NATURAL RESOURCE MODELING
Volume 17, Number 2, SUMMER 2004

TEMPORAL AGGREGATION AND TESTING FOR TIMBER PRICE BEHAVIOR

JEFFREY P. PRESTEMON
Forestry Sciences Laboratory
USDA Forest Service
P.O. Box 12254
Research Triangle Park, NC 27709
E-mail address: jprestemon@fs.fed.us

JOHN M. PYE
Forestry Sciences Laboratory
USDA Forest Service
P.O. Box 12254
Research Triangle Park, NC 27709

THOMAS P. HOLMES
Forestry Sciences Laboratory
USDA Forest Service
P.O. Box 12254
Research Triangle Park, NC 27709

ABSTRACT. Different harvest timing models make different assumptions about timber price behavior. Those seeking to optimize harvest timing are thus first faced with a decision regarding which assumption of price behavior is appropriate for their market, particularly regarding the presence of a unit root in the timber price time series. Unfortunately for landowners and investors, the literature provides conflicting guidance on this subject. One source for the ambiguous results of unit root tests of timber prices may involve data problems. We used Monte Carlo simulations to show that aggregating observations below their observed rate resulted in similar power reductions and empirical size distortions across three classes of unit root tests. Moving-average error structures can also affect power and sizes of tests on period-averaged data. Such error structures can also be created by the kind of temporal averaging common in reported timber prices. If we take timber prices at their face value and therefore ignore these sampling error and temporal aggregation complications, we find that unit root tests on southern timber prices support a unit root in 158 out of 208 product-deflation combinations tested, random walks in 38 of the series found to be nonstationary, and stationarity in none. However, if we recognize temporal aggregation errors, unit root tests more commonly favor stationarity, especially for pulpwood stumpage. Because price trends for sawtimber and pulpwood products may behave differently even in the same region, stochastic harvest

Copyright ©2004 Rocky Mountain Mathematics Consortium

timing models must be developed that allow their multiple products to follow different price paths.

KEY WORDS: Nonstationarity, stationarity, unit root, stumpage, temporal aggregation.

Introduction. Optimal timber harvest timing has received much attention in the literature in the past quarter century, e.g., Norstrøm [1975], Lohmander [1988], Brazee and Mendelsohn [1988], Clarke and Reed [1989], Haight and Holmes [1991], Haight and Smith [1991], Thomson [1992], Yin and Newman [1995a, 1995b, 1996, 1997], Forboseh et al. [1996], Abildtrup et al. [1997], Plantinga [1998], Gong [1998, 1999], Brazee and Bulte [2000], Reeves and Haight [2000], Saphores et al. [2002], Yoshimoto and Shoji [2002], and rightly so. Maximization of returns to timber management by carefully timing timber harvests and intermediate treatments depends in part on how prices are expected to evolve over time. Such expectations, however, depend on timber owner price perceptions (Burton and Love [1996], Gomez et al. [1999]) as well as actual price behavior. In choosing the "best" harvest timing model available, actual price behavior, especially the presence of a unit root in the timber price time series, is pivotal. If owners err in their perceptions, e.g., thinking that prices are stationary when, in fact, they are nonstationary, significant negative economic consequences may follow, including sub-optimal returns to timber production and underinvestment in timber as a land use (Brazee and Mendelsohn [1988], Haight and Holmes [1991], Yoshimoto and Shoji [2002]). The results of such research may not be useful for all timber producers. For example, industrial owners whose production is linked to a particular mill may not find these decisions critical. For many large landowners, however, more realistic perception of price behaviors could yield economic benefits both to themselves and to society in general.

Given its central importance for timber management decision-making, it is surprising that published studies on the time series behavior of timber prices, e.g., Washburn and Binkley [1990], Hultkrantz [1993], show such a confusing array of approaches. For example, Washburn and Binkley [1990] use turning point tests and examination of residuals of regressions (effective deflation by stock returns and by a commodity price index), while Hultkrantz [1993] uses Dickey-Fuller tests. Haight and Holmes [1991] use the augmented Dickey-Fuller test on one consumer price index-deflated timber price series, while Prestemon [2003] uses a lag order selection technique with the ADF as well as long run return regressions on similarly-deflated time series and applies them Southwide. Prestemon and Holmes [2000] use a lag order selection technique on nominal series. Most of these studies, and others not cited, also have different lengths of time series and evaluate price behavior with series reported at different frequencies (months or quarters). It may therefore be worthwhile to evaluate the individual effects of some of these assumptions, data structures, and tests on conclusions about price behavior.

These conflicts in testing across studies may have arisen from at least three data complications, in addition to assumption and test differences. These complications arise in the chain between the data gathering process used to develop a price series and the empirical result of a statistical test on the series. The complications affect the empirical size and the statistical power of unit root tests. One kind of complication arises from incomplete observation of the entire population, which most analysts cannot directly control. The resulting sampling error introduces a positive moving average term into the observed process (Brewer [1973], Harvey [1981, p. 43]), lowering the power and distorting the size of unit root-null unit root tests, e.g., Schwert [1987, 1989]. Two other complications which we will address in this research arise from the existing nuisance parameters in the true series (Said and Dickey [1984], Schwert [1987, 1989]) and from the introduction of moving parameters created by temporal aggregation (Working [1960], Brewer [1973], Harvey [1981], Haight and Holmes [1991], Maeso-Fernandez [1998], Taylor [2001]), both of which affect the empirical sizes of unit root tests, e.g., Phillips and Perron [1988], Schwert [1989], Hall [1994], Perron [1996]. Small degrees of temporal aggregation transform true ARIMA(p, d, q) series into IMA(d, d) observed series (Tiao [1972], Brewer [1973]), while large degrees can make them IMA$(d, 0)$ observed series (Rossana and Seater [1995]). All of these effects from data complications are enhanced, as well, by small samples and short data time spans, and some tests are more prone to these effects than others, see, for example, Shiller and Perron [1985], Schwert [1987, 1989], Lo and MacKinlay [1988, 1989], and Leybourne and McCabe [1994].

Finally, a common practice in evaluation of commodity prices is expressing nominal prices relative to other prices—for example deflating to account for general price level changes or the prices of alternative products in an investment portfolio. Harvey [1981] explains that this may create a time series process more complex than either of the combined original series, while Schnute [1987] further describes the spurious patterns that can arise from this.

The choice of statistical test can also affect the outcome of the hypothesis being tested. Each test has its power limitations and empirical size distortions when faced with various time series complexities. Below, we elaborate on and show with empirical examples how temporal aggregation, some kinds of nuisance parameters, deflation assumptions, and the statistical test itself can affect conclusions about price processes. By explicitly recognizing the effects of each of these issues, we can use available data and tests to draw more accurate conclusions about the applicability of the two classes of harvest timing rules.

Methods. Models of harvest timing under stochastic prices belong to either of two main strands, one assuming that prices are stationary (mean-reverting) and the other, nonstationary, Table 1. Economic doctrine that backward-looking asset price perceptions are not rational or imply market imperfections (Samuelson [1965], Fama [1970, 1991], LeRoy [1989]) suggests that timber prices should be a martingale (Abildtrup et al. [1997]), but this assessment only applies under strict circumstances. For example, more complex price behavior may arise if obtaining or responding to prices incurs costs.

Timber storage capabilities and nonzero storage costs (Williams and Wright [1991]) and (Deaton and Laroque [1996]) may also explain why prices may not behave according to martingale patterns even while markets are efficient. However, the underlying operation of a storage model implies that time series tests relying on Gaussian distributions of price innovations are invalid. The validity of the storage model for observed stumpage prices, on the other hand, depends on the possibility that the supply of the stored commodity sometimes comes close to being exhausted. Although McGough et al. [2002] have simulated the impacts of storage on an observed price series, we are not aware of research that empirically evaluates the exhaustion question for U.S. timber markets.

For the purposes of this study we are therefore assuming that price innovations are consistent with Gaussian behavior at the market level and that storage processes are unimportant. The implications of this assumption are themselves worthy of additional research.

We must acknowledge the simultaneity problem with evaluating timber price behavior. By acting on their assumptions, individual owners influence the market whose behavior they seek to understand. The smaller the geographic scope of evaluation, the greater the impact one's harvest has on the price received by that owner. This kind of fine-scaled harvest timing model is not the subject of our analysis. We focus instead on broader markets where an individual is a small part of a market and has little impact on market prices. This direction is consistent with the harvest timing literature whose conflicting results we seek to clarify.

Each strand of harvest timing models carries different sets of recommendations and implications for land use and land value and implies a different response when faced with a price change. In their starkest contrast, stationary-price harvest timing models require landowners to harvest when a price exceeds a time-dependent reservation price, while nonstationary-price models require landowners to harvest when the price falls below a time-dependent reservation price. See Haight and Holmes [1991] for an explanation of the reservation price differences and our Table 1 for a listing of some of the differences in assumed price processes in different harvest timing models.

Haight and Holmes [1991], reflecting a critique by Working [1960], recognized that temporal aggregation of data could be an important statistical barrier to understanding the "true" timber price generation process so critical to optimal harvest timing. The barrier derived from period-averaging. In line with a generalization by Taylor [2001] and works of Tiao [1972] and Brewer [1973], they found that temporal aggregation of a truly stationary underlying price series raises the probability that a stationary price would appear nonstationary. They also showed that spot-sampling from a series rather than period-averaging would not raise this probability. Harvey [1981] and Weiss [1984] explain that spot-sampling does not create the additional moving average term that changes this probability. Temporal averaging therefore would be expected to raise Type I error rates above the rate experienced with spot prices in stationarity-null unit root tests, e.g., Kwiatkowski et al. [1992], Leybourne and McCabe [1994]. However, we can find no evidence that this specific effect of temporal averaging has been quantified in the literature.

Our analysis of price behavior evaluates the outcomes of unit root tests of timber prices using simulation and empirical estimation. In both contexts, we examine three classes of unit root tests that, taken together, may enhance our confidence regarding the true nature of a time series. We use simulation to quantify the effects of period-averaging and spot-sampling observations from a longer underlying series and to generate adjusted critical values for the available tests that conform to a desired Type I error rate.

Augmented Dickey-Fuller test. The first class of unit root tests we examine is the augmented Dickey-Fuller test: the t-test (ADFT) and the rho-test (ADFR) (Dickey and Fuller [1979, 1981], Said and Dickey [1984]). The ADFT and the ADFR both begin by considering a series, $\{y_t\}$, in which the following relationship holds (Hall [1994]):

$$y_t = \alpha y_{t-1} + z_t$$
$$z_t = \sum_{j}^{J} \phi z_{t-j} + e_t \tag{1}$$

and where $-1 \leq \alpha \leq 1$ and $\{e_t\}$ are independently and normally-distributed about a mean zero. The augmented form of the ADF test is based on whether the coefficient b differs from zero in the following ordinary least-squares (OLS) regression:

$$dy_t = a + by_{t-1} + \sum_{k=1}^{K} c_k dy_{t-k} + \tau t + \varpi_t \tag{2}$$

where d is the first difference operator and where a, the c_k's and τ are additional (nuisance) parameters to be estimated. The ADFT test statistic is calculated as $ADFT = \hat{b}/\hat{\sigma}_b$, where $\hat{b}$ is the estimate of b in (2) and $\hat{\sigma}_b$ is its standard error. The ADFR test statistic is consistently calculated as $ADFR = T \cdot \hat{b}/(1 - \sum_{k=1}^{K} \hat{c}_k)$, where T is the number of observations in the regression and the $\hat{c}_k$'s are estimates of c_k's shown in (2) (Hamilton [1994, pp. 522–524]).

The size of K in (2) and the deterministic time trend, t, should be based on either *a priori* expectations or some model selection approach. The "general-to-specific" model selection strategy recommended by

Hall [1994], a consistent lag-order selection approach to the ADF, determines K in our research. The procedure begins at the longest probable lag and sequentially drops the last lag until the minimum of the Schwarz Information Criterion (SIC) (Schwarz [1978]) is found. We note that Agiakloglou and Newbold [1992] found that such a strategy does badly when the true data-generating process is an ARIMA(0,1,1), typically choosing K to be too small; this leads to values of the ADFT that are too large and high Type I error rates compared to alternative strategies. Nonetheless, if the initial number of lagged terms is sufficiently long and the selected order of the ADF is zero, then the ADFT and the ADFR are effectively random walk tests.

Variance-ratio test. From the second class, variance ratio tests, we use the variance ratio test for the random walk designed by Lo and MacKinlay [1988]. Variance ratio tests evaluate whether the variance of longer-term changes in the series rises in step with time, which would occur with a nonstationary series. The homoscedastic variance version of Lo and MacKinlay's [1988, 1989] variance ratio test to identify a random walk is as follows. Given $nq + 1$ observations of $\{y_t\}$, $(y_0, y_1, \dots, y_{nq})$ where both n and q are integers (> 1), parameters μ and σ_a^2 are estimated as u and s_a^2, an adjusted, i.e., unbiased, specification test statistic is based on

$$u = (y_{nq} - y_0)/(nq),$$
$$s_a^2 = (nq-1)^{-1} \sum_{t=1}^{nq} (y_t - y_{t-1} - u)^2. \tag{3}$$

Next, the variance of the qth differences of y_t, σ_b^2, under Lo and MacKinlay's [1988] null hypothesis, is q times the variance of first-differences. Hence, an estimate of σ_b^2, divided by q, should also converge to σ_a^2 under the null. Let $s_b^2(q)$ be the adjusted (unbiased) estimator of the variance of the qth-differences of y_t, $\sigma_b^2(q)$:

$$s_b^2(q) = \frac{1}{m} \sum_{t=q}^{nq} (y_t - y_{t-q} - qu)^2$$
$$m = q(nq - q + 1)\left(1 - \frac{q}{nq}\right). \tag{4}$$

Now calculate the dimensionless centered variance ratio, $M_r(q) = (s_b^2(q)/s_a^2) - 1$, which converges in probability to zero (Lo and MacKinlay [1988]). In their simulation article, Lo and MacKinlay [1989] show how the test is more powerful than the ADFT test, as long as q remains less than half of T. The normalized $N(0,1)$ version of $M_r(q)$ is $z_1(q)$:

$$z_1(q) = \sqrt{nq}M_r(q) \cdot \left(\frac{2(2q-1)(q-1)}{3q} \right)^{-1/2}. \tag{5}$$

Asymptotically, $z_1(q)$ is distributed normally. Note, the Lo-MacKinlay test is powerful against an AR(p) alternative and against ARIMA($p, 1, q$) alternatives. The Lo-MacKinlay is designed to be most powerful at detecting the random walk; comparing it with the ADFT and ADFR test outcomes becomes a check on the results of each.

Leybourne and McCabe test. Unlike the ADF and the Lo-MacKinlay test, a third class takes stationarity as the null time series process. In our analysis we implement a test by Leybourne and McCabe [1994], shown by those authors to have better size and power properties in most circumstances than an earlier test developed by Kwiatkowski et al. [1992]. Leybourne and McCabe's [1994] score-based stationarity test begins with the specification of a unit root process, wherein

$$\begin{aligned} \Phi(L)y_t &= \alpha_t + \beta_t + \varepsilon_t \\ \alpha_t &= \alpha_{t-1} + \eta_t \end{aligned} \tag{6}$$

where $\{\varepsilon_t\}$ is independently and identically distributed $(0, \sigma_\varepsilon^2)$ and $\{\eta_t\}$ is distributed independently and identically distributed $(0, \sigma_\eta^2)$. Also, L is the lag operator and $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ is a pth-order autoregressive polynomial with roots outside the unit circle. This model (6) is second-order equivalent in moments to the ARIMA($p, 1, 1$) process, $\Phi(L)(1-L)y_t = \beta + (1 - \theta L)\varsigma_t$, where $0 < \theta < 1$ and ς is independently and identically distributed $(0, \sigma_\varsigma^2)$. Leybourne and McCabe [1994] show that the existence of a nonzero σ_η^2 is evidence of a nonstationary ARIMA($p, 1, 1$) process against a null of an AR($p, 0, 0$) process. The test statistic developed by these authors begins with the maximum-likelihood estimate of the parameters $\phi = (\phi_1, \phi_2, \ldots, \phi_p)$, made by fitting the ARIMA model,

$$dy_t = \beta + \sum_{i=1}^{p} \phi_i dy_{t-i} + \varsigma_t + \theta \varsigma_{t-1} \tag{7}$$

and then constructs the series $y_t^* = y_t - \sum_{i=1}^{p} \phi_i^* y_{t-i}$, where the ϕ_i^* are estimates of ϕ_i from (7).

Two possible least-squares regressions can be estimated: regressing y_t^* on an intercept (no-trend case), or regressing y_t^* on an intercept and a trend (deterministic time trend case). The test is conducted on the residuals, $\{\varepsilon^*\}$, from either of these two possible regressions, by $s^* = \sigma_\varepsilon^{*-2} T^{-2} \varepsilon^{*\prime} V \varepsilon^*$, where $\sigma_\varepsilon^{*2} = \varepsilon^{*\prime}\varepsilon^*/T$ and V is the covariance matrix of a nonstationary series, where the elements v_{ij} of V are the $\min(i, j)$. Critical values for the statistic s^* are tabulated for both the no-trend and the deterministic time trend case by Kwiatkowski et al. [1992] and are applicable to the test of Leybourne and McCabe [1994]. Leybourne and McCabe showed in simulations how the number of lags of differenced prices in (7) did not appreciably affect the outcome of the test.

Because the three classes of unit root tests that we employ here have different strengths and weaknesses, combining them in a heuristic fashion can bolster confidence in our conclusions about the actual time series behavior of timber prices. Specifically, Dickey-Fuller type tests have been criticized as being more weakened by nuisance parameters than alternative tests, especially by the additional lags needed to test for the unit root. Lo and MacKinlay's [1988, 1989] variance ratio test avoids many of the specification issues inherent in ADF tests (Cecchetti and Lam [1994]). Finally, both the ADF and variance ratio test are weak in identifying stationary series with large but non-unitary roots, especially when moving average terms are present. Leybourne and McCabe's [1994] method is structured to be good at detecting stationary series with large moving average components. A weakness of the Leybourne-McCabe test is the severe size distortions in the test created by these very moving average components (Caner and Killian [2001]).

Data and empirical approach. Unit root and random walk behavior for timber prices were evaluated using the statistical tests described above for twenty-one U.S. southern timber sub-markets. These sub-markets are geographically delineated as two regions per state, from Virginia to Texas. Quarterly prices of southern pine and hardwood stumpage for these sub-markets were obtained from monthly and quarterly price reports of Timber Mart-South (TMS)

(Norris Foundation [2002]), from 1977 to 2002 (second quarter). TMS prices were reported monthly until 1987 and quarterly thereafter, so the monthly prices were converted to quarterly prices by period-averaging the monthly series. The quarterly series reported since 1987 are more akin to quarterly period-averaged prices (Harris et al. [1999]).

TMS sub-market boundaries changed once during the sample period. In order to make the pre-1992 series spatially consistent with prices reported 1992-onward, we applied the approach outlined by Prestemon and Pye [2000]. The Prestemon and Pye [2000] approach used average timber harvest activity in counties to derive weights for the pre-1992 prices corresponding to the old Timber Mart-South regions. A weighted sum of pre-1992 prices under the pre-1992 region definitions were transformed into prices corresponding to the new region boundaries.

Timber Mart-South sub-market boundaries ("areas" in the TMS reports) were defined by Timber Mart-South, based on their own studies of market areas. Region boundaries were chosen such that the forestland and markets contained within are similar in terrain, timber demand (mill types), harvest activity, and species mixes, among other factors (see Gunter and Cubbage [1987]).

Normality tests showed that logarithmic price series had distributions closer to normal than raw prices. Hence, prices—deflated and undeflated—were transformed by the natural logarithm. Such transformation could affect test results and is an area worthy of additional research. Nevertheless, if the dependent variable in a least squares regression is not normally distributed, statistical techniques such as the least-squares regression of the ADF would not be consistently estimated.

The United States Department of Commerce [2002a,b] was the source of PPI and CPI data, while stock market returns were taken as the average value of investments including reinvested dividends of the Standard and Poors 500 index (Anonymous [2002]). Quarterly PPI, CPI, and stock indices were generated in the same manner as were timber prices from monthly reporting periods-period-averaging across the three months of each quarter.

To better understand the effects of the kinds of period-averaging present in our data on the outcomes of the ADFT, ADFR, and Leybourne-McCabe tests, we conducted Monte Carlo simulations for a sample size similar to those available for empirical testing of timber price data, 100 observations. This sample size, coincidentally, is also a standard sample size evaluated in the unit root test simulation literature, so using this size makes comparisons with simulations by other analysts more straightforward.

The simulations of the ADFT and ADFR tests quantified their empirical sizes (rates of null hypothesis rejection) using the general-to-specific model selection strategy of Hall [1994] when applied to series that were period-averaged across the three periods, e.g., creating a "quarterly" observation out of three "monthly" observations, and series that were spot-sampled once every three periods, e.g., creating a "quarterly spot price" observation by sampling every three periods from the "monthly" observations. This monthly-to-quarterly temporal aggregation effect is what Haight and Holmes [1991] examined in their analysis, a logical departure point in our study. Nevertheless, our results will imply the directional impact of three-to-one aggregation of point observations, which might not accurately reflect the true rate of data generation for timber prices. These prices might, in fact, be best characterized as generated within a region by each timber sale, and such sales probably occur daily. We note, however, that the span (the actual time between the beginning and the end of a time series) is a dominant determinant of statistical power in time series tests of the unit root (Shiller and Perron [1985]). In our analysis, we do not directly address the effects of span, although this would be an area worthy of extensive analysis. We also leave to future research the issue of what is the "true" rate of data generation for the regional market price of timber.

In our simulations, we generated 10,000 replicates of an ARIMA series $y_t = \alpha + \rho y_{t-1} + \theta e_{t-1} + e_t$, where e_t was a Gaussian random variable; ρ was set at 1.0, 0.98, 0.95, 0.8, and 0.5; and θ was set at -0.8, 0, and 0.8 (simulations for 0.5 and -0.5 are available from the authors). The number of "monthly" observations in each replicate of the true ARIMA series was set at $1000 + [3 * \text{int}\{8(T/100)^{1/4}\} + 3] + 300 = 1327$. The first 1,000 observations of a simulated "monthly" series were not used to avoid start-up effects, while the $(3 * \text{int}\{8(T/100)^{1/4}\} + 3) + 300$ observations were used for generating an effective quarterly data sample size of 100 upon which to implement the Hall [1994] strategy with conditioning. The formula between the braces of this sum, the integer part of the ratio of $8T/100$, equal to 8 for an effective sample size of $T = 100$ "quarterly" observations, see Schwert [1989], was the number of lagged difference terms used in the ADF regression on the "quarterly" data. Conversion of the "monthly" data required three times that number. Eight lags reflected a compromise between potential over-specification (and hence power reductions) and under-specification (and hence inconsistency). Three observations were added to the braced term to generate an additional "quarterly" observation to accommodate the first lagged term in the ADF regression. The Hall [1994] general-to-specific procedure used the Schwarz Information Criterion (SIC) (Schwarz [1978]), starting with eight lagged difference terms. Simulations were done for both the trend- and the no-trend versions of the ADFT and ADFR tests. In the interest of clarity, we limit most of our discussion results to those of the no-trend versions of the ADFT and ADFR tests. In what follows, tables of results for simulations of the trend versions of the ADFT and ADFR tests are briefly discussed, and they are available from the authors upon request.

Simulations of the outcomes of the Leybourne-McCabe test (trend version) for period-averaged and sampled data were also done, with a table comparable to those generated for the ADFT and ADFR tests. The test was based on four lagged autoregressive terms (simulations with twelve such terms were more distorted and over-sized than tests with four). As in the ADFT and ADFR tests, the simulations dropped the first 1,000 observations of a simulated series to avoid start-up effects, and the subsequent 312 observations were used to generate the quarterly observations needed for the test, including the four lagged difference terms.

For the southern timber prices that we analyzed, unit root tests under the nonstationary null (ADFT, ADFR, Lo-MacKinlay) were done both with and without a deterministic time trend. Tables of results using the trend versions are available from the authors upon request. ADFT and ADFR tests with the Hall [1994] procedure began with a maximum of sixteen lagged difference terms, to account for potential AR or MA gaps (see Hall [1994]) and held the number of usable observations constant as the lags were reduced toward zero. Given our available sample size (102), the conditioning allowed for effective sample sizes of 85 for the majority of series. When we evaluated the outcomes of the ADFT and ADFR tests for a unit root in southern timber prices, we applied the values obtained by applying the algorithms of MacKinnon [1991] for the ADFT and found in Hamilton [1994, p. 761] or Fuller [1976, p. 371] for the ADFR.

Lo-MacKinlay tests examined the variance ratio at time lags of ten quarters. The ten quarter choice was determined to be a good compromise between power and nominal size of the test—very short and very long lags shorten the series and were shown by Lo and MacKinlay [1989] to be most powerful against the random walk null. Although we did not address the issue through simulation, we are not sure whether such a compromise between power and consistency would be effective or possible using period-averaged or sampled data.

Leybourne-McCabe tests incorporated four lags of price difference terms, sufficient to cover autoregressive behavior in a price series. This choice was also a compromise between statistical power and potential inferential mistakes associated with under-specification of the autoregressive price structure. The Leybourne-McCabe test included a time trend, as recommended by Caner and Killian [2001].

Assessments on whether series are random walks were made based on the results of the Lo and MacKinlay [1988, 1989] test, and whether the ADFT and ADFR test regressions selected using the Hall [1994] model selection approach included lagged difference terms. The null hypothesis of the Lo-MacKinlay test is the random walk, so a non-rejection of the Lo-MacKinlay test provides some empirical support of the random walk. The null hypothesis of the simple Dickey-Fuller test is that a series is a random walk. If the Hall [1994] model selection procedure indicates that no lagged difference terms are needed in the augmented Dickey-Fuller t-test, then the resulting statistical test for a unit root amounts to a test of the random walk conjecture. This latter approach on the ADFT was used by Prestemon [2003]. Missing data limited the tests that we could run in some Timber Mart-South sub-markets for the Lo-MacKinlay and Leybourne-McCabe tests. Consensus conclusions about the existence of a unit root and a random walk were confined to the continuous series.

TABLE 1. Some timber harvest timing research based on stochastic timber (stumpage) prices.

Citation	Price Process Assumed
Brazee & Bulte [2000]	Stationary, discrete time
Brazee & Mendelsohn [1988]	Stationary, discrete time
Clarke & Reed [1989]	Nonstationary, continuous time geometric Brownian motion (GBM)
Forboseh et al. [1996]	Stationary, discrete time
Gong [1998]	Stationary, discrete time
Gong [1999]	Stationary or random walk, discrete time
Haight & Holmes [1991]	Stationary, discrete time; nonstationary random walk, discrete time
Haight & Smith [1991]	Stationary, discrete time
Lohmander [1988]	Stationary, discrete time
Norstrøm [1975]	Stationary, discrete time
Plantinga [1998]	Stationary, nonstationary, discrete time
Reeves & Haight [2000]	Stationary, discrete time
Thomson [1992]	Nonstationary, random walk (discrete time)
Yin & Newman [1995a]	Nonstationary, GBM
Yin & Newman [1996]	Nonstationary, GBM with Poisson jumps
Yoshimoto & Shoji [2002]	Nonstationary or stationary continuous-time state-dependent volatility process

Results. We first present the results from the simulations that compare period averaging with spot sampling, and then explore results of empirical tests of softwood sawtimber, and softwood and hardwood pulpwood markets.

Simulations. Tables 2 and 3 summarize simulations of outcomes of ADFT and ADFR tests and support two principal findings. First, tests applied to either spot-sampled or temporally-averaged series have similar empirical sizes. Across all truly stationary models, in most cases period-averaging resulted in modest power reductions compared to spot-sampling, but the pattern of effect varies according to the size and sign of the moving average parameter. With a zero or positive $MA(1)$ parameter, period-averaging reduces power; with a negative and large moving-average parameter, period-averaging increases power. With large or positive moving average parameters, the ADFT correctly rejects a false null more often with sampled data; the opposite can occur when the moving average parameter is negative. Empirical sizes for truly nonstationary series ($\rho = 1$) at 5 percent nominal significance average about 0.01 closer to nominal significance for spot-sampled data; but the period-averaged nonstationary series produce better size properties than the comparable sampled series when the additional moving average parameter is negative and large. These sizes correspond closely to those generated by continuous datasets of lengths of about 250 observations (see Hall [1994]), validating a finding by Shiller and Perron [1985] that size distortions are more sensitive to span than frequency of observation. So even though averaging or spot-sampling data can weaken inference in testing through the introduction of nuisance parameters (Brewer [1973], Harvey [1981], Weiss [1984]), inferences are more accurate than those obtained using data modeled at the true rate of generation but covering a shorter span.

With the ADFR, the empirical size of the test is not much different between spot-sampled and period-averaged series, but both kinds of series result in over-sized ADFR tests. Spot-sampled nonstationary series produce sizes six times the nominal rate of 0.05 when the moving average parameter is -0.8, while period-averaged series produce ADFR tests that are less distorted when the MA(1) parameter is negative. The ability of the ADFR to detect large but non-unitary autoregressive roots differs little between spot-sampled and period-averaged series: for positive and zero MA(1) terms, period-averaged data produce more powerful ADFR tests; for negative MA(1) terms, ADFR tests on spot-sampled data are more powerful.

TABLE 2. Augmented Dickey-Fuller t-test (ADFT) empirical sizes for 10,000 simulated 324-observation pseudo-monthly series converted to pseudo-quarterly spot or quarterly averaged series of effective lengths of 100 observations. Sizes are based on the Hall [1994] SIC-based general-to-specific model selection strategy, eight lagged difference terms, and eight observations withheld for conditioning across models. Empirical sizes are evaluated at 5 percent significance using the nominal levels determined by MacKinnon [1991].

AR(1) Parameter	MA(1) Parameter	Empirical Sizes (Rejection Rates) at 5% Nominal Significance	
		Samples	Period-Averages
1.00	0.00	0.064	0.067
0.98	0.00	0.186	0.174
0.95	0.00	0.563	0.497
0.90	0.00	0.861	0.822
0.80	0.00	0.927	0.920
0.50	0.00	0.955	0.950
1.00	0.80	0.058	0.068
0.98	0.80	0.158	0.191
0.95	0.80	0.478	0.521
0.90	0.80	0.845	0.806
0.80	0.80	0.919	0.919
0.50	0.80	0.955	0.947
1.00	-0.80	0.210	0.124
0.98	-0.80	0.520	0.363
0.95	-0.80	0.816	0.731
0.90	-0.80	0.928	0.899
0.80	-0.80	0.958	0.959
0.50	-0.80	0.958	0.985

TABLE 3. Augmented Dickey-Fuller rho-test (ADFR) empirical sizes for 10,000 simulated 324-pseudo-monthly series converted to pseudo-quarterly spot or quarterly averaged series of effective lengths of 100 observations. Sizes are based on the Hall [1994] SIC-based general-to-specific model selection strategy, eight lagged difference terms and eight observations withheld for conditioning across models. Empirical sizes are evaluated at 5 percent significance using the nominal levels found in Fuller [1976, p. 371].

AR(1) Parameter	MA(1) Parameter	Empirical Sizes (Rejection Rates) at 5% Nominal Significance	
		Samples	Period-Averages
1.00	0.00	0.101	0.108
0.98	0.00	0.316	0.332
0.95	0.00	0.730	0.669
0.90	0.00	0.916	0.905
0.80	0.00	0.905	0.909
0.50	0.00	0.862	0.867
1.00	0.80	0.103	0.116
0.98	0.80	0.295	0.349
0.95	0.80	0.670	0.715
0.90	0.80	0.908	0.897
0.80	0.80	0.907	0.914
0.50	0.80	0.867	0.879
1.00	-0.80	0.303	0.193
0.98	-0.80	0.704	0.542
0.95	-0.80	0.909	0.868
0.90	-0.80	0.903	0.918
0.80	-0.80	0.849	0.846
0.50	-0.80	0.843	0.689

Comparing Tables 2 and 3, although the ADFR is more powerful than the ADFT in rejecting a false unit root null—almost twice as powerful with large positive moving average parameters and highly autoregressive series with non-unitary roots—the ADFT has empirical sizes that are closer to nominal levels. In other words, the ADFR's advantage of statistical power carries with it the greater risk of false rejection of a unit root, and the price of more accurate empirical sizes with the ADFT is its lower statistical power against a false null.

Second, our results show that inclusion of a moving-average parameter in sampled and period-averaged data affects the size of the ADFT and ADFR tests. When the moving-average parameter is negative and large, the differences in statistical outcomes between period-averaged and spot-sampled data are most pronounced. Validating results of Said and Dickey [1984] and simulations by Hall [1994], the SIC-based general-to-specific modeling strategy appears to have allowed inclusion of enough lagged difference terms in the selected ADF regression to capture and account for the moving average error terms and therefore preserve statistical power. In contrast, ADFR and ADFT tests conducted on a period-averaged process with a negative moving average error parameter have empirical sizes closer to nominal sizes than the same tests conducted on a sampled series.

We also conducted the same simulations on spot-sampled and period-averaged data for trend versions of the ADFT and the ADFR (results not shown). When confronted with either kind of data, power and empirical size differences between the ADFT and ADFR tests were qualitatively similar to differences found with the no-trend versions. We included no trend in our Monte Carlo-generated series, so including trends in the ADFT and ADFR tests had two effects. First, it inflated empirical sizes in the face of a true unit root null. At 5 percent nominal significance, the empirical sizes were 0.03 higher than the nominal rate in the ADFT and 0.08 higher than the nominal rate for the ADFR for positive and zero MA(1) parameters. Second, including trends lowered each test's statistical power to detect a false unit root null, typically by about 5 percent and 15 percent lower than the no-trend versions of the ADFT and ADFR, respectively. In the case of the ADFR, the reduction in statistical power compared to the no-trend version was greatest for sampled data, which fell by 15 to 20 percent more, compared to the no-trend version of the test.

Table 4 reports the empirical sizes of the Leybourne-McCabe test at 5 percent nominal significance. The power (the empirical size reported for

TABLE 4. Leybourne-McCabe stationarity (unit root) test (trend version) empirical sizes for 10,000 simulated 312-observation, e.g., pseudo-monthly, series converted to pseudo-quarterly spot or pseudo-quarterly averaged series of effective sample sizes of 100 observations. Size estimates are based on the four autoregressive terms (lags), evaluated at 5 percent significance using the nominal levels reported by Kwiatkowski et al. [1992].

AR(1) Parameter	MA(1) Parameter	Empirical Sizes (Rejection Rates) at 5% Nominal Significance	
		Samples	Period-Averages
1.00	0.00	0.862	0.875
0.98	0.00	0.835	0.844
0.95	0.00	0.693	0.718
0.90	0.00	0.508	0.539
0.80	0.00	0.295	0.348
0.50	0.00	0.149	0.182
1.00	0.80	0.836	0.862
0.98	0.80	0.839	0.834
0.95	0.80	0.706	0.725
0.90	0.80	0.517	0.556
0.80	0.80	0.316	0.365
0.50	0.80	0.124	0.047
1.00	-0.80	0.705	0.799
0.98	-0.80	0.605	0.727
0.95	-0.80	0.388	0.543
0.90	-0.80	0.219	0.305
0.80	-0.80	0.136	0.129
0.50	-0.80	0.125	0.051

$\rho = 1$) of this stationary-null test to detect a truly nonstationary series with no moving average term is 0.88 for period-averaged data and 0.86 for sampled data. With a moving average term it is often as correct as the ADFR. It also mimics the ADFR's weaker ability to detect a unit root using sampled data, with rejection rates of stationarity of 0.80 to 0.86 for period-averaged data and 0.71 to 0.84 for sampled data. Empirical sizes are greatly distorted for both averaged and sampled series when confronted with highly autocorrelated series with nonunitary roots, however. For period-averaged stationary series that have an autoregressive parameter above 0.9 in the original series, the empirical size of the Leybourne-McCabe test is ten to fifteen times above its nominal 5 percent significance level. In fact, the ability of the Leybourne-McCabe test to detect a truly stationary series with high non-unitary roots using either period-averaged or sampled data is comparable to that of the ADFT and worse than the ADFR. The test when applied to period-averaged data demonstrates empirical sizes that approach nominal sizes when the AR(1) parameter is low, e.g., 0.5, while the sampled series still has inflated empirical sizes. On the other hand, when the AR parameter is high, spot-sampled data produce empirical sizes slightly closer to the nominal level of the test (although still severely distorted). For example, when the AR parameter is 0.9, the empirical size of the period averaged data is about 0.05 larger. These size distortions mirror simulations of Caner and Killian [2001]. In summary, in the evaluation of period-averaged and sampled data the stationarity-null Leybourne-McCabe test (viewed in isolation) is no better than the ADFR at detecting a truly nonstationary series. It is substantially worse than the ADFR and comparable to the ADFT at correctly identifying series with large but non-unitary autoregressive roots.

Empirical results. Tables 5–7 summarize the results of our chosen representatives of the three classes of unit root tests of pine sawtimber stumpage, pine pulpwood stumpage, and mixed hardwood pulpwood stumpage price series, including nominal, PPI-deflated, CPI-deflated, and stocks-deflated. The penultimate column in Tables 5, 6 and 7 reports conclusions, rendered at 5 percent nominal significance, of whether the Leybourne-McCabe, ADFT, and ADFR tests agree that the series is stationary (S) (which, incidentally, was never found), nonstationary (N), or in conflict (I). The last column in these tables evaluates whether the Lo-MacKinlay and ADFT tests agree (Y) on whether the series is a random walk. Table 8 summarizes results of those

TABLE 5. Unit root tests on southern pine sawtimber stumpage nominal, PPI-deflated, CPI-deflated and S&P-500 earnings-deflated prices, 1977:I–2002:II.

Sub-Market [Transfor-mation]	Leybourne-McCabe Test with Trend	Lo-MacKinlay Test Without Trend	ADF Tests Without Trend				Form at 5%	RW?
			ADFT	ADFR	Obs.	Lags		
[Nominal]								
Alabama-1	1.29***	-1.50	-0.46	-0.63	85	7	N	
Alabama-2	1.16***	-0.79	-0.91	-2.05	85	6	N	
Arkansas-1	1.41***	-1.54	-0.95	-2.05	85	2	N	
Florida-1	0.90***	-2.02**	-1.37	-11.54*	85	16	N	
Florida-2	0.71***	-1.64	-1.86	-6.13	85	0	N	Y
Georgia-1	0.78***	-1.41	-1.55	-19.52***	85	16	I	
Georgia-2	1.36***	-1.21	-0.69	-0.89	85	5	N	
Louisiana-1	1.57***	-1.49	-0.93	-2.57	85	10	N	
Louisiana-2	1.48***	-1.55	-0.58	-0.95	85	6	N	
Mississippi-1	1.31***	-1.84*	-1.02	-2.46	85	1	N	
Mississippi-2	1.37***	-1.27	-1.05	-2.86	85	4	N	
No. Carolina-1	1.63***	-2.29**	1.82	3.03	85	11	N	
No. Carolina-2	1.92***	-2.00**	0.49	0.52	85	5	N	
So. Carolina-1	1.45***	-1.96*	-1.25	-4.19	85	11	N	
So. Carolina-2	1.51***	-1.16	-0.45	-0.51	85	6	N	
Tennessee-1			-2.15	-8.17	74	9		
Tennessee-2			-1.84	-5.94	65	0		
Texas-1	1.44***	-0.98	-1.41	-3.32	85	0	N	Y
Texas-2	1.40***	-1.18	-1.49	-12.75*	85	14	N	
Virginia-1			-0.09	-0.13	85	8		
Virginia-2	0.11	-1.99*	-0.34	-0.38	85	11	I	
[PPI-Deflated]								
Alabama-1	1.13***	-1.32	-0.64	-1.14	85	7	N	
Alabama-2	1.37***	-0.73	-1.74	-6.17	85	0	N	Y
Arkansas-1	1.78***	-1.41	-1.21	-3.41	85	2	N	
Florida-1	1.28***	-1.64	-1.34	-9.14	85	16	N	
Florida-2	0.83***	-1.39	-2.11	-8.37	85	0	N	Y
Georgia-1	0.82***	-1.46	-1.48	-6.29	85	11	N	
Georgia-2	1.43***	-1.28	-0.81	-1.38	85	5	N	
Louisiana-1	1.80***	-1.37	-0.79	-1.68	85	6	N	
Louisiana-2	1.80***	-1.46	-0.94	-2.30	85	5	N	
Mississippi-1	1.43***	-1.48	-1.68	-3.80	85	0	N	Y

TABLE 5 CONT'D.

Sub-Market [Transformation]	Leybourne-McCabe Test with Trend	Lo-MacKinlay Test Without Trend	ADF Tests Without Trend				Form at 5%	RW?
			ADFT	ADFR	Obs.	Lags		
[PPI-Deflated]								
Mississippi-2	1.62***	-1.13	-1.07	-2.90	85	2	N	
No. Carolina-1	1.75***	-2.19**	-0.25	-0.79	85	4	N	
No. Carolina-2	1.88***	-2.08**	0.09	0.13	85	5	N	
So. Carolina-1	1.43***	-2.01**	-1.25	-3.82	85	11	N	
So. Carolina-2	1.61***	-1.40	-1.67	-4.45	85	0	N	Y
Tennessee-1			-2.45	-11.88*	76	7		
Tennessee-2			-1.83	-5.64	65	0		
Texas-1	1.76***	-1.13	-1.87	-5.46	85	0	N	Y
Texas-2	1.76***	-1.13	-1.21	-3.53	85	2	N	
Virginia-1			-0.85	-1.98	85	3		
Virginia-2	0.06	-1.58	-2.25	-6.59	85	0	I	
[CPI-Deflated]								
Alabama-1	1.25***	-1.46	-0.89	-2.13	85	7	N	
Alabama-2	1.47***	-0.85	-2.13	-9.98	85	0	N	Y
Arkansas-1	1.83***	-1.41	-1.60	-5.36	85	2	N	
Florida-1	1.41***	-1.73*	-1.72	-16.42**	85	16	I	
Florida-2	0.91***	-1.49	-2.47	-13.05*	85	0	N	Y
Georgia-1	0.88***	-1.61	-1.61	-11.13*	85	11	N	
Georgia-2	1.51***	-1.47	-1.09	-2.80	85	5	N	
Louisiana-1	1.81***	-1.28	-1.37	-3.56	85	6	N	
Louisiana-2	1.83***	-1.44	-1.35	-4.15	85	5	N	
Mississippi-1	1.46***	-1.49	-2.19	-6.51	85	0	N	Y
Mississippi-2	1.69***	-1.15	-1.55	-5.50	85	2	N	
No. Carolina-1	1.82***	-2.24**	-1.49	-15.48**	85	13	I	
No. Carolina-2	1.94***	-2.16**	-0.75	-1.76	85	5	N	
So. Carolina-1	1.54***	-2.11**	-1.41	-6.14	85	11	N	
So. Carolina-2	1.15***	-1.49	-1.43	-6.48	85	11	N	
Tennessee-1			-2.69*	-21.07***	76	7		
Tennessee-2			-2.18	-7.83	65	0		
Texas-1	1.75***	-0.98	-2.25	-7.62	85	0	N	Y
Texas-2	1.77***	-1.02	-1.64	-5.46	85	2	N	
Virginia-1			-1.19	-3.59	85	3		
Virginia-2	0.08	-1.78*	-3.02**	-12.72*	85	0	I	

TABLE 5 CONT'D.

Sub-Market [Transformation]	Leybourne-McCabe Test with Trend	Lo-MacKinlay Test Without Trend	ADF Tests Without Trend				Form at 5%	RW?
			ADFT	ADFR	Obs.	Lags		
[Stocks-Deflated]								
Alabama-1	1.20***	-0.55	-1.31	-2.42	85	0	N	Y
Alabama-2	0.86***	0.26	-0.98	-1.48	85	6	N	
Arkansas-1	1.25***	0.09	-1.45	-2.53	85	0	N	Y
Florida-1	0.14*	0.45	-1.06	-1.14	85	0	I	
Florida-2	0.70***	0.25	-0.84	-1.35	85	0	N	Y
Georgia-1	0.50***	-0.16	-1.08	-2.35	85	0	N	Y
Georgia-2	0.66***	0.62	-1.10	-1.48	85	0	N	Y
Louisiana-1	1.29***	0.05	-1.56	-3.10	85	11	N	
Louisiana-2	1.29***	0.02	-1.37	-2.82	85	0	N	Y
Mississippi-1	0.67***	0.07	-1.59	-2.73	85	0	N	Y
Mississippi-2	1.08***	0.33	-1.37	-2.20	85	2	N	
No. Carolina-1	1.13***	-1.55	-1.79	-2.15	85	2	N	
No. Carolina-2	0.16**	-0.66	-1.70	-1.40	85	5	N	
So. Carolina-1	0.85***	-1.50	-1.31	-2.65	85	11	N	
So. Carolina-2	0.82***	0.81	-1.10	-1.64	85	0	N	Y
Tennessee-1			-1.25	-1.81	83	0		
Tennessee-2			-1.71	-3.35	65	0		
Texas-1	1.23***	0.44	-1.63	-2.80	85	0	N	Y
Texas-2	1.23***	0.35	-1.59	-3.64	85	11	N	
Virginia-1			-1.39	-2.30	85	3		
Virginia-2	0.06	-0.71	-1.21	-2.15	85	0	I	

Notes: Asterisks indicate rejection of null hypotheses (stationarity for Leybourne-McCabe and nonstationarity for Lo-MacKinlay, ADFT, and ADFR tests) at 1% (***), 5% (**), and 10% (*) significance (MacKinnon [1991], Hamilton [1994]; Kwiatkowski et al. [1992]). The form of the series reported in the penultimate column describes whether, at 5 percent nominal significance, the Leybourne-McCabe, ADFT, and ADFR tests agree that the series is stationary (S), nonstationary (N), or in conflict (I). The last column evaluates whether the Lo-MacKinlay and ADFT tests agree (Y) on whether the series is a random walk.

TABLE 6. Unit root tests on nominal, CPI-deflated and stocks-deflated southern pine pulpwood stumpage prices, 1977:I–2002:II.

Sub-Market [Transformation]	Leybourne-McCabe Test With Trend	Lo-MacKinlay Test Without Trend	ADF Tests Without Trend				Form at 5%	RW?
			ADFT	ADFR	Obs.	Lags		
[Nominal]								
Alabama-1	0.72***	-1.45	-1.73	-17.36**	85	12	I	
Alabama-2	0.79***	-0.80	-1.53	-6.70	85	6	N	
Arkansas-1	0.87***	-1.87*	-2.63*	-13.54*	85	1	N	
Florida-1	0.12*	-1.64	-2.88*	-10.28	85	0	I	
Florida-2	1.24***	-1.52	-1.65	-12.26*	85	8	N	
Georgia-1	0.92***	-1.64	-1.88	-1170.7***	85	15	I	
Georgia-2	0.81***	-1.60	-1.55	-10.62	85	12	N	
Louisiana-1	0.55***	-0.39	-1.72	-6.23	85	14	N	
Louisiana-2	0.48***	-1.73*	-2.02	-6.32	85	2	N	
Mississippi-1	0.75***	-0.80	-2.11	-45.91***	85	13	I	
Mississippi-2	0.64***	-1.43	-1.37	-3.64	85	6	N	
No. Carolina-1	1.06***	-2.38**	-1.58	-5.71	85	11	N	
No. Carolina-2	0.08	-2.01**	-1.08	-2.30	85	14	I	
So. Carolina-1	0.67***	-1.97*	-1.40	-3.92	85	4	N	
So. Carolina-2	0.63***	-1.59	-1.81	-6.09	85	2	N	
Tennessee-1			-1.74	-7.00	70	5		
Tennessee-2			-1.65	-6.05	55	7		
Texas-1	0.36***	-1.37	-2.25	-14.07**	85	7	I	
Texas-2	0.41***	-1.07	-2.51	-9.62	85	0	N	Y
Virginia-1			0.09	0.13	50	15		
Virginia-2	0.21**	-2.42**	0.25	0.57	85	15	N	
[PPI-Deflated]								
Alabama-1	0.74***	-1.34	-2.01	73.43	85	12	N	
Alabama-2	0.76***	-0.95	-1.60	-13.78**	85	6	I	
Arkansas-1	0.54***	-1.73*	-2.28	-12.10*	85	1	N	
Florida-1	0.63***	-1.58	-2.48	-9.20	85	0	N	Y
Florida-2	1.10***	-1.56	-1.73	237.61	85	8	N	
Georgia-1	0.94***	-1.34	-2.18	16.26	85	15	N	
Georgia-2	0.82***	-1.50	-1.54	-144.94***	85	12	I	
Louisiana-1	0.35***	-0.87	-2.47	-11.55*	85	0	N	Y
Louisiana-2	0.44***	-1.69*	-2.57	-10.49	85	2	N	
Mississippi-1	0.77***	-0.70	-2.18	-91.62***	85	13	I	

TABLE 6 CONT'D.

Sub-Market [Transformation]	Leybourne-McCabe Test with Trend	Lo-MacKinlay Test Without Trend	ADF Tests Without Trend				Form at 5%	RW?
			ADFT	ADFR	Obs.	Lags		
[PPI-Deflated]								
Mississippi-2	0.81***	-1.34	-1.97	-9.41	85	4	N	
No. Carolina-1	1.05***	-2.33**	-1.63	-4.66	85	2	N	
No. Carolina-2	0.39***	-1.86*	-2.00	-15.63**	85	11	I	
So. Carolina-1	0.89***	-1.91*	-1.44	-5.21	85	4	N	
So. Carolina-2	0.70***	-1.26	-1.98	-15.11**	85	7	I	
Tennessee-1			-2.08	-11.27*	75	2		
Tennessee-2			-1.77	-5.06	55	7		
Texas-1	0.34***	-1.37	-3.04**	-28.43***	85	4	I	
Texas-2	0.39***	-0.99	-2.58	-12.08*	85	0	N	Y
Virginia-1			0.10	0.18	50	15		
Virginia-2	0.06	-2.24**	-1.95	-81.40***	85	10	I	
[CPI-Deflated]								
Alabama-1	0.71***	-1.39	-2.17	16.46	85	12	N	
Alabama-2	0.75***	-0.96	-0.84	-6.86	85	6	N	
Arkansas-1	0.47***	-1.65	-1.41	-5.89	85	1	N	
Florida-1	0.58***	-1.51	-1.78	-4.70	85	0	N	Y
Florida-2	1.05***	-1.50	-0.50	-4.98	85	8	N	
Georgia-1	0.92***	-1.33	-2.41	11.78	85	15	N	
Georgia-2	0.76***	-1.43	-1.03	-2685.4***	85	12	I	
Louisiana-1	0.30***	-0.82	-2.22	-74.76***	85	7	I	
Louisiana-2	0.44***	-1.75*	-3.75***	-21.35***	85	0	I	
Mississippi-1	0.75***	-0.74	-2.60*	36.86	85	13	N	
Mississippi-2	0.85***	-1.38	-1.39	-6.07	85	6	N	
No. Carolina-1	0.86***	-2.43**	-1.60	-21.23***	85	13	I	
No. Carolina-2	0.32***	-2.03**	-3.08**	62.26	85	11	I	
So. Carolina-1	0.94***	-1.98*	-1.57	-9.05	85	4	N	
So. Carolina-2	0.74***	-1.37	-0.72	-3.43	85	3	N	
Tennessee-1			-3.96***	-32.29***	79	0		
Tennessee-2			-1.95	-7.80	55	7		
Texas-1	0.34***	-1.43	-3.08**	-46.49***	85	4	I	
Texas-2	0.38***	-1.02	-2.26	-11.04*	85	0	N	Y
Virginia-1			-0.50	-1.85	50	15		
Virginia-2	0.10	-2.36**	-2.31	28.08	85	12	I	

TABLE 6 CONT'D.

Sub-Market [Transformation]	Leybourne-McCabe Test with Trend	Lo-MacKinlay Test Without Trend	ADF Tests Without Trend				Form at 5%	RW?
			ADFT	ADFR	Obs.	Lags		
[Stocks-Deflated]								
Alabama-1	0.96***	-0.11	0.10	0.13	85	1	N	
Alabama-2	0.82***	0.63	-0.28	-0.21	85	0	N	Y
Arkansas-1	0.71***	0.12	-0.28	-0.34	85	1	N	
Florida-1	0.06	0.48	-0.30	-0.15	85	0	I	
Florida-2	1.16***	0.26	0.49	0.43	85	3	N	
Georgia-1	0.98***	0.32	-0.24	-0.55	85	12	N	
Georgia-2	0.88***	0.43	-0.31	-0.19	85	0	N	Y
Louisiana-1	0.66***	0.29	-0.24	0.05	85	0	N	Y
Louisiana-2	0.54***	-0.77	-0.75	-1.31	85	0	N	Y
Mississippi-1	0.71***	0.23	-0.03	-0.04	85	10	N	
Mississippi-2	0.61***	0.22	-0.20	-0.25	85	6	N	
No. Carolina-1	1.17***	-1.48	-0.70	-1.84	85	13	N	
No. Carolina-2	0.58***	-0.20	-0.52	-0.52	85	3	N	
So. Carolina-1	0.68***	-0.77	-0.32	-0.31	85	3	N	
So. Carolina-2	0.66***	0.22	0.13	0.13	85	3	N	
Tennessee-1			-0.01	-0.03	60	15		
Tennessee-2			-0.34	-0.78	45	12		
Texas-1	0.48***	-1.00	-0.39	-0.80	85	4	N	
Texas-2	0.56***	-0.37	-0.32	-0.26	85	0	N	Y
Virginia-1			-0.34	-0.07	50	15		
Virginia-2	0.11	-1.64	-0.93	-0.84	85	16	I	

Notes: Asterisks indicate rejection of null hypotheses (stationarity for Leybourne-McCabe and nonstationarity for Lo-MacKinlay, ADFT, and ADFR tests) at 1% (***), 5% (**), and 10% (*) significance (MacKinnon [1991], Hamilton [1994]; Kwiatkowski et al. [1992]). The form of the series reported in the penultimate column describes whether, at 5% nominal significance, the Leybourne-McCabe, ADFT, and ADFR tests agree that the series is stationary (S), nonstationary (N), or in conflict (I). The last column evaluates whether the Lo-MacKinlay and ADFT tests agree (Y) on whether the series is a random walk.

TABLE 7. Unit root tests on nominal, PPI-deflated, CPI-deflated and Standard and Poor's 500 earnings-deflated hardwood pulpwood stumpage prices, 1977:I–2002:II.

Sub-Market [Transformation]	Leybourne-McCabe Test With Trend	Lo-MacKinlay Test Without Trend	ADF Tests Without Trend				Form at 5%	RW?
			ADFT	ADFR	Obs.	Lags		
[Nominal]								
Alabama-1	0.69***	-2.12**	-1.17	-2.93	85	4	N	
Alabama-2	0.81***	-0.92	-1.10	-1.85	85	4	N	
Arkansas-1	0.12*	-1.90*	-0.46	-0.92	85	15	I	
Florida-1	0.10	-2.42**	-1.82	-6.74	85	2	I	
Florida-2	0.75***	-2.03**	-0.53	-1.03	85	13	N	
Georgia-1	0.33***	-2.01**	-0.84	-2.17	85	13	N	
Georgia-2	0.11	-1.76*	0.03	0.04	85	9	I	
Louisiana-1			-0.78	-1.58	77	6		
Louisiana-2	0.10	-2.31**	-1.39	-4.64	85	2	I	
Mississippi-1	1.01***	-2.01**	-1.06	-3.68	85	16	N	
Mississippi-2	0.62***	-2.07**	-1.18	-2.61	85	7	N	
No. Carolina-1	0.68***	-2.26**	-0.91	-0.87	85	7	N	
No. Carolina-2	0.47***	-2.29**	-1.21	-2.65	85	5	N	
So. Carolina-1	1.37***	-1.58	-0.80	-1.49	85	2	N	
So. Carolina-2	0.38***	-1.70*	-0.46	-0.70	85	9	N	
Tennessee-1			-0.74	-1.97	85	11		
Tennessee-2			-1.14	-3.26	85	15		
Texas-1	0.36***	-1.98*	-0.62	-1.91	85	8	N	
Texas-2			-1.44	-4.66	77	5		
Virginia-1			-0.75	-2.03	78	5		
Virginia-2	0.08	-2.32**	-0.35	-1.00	85	12	I	
[PPI-Deflated]								
Alabama-1	0.69***	-1.71*	-1.03	-2.08	85	4	N	
Alabama-2	0.70***	-0.81	-1.52	-3.16	85	0	N	Y
Arkansas-1	0.10	-1.71*	-0.87	-2.64	85	9	I	
Florida-1	0.10	-2.26**	-1.83	-7.15	85	2	I	
Florida-2	0.14*	-1.37	-1.70	-4.04	85	0	I	
Georgia-1	0.26***	-1.59	-0.79	-1.59	85	6	N	
Georgia-2	0.52***	-1.54	-0.32	-0.56	85	11	N	
Louisiana-1			-1.02	-2.23	73	10		
Louisiana-2	0.72***	-2.16**	-1.18	-3.49	85	2	N	
Mississippi-1	0.09	-1.86*	-1.13	-5.26	85	16	I	

TABLE 7 CONT'D.

Sub-Market [Transformation]	Leybourne-McCabe Test with Trend	Lo-MacKinlay Test Without Trend	ADF Tests Without Trend				Form at 5%	RW?
			ADFT	ADFR	Obs.	Lags		
[PPI-Deflated]								
Mississippi-2	0.80***	-2.00**	-1.74	-6.07	85	2	N	
No. Carolina-1	0.82***	-2.10**	-1.37	-1.66	85	7	N	
No. Carolina-2	0.54***	-2.29**	-1.19	-2.74	85	5	N	
So. Carolina-1	1.40***	-1.43	-0.74	-1.38	85	2	N	
So. Carolina-2	1.35***	-1.38	-1.00	-2.30	85	1	N	
Tennessee-1			-0.94	-2.01	85	5		
Tennessee-2			-1.88	-4.67	85	0		
Texas-1	0.43***	-2.06**	-0.43	-0.91	85	7	N	
Texas-2			-2.94**	-11.51*	82	0		
Virginia-1			-1.61	-7.34	80	3		
Virginia-2	0.11	-2.26**	-1.46	-15.20**	85	12	I	
[CPI-Deflated]								
Alabama-1	0.66***	-1.73*	-1.18	-2.97	85	4	N	
Alabama-2	0.64***	-0.90	-1.65	-4.24	85	0	N	Y
Arkansas-1	0.11	-1.72*	-1.20	-5.11	85	9	I	
Florida-1	0.09	-2.23**	-2.25	-11.25*	85	2	I	
Florida-2	0.17**	-1.36	-2.10	-6.22	85	0	N	Y
Georgia-1	0.09	-1.56	-1.04	-2.81	85	6	I	
Georgia-2	0.45***	-1.56	-0.55	-1.29	85	11	N	
Louisiana-1			-1.24	-4.00	73	10		
Louisiana-2	0.80***	-2.17**	-1.53	-5.59	85	2	N	
Mississippi-1	0.09	-1.84*	-1.51	-13.10*	85	16	I	
Mississippi-2	0.80***	-2.01**	-3.52***	-15.55**	85	0	I	
No. Carolina-1	0.79***	-2.23**	-1.38	-2.37	85	7	N	
No. Carolina-2	0.51***	-2.34**	-1.69	-12.51*	85	13	N	
So. Carolina-1	1.42***	-1.52	-0.95	-2.16	85	2	N	
So. Carolina-2	0.10	-1.47	-1.61	-3.36	85	0	I	
Tennessee-1			-1.13	-3.13	85	5		
Tennessee-2			-2.14	-6.21	85	0		
Texas-1	0.44***	-2.07**	-0.68	-1.85	85	7	N	
Texas-2			-3.39**	-15.70**	82	0		
Virginia-1			-2.24	-66.02***	68	15		
Virginia-2	0.12*	-2.26**	-2.77*	-66.33***	85	12	I	

TABLE 7 CONT'D.

Sub-Market [Transformation]	Leybourne-McCabe Test with Trend	Lo-MacKinlay Test Without Trend	ADF Tests Without Trend				Form at 5%	RW?
			ADFT	ADFR	Obs.	Lags		
[Stocks-Deflated]								
Alabama-1	0.62***	-1.07	-1.62	-4.97	85	1	N	
Alabama-2	0.67***	-0.01	-1.25	-2.39	85	0	N	Y
Arkansas-1	0.10	-1.08	-1.73	-7.41	85	9	I	
Florida-1	0.46***	-1.80*	-0.49	-0.87	85	15	N	
Florida-2	0.61***	-0.48	-1.93	-4.58	85	0	N	Y
Georgia-1	0.56***	-0.84	-1.84	-11.31*	85	12	N	
Georgia-2	0.63***	-0.77	-1.99	-7.84	85	7	N	
Louisiana-1			-1.00	-1.98	77	6		
Louisiana-2	0.08	-1.53	-2.23	-5.74	85	0	I	
Mississippi-1	0.78***	-1.16	-1.48	-4.03	85	1	N	
Mississippi-2	0.59***	-1.19	-1.65	-3.45	85	0	N	Y
No. Carolina-1	0.83***	-1.05	-1.00	-1.83	85	5	N	
No. Carolina-2	0.61***	-1.72*	-0.79	-1.33	85	7	N	
So. Carolina-1	0.74***	-0.75	-1.52	-5.90	85	14	N	
So. Carolina-2	0.98***	-0.58	-2.09	-5.34	85	0	N	Y
Tennessee-1			-1.93	-23.48***	85	11		
Tennessee-2			-2.16	-6.32	85	0		
Texas-1	0.10	-1.77*	-2.87*	-12.38*	85	0	I	
Texas-2			-2.52	-8.59	82	0		
Virginia-1			-1.49	-2.75	80	3		
Virginia-2	0.23***	-1.21	-0.67	-0.43	85	15	N	

Notes: Asterisks indicate rejection of null hypotheses (stationarity for Leybourne-McCabe and nonstationarity for Lo-MacKinlay, ADFT, and ADFR tests) at 1% (***), 5% (**), and 10% (*) significance (MacKinnon [1991], Hamilton [1994]; Kwiatkowski et al. [1992]). The form of the series reported in the penultimate column describes whether, at 5% nominal significance, the Leybourne-McCabe, ADFT, and ADFR tests agree that the series is stationary (S), nonstationary (N), or in conflict (I). The last column evaluates whether the Lo-MacKinlay and ADFT tests agree (Y) on whether the series is a random walk.

TABLE 8. Proportion of consensus conclusions about the time series properties of southern timber prices, by product and deflation method (1977:I–2002:II).

Product	Deflation Method	Consensus of Tests in Favor of		Consensus For or Against
		A Unit Root	Stationarity	Random Walk
Southern Pine Sawtimber	Nominal	0.89	0.00	0.28
	PPI-deflated	0.94	0.00	0.44
	CPI-deflated	0.83	0.00	0.39
	Stocks-deflated	0.89	0.00	0.50
Southern Pine Pulpwood	Nominal	0.67	0.00	0.22
	PPI-deflated	0.61	0.00	0.28
	CPI-deflated	0.61	0.00	0.28
	Stocks-deflated	0.89	0.00	0.28
Hardwood Pulpwood	Nominal	0.69	0.00	0.63
	PPI-deflated	0.69	0.00	0.50
	CPI-deflated	0.56	0.00	0.50
	Stocks-deflated	0.81	0.00	0.25

conclusions using nominal significance thresholds. For sawtimber stumpage, the ADFR rejected the null four times out of 84 at 5 percent nominal significance, while the ADFT test rejected a unit root once at this significance. The rejections for the ADFR were in Florida, North Carolina, and Tennessee. These results are consistent with the simple Dickey-Fuller t-test results reported by Yin et al. [2002] for southern pine sawtimber. That study used a shorter span, a smaller subset of sub-markets within the South, and different spatial aggregations compared to those that we examine here, where a unit root was supported for southern pine sawtimber stumpage prices. Given our findings in the simulations, the ADFR and the ADFT tend to reject the null too often with period-averaged data, so our results for timber prices might be surprising. On the other hand, the low rate of rejections just as well might be reflecting the low power of the ADF at detecting a high, nonunitary root.

The Leybourne-McCabe test provides additional support for the contention that sawtimber prices are nonstationary Southwide. In all but four cases (and just Virginia sub-market 2 under each deflation method), the Leybourne-McCabe test rejected the null of a stationary process at 5 percent significance. The Lo-MacKinlay variance ratio test rarely rejected the random walk null at the 5 percent nominal significance level. Rejections were found in nine cases out of 72 across deflation methods for southern pine sawtimber, most in the Carolinas.

The results of the Leybourne-McCabe, the ADFT, and the ADFR tests can be combined into a conclusion of stationary or nonstationary price behavior based on a consensus of the three. This shows that the majority of sawtimber price series are likely to be nonstationary at nominal significance levels (Table 8), regardless of deflation method. Nonstationarity was the consensus for 83 percent of CPI-deflated, 89 percent of stocks-deflated and nominal series, and 94 percent of PPI-deflated series. Stationarity was never the consensus. For sawtimber, the random walk rarely gained consensus support (Table 5), but the rate of acceptance of this hypothesis was higher for deflated series than for nominal series. Agreement was most common for stocks-deflated prices (50 percent) and least common for nominal prices (28 percent) (Table 8). An additional caveat is that the random walk may only have been accepted for these series because of the anomalous simplification of the sawtimber price process caused by a high degree of temporal aggregation in short time series (Rossana and Seater [1995]).

Results for southern pine pulpwood (Table 6) were similar to those of sawtimber stumpage. The ADFT favored the stationarity alternative hypothesis in five cases out of 84, while the ADFR favored this in 17 cases. Lo-MacKinlay tests rejected the random walk null hypothesis eight times out of 72, about the same rate as for sawtimber. Only for stocks-deflated timber prices was the random walk never rejected using this test. The Leybourne-McCabe test rejected the stationary null in every case except three, all in coastal Virginia. When the results of tests are combined we see that the nonstationary price consensus was most common for stocks-deflated and least common for CPI-deflated prices. Consensus for or against random walk pulpwood price behavior was even lower for pulpwood than for sawtimber prices, with the greatest conflict found for nominal prices.

Results for hardwood pulpwood prices (Table 7) are available for a smaller subset of sub-markets in the South because data limitations were more common than for southern pine. Results provide a more mixed picture than for southern pine products and are less consistent between the three tests. In three cases out of 84, the ADFT rejected the nonstationary null, and in five cases out of 84 the ADFR rejected the null. However, in 18 out of 64 cases the Leybourne-McCabe test could not reject the stationarity null, a much higher rate of disagreement than found for pine. Also different from southern pine pulpwood, the Lo-MacKinlay test rejected the random walk three times as often—24 out of 64 times. The random walk was favored by consensus in seven cases out of 64, with four of those in stocks-deflated prices. Still, in nearly 70 percent of cases at nominal significance thresholds, hardwood pulpwood stumpage prices appear to conform to unit root price behavior, with that behavior most commonly identified in stocks-deflated prices and least common in CPI-deflated prices.

Results of the tests on these products using the trend versions of the ADFT, ADFR, and the Lo-MacKinlay tests, not reported in tables, were similar to results without a trend. The primary difference is that the trend versions resulted in a slightly higher rate of rejection of the unit root null and the random walk. This higher rejection rate of the null of a unit root for the ADFT and ADFR is not surprising, given our simulation results for trend-versions of these tests, which are over-sized with period-averaged data and small samples. The ADFT rejection rates at 5 percent nominal significance were 4 out of 84 for southern pine sawtimber (1 in the no-trend case), 9 out of 84 for southern pine pulpwood (5 in the no-trend case), and 14 out of 84 for mixed hardwood pulpwood (3 in the no-trend case). Using the ADFR, these rates were 18 (9 with the no-trend), 16 (17 in the no-trend), and 37 (5 in the no-trend case) out of 84 for southern pine sawtimber, southern pine pulpwood, and mixed hardwood pulpwood, respectively. Using the trend version of the Lo-MacKinlay test, the corresponding rejection rates at 5 percent nominal significance were 13 out of 72 for southern pine sawtimber, 11 out of 72 for southern pine pulpwood, and 28 out of 64 for mixed hardwood pulpwood. When a trend is included, the likelihood of rejecting the null of a unit root or the random walk is double the rate without a trend. However, when evaluated using the nominal significance levels, the majority of series are still consistent in behavior with series containing a unit root when tests include a trend in the ADFT, ADFR, and the Lo-MacKinlay tests.

Caner and Killian [2001] state that simulations can be used to empirically identify critical values so that the nominal size of the test matches its empirical size. This strategy might be applied to any test. In our simulations, we obtained the critical values necessary to attain empirical sizes that match the nominal size of the Leybourne-McCabe test for series of effective sizes of 100 observations for monthly data that were period-averaged to quarterly data. At 5 percent significance, the critical value was 1.24 for ARIMA(1,0,1) series with an AR(1) parameter of 0.95 and an MA(1) parameter of 0.80. Using this threshold and referring to the values reported in Table 5 for southern pine sawtimber, it can be seen that the null is rejected for 13 out of 18 tested nominal series (Alabama-2, Florida-1, Florida-2, Georgia-1, and Virginia-2 were not), 14 tested PPI- and CPI-deflated series, and 5 stocks-deflated series.

For pulpwood, our results are more definitive. For southern pine pulpwood (Table 6), the corresponding rejection rates using 1.24 for this test at 5 percent significance are 1, 0, 0, and 0 for 18 tested undeflated, PPI-deflated, CPI-deflated, and stocks-deflated series, respectively. For hardwood pulpwood, those rejection rates out of 16 tested series are 1, 2, 1, and 0, given the test statistics reported in Table 7.

We similarly obtained the critical values necessary to attain empirical sizes that match the nominal size of the no-trend ADFT test for an effective sample size of 85 observations with eight possible lagged difference terms. The empirical size-adjusted 5 percent ADFT critical value based on 1,000 simulated "monthly" ARIMA(0,1,1) series that were period-averaged to "quarterly" observations with a "monthly" MA(1) parameter equal to -0.80 was -3.94. Based on this threshold, we found the following rejection rates for southern timber prices: no rejections out of 84 tests of the unit root null for southern pine sawtimber, one rejection out of 84 for southern pine pulpwood, and no rejections out of 64 for hardwood pulpwood.

If the true rate of data generation of timber prices is monthly and so the size-adjusted critical value of 1.24 for the Leybourne-McCabe test and -3.94 for the no-trend ADFT approximately correspond to the 5 percent nominal significance levels for our timber price series, then we can conclude that the ADFT test mostly conflicts with the Leybourne-McCabe test for pulpwood series but still commonly supports a unit root for southern pine sawtimber. Sawtimber series, however, in their majority reject the stationarity null in the Leybourne-McCabe and accept the unit root null in the ADFT, even when the empirical critical values are adjusted to more closely match their nominal sizes.

Conclusions. These results have documented the effects of infrequent data reporting on typical unit root tests, and in light of this the time series properties of southern timber prices with applicability to choice of alternative harvest timing models. We therefore have four principal conclusions.

First, the period-averaged data reported by Timber Mart-South and most other sources of timber price data readily available to managers pose substantial statistical challenges for identifying true price processes. Unit root tests based on data that are reported less frequently than the true process are statistically weaker and empirically oversized. Both period-averaging and spot-sampling render the common unit root tests less powerful and distorted in size, compared to results derived from data reported at their true rate of generation but over the same time span. Published research into timber price behavior that has used unit-root null statistical tests such as the ADFT has therefore suffered from reduced statistical power, leading to a failure to reject unit roots even when the true process was stationary. We found, notably, that there are more powerful and sometimes no more complicated alternatives to the ADFT (e.g., the ADFR) that analysts could apply. However, we caution the analyst against using the trend version of the ADFR on such data when no solid reason exists to include such a trend, given how much the superfluous trend reduces the statistical power of the test.

Second, comparing the results of tests with opposing nulls (unit-root null and stationarity-null) might enhance confidence regarding the presence of a unit root, rather than simply evaluating one kind of test with its known statistical limitations (Taylor [2001], Caner and

Killian [2001]). Improved confidence using this strategy can also come about through simulations that adjust the critical values to match the empirical and nominal sizes of the tests. However, improved confidence using this approach is by no means a certainty, as empirical size adjustment has the drawback of reducing statistical power against a false null. So, sometimes adjusting these sizes can create new conflicts while resolving others. This strategy also depends on knowledge of the true rate of data generation and whether nuisance parameters exist in the data generation process of the true series, knowledge few researchers have. In our case, the strategy of comparing results across tests worked to raise our confidence about the behavior of southern pine sawtimber prices but not pulpwood prices. In the latter case, the comparison yielded conflicts between the stationarity- and unit root-null tests. Our simulations also revealed that, applied on their own, stationarity-null tests are not better than unit root-null tests at detecting truly stationary series, in slight contradiction to Taylor's [2001] recommendation for period-averaged or sampled data.

Third, analysts and timber owners are urged to favor nonstationary price harvest timing models for managing southern pine stands for sawtimber in the Southern U.S. Recommendations for pulpwood management, both southern pine and mixed hardwood, must be more cautious based on the conflicting results for southern pine and hardwood pulpwood when empirical size-adjusted critical values were used. The best stationary price harvest timing models for these products would seem to be those that include large autoregressive parameters, e.g., the stationary model of Haight and Holmes [1991]. Additionally, because many stands derive value from a combination of products whose mix varies over time, future research should clarify how to adaptively time harvests for stands with mixes of products, each with a different price process. Recent work by Reeves and Haight [2000], which explicitly accounted for the different price processes of pulpwood and sawtimber, including their large autoregressive parameters using a Markowitz portfolio approach, is a step in that direction.

Fourth, although neither random walk nor stationary timber price behaviors were fully supported by consensus across our empirical tests, our results do not discount the usefulness of timing models based on random walk or diffusion price processes. The moving average terms introduced through data collection and temporal averaging may cause the random walk or martingale test to fail even though the true series conforms to these behaviors (Working [1960], Brewer [1973], Harvey [1981], Haight and Holmes [1991], Maeso-Fernandez [1998], Taylor [2001]). Nevertheless, the underlying nonstationary timber price series may still be more complex than a random walk. A model that assumes that timber prices follow a more general ARIMA$(p, 1, q)$ or a general diffusion process might better capture observed producer behavior, e.g., Gomez et al. [1999]; Reeves and Haight [2000], Yoshimoto and Shoji [2002]).

Our analysis used the temporal aggregation question addressed by Haight and Holmes [1991] as a departure point. A more fundamental question we have not addressed is, "What is the 'true' rate of timber price generation for a region?" We can only speculate that the true rate is more frequent than monthly or quarterly. Nonetheless, our research does lead us to several recommendations on what a timber manager or analyst can do to enhance certainty about how timber prices actually behave. First, the analyst should use multiple testing frameworks—different tests—to raise confidence about actual price behavior. Second, although no clear results were produced on the effects of deflation on our test outcomes, we take seriously the critique of Schnute [1987] and recommend that tests of the time series properties of timber prices be conducted on the undeflated series. Third, the span of a series should be as long as possible. Span may be a more important determinant of the power of a test than frequency of observation, temporal averaging, or the presence of nuisance parameters. Fourth, unit root tests should be selected to allow for the presence of moving average parameters, which are introduced through simple sampling error as well as temporal aggregation. This means applying augmented versions of the ADFT and ADFR and tests designed to be powerful when moving average terms are present.

Acknowledgments. The authors thank David Dickey, Barry Goodwin, Alastair Hall, Brian Murray, Sastry Pantula and Andrew Plantinga for their advice in various stages of this paper's development. Any remaining errors, however, are our responsibility.

REFERENCES

J. Abildtrup, J. Riis and B.J. Thorsen [1997], *The Reservation Price Approach and Informationally Efficient Markets*, J. For. Econ. **3**, 229–245.

C. Agiakloglou and P. Newbold [1992], *Empirical Evidence on Dickey-Fuller Type Tests*, J. Time Ser. Anal. **13**, 471–483.

Anonymous [2002], *S&P* 500 *Total Return*; *Monthly Dividend Reinvest* 1970–88 *then Daily* (EOP). Data accessed by authors August 15, 2002. Available at http://www.economagic.com/em-cgi/data.exe/fedstl/trsp500+1.

R.J. Brazee and E. Bulte [2000], *Optimal Harvesting and Thinning with Stochastic Prices*, For. Sci. **46**, 23–31.

R. Brazee and R. Mendelsohn [1988], *Timber Harvesting with Fluctuating Prices*, For. Sci. **34**, 359–372.

K.R.W. Brewer [1973], *Some Consequences of Temporal Aggregation and Systematic Sampling for ARMA and ARMAX Models*, J. Econometrics **1**, 133–154.

D.M. Burton and H.A. Love [1996], *A Review of Alternative Expectations Regimes in Commodity Markets*: *Specification, Estimation, and Hypothesis Testing Using Structural Models*, Agr. Resources Econ. Rev. **18**, 214–231.

M. Caner and L. Killian [2001], *Size Distortions of Tests of the Null Hypothesis of Stationarity*: *Evidence and Implications for the PPP Debate*, J. Int. Money Finance **20**, 639–657.

S.G. Cecchetti and P. Lam [1994], *Variance-Ratio Tests*: *Small-Sample Properties with an Application to International Output Data*, J. Bus. Econ. Statist. **12**, 177–186.

H.R. Clarke and W.J. Reed [1989], *The Tree-Cutting Problem in a Stochastic Environment*: *The Case of Age-Dependent Growth*, J. Econ. Dynam. Control **13**, 565–595.

A. Deaton and G. Laroque [1996], *Competitive Storage and Commodity Price Dynamics*, J. Polit. Econ. **104**, 896–923.

D.A. Dickey and W.A. Fuller [1979], *Distribution of the Estimators for Autoregressive Time Series with a Unit Root*, J. Amer. Stat. Assoc. **74**, 427–431.

D.A. Dickey and W.A. Fuller [1981], *Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root*, Econometrica **49**, 1057–1072.

E.F. Fama [1970], *Efficient Capital Markets*: *A Review of Theory and Empirical Work*, J. Finance **25**, 383–417.

E.F. Fama [1991], *Efficient Capital Markets*: II, J. Finance **46**, 1575–1617.

P.F. Forboseh, R.J. Brazee and J.B. Pickens [1996], *A Strategy for Multiproduct Stand Management with Uncertain Future Prices*, For. Sci. **42**, 58–66.

W.A. Fuller [1976], *Introduction to Statistical Time Series*, Wiley, London, U.K.

I.A. Gomez, H.A. Love and D.M. Burton [1999], *Alternative Price Expectations Regimes in Timber Markets*, J. For. Econ. **5**, 235–252.

P. Gong [1998], *Risk Preferences and Adaptive Harvest Policies for Even-Aged Stand Management*, For. Sci. **44**, 496–506.

P. Gong [1999], *Optimal Harvest Policy with First-Order Autoregressive Price Process*, J. For. Econ. **5**, 413–439.

J. Gunter and F. Cubbage [1987], *Data Collection for Timber Mart-South*, For. Farm. **46**, 17–18.

R.G. Haight and T.P. Holmes [1991], *Stochastic Price Models and Optimal Tree Cutting*: *Results for Loblolly Pine*, Natural Res. Modeling **5**, 423–443.

R.G. Haight and W.D. Smith, [1991], *Harvesting Loblolly Pine Plantations with Hardwood Competition and Stochastic Prices*, For. Sci. **37**, 1266–1282.

A. Hall [1994], *Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection*, J. Bus. Econ. Statist. **12**, 461–470.

J.D. Hamilton [1994], *Time Series Analysis Techniques*, Princeton University Press, Princeton, NJ.

T.G. Harris, Jr., J.F. Brooks and M.E. Aranow [1999], *Timber Mart-South Quarterly Market Report for Timber Products in the Southeastern U.S.*, in *Proceedings of the* 1998 *Southern Forest Economics Workshop*, March 25–27, 1998, Williamsburg, VA.

A.C. Harvey [1981], *Time Series Models*, Philip Allan Publishers Ltd., Oxford, U.K.

L. Hultkrantz [1993], *Informational Efficiency of Markets for Stumpage*: *Comment*, Amer. J. Agr. Econ. **75**, 234–238.

D. Kwiatkowski, P.C.B. Phillips, P. Schmidt and Y. Shin [1992], *Testing the Null of Stationarity Against the Alternative of a Unit Root*: *How Sure Are We that Economic Time Series Have a Unit Root?*, J. Econometrics **54**, 159–178.

S.F. LeRoy [1989], *Efficient Capital Markets and Martingales*, J. Econ. Lit. **27**, 1583–1621.

S.J. Leybourne and B.P.M. McCabe [1994], *A Consistent Test for a Unit Root*, J. Bus. Econ. Statist. **12**, 157–166.

A.W. Lo and A.C. MacKinlay [1988], *Stock Prices Do Not Follow Random Walks*: *Evidence from a Simple Specification Test*, Rev. Finan. Stud. **1**, 41–66.

A.W. Lo and A.C. MacKinlay [1989], *The Size and Power of the Variance Ratio Test in Finite Samples*: *A Monte Carlo Investigation*, J. Econometrics **40**, 203–238.

P. Lohmander [1988], *Pulse Extraction under Risk and a Numerical Forestry Application*, Syst. Anal. Model Simul. **5**, 339–354.

J.G. MacKinnon [1991], *Critical Values for Cointegration Tests*, in *Long Run Economic Relationships*: *Readings in Cointegration* (R.F. Engle and C.W.J. Granger, eds.), Oxford University Press, London.

F. Maeso-Fernandez [1998], *Econometric Methods and Purchasing Power Parity*: *Short- and Long-Run* PPP, Appl. Econ. **30**, 1443–1457.

B. McGough, A.J. Plantinga and W. Provencher [2002], *The Dynamic Behavior of Efficient Timber Prices*, Staff Paper No. 454, Dept. of Agr. and Appl. Econ., University of Wisconsin-Madison.

Norris Foundation [1977–2002], *Timber Mart-South*, The Daniel B. Warnell School of Forest Resources, University of Georgia, Athens.

C.J. Norstrøm [1975], *A Stochastic Model for the Growth Period Decision in Forestry*, Swed. J. Econ. **77**, 329–337.

P. Perron [1996], *The Adequacy of Asymptotic Approximations in the Near-Integrated Autoregressive Model with Dependent Errors*, J. Econometrics **70**, 317–350.

P.C.B. Phillips and P. Perron [1988], *Testing for a Unit Root in Time Series Regression*, Biometrika **75**, 535–547.

A.J. Plantinga [1998], *The Optimal Timber Rotation: An Option Value Approach*, For. Sci. **44**, 192–202.

J.P. Prestemon [2003], *Evaluation of Southern Pine Stumpage Market Informational Efficiency*, Can. J. For. Res. **33**, 561–572.

J.P. Prestemon and T.P. Holmes [2000], *Timber Price Dynamics Following a Natural Catastrophe*, Amer. J. Agr. Econ. **82**, 145–160.

J.P. Prestemon and J.M. Pye [2000], *A Technique for Merging Areas in Timber Mart-South Data*, South. J. Appl. For. **28**, 219–229.

L.H. Reeves and R.G. Haight [2000], *Timber Harvest Scheduling with Price Uncertainty Using Markowitz Portfolio Optimization*, Ann. Oper. Res. **95**, 229–250.

R. Rossana and J. Seater [1995], *Temporal Aggregation and Economic Time Series*, J. Bus. Econ. Statist. **13**, 441–450.

S.E. Said and D.A. Dickey [1984], *Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order*, Biometrika **71**, 599–607.

P.A. Samuelson [1965], *Proof that Properly Anticipated Prices Fluctuate Randomly*, Ind. Manage. Rev. **6**, 41–49.

J.-D. Saphores, L. Khalaf and D. Pelletier [2002], *On Jumps and Arch Effects in Natural Resource Prices: An Application to Pacific Northwest Stumpage Prices*, Amer. J. Agr. Econ. **84**, 387–400.

J. Schnute [1987], *Data Uncertainty, Model Ambiguity, and Model Identification*, Natural Res. Modeling **2**, 159–212.

G. Schwarz [1978], *Estimating the Dimension of a Model*, Ann. Stat. **6**, 461–464.

G.W. Schwert [1987], *Effect of Model Specification on Tests for Unit Roots in Macroeconomic Data*, J. Mon. Econ. **20**, 73–103.

G.W. Schwert [1989], *Tests for Unit Roots: A Monte Carlo Investigation*, J. Bus. Econ. Stat. **7**, 147–160.

R.J. Shiller and P. Perron [1985], *Testing the Random Walk Hypothesis: Power versus Frequency of Observation*, Econ. Letters **18**, 381–386.

A.M. Taylor [2001], *Potential Pitfalls for the Purchasing-Power-Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price*, Econometrica **69**, 473–498.

T.A. Thomson [1992], *Optimal Forest Rotation when Stumpage Prices Follow a Diffusion Process*, Land Econ. **68**, 329–342.

G.C. Tiao [1972], *Asymptotic Behavior of Temporal Aggregates of Time Series*, Biometrika **59**, 525–531.

United States Dept. of Commerce [2002a], CPI-*All Urban Consumers* (*Current Series*). Data accessed by authors August 15, 2002. Available at http://www.bls.gov/data/.

United States Dept. of Commerce [2002b], *Producer Price Index Commodity Data*. Data accessed by authors August 15, 2002. Available at http://www.bls.gov/data/.

C.L. Washburn and C.S. Binkley [1990], *Informational Efficiency of Markets for Stumpage*, Amer. J. Agr. Econ. **72**, 394–405.

A.A. Weiss [1984], *Systematic Sampling and Temporal Aggregation in Time Series Models*, J. Econometrics **26**, 271–281.

J.C. Williams and B.D. Wright [1991], *Storage and Commodity Markets*, Cambridge Univ. Press, Cambridge, U.K.

H. Working [1960], *Note on the Correlation of First Differences of Averages in a Random Chain*, Econometrica **28**, 916–918.

R. Yin and D.H. Newman [1995a], *A Note on the Tree-Cutting Problem in a Stochastic Environment*, J. For. Econ. **1**, 181–190.

R. Yin and D.H. Newman [1995b], *Optimal Timber Rotations with Evolving Prices and Costs Revisited*, For. Sci. **41**, 477–490.

R. Yin and D.H. Newman [1996], *The Effect of Catastrophic Risk on Forest Investment Decisions*, J. Environ. Econ. Manage. **31**, 186–197.

R. Yin and D.H. Newman [1997], *When to Cut a Stand of Trees?*, Natural Res. Modeling **10**, 251–261.

R. Yin, D.H. Newman and J. Siry [2002], *Testing for Market Integration among Southern Pine Regions*, J. For. Econ. **8**, 151–166.

A. Yoshimoto and I. Shoji [2002], *Comparative Analysis of Stochastic Models for Financial Uncertainty in Forest Management*, For. Sci. **48**, 755–766.