JOURNAL OF
Econometrics

ELSEVIER Journal of Econometrics 88 (1999) 283–299

Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series

Timothy J. Vogelsang*

Department of Economics, Cornell University, Uris Hall, Ithaca, NY 14853-7601, USA

Received 1 March 1997; received in revised form 1 January 1998

Abstract

Sources of nonmonotonic power are uncovered for a wide variety of tests for a shift in the mean of a dynamic time series. Two main sources of nonmonotonic power are found. The first source is the behavior of the estimate of error variance under the alternative hypothesis of a shift in mean. In particular if the error variance is estimated under the null hypothesis, nonmonotonic power can result. The second source is the presence of a lagged dependent variable in the estimated regression.

JEL classification: C12; C22

Keywords: Serial correlation; Wald test; Structural change; Slope shift; Unit root; Simulation

1. Introduction

In recent studies Perron (1991) and Vogelsang (1997a) documented that tests for structural change in the trend function of a dynamic time series can have nonmonotonic power functions. In other words, as the distance between the null and alternative increases, power is decreasing in some ranges. In fact, there are examples where power drops to near zero when trend shifts become very large in magnitude. Undetected shifts in trend bias estimates and tests of dynamic parameters that may be of interest. For small shifts in trend these biases are

* E-mail: tjv2@cornell.edu.

PII: S0304-4076(98)00034-7

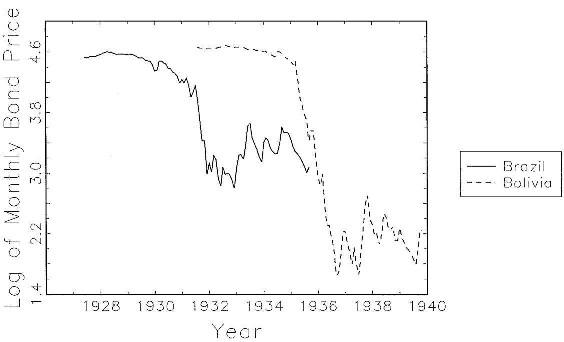

Fig. 1. Logarithm of monthly bond prices for Brazil and Bolivia.

likely to be small and of little consequence. However, as the trend shifts increase in magnitude, these biases can become substantial. Therefore, it seems a reasonable requirement to have trend function tests that can detect large shifts in trend with higher probability than smaller shifts in trend. This requirement is not satisfied by tests with nonmonotonic power.[1]

To illustrate potential practical problems caused by nonmonotonic power consider the two time series plotted in Fig. 1. These series are logarithms of monthly high bond prices for bonds issued by the Brazilian and Bolivian governments between the world wars. Each series spans about eight years and has 100 observations. The data was taken from various issues of the Commercial and Financial Chronicle. Both series clearly exhibit massive mean shifts at about the midpoint of their spans. A single break point was estimated for each series using least squares. For each possible break point the series were regressed on a constant and a mean shift dummy variable denoted by $DU_t = 1(t > T_b)$ where T_b is the shift date, T is the sample size and $1(\cdot)$ is the indicator function. The estimated break point is the break point that minimizes the sums of squared residuals over all possible break points. Denote by $\hat{T}_b$ the estimated shift date and $\hat{\lambda} = \hat{T}_b/T$ the estimated break point. From Bai (1994) it follows that $\hat{\lambda}$ converges to the true break point at rate T, and this result holds for serially correlated errors that are stationary linear processes. The magnitudes of the mean shifts were estimated by least squares by regressing each series on a constant and DU_t using $\hat{T}_b$ for the shift date. The estimated coefficient on DU_t

[1] Nonmonotonic power is not unique to tests for shifts in trend. Nelson and Savin (1990) showed that Wald tests in a model based on a simple one-parameter exponential response function can exhibit nonmonotonicities. Hauck and Donner (1977) pointed out that Wald statistics in simple logit models can have nonmonotonic power, and Nelson and Savin (1988) found nonmonotonic power of Wald tests in logit and Tobit models as well.

Table 1
Point estimates for bond price series

	T	$\hat{T}_b$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\gamma}/\hat{\sigma}_e$
Brazil	100	1931:8	0.51	− 1.18	0.74	− 69.8
Bolivia	100	1935:11	0.52	− 2.28	0.73	− 44.8

measures the magnitude of the mean shift and is denoted by $\hat{\gamma}$. Because power of tests for a shift in mean depend on the magnitude of the mean shift relative to the variance in the errors, $AR(1)$ models were fitted to the least squares residuals. The estimated $AR(1)$ coefficient is denoted by $\hat{\alpha}$, and the estimated standard deviation of the $AR(1)$ innovations is denoted by $\hat{\sigma}_e$. Point estimates are given in Table 1. The estimated break point is in the middle of each series and the magnitudes of the mean shifts relative to the standard deviation in the $AR(1)$ innovations are -69.8 and -44.8. These are extremely large mean shifts for series that have estimated $AR(1)$ coefficients of 0.74 and 0.73.

Given the obvious shift in mean in each of the series, one should expect any reasonable test statistic to indicate that a shift in mean is present even at very small significance levels. This, however, is not always the case. Several tests for a shift in mean that allow serial correlation in the errors were computed for the series. One set of statistics is based on a static regression which uses a constant and DU_t as regressors. The other set of statistics is based on a dynamic regression which uses a constant, DU_t and a lagged dependent variable as regressors. The definitions of the statistics are given in Sections 2.2 and 2.3. The statistics are reported in Table 2 which also provides asymptotic critical values. Interestingly, in only one case can the null hypothesis of no shift in mean be rejected at even the 10% level for the Bolivia series using the static regression statistics. The null is rejected more often for Brazil but not in every case. Using the dynamic regression there is also less evidence for a shift in mean than one would expect for both series with only four rejections out of eight possible at the 10% level (three and two rejections possible at the 5% and 1% levels). Fewer rejections are possible if the tests are used conservatively with critical values appropriate for $I(1)$ errors. What is the reason for so few rejections for series that have such large mean shifts?

In this paper Monte Carlo simulations are used to compute power functions of the mean shift tests used in Table 2. There are two main contributions of the paper. First, it is shown that a wide variety of tests can have nonmonotonic power functions indicating that nonmonotonic power is a serious problem in practice. Second, sources of nonmonotonic power are uncovered. The statistics are analyzed in a unified framework that allows direct comparisons. In particular, the statistics are expressed as functions of weighted Wald statistics. It is well known that because Wald statistics are scaled by estimates of variance

Table 2
Tests for a shift in mean of the bond price series

	Panel (*a*): Static regression					
	MWS	*SWS*	*MPSW**	*SPSW**	*CUSUMS*	*QS*
Brazil	2.04	4.81	0.55	4.84	1.10	0.44
Bolivia	1.84	4.06	0.05	0.53	1.01	0.40
10% cv	2.00	9.24	0.44	2.57	1.22	0.35
5% cv	2.66	10.85	0.55	3.31	1.36	0.46
1% cv	4.21	14.49	0.82	5.01	1.63	0.74

	Panel (*b*): Dynamic regression			
	MWD	*SWD*	*CUSUMD*	*QD*
Brazil	2.39	17.33	0.70	0.16
Bolivia	3.50	20.96	0.91	0.22
10% cv ($I(0)$)	2.00	9.24	1.22	0.35
5% cv ($I(0)$)	2.66	10.85	1.36	0.46
1% cv ($I(0)$)	4.21	14.49	1.63	0.74
10% cv ($I(1)$)	3.32	18.20	na	0.39
5% cv ($I(1)$)	3.91	20.23	na	0.48
1% cv ($I(1)$)	5.35	22.64	na	0.76

Note: The critical values are asymptotic and are taken from following sources: *MWS*, *SWS*, *MWD* and *SWD* from Vogelsang (1997a), *MPSW* and *SPSW* from simulations performed by the author, *CUSUMS* and *CUSUMD* from Ploberger and Kramer (1992), and *QS* and *QD* from MacNeill (1978) and Perron (1991). $I(0)$ and $I(1)$ denote critical values when the errors are $I(0)$ and $I(1)$, respectively.

*The *MPSW* and *SPSW* statistics were configured so that the 5% asymptotic critical values are the same for both $I(0)$ and $I(1)$ errors. By construction the values of these statistics are larger if configured for a 10% test and smaller if configured for a 1% test (see Vogelsang (1997b) for details). Therefore, the null hypothesis can be rejected at the 5% and 10% levels but not at the 1% level for Brazil using *MPSW* and *SPSW*.

parameters, their power functions are very sensitive to the behavior of the variance estimates. By examining the behavior of the Wald statistics and the weights as the null and alternative grow apart, two sources of nonmonotonic power are pinpointed. The first is the behavior of estimates of the variance (as expected). If such estimates are not invariant to the shift parameter, nonmonotonic power can result. The second is the inclusion of a lagged dependent variable in the estimated regression. In this case even if the variance is assumed known, nonmonotonic power can result in some cases.

The paper is organized as follows. In Section 2 the model is introduced and tests for a shift in mean are described. The tests include the well known *CUSUM* test, the partial sum test of Gardner (1969) which was extended by MacNeill

(1978) and Perron (1991), the average Wald tests of Andrews and Ploberger (1994) and the supremum test of Andrews (1993). Section 3 reports the results of power simulations and gives explanations for the sources of nonmonotonic power. Some practical implications of the power results are examined in Section 4.

2. Testing for a shift in mean

This section presents and analyzes tests for detecting an unstable mean of a univariate time series. Once the model is introduced, two sets of statistics are considered. The first set is based on a simple static regression model. The second set is based on a dynamic regression model (a regression with a lagged dependent variable).

2.1. The model

Consider the following very simple model for a univariate time series $\{y_t\}$, $t = 1,2,\ldots,T$,

$$y_t = \mu + \gamma DU_t^c + u_t, \tag{1}$$

$$u_t = \alpha u_{t-1} + e_t, \tag{2}$$

where $DU_t^c = 1(t > T_b^c)$ and $e_t \sim$ iid $N(0, \sigma_e^2)$. Model (1) can be interpreted as a regression model (location model) with serially correlated errors. The normality assumption is made because the finite sample simulations are based on pseudo random normal deviates. Based on the point estimates for the bond price series, $\alpha = 0.7$ is used throughout the paper. Therefore, $\{u_t\}$ is modeled as a stationary $AR(1)$ process. Many of the results that follow hold qualitatively for other values of α and for more general ARMA models of $\{u_t\}$. Under the hypothesis that the mean of $\{y_t\}$ is stable, $\gamma = 0$. When $\gamma \neq 0, \{y_t\}$ has a mean shift of magnitude γ at time T_b^c.

In order to test the hypothesis $\gamma = 0$ it is necessary to estimate serial correlation nuisance parameters associated with $\{u_t\}$. Given Eq. (2), if model (1) is estimated by OLS, then it is a standard result that the point estimates of μ and γ are asymptotically normal with variances that are functions of $\omega^2 = \sigma_e^2/(1-\alpha)^2$ where ω^2 is proportional to the spectral density of $\{u_t\}$ evaluated at frequency zero (see Hamilton, 1994, p. 195). Therefore, limiting distributions of statistics based on OLS estimates of Eq. (1) depend on ω^2, and estimates of ω^2 are required in practice. Given the $AR(1)$ assumption, ω^2 could be estimated parametrically using estimates of α and σ_e^2. In practice, however, the form of ω^2 is usually unknown and nonparametric methods are used to estimate

ω^2. In this paper ω^2 is estimated nonparametrically using $\hat{\omega}^2_{\mathrm{np}} = \sum_{j=-T+1}^{T-1}[\kappa(j/L)T^{-1}\sum_{t=1}^{T-j}\hat{u}_t\hat{u}_{t+j}]$ where $\{\hat{u}_t\}$ are OLS residuals from Eq. (1) imposing $\gamma = 0, \kappa(\cdot)$ is a kernel function, and L is a truncation lag parameter. Following Andrews (1991), the quadratic spectral kernel is used with L chosen using the data dependent method based on the $AR(1)$ plug-in method.

Another approach to account for serial correlation in $\{u_t\}$ is to add lags of $\{y_t\}$ to model (1). Formally, Eq. (1) can be transformed in the usual way to give the autoregression

$$y_t = \mu^* + \gamma^* DU_t^c + \theta D(T_b^c)_t + \alpha y_{t-1} + e_t, \tag{3}$$

where $D(T_b)_t = 1(t = T_b + 1)$, $\mu^* = (1-\alpha)\mu$, $\gamma^* = (1-\alpha)\gamma$ and $\theta = \alpha\gamma$. Regression (3) has the nice property that the regression error is iid. However, as shown below, the presence of $\{y_{t-1}\}$ in the regression can have important consequences for the power of some statistics.

2.2. Tests based on the static regression

Many statistics have been proposed that can be used to test for a shift in mean of $\{y_t\}$. Attention is focused on tests designed to detect a single shift in mean at an unknown date so that T_b^c is assumed unknown. In this scenario, Andrews and Ploberger (1994) proposed efficient tests based on functionals of the Wald statistic for testing $\gamma = 0$ in the static regression (1). Let $WS(T_b)$ denote this Wald statistic. Because $\{u_t\}$ has serial correlation, the asymptotic distribution of $WS(T_b)$ depends on ω^2. To obtain a limiting distribution that is invariant to ω^2, $WS(T_b)$ is computed in the usual way, but the usual OLS estimate of the variance of $\{u_t\}$ is replaced by $\hat{\omega}^2_{\mathrm{np}}$. In this case, simple algebra gives

$$WS(T_b) = \left(\sum_{t=T_b+1}^{T} \hat{y}_t\right)^2 \Bigg/ \left(\hat{\omega}^2_{\mathrm{np}} \sum_{t=T_b+1}^{T} \widehat{DU}_t^2\right),$$

where $\hat{y}_t$ and $\widehat{DU}_t$ are the residuals from the regressions of y_t and DU_t, respectively, on a constant, and $\hat{\omega}^2_{\mathrm{np}}$ is computed as described in Section 2.1 using $\hat{u}_t = \hat{y}_t$ so that $\hat{\omega}^2_{\mathrm{np}}$ is constructed by imposing the null hypothesis, $\gamma = 0$, on regression (1).

It is useful for later developments to define the Wald statistic in the case where the parameter ω is assumed known. This exact Wald statistic is defined as

$$WS^e(T_b) = \left(\sum_{t=T_b+1}^{T} \hat{y}_t\right)^2 \Bigg/ \left(\omega^2 \sum_{t=T_b+1}^{T} \widehat{DU}_t^2\right).$$

It immediately follows that $WS(T_b) = (\omega^2/\hat{\omega}^2_{\mathrm{np}})WS^e(T_b)$; therefore, statistics that are functions of $WS(T_b)$ can be written in terms of $WS^e(T_b)$. This exercise will be useful in pinpointing sources of power behavior.

The first statistic is in a class of statistics proposed by Andrews and Ploberger (1994). Consider the set of shift dates $\Lambda = (T_b^*, T_b^* + 1, \ldots, T - T_b^*)$ where $T_b^* = [\lambda^* T]$, $[x]$ is the integer part of x and $\lambda^* \in (0,1)$ (λ^* is called the amount of trimming, and $\lambda^* = 0.01$ is used throughout the paper). Andrews and Ploberger (1994) show that the mean of $WS(T_b)$ across T_b, defined as $MWS = T^{-1}\sum_{T_b \in \Lambda} WS(T_b)$, is an optimal test. A second statistic proposed by Andrews (1993) that is not in the optimal class is the supremum of $WS(T_b)$ defined as $SWS = \sup_{T_b \in \Lambda} WS(T_b)$.

Four additional statistics are considered and were chosen because they are closely related to $WS(T_b)$. The first is the $CUSUM$ statistic proposed by Brown et al. (1975) which is based on recursive residuals. In order to draw direct comparisons to MWS and SWS, the $CUSUM$ statistic based on OLS residuals proposed by Ploberger and Kramer (1992) is used here. Following the earlier literature on the $CUSUM$ statistic, Ploberger and Kramer (1992) consider the case where the regression errors are iid. Because of serial correlation in $\{u_t\}$, regression (1) does not fall into the standard $CUSUM$ framework. However, the results obtained by Ploberger and Kramer (1992) as they pertain to regression (1) remain unchanged as long as $\hat{\omega}_{np}^2$ is used in place of the OLS estimate of the variance of $\{u_t\}$. This slight modification gives the statistic $CUSUMS = \sup_{1 \leq T_b \leq T} |T^{-1/2}\sum_{t=1}^{T_b} \hat{y}_t / \hat{\omega}_{np}|$. The second statistic, which is similar in spirit to the $CUSUM$ statistic, was proposed by Gardner (1969) and extended by MacNeill (1978) and Perron (1991) and is defined as $QS = T^{-2}\sum_{T_b=1}^{T}(\sum_{t=1}^{T_b} \hat{y}_t)^2 / \hat{\omega}_{np}^2$. The third and fourth statistics are based on a class of statistics proposed by Vogelsang (1998b) and are similar to the mean shift statistic analyzed by Vogelsang (1998a). Let $z_t = \sum_{j=1}^{t} y_j$ and $S_t = \sum_{j=1}^{t} u_j$. Transforming (1) gives

$$z_t = \mu t + \gamma DT_t + S_t, \tag{4}$$

where $DT_t = 1(t > T_b)(t - T_b)$. Let $\hat{S}_t(T_b)$ denote the OLS residuals from Eq. (4). Define the statistic $J^* = \inf_{T_b \in \Lambda} J(T_b)$ where $J(T_b)$ is the standard Wald statistic divided by T for testing $\beta_1 = \beta_2 = \cdots = \beta_9 = 0$ in the regression $y_t = \mu + \gamma DU_t + \sum_{i=1}^{9} \beta_i t^i + u_t$. The J^* statistic is related to the class of unit root statistics proposed by Park and Choi (1988) and further analyzed by Park (1990). Let $c_i(T_b) = 100\exp(b_i J^*) T^{-2}\sum_{t=1}^{T} \hat{S}(T_b)^2$ for $i = 1, 2$ where $b_1 = 1.129$ and $b_2 = 1.261$. The statistics of interest are defined as

$$MPSW = T^{-1}\sum_{T_b \in \Lambda}\left(\sum_{t=T_b+1}^{T} \hat{y}_t\right)^2 \Bigg/ \left[\left(\sum_{t=T_b+1}^{T} \hat{D}U_t^2\right) c_1(T_b)\right],$$

$$SPSW = \sup_{T_b \in \Lambda}\left(\sum_{t=T_b+1}^{T} \hat{y}_t\right)^2 \Bigg/ \left[\left(\sum_{t=T_b+1}^{T} \hat{D}U_t^2\right) c_2(T_b)\right].$$

These statistics are functions of $WS(T_b)$ with $\hat{\omega}^2_{np}$ replaced by $c_1(T_b)$ and $c_2(T_b)$. The b_i's are chosen so that the 5% asymptotic critical values are the same for both $I(0)$ and $I(1)$ errors. These statistics have the property that ω^2 need not be estimated to carry out the tests. See Vogelsang (1998b) for details on the motivation behind the construction of these statistics.

To compare the power functions of the statistics in a unified framework, the statistics are expressed as functions of weighted $WS^e(T_b)$'s. Then, by examining the behavior of the weights, explanations can be given for the shapes of the power functions. It is not hard to see how MWS, SWS, $SPSW$ and $MPSW$ can be written as functions of $WS^e(T_b)$, but things are less obvious for $CUSUMS$ and QS. Because $\{\hat{y}_t\}$ is a demeaned or centered variable, it follows that $\sum_{t=1}^{T_b}\hat{y}_t = -\sum_{t=T_b+1}^{T}\hat{y}_t$. Squaring both sides of this relationship gives $(\sum_{t=1}^{T_b}\hat{y}_t)^2 = (\sum_{t=T_b+1}^{T}\hat{y}_t)^2$. Therefore, it follows that $WS^e(T_b) = (\sum_{t=1}^{T_b}\hat{y}_t)^2/(\omega^2\sum_{t=T_b+1}^{T}\hat{D}U_t^2)$, which implies that $(\omega^2\sum_{t=T_b+1}^{T}\hat{D}U_t^2)WS^e(T_b) = (\sum_{t=1}^{T_b}\hat{y}_t)^2$. $CUSUMS$ and QS can be written in terms of $WS^e(T_b)$ by substituting for $(\sum_{t=1}^{T_b}\hat{y}_t)^2$. Table 3 provides expressions for all six statistics written as functions of weighted $WS^e(T_b)$.

2.3. Tests based on the dynamic regression

Testing for a shift in mean can also be based on the dynamic regression (3). Four statistics are considered here. Vogelsang (1997a) extensively analyzed the statistics of Andrews and Ploberger (1994) and Andrews (1993) as applied to regression (3). The mean and supremum statistics are considered here. Define

$$WD(T_b) = \left(\sum_{t=T_b+1}^{T} \tilde{y}_t\right)^2 \Bigg/ \left(s_u^2(T_b) \sum_{t=T_b+1}^{T} \tilde{D}U_t^2\right)$$

Table 3
The test statistics written as functions of weighted $WS^e(T_b)$,$WD^e(T_b)$

Statistic	Weights
$MWS = T^{-1}\sum_{T_b\in\Lambda}[(\omega^2/\hat{\omega}^2_{np})WS^e(T_b)]$	$\omega^2/\hat{\omega}^2_{np}$
$SWS = \sup_{T_b\in\Lambda}[(\omega^2/\hat{\omega}^2_{np})WS^e(T_b)]$	$\omega^2/\hat{\omega}^2_{np}$
$CUSUMS = \sup_{1\leqslant T_b\leqslant T}[(\omega^2/\hat{\omega}^2_{np})(T^{-1}\sum_{t=T_b+1}^{T}\hat{D}U_t^2)WS^e(T_b)]^{1/2}$	$(\omega^2/\hat{\omega}^2_{np})T^{-1}\sum_{t=T_b+1}^{T}\hat{D}U_t^2$
$QS = T^{-1}\sum_{T_b=1}^{T}[(\omega^2/\hat{\omega}^2_{np})(T^{-1}\sum_{t=T_b+1}^{T}\hat{D}U_t^2)WS^e(T_b)]$	$(\omega^2/\hat{\omega}^2_{np})T^{-1}\sum_{t=T_b+1}^{T}\hat{D}U_t^2$
$MPSW = T^{-1}\sum_{T_b\in\Lambda}[\omega^2c_1(T_b)WS^e(T_b)]$	$\omega^2c_1(T_b)$
$SPSW = \sup_{T_b\in\Lambda}[\omega^2c_2(T_b)WS^e(T_b)]$	$\omega^2c_2(T_b)$
$MWD = T^{-1}\sum_{T_b\in\Lambda}[(\sigma_e^2/s_u^2(T_b))WD^e(T_b)]$	$\sigma_e^2/s_u^2(T_b)$
$SWD = \sup_{T_b\in\Lambda}[(\sigma_e^2/s_u^2(T_b))WD^e(T_b)]$	$\sigma_e^2/s_u^2(T_b)$
$CUSUMD = \sup_{1\leqslant T_b\leqslant T}[(\sigma_e^2/s_r^2)(T^{-1}\sum_{t=T_b+1}^{T}\tilde{D}U_t^2)WD^e(T_b)]^{1/2}$	$(\sigma_e^2/s_r^2)(T^{-1}\sum_{t=T_b+1}^{T}\tilde{D}U_t^2)$
$QD = T^{-1}\sum_{T_b=1}^{T}[(\sigma_e^2/s_r^2)(T^{-1}\sum_{t=T_b+1}^{T}\tilde{D}U_t^2)WD^e(T_b)]$	$(\sigma_e^2/s_r^2)(T^{-1}\sum_{t=T_b+1}^{T}\tilde{D}U_t^2)$

where $\tilde{y}_t$ and $\widetilde{DU}_t$ are the residuals from regressions of y_t and DU_t, respectively, on $\{1, D(T_b)_t, y_{t-1}\}$, and $s_u^2(T_b)$ is the OLS estimate of the variance of $\{e_t\}$ based on Eq. (3). $WD(T_b)$ is the standard Wald statistic for testing $\gamma = 0$ in Eq. (3). The mean and supremum statistics are defined as $MWD = T^{-1}\sum_{T_b \in \Lambda} WD(T_b)$ and $SWD = \sup_{T_b \in \Lambda} WD(T_b)$. Ploberger and Kramer (1992) showed that the OLS residual based $CUSUM$ test remains valid in regressions with lagged dependent variables[2] (note that Eq. (3) falls within their framework because $\{e_t\}$ is iid). Define $CUSUMD = \sup_{1 \leqslant T_b \leqslant T} |T^{-1/2}\sum_{t=1}^{T_b} \tilde{y}_t / s_r|$ where s_r^2 is the OLS estimate of the variance of $\{e_t\}$ based on regression (3) with the restriction $\gamma = 0$ imposed. Note that s_r^2 is not a function of T_b. Finally, Perron (1991) extended the results of Gardner (1969) and MacNeill (1978) to dynamic regressions and proposed the statistic $QD = T^{-2}\sum_{T_b=1}^{T}(\sum_{t=1}^{T_b}\tilde{y}_t)^2/s_r^2$.

Define $WD^e(T_b) = (\sum_{t=T_b+1}^{T}\tilde{y}_t)^2/(\sigma_e^2\sum_{t=T_b+1}^{T}\widetilde{DU}_t^2)$ which is the Wald statistic in the case where the variance of $\{e_t\}$ is assumed known. Using similar arguments as for the static regression statistics, the dynamic regression statistics can be expressed as functions of weighted $WD^e(T_b)$. These formulas are also given in Table 3.

3. Finite sample power

In this section finite sample power of the above statistics is analyzed with attention focused on the possibility of nonmonotonic power. In cases where power is nonmonotonic, sources of the nonmonotonicities are uncovered by examining the weighting functions given in Table 3. The data generating process (DGP) used is,

$$y_t = \gamma DU_t^c + u_t, \quad u_t = 0.7u_{t-1} + v_t, \quad u_0 = 0, \tag{5}$$

with v_t iid N(0, 1) random deviates and $T_b^c = 50$. The sample size was $T = 100$. The same set of random deviates was used in all simulations with an initial seed of 100. The random number generator was the rndns( ) random number generator from GAUSS. The number of replications was 2000 in all cases. The parameter μ was set to zero as the statistics are exactly invariant to μ. Power was simulated for $\gamma = 0, 1, 2, \ldots, 10$. Because the standard deviation of v_t is one, γ is measured in terms of standard deviation units.

The tests were carried out using 5% asymptotic critical values taken from Table 2. For the statistics based on the static regression (1), $I(0)$ asymptotic critical values were used. $I(0)$ critical values are appropriate because $\{u_t\}$ is an

[2] See Kramer et al. (1988) and Ploberger et al. (1989) for results on the standard $CUSUM$ test (based on recursive residuals) applied to models with lagged dependent variables.

$I(0)$ process. Note that with the exception of the $MPSW$ and $SPSW$ statistics, the statistics in the static regression model are invalid when the errors are $I(1)$ as the statistics diverge to ∞ as T increases. On the other hand, the $MPSW$ and $SPSW$ statistics are designed to be valid for both $I(0)$ and $I(1)$ errors. For the statistics based on the dynamic regression (3) the $I(0)$ critical values were used for all four statistics. Vogelsang (1997a) and Perron (1991) showed that when $I(1)$ critical values are used for MWD, SWD, and QD, the tests are conservative in the case where it is unknown whether $\{u_t\}$ is $I(0)$ or $I(1)$. Because this is a common situation in practice, power of these statistics was also simulated using $I(1)$ critical values but is reported only for MWD and SWD as power of QD is very similar for $I(0)$ and $I(1)$ critical values.

Power functions of the static regression statistics are plotted in Fig. 2, and power functions of the dynamic regression statistics are plotted in Fig. 3. The statistics $SPWS$, $MPSW$, SWD and MWD ($I(0)$ critical value) exhibit monotonic power while the statistics SWS, MWS, $CUSUMS$, QS, MWD ($I(1)$ critical value), $CUSUMD$, and QD exhibit nonmonotonic power. In some cases

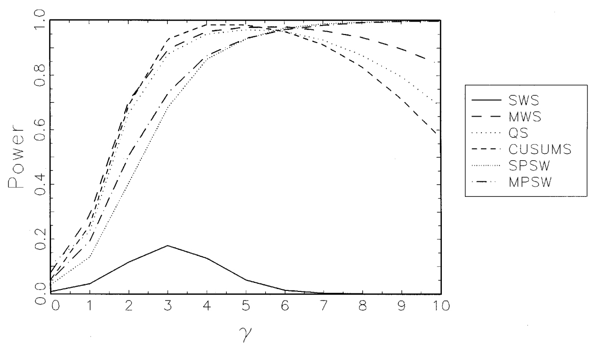

Fig. 2. Power against a mean shift at time 50, static regression.

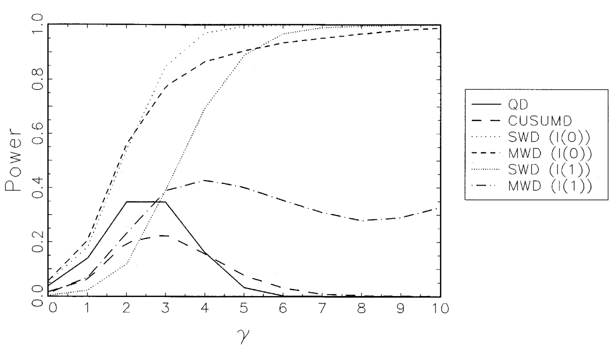

Fig. 3. Power against a mean shift at time 50, dynamic regression.

power actually drops to zero for moderate to large mean shifts (SWS, $CUSUMD$ and QD). In unreported simulations it was found that power of MWS, $CUSUMS$, and QS continues to fall as γ increases and eventually reaches zero for large γ. These power results explain many of the patterns seen in the empirical example.

What is the source of nonmonotonic power? It is well known that estimates of variance parameters can have important effects on power of test statistics. Harvey (1975) pointed out this fact when the $CUSUM$ statistic was first proposed. Indeed, in Fig. 2 the statistics which have nonmonotonic power require an estimate of ω^2, while the statistics that have monotonic power do not require such an estimate. In Fig. 3 no obvious pattern emerges because all the statistics require estimates of σ_e^2.

To illustrate the effects that estimates of variance parameters have on power, power was simulated under the assumption that ω^2 and σ_e^2 are known. The resulting power functions are depicted in Figs. 4 and 5. In the static regression model all of the statistics now have monotonic power indicating $\hat{\omega}_{np}^2$ is the

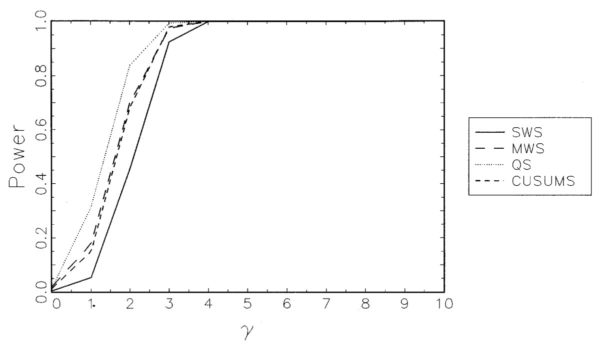

Fig. 4. Power against a mean shift at time 50, ω^2 assumed known, static regression.

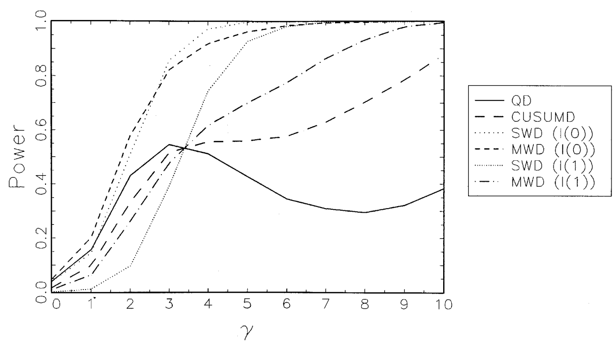

Fig. 5. Power against a mean shift at time 50, σ_e^2 assumed known, dynamic regression.

primary source of nonmonotonic power. In the dynamic regression all but the QD statistic have monotonic power. So, while estimates of σ_e^2 certainly contribute to the possibility of nonmonotonic power in the dynamic regression, other sources of nonmonotonic power remain.

Because the statistics are implicitly functions of weighted $WS^e(T_b)$ and $WD^e(T_b)$, sources of nonmonotonic power can be isolated by examining the behavior of $WS^e(T_b)$, $WD^e(T_b)$, and the weights as γ increases. To illustrate how $WS^e(T_b)$ and $WD^e(T_b)$ behave as γ increases the following simulations were performed. Given γ, a series was generated according to Eq. (5), and $WS^e(T_b)$ and $WD^e(T_b)$ were computed for each T_b. This was repeated 2000 times and the averages of $WS^e(T_b)$ and $WD^e(T_b)$ for each T_b were recorded. This experiment was carried out for $\gamma = 0, 2, 4, \ldots, 10$. The results are plotted in Figs. 6 and 7 for $WS^e(T_b)$ and $WD^e(T_b)$, respectively. As γ increases, $WS^e(T_b)$ increases on average regardless of T_b, and the increases are larger the closer T_b is to the true shift date of 50. A similar pattern holds for $WD^e(T_b)$ except that $WD^e(T_b)$ decreases slightly on average for $T_b > 50$, and $WD^e(T_b)$ is fairly insensitive to

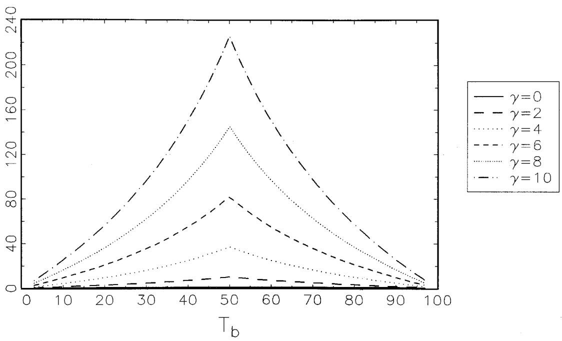

Fig. 6. Averages of $WS^e(T_b)$, static regression.

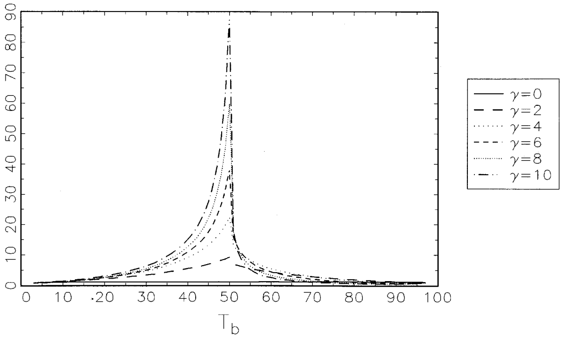

Fig. 7. Averages of $WD^e(T_b)$, dynamic regression.

γ for T_b far from 50. This suggests that as long as the weights are not decreasing in γ, then the statistics will be, on average, increasing in γ and power will be monotonic.

Consider the weights for the statistics in the static regression. Attention is focused on the *SWS*, *MWS*, *CUSUMS*, and *QS* statistics because they have nonmonotonic power. From Table 3 we see that the weights depend on γ only through $(\omega^2/\hat{\omega}^2_{np})$ which does not depend on T_b. Mean values of $\omega^2/\hat{\omega}^2_{np}$ were simulated for $\gamma = 0, 2, 4, \ldots, 10$ yielding 2.066, 0.9705, 0.2875, 0.1133, 0.0573 and 0.040, respectively. Therefore, power can be nonmonotonic because the weights are decreasing on average as γ increases. This result simply reiterates the fact that power of the static regression statistics is sensitive to the way in which ω^2 is estimated.

In the dynamic regression things are more complicated. The weights of *MWD* and *SWD* depend on γ through $s^2_u(T_b)$ and the dependence on γ is different for each value of T_b. The weights of *CUSUMD* and *QD* depend on γ through s^2_r which is constant across T_b, but the weights also depend on γ through $T^{-1}\sum_{t=T_b+1}^{T}\tilde{D}U_t^2$ because $\tilde{D}U_t^2$ depends on y_{t-1} (and y_{t-1} obviously depends on γ) and this dependence varies across T_b. In the same simulations used to examine $WS^e(T_b)$ and $WD^e(T_b)$, for each T_b (given γ), average weights (across the 2000 replications) were recorded for the dynamic regression statistics.

The average weights are plotted in Fig. 8 for *MWD* and *SWD*, in Fig. 9 for *CUSUMD* and *QD*, and in Fig. 10 for *CUSUMD* and *QD* assuming that σ^2_e is known. Several interesting patterns emerge from the figures. For *MWD* and *SWD* the weights are invariant to γ when $T_b = 50$ because $s^2_u(50)$ is exactly invariant to γ, but for $T_b \neq 50$ the weights are decreasing in γ (see Fig. 8). The reason that *SWD* has monotonic power, but *MWD* may have nonmonotonic power is now clear. *SWD* only depends on the largest weighted $WD^e(T_b)$. Because $WD^e(50)$ is increasing in γ, and the weight is invariant to γ, *SWD* is increasing in γ, and power is monotonic. *MWD*, on the other hand, is an average

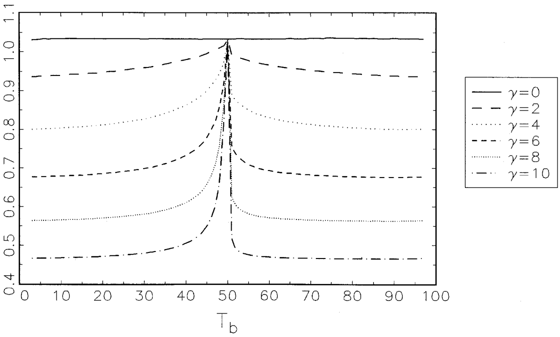

Fig. 8. Average weights of *MWD* and *SWD* mean shift at time 50, dynamic regression.

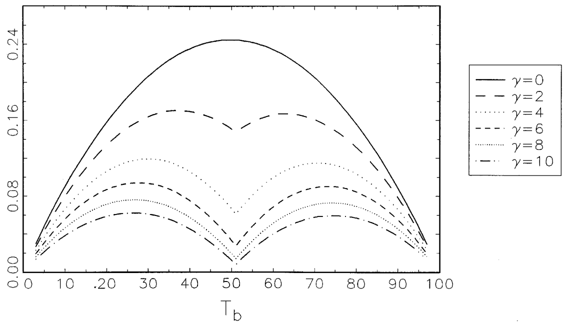

Fig. 9. Average weights of *QD* and *CUSUMD* mean shift at time 50, dynamic regression.

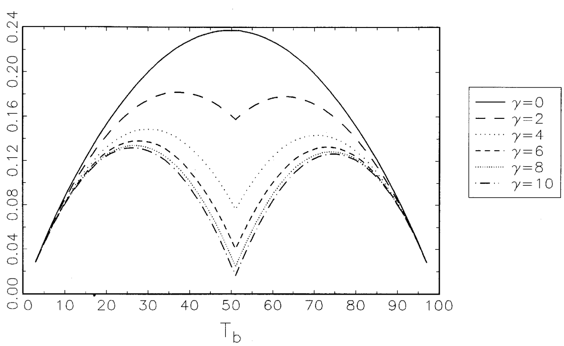

Fig. 10. Average weights of *QD* and *CUSUMD* mean shift at time 50, σ_e^2 assumed known, dynamic regression.

of all the weighted $WD^e(T_b)$'s. While the contribution to MWD from $WD^e(50)$ is increasing in γ, the contribution to MWD for $T_b \neq 50$ can be decreasing in γ because the weights are decreasing in γ while $WD^e(T_b)$ is not that sensitive to γ. Therefore, MWD can be decreasing in γ and have nonmonotonic power.

The story for the $CUSUMD$ and QD statistics is much simpler. For these statistics the weights are decreasing in γ on average for all values of T_b. Interestingly, the decrease occurs most rapidly at the true shift date of 50. So, as γ increases, less and less weight is placed on the $WD^e(T_b)$ statistic that increases most as γ increases. Therefore, almost any function of the weighted $WD^e(T_b)$'s (including the sup) can be decreasing in γ resulting in nonmonotonic power. The qualitative pattern of the weights remains the same even in the case where σ_e^2 is assumed known as shown in Fig. 10. This occurs because of the dependence of the weights on γ from the presence of $\{y_{t-1}\}$ in regression (6). This explains how power of QD can be nonmonotonic even when σ_e^2 is assumed known.

The role played by $\{y_{t-1}\}$ can also be explained in another way. From the results of Perron (1990) it is well known that ignoring a mean shift in an autoregression biases the estimate of α towards 1, and the estimated model appears $I(1)$ without a mean shift. If a shift in mean is allowed in the autoregression, but the date of the shift is different from the true shift date, then similar bias results. Bias is, therefore, present in the fitted dynamic regression when the wrong shift date is used. This bias can cause nonmonotonic power for statistics that are functions of estimates obtained from regressions using shift dates different from the true shift date.

4. Practical implications of the power results

The above results suggest that testing for a shift in mean can be problematic in practice depending on which of the statistics is used. In the static regression, the $MPSW$ and $SPSW$ statistics are useful as they have monotonic power and have comparable power to the other statistics in the range where power is monotonic. The other statistics are potentially problematic because they might not detect large mean shifts. Because in the static regression the source of nonmonotonic power is the estimate of ω^2, a potential solution is to use an estimate of ω^2 that is based on the model estimated under the alternative. If the true shift date were known, then an estimate of ω^2 could be constructed that is exactly invariant to γ. But, in practice the true shift date is often unknown. In this situation one approach is to consistently estimate the shift date and use residuals from Eq. (1) to construct $\hat{\omega}^2_{\mathrm{np}}$. However, in unreported simulations it was found that this approach results in size distorted tests with exact size in excess of 0.2. A potentially more serious problem when using the static regression with economic data is that the MWS, SWS, $CUSUMS$ and QS statistics diverge when the errors are $I(1)$. If the errors are $I(1)$ or nearly $I(1)$, the tests become oversized. The $MPSW$ and $SPSW$ statistics avoid this problem because they are designed to be robust to $I(1)$ errors.

In the dynamic regression, the SWD statistic is useful as it has monotonic power. And, the SWD statistic provides an estimate of the shift date. The MWD statistic is potentially problematic because it can have nonmonotonic power. On the other hand, Fig. 5 shows that MWD can be more powerful than SWD when the mean shift is small in magnitude. A statistic related to MWD that has monotonic power is the exponential average statistic of Andrews and Ploberger (1994). See Vogelsang (1997a) for a detailed discussion of this statistic applied to regression (3). The QD and $CUSUMD$ statistics are problematic as they suffer from nonmonotonic power and have power of nearly zero for very large breaks. The power properties are unlikely to be improved substantially by considering other estimates of the variance in place of s_r^2 because, as Fig. 5 shows, nonmonotonic power is still present even when the exact variance is assumed known.

Finally, it is important to note that nonmonotonic power can occur in models with different types of mean shifts, models with multiple means shifts, and models with shifts in higher-order trends. In the working paper, Vogelsang (1997b), it was shown that the MOSUM statistic proposed by Chu et al. (1995) can have nonmonotonic power in models with serially correlated errors. The MOSUM statistic is designed to detect a shift in mean that is temporary and lasts for only a finite time. If one allows for the possibility of two shifts in mean, then any of the statistics considered in this paper can exhibit nonmonotonic power. See Vogelsang (1997a,b) illustrations of this fact. It was also shown in Vogelsang (1997b) that tests for a shift in the slope of a trending time series can have nonmonotonic power (see in addition Perron (1991) and Vogelsang (1997a)). For example, the *QD* statistic, the mean Wald statistic, and the statistics proposed by Chu and White (1992) can have nonmonotonic power functions. Obviously, nonmonotonic power is a potential problem in general when testing for shifts in trend functions of series with serially correlated errors.

Acknowledgements

I am grateful to Terence Chong, Robin Lumsdaine, two anonymous referees and seminar participants at Rice University, University of Houston, Penn State University and the 1997 Winter Meetings of the Econometric Society for helpful comments on earlier drafts. I thank the Center for Analytic Economics at Cornell University.

References

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383–1414.

Bai, J., 1994. Least squares estimation of a shift in linear processes. Journal of Time Series Analysis 15, 453–472.

Brown, R.L., Durbin, J., Evans, J.M., 1975. Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society B 37, 149–163.

Commercial and Financial Chronicle, 1928–1940. William B. Dana Company, New York. (various issues).

Chu, C.-S.J., Hornik, K., Kuan, C.-M., 1995. MOSUM tests for parameter constancy. Biometrika 82, 603–617.

Chu, C.-S.J., White, H., 1992. A direct test for changing trend. Journal of Business and Economic Statistics 10, 289–300.

Gardner, L.A., 1969. On detecting changes in the mean of normal variates. Annals of Mathematic Statistics 40, 116–126.

Harvey, A.C., 1975. Comments on the paper by Brown, Durbin and Evans. Journal of the Royal Statistical Society B 37, 179–180.

Hauck, W.W., Donner, A., 1977. Wald's test as applied to hypotheses in logit analysis. Journal of the American Statistical Association 72, 851–853.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.

Kramer, W., Ploberger, W., Alt, R., 1988. Testing for structural change in dynamic models. Econometrica 56, 1355–1369.

MacNeill, I.B., 1978. Properties of sequences of partial sums of polynomial regression residuals with applications to test for change of regression at unknown times. Annals of Statistics 6, 422–433.

Nelson, F.D., Savin, N.E., 1988. The nonmonotonicity of the power of the Wald test in nonlinear models. Department of Economics Working Paper Series No. 88-7, University of Iowa.

Nelson, F.D., Savin, N.E., 1990. The danger of extrapolating asymptotic local power. Econometrica 58, 977–981.

Park, J.Y., 1990. Testing for unit roots and cointegration by variable addition. In: Fomby, B., Rhodes, F. (Eds.), Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots. Jai Press, London, pp. 107–134.

Park, J.Y., Choi, B., 1988. A new approach to testing for a unit root. Center for Analytic Economics, Working Paper #88-23, Cornell University.

Perron, P., 1990. Testing for a unit root in a time series with a changing mean. Journal of Business and Economic Statistics 8, 153–162.

Perron, P., 1991. A test for changes in a polynomial trend function for a dynamic time series. Mimeo. Princeton University Department of Economics.

Ploberger, W., Kramer, W., Alt, R., 1989. A modification of the CUSUM test in the linear regression model with lagged dependent variables. Empirical Economics 14, 65–75.

Ploberger, W., Kramer, W., 1992. The CUSUM test with OLS residuals. Econometrica 60, 271–285.

Vogelsang, T.J., 1997a. Wald-type tests for detecting shifts in the trend function of a dynamic time series. Econometric Theory 13, 818–849.

Vogelsang, T.J., 1997b. Sources of nonmonotonic power when testing for a shift in the trend of a dynamic time series. Center for Analytic Economics. Working Paper, #97-4. Cornell University.

Vogelsang, T.J., 1988a. Testing for a shift in mean without having to estimate serial correlation parameters. Journal of Business and Economic Statistics 16, 73–80.

Vogelsang, T.J., 1998b. Trend function hypothesis testing in the presence of serial correlation. Econometrica 66, 123–148.