On the Choice of the Unit Period in Time Series Models

Peter Fuleky *

August 14, 2011

Abstract

When estimating the parameters of a process, researchers can choose the reference unit of time (unit period) for their study. Frequently, they set the unit period equal to the observation interval. However, I show that decoupling the unit period from the observation interval facilitates the comparison of parameter estimates across studies with different data sampling frequencies. If the unit period is standardized (for example annualized) across these studies, then the parameters will represent the same attributes of the underlying process, and their interpretation will be independent of the sampling frequency.

Keywords: Unit Period, Sampling Frequency, Bias, Time Series.

JEL Codes: C13, C22, C51, C82.

*UHERO and Department of Economics, 540 Saunders Hall, University of Hawaii at Manoa, Honolulu, HI 96822. Email: fuleky@hawaii.edu.

1 Introduction

Since the seminal paper of Sargan (1974) there has been a vast amount of research conducted on the impact of a declining observation interval on the parameter estimates of continuous time models (see Bergstrom, 1988, for an early survey). However, in real world applications the observation interval is usually fixed at the data sampling frequency, for example monthly. When estimating a time series model, the researcher has to decide on a numerical value representing the length of an interval between observations. A monthly sampling frequency implies several possible designations for the observation interval: one twelfth of a year, one third of a quarter, one month, thirty days, and so on. Thus, the magnitude of the observation interval will depend on the choice of the reference unit of time, that is, the unit period (a year, a quarter, a month, and a day in the example above). It is important to recognize that, in contrast to the sampling frequency of most economic data, the choice of the unit period is at the discretion of the researcher. This paper illustrates the impact of this *choice* on model parameters and their estimates.

There are two widely used approaches to choosing the unit period. On the one hand, in discrete time models it is customary to set the unit period equal to the observation interval (see Hamilton, 1994). On the other hand, in continuous time models the unit period is usually set to a year so that it matches the reference period of the annualized underlying variable. The observation interval is then denoted as a fraction of a year, so that 1/12 represents monthly and 1/52 represents weekly observations (see Hurn et al., 2007; Phillips and Yu, 2009). However, when analyzing discrete approximations to continuous time processes, the choice of unit period is less clear, and some researchers follow the former, others the latter approach.

The unit period represents the reference unit of time for the underlying process. Using the same unit period across different studies facilitates the comparison of the results. A mismatch between unit periods across studies can occur in various settings: for example, through arbitrary choices in Monte Carlo experiments, or by using data sets with different sampling frequencies in empirical studies. For example, the parameters of popular interest rate models are frequently estimated from weekly or monthly data, but with the time elapsed between observations designated as unity (see for example Chan et al., 1992; Czellar et al., 2007). While equating the unit period with the observation interval can make the notation simpler, it can also make the comparison of results across studies more complicated. The choice of the unit period affects the interpretation of the parameters and the bias of their estimates, and this is true whether a study is conducted in discrete or continuous time.

In Section 2 and 3, I show the impact of the choice of the unit period on the interpretation of the model parameters and on the bias of parameter estimates, respectively. I use the exact solution of the Ornstein-Uhlenbeck (OU) process for the illustration of relevant issues. Section 4 concludes.

2 Interpretation of Model Parameters

The interpretation of the model parameters depends on the choice of the unit period. The idea can be illustrated by an example of weekly sampling from an OU process

$$F_\theta : dy = \theta_1(\theta_0 - y)dt + \theta_2 dW \ , \quad dW \sim \text{iid N}(0, dt) \ , \tag{1}$$

where θ_0 is the long run mean, θ_1 is the speed of mean reversion, and θ_2 is the volatility of the process. The sample of n weekly observations $\{y_t\}_{t=\Delta \ldots, n\Delta=T}$ can be generated from the exact solution of the OU process

$$F_\theta : y_t = \theta_0(1 - e^{-\theta_1\Delta}) + e^{-\theta_1\Delta} y_{t-\Delta} + \theta_2\sqrt{\frac{1 - e^{-2\theta_1\Delta}}{2\theta_1}}\epsilon_t \ , \quad \epsilon_t \sim \text{iid N}(0, 1) \ , \tag{2}$$

where Δ denotes the weekly observation interval. First, let the unit period of the OU process coincide with the observation interval so that the weekly observations are one unit of time apart (Δ=1). Then the θ_1 and θ_2 parameters represent *weekly* mean reversion and volatility.

Second, let the unit period of the OU process be a year, and assume that the *same* data set was generated by an alternative parameterization denoted with stars

$$F_{\theta^*} : dy = \theta_1^*(\theta_0^* - y)dt^* + \theta_2^* dW^* \ , \quad dW^* \sim \text{iid N}(0, dt^*) \ , \tag{3}$$

$$F_{\theta^*} : y_{t^*} = \theta_0^*(1 - e^{-\theta_1^*\Delta^*}) + e^{-\theta_1^*\Delta^*} y_{t^*-\Delta^*} + \theta_2^*\sqrt{\frac{1 - e^{-2\theta_1^*\Delta^*}}{2\theta_1^*}}\epsilon_{t^*} \ , \quad \epsilon_{t^*} \sim \text{iid N}(0, 1) \ , \tag{4}$$

where $\Delta^* = 1/52$ denotes the weekly observation interval when the unit period is a year. Now the

θ_1^* and θ_2^* parameters represent *annualized* mean reversion and volatility.

Note, I *chose* different unit periods above, but the data set I generated / observed were the same in both cases: $\{y_{t^*}\}_{t^*=\Delta^* \ldots, n\Delta^*=T^*} \equiv \{y_t\}_{t=\Delta \ldots, n\Delta=T}$. As long as the data generating process has a closed form solution, the parameters can be converted between models with different unit periods. To obtain the same weekly observations in (2) and (4), the following has to hold

$$y_t = y_{t^*} \quad \Longrightarrow \quad \theta_0^* = \theta_0, \quad \theta_1^* = \theta_1 \frac{\Delta}{\Delta^*}, \quad \theta_2^* = \theta_2 \sqrt{\frac{\Delta}{\Delta^*}}\ , \tag{5}$$

that is, the choice of a yearly unit period implies a 52 times larger mean reversion value in (3) than does a weekly one in (1). For example, if the annualized mean reversion is $\theta_1^* = 0.1$, then the weekly mean reversion is $\theta_1 = \theta_1^* \Delta^* \approx 0.002$. However, the long run mean, θ_0, and the volatility, θ_2, are not scaled by the same factor, and it can quickly become tedious to compare parameter estimates based on different unit periods.

The analysis can be extended to data sets sampled at different frequencies from a given process. Researchers often set the unit period equal to the observation interval thereby making the interpretation of the parameters dependent on the sampling frequency. If the unit period is not the same across studies, the comparison of parameter estimates in more complicated models that do not have closed form solutions may become infeasible. However, the interpretation of the parameters is associated with the unit period as opposed to the sampling frequency. If the unit period is standardized across studies, then the parameter estimates will be comparable despite the data being sampled at different frequencies. Thus a judicious choice of the unit period helps to disassociate the interpretation of the parameters from the sampling frequency and facilitates the comparability of parameter estimates across studies.

3 Bias of Parameter Estimates

Ball and Torous (1996) and Phillips and Yu (2009) showed that the finite sample estimates of the mean reversion parameter are severely biased in highly persistent time series. The persistency of the process is determined by $\theta_1 \Delta$ in (2), or $\theta_1^* \Delta^*$ in (4), where $\theta_1 \Delta = \theta_1^* \Delta^* = \alpha$ as shown in (5), and the finite sample bias of its estimate, $\hat{\alpha}$, is independent of the choice of the unit period. However, researchers are usually interested in the bias of the model-parameter estimate, which is

$$bias(\hat{\theta}_1) = \frac{1}{\Delta}(\hat{\alpha} - \alpha), \quad bias(\hat{\theta}_1^*) = \frac{1}{\Delta^*}(\hat{\alpha} - \alpha), \quad bias(\hat{\theta}_1^*) = \frac{\Delta}{\Delta^*} bias(\hat{\theta}_1) \ , \tag{6}$$

where $\hat{(\cdot)}$ denotes the estimate. Thus, the choice of the unit period influences the *absolute level* of the finite sample bias of the mean reversion estimate: the bias of the annualized parameter, $\hat{\theta}_1^*$, is 52 times larger than that of the weekly one, $\hat{\theta}_1$. However, the *relative* finite sample bias of $\hat{\theta}_1^*$ is equal to that of $\hat{\theta}_1$ for any choice of the unit period: $bias(\hat{\theta}_1^*)/\theta_1^* = bias(\hat{\theta}_1)/\theta_1 = (\hat{\alpha} - \alpha)/\alpha$.

In practice, the continuous time model (1) is often approximated by a crude Euler discretization

$$F_\mu : y_t = \mu_0 \mu_1 \Delta + (1 - \mu_1 \Delta) y_{t-\Delta} + \mu_2 \sqrt{\Delta} \xi_t \ , \quad \xi_t \sim \text{iid N}(0, 1) \ , \tag{7}$$

but here the naïve parameter estimates, $\hat{\mu}$, are asymptotically biased (Lo, 1988). Using (2) one can show that $\hat{\mu}$ asymptotically converges to $\mu(\theta)$

$$\mu_0(\theta) = \theta_0, \quad \mu_1(\theta) = \frac{1}{\Delta}(1 - e^{-\theta_1 \Delta}), \quad \mu_2(\theta) = \theta_2 \sqrt{\frac{1 - e^{-2\theta_1 \Delta}}{2\theta_1 \Delta}} \ , \tag{8}$$

and the asymptotic discretization bias is given by the difference $\mu(\theta) - \theta$. Similarly, $\hat{\mu}^*$ asymptotically converges to $\mu^*(\theta^*)$

$$\mu_0^*(\theta^*) = \theta_0^*, \quad \mu_1^*(\theta^*) = \frac{1}{\Delta^*}(1 - e^{-\theta_1^* \Delta^*}), \quad \mu_2^*(\theta^*) = \theta_2^* \sqrt{\frac{1 - e^{-2\theta_1^* \Delta^*}}{2\theta_1^* \Delta^*}} \ , \tag{9}$$

and using (5), the asymptotic discretization bias of $\hat{\mu}^*$ can be written as

$$asbias(\hat{\mu}_0^*) = \mu_0(\theta) - \theta_0, \quad asbias(\hat{\mu}_1^*) = \frac{\Delta}{\Delta^*} \ (\mu_1(\theta) - \theta_1), \quad asbias(\hat{\mu}_2^*) = \sqrt{\frac{\Delta}{\Delta^*}} \ (\mu_2(\theta) - \theta_2) \ . \tag{10}$$

Thus, the choice of the unit period influences the *absolute level* of the asymptotic discretization bias of the naïve estimates. For example, the asymptotic bias of the annualized mean reversion, $\hat{\mu}_1^*$, is 52 times larger than that of the weekly one, $\hat{\mu}_1$. (However, note that the scaling factor is not equal across all parameters.) Nevertheless, as in the case of finite sample bias, the *relative* asymptotic bias of $\hat{\mu}^*$ is equal to that of $\hat{\mu}$ for any choice of the unit period: $asbias(\hat{\mu}^*)/\theta^* = asbias(\hat{\mu})/\theta$.

Figure 1 displays the impact of the choice of different unit periods on the bias of the discrete mean reversion estimates $\hat{\mu}_1(\theta_1)$ for different θ_1 values: the choice of the unit period only has a scaling effect and the relative bias is the same for both unit period choices.

Therefore, assuming a given underlying process, it might be meaningful to compare the relative bias but not the absolute bias of parameter estimates across studies with different unit periods. Different unit periods usually arise from setting them equal to the observation interval in data sets with different sampling frequencies. But because the unit period can be chosen independently of the sampling frequency, absolute comparability of parameter estimates can be achieved by choosing the same unit period across studies.

4 Conclusion

I use a simple OU process to illustrate the impact of the *choice* of the unit period on parameter estimates. The choice of the unit period determines the notation for the length of the observation interval, and it affects the interpretation and absolute bias of parameter estimates. If the unit period is set equal to the observation interval, then parameter estimates will not necessarily be comparable across studies with different data sampling frequencies. However, if the unit period is decoupled from the observation interval, and is set equal across these studies, then parameters will represent the same attributes of the underlying process, and their interpretation will be independent of the sampling frequency. A common unit period serves as a reference unit across studies: it affords simple comparison of estimation results including conclusions about the effect of sample size or sampling frequency on the absolute bias of parameter estimates.

References

Ball, C. and Torous, W. (1996). Unit roots and the estimation of interest rate dynamics. *Journal of Empirical Finance*, 3(2):215–238.

Bergstrom, A. (1988). The history of continuous-time econometric models. *Econometric Theory*, 4(3):365–383.

Chan, K., Karolyi, G., Longstaff, F., and Sanders, A. (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, 47(3):1209–1227.

Czellar, V., Karolyi, G., and Ronchetti, E. (2007). Indirect Robust Estimation of the Short-Term Interest Rate Process. *Journal of Empirical Finance*, 14(4):546–563.

Hamilton, J. (1994). *Time series analysis.* Princeton University Press.

Hurn, A., Jeisman, J., and Lindsay, K. (2007). Seeing the Wood for the Trees: A Critical Evaluation of Methods to Estimate the Parameters of Stochastic Differential Equations. *Journal of Financial Econometrics*, 5(3):390.

Lo, A. (1988). Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data. *Econometric Theory*, 4(2):231–247.

Phillips, P. and Yu, J. (2009). Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance. *Handbook of Financial Time Series*, 497:530.

Sargan, J. (1974). Some discrete approximations to continuous time stochastic models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(1):74–90.

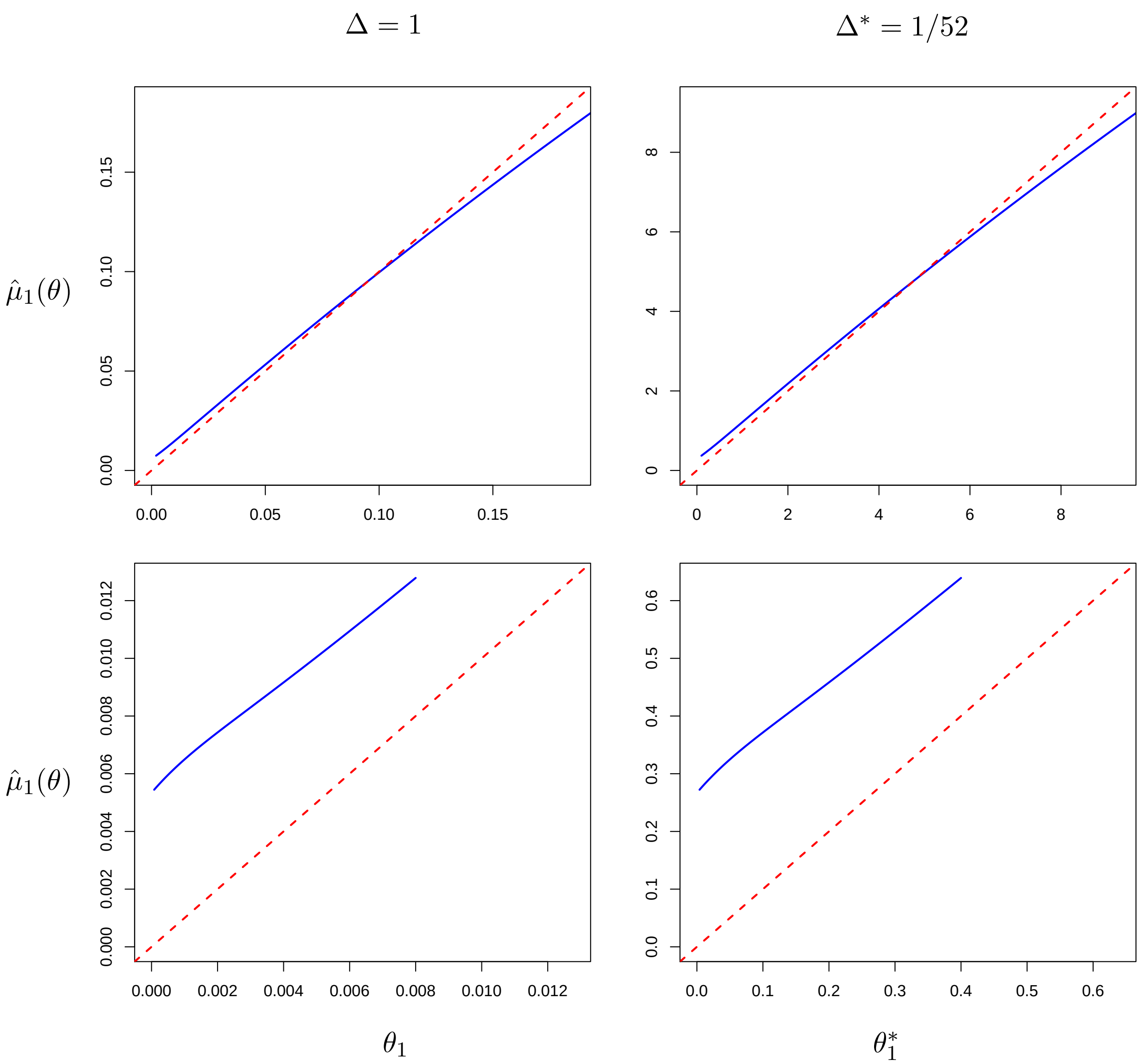

Figure 1: Impact of unit period choice on parameter estimates and bias.

Note: The left and right diagrams are based on the same weekly pseudo-data generated from the OU process (2) with the range of θ_1 and θ_1^* as shown on the horizontal axis, and the remaining parameters held constant at $\theta_0 = 0.1$, $\theta_2 = 0.1/\sqrt{52}$, and $\theta_0^* = 0.1$, $\theta_2^* = 0.1$. The difference between the left and right diagrams is the choice of the unit period - left: $\Delta = 1$ (parameter values in weekly terms), right: $\Delta = 1/52$ (parameter values in annual terms). The sample size is set to 1000 observations. The estimated model is given by (7). The red dashed 45° line represents the theoretical unbiased estimator, and the blue solid line indicates the mean of 1000 least squares estimates of the mean reversion parameter μ_1 as a function of θ_1. The bottom diagrams zoom in on mean reversion values close to 0, the range of realistic parameter values for interest rate models and where the finite sample bias is the strongest. As apparent from the diagrams, the choice of the unit period only has a scaling effect and the relative bias is the same for both unit period choices. (The top diagrams also illustrate that the positive finite sample bias dominates close to 0, but at $\theta_1 = 0.1$ or $\theta_1^* = 5.2$ the negative discretization bias takes over.)