value for the Dow Jones Industrial Average.

(vii) *Net equity expansion* (NTIS): The ratio of 12-month moving sums of net issues by NYSE-listed stocks to the total end-of-year market capitalization of NYSE stocks.

(viii) *Treasury bill rate* (TBL): The 3-month Treasury bill rate.

(ix) *Long-term yield* (LTY): The long-term government bond yield.

(x) *Long-term return* (LTR): The return on long-term government bonds.

(xi) *Term spread* (TMS): The difference between the long-term yield and the Treasury bill rate.

(xii) *Default yield spread* (DFY): The difference between BAA- and AAA-rated corporate bond yields.

(xiii) *Default return spread* (DFR): The difference between long-term corporate bond and long-term government bond returns.

(xiv) *Inflation* (INFL): The consumer price index (CPI)-based inflation rate in the previous period.

All data are from Welch and Goyal (2008).[18]

We apply our tests to monthly and quarterly stock returns and compare them with the UD max and qLL tests, which have displayed overall good finite sample performance in our simulation study and have been used in Paye and Timmermann (2006) and Rapach and Wohar (2006). We consider the postwar sample, January 1947–December 2005, and the post-oil-shocks subsample, January 1976–December 2005.

Table III reports the bootstrap $p$-values of the heteroscedasticity-robust version of our tests, UD max, and qLL for univariate predictor regressions with each of the above 14 predictors using monthly and quarterly data. The bootstrap $p$-values, based on 9999 bootstrap iterations, are computed as described in Section 6. For the whole sample, we find strong evidence against the model stability for all predictors considered: all bootstrap $p$-values of our tests and qLL tests are smaller than 1%. It is also evident from the figures in the Supplemental Material that the nonparametric estimators of the slope coefficient $\beta$ for all univariate predictor regressions do change over time and smooth structural changes should not be ruled out. For the subsample of 1976–2005, our $\hat{H}$ test is able to reject the model stability of all predictors except E/P at the 10% level and all except D/P, D/Y, E/P, and LTY at the 5% level. Bai and Perron's (1998) UD max and Elliott and Müller's (2006) qLL tests also yield strong rejection in the whole sample. However, for the post-oil-shocks subsample, UD max cannot reject D/Y and E/P, and qLL cannot reject D/P, D/Y, and E/P at the 10% level.

Next, we test the stability of popular multivariate predictor models, including two bivariate predictor regressions with D/P and TBL and D/P and E/P, one trivariate predictor regression with D/P, E/P, and TBL, and one quadrivariate predictor regression with D/P, TBL, TMS, and DFR. The bivariate and trivariate models were studied by Ang and Bekaert (2007), and the quadrivariate model was studied by Paye and Timmermann (2006). The strong evidence of model instability in the whole sample carries over to the multivariate predictor regressions. Our tests reject the stability of all multivariate models at all conventional significance levels (e.g., the 1% level); UD max is able to reject all multivariate models at the 5% level, but qLL does not find structural break in the trivariate predictor regression. For the post-oil-shocks subsample, our tests and UD max have similar $p$-values: they reject the bivariate predictor regression with D/P and TBL, the trivariate and quadrivariate predictor regressions at the 5% level, and the bivariate predictor regression with D/P and E/P at the 10% level. On the other hand, the qLL test finds no evidence against model stability for bivariate predictor regressions.

[18]Some variables are potentially nonstationary (e.g., the Treasury bill rate), which may complicate our testing results. Extending our tests to the nonstationary case requires very different asymptotic derivation and is left for future study. Here we follow the literature to consider the predictive regression under stationarity.

TABLE III

STABILITY TEST FOR EXCESS RETURN[a]

| | Monthly Excess Return | | | | Quarterly Excess Return | | | |
|---|---|---|---|---|---|---|---|---|
| | $\hat{H}$-het | $\hat{C}$-het | UDmax-het | qLL-het | $\hat{H}$-het | $\hat{C}$-het | UDmax-het | qLL-het |
| | Univariate Predictor Regressions | | | | | | | |
| | 1947 January–2005 December | | | | 1947Q1–2005Q4 | | | |
| D/P | 0.0006 | 0.0001 | 0.0006 | 0.0006 | 0.0066 | 0.0014 | 0.0390 | 0.0367 |
| D/Y | 0.0019 | 0.0007 | 0.0010 | 0.0004 | 0.0177 | 0.0037 | 0.0477 | 0.0495 |
| E/P | 0.0069 | 0.0086 | 0.0226 | 0.0010 | 0.2179 | 0.0428 | 0.0867 | 0.3697 |
| D/E | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.1931 | 0.1760 | 0.0779 | 0.0059 |
| SVAR | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0006 | 0.0034 | 0.0037 | 0.0098 |
| B/M | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0126 | 0.0007 | 0.0149 | 0.0855 |
| NTIS | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0222 | 0.0431 | 0.0013 | 0.0001 |
| TBL | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0103 | 0.0025 | 0.0034 | 0.0008 |
| LTY | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0334 | 0.0231 | 0.0012 | 0.0020 |
| LTR | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0712 | 0.1506 | 0.0227 | 0.0008 |
| TMS | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0048 | 0.0051 | 0.0079 | 0.0012 |
| DFY | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0074 | 0.0071 | 0.0007 | 0.0021 |
| DFR | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0233 | 0.0367 | 0.0150 | 0.0010 |
| INFL | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.2552 | 0.2848 | 0.1060 | 0.0025 |
| | 1976 January–2005 December | | | | 1976Q1–2005Q4 | | | |
| D/P | 0.0583 | 0.0222 | 0.0765 | 0.4318 | 0.0368 | 0.0126 | 0.2317 | 0.2346 |
| D/Y | 0.0911 | 0.0456 | 0.1072 | 0.4963 | 0.1549 | 0.0594 | 0.4085 | 0.3779 |
| E/P | 0.3291 | 0.5726 | 0.1580 | 0.1934 | 0.1811 | 0.8289 | 0.3740 | 0.1763 |
| D/E | 0.0011 | 0.0030 | 0.0014 | 0.0000 | 0.6413 | 0.6473 | 0.6697 | 0.4440 |
| SVAR | 0.0001 | 0.0001 | 0.0113 | 0.0001 | 0.2520 | 0.3279 | 0.7433 | 0.4178 |
| B/M | 0.0395 | 0.0148 | 0.0832 | 0.0330 | 0.2130 | 0.1261 | 0.6351 | 0.2173 |
| NTIS | 0.0010 | 0.0010 | 0.0038 | 0.0000 | 0.2059 | 0.2553 | 0.5785 | 0.2536 |
| TBL | 0.0177 | 0.0002 | 0.0229 | 0.0098 | 0.4301 | 0.0771 | 0.4730 | 0.3593 |
| LTY | 0.0735 | 0.0419 | 0.0098 | 0.0477 | 0.4211 | 0.2712 | 0.1614 | 0.4385 |
| LTR | 0.0002 | 0.0020 | 0.0012 | 0.0000 | 0.7391 | 0.9013 | 0.7003 | 0.6214 |
| TMS | 0.0003 | 0.0000 | 0.0020 | 0.0001 | 0.6264 | 0.6361 | 0.6080 | 0.4592 |
| DFY | 0.0077 | 0.0316 | 0.0094 | 0.0042 | 0.5718 | 0.8576 | 0.8013 | 0.2391 |
| DFR | 0.0000 | 0.0001 | 0.0024 | 0.0000 | 0.1433 | 0.1712 | 0.1009 | 0.2716 |
| INFL | 0.0014 | 0.0086 | 0.0028 | 0.0003 | 0.7536 | 0.7311 | 0.6857 | 0.6311 |
| | Multivariate Predictor Regressions | | | | | | | |
| | 1947 January–2005 December | | | | 1947Q1–2005Q4 | | | |
| Bi 1 | 0.0000 | 0.0000 | 0.0082 | 0.0170 | 0.0000 | 0.0000 | 0.1391 | 0.0779 |
| Bi 2 | 0.0005 | 0.0005 | 0.0006 | 0.0020 | 0.0014 | 0.0005 | 0.0770 | 0.0293 |
| Tri | 0.0000 | 0.0000 | 0.0062 | 0.1456 | 0.0000 | 0.0000 | 0.1805 | 0.0912 |
| Quadri | 0.0000 | 0.0000 | 0.0142 | 0.0147 | 0.0000 | 0.0001 | 0.1361 | 0.1136 |
| | 1976 January–2005 December | | | | 1976Q1–2005Q4 | | | |
| Bi 1 | 0.0022 | 0.0009 | 0.0418 | 0.1767 | 0.0102 | 0.0012 | 0.2253 | 0.1845 |
| Bi 2 | 0.0810 | 0.0596 | 0.0751 | 0.5315 | 0.0658 | 0.1181 | 0.2102 | 0.3687 |
| Tri | 0.0192 | 0.0129 | 0.0189 | 0.0181 | 0.0276 | 0.0142 | 0.3679 | 0.0385 |
| Quadri | 0.0375 | 0.0296 | 0.0180 | 0.0289 | 0.0105 | 0.0248 | 0.2592 | 0.0271 |

[a] $\hat{H}$ and $\hat{C}$ are the generalized Hausman and Chow tests; UDMax is Bai and Perron's (1998) double maximum test; qLL is Elliott and Müller's (2006) efficient test based on a quasilocal level model. *-het denotes the heteroscedasticity-robust version of the corresponding * test. The bootstrap $p$-values are based on $B = 9999$.

For quarterly data, the evidence against stability is a bit weaker for univariate predictor regressions. For the whole sample, our $\hat{H}$ test rejects all univariate predictor regressions except E/P, D/E, LTR, and INFL at the 5% level; UD max cannot reject E/P, D/E, and INFL, and qLL cannot reject E/P and B/M at the 5% level. For the subsample, all tests can barely find evidence against model stability for univariate predictor regressions: our tests can only reject the null with D/P, UD max can only marginally reject the stability of DFR, and qLL rejects none. We conjecture that the weak evidence against model stability is mainly due to the small sample size. However, our tests firmly reject the stability hypothesis for all multivariate models in both periods considered. In particular, the $p$-values of our tests are essentially 0 for the whole sample. In comparison, qLL is able to reject all except the quadrivariate model and UD max is only able to reject the bivariate predictor regression with D/P and E/P at the 10% level for the whole sample; for the subsample, qLL is able to reject the trivariate and quadrivariate models, but UD max does not find any evidence against model stability.

To sum up, we find *stronger* evidence against stability in both univariate and multivariate regressions than the existing literature in asset return predictability. Our tests strongly reject the stability of univariate and multivariate return prediction models in the postwar and the post-oil-shocks sample periods. Our findings support the argument of Rapach, Strauss, and Zhou (2010) that "model uncertainty and instability seriously impair the forecasting ability of individual predictive regression models." Our nonparametric estimation suggests that smooth structural changes are a possibility (see, e.g., Figures 8 and 11 in the Supplemental Material). We also find evidence that our local alternative $\mathbb{H}_{2A}(b_T, r_T)$ may be relevant in practice (see, e.g., Figures 1 and 3 in the Supplemental Material). Of course, the rejection may be due to model misspecification, and how to reconstruct the models needs further investigation.

## 8. CONCLUSION

Detection and identification of structural breaks have attracted a lot attention in econometrics over the past several decades. We have contributed to this literature by proposing a nonparametric Wald-type test for smooth structural changes as well as abrupt structural breaks. Our test has intuitive appeal because it can be regarded as the generalization of Hausman's (1978) test from a parametric context to a nonparametric context. It is asymptotically pivotal, does not require trimming data, does not require prior information on the alternative, and is consistent against all smooth structural changes as well as multiple abrupt structural breaks. Simulation studies show that the proposed test performs reasonably in finite samples. We apply the proposed test to stock return prediction models and find strong evidence against model stability.

We can extend the proposed method to linear regression models with nonstationary regressors or serially correlated and endogenous errors with time-varying variances. Our approach can also be adopted to test whether an autoregressive moving average model or a generalized autoregressive conditional heteroscedasticity model has smooth structural changes, using the log-likelihood criterion. Moreover, it can be used to test whether a time trend follows a polynomial of time, with the stochastic component being a weakly stationary but not necessarily m.d.s. All these topics are left for future research.

## REFERENCES


ANDREWS, D. W. K. (1993): "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, 821–856. [1159,1168-1172,1174]

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414. [1171]

ANG, A., AND G. BEKAERT (2007): "Return Predictability: Is It There?" *Review of Financial Studies*, 20, 651–707. [1178]

BAHADUR, R. R. (1960): "Stochastic Comparison of Tests," *Annals of Mathematical Statistics*, 31, 276–295. [1169]

BAI, J., AND P. PERRON (1998): "Estimating and Testing Linear Models With Multiple Structural Changes," *Econometrica*, 66, 47–78. [1159,1168-1172,1174,1178,1179]

BROWN, R. L., J. DURBIN, AND J. M. EVANS (1975): "Techniques for Testing the Constancy of Regression Relationship Over Time," *Journal of Royal Statistical Society, Ser. B*, 37, 149–192. [1159,1171]

CAI, Z. (2007): "Trending Time-Varying Coefficient Time Series Models With Serially Correlated Errors," *Journal of Econometrics*, 136, 163–188. [1158,1160-1163]

CHEN, B., AND Y. HONG (2008): "Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression," Working Paper, University of Rochester and Cornell University. [1158,1169]

——— (2010): "Characteristic Function-Based Testing for Multifactor Continuous-Time Markov Models via Nonparametric Regression," *Econometric Theory*, 26, 1115–1179. [1166]

——— (2012): "Supplement to 'Testing for Smooth Structural Changes in Times Series Models via Nonparametric Regression'," *Econometrica Supplemental Material*, 80, http://www.econometricsociety.org/ecta/Supmat/7990_proof.pdf; http://www.econometricsociety.org/ecta/Supmat/7990_data and programs.zip. [1159]

CHOW, G. C. (1960): "Tests of Equality Between Subsets of Coefficients in Two Linear Regressions," *Econometrica*, 28, 591–605. [1158,1160,1161,1169]

CHU, C. J., K. HORNIK, AND C. M. KUAN (1995): "MOSUM Tests for Parameter Constancy," *Biometrika*, 82, 603–617. [1162]

CLARK, T. E., AND M. W. MCCRACKEN (2005): "The Power of Tests of Predictive Ability in the Presence of Structural Breaks," *Journal of Econometrics*, 124, 1–31. [1176]

CLINE, D. B. H., AND J. D. HART (1991): "Kernel Estimation of Densities With Discontinuities or Discontinuous Derivative," *Statistics*, 22, 69–84. [1162]

DAHLHAUS, R. (1996): "On the Kullback–Leibler Information Divergence of Locally Stationary Processes," *Stochastic Process Application*, 62, 139–168. [1161]

ELLIOTT, G., AND U. MÜLLER (2006): "Efficient Tests for General Persistent Time Variation in Regression Coefficients," *Review of Economic Studies*, 73, 907–940. [1170-1172,1174, 1177-1179]

FARLEY, J. U., M. J. HINICH, AND T. W. MCGUIRE (1975): "Some Comparisons of Tests for a Shift in the Slopes of a Multivariate Linear Time Series Model," *Journal of Econometrics*, 3, 297–318. [1158]

FRYZLEWICZ, P., T. SAPATINAS, AND S. SUBBA RAO (2008): "Normalized Least-Squares Estimation in Time-Varying ARCH Models," *The Annals of Statistics*, 36, 742–786. [1165]

HACKL, P. (1980): *Testing the Constancy of Regression Models Over Time*. Gottingen: Vandenhoeck & Ruprecht. [1171]

HALL, P., AND J. HART (1990): "Nonparametric Regression With Long-Range Dependence," *Stochastic Process Application*, 36, 339–351. [1160]

HALL, P., AND T. E. WEHRLY (1991): "A Geometrical Method for Removing Edge Effects From Kernel-Type Nonparametric Regression Estimators," *Journal of the American Statistical Association*, 86, 665–672. [1162]

HANSEN, B. (2001): "The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity," *Journal of Economic Perspectives*, 15, 117–128. [1157]

HÄRDLE, W., AND E. MAMMEN (1993): "Comparing Nonparametric versus Parametric Regression Fits," *Annals of Statistics*, 21, 1926–1947. [1168]

HAUSMAN, J. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271. [1158, 1163,1180]

KIM, C., J. C. MORLEY, AND C. R. NELSON (2005): "The Structural Break in the Equity Premium," *Journal of Business & Economic Statistics*, 23, 181–191. [1176]

LIN, C. J., AND T. TERÄSVIRTA (1994): "Testing the Constancy of Regression Parameters Against Continuous Structural Change," *Journal of Econometrics*, 62, 211–228. [1157,1158,1160,1165, 1170-1172,1174,1175]

ORBE, S., E. FERREIRA, AND J. RODRIGUEZ-PÓO (2000): "A Nonparametric Method to Estimate Time Varying Coefficients Under Seasonal Constraints," *Journal of Nonparametric Statistics*, 12, 779–806. [1158,1160]

——— (2005): "Nonparametric Estimation of Time Varying Parameters Under Shape Restrictions," *Journal of Econometrics*, 126, 53–77. [1158,1160]

PAYE, B., AND A. TIMMERMANN (2006): "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13, 274–315. [1176-1178]

PERRON, P. (2006): "Dealing With Structural Breaks," in *Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory*, ed. by K. Patterson and T. C. Mills. New York: Palgrave Macmillan, 278–352. [1175]

PHILLIPS, P. C. B., AND B. E. HANSEN (1990): "Statistical Inference in Instrumental Variables Regression With $I(1)$ Processes," *Review of Economic Studies*, 57, 99–125. [1160]

PITMAN, E. J. G. (1979): *Some Basic Theory for Statistical Inference*. New York: Chapman and Hall. [1169]

RAPACH, D., AND M. WOHAR (2006): "Structural Breaks and Predictive Regression Models of Aggregate U.S. Stock Returns," *Journal of Financial Econometrics*, 2, 238–274. [1177,1178]

RAPACH, D., J. STRAUSS, AND G. ZHOU (2010): "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy," *Review of Financial Studies*, 23, 821–862. [1177,1180]

ROBINSON, P. M. (1989): "Nonparametric Estimation of Time-Varying Parameters," in *Statistical Analysis and Forecasting of Economic Structural Change*, ed. by P. Hackl. Springer, 253–264. [1158,1160,1165]

——— (1991): "Time-Varying Nonlinear Regression," in *Economic Structural Change Analysis and Forecasting*, ed. by P. Hackl. Berlin: Springer, 179–190. [1158,1160,1161]

ROSENBLATT, M. (1975): "A Quadratic Measure of Deviation of Two-Dimensional Density Estimates and a Test of Independence," *Annals of Statistics*, 3, 1–14. [1167]

SUN, Y., P. PHILLIPS, AND S. JIN (2008): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica*, 76, 175–194. [1166]

WELCH, I., AND A. GOYAL (2008): "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, 1455–1508. [1176-1178]

WONG, W. H. (1983): "On the Consistency of Cross-Valiation in Kernel Nonparametric Regression," *Annals of Statistics*, 11, 1136–1141. [1165]

*Dept. of Economics, University of Rochester, Rochester, NY 14627, U.S.A.; bchen8@mail.rochester.edu* *and*

*Dept. of Economics and Dept. of Statistical Science, Cornell University, Ithaca, NY 14850, U.S.A. and Wang Yanan Institute for Studies in Economics (WISE) & MOE Key Laboratory of Econometrics, Xiamen University, Xiamen 361005, China; yh20@cornell.edu.*

*Manuscript received June, 2008; final revision received June, 2011.*