Reducing the Bias in Empirical Studies Due to Limit Moves

Kenneth H. Sutrick

INTRODUCTION

A number of different analytical techniques are required to do empirical research in finance. Regression analysis, in particular, is important. Regression analysis is used in futures research to study relationships between sets of variables such as price volatility and volume, to determine hedge ratios, and in asset pricing models such as testing whether futures prices are unbiased predictors of future spot prices.

With futures data, econometric tools are complicated by the fact that many futures contracts are subject to price limits. Once a price limit is hit, the true equilibrium price becomes unobservable. One must decide how the limits are to be treated. Numerous approaches to handling the limits have already been used. Researchers have: thrown them away, left them in as they are, aggregated them by using weekly or monthly returns to reduce the effects of the limits, and estimated equilibrium prices to be used in place of the limits.

It is obvious that these limit moves have the potential to introduce bias into a study. In particular any procedure which utilizes estimates of variability will be biased because the limits will cause variance estimates to be too small on average. The purpose of this article is to systematically look at this bias in the context of regression analysis and investigate ways to reduce it.[1]

Previous articles have noted that inferences can differ when a limit structure is incorporated into the analysis. McCurdy and Morgan (1987) in studying the unbiasedness hypothesis above for foreign exchange found that the limits affected their results. Kodres (1988, 1990) found fewer rejections of unbiasedness for foreign exchange than do Hodrick and Srivastava (1987) or McCurdy and Morgan (1987), once a limit structure was explicitly considered. Roll (1984) found that a significant regression coefficient was rendered insignificant by replacing the limits with estimates of equilibrium prices along with using a slope dummy variable for limit move days.

In all these articles, one or more variables were subject to price limits. In the foreign exchange papers, the dependent variable was daily return, the independent variables were lagged daily returns, and all were prone to limits. In Roll's article on weather and Orange Juice Futures, the dependent variable was the percentage difference between the actual temperature and the forecasted temperature from the National Weather Service,

[1]Regression calculations are affected by variance estimates. For example, the slope is $\mathrm{Cov}(X,Y)/\mathrm{Var}(X)$ in a simple regression.

Kenneth H. Sutrick is an Associate Professor at Murray State University.

The Journal of Futures Markets, Vol. 13, No. 5, 527–543 (1993)

CCC 0270-7314/93/050527–17

the independent variables were daily return and lagged returns. Only the independent variables were subject to limit moves in this article.

The method of approach to the limits was different in each of the articles. Hodrick and Srivastava did not take the limits into account. McCurdy and Morgan sampled weekly on Wednesdays, while Kodres and Roll both essentially used estimates of equilibrium prices in parts of their analysis (although neither used that terminology specifically). In the Kodres' article, price estimates were used in place of the limits for the independent variables and a limit structure analogous to a truncated distribution was used for the dependent variable. In Roll's article, the focus had nothing to do with price limits, but he found that when the limits were used without modification, a coefficient for lagged returns was a significant predictor of the temperature forecast error. He reasoned that prices during limit moves could not adequately reflect all available information. He then used a rough estimate of equilibrium prices before proceeding with the analysis.

The method in this article uses price estimates in place of the limited prices as inputs into a regression procedure. Kodres (1988, 1990) used covered interest rate arbitrage to get a proxy for the limited futures price, while Burghardt (1988) employed put–call parity. Roll (1984) used the following rough estimate, paraphrased from his article:

•If a particular day registered a limit price move, the price for that day was assumed to be the price on the next subsequent day without a limit move. If multiple limits in the same direction occurred, the price on the first day with no limit was brought back to the day of the first limit move and all intervening days were discarded. If an up-limit was followed by a down-limit, the price change on day 1 was taken to be zero and day 2 was discarded. The next included observation was then day 3 if it were non-limit.

In this article equilibrium price estimates are based on utilizing the price process itself, as is done in Roll's article, as opposed to using independent information such as option prices. It is a general method based on maximum likelihood and conditional expectations. This technique can be used when others are not available, i.e., when options don't exist or are thinly traded or when there is not a theoretical model that can be relied on. The advantage of this approach over others is that it is based on the joint distribution of the prices subject to limit moves, allowing the maximal amount of information to be extracted. The joint distribution takes the process volatility into account where the other methods do not.

Whether one estimates equilibrium prices or handles the limits in a different fashion it will affect parameter estimates and possible conclusions. The potential bias this causes in regression estimates for the various approaches above are determined via a Monte Carlo simulation. With limits, a simulation has one advantage over real data in that the true parameters are known and one can compare parameter estimates to them under different scenarios. The first section discusses estimation of equilibrium prices by maximum likelihood and conditional expectations; the second section considers biases in regression; and in the third section an application to Treasury Bond Futures is analyzed. It is found that conditional expectations represent the unobservable equilibrium prices reasonably well and reduce the bias for regression.

Assumptions must be made about the variables to carry out the simulations. It is assumed that at least one of the variables in a regression model is a daily price change subject to limits. It is also assumed that the equilibrium price process is unaffected by the imposition of the limits as is done in Kodres (1988, 1990). This implies that the observed price on a non-limit day following one or more limit move days is the same as it would be if the limits did not exist. Without this assumption a model for how the limits affect prices would need to be specified which is an additional complication. All variables are assumed, at least conditionally, to have independent mean zero increments. For the sake of exposition, price changes are taken to have a normal distribution. The normality assumption is unnecessary but is analytically simpler since the joint distribution of normals is also a normal. Also for simplicity, the limits are assumed constant even though real limits can change from day to day. The methods can be modified easily to handle limits that change or where only down-limits exist, such as in the Japanese stock market. The modifications are taken up in the appendix.

ESTIMATION OF EQUILIBRIUM PRICES

In this section two methods of estimating the unobservable equilibrium prices that would occur under a limit move are detailed. The first method is based on maximum likelihood and the second involves the conditional distribution of the unobservable variables. Technically, price changes instead of prices are estimated. For Day i, let X_i denote the equilibrium price change which is sometimes unobservable and X_i^* the always observable price change which is the X process subjected to a price limit denoted by M.[2] A particular realization of X_i^* is denoted x_i^* and the volatility of X_i, $SD(X_i)$ is denoted σ_i. The variability is allowed to change from day to day, but it is assumed that σ_i is known or at least known conditionally as one would in an ARCH model [Engle (1982)].

Maximum likelihood estimates, MLE, under these circumstances can be computed by treating the unobserved equilibria as parameters of the problem. The MLE equilibrium price changes are derived from that sample path which has the highest probability density in the conditional probability set determined by the limits.[3] Consider a sequence of limit moves starting on Day 1 and ending non-limit on Day j. Let $\hat{X} = (\hat{X}_1, \ldots, \hat{X}_j)$ denote the maximum likelihood sample path changes, $x = (x_1^*, \ldots, x_j^*)$ denote the observed price changes subject to the limits, and L the likelihood function; then,

$$\hat{X} = \arg \max_{x_1,\ldots,x_j} L(X_1, \ldots, X_j \big| x_1^*, \ldots, x_j^*)$$

The above formula is general and applies to the case where the limits can change from day to day, or are not symmetric between up- and down-limits, or if only down-limits exist.[4]

As an example of the MLE, the simplest case is where the price finishes limit-up on Day 1 and non-limit on Day 2. It follows that $X_1^* = M$ and $X_2^* = x_2^*$ where $-M < x_2^* < M$[5] and with price changes that are normal the MLE is:

$$\begin{aligned}\hat{X} = (\hat{X}_1, \hat{X}_2) &= \arg \max_{\substack{X_1 \geq M \\ X_1 + X_2 = M + x_2^*}} \frac{1}{2\pi\sigma_1\sigma_2} e^{-x_1^2/2\sigma_1^2 - x_2^2/2\sigma_2^2} \\ &= \left(\max\left(M, \frac{\sigma_1^2(M + x_2^*)}{\sigma_1^2 + \sigma_2^2} \right), \min\left(x_2^*, \frac{\sigma_2^2(M + x_2^*)}{\sigma_1^2 + \sigma_2^2} \right) \right) \qquad (1)\end{aligned}$$

Since this is a constrained maximization, the MLE is found by setting the derivative to zero and also by checking the endpoints of the parameter space determined by the limits. The reason for the normality assumption is evident here, a nice closed form expression

[2]By assumption, the limits do not affect equilibrium prices. If the price hits the limit one day and is non-limit the next, this implies that $X_1 + X_2 = X_1^* + X_2^*$ with analogous relationships for multiple limit days.

[3]The conditional probability set is explained in the appendix.

[4]Only down-limits that change would imply that the x_i^* are all negative and have different magnitudes for $i < j$.

[5]If $X_1^* = M$ is an actual transaction price, as opposed to just the settlement price, Day 1 would be treated as a non-limit. According to the theoretical model here with a continuous distribution, this would have probability zero.

for the estimates. In addition to M and x_2^* the MLE is seen to be a function of the ratio of σ_1 to σ_2. In this situation, Roll (1984) uses the estimate $\hat{X}_1 = M + x_2^*$ and eliminates the second day. This $\hat{X}_1$ is a maximum likelihood estimate under the limiting condition $\sigma_1/\sigma_2 \to \infty$.

In the case where the price finishes down-limit one day and non-limit the next, the maximum likelihood estimated price change would be the negative of the up-limit situation. For multiple limit moves, MLE's are computed in a similar manner but are more complicated due to the interplay of the boundary conditions and the standard deviations. However, if the price finishes limit-up for $(j - 1)$ days and non-limit at Day j, then

$$\hat{X}_1 = \max\left(M, \frac{\sigma_1^2((j-1)M + x_j^*)}{\sigma_1^2 + \ldots + \sigma_j^2}\right)$$

When variability is stationary so that $\sigma_1 = \cdots = \sigma_j$, the estimated price change on the first limit day reduces to $\hat{X}_1 = M$, which is an uninteresting result. Intuitively, the price change should be above the limit. In general, if daily volatility is constant, the maximum likelihood estimates of prices on limit days is the limit itself.

A second method which gives an intuitively more appealing solution is based on the conditional distribution of the price change sample path given the limits. The estimates are defined by the conditional expectation, CE:

$$\hat{\hat{X}}_i = E(X_i | x_1^*, \ldots, x_j^*)$$

The general form of the conditional distribution is given in the appendix; but, if Day 1 is up-limit and the next day non-limit, then:

$$\hat{\hat{X}}_1 = \frac{\int_M^\infty x_1 e^{-x_1^2/2\sigma_1^2 - (M + x_2^* - x_1)^2/2\sigma_2^2} dx_1}{\int_M^\infty e^{-x_1^2/2\sigma_1^2 - (M + x_2^* - x_1)^2/2\sigma_2^2} dx_1} \tag{2}$$

The integrals in (2) must be evaluated with numerical integration. More on this is said in the appendix.

To get a feel for what the various estimators look like, Table I contains some sample values for the estimates of the limited price X_1 under different scenarios. The prices are assumed to have finished limit-up on Day 1 at a limit $M = 2.00$ and non-limit on Day 2 with an observed price change of $x_2^* = 1.00$ or -1.00. The standard deviations are taken to be around 1.0 so that the limits are hit about 5% of the time. The MLE and Conditional Expectation given are obtained by plugging the values of the parameters into eqs. (1) and (2). Roll's Value is $M + x_2^*$. To determine which procedure is best, one must compare the values in Table I to the true value x_1. This is done in the simulation below. However, a perusal of Table I shows that the conditional expectation furnishes estimates which have better smoothness properties and correspond better to intuition. For example if the first day's variability is higher relative to the second day, the estimate of X_1 should be higher. If x_2^* is higher with everything else constant, the estimate should be higher. These are always true for conditional expectation but is not necessarily true for the others. Also note that even when σ_1 is not equal to σ_2 the MLE tends to give the limit as an estimate. This is a function of the assumptions here, namely symmetric mean zero unimodal price changes where the density is highest at zero and decreases monotonically away from zero. Under the normality assumption in maximizing the observed densities the MLE will tend to give estimates as close to the origin as possible subject to the constraints (i.e., the limits). Only if the first day's variability is large enough compared to the second day's will the estimate be above the limit. The conditional expectation, unlike maximum likelihood, is not just a function of the ratio of σ_1 to σ_2 but depends somewhat on their magnitudes. Roll's value is also a conditional expectation under the limiting condition $\sigma_1/\sigma_2 \rightarrow \infty$.

Table I gives estimates of equilibrium prices in a few selected examples but does not demonstrate which procedure is optimal in any sense. This is done with a simulation. A random set of equilibrium price changes X are generated and whenever an X_i is above the limit, an estimate of it $\hat{X}_i$ is computed based on the observed price change X_i^* and the volatility σ_i that day. To decide which procedure gives estimates which are closest to the true equilibrium price the root mean square error

$$\text{RMSE} = \left[\frac{1}{n}\sum (X_i - \hat{X}_i)^2\right]^{1/2} \tag{3}$$

for each estimator is computed. The estimates will depend on whether one, two, or more consecutive limits are hit, but only the case where the price is limit-up or -down one day and non-limit the next is considered here. The details on how the simulation is carried out are discussed in the next few paragraphs.

To use either maximum likelihood or conditional expectations one must know or know conditionally the variability of the price process, which can be constant or changing. There are innumerable possible models of price volatility which can be used and a discussion of this is contained in Pagan and Schwert (1990). Examples of models for nonconstant volatility include ARCH, Engle (1982), and GARCH, Bollerslev (1986). In an ARCH model the process is given by

$$X_i = U_i[\alpha^0 + \alpha^1 X_{i-1}^2 + \cdots + \alpha^k X_{i-k}^2]^{1/2} \tag{4}$$

where U_i is standard normal independent of all previous U_k and X_k when $k < i$. The ARCH model has the property

$$\alpha_i^2 = \text{Var}[X_i \mid X_{i-1}] = \alpha^0 + \alpha^1 X_{i-1}^2 + \cdots + \alpha^k X_{i-k}^2$$

This was generalized by Bollerslev (1986) to a GARCH model where

$$\sigma_i^2 = \alpha^0 + \alpha^1 X_{i-1}^2 + \cdots + \alpha^k X_{i-k}^2 + \beta^1 \sigma_{i-1}^2 + \cdots + \beta^l \sigma_{i-l}^2$$

For this study, data for only a few simple processes are generated for the X variable. The first process run assumes a conditionally and unconditionally constant standard deviation and the second process is determined with an ARCH(1) model where $k = 1$ in (4). The GARCH model is not run. The ARCH and GARCH representations have the property that high volatility will be followed by high volatility. This feature makes it less attractive for modeling limits. Ma, Rao, and Sears (1989) found that futures volatility was reduced after limit moves for Treasury Bond Futures, also a large shock might mean

Table I
VALUES OF ESTIMATES OF X_I ON A LIMIT DAY

σ_1	σ_2	x_1^*	x_2^*	Cond. Ex.	MLE	Roll's Value
1.00	1.00	2.000	1.000	2.416	2.000	3.000
1.50	0.50	2.000	1.000	2.768	2.700	3.000
0.50	1.50	2.000	1.000	2.117	2.000	3.000
1.00	1.00	2.000	−1.000	2.254	2.000	1.000
1.50	0.50	2.000	−1.000	2.161	2.000	1.000
0.50	1.50	2.000	−1.000	2.107	2.000	1.000

large volatility for the equilibrium price on that one day followed by normal variability thereafter. Meanwhile, the shock may cause multiple limit moves. However, there are also models where a large increase implies smaller volatility. Cox's constant elasticity of variance model (CEV) for option pricing is an example of this [Cox (1975)]. To model this behavior the third process is chosen to be of the form

$$X_i = U_i/[\alpha^0 + \alpha^1 X_{i-1}^2]^{1/2}$$

which could be considered an Inverse ARCH(1) model.

In using maximum likelihood and conditional expectations, one must be concerned with what effect a misspecification of the conditional variance process will have on the ability to predict equilibrium prices. Even if the process is correctly specified, the parameters may not be known with perfect precision. To study this, three steps are taken in this part of the simulation. First, the three equilibrium price change processes above are generated with the following parameters: a constant standard deviation of $\sigma = 1$, then ARCH(1) and Inverse ARCH(1) with parameters $\alpha^0 = 0.5$ and $\alpha^1 = 0.5$.[6] Second, for each process estimates of equilibrium prices are made, assuming the variance is correctly specified and also misspecified. For example, when the price changes were generated with a constant standard deviation, estimates are made by taking $\sigma_i \equiv 1$ in eqs. (1) and (2), but also by calculating the σ_i according to the ARCH and Inverse ARCH models before substituting into the two equations. A similar program is carried out for the other two processes. Third, to study precision, the parameters in the variance models are taken to be known or unknown. In the simulation all combinations are examined so that the equilibrium price changes might be generated by an ARCH process but equilibrium prices estimated with σ_i computed from an Inverse ARCH model where the parameters are estimated from the data, etc. In the cases where the variance model parameters are estimated, it is done by maximum likelihood from the observed price change process $x^* = (x_i^*, \ldots, x_{n_0}^*)$ where n_0 is chosen to be 20 days. The time period is chosen arbitrarily but is small enough so that parameter estimates would not be highly accurate. A time period of $n_0 = 50$ is also run but the results are virtually identical and are not reported.

The root mean square errors, eq. (3), for the estimators of equilibrium prices when there is a limit one day and no limit the next day are presented in Table II. The numbers in the table are based on $n = 2{,}000$ estimates of X_1. Prices are sampled 20 days at a time, estimates of the process parameters are made, then estimates of equilibrium prices are calculated whenever one day limits occur. Another 20 days are sampled, and the process repeated until X_1 is estimated 2,000 times. The entries in the table correspond to how the prices are generated and how the standard deviations for eqs. (1) and (2) are computed. For example, the entry under *ConditionalExpectation, ARCH*.5, .5, and *Process*1 means that any one day limits are estimated with the conditional expectation assuming that the process is an ARCH process with parameters .5 and .5 even though the data X are actually generated with a constant standard deviation of 1.0. A typical prediction error was .320 in that case. The entry under *ConditionalExpectation, SigmaEstimated,* and *Process*3 means that limits are estimated with conditional expectations when the variance is assumed to be unknown and stationary[7] when in reality the equilibrium prices come from an inverse ARCH process. The standard deviation is estimated from the observed data and then the condition expectation is computed with a typical prediction error .404 resulting. The row for *Limit* corresponds to using the value M as an estimate of the unobservable price.

[6]By using these parameters the second and third processes are conditionally heteroscedastic but unconditionally have variance equal to one. The first process is conditionally and unconditionally homoscedastic.

[7]Conditionally and unconditionally constant.

Not all the simulations for the maximum likelihood estimates are reported because, in general, they did poorly. The only numbers reported are the MLE assuming that the correct process is known and the correct parameters are known. In other places the RMSE's for maximum likelihood are higher.

To summarize Table II, prediction errors are the smallest when the correct process and the correct parameters are known. However, some procedures do well across all simulations. The best method overall is to use conditional expectation with volatility computed by estimating an inverse ARCH model. Also, assuming a constant standard deviation and estimating it before inputing into the conditional expectation model works reasonably well. Maximum likelihood does not do well and the limits themselves do not represent equilibrium prices very well.

REGRESSION BIAS

The bias in regression coefficients is now considered for the different ways of handling the limits. The bias, for example, for a slope estimator $\hat{\beta}$ of a slope parameter β is defined to be $E(\hat{\beta}) - \beta$. Since, under limit moves, the computation of this expectation is intractable, the bias is determined with another simulation. In the simulation each approach to the limits defines a different set of inputs to the regression equation and a different estimator to be compared. When the conditional expectation is used in place of the limits, the effect of a misspecification of the variance and lack of precision in the variance parameters is checked.

To keep things manageable, only simple regression is considered. Even with simple regression there are several permutations of the assumptions on the variables. The dependent and independent variables may or may not be subject to limits, they may be conditionally heteroscedastic or not, the error term may be stationary or heteroscedastic, etc. Since the focus here is on the bias and not on the correct specification of the econometric model, only a situation similar to that faced by Roll (1984) is considered.

Table II
ROOT MEAN SQUARE ERRORS, EQ. (3), FOR ESTIMATING X_l WHEN DAY 1 IS A LIMIT DAY AND DAY 2 IS NOT

Estimator and Variance Process Assumption	Process 1 X Generated as Stationary Normal	Process 2 X Generated as ARCH .5,.5	Process 3 X Generated as Inverse ARCH .5,.5
Conditional Expectation			
Sigma = 1.0	.295	.403	.416
Sigma Estimated	.305	.401	.404
ARCH .5, .5	.320	.379	.502
ARCH Parameters Estimated	.319	.397	.388
Inv. ARCH .5, .5	.342	.464	.331
Inv. ARCH Parameters Est.	.312	.386	.378
Limit	.462	.534	.668
MLE Parameters Known	.462	.564	.443

The dependent variable Y is assumed to be unconstrained, and generated from an unconstrained independent variable X and an error term. The X is considered to be a price change variable subject to a daily limit M and would not be observable in real data during a limit move. The variable X^* which is the X process subject to the limits is always observable and is the input into the regression.[8] Three different variances for the error term are chosen to determine how the bias is affected by the amount of noise in the data. For the simulation the X-process is chosen to have two distributions, a stationary standard normal distribution and an ARCH .9,.1 process which is conditionally normal. The error term is a function of a variable ϵ which is also standard normal. All variables are derived mutually independent except for ARCH where they are conditionally independent. The Y variable is generated according to a model

$$Y_i = X_i + \text{Damp} \cdot \epsilon_i$$

for a period of $n_0 = 20$ days and the damping constant is chosen to be .25, 1.0, or 2.0. Thus, the true value of the intercept is zero, the true value of the slope is one, and the damping constant defines the signal to noise ratio. The advantage of a simulation is that any estimator $\hat{\beta} = \hat{\beta}\left(X_1^*, \cdots, X_{n_0}^*\right)$ can be compared to the unbiased estimator $\hat{\beta}_u = \hat{\beta}(X_1, \ldots, X_{n_0})$. The unbiased estimator is observable in the simulation but not always for real data.

There are $n = 10{,}000$ sets of $n_0 = 20$ generated for each value of the damping constant. Slopes $\hat{\beta}$ and intercepts $\hat{\alpha}$ for each approach to the limits for each set of 20 days are computed. The estimators simulated are those approaches to the limits discussed in the Introduction and where the limits are replaced by conditional expectations. The approaches and the results of the simulation are found in Table III. The *Unbiased* estimator is the regression of Y on X and the *LimitsAsIs* estimator is the regression of Y on X^*. The *LimitDaysDiscarded* reduces the 20-day sampling period by eliminating all limit days from the sample before computing the regression parameters. The *Aggregated* estimator groups days together in both the dependent and independent variables. Since n_0 is so small the data are grouped only two days at a time as opposed to weekly for real data. *Roll's* and *ConditionalExpectation* replace limit days with the corresponding estimates of equilibrium prices before regresssion. Using maximum likelihood estimates of equilibrium prices before regression is identical to the Limits As Is estimator when X has the stationary normal distribution and almost identical to it when X is ARCH. For the conditional expectation a specification of the variance process for the independent variable is required. The X process is chosen to have two distributions to determine how a misspecification of the variance process affects the regression bias. The variance misspecification is taken to occur during the ARCH process. There it is incorrectly assumed to be conditionally homoscedastic. To check on precision the standard deviation inputs for eq. (2) are $\sigma_i \equiv 1$ or a constant standard deviation equal to $s_{x^*} = \left[\sum(x^* - \overline{x}^*)^2\right]^{1/2}/(n_0 - 1)$ which is estimated from the data. These inputs are used for both distributions of the independent variable. The means $\overline{\alpha} = \sum \hat{\alpha}/n$, $\overline{\beta} = \sum \hat{\beta}/n$ and standard deviations $s_\alpha = \left[\sum(\hat{\alpha} - \overline{\alpha})^2\right]^{1/2}/(n - 1)$, $s_\beta = \left[\sum(\hat{\beta} - \overline{\beta})^2\right]^{1/2}/(n - 1)$ for each approach are reported in Table III. Sample sizes of $n_0 = 50$ are also run but the results are similar and are not reported.[9] There are two possible ways to compare the results of Table III for the

[8] Note that Y is determined from X not X^*.

[9] Some difficulties with using conditional expectations come up during simulations. If limits are hit on three consecutive days the conditional expectation requires the evaluation of a triple integral. Evaluating a triple

Table III

MEANS AND STANDARD DEVIATIONS OF 10,000 SIMULATED ESTIMATES OF SIMPLE REGRESSION INTERCEPTS AND SLOPES FOR VARIOUS WAYS OF HANDLING AN INDEPENDENT VARIABLE X WHICH IS SUBJECT TO LIMITS

(a) X is generated with a stationary normal $\sigma = 1$ distribution.

Estimator	Mean Intercept	SD Intercept	Simulation Std. Error $\overline{\alpha}$	Mean Slope	SD Slope	Simulation Std. Error $\overline{\beta}$
Damp = .25						
Unbiased	.000998	.057547	.000575	1 − .000034	.060463	.000605
Limits As Is	.001133	.058888	.000589	1 + .025626	.071837	.000718
Limit Days Discarded	.000945	.063959	.000640	1 − .009727	.076941	.000769
Aggregated	.002321	.121799	.001218	1 + .014296	.102187	.001022
Roll's	.001453	.074682	.000747	1 − .018035	.110848	.001108
Cond. Ex. SD Known	.000992	.058134	.000581	1 − .000148	.068153	.000682
Cond. Ex. SD Estimated	.001043	.058279	.000583	1 − .001040	.069356	.000694
Damp = 1.00						
Unbiased	.003338	.232858	.002329	1 − .002344	.241092	.002411
Limits As Is	.003269	.233801	.002338	1 + .023547	.250946	.002509
Limit Days Discarded	.003708	.240968	.002410	1 − .009401	.280849	.002808
Aggregated	.007859	.487611	.004876	1 + .012533	.385646	.003856
Roll's	.003853	.244446	.002444	1 − .020027	.265682	.002657
Cond. Ex. SD Known	.003331	.233240	.002332	1 − .002318	.243846	.002438
Cond. Ex. SD Estimated	.003334	.233301	.002333	1 − .003131	.244566	.002446

(continued)

Table III (continued)

(a) X is generated with a stationary normal $\sigma = 1$ distribution.

Estimator	Mean Intercept	SD Intercept	Simulation Std. Error $\overline{\alpha}$	Mean Slope	SD Slope	Simulation Std. Error $\overline{\beta}$
Damp = 2.00						
Unbiased	.007504	.456410	.004564	1 + .001455	.479264	.004793
Limits As Is	.007642	.457499	.004575	1 + .026889	.493972	.004940
Limit Days Discarded	.007644	.471319	.004713	1 − .007643	.552531	.005525
Aggregated	.012717	.946993	.009470	1 + .013219	.770083	.007701
Roll's	.005929	.470305	.004703	1 − .016330	.498108	.004981
Cond. Ex. SD Known	.007663	.456586	.004566	1 + .001832	.480978	.004810
Cond. Ex. SD Estimated	.007617	.456779	.004568	1 + .000832	.481432	.004814

(b) X is generated as an ARCH .9,.1 process.

Estimator	Mean Intercept	SD Intercept	Simulation Std. Error $\overline{\alpha}$	Mean Slope	SD Slope	Simulation Std. Error $\overline{\beta}$
Damp = .25						
Unbiased	.000033	.058056	.000581	1 + .000163	.060980	.000610
Limits As Is	.000187	.059355	.000594	1 + .026844	.074130	.000741
Limit Days Discarded	.000440	.064818	.000648	1 − .007662	.077938	.000779
Aggregated	.000408	.123487	.001235	1 + .014404	.107504	.001075
Roll's	−.005350	.078980	.000790	1 − .024549	.121821	.001218
Cond. Ex. SD Known	.000083	.058667	.000587	1 + .001551	.070057	.000701
Cond. Ex. SD Estimated	.000077	.058918	.000589	1 − .000179	.072552	.000726

(continued)

Table III (continued)

(b) X is generated as an ARCH .9,.1 process.

Estimator	Mean Intercept	SD Intercept	Simulation Std. Error $\overline{\alpha}$	Mean Slope	SD Slope	Simulation Std. Error $\overline{\beta}$
Damp = 1.00						
Unbiased	−.001811	.230662	.002307	1 − .000542	.245597	.002456
Limits As Is	−.002063	.231434	.002314	1 + .025519	.254421	.002544
Limit Days Discarded	−.002296	.237479	.002375	1 − .007366	.282736	.002827
Aggregated	−.004545	.481663	.004817	1 + .018774	.392304	.003923
Roll's	−.002539	.242196	.002422	1 − .023517	.273118	.002731
Cond. Ex. SD Known	−.001992	.230878	.002309	1 + .000763	.248592	.002486
Cond. Ex. SD Estimated	−.002093	.230980	.002310	1 − .000321	.248757	.002488
Damp = 2.00						
Unbiased	.001639	.458819	.004588	1 + .001422	.487142	.004871
Limits As Is	.001167	.459723	.004597	1 + .028168	.501612	.005016
Limit Days Discarded	.003155	.472909	.004729	1 − .002956	.558815	.005588
Aggregated	.002846	.955514	.009555	1 + .022052	.779693	.007797
Roll's	.002659	.471743	.004717	1 − .020396	.512343	.005123
Cond. Ex. SD Known	.001474	.458931	.004589	1 + .002435	.489078	.004891
Cond. Ex. SD Estimated	.001597	.458980	.004590	1 + .001392	.489757	.004898

various methods. One is in terms of their means $E(\hat{\alpha})$, $E(\hat{\beta})$ and then in terms of their proximity to the unbiased estimator which is observable in the simulation. The means is taken up first. It should be noted that $\overline{\alpha}$, $\overline{\beta}$ are estimates of $E(\hat{\alpha})$, $E(\hat{\beta})$ respectively subject to the randomness in the simulation. The estimation error is evident since by construction the unbiased estimator has $E(\hat{\beta}_u) = 1$ but $\overline{\beta}_u \neq 1$ for all the cases studied. However the error in these estimated means can be measured and is represented by the simulation standard error $s/\sqrt{10000}$. These are also reported in Table III. In all six cases the slope means for the unbiased estimator are well within the sampling error of the simulation from the true value of $\beta = 1$. In some cases the conditional expectation estimated mean comes closer to the true value than the estimated mean for the unbiased procedure. This is only due to the randomness in the simulation but suggests that the CE is nearly unbiased.[10]

For the problems discussed in the Introduction, the slope is the more important regression coefficient. The relative order of the simulated means for the slopes from the true value of 1 is consistent across all damping constants and thus is not a function of the amount of noise in the data. The relative order is also consistent between the two distributions for the independent variable and not dependent on that assumption either. Table III shows that using the limits as they are has the largest bias and other procedures are biased as well. This is apparent since a number of the means are a large number of simulation standard errors away from 1. For example, in part (a) of the table under the .25 damping constant, the error for the slope means is roughly .0007 and the Limits As Is mean of 1.025626 is at least 35 standard errors away. It can be concluded that a number of procedures are biased even with the randomness of the Monte Carlo technique.

There are other inferences that can be made from Table III. For the procedure which aggregates data two days at a time, the bias is reduced compared to Limits As Is estimator but at the expense of precision since the sample size is significantly reduced. Sampling weekly rather than every other day, as has been done in actual studies, reduces this precision even more. If limit days are removed from the data set, bias can still occur because the effects of the limits carry over into the next non-limit day. This is essentially what Roll found in the orange juice and weather data. Therefore, if the limit days are to be removed, the first non-limit day should be removed also. Finally, one can conclude that when equilibrium prices are estimated with conditional expectations and used in place of the limits, the bias is significantly reduced and reduced more if price volatility is known precisely. However, part (b) of the table indicates that if the variance process is misspecified, you are better off estimating variance parameters from the data than using what you think are the correct parameters. This implies that if one is not sure about the variance process, one should assume the parameters are unknown.

A second criterion for deciding which approach to the limits is optimal is to see which procedure is closest to the simulation observable unbiased estimator. The results of Table III indicate, on average, what happens for each method but do not demonstrate the relationship between them on any one particular trial. For any one particular realization $x^* = (x_1^*, \ldots, x_{n_0}^*)$ all the approaches give a different estimate of the parameters and any

indefinite integral accurately takes a prohibitive amount of computer time. If three consecutive limits are hit, a proxy is used: the price change on the fourth day is taken to be zero, the price on the third day is assumed to be non-limit finishing exactly at the limit. This allows only a double integral to be evaluated. If four of more consecutive limits are hit that set of $n_0 = 20$ is eliminated and another generated. This results in only 11 sets discarded out of the 30,000 generated when X is normal and 75 out of 30,000 when X is ARCH. The bias these two changes produce in the simulation is probably negligible.

[10]A truly unbiased estimator which is a function of the limited prices is analytically intractable.

one of them could be closest to the true values. Even the unbiased estimator can be beaten at times. While this is all observable in a simulation, it cannot be done in a real data situation since the true parameter value is unknown. Therefore, it is reasonable, a priori, to find an approach which is as close as possible trial to trial from a procedure with desirable properties such as unbiasedness. The relative distance from the unbiased procedure can be calculated in the simulation by constructing confidence intervals for the difference between each estimator and the unbiased estimator. A set of 95% confidence intervals, $\bar{d} \pm 1.96 s_{\bar{d}}/\sqrt{n}$ where $d = \hat{\alpha} - \hat{\alpha}_u$ or $d = \hat{\beta} - \hat{\beta}_u$, are given in Table IV. Only the confidence intervals from Table III where X is generated as stationary normal and the damping constant was .25 are reported. Other confidence intervals give qualitatively similar results.[11]

The confidence intervals in Table IV are centered closest to zero and are narrowest for the conditional expectation. It would be considered an optimal procedure under the second criterion. For the slope parameter using the limits as they are is the farthest from the unbiased procedure and consistently above it. It is not a good procedure under this criterion. In terms of both of the criteria considered above, the CE is the method that recovers the information available in the unbiased procedure to the largest extent.

AN APPLICATION TO TREASURY BOND FUTURES

Samuelson (1965) provides a theoretical model that futures price variability should increase as the time to maturity of the contract decreases. This issue has been studied by a number of researchers [Rutledge (1976); Miller (1979); Castelino (1981); Anderson (1985); Milonas (1986); Grammatikos and Saunders (1986)]. Castelino, Anderson, and Milonas found strong evidence for this maturity hypothesis for many futures contracts while the others had mixed results. This study investigates how using conditional expectation estimates of the equilibrium prices, as opposed to the limited data, can strengthen or weaken these propositions. This is illustrated with settlement price data for Treasury Bond Futures obtained from the Chicago Board of Trade. Daily data are used which many of the papers have been reluctant to use because of the limits.

Table IV
95% CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN VARIOUS ESTIMATORS AND THE UNBIASED ESTIMATOR IN SIMPLE REGRESSION

	Intercepts		Slopes	
Comparison Estimator	Lower Confidence Limit	Upper Confidence Limit	Lower Confidence Limit	Upper Confidence Limit
Limits As Is	−.0001686	.0006146	.0411257	.0435761
Limits Discarded	−.0009893	.0008160	−.0175536	−.0144388
Aggregated	−.0005819	.0030906	.0203146	.0245347
Rolls	−.0007837	.0022866	−.0327249	−.0266941
Cond. Exp. SD Known	−.0002922	.0002740	−.0012366	.0008613
Cond. Exp. SD Unknown	−.0002379	.0003871	−.0028032	−.0005175

[11] When no limits are hit all methods are the same and the difference between them is zero. The confidence intervals are based on the 6,059 times out of 10,000 that the 20-day sampling period has at least one limit.

To apply the techniques of the first section to Treasury Bond contracts, it should be noted that there are about 12 different maturities that are traded at any one particular time. If on a given day a limit is hit for one contract, it tends to be hit on all the other maturity contracts as well; so that all contracts tend to act alike in this regard. It also turns out that the limits have the same relative effect no matter what the year is for the problem of maturity and volatility. For this reason it is sufficient to look at only one contract for one specific period of time. The December, 1986 contract for trading days from February, 1986 thru September, 1986 is used to illustrate the effect of the limits. This contract is chosen arbitrarily except for the fact that it has a relatively large number of limit moves (13) during this period.

The effect of the limits is found by running two separate regressions on the data, one with the limited prices and the second with the limits adjusted by conditional expectations. The dependent variable chosen to be price variability and the independent variable are chosen so that an increase in volatility as maturity approaches will result in a positive slope coefficient. There are many different ways to define price volatility [Grammatikos and Saunders (1986)] but here the dependent variable is chosen to be $Y_i = [\log(P_i/P_{i-1})]^2$ where P_i represents the settlement price on Day i. The independent variable is taken as X_i = # days from February 1, 1986 so that a positive slope and larger X implies a larger Y. In the first regression, the P_i are the Board of Trade settlement prices and in the second regression the P_i for those settlement prices which are at the limit are replaced by the conditional expectation estimate of the equilibrium price. The volatility estimates required for the conditional expectation are estimated from the data based on the limit day and the previous 24 trading days.[12] The results of the regressions are reported in Table V.

Table V reveals the slope coefficient in the regression is small and negative indicating that for this contract volatility decreased slightly on average instead of increasing as the time maturity decreased. The slope is more negative when estimates of equilibrium prices for the limits are used. This will always be the effect of the limits in this problem. The

Table V

REGRESSION OUTPUTS FOR TESTING PRICE VOLATILITY VS. TIME TO MATURITY FOR THE DECEMBER 1986 TREASURY BOND FUTURES CONTRACT

$Y_i = [\log(P_i/P_{i-1})]^2$, X_i = **# days from Feb. 1, 1986**

	Regression 1 Limits As Is	**Regression 2 Limits Adjusted**
Constant	.0001400566	.0001680013
X Coeficient	−.0000000543	−.0000000942
Std. Err. of X Coef.	.0000001676	.0000002426
R-Squared	.0006361885	.0009131333
Std. Err. of Y Est.	.001504523	.0002178843
Degrees of Freedom	165	165

[12]This short time period for volatility estimates is chosen because the articles mentioned have documented various nonstationarities such as a "month effect" and a "year effect." The 25 days is approximately one month of trading. At some points short-term price trends appear in the data resulting in large estimates of a standard deviation when the data is used as is, sometimes even larger than the limits itself. These short term trends are removed before computing the standard deviations to get appropriate estimates of daily volatility.

limits restrict the volatility on the limit day and on average increase it on the next day. Adjusting for the limits will give a relatively higher volatility one day earlier and less the next leading to a more negative slope coefficient when X is defined as above. Using the limited data as is will always tend to bias a study in favor of the maturity hypothesis. Even though the change in the slope here is small, since there are only 13 limit days compared to 167 trading days, and not enough to make the slope statistically significant it is always useful to know the direction of the bias in whatever procedure used. While the effect is small for this problem, other articles have found a stronger effect due to the limits.

CONCLUSIONS

Limits have the potential to bias empirical studies. This study investigates ways to reduce estimation bias by estimating the equilibrium prices and using these as inputs. When the limited prices are used as is, bias results. This bias can be reduced by using estimates of equilibrium prices in place of the limits. The use of estimated prices can indicate the size and direction of this bias. If the errors in the regression model are uncorrelated, perhaps the best thing to do is to remove all limit days and the first non-limit day after the limits. This reduces the effective sample size somewhat but is unbiased. This study finds that one should never aggregate the data to reduce the effect of the limits. Even though the bias is reduced, the effective sample size is greatly reduced and a large amount of statistical power is lost.

Two different methods of estimating equilibrium prices are presented, maximum likelihood and conditional expectations. Maximum likelihood estimation, which, in general, is an excellent technique does not work well here. It is useful to find out that the MLE is a function of the ratio of daily volatilities and perhaps volatility should be taken into account. The MLE is, under certain assumptions, equal to the limit plus the next day's price change, which was used in Roll (1984). Estimates based on conditional expectations give intuitively appealing answers and when used in place of the limits as inputs into a regression procedure reduce the bias. Conditional expectations are higher when you expect them to be higher such as when volatility is higher and in regression reproduce more of the information from the unobservable equilibrium prices than any other method. The CE helps even if you assume that price variability is constant[13] but will work better when the structure of this variability is identified.

A final reason for estimating equilibrium prices, as opposed to eliminating days or using the limits as is, is that in studying price volatility, the limits are the most interesting part of the data. The conditional expectation can then be used to get a handle on this. An application to Treasury Bond Futures demonstrates that using the limited prices without adjustment is biased in favor of accepting the hypothesis that volatility increases as the time to maturity approaches.

APPENDIX

The general form of the conditional distribution is given here and allowances are made for limits that change from day to day. Consider a sequence of limit moves for $(j - 1)$ days followed by a non-limit on Day j and let the observed price change process be denoted by $x^* = (x_1^*, \ldots, x_j^*)$ with M_i being the limits.

First the conditional probability set determined by the limits must be found. For example, if Days 1 and 2 are up-limit while Day 3 is limit-down and Day 4 is

[13]Conditionally and unconditionally.

non-limit, then $x^* = (M_1, M_2, -M_3, x_4^*)$ and the equilibrium price change process x_i must satisfy $x_1 > M_1, x_1 + x_2 > M_1 + M_2, x_1 + x_2 + x_3 < M_1 + M_2 - M_3, x_1 + x_2 + x_3 + x_4 = M_1 + M_2 - M_3 + x_4^*$ and the conditional probability set is

$$\left[x_1, x_2, x_3, x_4 : x_1 > M_1, x_2 > M_1 + M_2 - x_1, x_3 < M_1 + M_2 - M_3 - x_1 - x_2,\right.$$
$$\left. x_4 = M_1 + M_2 - M_3 + x_4^* - x_1 - x_2 - x_3\right].$$

If there are only one-sided limits, such as only lower limits, the conditional probability set will not have any greater than terms. The general form of the conditional probability distribution is derived from the definition of conditional probability and the conditional probability set. The general form is given by

$$P\left[x_1, \ldots, x_j \,\middle|\, x_1^*, \ldots, x_j^*\right] =$$
$$\frac{e^{-\sum_{i=0}^{j-1} x_i^2/\sigma_i^2 - (x_1^* + \ldots + x_j^* - x_1 - \ldots - x_{j-1})^2/2\sigma_j^2}}{\int_{a_1(x^*)}^{b_1(x^*)} \cdots \int_{a_j(x^*)}^{b_j(x^*)} e^{-\sum_{i=0}^{j-1} x_i^2/\sigma_i^2 - (x_1^* + \ldots + x_j^* - x_1 - \ldots - x_{j-1})^2/2\sigma_j^2} dx_{j-1} \cdots dx_1}$$

where

$$b_k(x^*) = \infty \cdot I(x_k^* = M_k) + \sum_{i=0}^{k-1} x_i^* - \sum_{i=1}^{k-1} x_{i-1}$$

$$a_k(x^*) = -\infty \cdot I(x_k^* = -M_k) + \sum_{i=0}^{k-1} x_i^* - \sum_{i=1}^{k-1} x_{i-1}$$

where $0 \cdot \infty = 0$ and where $I(A)$ is the indicator function, $I(A) = 1$ if A is true and 0 otherwise. When $k = 1$ the second summation in the formulas for a_k and b_k can have a lower limit which is larger than the upper limit, the summations are defined to zero in that case. Note that limits that can change daily come into play in the limits of integration $a_k(x^*), b_k(x^*)$ since $x_k^* = \pm M_k$ for $k < j$. The conditional expectation is just the integral of x_i with respect to that probability measure over the conditional probability set determined by the limits. These integrals must be evaluated numerically since closed form expressions do not exist. Since they are also indefinite integrals, extra effort must be taken to compute them. For this article the region of integration is split into finite segments, either intervals if there is one limit day or squares for two consecutive limit days. Each segment is estimated to eight decimal places using Romberg's method, a standard numerical integration technique. Segments are evaluated sequentially starting at the origin until their overall contribution is negligible.

Bibliography

Anderson, R (1985): "Some Determinants of the Volatility of Futures Prices," *The Journal of Futures Markets,* 5:331–348.

Barclay, M., Warner, J., and Litzenberger, R. (1990): "Private Information, Trading Volume and Stock Return Variances," *Review of Financial Studies,* 3(2):233–254.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics,* 31:307–327.

Bollerslev, T. (1987, August): "A Conditionally Heteroscedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics,* 69:542–547.

Brennan, M. (1986): "A Theory of Price Limits in Future Markets," *Journal of Financial Economics,* 16:213–233.

Burghardt, G. (1988): "Comments to Kodres (1988)," *The Review of Futures Markets,* 7(1):172–174.

Castelino, M. (1981): "Futures Markets: The Maturity Effect on Risk and Return," Unpublished Ph.D. Dissertation, The City University of New York.

Cox, J. (1975): "Notes on Option Pricing I; Constant Elasticity of Variance Diffusions," Working Paper, Stanford University, Stanford, CA.

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica,* 50:987–1007.

Garcia, P., Leuthold, R., and Zapata, H. (1986): "Lead–Lag Relationships Between Trading Volume and Price Variability: New Evidence," *The Journal of Futures Markets,* 6(1):1–10.

Garman, M., and Klass, M. (1980): "On the Estimation of Security Price Volatilities from Historical Data," *Journal of Business,* 53(1):67–78.

Grammatikos, T., and Saunders, A. (1986): "Futures Price Variability: A Test of Maturity and Volume Effects," *Journal of Business,* 59(2, pt. 4):319–330.

Greenwald, B., and Stein, J. (1989): "Transactional Risk, Market Crashes, and the Rolé of Circuit Breakers," Working Paper, Harvard University.

Harvey, C., and Huang, R. (1900): "Inter and Intraday Volatility in the Foreign Currency Futures Markets," Working Paper, Fuqua School of Business, Duke University.

Hodrick, R., and Srivastava, S. (1987): "Foreign Currency Futures," *Journal of International Economics,* 22:1–24.

Kodres, L. (1988): "Tests of Unbiasedness in Foreign Exchange Futures Markets: The Effects of Price Limits," *The Review of Futures Markets,* 7(1):138–166.

Kodres, L. (1990): "Tests of Unbiasedness in Foreign Exchange Futures Markets: An Examination of Price Limits and Conditional Heteroscedasticity," Working Paper, School of Business Administration, University of Michigan.

Ma, C. (1991): "The Information Content of Price Limit Moves," Working Paper, Texas Tech University.

Ma, C., Rao, R., and Sears, R. (1989): "Limit Moves and Price Resolution: The Case of the Treasury Bond Futures Market," *The Journal of Futures Markets,* 9(4):321–335.

Miller, K. (1979): "The Relation between Volatility and Maturity in Futures Contracts," in *Commodity Markets and Futures Prices,* Leuthold, R. M. (ed.), Chicago Mercantile Exchange, pp. 25–36.

Milonas, N. (1986): "Price Variability and the Maturity Effect in Futures Markets," *The Journal of Futures Markets,* 6(3):443–460.

McCurdy, T., and Morgan, I. (1987): "Tests of the Martingale Hypothesis for Foreign Currency Futures," *International Journal of Forecasting,* 3:131–148.

Pagan, A., and Schwert, W. (1990): "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics,* 45:267–290.

Roll, R. (1984, Dec.): "Orange Juice and Weather," *The American Economic Review,* 74(5):861–880.

Rutledge, D. (1976): "A Note on the Variability of Futures Prices," *Review of Economics Statistics,* 58:118–120.

Samuelson, P. (1965): "Proof that Properly Anticipated Prices Fluctuate Randomly," *Industrial Management Review,* 6:41–49.

Schwert, W. (1989): "Why Does Stock Market Volatility Change Over Time?," *Journal of Finance,* 44(5):1115–1153.

Schwert, W., and Sequin, P. (1989): "Heteroscedasticity in Stock Returns," *Journal of Finance,* 45(4):1129–1155.