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

\[ y_t = \beta_t X_t + u_t, \]  
\[ \beta_t = \beta_{t-1} + v_t, \]  
\[ a(L) u_t = \varepsilon_t, \]

and

\[ v_t = \tau \nu_t, \quad \text{where} \quad \nu_t = B(L) \eta_t, \]

where \{(y_t, X_t), t = 1, \ldots, T\} are observed, \(X_t\) is an exogenous \(k\)-dimensional regressor, \(\beta_t\) is a \(k \times 1\) vector of unobserved time-varying coefficients, \(\tau\) is a scalar, \(a(L)\) is a scalar lag polynomial, \(B(L)\) is a \(k \times k\) matrix lag polynomial, and \(\varepsilon_t\) and \(\eta_t\) 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 \(B(L)\) and \(E(\eta_t \eta_t')\) that are sufficient to identify the scale factor \(\tau\).) An important special case of this model is when \(X_t = 1\) and \(B(L) = 1\); following Harvey (1985), we refer to this case as the “local-level” unobserved components model.

We consider the problem of estimation of the scale parameter \(\tau\). If \((as is common)\eta_t\) and \(\eta_t\) are assumed to be jointly normal and independent of \(\{X_t, t = 1, \ldots, T\}\), 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 \(\tau\) is small, then it has point mass at 0. In the case \(X_t = 1\), 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 \(X_t\) and can be large. The pile-up probability is a particular problem when \(\tau\) is small and thus is readily mistaken for 0. Arguably, small values of \(\tau\) are appropriate for many empirical applications; indeed, if \(\tau\) 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 \(X_t = 1\).)

We thus focus on the estimation of \(\tau\) when it is small. In particular, we consider the nesting

\[ \tau = \lambda / T. \]

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 \(\lambda\) and asymptotically valid confidence intervals for \(\lambda\) in the model (1)–(5). These are obtained by inverting asymptotic quantile functions of statistics that test the hypothesis \(\lambda = 0\). The
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 \( \lambda \). 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 \( \beta_t \). Both MLEs are biased and have large pile-ups at \( \lambda = 0 \). When \( \lambda \) is small, the median unbiased estimators are more tightly concentrated around the true value of \( \lambda \) 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 \( X_t \), 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) \( \tilde{y}_t = \hat{a}(L)y_t \) is regressed on \( \tilde{X}_t \) to yield the GLS estimator \( \hat{\beta} = T^{-1} \sum_{t=1}^{T} \tilde{X}_t \tilde{y}_t \), residuals \( \tilde{e}_t \) and moment matrix \( \tilde{V} \):

\[
\tilde{e}_t = \tilde{y}_t - \hat{\beta} \tilde{X}_t
\]

and

\[
\tilde{V} = T^{-1} \sum_{t=1}^{T} \tilde{X}_t \tilde{X}_t' \tilde{\sigma}_e^2,
\]

where \( \tilde{\sigma}_e^2 = (T-k)^{-1} \sum_{t=1}^{T} \tilde{e}_t^2 \). If \( a(L) = 1 \), then steps (a) and (b) are omitted and the OLS and GLS regressions of \( y_t \) on \( X_t \) 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, \( F_T(s)(0 \leq s \leq 1) \), which test for a break at date \([T_s]\), where \([\cdot]\) denotes the greatest lesser integer. Let \( \text{SSR}_{T_s,t_s} \) denote the sum of squared residuals from the GLS regression of \( \tilde{y}_t \) onto \( \tilde{X}_t \) over observations \( t_1 \leq t \leq t_2 \), and let \( \xi_T(s) = T^{-1/2} \sum_{t=1}^{T_s} \tilde{X}_t \tilde{e}_t \). The \( L_T \) and \( F_T \) statistics are

\[
L_T = T^{-1} \sum_{t=1}^{T} \xi_T(t/T)' \tilde{V}^{-1} \xi_T(t/T)
\]

and

\[
F_T(s) = \frac{\text{SSR}_{1,T} - \text{SSR}_{1,\lceil Ts \rceil} - \text{SSR}_{\lceil Ts \rceil + 1,T}}{[k(\text{SSR}_{1,\lceil Ts \rceil} + \text{SSR}_{\lceil Ts \rceil + 1,T})/(T-k)]}. \quad (9)
\]

(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), \( QLR_T = \sup_{s \leq (s_0, s_1)} F_T(s) \); the mean Wald statistic of Andrews and Ploberger (1994) and Hansen (1992), \( MW_T = \int_{s_0}^{s_1} F_T(r) \, dr \); and the Andrews–Ploberger (1994) exponential Wald statistic, \( EW_T = \ln \left( \frac{\int_{s_0}^{s_1} \exp \left( \frac{1}{2} F_T(r) \right) \, dr}{s_1 - s_0} \right) \), where \( 0 < s_0 < s_1 < 1 \).

Three assumptions are used to obtain the asymptotic results. For a stationary process \( x_t \), let \( c_{i_1, \ldots, i_n} (r_1, \ldots, r_{n-1}) \) denote the \( n \)th joint cumulant of \( x_{t_1+i_1}, \ldots, x_{t_n+i_n} \), 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.** \( X_t \) is stationary with eighth order cumulants that satisfy \( \sum_{r_1, \ldots, r_{n-1}} |C(r_1, \ldots, r_{n-1})| < \infty \).

**Assumption B.** \( \{X_t, t = 1, \ldots, T\} \) is independent of \( \{u_t, \nu_t, t = 1, \ldots, T\} \).

**Assumption C.** \( (\varepsilon_t, \eta' t)^\prime \) is a \( (k+1) \times 1 \) vector of iid errors with mean 0 and four moments; \( \varepsilon_t \) and \( \eta_t \) are independent; \( a(L) \) has finite-order \( p \); and \( B(L) \) is one-summable with \( B(1) \neq 0 \).

Assumption A requires that \( X_t \) 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 \( X_t \) not be integrated of order 1 (\( I(1) \)) or higher is essential for our results.

Assumption B requires \( X_t \) to be strictly exogenous. This assumption permits estimation of (1), under the null \( \beta_t = \beta_0 \), 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 \( \varepsilon_t \) and \( \eta_t \) are independent ensures that \( u_t \) and \( v_t \) have a zero cross-spectral density matrix. This is a basic identifying assumption of the TVP model (Harvey 1989). To construct the Gaussian MLE, \( \varepsilon_t \) and \( \eta_t \) are modeled as independent iid normal random variables.

The assumption that \( X_t \) is independent of the errors can be unappealing in some applications. For example, in some econometric applications \( X_t \) is predetermined but not strictly exogenous, \( u_t \) is plausibly serially uncorrelated, but there is feedback from \( u_t \) to future \( X_t \). In lieu of assumptions B and C, we thus introduce an alternative assumption to handle regressors that are predetermined but not exogenous.
Assumption D. \( (e_t, \eta_t') \) is a \((k + 1) \times 1\) vector of iid errors with mean 0 and four moments; \( e_t \) and \( \eta_t \) are independent; \( \alpha(L) = 1; B(L) \) is one-symmetric with \( B(1) \neq 0; \eta_t \) is independent of \( \{X_t, X_{t+1}, X_{t+2}, \ldots\} \); and \( u_t \) is independent of \( \{X_t, X_{t-1}, X_{t-2}, \ldots\} \).

This permits feedback from \( u_t \) to future \( X_t \), but not from \( v_t \) to \( X_t \), and thus rules out \( X_t \) 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 \( W_1 \) and \( W_2 \) be independent standard Brownian motions on \([0,1]^k\), and let \( \Gamma = E\{ [\alpha(L)X_t] [\alpha(L)X_t']\}, \) \( \Omega = B(1)E\{ \eta_t \eta_t' B(1)' \}, \) and \( D = \Gamma^{1/2} \Omega^{1/2}/\sigma_\varepsilon. \)

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

\[ a. \quad \tilde{V}^{-1/2} \xi_t \Rightarrow h_0, \text{ where } h_0(s) = h_\lambda(s) - s h_\lambda(1), \text{ where } h_\lambda(s) = W_1(s) + \lambda \int_0^s W_2(r) \, dr; \]

\[ b. \quad L_T \Rightarrow \int_0^t h_0(s) \, ds; \text{ and} \]

\[ c. \quad F_T \Rightarrow F^*, \text{ where } F^*(s) = h_0(s)/h_0'(s)/(ks(1-s)). \]

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 \( QLR_T \Rightarrow \sup_{s_0 \leq s \leq s_1} F^*(s), \) \( MW_T \Rightarrow \int_{s_0}^{s_1} F^*(r) \, dr, \) and \( EW_T \Rightarrow \ln \left( \int_{s_0}^{s_1} \exp \left( \frac{1}{2} F^*(r) \right) \, dr \right). \) 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 \( h_0 \) 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, \( \lambda \) and \( D \). The limiting representations in Theorem 1 are used for three purposes: to compute local asymptotic power functions, to construct median unbiased estimators of \( \lambda \), and to construct asymptotically valid confidence intervals for \( \lambda \). To do so, \( D \) must either be known or be consistently estimable, so that asymptotically \( \lambda \) is the only unknown parameter entering these distributions.

The determination of \( D \) raises issues of identification and modeling strategy. Evidently \( \alpha(L) \) and \( \text{var}(e_t) \) 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, \( B(L) \) and \( E_{\eta_t} \eta_t' \) are not separately identified; we thus adopt the conventional assumptions that \( B_0 = I_k \) and the roots of \( |B(z)| \) are outside the unit circle. Even with these assumptions, however, inspection of (1)-(5) reveals that \( \lambda \) and \( \Omega \) are not separately identified: The parameter combinations \( (\lambda, \Omega) = (\tilde{\lambda}, \tilde{\Omega}) \) and \( (\lambda, \Omega) = (1, \tilde{\lambda}^2 \tilde{\Omega}) \) are observationally equivalent for fixed \( (\tilde{\lambda}, \tilde{\Omega}) \).

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 \( D = 1. \) When \( X_t = 1, \) under this normalization, \( \lambda \) is \( T \) times the ratio of the long-run standard deviation of \( \Delta \beta_t \) to the long run standard deviation of \( u_t. \)

When \( k > 1, \) \( \Omega \) is identified upon making a single suitable normalization; for example, that the trace of \( \Omega \) is unity. However, the local-to-0 variation in \( \Delta \beta_t \) makes it impossible to estimate the free elements of \( \Omega \) consistently without further restrictions. In this case two types of further restrictions suggest themselves. First, \( \Omega \) 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, \( \Omega \) may be parameterized as a function of \( \Gamma \) and \( \sigma_\varepsilon^2 \) (which are consistently estimable). As in the \( k = 1 \) case, a convenient parameterization sets \( \Omega = \sigma_\varepsilon^2 \Gamma^{-1}, \) for this implies that \( D = I_k. \) This choice of \( \Omega \) 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 \( \lambda \) 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 \( \alpha(L) \) and \( \beta(L), \) long-run standard deviation units). With the additional restrictions \( \alpha(L) = 1 \) and \( B(L) = 1, \) 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 \( X_t \) 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 \( \lambda \)

Median unbiased estimators of \( \lambda \) 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 \( \lambda \) and \( D. \) Let \( m_\lambda \) denote the median of \( g(F^*) \) as a function of \( \lambda \) for a given matrix of nuisance parameters \( D. \) Suppose that \( m_\lambda(\cdot) \) is monotone increasing and continuous in \( \lambda. \) Then the inverse function \( m_\lambda^{-1} \) exists, and for \( D \) known, \( \lambda \) can be es-
timated by
\[ \hat{\lambda}_g = m_{D}^{-1}(g(F_T)). \] (10)

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

In practice \( D \) is not known, so the estimator (10) is infeasible. However, as discussed earlier, \( D \) generally can be consistently estimated for a given choice of \( \Omega \). If in addition \( m_{D}^{-1}(\cdot) \) is continuous in \( D \) (which it is for the functionals considered in this article), then (10) can be computed with \( D \) replaced by a consistent estimator \( \hat{D} \), and the same asymptotic distribution obtains. Note, however, that this is computationally cumbersome, as it requires computing the inverse median function \( m_{D}^{-1}(\cdot) \) for every estimate \( \hat{D} \) 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 \( D \).) When \( \Omega \) is chosen so that \( \hat{D} = I_k \), the limiting distributions of \( L_T \) and \( F_T \) depend only on \( \lambda \) 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 \( \lambda \)

Suppose that \( D = I_k \), in which case the local asymptotic representations in Theorem 1 depend only on \( \lambda \) 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 \( \lambda_0 \) that are not rejected. This set constitutes a 95% confidence set for \( \lambda_0 \). 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.

3. NUMERICAL RESULTS FOR THE UNIVARIATE LOCAL-LEVEL MODEL

In the univariate of local-level model, \( X_t = 1 \) and \( B(L) = 1 \), 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 \( \Delta y_t \) 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 Luukkonen 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 \( \lambda \) 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 \( \lambda \)

Each of the tests examined in Figure 1 has a power function that depends only on \( \lambda \) and has a median function that is monotone and continuous in \( \lambda \). Asymptotically median unbiased estimators of \( \lambda \) 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 \( \beta_0 \). The first estimator, the maximum profile (or concentrated) likelihood estimator (MLE), treats \( \beta_0 \) as an unknown nuisance parameter that is concentrated out of the likelihood. The second estimator, the maximum marginal likelihood estimator (MMLE), treats \( \beta_0 \) as a \( N(\bar{\beta}, \kappa) \) random variable that is independent of \( \{u_t, v_t, t = 1, \ldots, T\} \), so that \( \beta_0 \) is integrated out of the likelihood. When \( \kappa \to \infty \), this produces the “diffuse prior” likelihood function (see Shephard 1993.

![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).](image-url)
Table 1. Pile-Up Probability That $\lambda = 0$ for MLEs and Various Median-Unbiased Estimators

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>MPLE</th>
<th>MMLE</th>
<th>L</th>
<th>MW</th>
<th>EW</th>
<th>QLR</th>
<th>POI(7)</th>
<th>POI(17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.96</td>
<td>.66</td>
<td>.50</td>
<td>.50</td>
<td>.50</td>
<td>.50</td>
<td>.50</td>
<td>.50</td>
</tr>
<tr>
<td>1</td>
<td>.91</td>
<td>.60</td>
<td>.47</td>
<td>.47</td>
<td>.48</td>
<td>.47</td>
<td>.47</td>
<td>.47</td>
</tr>
<tr>
<td>2</td>
<td>.88</td>
<td>.57</td>
<td>.42</td>
<td>.42</td>
<td>.43</td>
<td>.44</td>
<td>.43</td>
<td>.43</td>
</tr>
<tr>
<td>3</td>
<td>.81</td>
<td>.47</td>
<td>.34</td>
<td>.34</td>
<td>.35</td>
<td>.35</td>
<td>.35</td>
<td>.35</td>
</tr>
<tr>
<td>4</td>
<td>.72</td>
<td>.40</td>
<td>.28</td>
<td>.28</td>
<td>.29</td>
<td>.29</td>
<td>.29</td>
<td>.29</td>
</tr>
<tr>
<td>5</td>
<td>.65</td>
<td>.35</td>
<td>.24</td>
<td>.24</td>
<td>.24</td>
<td>.24</td>
<td>.24</td>
<td>.24</td>
</tr>
<tr>
<td>7</td>
<td>.48</td>
<td>.24</td>
<td>.15</td>
<td>.15</td>
<td>.15</td>
<td>.15</td>
<td>.15</td>
<td>.15</td>
</tr>
<tr>
<td>9</td>
<td>.37</td>
<td>.17</td>
<td>.11</td>
<td>.11</td>
<td>.11</td>
<td>.11</td>
<td>.11</td>
<td>.11</td>
</tr>
<tr>
<td>10</td>
<td>.30</td>
<td>.13</td>
<td>.09</td>
<td>.09</td>
<td>.09</td>
<td>.09</td>
<td>.09</td>
<td>.09</td>
</tr>
<tr>
<td>12</td>
<td>.24</td>
<td>.09</td>
<td>.06</td>
<td>.06</td>
<td>.06</td>
<td>.06</td>
<td>.06</td>
<td>.06</td>
</tr>
<tr>
<td>14</td>
<td>.15</td>
<td>.03</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
</tr>
<tr>
<td>16</td>
<td>.13</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td>18</td>
<td>.09</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>20</td>
<td>.07</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>25</td>
<td>.03</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>30</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
</tbody>
</table>

NOTE: Entries for MPLE and MMLE for $\lambda = 0$ are from Shepard and Harvey (1990). Entries for other values of $\lambda$ 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 $0 \leq \lambda \leq 60$, 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 $\lambda$ 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 $\lambda$ increases, both in absolute terms (it is above 50% for $\lambda \leq 6$) 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 unbiasedness 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 $i$th estimator $\hat{\tau}_i$ relative to the MMLE, $\hat{\tau}_{\text{MMLE}}$ (denoted by $\text{ARE}_{i,\text{MMLE}}$), as the limit of the ratio of observations $T_{\text{MMLE}}/T_i$ needed for $\text{Pr}[\hat{\tau}_i \in G(\tau); T_i] = \text{Pr}[\hat{\tau}_{\text{MMLE}} \in G(\tau); T_{\text{MMLE}}]$, where $T_i$ and $T_{\text{MMLE}}$ denote the number of observations used to compute $\hat{\tau}_i$ and $\hat{\tau}_{\text{MMLE}}$. The AREs reported here were for the sets $G(\tau) = \{x: .5\tau \leq x \leq 1.5\tau\}$, so $\text{Pr}[\hat{\tau}_i \in G(\tau); T_i] = \text{Pr}[|T_i \hat{\tau}_i - T_i \tau| \leq .5T_i \tau] \to p_\lambda(T_i \tau)$, say, and similarly for $\hat{\tau}_{\text{MMLE}}$. Using (5), set $\lambda = T_{\text{MMLE}}/\tau$; then $\text{ARE}_{i,\text{MMLE}} = \lim T_{\text{MMLE}}/T_i$ can be computed by solving $p_\lambda(\lambda | \text{ARE}_{i,\text{MMLE}}) = p_\lambda(T_i \tau)$. In general, the ARE depends on $\lambda$.

Table 2 reports these AREs for the MPLE and six median unbiased estimators for various values of $\lambda$; 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 $\lambda$ shown and is considerably more efficient for small to moderate values of $\lambda$. In contrast, the median unbiased estimators perform slightly better than MMLE for small values of $\lambda$ ($\lambda \leq 4$) and comparably for moderate

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

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>MPLE</th>
<th>L</th>
<th>MW</th>
<th>EW</th>
<th>QLR</th>
<th>POI(7)</th>
<th>POI(17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.13</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.07</td>
<td>1.07</td>
</tr>
<tr>
<td>2</td>
<td>.19</td>
<td>1.09</td>
<td>1.07</td>
<td>1.07</td>
<td>1.07</td>
<td>.96</td>
<td>1.07</td>
</tr>
<tr>
<td>3</td>
<td>.52</td>
<td>1.08</td>
<td>1.10</td>
<td>1.08</td>
<td>1.04</td>
<td>1.02</td>
<td>.94</td>
</tr>
<tr>
<td>4</td>
<td>.62</td>
<td>1.06</td>
<td>1.06</td>
<td>1.06</td>
<td>1.02</td>
<td>1.06</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>.65</td>
<td>.93</td>
<td>.93</td>
<td>.98</td>
<td>.97</td>
<td>1.11</td>
<td>1.14</td>
</tr>
<tr>
<td>6</td>
<td>.71</td>
<td>.94</td>
<td>.92</td>
<td>1.00</td>
<td>.99</td>
<td>1.03</td>
<td>1.08</td>
</tr>
<tr>
<td>7</td>
<td>.76</td>
<td>.79</td>
<td>.79</td>
<td>.85</td>
<td>.88</td>
<td>.96</td>
<td>1.04</td>
</tr>
<tr>
<td>8</td>
<td>.78</td>
<td>.77</td>
<td>.77</td>
<td>.85</td>
<td>.85</td>
<td>.88</td>
<td>.91</td>
</tr>
<tr>
<td>9</td>
<td>.75</td>
<td>.69</td>
<td>.69</td>
<td>.75</td>
<td>.77</td>
<td>.86</td>
<td>.89</td>
</tr>
<tr>
<td>10</td>
<td>.80</td>
<td>.65</td>
<td>.65</td>
<td>.71</td>
<td>.74</td>
<td>.80</td>
<td>.80</td>
</tr>
<tr>
<td>12</td>
<td>.67</td>
<td>.56</td>
<td>.56</td>
<td>.65</td>
<td>.67</td>
<td>.67</td>
<td>.67</td>
</tr>
<tr>
<td>14</td>
<td>.57</td>
<td>.50</td>
<td>.49</td>
<td>.57</td>
<td>.57</td>
<td>.57</td>
<td>.57</td>
</tr>
<tr>
<td>16</td>
<td>.50</td>
<td>.42</td>
<td>.42</td>
<td>.49</td>
<td>.50</td>
<td>.50</td>
<td>.50</td>
</tr>
<tr>
<td>18</td>
<td>.44</td>
<td>.38</td>
<td>.38</td>
<td>.44</td>
<td>.44</td>
<td>.44</td>
<td>.44</td>
</tr>
<tr>
<td>20</td>
<td>.40</td>
<td>.33</td>
<td>.33</td>
<td>.40</td>
<td>.40</td>
<td>.40</td>
<td>.40</td>
</tr>
<tr>
<td>25</td>
<td>.32</td>
<td>.28</td>
<td>.28</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
</tr>
<tr>
<td>30</td>
<td>.27</td>
<td>.22</td>
<td>.23</td>
<td>.27</td>
<td>.27</td>
<td>.27</td>
<td>.27</td>
</tr>
</tbody>
</table>

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 $r \pm \tau$ as the candidate estimator, as a function of $\lambda = rT$, 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 $\lambda$ shown in column 1. These probabilities were estimated using 5,000 replications and $T = 500$ for each value of $\lambda$. 

Figure 2. Cumulative Asymptotic Distributions of the Gaussian MLEs and Six Median-Unbiased Estimators of $\lambda$ When $\lambda = 5$. ---, MPLE; --------, MMLE; ---, L; · · ·, MW; · · ·, EW; ---, QLR; · · ·, POI(7); · · ·, POI(17).
values of $\lambda$ ($5 \leq \lambda \leq 8$). However, their performance deteriorates for large values of $\lambda$ ($\lambda > 10$).

One way to calibrate the magnitude of $\lambda$ 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 $\lambda$.

### 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.

The data used are real quarterly values of GDP per capita from 1947:II–1995:IV. The data from 1959:1–1995:IV are the GDP chain-weighted quantity index, quarterly, seasonally adjusted (Citibase series GDPFC). The data from 1947:II–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 1959:1. 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 $GY_t = 400 \Delta \ln(real per capita GDP)$. The model is

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

| Parameter estimates | MPLE | MMLE | Estimates with fixed $\lambda$
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{\lambda} \rho$</td>
<td>0</td>
<td>.04</td>
<td>.13</td>
</tr>
<tr>
<td>$\sigma_{\epsilon}$</td>
<td>3.85 (17)</td>
<td>3.86 (17)</td>
<td>3.85 (17)</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>.33 (06)</td>
<td>.34 (07)</td>
<td>.34 (07)</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>.13 (06)</td>
<td>.13 (07)</td>
<td>.13 (06)</td>
</tr>
<tr>
<td>$p_3$</td>
<td>-.01 (07)</td>
<td>-.01 (08)</td>
<td>-.01 (07)</td>
</tr>
<tr>
<td>$p_4$</td>
<td>-.09 (06)</td>
<td>-.08 (06)</td>
<td>-.09 (06)</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>1.80 (46)</td>
<td>2.44 (84)</td>
<td>2.67 (22)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic ($p$-value)</th>
<th>$\lambda$</th>
<th>$\delta_{\lambda}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>.21 (.29)</td>
<td>4.1</td>
<td>(19.4) .13</td>
</tr>
<tr>
<td>MW</td>
<td>1.16 (.29)</td>
<td>3.4</td>
<td>(18.8) .11</td>
</tr>
<tr>
<td>EW</td>
<td>.68 (.32)</td>
<td>3.0</td>
<td>(17.0) .10</td>
</tr>
<tr>
<td>QLR</td>
<td>3.31 (.48)</td>
<td>8.0</td>
<td>(13.3) .03</td>
</tr>
<tr>
<td>POI(7)</td>
<td>2.90 (.45)</td>
<td>1.7</td>
<td>(12.9) .05</td>
</tr>
<tr>
<td>POI(17)</td>
<td>7.52 (.54)</td>
<td>0.0</td>
<td>(11.3) .00</td>
</tr>
</tbody>
</table>

$\delta_{\lambda}$ is the estimate of the standard deviation of $\Delta \beta_{t}$ in (11); that is, $\delta_{\lambda} = \tau^{-1} \lambda$. The two sets of columns report estimates by restricted MLE, with $\lambda$ fixed to the indicated values. The column labeled $GY_t$ in the second part of the table is the sample mean of $GY_t$; the other entries are average values of $\delta_{\lambda}$ over the indicated subsample for the indicated model, where $\beta_{t}$ are the estimates of $\beta_{t}$ 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; ----, MPL; ---, 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 \( \lambda \) are constructed using the normalization \( D = 1 \) (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 \( \Delta GY_t \) 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 \( \lambda \) 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 \( \lambda \)—and indeed the median-unbiased estimates are, with only one exception, nonzero. The median unbiased estimates of \( \lambda \) 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 \( \Delta \beta_t \), 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 \( \beta_{1995:IV} - \beta_{1947:II} \) of 1.4 percentage points.

Estimates of the model parameters are presented in the top part of Table 5, for various values of \( \lambda \): the MPL, 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 \( \lambda \) based on \( L \) (the largest such value for the 90% confidence intervals). Consistent with large pile-up probability discussed in Section 3, \( \lambda_{\text{MLE}} = 0 \). The MMLE produces a small but nonzero estimate of \( \sigma_{\Delta \beta} \) equal to .04%, which corresponds to a point estimate of \( \lambda \) 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 \( \beta_{1T} \) 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. ---, MPL; ----, 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 (GY2); 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 $\sigma_{2t}\beta$ 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 9.9% 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 $X_t$. 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 $D = I_k$ 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 $X_t = 1$, 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 $\bar{U}_{t-1} = (u_{t-1}, \ldots, u_{t-p})'$, $A = (-a_1, -a_2, \ldots, -a_p)'$, and $\bar{A} = (\bar{U}_t' U_{t-1})^{-1} (\bar{U}_t' \bar{u})$ using the usual matrix notation.

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

Proof. The result follows by showing $T^{1/2}(\bar{A} - A) \xrightarrow{p} 0$, where $\bar{A} = (\bar{U}_t' U_{t-1})^{-1} (\bar{U}_t' \bar{u})$, where $U_{t-1} = (u_{t-1}, \ldots, u_{t-p})'$. After straightforward algebra, it is seen that this follows if $\mu_{1T} \xrightarrow{p} 0$ and $\mu_{2T} \xrightarrow{p} 0$, where $\mu_{1T}$ and $\mu_{2T}$ are matrices with $(i, j)$ elements, $\mu_{1T}(i, j) = T^{-1/2} \sum_{t=1}^{T} u_{t-1} X_{t-1}'(\beta_{i-1} - \bar{\beta})$ and $\mu_{2T}(i, j) = T^{-1/2} \sum_{t=1}^{T} u_{t-1} X_{t-1}'(\beta_{i-1} - \bar{\beta})$. These limits follow using the Markov and Chebyshev inequalities and applying assumptions A–C, assuming that $T^{1/2}(\bar{\beta} - \beta_0) = O_p(1)$. An $O_p(1)$ limiting representation for $T^{1/2}(\bar{\beta} - \beta_0)$ 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 Chebyshev's inequality.

Lemma A2. Let $z_t$ be a mean 0 stationary vector stochastic process with fourth-order cumulants that satisfy $\sum_{r_1, r_2, r_3, r_4 = -\infty}^{\infty} |C(r_1, r_2, r_3)] < \infty$. Let $w_t$ be either a scalar nonrandom sequence or a random variable that is independent of $z_t$ for which $\sup_s \sup_{t \geq 1} E|z_{st}|^4 < \infty$ and $\sup_{t \geq 1} E|w_t|^4 < \infty$. Then $T^{-1} \sum_{t=1}^{T} z_t w_t \xrightarrow{p} 0$ uniformly in $s$.

Proof. First let $z_t$ be a scalar. For $\delta > 0$,\begin{equation}
\Pr \left[ \sup_s T^{-1} \sum_{t=1}^{T} z_t w_t > \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) + 3 \sum_{t_1, t_2 = 1}^{T} |C(t_1 - t_2)|^2 \right\} \leq \delta^{-4} T^{-3} \sup_{t} E|w_t|^4 \left\{ \sum_{r_1, r_2, r_3 = -\infty}^{\infty} |C(r_1, r_2, r_3)| + 3T^2 \sum_{r_3 = -\infty}^{\infty} |C(r_1)|] \right\}.
\end{equation}

Under the stated assumptions, the final expression $\rightarrow 0$, proving uniform consistency. This extends to vector $z_t$ by replacing $z_t$ by $z_{st}$.

Proof of Theorem 1

Let $a(L) = \sum_{i=0}^{p} a_i L^i$, with $a_0 = 1$, and let $X_t^{\dagger} = a(L)X_t$. An implication of assumptions A, B, and C, or alternatively of assumptions A and D, is that \begin{equation}
(\sigma_{\epsilon, \eta}^{-1} - T^{-1/3} \sum_{t=1}^{T} X_t^{\dagger} \epsilon T^{-1/2} T^{-1/3} \sum_{t=1}^{T} \nu_t) \Rightarrow (W_1, W_2),
\end{equation}
where $W_1$ and $W_2$ are independent $k$-dimensional standard Brownian motions.

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

Proof of Part a

Let $\tilde{u}_t = \tilde{a}(L)u_t$ and $\bar{w}_t = -\sum_{j=1}^{p} \tilde{a}_j X_{t-j} \sum_{i=t-1}^{j-1} v_{t-i}$, so that
\[ \hat{y}_t = \beta_0 \tilde{X}_t + \left( \sum_{r=1}^t \nu_r \right) \tilde{X}_t + \tilde{\omega}_t + \tilde{\omicron}_t. \]

Accordingly,
\[ \xi_T(s) = \xi_{1T}(s) + \lambda s \xi_T(s) + \xi_{1T}(s) - \kappa_T(s) \{ \xi_{1T}(1) + \lambda \xi_{1T}(1) + \xi_{1T}(1) \}. \] (A.2)

where
\[ \xi_{1T}(s) = T^{-1/2} \sum_{r=1}^{[Ts]} \tilde{X}_r \tilde{u}_r, \]
\[ \xi_{1T}(s) = T^{-3/2} \sum_{r=1}^{[Ts]} \tilde{X}_r \tilde{X}_r' \sum_{r=1}^{[Ts]} \nu_r, \]
and
\[ \kappa_T(s) = \left[ T^{-1} \sum_{t=1}^{[Ts]} \tilde{X}_r \tilde{X}_r' \right] \left[ T^{-1} \sum_{t=1}^{T} \tilde{X}_r \tilde{X}_r' \right]^{-1}. \]

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

1. Write \( \xi_{1T}(s) = \Delta_{1T}(s) + \Delta_{2T}(s) + \xi_{1T}^0(s) \), where
   \( \Delta_{1T}(s) = \sum_{j=0}^{p} \delta_j s \sum_{r=1}^{[Ts]} T^{1/2} ( \tilde{u}_r - \tilde{\alpha}_r ) ( T^{-1} \sum_{t=1}^{[Ts]} X_{t-j} \tilde{u}_{t-i} ). \)
   \( \Delta_{2T}(s) = T^{-1/2} \sum_{t=1}^{[Ts]} \tilde{X}_r \tilde{X}_r' \sum_{r=1}^{[Ts]} \nu_r, \) and \( \xi_{1T}(s) = T^{-3/2} \sum_{r=1}^{[Ts]} \tilde{X}_r \tilde{X}_r' \sum_{r=1}^{[Ts]} \nu_r. \)

To show \( \Delta_{1T}(s) \stackrel{p}{\rightarrow} 0 \) and \( \Delta_{2T}(s) \stackrel{p}{\rightarrow} 0 \), consider for notational simplicity the case \( k = 1 \). (The argument for \( k > 1 \) is similar.)

2. Write \( \xi_{1T}(s) = \Delta_{3T}(s) + \Delta_{4T}(s) + \xi_{1T}^0(s) \), where
   \( \Delta_{3T}(s) = T^{-3/2} \sum_{t=1}^{[Ts]} X_{t-j} \tilde{X}_r \tilde{X}_r' - \Gamma \sum_{r=1}^{[Ts]} \nu_r, \) and \( \Delta_{4T}(s) = T^{-3/2} \sum_{r=1}^{[Ts]} \tilde{X}_r \tilde{X}_r' \sum_{r=1}^{[Ts]} \nu_r. \)

3. Write \( \xi_{1T}(s) = -\lambda \sum_{j=0}^{p} \sum_{k=0}^{[Ts]} \sum_{l=0}^{[Ts]} \xi_{1T}(s) \xi_{1T}(s), \) where
   \( \xi_{1T}(s) = \left( T^{-1/2} \sum_{t=1}^{[Ts]} X_{t-j} \tilde{X}_r \tilde{X}_r' \right) \left( T^{-1/2} \sum_{t=1}^{[Ts]} X_{t-j} \tilde{X}_r \tilde{X}_r' \right) \left( T^{-1/2} \sum_{t=1}^{[Ts]} X_{t-j} \tilde{X}_r \tilde{X}_r' \right) \).

4. Let \( \Delta_{5T}(s) = T^{-1} \sum_{t=1}^{[Ts]} ( \tilde{X}_r \tilde{X}_r' - \tilde{X}_r \tilde{X}_r' ) \) and \( \Delta_{6T}(s) = T^{-1} \sum_{t=1}^{[Ts]} ( \tilde{X}_r \tilde{X}_r^f - \Gamma ) \), and let \( \Delta_{5T} = \Delta_{5T} + \Delta_{6T} = T^{-1} \sum_{t=1}^{[Ts]} ( \tilde{X}_r \tilde{X}_r' - \Gamma ) \), the argument that \( \Delta_{1T} \stackrel{p}{\rightarrow} 0 \) follows the argument that \( \Delta_{4T} \stackrel{p}{\rightarrow} 0 \) with \( T^{-1/2} \sum_{r=1}^{[Ts]} \nu_r \) replaced by 1, and the argument that \( \Delta_{6T} \stackrel{p}{\rightarrow} 0 \) follows the argument that \( \Delta_{5T} \stackrel{p}{\rightarrow} 0 \) with the same replacement.

Similiar calculations imply that \( \Delta_{5T} \stackrel{p}{\rightarrow} 0 \) and \( \Delta_{6T} \stackrel{p}{\rightarrow} 0 \) by collecting terms and using (A.2), it follows that \( \hat{v}_{T-1/2} \xi_{1T}(s) \Rightarrow \hat{v}_{\Delta T}(s) - s \hat{h}_{\Delta T}(s) \), where \( h_{\Delta T}(s) = \hat{W}_1(s) + \lambda T^{1/2} \Omega_{1/2}^f \hat{W}_2(r) dr \).

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, \( \delta(L) = \delta(L) = 1 \), so \( \tilde{X}_r = X_{t-j} \tilde{X}_r' = \tilde{X}_r' \).

REFERENCES


