Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model

James H. Stock and Mark W. Watson

This article considers inference about the variance of coefficients in time-varying parameter models with stationary regressors. The Gaussian maximum likelihood estimator (MLE) has a large point mass at 0. We thus develop asymptotically median unbiased estimators and asymptotically valid confidence intervals by inverting quantile functions of regression-based parameter stability test statistics, computed under the constant-parameter null. These estimators have good asymptotic relative efficiencies for small to moderate amounts of parameter variability. We apply these results to an unobserved components model of trend growth in postwar U.S. per capita gross domestic product. The MLE implies that there has been no change in the trend growth rate, whereas the upper range of the median-unbiased point estimates imply that the annual trend growth rate has fallen by 0.9% per annum since the 1950s.

KEY WORDS: Stochastic coefficient model; Structural time series model; Unit moving average root; Unobserved components.

1. INTRODUCTION

Since its introduction in the early 1970s by Cooley and Prescott (1973a,b, 1976), Rosenberg (1972, 1973), and Sarris (1973), the time-varying parameter (TVP), or "stochastic coefficients," regression model has been used extensively in empirical work, especially in forecasting applications. Chow (1984), Engle and Watson (1987), Harvey (1989), Nichols and Pagan (1985), Pagan (1980), and Stock and Watson (1996) have provided references and discussion of this model. The appeal of the TVP model is that by permitting the coefficients to evolve stochastically over time, it can be applied to time series models with parameter instability.

The TVP model considered in this article is

$$y_t = \boldsymbol{\beta}_t' \mathbf{X}_t + u_t, \tag{1}$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \mathbf{v}_t, \tag{2}$$

$$a(L)u_t = \varepsilon_t, \tag{3}$$

and

$$\mathbf{v}_t = \tau \boldsymbol{\nu}_t, \quad \text{where} \quad \boldsymbol{\nu}_t = \mathbf{B}(L)\boldsymbol{\eta}_t, \tag{4}$$

where $\{(y_t, \mathbf{X}_t), t = 1, \ldots, T\}$ are observed, $\mathbf{X}_t$ is an exogenous k-dimensional regressor, $\boldsymbol{\beta}_t$ is a $k \times 1$ vector of unobserved time-varying coefficients, τ is a scalar, $a(L)$ is a scalar lag polynomial, $\mathbf{B}(L)$ is a $k \times k$ matrix lag polynomial, and ε_t and $\boldsymbol{\eta}_t$ are serially and mutually uncorrelated mean 0 random disturbances. (Additional technical conditions used for the asymptotic results are given in Section 2, where we also discuss restrictions on $\mathbf{B}(L)$ and $E(\boldsymbol{\eta}_t \boldsymbol{\eta}_t')$ that are sufficient to identify the scale factor τ.) An important special case of this model is when $\mathbf{X}_t = 1$ and $\mathbf{B}(L) = 1$; following Harvey (1985), we refer to this case as the "local-level" unobserved components model.

James H. Stock is Professor of Political Economy, Kennedy School of Government, Harvard University, Cambridge, MA 02138, and Research Associate at the National Bureau of Economic Research (NBER). Mark W. Watson is Professor of Economics and Public Affairs, Woodrow Wilson School, Princeton University, Princeton, NJ 08544, and Research Associate at the NBER. The authors thank Bruce Hansen, Andrew Harvey and two referees for comments on an earlier draft, and Lewis Chan and Jonathan Wright for research assistance. This research was supported in part by National Science Foundation grant SBR-9409629.

We consider the problem of estimation of the scale parameter τ. If (as is common) ε_t and $\boldsymbol{\eta}_t$ are assumed to be jointly normal and independent of $\{\mathbf{X}_t, t = 1, \ldots, T\}$, then the parameters of (1)–(4) can be estimated by maximum likelihood implemented by the Kalman filter. However, the maximum likelihood estimator (MLE) has the undesirable property that if τ is small, then it has point mass at 0. In the case $\mathbf{X}_t = 1$, this is related to the so-called pile-up problem in the first-order moving average [MA(1)] model with a unit root (Sargan and Bhargava 1983; Shephard and Harvey 1990). In the general TVP model (1)–(4), the pile-up probability depends on the properties of $\mathbf{X}_t$ and can be large. The pile-up probability is a particular problem when τ is small and thus is readily mistaken for 0. Arguably, small values of τ are appropriate for many empirical applications; indeed, if τ is large, then the distribution of the MLE can be approximated by conventional $T^{1/2}$-asymptotic normality, but Monte Carlo evidence suggests that this approximation is poor in many cases of empirical interest. (See Davis and Dunsmuir 1996 and Shephard 1993 for discussions in the case of $\mathbf{X}_t = 1$.)

We thus focus on the estimation of τ when it is small. In particular, we consider the nesting

$$\tau = \lambda / T. \tag{5}$$

Order of magnitude calculations suggest that this might be an appropriate nesting for certain empirical problems of interest, such as estimating stochastic variation in the trend component in the growth rate of U.S. real gross domestic product (GDP), as we discuss in Section 4. This is also the nesting used to obtain local asymptotic power functions of tests of $\tau = 0$, a fact suggesting that if the researcher is in a region in which tests yield ambiguous conclusions about the null hypothesis $\tau = 0$, then the nesting (5) is appropriate.

The main contribution of this article is the development of asymptotically median unbiased estimators of λ and asymptotically valid confidence intervals for λ in the model (1)–(5). These are obtained by inverting asymptotic quantile functions of statistics that test the hypothesis $\lambda = 0$. The

Journal of the American Statistical Association
March 1998, Vol. 93, No. 441, Theory and Methods

test statistics are based on generalized least squares (GLS) residuals, which are readily computed under the null. As part of the calculations, we obtain asymptotic representations for a family of tests under the local alternative (5). These representations can be used to compute local asymptotic power functions against nonzero values of λ. Section 2 presents these theoretical results.

Section 3 provides numerical results for the special cases of the univariate local-level model. Properties of the median unbiased estimators are compared to two MLEs, which alternatively maximize the marginal and the profile (or concentrated) likelihoods; these MLEs differ in their treatment of the initial value for $\boldsymbol{\beta}_t$. Both MLEs are biased and have large pile-ups at $\hat{\lambda} = 0$. When λ is small, the median unbiased estimators are more tightly concentrated around the true value of λ than either MLE.

Section 4 presents an application to the estimation of a long-run stochastic trend for the growth rate of postwar real per capita GDP in the United States. Point estimates from the median unbiased estimators suggest a slowdown in the average trend rate of growth; the largest point estimate suggests a slowdown of approximately .9% per annum from the 1950s to the 1990s. The MLEs suggest a much smaller decline, with point estimates ranging from 0 to .2%. Section 5 concludes.

2. THEORETICAL RESULTS

We assume that $a(L)$ has known finite order p and thus consider statistics based on feasible GLS. Specifically, (a) y_t is regressed on $\mathbf{X}_t$ by ordinary least squares (OLS), producing residuals $\hat{u}_t$; (b) a univariate AR(p) is estimated by OLS regression of $\hat{u}_t$ on $(1, \hat{u}_{t-1}, \ldots, \hat{u}_{t-p})$, yielding $\hat{a}(L)$; and (c) $\tilde{y}_t \equiv \hat{a}(L)y_t$ is regressed on $\tilde{\mathbf{X}}_t \equiv \hat{a}(L)\mathbf{X}_t$, yielding the GLS estimator $\tilde{\boldsymbol{\beta}} = (T^{-1}\sum_{t=1}^{T}\tilde{\mathbf{X}}_t\tilde{\mathbf{X}}_t')^{-1}T^{-1}\sum_{t=1}^{T}\tilde{\mathbf{X}}_t\tilde{y}_t$, residuals $\tilde{e}_t$ and moment matrix $\tilde{\mathbf{V}}$:

$$\tilde{e}_t = \tilde{y}_t - \tilde{\boldsymbol{\beta}}'\tilde{\mathbf{X}}_t \tag{6}$$

and

$$\tilde{\mathbf{V}} = \left(T^{-1}\sum_{t=1}^{T}\tilde{\mathbf{X}}_t\tilde{\mathbf{X}}_t'\right)\tilde{\sigma}_{\varepsilon}^2, \tag{7}$$

where $\tilde{\sigma}_{\varepsilon}^2 = (T-k)^{-1}\sum_{t=1}^{T}\tilde{e}_t^2$. If $a(L) = 1$, then steps (a) and (b) are omitted and the OLS and GLS regressions of y_t on $\mathbf{X}_t$ are equivalent.

Two test statistics are considered: Nyblom's (1989) L_T statistic (modified to use GLS residuals) and the sequential GLS Chow F statistics, $F_T(s)(0 \le s \le 1)$, which test for a break at date $[Ts]$, where $[\cdot]$ denotes the greatest lesser integer. Let SSR_{t_1,t_2} denote the sum of squared residuals from the GLS regression of $\tilde{y}_t$ onto $\tilde{X}_t$ over observations $t_1 \le t \le t_2$, and let $\boldsymbol{\xi}_T(s) = T^{-1/2}\sum_{t=1}^{[Ts]}\tilde{\mathbf{X}}_t\tilde{e}_t$. The L_T and F_T statistics are

$$L_T = T^{-1}\sum_{t=1}^{T}\boldsymbol{\xi}_T(t/T)'\tilde{\mathbf{V}}^{-1}\boldsymbol{\xi}_T(t/T) \tag{8}$$

and

$$\begin{aligned} F_T(s) = (&\mathrm{SSR}_{1,T} - \mathrm{SSR}_{1,[Ts]} - \mathrm{SSR}_{[Ts]+1,T}) \\ &\div [k(\mathrm{SSR}_{1,[Ts]} + \mathrm{SSR}_{[Ts]+1,T})/(T-k)]. \end{aligned} \tag{9}$$

(For other tests in versions of this model, see Franzini and Harvey 1983; Harvey and Streibel 1997; King; and Hillier 1985; Nabeya and Tanaka 1988; Nyblom 1989; Reinsel and Tam 1996; Shively 1988.)

The F_T statistic is an empirical process, and inference is performed using one-dimensional functionals of F_T. We consider three such functionals: the maximum F_T statistic (the Quandt [1960] likelihood ratio statistic), $\mathrm{QLR}_T = \sup_{s\in(s_0,s_1)} F_T(s)$; the mean Wald statistic of Andrews and Ploberger (1994) and Hansen (1992), $\mathrm{MW}_T = \int_{s_0}^{s_1} F_T(r)\,dr$; and the Andrews–Ploberger (1994) exponential Wald statistic, $\mathrm{EW}_T = \ln\{\int_{s_0}^{s_1}\exp(\frac{1}{2}F_T(r))\,dr\}$, where $0 < s_0 < s_1 < 1$.

Three assumptions are used to obtain the asymptotic results. For a stationary process $\mathbf{z}_t$, let $c_{i_1\ldots i_n}(r_1, \ldots, r_{n-1})$ denote the nth joint cumulant of $\mathbf{z}_{i_1t_1}, \ldots, \mathbf{z}_{i_nt_n}$, where $r_j = t_j - t_n, j = 1, \ldots, n-1$ (Brillinger 1981), and let $C(r_1, \ldots, r_{n-1}) = \sup_{i_1,\ldots,i_n} c_{i_1\ldots i_n}(r_1, \ldots, r_{n-1})$.

Assumption A. $\mathbf{X}_t$ is stationary with eighth order cumulants that satisfy $\sum_{r_1,\ldots,r_7}^{\infty}|C(r_1, \ldots, r_7)| < \infty$.

Assumption B. $\{\mathbf{X}_t, t = 1, \ldots, T\}$ is independent of $\{u_t, \boldsymbol{\nu}_t, t = 1, \ldots, T\}$.

Assumption C. $(\varepsilon_t, \boldsymbol{\eta}_t')'$ is a $(k+1)\times 1$ vector of iid errors with mean 0 and four moments; ε_t and $\boldsymbol{\eta}_t$ are independent; $a(L)$ has finite-order p; and $\mathbf{B}(L)$ is one-summable with $\mathbf{B}(1) \neq 0$.

Assumption A requires that $\mathbf{X}_t$ have bounded moments or, if nonstochastic, that it not exhibit a trend. The assumption of stationarity is made for convenience in the proofs and could be relaxed somewhat. However, the requirement that $\mathbf{X}_t$ not be integrated of order 1 ($I(1)$) or higher is essential for our results.

Assumption B requires $\mathbf{X}_t$ to be strictly exogenous. This assumption permits estimation of (1), under the null $\boldsymbol{\beta}_t = \boldsymbol{\beta}_0$, by GLS.

The assumption that $a(L)$ has finite-order p in assumption C is made to simplify estimation by feasible GLS. The assumption that ε_t and $\boldsymbol{\eta}_t$ are independent ensures that u_t and $\mathbf{v}_t$ have a zero cross-spectral density matrix. This is a basic identifying assumption of the TVP model (Harvey 1989). To construct the Gaussian MLE, ε_t and $\boldsymbol{\eta}_t$ are modeled as independent iid normal random variables.

The assumption that $\mathbf{X}_t$ is independent of the errors can be unappealing in some applications. For example, in some econometric applications $\mathbf{X}_t$ is predetermined but not strictly exogenous, u_t is plausibly serially uncorrelated, but there is feedback from u_t to future $\mathbf{X}_t$. In lieu of assumptions B and C, we thus introduce an alternative assumption to handle regressors that are predetermined but not exogenous.

Assumption D. $(\varepsilon_t, \boldsymbol{\eta}_t')'$ is a $(k+1) \times 1$ vector of iid errors with mean 0 and four moments; ε_t and $\boldsymbol{\eta}_t$ are independent; $a(L) = 1; \mathbf{B}(L)$ is one-summable with $\mathbf{B}(1) \neq 0; \boldsymbol{\eta}_t$ is independent of $\{\mathbf{X}_t, \mathbf{X}_{t\pm 1}, \mathbf{X}_{t\pm 2}, \ldots\}$; and u_t is independent of $\{\mathbf{X}_t, \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots\}$.

This permits feedback from u_t to future $\mathbf{X}_t$, but not from $\mathbf{v}_t$ to $\mathbf{X}_t$, and thus rules out $\mathbf{X}_t$ containing lagged y_t when $\lambda \neq 0$.

Our main theoretical results are given in the following theorem. Let "$\Rightarrow$" denote weak convergence on $D[0,1]$, let $\mathbf{W}_1$ and $\mathbf{W}_2$ be independent standard Brownian motions on $[0,1]^k$, and let $\boldsymbol{\Gamma} = E\{[a(L)\mathbf{X}_t][a(L)\mathbf{X}_t]'\}, \boldsymbol{\Omega} = \mathbf{B}(1)E(\boldsymbol{\eta}_t\boldsymbol{\eta}_t')\mathbf{B}(1)'$, and $\mathbf{D} = \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Omega}^{1/2}/\sigma_\varepsilon$.

Theorem 1. Let y_t obey (1)–(5), and suppose either that assumptions A, B, and C hold or that assumptions A and D hold. Then

a. $\tilde{\mathbf{V}}^{-1/2}\boldsymbol{\xi}_T \Rightarrow \mathbf{h}_\lambda^0$, where $\mathbf{h}_\lambda^0(s) = \mathbf{h}_\lambda(s) - s\mathbf{h}_\lambda(1)$, where $\mathbf{h}_\lambda(s) = \mathbf{W}_1(s) + \lambda\mathbf{D}\int_0^s \mathbf{W}_2(r)\,dr$;
b. $L_T \Rightarrow \int_0^1 \mathbf{h}_\lambda^0(s)'\mathbf{h}_\lambda^0(s)\,ds$; and
c. $F_T \Rightarrow F^*$, where $F^*(s) = \mathbf{h}_\lambda^0(s)'\mathbf{h}_\lambda^0(s)/(ks(1-s))$.

The proof is given in the Appendix.

Limiting representations of the QLR, mean Wald, and exponential Wald statistics are obtained from part (c) of Theorem 1 and the continuous mapping theorem. Thus $\mathrm{QLR}_T \Rightarrow \sup_{s_0 \le s \le s_1} F^*(s)$, $\mathrm{MW}_T \Rightarrow \int_{s_0}^{s_1} F^*(r)\,dr$, and $\mathrm{EW}_T \Rightarrow \ln\{\int_{s_0}^{s_1} \exp(\frac{1}{2}F^*(r))dr\}$. Note that the limiting representation for L_T can be written as $L_T \Rightarrow k\int_0^1 (r(1-r))F^*(r)\,dr$.

When $\lambda = 0$, the process h_λ^0 is a k-dimensional Brownian bridge, and the representations for the statistics L_T and F_T reduce to their well-known null representations as functionals of a Brownian bridge (Andrews and Ploberger 1994; Nabeya and Tanaka 1988; Nyblom 1989).

When $\lambda \neq 0$, the limiting distributions of L_T and F_T depend on two parameters, λ and $\mathbf{D}$. The limiting representations in Theorem 1 are used for three purposes: to compute local asymptotic power functions, to construct median unbiased estimators of λ, and to construct asymptotically valid confidence intervals for λ. To do so, $\mathbf{D}$ must either be known or be consistently estimable, so that asymptotically λ is the only unknown parameter entering these distributions.

The determination of $\mathbf{D}$ raises issues of identification and modeling strategy. Evidently $a(L)$ and $\mathrm{var}(\varepsilon_t)$ are not separately identified, but this is resolved without loss of generality by adopting the normalization $a_0 = 1$. Similarly, for standard reasons associated with moving average models, $\mathbf{B}(L)$ and $E\eta_t\eta_t'$ are not separately identified; we thus adopt the conventional assumptions that $\mathbf{B}_0 = \mathbf{I}_k$ and the roots of $|\mathbf{B}(z)|$ are outside the unit circle. Even with these assumptions, however, inspection of (1)–(5) reveals that λ and $\boldsymbol{\Omega}$ are not separately identified: The parameter combinations $(\lambda, \boldsymbol{\Omega}) = (\bar{\lambda}, \bar{\boldsymbol{\Omega}})$ and $(\lambda, \boldsymbol{\Omega}) = (1, \bar{\lambda}^2\bar{\boldsymbol{\Omega}})$ are observationally equivalent for fixed $(\bar{\lambda}, \bar{\boldsymbol{\Omega}})$.

When $k = 1$, this identification problem can be solved without loss of generality by adopting an arbitrary normalization. Henceforth, when $k = 1$, we thus set $\mathbf{D} = 1$. When $\mathbf{X}_t = 1$, under this normalization, λ is T times the ratio of the long-run standard deviation of $\Delta\boldsymbol{\beta}_t$ to the long run standard deviation of u_t.

When $k > 1, \boldsymbol{\Omega}$ is identified upon making a single suitable normalization; for example, that the trace of $\boldsymbol{\Omega}$ is unity. However, the local-to-0 variation in $\Delta\boldsymbol{\beta}_t$ makes it impossible to estimate the free elements of $\boldsymbol{\Omega}$ consistently without further restrictions. In this case two types of further restrictions suggest themselves. First, $\boldsymbol{\Omega}$ may be set equal to a prespecified constant matrix chosen by the researcher in a manner appropriate for the specific empirical model under study. Second, $\boldsymbol{\Omega}$ may be parameterized as a function of $\boldsymbol{\Gamma}$ and σ_ε^2 (which are consistently estimable). As in the $k = 1$ case, a convenient parameterization sets $\boldsymbol{\Omega} = \sigma_\varepsilon^2\boldsymbol{\Gamma}^{-1}$, for this implies that $\mathbf{D} = \mathbf{I}_k$. This choice of $\boldsymbol{\Omega}$ implies that the regression coefficients evolve as mutually independent random walks after rotating the regressors so that they are mutually uncorrelated. This is the parameterization used by Nyblom (1989) in his development of the LMPI test for $\lambda = 0$. From a computational perspective, this assumption is attractive because it simplifies the calculation of median unbiased estimators of λ and the construction of confidence intervals. From a modeling perspective, the restriction is arguably appealing because it makes the magnitude of the time variation comparable across variables when measured in standard deviation units (or, for general $a(L)$ and $\boldsymbol{\beta}(L)$, long-run standard deviation units). With the additional restrictions $a(L) = 1$ and $\mathbf{B}(L) = 1$, this restriction was used by Stock and Watson (1996) in their investigation of time variation in empirical macroeconomic relationships. Whether this assumption is desirable for general $\mathbf{X}_t$ is a matter of modeling strategy in a particular empirical application.

2.1 Local Asymptotic Power

The representations can be used to compute the distribution of the tests under the local alternative (5) and thus to compute the local asymptotic power of tests of the null $\tau = 0$. The various test statistics have limiting representations under the local alternative that are qualitatively similar. This is interesting, because the F_T-based statistics are typically motivated by considering the single break model, whereas Nyblom (1989) derived the L_T statistic as the LMPI test statistic for the seemingly rather different Gaussian TVP model.

2.2 Median Unbiased Estimation of $\boldsymbol{\lambda}$

Median unbiased estimators of λ can be computed from L_T or from a scalar functional of F_T. Consider, for example, the scalar functional $g(F_T)$, which is assumed to be continuous. By the continuous mapping theorem, $g(F_T) \Rightarrow g(F^*)$, the distribution of which depends on λ and $\mathbf{D}$. Let $m_{\mathbf{D}}(\lambda)$ denote the median of $g(F^*)$ as a function of λ for a given matrix of nuisance parameters $\mathbf{D}$. Suppose that $m_{\mathbf{D}}(\cdot)$ is monotone increasing and continuous in λ. Then the inverse function $m_{\mathbf{D}}^{-1}$ exists, and for $\mathbf{D}$ known, λ can be estimated by

$$\hat{\lambda}_g = m_{\mathbf{D}}^{-1}(g(F_T)). \qquad (10)$$

Asymptotically, $\hat{\lambda}_g \Rightarrow m_{\mathbf{D}}^{-1}(g(F^*))$. By construction, $\Pr[\hat{\lambda}_g < \lambda] \to \Pr[m_{\mathbf{D}}^{-1}(g(F^*)) < \lambda] = \Pr[g(F^*) < m_{\mathbf{D}}(\lambda)] = .5$, so $\hat{\lambda}_g$ is asymptotically median unbiased.

In practice $\mathbf{D}$ is not known, so the estimator (10) is infeasible. However, as discussed earlier, $\mathbf{D}$ generally can be consistently estimated for a given choice of Ω. If in addition $m_{\mathbf{D}}^{-1}(\cdot)$ is continuous in $\mathbf{D}$ (which it is for the functionals considered in this article), then (10) can be computed with $\mathbf{D}$ replaced by a consistent estimator $\hat{\mathbf{D}}$, and the same asymptotic distribution obtains. Note, however, that this is computationally cumbersome, as it requires computing the inverse median function $m_{\hat{\mathbf{D}}}^{-1}(\cdot)$ for every estimate $\hat{\mathbf{D}}$ under consideration. (However, some simplifications are possible because, as pointed out by a referee, the distribution of the test statistics depends only on the eigenvalues of $\mathbf{D}$.) When $\boldsymbol{\Omega}$ is chosen so that $\mathbf{D} = I_k$, the limiting distributions of L_T and F_T depend only on λ and k under the local alternative.

It would be of interest to obtain theoretical results comparing the efficiency of median-unbiased estimators based on the various functionals of F_T. However, the limiting distributions are nonstandard and do not appear to have any simple relation to each other. Thus these efficiency comparisons are undertaken numerically and are reported in the next section.

2.3 Confidence Intervals for $\boldsymbol{\lambda}$

Suppose that $\mathbf{D} = \mathbf{I}_k$, in which case the local asymptotic representations in Theorem 1 depend only on λ and k. For a given scalar test statistic, its representation can then be used to compute a family of asymptotic 5% critical values of tests of $\lambda = \lambda_0$ against a two-sided alternative, and in turn these critical values can be used to construct the set of λ_0 that are not rejected. This set constitutes a 95% confidence set for λ_0. This process of inverting the test statistic can be done graphically by the method of confidence belts or by interpolation of a lookup table. The details parallel those for construction of confidence intervals for autoregressive roots local to unity (Stock 1991) and are omitted.

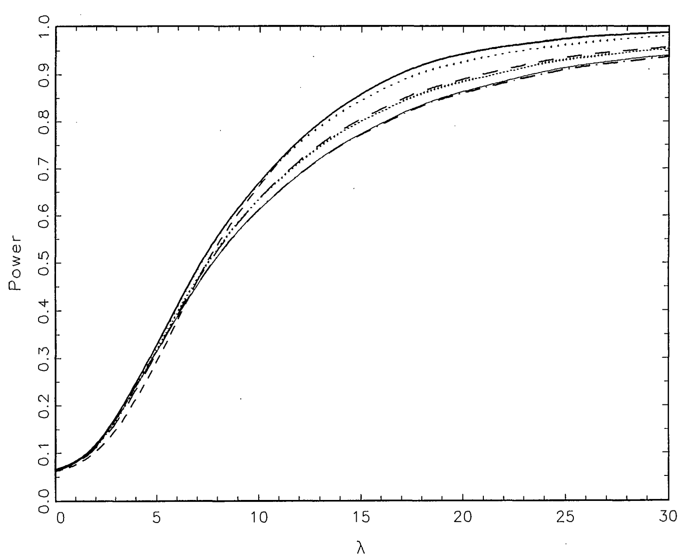

Figure 1. Asymptotic Power Functions of 5% Tests of $\tau = 0$ Against Alternatives $\tau = \lambda/T$. ——, envelope; ——, L; · – ·, MW; · · · · ·, EW; – – –, QLR; · · ·, POI(7); - - -, POI(17).

3. NUMERICAL RESULTS FOR THE UNIVARIATE LOCAL-LEVEL MODEL

In the univariate of local-level model, $\mathbf{X}_t = 1$ and $\mathbf{B}(L) = 1$, so that y_t is the sum of an $I(0)$ component and an independent random walk, which under the parameterization (5) has a small disturbance variance. In this model Δy_t follows a moving average (MA) process, with largest MA root $(1 - \lambda/T)^{-1} + o(T^{-1})$. In this section we first compare numerically the power of the tests in Section 2 and of some other previously proposed tests against the local alternative, then turn to an analysis of the properties of median unbiased estimators. All computations of asymptotic distributions are based on simulation of the limiting representations, with $T = 500$ and 5,000 Monte Carlo replications.

3.1 Asymptotic Power of Tests

A great deal of work has been on tests of $\lambda = 0$ in the local-level model and of a unit MA root in the related MA(1) model (see Nyblom and Mäkeläinen 1983; Saikkonen and Luukonen 1993; Shively 1988; Stock 1994; Tanaka 1990). In addition to the tests discussed in Section 2, local power functions are computed for two point-optimal invariant (POI) tests (Saikkonen and Luukkonen 1993; Shively 1988), for $\lambda = 7$ and $\lambda = 17$, denoted by POI(7) and POI(17). As a basis of comparison, we also computed the asymptotic Gaussian power envelope.

Asymptotic powers of various 5% tests are summarized in Figure 1. Evidently, for small values of λ all tests effectively lie on the asymptotic Gaussian power envelope. For more distant alternatives, the MW and L tests lose power and, to a lesser degree, so do the EW and QLR. The asymptotic power functions of the EW and QLR tests are essentially indistinguishable, consistent with findings elsewhere (Andrews, Lee, and Ploberger 1996; Stock and Watson 1996) that these tests perform similarly.

3.2 Estimators of $\boldsymbol{\lambda}$

Each of the tests examined in Figure 1 has a power function that depends only on λ and has a median function that is monotone and continuous in λ. Asymptotically median unbiased estimators of λ based on each of these tests thus can be constructed as described in Section 2. In addition, results are reported for two versions of the Gaussian MLE that differ in their assumptions concerning the initial value of β_0. The first estimator, the maximum profile (or concentrated) likelihood estimator (MPLE), treats β_0 as an unknown nuisance parameter that is concentrated out of the likelihood. The second estimator, the maximum marginal likelihood estimator (MMLE), treats β_0 as a $\mathrm{N}(\bar{\beta}, \kappa)$ random variable that is independent of $\{u_t, \mathbf{v}_t, t = 1, \ldots, T\}$, so that β_0 is integrated out of the likelihood. When $\kappa \to \infty$, this produces the "diffuse prior" likelihood function (see Shephard 1993

Table 1. Pile-Up Probability That $\hat{\lambda}$ = 0 for MLEs and Various Median-Unbiased Estimators

λ	MPLE	MMLE	L	MW	EW	QLR	POI(7)	POI(17)
0	.96	.66	.50	.50	.50	.50	.50	.50
1	.91	.60	.47	.47	.47	.46	.47	.47
2	.88	.57	.42	.42	.42	.43	.44	.43
3	.81	.47	.34	.34	.34	.35	.35	.37
4	.72	.40	.28	.28	.29	.29	.29	.30
5	.65	.35	.24	.24	.24	.24	.24	.26
6	.56	.28	.19	.19	.19	.19	.18	.20
7	.48	.24	.15	.16	.16	.16	.14	.15
8	.42	.19	.13	.13	.13	.13	.12	.13
9	.37	.17	.11	.12	.12	.12	.09	.10
10	.30	.13	.09	.09	.09	.09	.07	.07
12	.24	.09	.06	.07	.07	.06	.05	.05
14	.15	.06	.03	.04	.04	.04	.03	.02
16	.13	.04	.03	.03	.03	.03	.01	.01
18	.09	.03	.02	.03	.02	.02	.01	.01
20	.07	.02	.01	.01	.01	.01	.01	.01
25	.03	.01	.01	.01	.01	.01	.01	.01
30	.01	.01	.01	.01	.01	.01	.01	.01

NOTE: Entries for MPLE and MMLE for λ = 0 are from Shepard and Harvey (1990). Entries for other values of λ are estimated using 5,000 replications with T = 500. To facilitate the computations, the likelihoods were computed on a discrete grid of 240 equally spaced values of $0 \le \lambda \le 60$, and the MLEs were computed by a search over this grid.

and Shephard and Harvey 1990). The MMLE is equivalent, after reparameterization on a restricted parameter space, to the MA(1) MLE analyzed by Davis and Dunsmuir (1996), and their local-to-unity asymptotic results apply here.

Pile-up probabilities that λ is estimated to be exactly 0 are reported in Table 1. The mass of the median unbiased estimators at 0 is similar for all estimators. The pile-up probability for the MPLE remains large as λ increases, both in absolute terms (it is above 50% for $\lambda \le 6$) and relative to the median unbiased estimators. As pointed out by Shephard and Harvey (1990), the pile-up probability for the MMLE is smaller than for MPLE.

Cumulative distribution functions of the various estimators for $\lambda = 5$ are plotted in Figure 2. As expected, both MLEs are biased and median biased. For $\lambda = 5$, 77% of the mass of the distribution of the MPLE is below the true value and the median is 0; the MMLE performs better, with 64% of its mass below $\lambda = 5$ and a median bias of approximately -1. The cdfs of the median unbiased estimators are fairly similar to each other, but markedly different than the MLE. One apparent cost of unbiasness is their longer right tail relative to the MLEs.

We compare the estimators by computing their asymptotic relative efficiencies (AREs). Because the distributions are nonnormal and not proportional, conventional methods of computing AREs do not apply. Instead, we measure the ARE of the ith estimator $\hat{\tau}_i$ relative to the MMLE, $\hat{\tau}_{\mathrm{MMLE}}$ (denoted by $\mathrm{ARE}_{i,\mathrm{MMLE}}$), as the limit of the ratio of observations T_{MMLE}/T_i needed for $\Pr[\hat{\tau}_i \in G(\tau); T_i] = \Pr[\hat{\tau}_{\mathrm{MMLE}} \in G(\tau); T_{\mathrm{MMLE}}]$, where T_i and T_{MMLE} denote the number of observations used to compute $\hat{\tau}_i$ and $\hat{\tau}_{\mathrm{MMLE}}$. The AREs reported here were for the sets $G(\tau) = \{x: .5\tau \le x \le 1.5\tau\}$, so $\Pr[\hat{\tau}_i \in G(\tau); T_i] = \Pr[|T_i\hat{\tau}_i - T_i\tau| \le .5T_i\tau] \to p_i(T_i\tau)$, say, and similarly for $\hat{\tau}_{\mathrm{MMLE}}$. Using (5), set $\lambda = T_{\mathrm{MMLE}}\tau$; then $\mathrm{ARE}_{i,\mathrm{MMLE}} = \lim T_{\mathrm{MMLE}}/T_i$ can be computed by solving $p_i(\lambda/\mathrm{ARE}_{i,\mathrm{MMLE}}) = p_{\mathrm{MMLE}}(\lambda)$. In general, the ARE depends on λ.

Table 2 reports these AREs for the MPLE and six median unbiased estimators for various values of λ; all AREs are relative to the MMLE. For example, when $\lambda = 4$, the ARE of the QLR-based median unbiased estimator, relative to the MMLE, is 1.02, which indicates that in large samples the MMLE requires 1.02 times as many observations as the QLR-based estimator to achieve the same probability of falling in the set $G(\tau)$. Evidently, MMLE dominates MPLE for all values of λ shown and is considerably more efficient for small to moderate values of λ. In contrast, the median unbiased estimators perform slightly better than MMLE for small values of λ ($\lambda \le 4$) and comparably for moderate

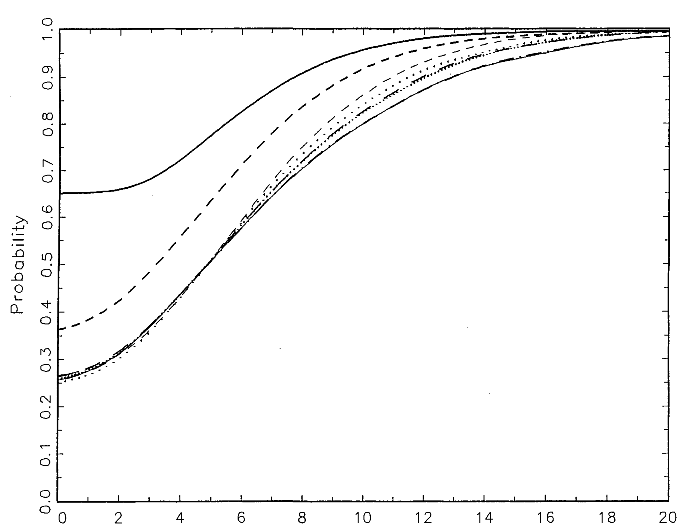

Figure 2. Cumulative Asymptotic Distributions of the Gaussian MLEs and Six Median-Unbiased Estimators of λ When λ = 5. ——, MPLE; - - -, MMLE; ——, L; · – ·, MW; · · · · ·, EW; — —, QLR; · · ·, POI(7); - - ·, POI(17).

Table 2. Asymptotic Relative Efficiencies of Median-Unbiased Estimators Relative to the MMLE

λ	MPLE	L	MW	EW	QLR	POI(7)	POI(17)
1	.13	1.00	1.00	1.00	1.00	1.07	1.07
2	.19	1.09	1.07	1.07	.96	1.07	1.02
3	.52	1.08	1.10	1.08	1.04	1.02	.94
4	.62	1.06	1.06	1.06	1.02	1.06	1.00
5	.65	.93	.93	.98	.97	1.11	1.14
6	.71	.94	.92	.99	1.00	1.03	1.08
7	.76	.79	.79	.85	.88	.96	1.04
8	.77	.77	.77	.85	.85	.91	.98
9	.75	.69	.69	.75	.77	.86	.89
10	.80	.65	.65	.71	.74	.80	.80
12	.67	.56	.56	.65	.67	.67	.67
14	.57	.50	.49	.57	.57	.57	.57
16	.50	.42	.42	.49	.50	.50	.50
18	.44	.38	.38	.44	.44	.44	.44
20	.40	.33	.33	.40	.40	.40	.40
25	.32	.28	.28	.32	.32	.32	.32
30	.27	.22	.23	.27	.27	.27	.27

NOTE: The reported AREs are the limiting ratio of the number of observations necessary for the MLE to achieve the same probability of being in the region $\tau \pm .5\tau$ as the candidate estimator, as a function of $\lambda = \tau/T$, as described in the text. AREs exceeding 1 indicate greater efficiency than the MLE. Entries are estimates based on interpolating probabilities from the values of λ shown in column 1. These probabilities were estimated using 5,000 replications and T = 500 for each value of λ.

Table 3. Lookup Table for Constructing Median-Unbiased Estimator of λ for Various Test Statistics When $X_t = 1$ and $D = 1$

λ	L	MW	EW	QLR	POI7	POI17
0	.118	.689	.426	3.198	2.693	7.757
1	.127	.757	.476	3.416	2.740	7.825
2	.137	.806	.516	3.594	2.957	8.218
3	.169	1.015	.661	4.106	3.301	8.713
4	.205	1.234	.826	4.848	3.786	9.473
5	.266	1.632	1.111	5.689	4.426	10.354
6	.327	2.018	1.419	6.682	4.961	11.196
7	.387	2.390	1.762	7.626	5.951	12.650
8	.490	3.081	2.355	9.160	6.689	13.839
9	.593	3.699	2.910	10.660	7.699	15.335
10	.670	4.222	3.413	11.841	8.849	16.920
11	.768	4.776	3.868	13.098	10.487	19.201
12	.908	5.767	4.925	15.451	11.598	20.570
13	1.036	6.586	5.684	17.094	13.007	22.944
14	1.214	7.703	6.670	19.423	14.554	24.962
15	1.360	8.683	7.690	21.682	16.153	27.135
16	1.471	9.467	8.477	23.342	18.073	30.030
17	1.576	10.101	9.191	24.920	19.563	32.209
18	1.799	11.639	10.693	28.174	21.662	35.426
19	2.016	13.039	12.024	30.736	24.160	38.465
20	2.127	13.900	13.089	33.313	25.479	40.583
21	2.327	15.214	14.440	36.109	27.687	44.104
22	2.569	16.806	16.191	39.673	30.260	47.239
23	2.785	18.330	17.332	41.955	32.645	50.881
24	2.899	19.020	18.699	45.056	35.011	54.426
25	3.108	20.562	20.464	48.647	37.481	58.172
26	3.278	21.837	21.667	50.983	39.907	60.842
27	3.652	24.350	23.851	55.514	41.146	63.561
28	3.910	26.248	25.538	59.278	43.212	66.782
29	4.015	27.089	26.762	61.311	47.135	71.577
30	4.120	27.758	27.874	64.016	50.134	76.343

NOTE: Entries are the value of the test statistic, for which the value of λ given in the first column is the median-unbiased estimator. Care must be taken to impose the normalization D = 1 when using these estimates of λ. Estimates of τ are computed as λ/T. If the test statistic takes on a value smaller than that in the first row, then the median-unbiased estimate is 0. Estimates for other values of the test statistics can be obtained by interpolation. For example, suppose that QLR = 5.0 is obtained empirically; using linear interpolation, the median unbiased estimator of λ is 4 + (5.0 − 4.848)/(5.689 − 4.848). A software implementation that handles general $\mathbf{X}_t$ for the case $D = \mathbf{I}_k$ is available from the authors by request. All entries in the table were estimated using 5,000 replications and T = 500.

values of λ $(5 \leq \lambda \leq 8)$. However, their performance deteriorates for large values of λ $(\lambda > 10)$.

One way to calibrate the magnitude of λ is to compare it to the asymptotic powers given in Figure 1. When $\lambda = 4$, the tests have rejection probabilities of approximately 25%; when $\lambda = 7$, the rejection probabilities are approximately 50%. For $\lambda \geq 14$, the power exceeds 80%. As an empirical guideline, this suggests that the median unbiased estimators will be roughly as efficient as the MMLE when the results of stability tests are ambiguous. When there is substantial instability, the MMLE will be more efficient than the median unbiased estimators.

Table 3 is a lookup table that permits computing median unbiased estimates, given a value of the test statistic. The normalization used in Table 3 is that D = 1, and users of this lookup table must be sure to impose this normalization when using the resulting estimator of λ.

4. APPLICATION TO TREND GROSS OF U.S. GROSS DOMESTIC PRODUCT

The issues of whether there has been a decline in the long-run growth rate of output in the United States, when this decline took place, how large the decline has been, and whether it has recently been reversed are of considerable practical and policy interest. Following Harvey (1985), we examine these issues using the local-level model in which the growth rate of output is allowed to have a small random-walk component. This introduces the possibility of a persistent decline in mean output growth, consistent with the productivity slowdown.

Table 4. Postwar U.S. GDP Growth, 1947:II–1995:IV: Tests of $\tau = 0$, Median-Unbiased Estimates, and 90% Confidence Intervals

Test	Statistic (p-value)	$\hat{\lambda}$	(90% CI)	$\hat{\sigma}_{\Delta\beta}$	(90% CI)
L	.21 (.25)	4.1	(0, 19.4)	.13	(0, .62)
MW	1.16 (.29)	3.4	(0, 18.8)	.11	(0, .60)
EW	.68 (.32)	3.1	(0, 17.0)	.10	(0, .55)
QLR	3.31 (.48)	.8	(0, 13.3)	.03	(0, .41)
POI(7)	2.90 (.45)	1.7	(0, 12.9)	.05	(0, .37)
POI(17)	7.52 (.54)	0	(0, 11.3)	0	(0, .36)

NOTE: $\hat{\sigma}_{\Delta\beta}$ is the estimate of the standard deviation of $\Delta\beta_t$ in (11); that is, $\hat{\sigma}_{\Delta\beta} = T^{-1}\hat{\lambda}\hat{\sigma}_\varepsilon/\hat{a}(1)$.

The data used are real quarterly values of GDP per capita from 1947:II–1995:IV. The data from 1959:I–1995:IV are the GDP chain-weighted quantity index, quarterly, seasonally adjusted (Citibase series GDPFC). The data from 1947:I–1958:IV are real GDP in 1987 dollars, seasonally adjusted (Citibase series GDPQ, in releases prior to 1996) and proportionally spliced to the GDP chain-weighted quantity index in 1959:I. These series were deflated by the civilian population (Citibase series P16). This GDP series was transformed to (approximate) percentage growth at an annual rate, GY_t, by setting $GY_t = 400\Delta\ln(\text{real per capita GDP})$. The model is

Table 5. Estimates of Parameters in (11)–(13) for Various Values of λ and Implied Subsample Trend Growth Rates

Parameter	MPLE	MMLE	Estimates with fixed λ	
Parameter estimates				
$\sigma_{\Delta\beta}$	0	.04	.13	.62
σ_ε	3.85 (.17)	3.86 (.17)	3.85 (.17)	3.78 (.20)
ρ_1	.33 (.06)	.34 (.07)	.34 (.07)	.32 (.08)
ρ_2	.13 (.06)	.13 (.07)	.13 (.06)	.12 (.07)
ρ_3	−.01 (.07)	−.01 (.08)	−.01 (.07)	−.01 (.08)
ρ_4	−.09 (.06)	−.08 (.06)	−.09 (.06)	−.09 (.07)
β_0	1.80 (.46)		2.44 (.84)	2.67 (2.25)

Estimated average trend growths

Date	$\overline{GY}$	MPLE	MMLE	$\sigma_{\Delta\beta}$ = .13	$\sigma_{\Delta\beta}$ = .62
1947–1995	1.80	1.80	1.80	1.80	1.80
1947–1970	2.46	1.80	1.89	2.16	2.43
1970–1995	1.22	1.80	1.71	1.47	1.23
1950–1960	2.75	1.80	1.91	2.25	2.27
1960–1970	2.39	1.80	1.84	1.98	2.39
1970–1980	1.20	1.80	1.75	1.56	1.07
1980–1990	1.58	1.80	1.70	1.45	1.50
1990–1995	.62	1.80	1.68	1.36	1.04

NOTE: Estimates were computed by Gaussian maximum likelihood, with numerical standard errors computed from the inverse of the outer product estimate of the Hessian. Unrestricted MLEs (standard errors in parentheses) are reported in the first two columns. (Because of the nonnormal distribution of the MLE of λ, the standard error for $\sigma_{\Delta\beta}$ is not reported.) The last two sets of columns report estimates by restricted MLE, with λ fixed to the indicated values. The column labeled $\overline{GY}$ in the second part of the table is the sample mean of GY; the other entries are average values of $\beta_{t|T}$ over the indicated subsample for the indicated model, where $\beta_{t|T}$ are the estimates of β_t obtained from the Kalman smoother.

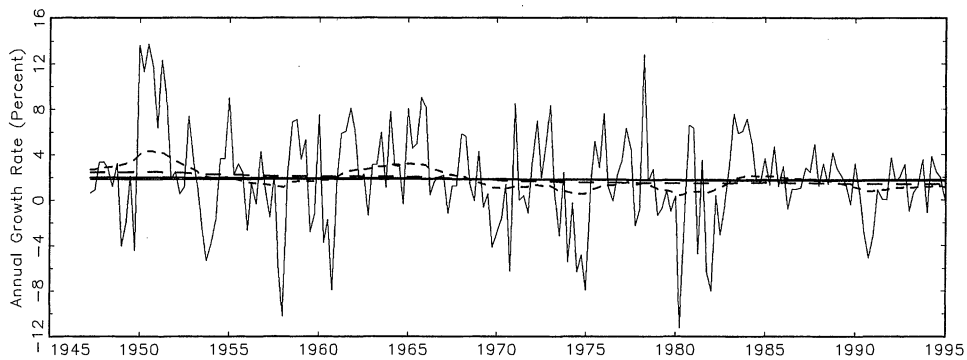

Figure 3. Growth Rate in U.S. Real per Capita GDP and Estimated Trends Based On the Four Models in Table 5. ——, GY; ——, MPLE; ——, MMLE; – – –, $\sigma_{\delta\beta}$ = .13; - - -, $\sigma_{\delta\beta}$ = .62.

$$GY_t = \beta_t + u_t, \tag{11}$$

$$\Delta\beta_t = (\lambda/T)\eta_t, \tag{12}$$

and

$$a(L)u_t = \varepsilon_t, \tag{13}$$

where the order $p = 4$ is used for $a(L)$. (The results are insensitive to choice of the AR order or to substituting an ARMA(2, 3) parameterization for $a(L)$, the latter being consistent with Harvey's [1985] original unobserved components formulation.) Estimates of λ are constructed using the normalization $\mathrm{D} = 1$ (i.e., $\sigma_\eta^2 = \sigma_\varepsilon^2/a(1)^2$) as discussed in Section 2.

It is worth digressing to discuss the implications of this model for orders of integration and unit roots. If there is a random-walk component in GY_t, then the logarithm of real per capita GDP is $I(2)$. This hypothesis is soundly rejected by unit root tests applied to these data. However, when the variance of $\Delta\beta_t$ is small, the model implies that ΔGY_t has a nearly unit MA root. Because it is well known that tests for a unit AR root have a high false-rejection rate under the null of a unit AR root when there is a nearly unit MA root (Pantula 1991; Schwert 1989), these rejections are consistent with the postulated model.

Test statistics, median unbiased estimates, and equal-tailed confidence intervals for λ and the standard deviation of $\Delta\beta$ are presented in Table 4. None of the tests rejects at the 10% level. Of course, this could mean that the tests have insufficient power to detect a small but nonzero value of λ—and indeed the median-unbiased estimates are, with only one exception, nonzero. The median unbiased estimates of λ are all small, ranging from 0 (POI(17)) to 4.1 (L). These correspond to point estimates of $\sigma_{\Delta\beta}$, the standard deviation of $\Delta\boldsymbol{\beta}_t$, ranging from 0% to .13%. This range of estimates is consistent with intuition. For example, a value of $\sigma_{\Delta\beta} = .1$ corresponds to a standard deviation of $\beta_{1995:\text{IV}}-\beta_{1947:\text{II}}$ of 1.4 percentage points.

Estimates of the model parameters are presented in the top part of Table 5, for various values of λ: the MPLE and the MMLE, the median-unbiased estimate based on the L (which is the largest of the point estimates in Table 4), and the upper end of the 90% confidence interval for λ based on L (the largest such value for the 90% confidence intervals). Consistent with the large pile-up probability discussed in Section 3, $\hat{\lambda}_{\text{MPLE}} = 0$. The MMLE produces a small but nonzero estimate of $\sigma_{\Delta\beta}$ equal to .04%, which corresponds to a point estimate of λ of 1.4. Estimates of parameters of the u_t process change little for this range of value of $\sigma_{\Delta\beta}$, although estimates of the initial value of the trend growth rate increase (as do their standard errors) as $\sigma_{\Delta\beta}$ increases. These results are broadly consistent with other results reported in the literature. For example, Harvey (1985) reported a MMLE point estimate of $\sigma_{\Delta\beta} = 0$ for annual U.S. real GNP data from 1909–1970. Harvey and Jaeger (1993) reported a larger point estimate ($\hat{\sigma}_{\Delta\beta} = .36$) constructed from a frequency domain estimator and quarterly real GNP data from 1954–1989.

Estimates of the trend growth rates $\beta_{t|T}$ based on these models over various time spans (computed using the

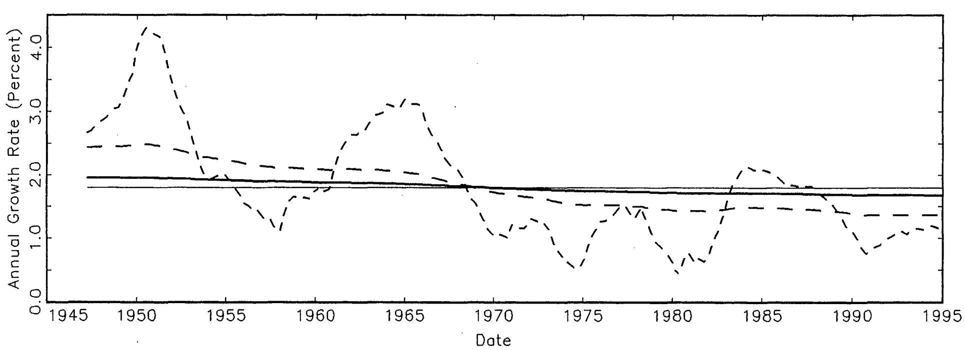

Figure 4. Estimated trends of U.S. Real GDP Growth Based on the Four Models in Table 5. ——, MPLE; ——, MMLE; – – –, $\sigma_{\delta\beta}$ = .13; - - -, $\sigma_{\delta\beta}$ = .62.

Kalman smoother) are given in the bottom part of Table 5, and these series are plotted in Figures 3 and 4. Figure 3 includes the raw data (GY_t); this series is omitted from Figure 4, in which the scale is enlarged. No large mean shift is evident in the raw data, consistent with the small estimates of $\sigma_{\Delta\beta}$ found using the various methods. The estimate of trend per capita GDP growth based on the MPLE is, of course, a horizontal line showing the mean of the raw data. In contrast, the other estimates reflect, to a varying degree, a slowdown in mean GDP growth over this period. The point estimate based on L implies a slowdown in the annual trend growth rate of approximately .9% per annum from the 1950s to the 1990s. Finally, none of the methods detects any substantial increase in trend GDP growth over the 1990s relative to the 1980s; indeed, all of the point estimates suggest a modest decrease.

5. DISCUSSION AND CONCLUSIONS

The median unbiased estimators developed here provide empirical researchers with a device to circumvent the undesirable pile-up problem and bias of the MLE in the TVP model when coefficient variation is small. The L_T- and F_T-based test statistics are easily computed using statistics from the GLS regression of y_t on $\mathbf{X}_t$. Given these statistics, the median unbiased estimator can be obtained by interpolating the entries in a lookup table. Such a lookup table is provided here for the univariate local levels model (Table 3), and lookup tables in higher dimensions for the normalization $D = \mathbf{I}_k$ are available from the authors on request.

In the special case of the univariate local-level model, we examined six asymptotically median unbiased estimators and two MLEs and found considerable differences among them. The MLEs were badly biased, particularly the MPLE. When the variance of the coefficients is small, the median unbiased estimators based on the QLR and POI(17) test statistics had good AREs. Because no asymptotic theory for the POI tests in the TVP model appears to be available outside the case $\mathbf{X}_t = 1$, and because the POI tests are somewhat cumbersome to compute even in the local-level model, these results provide support for using the QLR-based median unbiased estimators in the general TVP model when the coefficient instability is small.

APPENDIX: PROOF OF THEOREM 1

Before proving Theorem 1, we state and prove two preliminary lemmas. Let $\hat{\mathbf{U}}_{t-1} = (\hat{u}_{t-1}, \ldots, \hat{u}_{t-p})'$, $\mathbf{A} = (-a_1, -a_2, \ldots, -a_p)'$, and $\hat{\mathbf{A}} = (\hat{\mathbf{U}}'_{-1}\hat{\mathbf{U}}_{-1})^{-1}(\hat{\mathbf{U}}'_{-1}\hat{\mathbf{u}})$ using the usual matrix notation.

Lemma A1. Under assumptions A–C, $T^{1/2}(\bar{\mathbf{A}} - \mathbf{A}) = O_p(1)$.

Proof. The result follows by showing $T^{1/2}(\hat{\mathbf{A}} - \bar{\mathbf{A}}) \xrightarrow{p} 0$, where $\bar{\mathbf{A}} = (\mathbf{U}'_{-1}\mathbf{U}_{-1})^{-1}(\mathbf{U}'_{-1}\mathbf{u})$, where $\mathbf{U}_{t-1} = (u_{t-1}, \ldots, u_{t-p})'$. After straightforward algebra, it is seen that this follows if $\boldsymbol{\mu}_{1T} \xrightarrow{p} 0$ and $\boldsymbol{\mu}_{2T} \xrightarrow{p} 0$, where $\boldsymbol{\mu}_{1T}$ and $\boldsymbol{\mu}_{2T}$ are matrices with (i,j) elements, $\boldsymbol{\mu}_{1t,ij} = T^{-1/2}\sum_{t=1}^{T} u_{t-j}\mathbf{X}'_{t-i}(\boldsymbol{\beta}_{t-i} - \hat{\boldsymbol{\beta}})$ and $\boldsymbol{\mu}_{2t,ij} = T^{-1/2}\sum_{t=1}^{T}(\boldsymbol{\beta}_{t-i} - \hat{\boldsymbol{\beta}})'\mathbf{X}_{t-i}\mathbf{X}'_{t-j}(\boldsymbol{\beta}_{t-j} - \hat{\boldsymbol{\beta}})$. These limits follow using the Markov and Chebyschev inequalities and applying assumptions A–C, assuming that $T^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \Rightarrow O_p(1)$. An $O_p(1)$ limiting representation for $T^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ can be obtained using the methods in the proof of Theorem 1, but showing the $T^{1/2}$ rate (which is all that is required here) can be verified directly using Chebyschevs' inequality.

Lemma A2. Let $\mathbf{z}_t$ be a mean 0 stationary vector stochastic process with fourth-order cumulants that satisfy $\sum_{r_1,r_2,r_3=-\infty}^{\infty} |C(r_1, r_2, r_3)| < \infty$. Let w_t be either a scalar nonrandom sequence or a random variable that is independent of $\mathbf{z}_t$ for which $\sup_i \sup_{t\geq 1} E|\mathbf{z}_{it}|^4 < \infty$ and $\sup_{t\geq 1} E|w_t|^4 < \infty$. Then $T^{-1}\sum_{t=1}^{[Ts]} \mathbf{z}_t w_t \xrightarrow{p} 0$ uniformly in s.

Proof. First let z_t be a scalar. For $\delta > 0$,

$$
\begin{aligned}
\Pr\left[\sup_s \left| T^{-1}\sum_{t=1}^{[Ts]} z_t w_t \right| > \delta \right] &\leq \delta^{-4} E \max_{1\leq r\leq T} \left(T^{-1}\sum_{t=1}^{r} z_t w_t \right)^4 \\
&\leq \delta^{-4} E \sum_{r=1}^{T} \left(T^{-1}\sum_{t=1}^{r} z_t w_t \right)^4 \\
&\leq \delta^{-4} T^{-3} \max_{1\leq r\leq T} E \left(\sum_{t=1}^{r} z_t w_t \right)^4 \\
&\leq \delta^{-4} T^{-3} \sup_t E|w_t|^4 \sum_{t_1,t_2,t_3,t_4=1}^{T} |E(z_{t_1}\cdots z_{t_4})| \\
&\leq \delta^{-4} T^{-3} \sup_t E|w_t|^4 \left\{ \sum_{t_1,t_2,t_3,t_4=1}^{T} |C(t_1 - t_4, t_3 - t_4, t_3 - t_4)| \right. \\
&\qquad \left. + 3\left[\sum_{t_1,t_2=1}^{T} |C(t_1 - t_2)| \right]^2 \right\} \\
&\leq \delta^{-4} T^{-3} \sup_t E|w_t|^4 \left\{ T \sum_{r_1,r_2,r_3=-\infty}^{\infty} |C(r_1, r_3, r_3)| \right. \\
&\qquad \left. + 3T^2 \left[\sum_{r_1=-\infty}^{\infty} |C(r_1)| \right]^2 \right\}.
\end{aligned}
$$

Under the stated assumptions, the final expression $\to 0$, proving uniform consistency. This extends to vector $\mathbf{z}_t$ by replacing $\mathbf{z}_t$ by $\mathbf{z}_{it}$.

Proof of Theorem 1

Let $a(L) = \sum_{i=0}^{p} a_i L^i$ with $a_0 = 1$, and let $\mathbf{X}_t^{\dagger} = a(L)\mathbf{X}_t$. An implication of assumptions A, B, and C, or alternatively of assumptions A and D, is that

$$
\left(\sigma_\varepsilon^{-1}\boldsymbol{\Gamma}^{-1/2} T^{-1/2} \sum_{t=1}^{[Ts]} \mathbf{X}_t^{\dagger}\varepsilon_t, \boldsymbol{\Omega}^{-1/2} T^{-1/2} \sum_{t=1}^{[Ts]} \boldsymbol{\nu}_t \right) \Rightarrow (\mathbf{W}_1, \mathbf{W}_2), \quad \text{(A.1)}
$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are independent k-dimensional standard Brownian motions.

We first prove the theorem under assumptions A, B, and C.

Proof of Part a

Let $\tilde{u}_t = \hat{a}(L)u_t$ and $\tilde{\omega}_t = -\sum_{j=1}^{p} \hat{a}_j \mathbf{X}'_{t-j} \sum_{i=0}^{j-1} \mathbf{v}_{t-i}$, so that

$\tilde{y}_t = \beta_0' \tilde{\mathbf{X}}_t + (\sum_{r=1}^{t} \mathbf{v}_r)' \tilde{\mathbf{X}}_t + \tilde{\omega}_t + \tilde{u}_t$. Accordingly,

$$\boldsymbol{\xi}_T(s) = \boldsymbol{\xi}_{1T}(s) + \lambda \boldsymbol{\xi}_{2T}(s) + \boldsymbol{\xi}_{3T}(s) - \boldsymbol{\kappa}_T(s)\{\boldsymbol{\xi}_{1T}(1) + \lambda \boldsymbol{\xi}_{2T}(1) + \boldsymbol{\xi}_{3T}(1)\}, \quad \text{(A.2)}$$

where

$$\boldsymbol{\xi}_{1T}(s) = T^{-1/2} \sum_{t=1}^{[Ts]} \tilde{\mathbf{X}}_t \tilde{u}_t,$$

$$\boldsymbol{\xi}_{2T}(s) = T^{-3/2} \sum_{t=1}^{[Ts]} \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t' \sum_{r=1}^{t} \boldsymbol{\nu}_r,$$

$$\boldsymbol{\xi}_{3T}(s) = T^{-1/2} \sum_{t=1}^{[Ts]} \tilde{\mathbf{X}}_t \tilde{\omega}_t,$$

and

$$\boldsymbol{\kappa}_T(s) = \left[T^{-1} \sum_{t=1}^{[Ts]} \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t' \right] \left[T^{-1} \sum_{t=1}^{T} \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t' \right]^{-1}.$$

Limits are obtained for these terms in turn. All limits are uniform in $s \in [0, 1]$.

1. Write $\boldsymbol{\xi}_{1T}(s) = \boldsymbol{\Delta}_{1T}(s) + \boldsymbol{\Delta}_{2T}(s) + \boldsymbol{\xi}_{1T}^{\dagger}(s)$, where $\boldsymbol{\Delta}_{1T}(s) = \sum_{j=0}^{p} \hat{a}_j \sum_{i=1}^{p} T^{1/2}(\hat{a}_i - a_i)(T^{-1} \sum_{t=1}^{[Ts]} \mathbf{X}_{t-j} u_{t-i})$, $\Delta_{2T}(s) = T^{-1/2} \sum_{t=1}^{[Ts]} (\tilde{\mathbf{X}}_t - \mathbf{X}_t^{\dagger}) \varepsilon_t$, and $\boldsymbol{\xi}_{1T}^{\dagger}(s) = T^{-1/2} \sum_{t=1}^{[Ts]} \mathbf{X}_t^{\dagger} \varepsilon_t$. Assumptions A, B, and C imply that $T^{-1} \sum_{t=1}^{[Ts]} \mathbf{X}_{t-j} u_{t-i}$ satisfies the conditions of Lemma A.2 with $\mathbf{z}_t = \mathbf{X}_{t-j} u_{t-i}$ and $w_t = 1$; so, because p is fixed, $\boldsymbol{\Delta}_{1T} \xrightarrow{p} 0$. By an analogous argument, $\boldsymbol{\Delta}_{2T} \xrightarrow{p} 0$. Using the limit in (A.1), we have $\boldsymbol{\xi}_{1T}(s) \Rightarrow \sigma_\varepsilon \boldsymbol{\Gamma}^{1/2} \mathbf{W}_1(s)$.

2. Write $\boldsymbol{\xi}_{2T}(s) = \boldsymbol{\Delta}_{3T}(s) + \boldsymbol{\Delta}_{4T}(s) + \boldsymbol{\xi}_{2T}^{\dagger}(s)$, where $\boldsymbol{\Delta}_{3T}(s) = T^{-3/2} \sum_{t=1}^{[Ts]} (\mathbf{X}_t^{\dagger} \mathbf{X}_t^{\dagger\prime} - \boldsymbol{\Gamma}) \sum_{r=1}^{t} \boldsymbol{\nu}_r$, $\boldsymbol{\Delta}_{4T}(s) = T^{-3/2} \sum_{t=1}^{[Ts]} (\tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t' - \mathbf{X}_t^{\dagger} \mathbf{X}_t^{\dagger\prime}) \sum_{r=1}^{t} \boldsymbol{\nu}_r$, and $\boldsymbol{\xi}_{2T}^{\dagger}(s) = \boldsymbol{\Gamma} T^{-3/2} \sum_{t=1}^{[Ts]} \sum_{r=1}^{t} \boldsymbol{\nu}_r$. To show $\boldsymbol{\Delta}_{3T} \xrightarrow{p} 0$ and $\boldsymbol{\Delta}_{4T} \xrightarrow{p} 0$, consider for notational simplicity the case $k = 1$. (The argument for $k > 1$ is similar.) Note that $T^{-2} \max_{t_1 \cdots t_4} E| \sum_{r_1=1}^{t_1} \boldsymbol{\nu}_{r_1} \cdots \sum_{r_4=1}^{t_4} \boldsymbol{\nu}_{r_4}| \Rightarrow \sup_{s_1 \cdots s_4} E|\mathbf{W}_2(s_1) \cdots \mathbf{W}_2(s_4)| \boldsymbol{\Omega}^2 < \infty$. Because $\mathbf{X}_t$ is stationary with absolutely summable eighth-order cumulants, $\mathbf{X}_t^{\dagger 2} - \boldsymbol{\Gamma}$ is stationary with absolutely summable fourth-order cumulants. Thus $\boldsymbol{\Delta}_{3T}$ satisfies the conditions of Lemma A.2 with $\mathbf{z}_t = \mathbf{X}_t^{\dagger 2} - \boldsymbol{\Gamma}$ and $\mathbf{w}_t = T^{-1/2} \sum_{r=1}^{t} \boldsymbol{\nu}_r$, so $\boldsymbol{\Delta}_{3T} \xrightarrow{p} 0$. Turning to $\boldsymbol{\Delta}_{4T}, \boldsymbol{\Delta}_{4T} = \sum_{j=0}^{p} \sum_{i=0}^{p} (\hat{a}_i + a_i) T^{1/2} (\hat{a}_j - a_j) \boldsymbol{\Delta}_{4T,ij}(s)$, where $\boldsymbol{\Delta}_{4T,ij}(s) = [T^{-3/2} \sum_{t=1}^{[Ts]} \mathbf{X}_{t-j} \mathbf{X}_{t-i} T^{-1/2} \sum_{r=1}^{t} \boldsymbol{\nu}_r]$. An argument analogous to that used for $\boldsymbol{\Delta}_{3T}$ shows that $\boldsymbol{\Delta}_{4T,ij} \xrightarrow{p} 0$ and, because p is finite, $\boldsymbol{\Delta}_{4T} \xrightarrow{p} 0$. The limit of $\boldsymbol{\xi}_{2T}^{\dagger}$ follows from (A.1). Thus $\boldsymbol{\xi}_{2T}(s) \Rightarrow \boldsymbol{\Gamma} \boldsymbol{\Omega}^{1/2} \int_0^s \mathbf{W}_2(r)\, dr$.

3. Write $\boldsymbol{\xi}_{3T}(s) = -\lambda \sum_{j=0}^{p} \sum_{l=0}^{p} \sum_{i=0}^{j-1} \boldsymbol{\xi}_{3T,ilj}(s)$, where $\boldsymbol{\xi}_{3T,ilj}(s) = (T^{-3/2} \sum_{t=1}^{[Ts]} \mathbf{X}_{t-l} \mathbf{X}_{t-j}' \mathbf{v}_{t-i})$. As before, consider the case $k = 1$. Now, $T^{1/2} \boldsymbol{\xi}_{3T,ilj}(s)$ satisfies Lemma A.2 with $\mathbf{z}_t = \mathbf{X}_{t-l} \mathbf{X}_{t-j}' \mathbf{v}_{t-i}$ and $w_t = 1$; thus $\boldsymbol{\xi}_{3T} \xrightarrow{p} 0$.

4. Let $\boldsymbol{\Delta}_{5T}(s) = T^{-1} \sum_{t=1}^{[Ts]} (\tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t' - \mathbf{X}_t^{\dagger} \mathbf{X}_t^{\dagger\prime})$ and $\boldsymbol{\Delta}_{6T}(s) = T^{-1} \sum_{t=1}^{[Ts]} (\mathbf{X}_t^{\dagger} \mathbf{X}_t^{\dagger\prime} - \boldsymbol{\Gamma})$, and let $\boldsymbol{\Delta}_{7T} = \boldsymbol{\Delta}_{5T} + \boldsymbol{\Delta}_{6T} = T^{-1} \sum_{t=1}^{[Ts]} (\tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t' - \boldsymbol{\Gamma})$. The argument that $\boldsymbol{\Delta}_{5T} \xrightarrow{p} 0$ follows the argument that $\boldsymbol{\Delta}_{4T} \xrightarrow{p} 0$ with $T^{-1/2} \sum_{r=1}^{t} \boldsymbol{\nu}_r$ replaced by 1, and the argument that $\boldsymbol{\Delta}_{6T} \xrightarrow{p} 0$ follows the argument that $\boldsymbol{\Delta}_{3T} \xrightarrow{p} 0$ with the same replacement. Thus $\boldsymbol{\Delta}_{7T} \xrightarrow{p} 0$, so $\boldsymbol{\kappa}_T(s) \xrightarrow{p} s\mathbf{I}_k$.

Similar calculations imply that $\tilde{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_\varepsilon^2$ so $\tilde{\mathbf{V}} \xrightarrow{p} \boldsymbol{\Gamma} \sigma_\varepsilon^2$. By collecting terms and using (A.2), it follows that $\tilde{\mathbf{V}}^{-1/2} \boldsymbol{\xi}_T(s) \Rightarrow \mathbf{h}_\lambda(s) - s\mathbf{h}_\lambda(1)$, where $\mathbf{h}_\lambda(s) = \mathbf{W}_1(s) + \lambda(\boldsymbol{\Gamma}^{1/2} \boldsymbol{\Omega}^{1/2} / \sigma_\varepsilon) \int_0^s \mathbf{W}_2(r)\, dr$.

Proof of Part b

This follows from the continuous mapping theorem.

Proof of Part c

This follows by straightforward but tedious manipulations using the previous limiting results.

Next, turn to the proof under assumptions A and D. Under assumption D, $a(L) = \hat{a}(L) = 1$, so $\tilde{\mathbf{X}}_t = \mathbf{X}_t^{\dagger} = \mathbf{X}_t$ and $\tilde{e}_t = y_t - \hat{\beta}' \mathbf{X}_t$ (where $\hat{\beta}$ remains the OLS estimator). The proof under these conditions follows the foregoing proof but is simpler. In particular, (A.2) now holds with $\boldsymbol{\xi}_{1T}(s) = T^{-1/2} \sum_{t=1}^{[Ts]} \mathbf{X}_t \varepsilon_t, \boldsymbol{\xi}_{2T}(s) = T^{-3/2} \sum_{t=1}^{[Ts]} \mathbf{X}_t \mathbf{X}_t' \sum_{r=1}^{t} \boldsymbol{\nu}_r, \boldsymbol{\xi}_{3T}(s) = 0$, and $\boldsymbol{\kappa}_T(s) = [T^{-1} \sum_{t=1}^{[Ts]} \mathbf{X}_t \mathbf{X}_t'][T^{-1} \sum_{t=1}^{T} \mathbf{X}_t \mathbf{X}_t']^{-1}$. The limit of $\boldsymbol{\xi}_{1T}$ follows from (A.1). Write $\boldsymbol{\xi}_{2T}(s)$ as $\boldsymbol{\xi}_{2T}(s) = T^{-1} \sum_{t=1}^{[Ts]} (\mathbf{X}_t \mathbf{X}_t' - \boldsymbol{\Gamma})(T^{-1/2} \sum_{r=1}^{t} \boldsymbol{\nu}_r) + \boldsymbol{\Gamma} T^{-3/2} \sum_{t=1}^{[Ts]} \sum_{r=1}^{t} \boldsymbol{\nu}_r$. The first term in this expression $\xrightarrow{p} 0$ as a consequence of Lemma A.2 and the independence of $\{\boldsymbol{\nu}_t, t = 1, \ldots, T\}$ and $\{\mathbf{X}_t, t = 1, \ldots, T\}$, as discussed earlier for the term $\boldsymbol{\Delta}_{3T}$. The limit of the second term follows from (A.1) and the continuous mapping theorem. The argument given earlier for $\boldsymbol{\kappa}_T(s) \xrightarrow{p} s\mathbf{I}_k$ applies under these assumptions, and $\tilde{\mathbf{V}} \xrightarrow{p} \sigma_\varepsilon^2 \boldsymbol{\Gamma}$. This proves part (a) under assumptions A and D; the proof of parts (b) and (c) follows accordingly.

[Received August 1996. Revised June 1997.]

REFERENCES

Andrews, D. W. K., Lee, I., and Ploberger, W. (1996), "Optimal Changepoint Tests for Normal Linear Regression," *Journal of Econometrics*, 70, 9–38.

Andrews, D. W. K., and Ploberger, W. (1994), "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414.

Brillinger, D. R. (1981), *Time Series: Data Analysis and Theory*, San Francisco: Holden-Day.

Chow, G. C. (1984), "Random and Changing Coefficient Models," in *Handbook of Econometrics*, eds. Z. Griliches and M. D. Intrilligator, Amsterdam: North-Holland, chap. 21.

Cooley, T. F., and Prescott, E. C. (1973a), "An Adaptive Regression Model," *International Economic Review*, 14, 364–371.

——— (1973b), "Tests of an Adaptive Regression Model," *Review of Economics and Statistics*, 55, 248–256.

——— (1976), "Estimation in the Presence of Stochastic Parameter Variation," *Econometrica*, 44, 167–184.

Davis, R. A., and Dunsmuir, W. T. M. (1996), "Maximum Likelihood Estimation for MA(1) Processes with a Root on or Near the Unit Circle," *Econometric Theory*, 12, 1–29.

Engle, R. F., and Watson, M. W. (1987), "The Kalman Filter Model: Applications to Forecasting and Rational Expectations Models," in *Advances in Econometrics: Fifth World Congress of the Econometric Society*, ed. T. Bewley, Cambridge, U.K.: Cambridge University Press.

Franzini, L., and Harvey, A. C. (1983), "Testing for Deterministic Trend and Seasonal Components in Time Series Models," *Biometrika*, 70, 673–682.

Hansen, B. E. (1992), "Tests for Parameter Instability in Regressions with $I(1)$ Processes," *Journal of Business and Economic Statistics*, 10, 321–336.

Harvey, A. C. (1985), "Trends and Cycles in Macroeconomic Time Series," *Journal of Business and Economic Statistics*, 3, 216–227.

——— (1989), *Forecasting, Structural Time Series Models, and the Kalman Filter*, Cambridge, U.K.: Cambridge University Press.

Harvey, A. C., and Jaeger, A. (1993), "Detrending, Stylized Facts and the Business Cycle," *Journal of Applied Econometrics*, 8, 231–248.

Harvey, A. C., and Streibel, M. (1997), "Testing Against Nonstationary Unobserved Components," unpublished manuscript, University of Cambridge, Faculty of Economics and Politics.

King, M. L., and Hillier, G. H. (1985), "Locally Best Invariant Tests of the Error Covariance Matrix of the Linear Regression Model," *Journal of the Royal Statistical Society*, Ser. B, 47, 98–102.

Nabeya, S., and Tanaka, K. (1988), "Asymptotic Theory of a Test for the Constancy of Regression Coefficients Against the Random Walk Alternative," *The Annals of Statistics*, 16, 218–235.

Nichols, D. F., and Pagan, A. R. (1985), "Varying Coefficient Regression," in *Handbook of Statistics*, Vol. 5, eds. E. J. Hannan, P. R. Krishnaiah, and M. M. Rao, Amsterdam: North-Holland, pp. 413–449.

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

Nyblom, J., and Mäkeläinen, 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. (1980), "Some Identification and Estimation Results for Regression Models With Stochastically Varying Coefficients," *Journal of Econometrics*, 12, 341–363.

Pantula, S. G. (1991), "Asymptotic Distributions of the Unit-Root Tests When the Process is Nearly Stationary," *Journal of Business and Economic Statistics*, 9, 325–353.

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

Reinsel, G. C., and Tam, W.-K. (1996), "Tests for Seasonal Moving Average Unit Root in ARIMA Models," unpublished manuscript, University of Wisconsin-Madison, Dept. of Statistics.

Rosenberg, B. (1972), "The Estimation of Stationary Stochastic Regression Parameters Re-Examined," *Journal of the American Statistical Association*, 67, 650–654.

——— (1973), "The Analysis of a Cross-Section of Time Series by Stochastically Convergent Parameter Regression," *Annals of Economic and Social Measurement*, 2, 461–484.

Saikkonen, P,. and Luukonen, R. (1993), "Point Optimal Test for Testing the Order of Differencing in ARIMA Models," *Econometric Theory*, 9, 343–362.

Sarris, A. H. (1973), "A Bayesian Approach to Estimation of Time-Varying Regression Coefficients," *Annals of Economic and Social Measurement*, 2, 501–523.

Sargan, J. D., and Bhargava, A. (1983), "Maximum Likelihood Estimation of Regression Models With First-Order Moving Average Errors When the Root Lies on the Unit Circle," *Econometrica*, 51, 799–820.

Schwert, G. W. (1989), "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147–160.

Shephard, N. H. (1993), "Distribution of the ML Estimator of an MA(1) and a Local Level Model," *Econometric Theory*, 9, 377–401.

Shephard, N. G., and Harvey, A. C. (1990), "On the Probability of Estimating a Deterministic Component in the Local Level Model," *Journal of Time Series Analysis*, 11, 339–347.

Shively, T. S. (1988), "An Exact Test of a Stochastic Coefficient in a Time Series Regression Model," *Journal of Time Series Analysis*, 9, 81–88.

Stock, J. H. (1991), "Confidence Intervals for the Largest Autoregressive Root in U.S. Economic Time Series," *Journal of Monetary Economics*, 28, 435–460.

——— (1994), "Unit Roots, Structural Breaks, and Trends," in *The Handbook of Econometrics*, Vol. IV, eds. R. F. Engle and D. McFadden, Amsterdam: Elsevier.

Stock, J. H., and Watson, M. W. (1996), "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11–30.

Tanaka, K. (1990), "Testing for a Moving Average Unit Root," *Econometric Theory*, 6, 433–444.