Testing for Parameter Instability in Linear Models

Bruce E. Hansen, *University of Rochester*

Simple tests for parameter instability are presented and discussed. These tests have locally optimal power and do not require *a priori* knowledge of "the breakpoint." Two empirical examples are presented to illustrate the use of the tests. The first examines whether an AR(1) model for annual U.S. output growth rates has remained stable over 1889–1987. The second examines the stability of an error correction model for an aggregate life cycle model of consumption.

1. INTRODUCTION

Model stability is necessary for prediction and econometric inference. Because a parametric econometric model is completely described by its parameters, model stability is equivalent to parameter stability. Model instability may be caused simply by the omission of an important variable, or be due to some kind of "regime shift." While model instability generically makes it difficult to interpret regression results, it is of particular importance in policy analysis to know if econometric models are invariant to possible policy interventions. Engle et al. (1983) incorporated parameter invariance to interventions in their definition of super exogeneity, a condition they argued was a necessary precondition for predictive policy experiments. A necessary condition for super exogeneity is within sample parameter constancy.

Because of the well-recognized need for stable models, a large literature has emerged developing tests of model stability. The number and variety of testing procedures are quite surprising. Unfortunately, not all tests are equal, and many, developed from ad hoc criteria, are quite poor. Ideally, a test should have known size and possess the

---

Address correspondence to Bruce E. Hansen, Department of Economics, University of Rochester, Rochester, NY 14627.

I would like to thank Neil Ericsson and a referee for helpful comments on an earlier draft which improved the exposition and analysis. Research support was provided by NSF Grant No. SES 9022176.

Received January 1991; final draft accepted August 1991.

0161-8938/92/$5.00

maximal power against the alternative of interest for all tests of the same size. In practice, ideal tests rarely exist. Asymptotic theory may be necessary to approximate the null distribution, and direct power comparisons may be impossible. A useful criterion is *local power*, that is, the slope of the power function at the null hypothesis (in the direction of interest). Tests that have maximal local power can be derived (see Cox and Hinkley, 1974), and the local power function can be approximated via the asymptotic local power function.

One of the most common tests in applied econometrics is Chow's (1960, pp. 595–599) simple split-sample test.[1] This test is designed to test the null hypothesis of constant parameters against an alternative of a one-time shift in the parameters at some *known* time. There are several simple ways to calculate this test statistic, each of which involves splitting the sample at the hypothesized time of structural change. One method is to compare (using the appropriate covariance matrix) the estimates obtained from each subsample. A second method is to calculate the estimates using just the data from one subsample, and compare this with the estimates using the full sample. A third method uses dummy variables (intercept and slope) for one subsample, and then tests the significance of the dummy variables. These methods are essentially equivalent and easy to use in practice.

A severe problem arises with this Chow test, however, in the need to select the timing of the structural change that occurs under the alternative hypothesis. The problem is that the date of structural change is not defined (has no meaning) under the null hypothesis, and standard testing theory is not applicable; see Davies (1977, 1987) on this point. A researcher has several options open. First, the timing can be selected in an arbitrary way, such as at the sample midpoint. This solution effectively eliminates the dating question, but is ad hoc and would not be expected to have particularly good power against many alternatives of interest. Second, the data and/or regression residuals can be plotted for indications of structural change. If this is done, the timing is selected conditional on the data and the conventional χ^2 approximation for the distribution of the resulting test statistic is invalid. What may in effect be done is to select a candidate breakpoint that suggests a structural change, when in fact none may have occurred. Third, the date of structural change may be selected by appeal to events known a priori. If this approach is adopted, it is essential that the researcher can argue

[1]Chow's test is a straightforward application of the analysis of covariance, a standard statistical method dating back to Fisher (1922). Chow (1960) appears to have been the first, however, to apply the principle to testing stability in a time series context.

that the events are selected exogenously. For example, the oil shock of 1974 is "known" to be associated with a slowdown in aggregate output in many countries. This is known simply because a slowdown did occur after the oil shock. Thus it is impossible to test (using conventional theory) whether the oil shock had an effect upon the GNP process, because the selection of 1974 has been made *after* the data has been informally examined. Fourth, tests for structural change for every breakpoint could be calculated, and the largest test statistic examined. This is the test proposed originally by Quandt (1960), but not used much because of the lack of a distributional theory. This theory has been given recently in Andrews (1990), Chu (1989), and Hansen (1990). This procedure is theoretically sound, but may be a computational burden in some cases.

Another commonly applied stability test is the "predictive-failure" test derived by Chow (1960, pp. 594–595)[2] and routinely used by Hendry and his coauthors.[3] This test also requires an a priori selection of a breakpoint, and therefore suffers the same problem as the Chow test.

Recognizing the need for tests that reveal model instability of general form, Brown, Durbin and Evans (1975) proposed the CUSUM test, which was fairly widely programmed and used in the late 1970s and early 1980s. Theoretical investigations eventually revealed that the CUSUM test is essentially a test to detect instability in the intercept alone (see, for example, Krämer, Ploberger, and Alt, 1988). Another test proposed from a similar motivation is the CUSUM of squares test. This test, however, has poor asymptotic power against instability in the regression coefficients (Ploberger and Krämer, 1990). Instead, the CUSUM of squares test can be viewed as a test for detecting instability in the variance of the regression error. For an analysis of the power of these and other tests, see Hansen (1991a).

Below we describe a simple yet powerful test for parameter instability. The statistic has a long history in theoretical statistics and econometrics. The test was independently proposed for the Gaussian linear model by Gardner (1969), Pagan and Tanaka (1981), Nyblom and Makelainen (1983), and King (1987). The extensions to nonlinear maximum likelihood and general econometric problems were made by

[2]Although Chow developed the predictive failure test, he argued (1960, p. 598) that it would not be as powerful as the analysis-of-covariance test when the number of observations in the prediction interval exceeds the number of coefficients.

[3]See, for example, Hendry and Richard (1982), Engle and Hendry (1990), Hendry (1989), and Hendry and Ericsson (1991).

Nyblom (1989) and Hansen (1990), respectively. The test is approximately the Lagrange multiplier test (or locally most powerful test) of the null of constant parameters against the alternative that the parameters follow a martingale. This alternative incorporates simple structural breaks of unknown timing as well as random walk parameters. These tests can be developed for any econometric model, although this paper concentrates on least squares regression. The analysis includes both static and dynamic regression, for no special treatment of lagged dependent variables is required. It is necessary to exclude, however, nonstationary regressors. That is, we exclude unit root processes and deterministic trends; otherwise, a different distributional theory applies (Hansen, 1991b). One caveat should be noted. The tests discussed here simply test the null of constancy. They are not designed for determining the timing of a "structural break" if one has occurred. Other methods should be used for this purpose, and will not be discussed in this paper.

Section 2 describes the test statistics and discusses their interpretation. Section 3 uses these tests to study U.S. GNP. Section 4 examines error correction models and a simple aggregate consumption equation. Section 5 concludes.

2. INSTABILITY TESTS

We will examine the standard linear regression model

$$
\begin{aligned}
y_t &= \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots \beta_m x_{mt} + e_t, \qquad (1)\\
&= \beta' x_t + e_t\\
E(e_t|x_t) &= 0,\\
E(e_t^2) &= \sigma_t^2,\\
\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \sigma_t^2 &= \sigma^2.
\end{aligned}
$$

Throughout this paper, we will maintain the assumption that the variables $\{x_t, e_t\}$ are weakly dependent processes. That is, the variables do not contain deterministic or stochastic trends (such as unit roots). See Andrews (1990) or Hansen (1990) for the technical details, and Hansen (1991b) for an analysis in the presence of trends in the regressors. We are interested in testing the assumed constancy of the model parameters (β, σ^2). Nyblom (1989) has shown how to derive the locally most powerful test of the hypothesis of constancy against the alternative that the parameters follow a martingale process. We follow this approach here.

Equation 1 is estimated by least squares, yielding the parameter estimates $(\hat{\beta}, \hat{\sigma}^2)$ and the first-order conditions

$$
\begin{aligned}
0 &= \sum_{t=1}^{n} x_{it}\hat{e}_t \qquad i = 1, \ldots, m \\
0 &= \sum_{t=1}^{n} (\hat{e}_t^2 - \hat{\sigma}^2)
\end{aligned} \tag{2}
$$

where $\hat{e}_t = y_t - x_t'\hat{\beta}$. We can rewrite this by defining

$$
f_{it} = \begin{cases} x_{it}\hat{e}_t, & i = 1, \ldots, m \\ \hat{e}_t^2 - \hat{\sigma}^2, & i = m + 1 \end{cases} \tag{3}
$$

so Equation 2 is equivalent to

$$
0 = \sum_{t=1}^{n} f_{it}, \qquad i = 1, \ldots, m + 1. \tag{4}
$$

The variables $\{f_{it}\}$ are the first-order conditions (*scores* in a maximum likelihood context). Our test statistics are based upon the cumulative first-order conditions, given by

$$
S_{it} = \sum_{j=1}^{t} f_{ij}. \tag{5}
$$

Note that $S_{in} = 0$ by the first-order condition Equation 4. Also note that the sums S_{it} are functions of the full sample estimates, unlike the classic CUSUM of Brown, Durbin, and Evans (1975) (BDE).

We are interested in testing the stability of each parameter individually and the stability of all the parameters jointly. First, the individual stability test statistics[4] are given by

$$
L_i = \frac{1}{nV_i} \sum_{t=1}^{n} S_{it}^2 \tag{6}
$$

where

$$
V_i = \sum_{t=1}^{n} f_{it}^2. \tag{7}
$$

To obtain the statistic for the joint stability test, it will be most convenient to use matrix notation. Define the vectors

[4]See Nyblom (1989) or Hansen (1990) for the derivations of the test statistic.

$$f_t = (f_{1t}, \ldots, f_{m+1t})' \qquad (8)$$
$$S_t = (S_{1t}, \ldots, S_{m+1t})'.$$

The joint stability test statistic is

$$L_C = \frac{1}{n} \sum_{t=1}^{n} S_t' V^{-1} S_t, \qquad (9)$$

where

$$V = \sum_{t=1}^{n} f_t f_t'. \qquad (10)$$

Note that the V_i in Equation 7 are the diagonal elements of V in Equation 10.

Expression 9 shows that the test statistic is essentially an average of the squared cumulative sums of first-order conditions. Under the null hypothesis, the first-order conditions are mean zero, and their cumulative sums will tend to wander around zero (in the manner of a tied-down random walk). Under the alternative hypothesis of parameter instability, however, the cumulative sums will develop a nonzero mean in parts of the sample, so S_n will not behave like a random walk, and the test statistic will tend to be large. Thus, the test is to reject the null hypothesis of stability for large values of L_C. The test can be shown to have asymptotic local power against any nonstationary (long-run) movements in β and/or σ^2. This is unlike the CUSUM test, which has asymptotic local power only against movements in $\beta' E(x)$, or the CUSUM of squares test, which has asymptotic local power only against movements in σ^2.[5] In fact, if one regressor is a constant (say x_{1t}), then its associated test statistic (L_1) is an analog of the BDE CUSUM test. Similarly, the test for the stability of the error variance (L_{m+1}) is an analog of the BDE CUSUM of squares test. We can therefore think of the L_i family of tests as expanding, rather than replacing, the CUSUM family of tests.

The following facts about the above construction should be kept in mind. The L_C test statistic given in Equations 9 and 10 is asymptotically robust to heteroskedasticity because the matrix V is exactly the central component in the heteroskedasticity-robust covariance matrix estimator of White (1980). Similarly, the test statistic could be made robust to residual serial correlation as in White and Domowitz (1984). We are

[5]See Krämer, Ploberger, and Alt (1988), Ploberger and Krämer (1990), and Hansen (1991a) for analytic studies of the power of these tests

explicitly discussing testing via ordinary least squares (OLS) estimation, while in applications generalized least squares (GLS) methods (either for heteroskedasticity or serial correlation) are routinely used. If Equation 1 represents the *transformed* model (after a Cochrane–Orcutt correction or weighting has been made) then all the above analysis is still valid, if the tests are applied to the transformed data. Similarly, in an application involving two-stage least squares (2SLS) estimation, Equation 1 could represent the *transformed* model (after the endogenous regressors have been replaced by their predicted values). The only caution in this case is that the covariance matrix estimate V needs to be calculated differently.

The asymptotic distribution theory for the stability tests has been given by Nyblom (1989) and Hansen (1990). The distributional theory is nonstandard, but depends only upon the number of parameters tested for stability. In the joint tests there are $m + 1$ parameters: m regression parameters (which include the intercept if one of the regressors is a constant) and the error variance. Asymptotic critical values are presented in Table 1. The first line of Table 1 gives the relevant critical values for the individual stability tests: for these tests there is only one "degree of freedom." Note that the 5 percent significance level for the individual stability test is 0.470. This suggests the informal rule of finding an individual stability test "significant" if the test statistic exceeds one-half, much as we commonly find a t statistic "significant" if its value exceeds 2.0.

One may interpret the individual and joint statistics quite similarly to the interplay between individual and joint tests of significance of regression coefficients. In the latter case, t statistics for each regression coefficient give information regarding the significance of each individual variable, while the F statistic gives the joint significance of all the variables. The difference in the present context is that a significant statistic is bad news, indicating possible instability. It is important as well not to abuse such information. If a large number of parameters are estimated, it should not be surprising to find a small number of "significant" test statistics for individual instability. The joint significance test is a more reliable guide in this context.

One important question that most analysts ask is: If we reject stability, what then? It is important to emphasize that there can be no universal answer, or solution, to the problem of unstable coefficients. Frequently, a significant instability test indicates some form of model misspecification. For example, omitted variables can induce parameter variation. If an alternative specification appears free of this problem, then it seems reasonable to adopt the alternative specification. In prac-

Table 1: *Asymptotic Critical Values for L_c*

	Significance level					
Degrees of freedom (m + 1)	**1%**	**2.5%**	**5%**	**7.5%**	**10%**	**20%**
1	0.748	0.593	0.470	0.398	0.353	0.243
2	1.07	0.898	0.749	0.670	0.610	0.469
3	1.35	1.16	1.01	0.913	0.846	0.679
4	1.60	1.39	1.24	1.14	1.07	0.883
5	1.88	1.63	1.47	1.36	1.28	1.08
6	2.12	1.89	1.68	1.58	1.49	1.28
7	2.35	2.10	1.90	1.78	1.69	1.46
8	2.59	2.33	2.11	1.99	1.89	1.66
9	2.82	2.55	2.32	2.19	2.10	1.85
10	3.05	2.76	2.54	2.40	2.29	2.03
11	3.27	2.99	2.75	2.60	2.49	2.22
12	3.51	3.18	2.96	2.81	2.69	2.41
13	3.69	3.39	3.15	3.00	2.89	2.59
14	3.90	3.60	3.34	3.19	3.08	2.77
15	4.07	3.81	3.54	3.38	3.26	2.95
16	4.30	4.01	3.75	3.58	3.46	3.14
17	4.51	4.21	3.95	3.77	3.64	3.32
18	4.73	4.40	4.14	3.96	3.83	3.50
19	4.92	4.60	4.33	4.16	4.03	3.69
20	5.13	4.79	4.52	4.36	4.22	3.86

Source: Hansen (1990), Table 1.

tice, researchers always estimate a variety of specifications in an attempt to "fit" the data. Instability test statistics can be added to the standard bag of tricks to determine the worth of a particular specification, if used cautiously. Of course, if one *searches* over a variety of specifications to find one whose stability test statistic is "insignificant," a form of data mining has been performed, and the credibility of the resulting regression is suspect.

A more troubling alternative is to explicitly allow the parameters to change over the course of the sample. The most common method is to use intercept and slope dummies to capture "regime shifts." By including dummies, the analyst is admitting that the estimated relationship is not constant, but is attempting to go ahead with the analysis. The immediate questions are the following: Of what use are these regression results? Do they have *any* predictive content? If events can arise which shift the regression slopes in an arbitrary way, how can we exclude such events from arising in the future?

If it is believed that the regression coefficients are shifting because of economic events, then it is (at least in principle) possible to explicitly model these events. Perhaps simply allowing for interaction among regressors (which can be thought of as allowing the regression coefficients to depend linearly upon other regressors) will adequately capture the "parameter shifts," but more complicated interactions may be required in particular applications.

A final alternative is to use a so-called random coefficient model such as that of Cooley and Prescott (1976). Cooley and Prescott model the regression coefficients as random walks that are independent of the regression error. Such models are highly nonlinear and can be estimated by maximum likelihood. A fundamental feature of this approach is that the coefficient variation is not modeled explicitly. Indeed, the specification is quite nonparametric in spirit. This allows for a large degree of flexibility, but with a reduction of precision vis-à-vis more structural approaches. If the alternative is to include slope dummies for the points of "structural change", then the Cooley–Prescott approach has a natural advantage as it incorporates uncertainty over future values of the parameters into estimated prediction intervals, while inclusion of slope dummies does not. Incidentally, it should be noted that the "meta-parameters" in random coefficient models are of course assumed constant, and the methods described in Nyblom (1989) could be used to design powerful tests of this hypothesis, if desired.

3. HAS THE NATURE OF OUTPUT FLUCTUATIONS CHANGED?

A major debate in macroeconomics concerns the stability of the process describing aggregate output. A common argument by Keynesians is that institutional changes have decreased output fluctuations in the postwar period (see, for example, Tobin, 1980, p. 47). This view has been challenged recently by Romer (1986a, 1986b, 1989), who argues that the perceived decrease in output fluctuations may be a figment of poorly measured prewar data. It is also commonly asserted that the behavior of output during the great depression is fundamentally different than in neighboring years. DeLong and Summers (1988), for example, argue that shocks to GNP were more persistent during the depression than in the predepression and post-World War II years. All of these debates concern whether or not the distribution of the output process has changed. We follow DeLong and Summers in using the corrected GNP data provided by Romer (1989) and excluding the pre-1888 data as unreliable.

Table 2: GNP Equation

	α_0	α_1	σ^2
Sample period			
1889–1987	0.012	0.36	0.0022
	(0.005)	(0.11)	(0.0004)
1889–1929	0.021	−0.12	0.0013
	(0.006)	(0.13)	(0.0003)
1930–1947	0.011	0.60	0.0054
	(0.018)	(0.14)	(0.0015)
1948–1987	0.019	0.05	0.0006
	(0.005)	(0.15)	(0.0001)
Individual L_C	0.03	0.19	0.45*
Joint L_C = 0.72	$R^2 = 0.13$		

*Rejects stability at the asymptotic 10 percent level.

An AR(1) model seems a reasonable univariate description for growth rates of U.S. output:

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + e_t \qquad E(e_t^2) = \sigma^2,$$

where Δy_t denotes the first difference of the log of annual per capita real GNP.

DeLong and Summers (1988) argue that the parameter describing persistence, α_1, has changed over time (most importantly, was substantially higher during the Depression and World War II). Table 2 reports model estimates for the periods 1889–1987, 1889–1929, 1930–1947, and 1948–1987 as recommended by DeLong and Summers. The standard errors are calculated using White's heteroskedasticity-robust covariance matrix estimator. The L stability tests are calculated on the full sample.

At first sight, the point estimates from the sample subperiods and the L statistics appear to be partially in conflict. The estimates and L statistic agree that σ^2 is not constant, but the estimates of the autoregressive parameter indicate a substantial shift over the periods, while the L_1 statistic fails to reject the null that α_1 is constant. How should we interpret this information?

We first need to think about the finite sample properties of the test statistics. Because the distributional theory presented in the previous section is based upon asymptotic approximations, we should be careful to check if the approximation is useful in the present context. I generated 2,000 samples from the full sample model reported in the first

Table 3: Finite Sample Rejection Frequencies of Asymptotic 10 per cent Tests

	L_1	$L_2\ (\sigma^2)$	Joint L_C
Null	0.11	0.08	0.09
α_1 and σ^2 varying	0.11	0.61	0.25
α_1 varying	0.30	0.10	0.17

line of Table 2 (under the assumption of independently and identically distributed normal errors), and applied the individual stability tests for the autoregressive coefficient and error variance, as well as the joint stability test, using the asymptotic 10 percent critical values. Table 3 presents some simulation results. The first line reports a measure of finite sample size distortion. The table reports rejection frequencies. The tests rejected the null hypothesis at rates close to the nominal values, indicating quite mild size distortion. The second and third lines of Table 3 report the power of the test against particular alternatives. The numbers reported are rejection frequencies of the test statistics using 1000 samples and the asymptotic 10 percent critical values. The second line used data generated from the same model as used under the null, except that the autoregressive parameter and error variance shift according to the parameter estimates reported in Table 2. The third line of Table 3 has only the autoregressive parameter shifting.

These experiments are designed to answer the question: If the data is generated as suggested by the split-sample estimates, what is the probability that the test statistics will reject the null hypothesis? The answer is quite startling. If both the autoregressive *and* the error variance shift over time, the test will only be able to detect the shift in the error variance. The reason apparently is that the shifting error variance induces too much noise into the series for the test to be able to distinguish parameter variation from sampling variation. But the test is able to reject the constant variance hypothesis 61 percent of the time. When the error variance is held fixed, then the test on the autoregressive coefficient displays some ability to reject the null hypothesis, rejecting 30 percent of the time.

The last simulations suggest that it is difficult to test the stability of one set of parameters, if we allow another subset to be shifting over time (the null hypothesis for all of the tests is that *all* of the parameters are stable, and this is important in the derivation of the distributional theory). Although the results so far seem to suggest strongly that the error variance has changed, we just do not know if the autoregressive parameter was stable or not. We can partially circumvent this problem by using generalized least squares to eliminate the shifts in the error variance. Under the assumption that the error variance shifted twice, in 1930 and 1948, we can use the estimates from Table 2 to re-estimate the equation. This regression is not reported here, but the main results do not change (the L statistic for the AR parameter is 0.18). We conclude that the data are not sufficiently informative to determine whether or not the autoregressive parameter is stable, *given* that the error variance is unstable.

This exercise illustrates the limitations of econometric techniques and the potential dangers of "ocular" or "eyeball" econometrics. The natural impulse to split the sample at some known important date, as done by DeLong and Summers, must be resisted. If conventional critical values are used (which ignore the fact that the sample split point was *selected*), then the tests will be biased towards *spurious rejection* of the stability hypothesis. Similarly, visual displays of recursive estimates or split-sample estimates such as those in Table 2 must be tempered with a large dose of caution. What appears as a significant difference may not in fact be significant once the selection of the sample split is taken into account. The tests advocated in this paper are immune to such criticism because they do not require the selection of a breakpoint. On the other hand, these general tests may have relatively low power against particular alternatives. As a result, we may be unable to extract definitive conclusions from time-series data.

4. TESTING STABILITY IN ERROR-CORRECTION MODELS

Many applied time-series regressions take the form of error-correction models (ECMs). We can use the L test statistics to assess the stability of ECMs. ECMs can generally be written in the form

$$\Delta y_{1t} = \mu + \gamma(y_{1t-1} - \alpha' y_{2t-1}) + \beta' x_t + e_t, \tag{11}$$

where y_{1t} and y_{2t} are individually $I(1)$ yet jointly cointegrated. The x_t variables should be $I(0)$, such as lagged values of Δy_{1t} and Δy_{2t}.

The difficulty in applying stability tests in the context of ECMs is that the levels data y_{1t} and y_{2t} contain stochastic or deterministic trends, thus invalidating the distributional theory used to justify the critical values reported in Table 1. Therefore, testing the stability of the cointegrating parameter α requires a different theory, which is given in Hansen (1991b). If we are interested in the *dynamics* of an ECM, however, we can use our stability tests to test the stability of the coefficients on $I(0)$ variables. In Equation 11, if α were known, then the remaining parameters are all coefficients of $I(0)$ variables. If α is not known but is consistently estimated at a rate faster than the square root of the sample size, then the stability test applied to the remaining coefficients can proceed as before.

The easiest way to conduct this test is by using a two-step estimator in the spirit of Engle and Granger (1987). First estimate the cointegrating vector α using either OLS or an asymptotically efficient estimator such as the fully modified (FM) estimator of Phillips and Hansen (1990). Then take the residuals from this first-stage regression and use them in the ECM, Equation 11. Because all the variables in this second stage are $I(0)$, we can apply the testing procedures of Section 2.

This procedure is illustrated by the following aggregate consumption model using quarterly U.S. data:

$$\Delta c_t = \beta_1 u_{t-1} + \beta_2 \Delta i_t + \beta_3 \Delta \pi_t + \beta_4 + e_t, \tag{12}$$

$$u_t = c_t - \alpha_1 i_t - \alpha_2 \pi_t - \alpha_3. \tag{13}$$

Here, c_t is aggregate consumption expenditure, i_t is aggregate total disposable income, and π_t is the inflation rate. In relation to Equation 11, y_{1t} is c_t, y_{2t} is (i_t, π_t), and x_t is $(\Delta i_t, \Delta \pi_t)$. The sample is 1953:2–1984:4. The consumption and income data are taken from Blinder and Deaton (1985), and the inflation rate is calculated from the implicit GNP deflator in the Citibase data base. We will consider Equations 12 and 13 in both levels and logarithmic specifications for the consumption and income series.

The cointegrating regression (Equation 13) in levels without the inflation rate was proposed and analyzed in Campbell (1987) because the model presented in his paper implied that this should be a cointegrating relationship. Deaton (1977) argued that the inflation rate should enter into an aggregate consumption function. He pointed out that a rate of inflation higher than expected is likely to reduce consumption expenditure. This is a disequilibrium (or short-run) mechanism, and suggests that inflation should enter in the dynamic relationship (Equation 12), but not in the long-run relationship (Equation 13). Deaton's empirical results, however, suggest that inflation is significant in a regression similar to Equation 13. This empirical finding was confirmed by the more extensive study of Davidson et al. (1978). Although there does not appear to be a good theory to explain the presence of inflation in the cointegrating relationship, it seems reasonable to test its presence empirically.

We will estimate and evaluate four competing specifications for the cointegrating relationship:

Model A: consumption and income in levels; inflation rate excluded
Model B: consumption and income in levels; inflation rate included
Model C: consumption and income in logs; inflation rate excluded
Model D: consumption and income in logs; inflation rate included.

We now explore these alternative specifications using the tests for cointegration and instability. The parameters of Equation 13 were estimated using the fully modified estimator of Phillips and Hansen (1990).[6] Phillips' $Z(t)$ test of the null of no cointegration was applied to the cointegrating residuals $\hat{u}_t$. These same cointegrating residuals were used in the second step estimation of Equation 12 by OLS. These results are reported in Table 4. For each parameter in Equation 12, the estimate, standard error, and instability statistic are reported.

We first examine the long-run relationship (Equation 13). All four specifications yield reasonable parameter estimates. Particularly interesting are the near-unity values for the income elasticities when the model is estimated in logs. The inflation rate coefficient is negative and significant in both specifications. All four specifications reject the null of no cointegration at the 5 percent level, but the specifications with the inflation rate included reject no cointegration at the 1 percent level.

We now turn to the estimates of the dynamic equation, (12). In all specifications, the error-correction term is negative and significant, as expected. As predicted by Deaton's theory, when the inflation rate is included, its coefficient has a negative sign, but it is not significantly different from zero. In all four equations, none of the regression coefficients display any evidence of instability. The error variance, however, is apparently unstable in the levels equations (rejecting stability at the 1 percent level), but not in the logarithmic equations.

In summary, the empirical evidence appears to favor the logarithmic specification with the inflation rate included. This model strongly rejects no cointegration, has estimated parameters of the proper sign and magnitude, and passes the stability tests applied to the dynamic equation. It is puzzling, however, that the inflation rate appears significantly in the long-run relationship, but not in the short-run relationship. An explanation of this phenomenon would be a useful enterprise for future research.

[6]To calculate the long-run covariance parameters, the residuals were first pre-whitened as suggested by Andrews and Monahan (1990), and then a quadratic kernel applied using the plug-in bandwidth suggested by Andrews (1991).

Table 4: Aggregate Consumption Functions: Estimates and Tests

Parameter	Model A	Model B	Model C	Model D
Equation 13				
α_1	0.93	0.96	1.00	1.02
	(0.01)	(0.01)	(0.01)	(0.01)
α_2		−8.8		−0.002
		(3.6)		(0.001)
α_3	−6.9	−54.6	−0.09	−0.23
	(43.8)	(41.1)	(0.10)	(0.12)
$Z(t)$	−3.51**	−4.91***	−3.68**	−4.68***
Equation 12				
β_1	−0.09	−0.14	−0.11	−0.13
	(0.05)	(0.05)	(0.05)	(0.05)
L_1	0.06	0.07	0.09	0.09
β_2	0.65	0.65	0.68	0.67
	(0.08)	(0.07)	(0.08)	(0.07)
L_2	0.28	0.29	0.15	0.15
β_3		−0.63		−0.00015
		(0.69)		(0.00021)
L_3		0.04		0.04
β_4	5.70	5.78	0.0019	0.0019
	(2.52)	(2.43)	(0.0007)	(0.0007)
L_4	0.18	0.19	0.07	0.07
σ^2	393	382	0.000033	0.000032
	(63)	(59)	(0.000004)	(0.000004)
L_5	1.14***	1.09***	0.21	0.16
R^2	0.47	0.48	0.46	0.47
Joint L_c	1.41**	1.42*	0.44	0.45

* Rejects at the asymptotic 10 per cent level.
** Rejects at the asymptotic 5 per cent level.
***Rejects at the asymptotic 1 per cent level.

5. CONCLUSION

This paper has presented a simple yet powerful test for parameter instability. No sample-split points or forecast intervals need to be chosen. The test only requires that the model be estimated once over the full sample. The asymptotic distribution is nonstandard, depending only upon the number of coefficients tested for stability.

If the test statistics are insignificant, then the investigator can be reasonably confident that either the model has been constant over that sample or the data is not sufficiently informative to reject this hypothesis. On the other hand, a significant test statistic suggests the presence of model misspecification. It appears that this test statistic can provide useful information in practice.

REFERENCES

Andrews, D.W.[illegible]990) Tests for Parameter Instability and Structural Change With Unknown Change Poi[illegible] Discussion Paper No. 943, Yale University.

Andrews, D.W.K. (1991[illegible]oskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica* 59: 817–858.

Andrews, D.W.K., and Monahan, J.C. (1990) An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. Forthcoming in *Econometrica*.

Blinder, A.S., and Deaton, A.S. (1985) The Time Series Consumption Function Revisited, *Brookings Papers on Economic Activity* 1985: 465–511.

Brown, R.L., Durbin, J., and Evans, J.M. (1975) Techniques for Testing the Constancy of Regression Relationships Over Time, *Journal of the Royal Statistical Society, Series B* 37: 149–163.

Campbell, J.Y. (1987) Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis, *Econometrica* 55: 1249–1273.

Chow, G.C. (1960) Tests of Equality Between Sets of Coefficients in Two Linear Regressions, *Econometrica* 28: 591–605.

Chu, C.-S.J. (1989) New Tests for Parameter Constancy in Stationary and Nonstationary Regression Models, University of California at San Diego.

Cooley, T.F., and Prescott, E.C. (1976) Estimation in the Presence of Stochastic Parameter Variation, *Econometrica* 44: 167–184.

Cox, D.R., and Hinkley, D.V. (1974) *Theoretical Statistics*. London: Chapman and Hall.

Davidson, J.E.H., Hendry, D.F., Srba, F., and Yeo, S. (1978) Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom, *Economic Journal* 88: 661–692.

Davies, R.B. (1977) Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative, *Biometrika* 64: 247–254.

Davies, R.B. (1987) Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative, *Biometrika* 74: 33–43.

Deaton, A. (1977) Involuntary Saving Through Unanticipated Inflation, *American Economic Review* 67: 899–910.

DeLong, J.B., and Summers, L.H. (1988) On the Existence and Interpretation of a 'Unit Root' in U.S. GNP. NBER Working Paper No. 2716.

Engle, R.F., and Granger, C.W.J. (1987) Co-integration and Error Correction: Representation, Estimation and Testing, *Econometrica* 55: 251–276.

Engle, R.F., and Hendry, D.F. (1990) Testing Super Exogeneity and Invariance in Regression Models. Forthcoming in *Journal of Econometrics*.

Engle, R.F., Hendry, D.F., and Richard, J.-F. (1983) Exogeneity, Econometrica 51: 277–304.

Fisher, R.A. (1922) The Goodness of Fit of Regression Formula, and the Distribution of Regression Coefficients, *Journal of the Royal Statistical Society* 85: 597–612.

Gardner, L.A., Jr. (1969) On Detecting Changes in the Mean of Normal Variates, *Annals of Mathematical Statistics* 40: 116–126.

Hansen, B.E. (1990) Lagrange Multiplier Tests for Parameter Instability in Non-Linear Models, University of Rochester.

Hansen, B.E. (1991a) A Comparison of Tests for Parameter Instability: An Examination of Asymptotic Local Power, University of Rochester.

Hansen, B.E. (1991b) Tests for Parameter Instability in Regressions with $I(1)$ Processes. Forthcoming in *Journal of Business and Economic Statistics*.

Hendry, D.F. (1989) *PC-GIVE: An Interactive Econometric Modeling System*, Oxford: University of Oxford, Institute of Economics and Statistics and Nuffield College.

Hendry, D.F., and Ericsson, N.R. (1991) An Econometric Analysis of UK Money Demand in *Monetary Trends in the United States and the United Kingdom* by Milton Friedman and Anna J. Schwartz, *American Economic Review* 81: 8–38.

Hendry, D.F., and Richard, J.-F. (1982) On the Formulation of Empirical Models in Dynamic Econometrics, *Journal of Econometrics* 20: 3–33.

King, M.L. (1987) An Alternative Test for Regression Coefficient Stability, *Review of Economics and Statistics* 69: 379–381.

Krämer, W., Ploberger, W., and Alt, R. (1988) Testing for Structural Change in Dynamic Models, *Econometrica* 56: 1355–1369.

Nyblom, J. (1989) Testing for the Constancy of Parameters Over Time, *Journal of the American Statistical Association* 84: 223–230.

Nyblom, J., and Makelainen, T. (1983) Comparisons of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model, *Journal of the American Statistical Association* 78: 856–864.

Pagan, A.R., and Tanaka, K. (1981) A Further Test for Assessing the Stability of Regression Coefficients, Australian National University.

Phillips, P.C.B., and Hansen, B.E. (1990) Statistical Inference in Instrumental Variables Regression With I(1) Processes, *Review of Economic Studies* 57: 99–125.

Phillips, P.C.B., and Ouliaris, S. (1990) Asymptotic Properties of Residual-Based Tests for Cointegration. *Econometrica* 58: 165–193.

Ploberger, W., and Krämer, W. (1990) The Local Power of the CUSUM and CUSUM of Squares Tests, *Econometric Theory* 6: 335–347.

Quandt, R. (1960) Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes, *Journal of the American Statistical Association* 55: 324–330.

Romer, C.D. (1986a) Spurious Volatility in Historical Unemployment Data, *Journal of Political Economy* 94: 1–37.

Romer, C.D. (1986b) Is the Stabilization of the Postwar Economy a Figment of the Data, *American Economic Review* 76: 314–334.

Romer, C.D. (1989) The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product, 1869–1908, *Journal of Political Economy* 97: 1–37.

Tobin, J. (1980) *Asset Accumulation and Economic Activity*. Chicago: Basil Blackwell.

White, H. (1980) A Heteroskedastic-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica* 48: 817–838.

White, H. and Domowitz, I. (1984) Nonlinear Regression With Dependent Observations, *Econometrica* 52: 143–162.