Received: 30 September 2016 | Accepted: 4 February 2018
DOI: 10.1002/fut.21912

RESEARCH ARTICLE

WILEY THE JOURNAL OF FUTURES MARKETS

The impact of data frequency on market efficiency tests of commodity futures prices

Xuedong Wu | Jeffrey H. Dorfman | Berna Karali

The University of Georgia, Athens, Georgia

Correspondence
Jeffrey H. Dorfman, University of Georgia, 312 Conner Hall, Athens, GA 30602.
Email: jdorfman@uga.edu

We investigate the impacts of sampling frequency and model specification uncertainty on the outcome of unit root tests, commonly employed as market efficiency tests, using a new, robust Bayesian test on seven commodity futures prices at three different sample frequencies (daily, weekly, and monthly). Using Bayesian model averaging to account for different possible mean and error variance specifications, we show that sample frequency does affect the unit root test results: the higher the frequency, the higher the support for stationarity. We further show that not accounting for model specification uncertainty can produce unit root test results that are not robust.

KEYWORDS
Bayesian model averaging, commodity futures, GARCH, model uncertainty, stationarity, unit root tests

JEL CLASSIFICATION
C11, C58, Q11

1 | INTRODUCTION

When analyzing asset prices, the time series properties of the data have a very important economic meaning. If an asset price series follows a unit root, then the market is deemed efficient in the sense that profitable predictions are unlikely to be possible. Alternative tests of market efficiency are many, for example, based on cointegration. Researchers have tested whether spot and futures prices are cointegrated to investigate price discovery and market efficiency in a variety of markets. Examples include Bessler and Covey (1991), Chowdhury (1991), Lai and Lai (1991), Fortenbery and Zapata (1993), Schwarz and Szakmary (1994), Chow (1998), and Yang, Bessler, and Leatham (2001). However, such cointegration tests are based on nonstationarity of price series, and therefore require dependable tests of a unit root in order to proceed. Thus, for many empirical studies of asset prices, model specification depends crucially on whether such market efficiency is assumed; for two recent examples in this *Journal*, see Tong, Wang, and Yang (2016) and Fan, Li, and Park (2016). For these reasons, economists need a reliable, accurate, and statistically efficient manner of testing such series for nonstationarity. Complicating the matter is the fact that existing tests have low statistical power, which has led many researchers to seek larger samples of data to test.

To obtain more data one has two choices: a longer time span of data or a higher frequency of sampling within the same time span. Using a longer time series (in calendar terms) is commonly believed to provide more information related to the stationarity, thus leading to a more reliable testing result. Using higher frequency data while maintaining the same time span is generally believed not to provide much additional information since intuitively, stationarity requires a series to pass its mean regularly within the test sample, and increasing the sampling frequency may not change this mean reversion within the sample (Boswijk & Klaassen, 2012). However, this is not necessarily the case if the low frequency data is constructed by systematic sampling, that is, skipping certain intermediate observations from a high frequency process, because systematic sampling at a lower frequency can reduce observed mean reversions as well as impacting sample moments such as the mean and variance. This is worrisome because systematic sampling is exactly the method employed to produce commonly used asset price series such as monthly (weekly, daily) futures prices.

For example, researchers often pick the price of one day each week to construct weekly data from daily data. Choi (1992) demonstrated by simulation that this kind of data aggregation will lower the power of augmented Dickey–Fuller (ADF) and Phillips–Perron (PP) tests, although Chambers (2004) showed that this is a finite sample effect and asymptotically it is still possible to consistently test for a unit root when sampling frequency varies.

Recently, Boswijk and Klaassen (2012) proved that the effects of systematic sampling on unit root testing is not negligible, when a high-frequency sample has volatility clustering with fat-tailed innovations, the famous autoregressive conditional heteroskedasticity which is typical of financial market data. Using simulated data they showed that likelihood ratio-based tests were more powerful than the traditional ADF test on data processes displaying the aforementioned behavior characteristics. This leads to a second way to test for unit roots more accurately: improve the testing method.

Unit root tests which focus on non-normal errors can be found in Lucas (1995) and Rothenberg and Stock (1997). Seo (1999), Boswijk (2001), and Ling and Li (2003) present unit root tests when the time series being tested has Gaussian GARCH (Generalized Autoregressive Conditional Heteroskedasticity) stochastics. Although these tests have increased power when testing financial data, a common trait for the existing testing methods is that they all require some specific model specification assumption, either in terms of mean functional form (e.g., the ADF test requires the number of autoregressive lags to be specified) or the error term distribution (ARCH–Autoregressive Conditional Heteroscedasticity–, GARCH, normal, *t*, etc.). This is a non-trivial issue, because while the question of interest is the presence or absence of a unit root, not the specification of the time series model, Moral-Benito (2015) showed that inappropriate model assumptions will produce erroneous unit root test results—and that is certainly something with which economists are concerned.

In this paper, we present an improved method of testing for market efficiency using a Bayesian unit root test which averages over multiple possible specifications of the underlying time series model, thus providing robust test results. Our method is demonstrated using futures price data on seven futures prices: five agricultural commodities (corn, soybean, cotton, live cattle, and lean hog) and two industrial (crude oil and silver), all of which display typical financial data characteristics. We first show that systematic sampling can have significant effects on the results of unit root testing using three different frequency samples (daily, weekly, and monthly) using the traditional testing methods. Then, more importantly, we test the stationarity of the seven series by averaging 24 diverse models using our Bayesian model averaging (BMA) unit root test and compare the results with traditional unit root test results to show the performance of the BMA method, as well as its ability to handle the model specification issue.

The rest of the paper is organized as follows. The next section introduces the robust Bayesian unit root test and the specific models averaged for this application. Next, the data used in the analysis are introduced, followed by the priors, posterior distributions, and the sampling methods. We then present and discuss the results of the tests, followed by conclusions.

2 | A ROBUST BAYESIAN UNIT ROOT TEST ACCOUNTING FOR MODEL UNCERTAINTY

Because traditional unit root tests have very low power and use a null hypothesis of a unit root, they are biased in favor of finding market efficiency (i.e., nonstationarity). Further, when testing financial asset price series it is important to incorporate the distributional features commonly found in such time series, such as fatter tails and conditional heteroskedasticity. To this end, we introduce a new testing approach which relies on Bayesian model averaging to produce probabilities in favor of and against a unit root while accounting for uncertainty over both model and error specifications.

2.1 | Model parameterization

Although the error specification is a key for financial data such as the futures prices considered in this paper, it is also of equal importance to specify the mean function accurately as model uncertainty arises in both areas. In this paper, we adopt a standard autoregressive model with maximum lag p as the mean function. This can be written as

$$A(L)x_t = \varepsilon_t \tag{1}$$

where

$$A(L) = 1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_p L^p. \tag{2}$$

In equation (2), L is the lag operator such that $Lx_t = x_{t-1}$, and in general $L^k x_t = x_{t-k}$. The stochastic term ε follows some distribution, as yet unspecified, with mean 0 and possibly time-varying variance $\sigma_t^2 : f\left(0,\sigma_t^2\right)$. Given the mean function specification above, we consider models with 1 to 6 autoregressive lags in our tests for futures price (non) stationarity.[1]

Because financial time series usually contain stochastic terms with fat-tails and volatility clustering (Boswijk & Klaassen, 2012), both of which can affect the results of unit root tests (Choi & Chung, 1995), we need to incorporate such features in the set of models considered within our test. Widely used models to consider include the ARCH model (Engle, 1982) and the GARCH model (Bollerslev, 1986). After these two seminal papers many related variations of these specifications have been developed, such as NGARCH (Engle & Ng, 1993) and EGARCH (Nelson, 1991). For the seven futures prices tested here, we decided on the following four error specifications.

2.1.1 | GARCH (1,1) with student's *t* distribution (GARCH-T)

The Student's *t* distribution with small degrees of freedom can capture the heavy-tails characteristic of the analyzed data while the clustered volatility can be modeled by a GARCH specification, which has the following form:

$$\begin{gathered} x_t = E(x_t|x_{t-1},\ldots) + \varepsilon_t \\ \varepsilon_t = Z_t\sqrt{h_t} \\ h_t = \alpha_0 + a_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1} \end{gathered} \tag{3}$$

and z_t follows a Student's *t* distribution. Model parsimony is one concern here because while the goal is to yield more accurate unit root test results, the procedure must be something practitioners can implement in terms of complexity and computing time. Further justifying the model specification choices that follow, Hansen and Lunde (2005) compared 330 different ARCH-type models in terms of their ability to describe the conditional variance and concluded that GARCH (1,1) outperformed other higher level GARCH models.

2.1.2 | GARCH (1,1) with standard normal distribution (GARCH-N)

This is exactly as above except z_t follows a normal distribution instead of a *t* distribution.

2.1.3 | ARCH (1) with Student's *t* distribution (ARCH-T)

The ARCH model is another common model used to describe highly volatile error behavior and matches the GARCH-T model except that

$$h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2. \tag{4}$$

[1]It should be noted that if this mean specification misses some minor nonlinearities, the GARCH error specification will help to capture that behavior. Also, while we do not test for nonlinear means directly, mean shifts (such as seasonal effects or time-to-maturity effects) or other structural breaks, adding such features will uniformly lower the probability of a unit root by accounting for at least some share of mean non-reverting behavior. Thus, it is possible we overstate the posterior probability of market efficiency here; however, because we are specifically interested in how those probabilities change with sampling frequency, we believe a little simplicity is in order here. Such specification features should not impact the changes in probabilities of market efficiency as the sampling frequency changes. While different first moment specifications will result in different second moment specifications (the GARCH parameters will change), the goal of this paper is studying the results of unit root tests as data frequency changes, not identifying GARCH effects perfectly. Further, it is important to remember that in a Bayesian context, results are exact conditional on the model, so model misspecification is not a problem the way it is in frequentist estimation.

2.1.4 | Autoregressive mean with Student's *t* distribution (AR-T)

In this case, we do not add any conditional heteroskedasticity to the variance specification and focus only on modeling the heavy-tailed behavior. The mean function is the autoregressive model shared by all the models considered above and the error term, ϵ_t, follows a *t* distribution.

Each of the above four error specifications is then paired with six different autoregressive mean models (with 1 to 6 lags), yielding 24 models in total that will be averaged to test one data series for a unit root (and, thereby, market efficiency). Now we turn to the mechanics of the Bayesian model averaging unit root test.

2.2 | The Bayesian unit root test for a single model

To test for a unit root, we need the distribution of the roots of the characteristic equation controlling the dynamic properties of the time series being tested. For autoregressive models such as we study here, this can be done through examination of the following matrix:

$$A = \begin{bmatrix} \rho_1 & 1 & 0 & 0 & \cdots & 0 \\ \rho_2 & 0 & 1 & 0 & \cdots & 0 \\ \rho_3 & 0 & 0 & \ddots & & \vdots \\ \vdots & 0 & 0 & \cdots & 0 & 1 \\ \rho_p & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{5}$$

The first column is composed of the autoregressive parameters from the model being tested. The roots, or eigenvalues of this matrix, determine the behavior of the modeled process. If any eigenvalues of $\boldsymbol{A}$ have moduli equal to (greater than) one, the time series defined by these coefficients has a unit root (thus, explosive) and is nonstationary. Let $\phi_i = \|\lambda_i\|$ be the modulus of the ith eigenvalue of A sorted by their magnitude where $\|\cdot\|$ stands for Euclidian distance. Given this definition, the statistical hypothesis of a unit root can be expressed as:

$$H_0 : \phi_1 \geq 1 \text{ vs. } H_1 : \phi_1 < 1. \tag{6}$$

Unlike the classical unit root test method which focuses on deriving the asymptotic distribution of the test statistic conditional on the null hypothesis, in the Bayesian framework both the null and alternative hypotheses are treated equally and test decisions are based on the relative posterior probabilities in support of each hypothesis.

Define p_0 and p_1 as posterior probabilities of support for the null (Θ_0) and alternative (Θ_1) hypotheses, while π_0 and π_1 are prior probabilities for the same hypotheses implicitly defined by the prior distributions of the roots. Let $\Phi = \left(\phi_1, \phi_2, \ldots, \phi_p\right)'$ be the vector of the moduli of the eigenvalues of the matrix A and let $\boldsymbol{\eta}$ stand for all the other parameters which vary across different model specifications (for example, in GARCH-T, $\boldsymbol{\eta} = (\alpha_0, \alpha_1, \beta_1, v)'$; in GARCH-N, $\boldsymbol{\eta} = (\alpha_0, \alpha_1, \beta_1)'$). We can then denote the prior distributions as $\pi(\boldsymbol{\Phi}) = \pi\left(\phi_1, \phi_2, \ldots, \phi_p\right)$ and $h(\boldsymbol{\eta})$, respectively (the detailed discussions on prior distribution are given in the next section), and the likelihood function can be denoted as $f(x; \boldsymbol{\eta}, \boldsymbol{\Phi})$. Given these definitions, the marginal posterior distribution of the dominant root would be:

$$p(\phi_1|x) = \int_{\phi_2} \cdots \int_{\phi_p} \int_{\eta} f(x; \boldsymbol{\eta}, \boldsymbol{\Phi}) \pi(\boldsymbol{\Phi}) h(\boldsymbol{\eta}) d\boldsymbol{\eta} d\phi_2 \cdots d\phi_p. \tag{7}$$

Because characteristics of the marginal posterior density in (7) cannot be analytically computed, numerical approximations are required to compute the probabilities for each hypothesis. Using numerical methods (to be discussed later) samples can be drawn from this posterior distribution. Define an indicator function:

$$D\left(\phi_1^{(i)}\right) = \begin{cases} 1 \; \phi_1 < 1 \text{ in the } i\text{th draw(support } H_1), \\ 0 \qquad\qquad \text{otherwise (support } H_0). \end{cases} \tag{8}$$

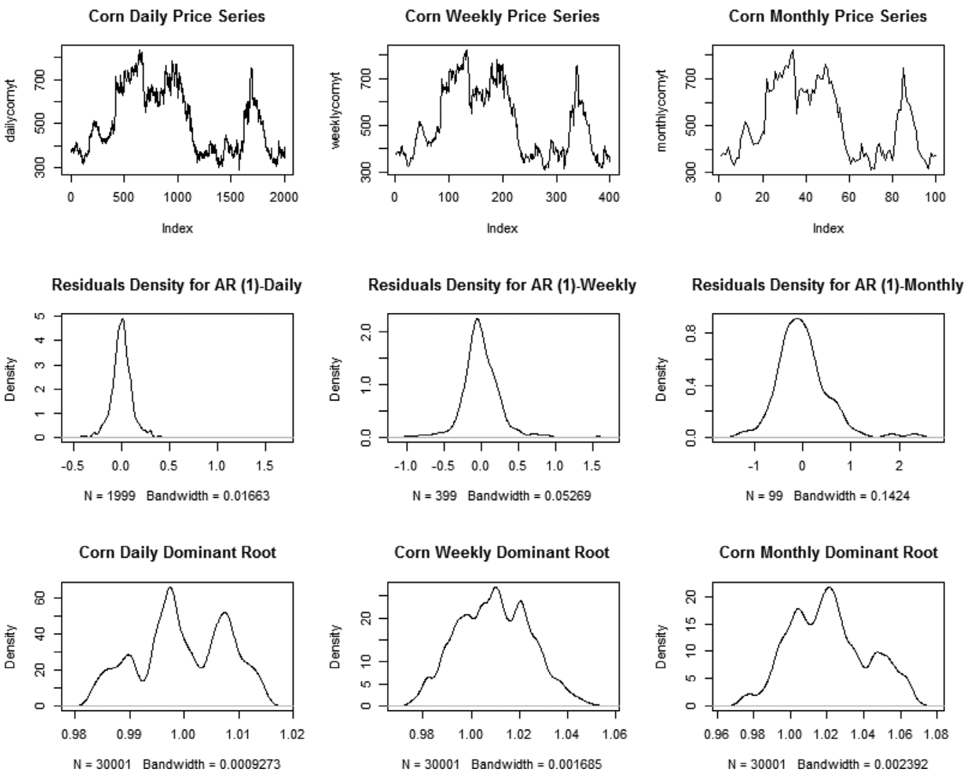

FIGURE 1 Corn futures

The posterior probability in support H_1 (stationarity) is therefore given by:

$$K_1 = p(H_1|\boldsymbol{y}) = \frac{\Sigma_{i=1}^{B}\ D\left(\phi_1^{(i)}\right) p\left(\phi_1^{(i)}|\boldsymbol{\eta}^{(i)},\phi_2^{(i)},\ldots,\phi_p^{(i)};y\right)}{\sum_{i=1}^{B} p\left(\phi_1^{(i)}|\boldsymbol{\eta}^{(i)},\phi_2^{(i)},\ldots,\phi_p^{(i)};\boldsymbol{y}\right)} \tag{9}$$

where B is the number of draws used in the construction of the numerical approximation to the marginal posterior distribution.[2] Defining K_0 similarly gives:

$$K_0 = p(H_0|\boldsymbol{y}) = \frac{\sum_{i=1}^{B}\left[1 - D\left(\phi_1^{(i)}\right)\right] p\left(\phi_1^{(i)}|\boldsymbol{\eta}^{(i)},\phi_2^{(i)},\ldots,\phi_p^{(i)};\boldsymbol{y}\right)}{\sum_{i=1}^{B} p\left(\phi_1^{(i)}|\boldsymbol{\eta}^{(i)},\phi_2^{(i)},\ldots,\phi_p^{(i)};\boldsymbol{y}\right)}. \tag{10}$$

These probabilities in support of each hypothesis are then used to construct the posterior odds ratio in favor of a unit root (i.e., nonstationarity):

$$K_{01} = \frac{\sum_{i=1}^{B}\left[1 - D\left(\phi_1^{(i)}\right)\right] p\left(\phi_1^{(i)}|\boldsymbol{\eta}^{(i)},\phi_2^{(i)},\ldots,\phi_p^{(i)};\boldsymbol{y}\right)}{\sum_{i=1}^{B} D\left(\phi_1^{(i)}\right) p\left(\phi_1^{(i)}|\boldsymbol{\eta}^{(i)},\phi_2^{(i)},\ldots,\phi_p^{(i)};\boldsymbol{y}\right)} = \frac{K_0}{K_1}. \tag{11}$$

When combined with a loss function that summarizes the penalties for incorrect decisions in choosing which hypothesis to support post-data analysis, the posterior odds ratio provides a test result (Berger, 1985).

[2]Note that if the draws are generated by a process which yields an unweighted approximation to the posterior distribution, then the posterior probabilities can be computed simply as the proportion of draws in each region. Because the draws here are not drawn from the full posterior, weighting is needed to get accurate estimates of the posterior probabilities.

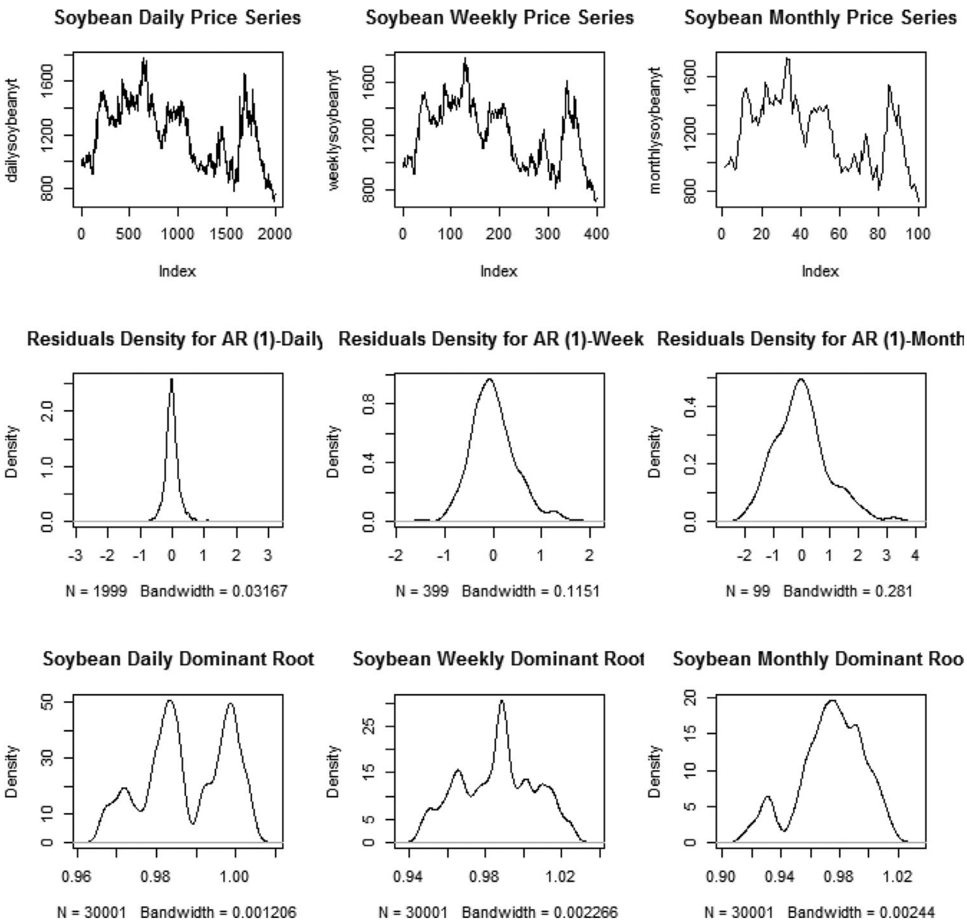

FIGURE 2 Soybean futures

For example, if the loss from incorrect decisions is symmetric (it is equally bad to choose stationarity or nonstationarity when the truth is the opposite), then the decision will be to choose the hypothesis that has greater than 50 percent posterior support (i.e., choose nonstationarity if $K_{01} > 1$). If the loss from test decisions is asymmetric (say, it is worse to declare a series stationary when it is actually nonstationary), then the threshold of support necessary to select nonstationarity gets larger (smaller) as the loss from incorrectly choosing nonstationarity gets smaller (larger). The general decision rule is that if the loss from supporting stationarity when that is false is given by c and the loss from believing in a unit root when the series is truly stationary equals d, then one would support stationarity if and only if $K_{01} < d/c$.

2.3 | Adding model specification uncertainty

The above discussion explains how to infer the stationarity properties for a series using a numerical Bayesian unit root test conditional on a single, known model. To incorporate model specification uncertainty one simply repeats the above procedure for all models under consideration and then averages the results from all candidate models to reach a single, robust conclusion.

Suppose there are k possible candidate models which can differ in various aspects (in this paper, models vary in the autoregressive model order, the specification of the variance terms, and the distribution of the errors). We assign each model a prior probability, $p(M_l)$, which leads to the posterior probability for model M_l as:

$$p(M_l|x) = \frac{p(x|M_l)p(M_l)}{\sum_{j=1}^{k} p(x|M_j)p(M_j)} \tag{12}$$

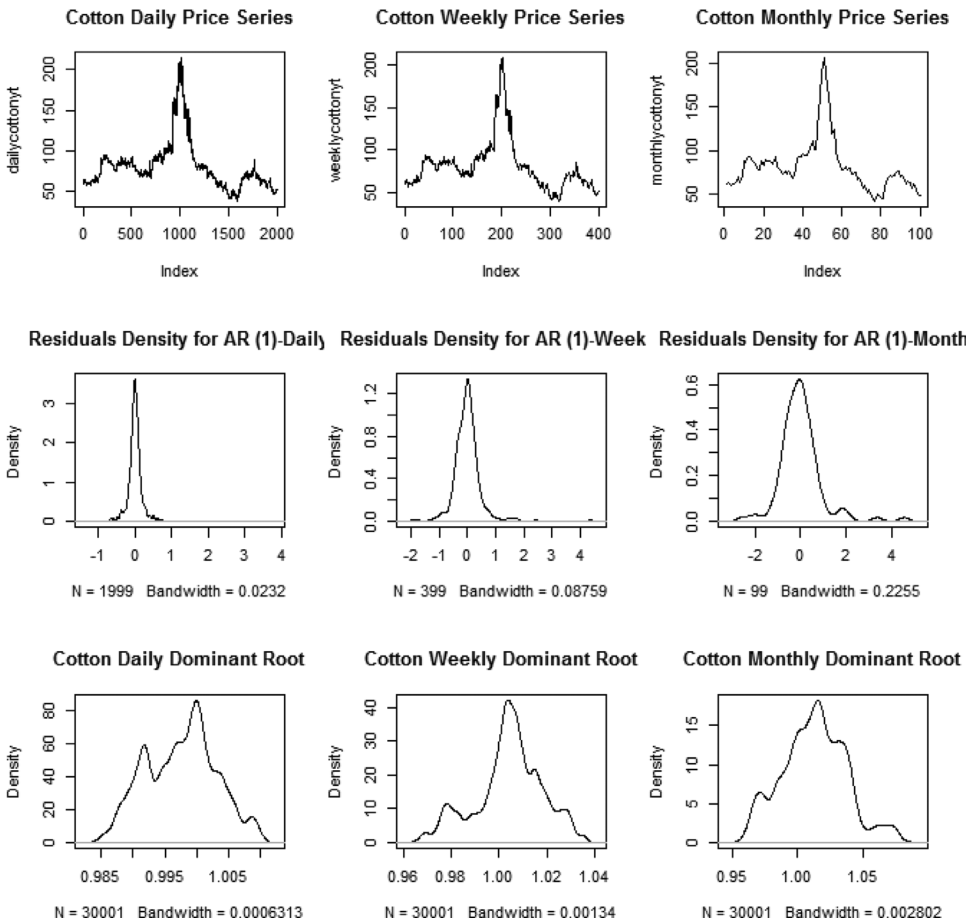

FIGURE 3 Cotton futures

where $p(x|M_l)$ is the marginal likelihood value for model M_l which is given by

$$p(x|M_l) = \int_S p(x|\mathbf{\Omega}_l, M_l)p(\mathbf{\Omega}_l|M_l)d\mathbf{\Omega}_l. \tag{13}$$

In the above equation, $\mathbf{\Omega}_l$ is all the parameters in the model M_l, $p(\mathbf{\Omega}_l|M_l)$ is the prior distribution of parameters $\mathbf{\Omega}$ under model M_l, and S is the support of $\mathbf{\Omega}$. Note that the marginal likelihood is the average likelihood function value over the supported parameter space, not the maximal value. Then the final comprehensive probability of a possible unit root across the entire model space is the weighted average of the results from the individual models:

$$p(\phi_1 \geq 1|x) = \sum_{l=1}^{k} p(\phi_1 \geq 1|M_l, x)p(M_l|x). \tag{14}$$

Decisions based on this robust posterior probability would follow in the identical manner to that discussed above in the single model case.

3 | DATA

In this paper, seven commodity futures price series are used to test and compare the impact of data frequency on unit root test results: corn, soybean, cotton, live cattle, lean hog, crude oil, and silver. To test for any sampling frequency effect, three different frequencies are used for each series: daily, weekly, and monthly. The high-frequency sample is the daily settlement prices from the Chicago Mercantile Exchange Group (corn, soybean, live cattle, lean hog, crude oil, and silver) and the Intercontinental

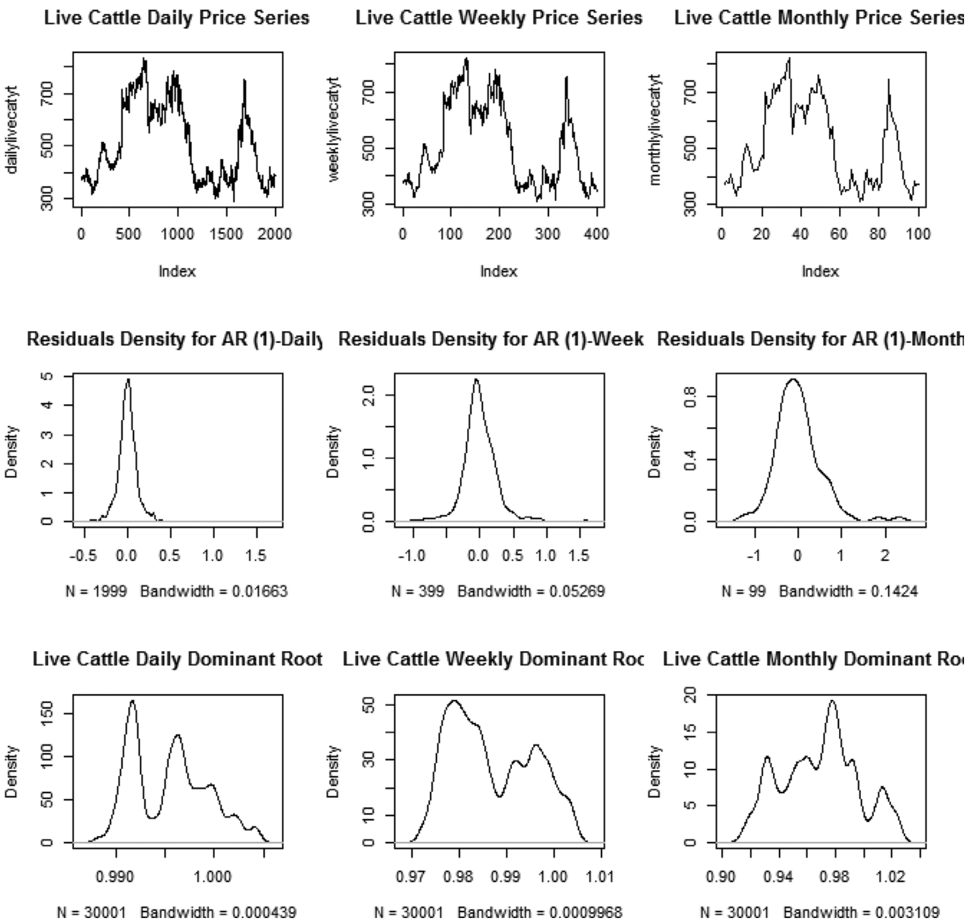

FIGURE 4 Live cattle futures

Exchange (cotton).[3] Each daily data sample size is 2,000, spanning from March 2007 to March 2015. Data are downloaded from Quandl.com and consist of prices on the nearby contract, rolling over on the day before each contract expires. The low-frequency sample is constructed from the daily data by what is usually referred to as systematic sampling. Denoting the daily sample as Y_t, lower frequency data sets are constructed by skipping observations in a set pattern (Boswijk & Klaassen, 2012) as

$$\boldsymbol{Y}_j^* = \boldsymbol{Y}_{\mathrm{mj}},\ j = 0,\ldots,n^* = \frac{n}{m} \tag{15}$$

where for weekly data we take $m = 5$ and for monthly, $m = 20$. Given the vagaries of the trading calendar these lower-frequency data sets are not precisely weekly or monthly as we have chosen a precise change in sampling frequency over calendar precision. Since the daily data sample has 2,000 observations, the constructed weekly sample has 400 and the monthly data sample has 100 observations. The first rows of Figures 1–7 show the time series plot of each commodity price with different sample frequencies.

4 | TESTING FOR UNIT ROOTS IN COMMODITIES FUTURES DATA AT DIFFERENT FREQUENCIES

4.1 | Priors, likelihood, and posterior distributions

The choice of priors for the parameters is always an important issue in Bayesian econometric analysis since these prior beliefs have some influence on final estimates. Considerable research has been devoted to studying the properties and suitability of

[3]The continuous futures data are available for downloading from the Open Financial Data Project in Quandl: https://www.quandl.com. The front month contract price is used to build the continuous data.

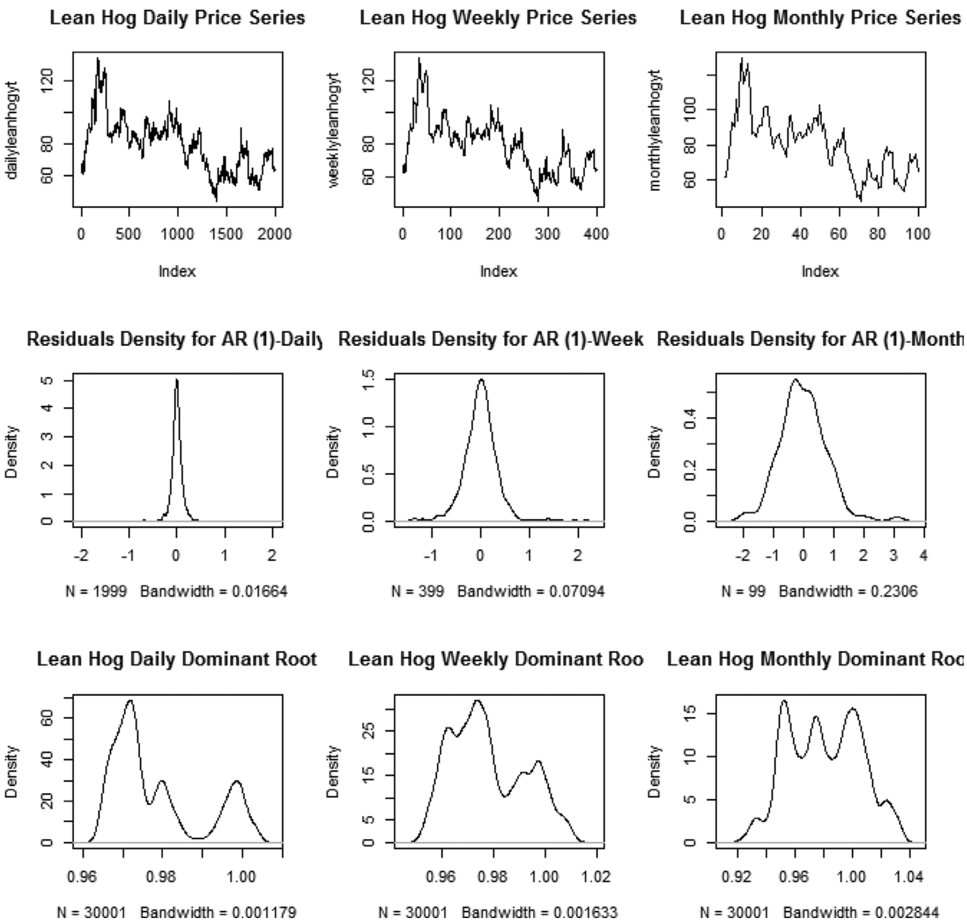

FIGURE 5 Lean hog futures

different priors. Berger and Yang (1994) compared various approaches to the development of a uninformative prior for the AR(1) model, and recently Griffiths and Xia (2012) performed a Monte Carlo experiment as well as an application to real world data to examine the effects of choices of different priors on Bayesian unit root test outcomes. In light of this continuing debate, we do not spend much time on judging the priors since this is not the main goal of this research and acknowledge that it is hard or impossible to find a perfect prior that works well under all circumstances due to the fact that it may well depend on the data and topic (pre-knowledge). Instead we follow an approach similar to Dorfman (1993) and assign independent Beta priors to the moduli of the characteristic roots while maintaining diffuse, relatively uninformative priors on other parameters. Meanwhile, a slightly explosive situation is allowed for the value of the dominant root in order to account for possible upward bias due to sampling error (Dorfman, 1993). Specifically, the priors on all the roots are defined as:

$$\begin{gathered}\pi(\Phi) = \pi\left(\phi_1,\phi_2,\ldots,\phi_p\right) = \pi_1(\phi_1)\pi_2(\phi_2)\ldots\pi_p\left(\phi_p\right)\\ \pi_1(\phi_1-\tau)\sim\text{Beta}(30,2);\ \pi_2(\phi_2)=\pi_3(\phi_3)=\cdots=\pi_p\left(\phi_p\right)\sim\text{Beta}(1.1,1.1).\end{gathered} \tag{16}$$

This prior is quite informative on the moduli of the dominant root since it is severely skewed to the right, and ranges only from τ to $1+\tau$. The priors on all the other roots are quite weakly informative with very flat, round curves ranging over [0, 1]. The parameter τ controls the degree to which the dominant root is allowed to be explosive (above unity) and it is also assigned a prior in our Bayesian framework. Following the suggestion of a hierarchical approach by Griffiths and Xia (2012), we assign an exponential distribution with mean $E(\tau)=\kappa^{-1}$ as a prior for τ:

$$f(\tau|\kappa) = \kappa\exp(-\kappa\tau) \tag{17}$$

where κ is chosen such that $p(\tau > 0.15) = 0.05$.

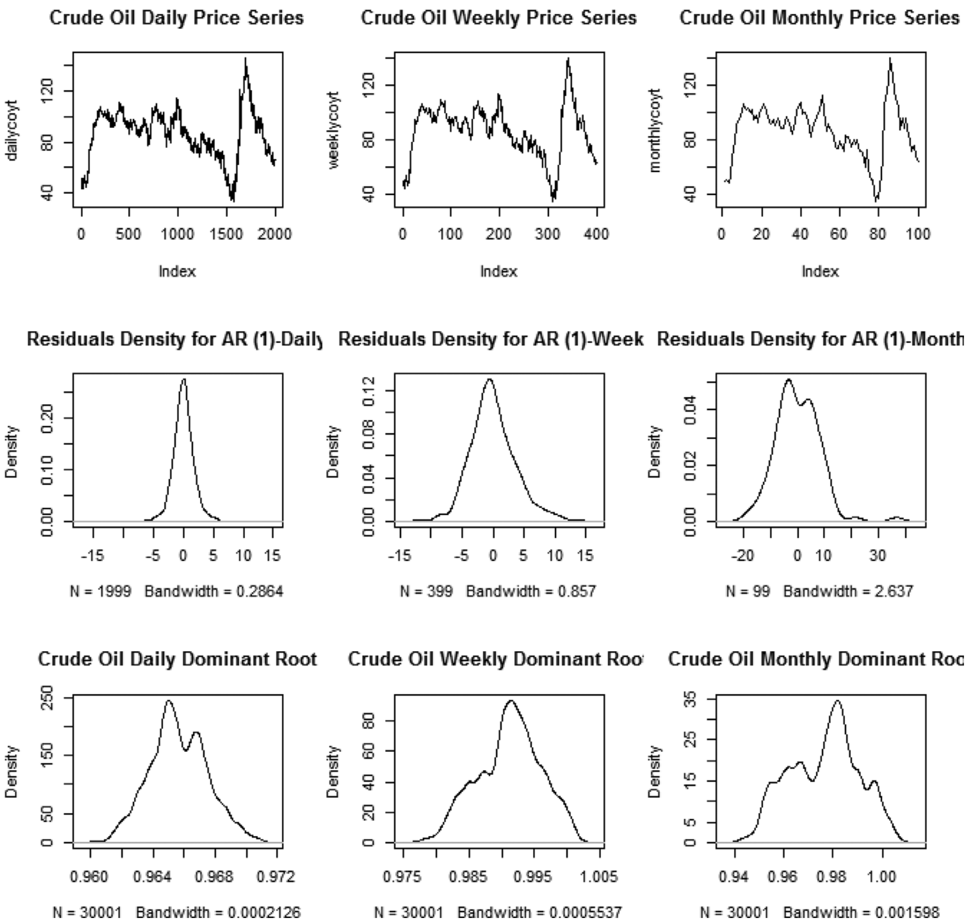

FIGURE 6 Crude oil futures

For the parameters in the conditional variance specification, the priors can be relatively uninformative as long as they satisfy positivity in all circumstances. The common past practice has been to restrict all the parameters to be non-negative and the constant to be positive. Instead of following this tradition, we follow Dorfman and Park (2011) and impose positivity directly on the conditional volatility. By doing so we allow negative individual parameters in the ARCH and GARCH specifications, which widens the supported parameter space, increasing the opportunity of achieving accurate estimates of all parameters in our system. Thus, all these priors are set to have a normal distribution with zero mean and are mutually independent. The prior distributions of the parameters in the volatility models can be represented as:

GARCH (1,1): $p(\alpha_0,\alpha_1,\beta_1) = I(h_t)N_{\alpha_0}(0,3)N_{\alpha_1}(0,3)N_{\beta_1}(0,3)$,

ARCH (1): $p(\alpha_0,\alpha_1) = I(h_t)N_{\alpha_0}(0,3)N_{\alpha_1}(0,3)$.

The $I(h_t)$ here is an indicator function used to control the positivity of volatility which equals one if a sampled parameter vector provides for a positive conditional variance in all time periods and equals zero otherwise.

Another parameter to consider is the degrees of freedom ν of the Student's t distribution. To model the fat-tail behavior well, the large part of the density should be kept around a small value, yet some degree of variance is needed to incorporate more possible parameter values under consideration. To balance these two ends, we assign a truncated exponential distribution as the prior for ν which has the following density form:

$$p(v|\lambda,\delta) = \lambda \exp(-\lambda(v-\delta))I(v>\delta). \tag{18}$$

The values of λ and δ can be set to control the shape of the exponential distribution density. The indicator function and parameter δ are used to make sure the generated t distribution is well defined (hence, $\delta \geq 2$); it can also be used to approximate a normal distribution if the value of δ is set to be relatively large. In our application we set $\lambda = 0.4$ and $\delta = 2$. This is a fairly uninformative prior while still proper (having finite moments).

Now we can write the full form of the joint posterior distribution for each model as follows.

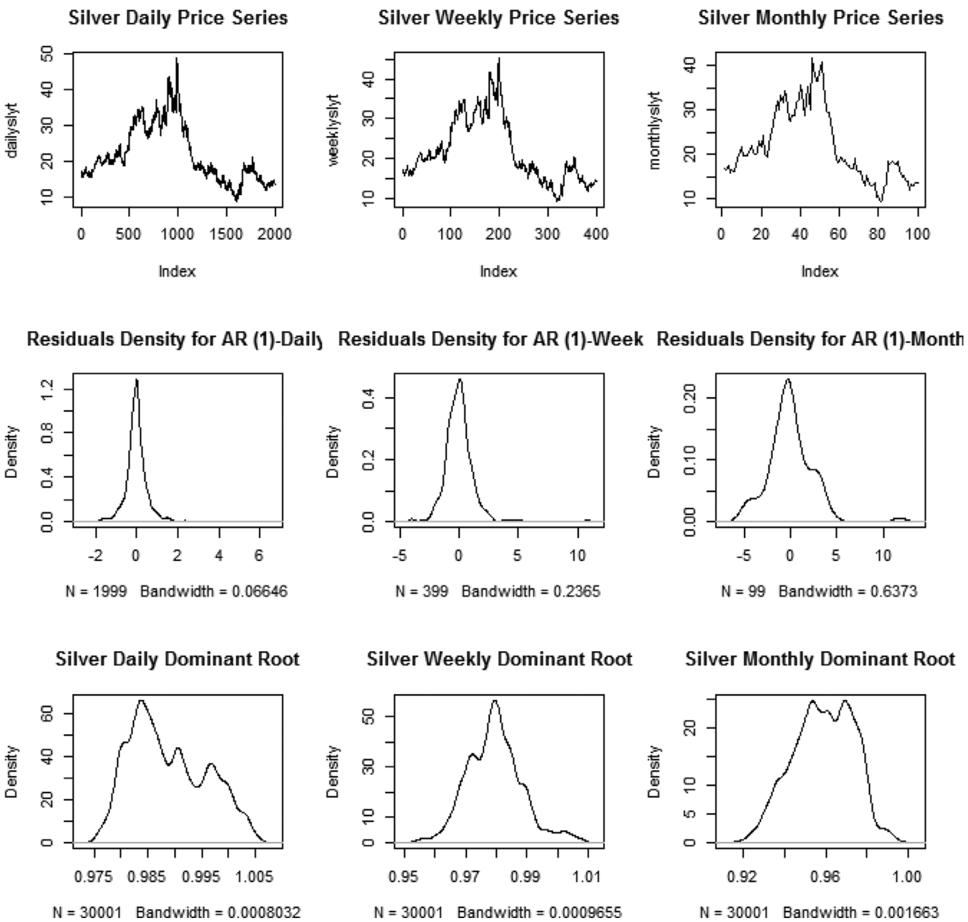

FIGURE 7 Silver futures

4.1.1 | GARCH (1,1) with Student's *t* distribution (GARCH-T)

$$p(\boldsymbol{\Phi},\alpha_0,\alpha_1,\beta_1,v,\tau|\boldsymbol{x}) \propto L(\boldsymbol{x}|\boldsymbol{\rho},\alpha_0,\alpha_1,\beta_1,\nu) \times \pi(\boldsymbol{\Phi}) \times p(\alpha_0,\alpha_1,\beta_1) \times p(\nu|\lambda,\delta) \times f(\tau|\kappa) \tag{19}$$

The term $L(\boldsymbol{x}|\boldsymbol{\rho},\alpha_0,\alpha_1,\beta_1,v)$ is the likelihood function of a Student's *t* distribution which has the following form (assuming sample size T):

$$L(x|\rho,a_0,a_1,\beta_1,v) = \left[\frac{\Gamma\left[\frac{v+1}{2}\right]}{\sqrt{\pi(v-2)\Gamma\left(\frac{v}{2}\right)}}\right]^T \exp\left\{-\frac{1}{2}\prod_{t=1}^{T} h_t - \prod_{t=1}^{T}\left[1+\frac{(x_t - x'\rho)^2}{h_t(v-2)}\right]^{\frac{v+1}{2}}\right\} \tag{20}$$

where $h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\ h_{t-1}$ and all the other terms in were defined previously.

4.1.2 | GARCH (1,1) with standard normal distribution (GARCH-N)

$$p(\boldsymbol{\Phi},\alpha_0,\alpha_1,\beta_1,\tau|\boldsymbol{x}) \propto L(\boldsymbol{x}|\boldsymbol{\rho},\alpha_0,\alpha_1,\beta_1) \times \pi(\boldsymbol{\Phi}) \times p(\alpha_0,\alpha_1,\beta_1) \times f(\tau|\kappa) \tag{21}$$

Here, the term $L(\boldsymbol{x}|\boldsymbol{\rho},\alpha_0,\alpha_1,\beta_1)$ is the likelihood function of normal distribution which has the following form:

$$L(x|\rho,a_0,a_1,\beta_1) = (2\pi)^{-T/2}\ \prod_{t=1}^{T} h_t^{-1/2} \exp\left[-\frac{1}{2h_t}\left(\prod_{t=1}^{T}(x_t - x'\rho)^2\right)\right] \tag{22}$$

where $h_t = a_0 + a_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$. The posterior densities of the other two models (ARCH-T and AR-T) are omitted here since they have similar formulas to those above with only minor parameter changes.

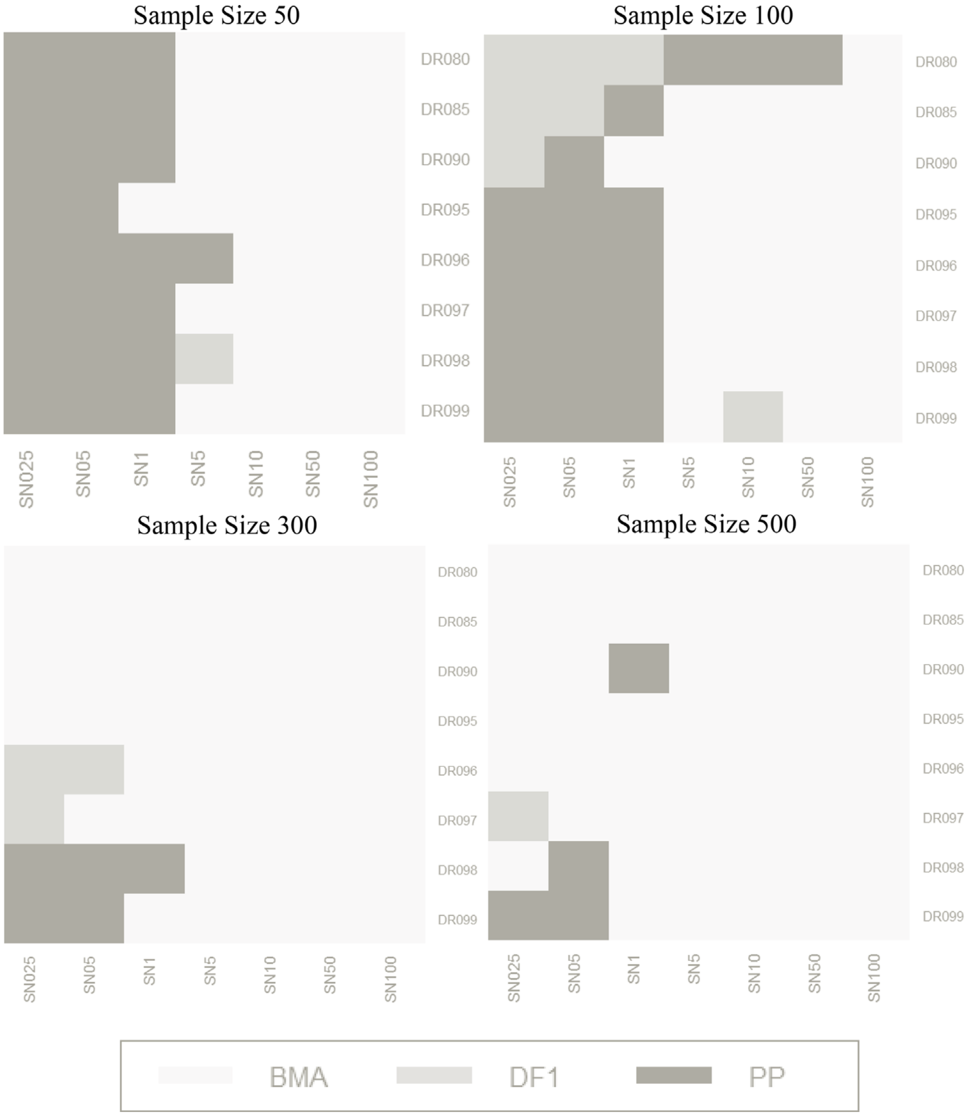

FIGURE 8 Best testing method by sample size, signal to noise ratio, and dominant root

4.2 | Analyzing the posterior

Having established the models to be considered, the next step to estimating the probability of stationarity for the series being studied is to find the marginal posterior of the dominant root using equation (7) by integrating out all the nuisance parameter for each model, then averaging the results of all 24 models considered. As is common, in our case the first task is impossible to do analytically since the use of several non-conjugate prior densities results in posterior distributions that do not have a standard distributional form. Furthermore, the Beta prior on the moduli of the eigenvalues of the matrix $\boldsymbol{A}$ in equation (5), while logically straightforward, would require a Jacobian in order to allow an analytical calculation of likelihood function. Adopting numerical methods allows us to generate an empirical sample from the posterior distribution without involving a Jacobian as the value of the prior distribution can be computed after drawing parameters.

To generate a numerical approximation to the integrals that represent the posterior measures of interest, we use Markov Chain Monte Carlo (MCMC) methods, specifically a Gibbs sampler with Metropolis–Hastings steps. This approach utilizes two methods. First, we draw random samples from a series of conditional distributions of subsets of parameters where such subsets can be chosen in order to provide distributional forms which are simple to use (the Gibbs sampler part). Then, in the remaining cases we produce conditional draws from alternative densities and then either accept or reject with a specific probability in order to adjust their sampling distribution to the true density (the Metropolis–Hastings steps). Together, these methods produce a Markov chain which converges to the joint posterior distribution of all the parameters and allow us to compute posterior probabilities in favor or against stationarity (Chib, 1995; Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953; Tanner, 1998). In our case, Metropolis–Hastings steps are necessary

TABLE 1 Unit root test results

	(1) BMA	(2) DF 1	(3) DF 2	(4) DF 3	(5) DF 4	(6) DF 5	(7) DF 6	(8) PP
Corn								
Daily	0.503	0.576	0.538	0.475	0.339	0.390	0.315	0.542
Weekly	0.717	0.530	0.535	0.497	0.577	0.648	0.622	0.552
Monthly	0.784	0.576	0.538	0.475	0.338	0.390	0.315	0.519
Soybean								
Daily	0.372	0.271	0.312	0.282	0.332	0.359	0.357	0.298
Weekly	0.406	0.346	0.381	0.319	0.296	0.277	0.256	0.317
Monthly	0.481	0.227	0.098	0.163	0.113	0.0431	0.0264	0.263
Cotton								
Daily	0.498	0.622	0.627	0.628	0.624	0.579	0.647	0.621
Weekly	0.535	0.593	0.651	0.662	0.641	0.550	0.499	0.616
Monthly	0.685	0.489	0.320	0.216	0.468	0.463	0.402	0.519
Live cattle								
Daily	0.222	0.537	0.570	0.545	0.559	0.546	0.553	0.542
Weekly	0.254	0.530	0.535	0.497	0.577	0.648	0.622	0.552
Monthly	0.408	0.576	0.538	0.475	0.338	0.390	0.315	0.519
Lean hog								
Daily	0.141	0.084	0.060	0.010	0.010	0.010	0.010	0.010
Weekly	0.188	0.010	0.010	0.010	0.010	0.010	0.0188	0.010
Monthly	0.396	0.010	0.010	0.010	0.010	0.010	0.0262	0.010
Crude oil								
Daily	0.103	0.324	0.372	0.339	0.287	0.307	0.277	0.319
Weekly	0.0784	0.320	0.208	0.232	0.202	0.116	0.0813	0.290
Monthly	0.0588	0.126	0.010	0.010	0.0127	0.080	0.174	0.131
Silver								
Daily	0.246	0.584	0.608	0.594	0.607	0.595	0.593	0.581
Weekly	0.273	0.599	0.524	0.590	0.610	0.655	0.615	0.614
Monthly	0.251	0.651	0.679	0.560	0.573	0.511	0.530	0.633

Notes: Column (1) shows the probability that the modulus of the dominant root is equal to or greater than 1, $p(\phi_1 \geq 1|x)$. Values equal to or greater than 0.5 indicate evidence of nonstationarity. Columns 2–8 list the p-values from augmented Dickey–Fuller tests with lag length of 1–6 (DF1-DF6) and the Phillips–Perron (PP) test. Usually a p-value >0.1 is taken as evidence of nonstationarity.

because the Beta prior on the roots produces a conditional posterior distribution of the autoregressive parameters in an unknown form. However, once draws from the proposal density (in our case a uniform distribution) are made, the posterior distribution can be computed since it is straightforward to take a vector of proposed S_t parameters, compute the eigenvalues of the $\boldsymbol{A}$ matrix from above, then compute the prior distribution value for those parameters, and finally arrive at the posterior value for that draw of coefficient values.

The Metropolis–Hastings algorithm is very beneficial in dealing with complex distributions, but there is a price for this convenience (Jing, 2010). If the proposal density is not chosen appropriately, the acceptance rate can be either too high or too low, which means that the sample we create may include too much or too little variation leading to insufficient mixing in the chain (not visiting all parts of the posterior distribution). We control the acceptance rate through tuning parameters on the mean and variance of the proposal density, as well as using different sets of starting values, to achieve a rate of acceptance that guarantees an accurate approximation of the posterior density. See Tanner (1998) for more details.

4.3 | Some Monte Carlo results

To demonstrate that the robust Bayesian unit root test presented above performs well, a Monte Carlo experiment was conducted. Data sets were generated using 24 different models that included or excluded trend and drift, that were in levels or in logs, and

TABLE 2 AR (1) residual kurtosis

Commodity	Frequency	Residual Kurtosis
Corn	Daily	23.97
	Weekly	5.32
	Monthly	3.53
Soybean	Daily	26.40
	Weekly	1.03
	Monthly	0.70
Cotton	Daily	60.22
	Weekly	17.80
	Monthly	6.31
Live cattle	Daily	23.97
	Weekly	5.32
	Monthly	3.53
Lean hog	Daily	23.97
	Weekly	5.32
	Monthly	3.53
Crude oil	Daily	6.96
	Weekly	0.99
	Monthly	2.95
Silver	Daily	17.62
	Weekly	15.24
	Monthly	4.70

that had from one to six lags. The dominant root of the data generating process (DGP) was set to values of 0.80, 0.85, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, and 1. The signal to noise (SN) ratio was varied across the range of 0.25, 0.5, 1, 5, 10, 50, and 100. For each combination of dominant root modulus and signal to noise ratio, data sets were generated for sample sizes of 50, 100, 300, and 500 to allow both small and large sample properties to be analyzed. All possible combinations of these settings produce 252 different DGPs to be analyzed. For more stationary DGPs, 200 data sets were analyzed; for the DGPs with a dominant root above 0.95, 500 data sets were tested.

The robust Bayesian unit root test was performed on each data set and compared to the results of Dickey–Fuller and Phillips–Perron tests. Figure 8 shows which of these three tests performed best in different areas of this data space.[4] Note that the Bayesian test performs quite well, using the precise prior we employ in this paper. None of the tests is always the best and all three tests are sometimes the best, but the Bayesian test is clearly admissible in the statistical sense. Inclusion of augmented Dickey–Fuller tests (with lags from 2 to 6) in the figure would show that they are also sometimes the best, but would not eliminate areas in which the Bayesian test is superior. Complete results are available from the authors, but we believe this to be sufficient to have confidence in the tests results presented next.

5 | EMPIRICAL RESULTS

We test for unit roots in the futures prices of all seven commodities at the three different sampling frequencies (daily, weekly, and monthly) by averaging the four types of conditional variance structures introduced above: GARCH-T, GARCH-N, ARCH-T, and AR-T. For each type of variance structure, the mean functions are assumed to be autoregressive models and are estimated with lag lengths from 1 to 6 to incorporate model specification uncertainty in the mean process. Thus, for each data series the test for stationarity is based on an average over 24 models. For each model, we produce 51,000 MCMC draws to assure convergence

[4]Note in the figure that SN on the horizontal axis of each box is the signal to noise ratio, so higher values mean less noise and easier data to get the right answer. DR on the vertical axis of each box is the dominant root so as one moves lower in each box the DGP is getting closer to nonstationary.

TABLE 3 Number of unit roots detected by Bayesian test

Commodity	Frequency	Model GARCH-T	GARCH-N	ARCH-T	AR-T	Total
Corn	Daily	0	6	0	6	12
	Weekly	0	6	0	6	12
	Monthly	0	6	0	6	12
Soybean	Daily	0	6	0	0	6
	Weekly	0	4	0	6	10
	Monthly	0	6	0	1	7
Cotton	Daily	0	6	0	6	12
	Weekly	0	6	0	6	12
	Monthly	2	6	0	6	14
Live cattle	Daily	0	6	0	5	11
	Weekly	0	3	0	6	9
	Monthly	0	6	0	6	12
Lean hog	Daily	0	0	0	0	0
	Weekly	0	0	0	0	0
	Monthly	3	0	0	0	3
Crude oil	Daily	0	0	0	0	0
	Weekly	0	0	0	0	0
	Monthly	0	0	0	0	0
Silver	Daily	0	1	0	0	1
	Weekly	0	3	0	0	3
	Monthly	0	0	1	0	1

Notes: Numbers indicate how many unit roots were detected under each variance structure specification; "Total" column shows the sum across all four variance structures.

and discard the first 21,000 draws to remove dependence on the initial conditions. All samples are tested for convergence using the Geweke test (Geweke, 1992) and they all pass the test.

Table 1 lists the probability of nonstationarity (a dominant root greater than 1) using the Bayesian model averaging test, where a value greater than or equal to 0.5 is considered to support nonstationarity (implying a symmetric loss function). The posterior distributions of the dominant roots are shown in the bottom rows of Figures 1–7 for all three sampling frequencies. For comparison, the *p*-values from two commonly used unit root tests, the ADF and PP tests, are also shown in the table.[5] While examining the comparative results, one thing to remember is that Bayesian and frequentist tests are completely different logically. For instance, Bayesian tests do not conform to the conventional 0.05 significance level from the classical sampling theory and common standards from frequentist econometrics like "power" do not translate perfectly to Bayesian tests which are based on the sample at hand rather than many hypothetical future samples.

Examining the results in Table 1, we find that the qualitative outcome of the Bayesian unit root test is generally unaffected by the sampling frequency. For our new test, cotton is the only commodity where the decision of the test changes with sampling frequency. We find the daily cotton data stationary but the weekly and monthly data nonstationary. Soybean, cattle, lean hog, crude oil, and silver futures are stationary at all three frequencies, while corn is (barely) nonstationary at all three frequencies. In contrast, the ADF and PP tests find all commodities except lean hog and crude oil to be nonstationary at all sampling frequencies. This result is likely due to the lower power of the ADF and PP tests and their placement of nonstationarity as the null hypothesis.

While the results are not greatly affected by the sampling frequency in terms of deciding between stationarity and nonstationarity, we do find a clear pattern in the posterior support for nonstationarity as the sampling frequency changes. For all five agricultural commodities the less frequent the observation, the higher the support for nonstationarity (interestingly, the two non-agricultural futures prices display no such trend [silver] or a slight opposite one [crude oil]). Selective sampling appears to bias results toward finding a unit root, especially in agricultural futures prices. This indicates that more mean-reversion

[5]While improved unit root tests have been developed since the augmented Dickey–Fuller and Phillips–Perron tests were first published, they are rarely used in actual empirical research. Applied econometricians almost exclusively perform ADF or PP tests.

TABLE 4 Posterior model weights

		Model			
Commodity	**Frequency**	**GARCH-T**	**GARCH-N**	**ARCH-T**	**AR-T**
Corn	Daily	0.244	0.277	0.231	0.248
	Weekly	0.031	0.521	0.019	0.429
	Monthly	0.189	0.361	0.182	0.268
Soybean	Daily	0.089	0.462	0.076	0.373
	Weekly	0.169	0.172	0.176	0.483
	Monthly	0.192	0.158	0.181	0.469
Cotton	Daily	0.173	0.368	0.152	0.307
	Weekly	0.305	0.417	0.135	0.143
	Monthly	0.263	0.249	0.261	0.227
Live cattle	Daily	0.388	0.115	0.385	0.112
	Weekly	0.351	0.158	0.347	0.144
	Monthly	0.308	0.196	0.312	0.184
Lean hog	Daily	0.305	0.321	0.205	0.169
	Weekly	0.218	0.433	0.215	0.134
	Monthly	0.223	0.322	0.221	0.234
Crude oil	Daily	0.224	0.314	0.229	0.234
	Weekly	0.205	0.345	0.210	0.236
	Monthly	0.171	0.171	0.304	0.354
Silver	Daily	0.151	0.388	0.168	0.292
	Weekly	0.221	0.327	0.223	0.229
	Monthly	0.223	0.324	0.225	0.228

information is provided by using the high frequency data which is consistent with the conclusion in Boswijk and Klaassen (2012). To be more specific, high frequency samples carry more information through changing volatility and fat-tailed innovations which can be captured by GARCH and ARCH models with a Student's t distribution in our Bayesian model averaging test. This information is lost when constructing a lower frequency sample by skipping intermediate observations. Thus, it appears important to use daily futures price data when testing for nonstationarity (equivalent to efficient markets), even if later modeling will proceed at a lower data frequency.

Table 2 shows the residual kurtosis from the models with an AR (1) mean function. The different mean functions used in our model averaging do not have a significant effect on the variability of the residuals so only the AR (1) residuals are shown here for

TABLE 5 Sensitivity of unit root results to the prior on the dominant root

		Prior			
Commodity	**Frequency**	**Beta (30,1.5)**	**Beta (40,2)**	**Beta (20,2)**	**Beta (30,4)**
Corn	Daily	0.642	0.606	0.501	0.493
	Weekly	0.852	0.794	0.603	0.496
	Monthly	0.890	0.854	0.752	0.538
Cotton	Daily	0.603	0.542	0.426	0.396
	Weekly	0.692	0.603	0.499	0.452
	Monthly	0.723	0.679	0.560	0.525
Live Cattle	Daily	0.310	0.298	0.198	0.200
	Weekly	0.368	0.330	0.235	0.120
	Monthly	0.522	0.486	0.331	0.161

Notes: The values are the probability that the modulus of the dominant root is equal to or greater than 1, $p(\phi_1 \geq 1|x)$., under four different priors on the dominant root.

simplicity. Kurtosis is a statistic which primarily describes the peakedness (width of peak), tail weight, and lack of shoulders of a probability distribution. The larger the kurtosis is, the higher and sharper the central peak is, and the longer and fatter the tails are. It is clear from the middle row of Figures 1–7 that the high frequency daily data displays leptokurtosis (meaning the tails are fatter than in the normal distribution), while for the lower-frequency weekly and monthly data the kurtosis value drops significantly. This is concrete evidence of how changing the sampling frequency can affect the information in the data. This extra information in daily sample can be captured well by the GARCH and ARCH models included in our Bayesian model averaging but is mostly ignored by the ADF and PP tests. Referring back to Table 1, one sees that the ADF and PP test results do not vary much across different sample frequencies.

Of course, separate from this ability to capture dynamic features such as changing volatility, the most desirable property of our Bayesian model averaging test is that it deals with model specification uncertainty within a unit root test. This issue exists in all empirical modeling work since we need some assumptions to begin analysis, but the importance of this issue in unit root tests appears to be high. In this paper, one such example is found with the monthly soybean sample. Using the ADF test on the monthly data and applying the commonly used 10% significance level, the soybean futures price is confirmed nonstationary if the model is specified with 1, 3, or 4 lags, while with either 2, 5, or 6 lags the ADF test finds the data to be stationary. A similar sensitivity to the number of lags in the ADF test is also shown for crude oil futures. Our Bayesian model averaging surmounts this specification sensitivity by averaging over six possible lag specifications (and can do more if the researcher wishes) to reach a final comprehensive test result.

Table 3 lists the number of unit roots detected in each variance structure specification with our BMA test. While Table 3 focuses on the impact of error specification on the outcome of unit root testing, there were also cases where test results varied with the autoregressive order of the mean equation (seen in Table 3 whenever the number of detected roots is between 0 and 6). Still, Table 3 focuses on how dramatically results change based on the error term specification. Test results favoring a unit root are concentrated in the GARCH model with normal errors (GARCH-N) and the regular AR model with Student's *t* errors (AR-T), implying any single model specification may not fully describe the behavior of the data and erroneous test results can be obtained by a single feature of the model specification.

Table 4 shows the posterior probability of each variance specification. Specifically, each of these probabilities is the sum of the probabilities for the six models with different lag lengths in the mean equation. The probabilities show a fairly even distribution of posterior support with only a single case where a variance specification secures a majority of the posterior support. Thus, although some model specifications are clearly more favored than others, there is no definitive, best model, making these commodity futures prices an excellent example of the need for tests that account for model specification uncertainty.

Table 5 presents some results on the sensitivity of the robust Bayesian unit root results to changes in the prior distribution, specifically the prior on the dominant root. In Table 5, the first two columns of results are for somewhat more concentrated priors, with less weight allocated to clearly stationary values of the dominant root. The last two columns of results are for somewhat less concentrated priors, with more weight allocated to more clearly stationary values of the dominant root. Not surprisingly, when the prior favors stationarity more (columns [3] and [4]), the posterior probability of nonstationarity is lower. However, the range of prior distributions shown in Table 5 is much wider than the variation in results. For example, a Beta (30,1.5) distribution places about 40% of the probability in the last 0.03 of the range while a Beta (30,4) only places 1.7% of the probability in that same region. Thus, while prior probabilities of nonstationarity are varied by an order of 20, posterior probabilities of the same measure vary by an order of 3 or less. This seems sufficient to show that only very large perturbations in the prior move the results by margins that change the results of the test.

6 | CONCLUSIONS

Financial data series usually display high volatility clustering with heavy tail distributions, which may affect the results of traditional asymptotic theory-based unit root tests. Some recent research has suggested that the sampling frequency of the data can also affect the outcome of unit root tests (Boswijk & Klaassen, 2012; Moral-Benito, 2015). This paper investigated these issues by testing seven commodity futures prices: corn, soybean, cotton, live cattle, lean hog, crude oil, and silver. We performed a new robust, Bayesian unit root test on three different sample frequencies: daily, weekly, and monthly prices with lower frequency series constructed by skipping intermediate observations of the daily data. Using Bayesian model averaging to account for different possible autoregressive lag lengths and different error variance specifications, we showed that sample frequency does have effects on the unit root test results. It appears that the high frequency data contains more mean-reversion information as well as having more leptokurtotic errors so that unit root tests on the lower frequency data can reach different conclusions because of information which is lost or altered by skipping observations.

Results from the empirical application offer several useful insights. First, allowing non-normality and conditional heteroskedasticity has significant impacts on the test results. Second, we uncovered clear evidence that not accounting for model specification uncertainty in unit root tests can sometimes produce results that are not robust. This could be seen from the augmented Dickey–Fuller tests on monthly soybean and crude oil futures price data. When a different number of lags is included in the test completely different decisions are reached on the stationarity of the series. Employing our Bayesian model averaging unit root test, however, the results are consistent across different sampling frequency, with the unit root test result changing only once for one commodity (cotton).

In many situations in applied finance and economics, econometric model specifications depend on the outcome of unit root tests on variables that include commodity prices. The results presented here should encourage researchers to perform those tests with daily data to preserve as much mean reversion information as possible and to think carefully about their error specification. Researchers now need to factor into their preliminary testing both that the sampling frequency and (not) specifying an error distribution that accounts for the characteristics which are common to financial asset prices may be determining the outcome of unit root testing. Such sensitivity of test results to model specification features show the usefulness of the model averaging approach taken here. When so much depends on getting this important facet of model specification correct, protecting against incorrect decisions on whether the commodity futures prices being modeled are stationary or not seems like insurance well worth having.

ACKNOWLEDGMENTS

The authors acknowledge seminar participants at Iowa State University, attendees at the NCCC-134 Conference on Applied Commodity Price Forecasting and Risk Management and the Agricultural and Applied Economics Association Annual Meetings, and two referees for helpful comments that improved the paper.

ORCID

Jeffrey H. Dorfman http://orcid.org/0000-0001-9568-1248

REFERENCES

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer.

Berger, J. O., & Yang, R. (1994). Noninformative priors and Bayesian testing for the AR(1) model. *Econometric Theory*, *10*(3/4), 461–482.

Bessler, D. A., & Covey, T. (1991). Cointegration: Some results on U. S. cattle prices. *The Journal of Futures Markets*, *11*, 461–474.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Boswijk, H. P. (2001). *Testing for a unit root with near-integrated volatility* (Discussion Paper). Tinbergen Institute. Retrieved from http://www.econstor.eu/handle/10419/85890

Boswijk, H. P., & Klaassen, F. (2012). Why frequency matters for unit root testing in financial time series. *Journal of Business & Economic Statistics*, *30*(3), 351–357.

Chambers, M. J. (2004). Testing for unit roots with flow data and varying sampling frequency. *Journal of Econometrics*, *119*(1), 1–18.

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, *90*(432), 1313–1321.

Choi, I. (1992). Effects of data aggregation on the power of tests for a unit root: A simulation study. *Economics Letters*, *40*(4), 397–401.

Choi, I., & Chung, B. S. (1995). Sampling frequency and the power of tests for a unit root: A simulation study. *Economics Letters*, *49*(2), 131–136.

Chow, Y. F. (1998). Regime switching and cointegration tests of the efficiency of futures markets. *Journal of Futures Markets*, *18*(8), 871–901.

Chowdhury, A. R. (1991). Futures market efficiency: Evidence from cointegration tests. *Journal of Futures Markets*, *11*(5), 577–589.

Dorfman, J. H. (1993). Bayesian efficiency tests for commodity futures markets. *American Journal of Agricultural Economics*, *75*(5), 1206–1210.

Dorfman, J. H., & Park, M. D. (2011). Estimating the risk-return tradeoff in agribusiness stocks: Linkages with the broader stock market. *American Journal of Agricultural Economics*, *93*(2), 426–433.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007.

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, *48*(5), 1749–1778.

Fan, R., Haiqi, L., & Park, S. Y. (2016). Estimation and hedging effectiveness of time-varying hedge rato: Nonparametric approaches. *Journal of Futures Markets*, *36*(10), 968–991.

Fortenbery, T. R., & Zapata, H. O. (1993). An examination of cointegration relations between futures and local grain markets. *Journal of Futures Markets*, *13*(8), 921–932.

Geweke, J, (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernardo, J. O. Berger, A. P Dawid, & J. F. M. Smith (Eds.), *Bayesian statistics 4* (pp. 169–193). Oxford: Oxford University Press.

Griffiths, W, & Xia, C. (2012). *Bayesian unit root testing: The effect of choice of prior on test outcomes*. Melbourne, Australia: The University of Melbourne.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, *20*(7), 873–889.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57*(1), 97–109.

Jing, L (2010). *Hastings-within-gibbs algorithm: Introduction and application on hierarchical model*. San Antonio: University of Texas.

Lai, K. S., & Lai, M. (1991). A cointegration test for market efficiency. *Journal of Futures Markets*, *11*(5), 567–575.

Ling, S., & Li, W. K. (2003). Asymptotic inference for unit root processes with GARCH (1, 1) errors. *Econometric Theory*, *19*(04), 541–564.

Lucas, A. (1995). Unit root tests based on M estimators. *Econometric Theory*, *11*(2), 331–346.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092.

Moral-Benito, E. (2015). Model averaging in economics: An overview. *Journal of Economic Surveys*, *29*(1), 46–75.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347–370.

Rothenberg, T. J., & Stock, J. H. (1997). Inference in a nearly integrated autoregressive model with nonnormal innovations. *Journal of Econometrics*, *80*(2), 269–286.

Schwarz, T. V., & Szakmary, A. C. (1994). Price discovery in petroleum markets: Arbitrage, cointegration, and the time interval of analysis. *Journal of Futures Markets*, *14*(2), 147–167.

Seo, B. (1999). Distribution theory for unit root tests with conditional heteroskedasticity. *Journal of Econometrics*, *91*(1), 113–144.

Tanner, M. A. (1998). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions* (3rd ed.). New York: Springer.

Tong, J., Wang, Z., & Yang, J. (2016). Information flow between forward and spot markets: Evidence from the chinese renminbi. *Journal of Futures Markets*, *36*(7), 695–718.

Yang, J., Bessler, D. A., & Leatham, D. J. (2001). Asset storability and price discovery in commodity futures markets: A new look. *Journal of Futures Markets*, *21*(3), 279–300.

How to cite this article: Wu X, Dorfman JH, Karali B. The impact of data frequency on market efficiency tests of commodity futures prices. *J Futures Markets*. 2018;38:696–714. https://doi.org/10.1002/fut.21912