The Distribution of Gold Futures Spreads

Geoffrey Poitras

A sizable body of empirical research exists which examines the distributional properties of speculative price changes. In this literature, the distributions for combinations of prices are not directly examined. This is a significant omission because in many financial markets trades are based on combining prices. In foreign exchange markets, for example, swap rates are quoted which implicitly combine spot and forward exchange rates. In futures markets, price spreads for contracts with different delivery dates are directly traded in the pits. By construction, the distribution for changes in the difference between two prices is a mixture of component distributions. In the case of futures price spread changes, the distribution is a weighted combination of the distributions for changes in futures prices and implied carry costs. In addition to providing information useful for statistical purposes (e.g., testing hypotheses involving futures spreads), the distributional properties of futures price spreads (*fps*) can also provide indirect information on the variables which determine *fps*.[1]

This article examines the distributional properties of the gold *fps*: the traded difference between gold futures prices for two different delivery months. A number of competing distributional alternatives for describing the behavior of price changes are examined: the lognormal, heteroskedastic normal, and the stable. Theoretical motivations for testing specific fps distributional hypotheses are provided. Testing procedures include examining the behavior of parametric distributional tests under temporal aggregation to identify the appropriate distributional hypotheses.[2]

DISTRIBUTIONAL HYPOTHESES APPLIED TO FPS

Specific factors determining the difference between prices of futures contracts for the same commodity with different delivery dates depend on the commodity under consideration (e.g., Jones (1981), Rentzler (1986), and Yano (1989)). For example, if the commodity is harvestable (e.g., grain futures), then the harvest month usually

This article is based on a section of the author's Ph.D. dissertation. In addition to the useful direction of P. Dhrymes, B. Lehmann, and D. Modest the helpful comments of the referees is acknowledged. Financial support from the Social Sciences and Humanities Research Council of Canada as well as data support from the Center for the Study of Futures Markets is acknowledged.

[1]Cornew, Town, and Crowson (1984) provides a summary of a number of implications of the failure of the normality assumption. Nonnormality also affects statistical hypothesis testing in least squares regressions involving financial variables (e.g., Knight (1986)).

[2]Generally, temporal aggregation involves widening the sampling interval for the time series of a flow variable (e.g., Weiss (1984), and Stram and Wei (1986)).

Geoffrey Poitras is an Associate Professor, Faculty of Business Administration, Simon Fraser University.

The Journal of Futures Markets, Vol. 10, No. 6, 643–659 (1990)

CCC 0270-7314/90/060643-17$04.00

provides a discontinuity in the futures price structure. Another aspect is whether the commodity earns a carry return. For example, in the case of T-Bond or T-Bill futures, the presence of a carry return dictates that the term structure of financial futures prices depends on the relationship between carry cost and carry return, i.e., two interest rate processes must be taken into account in determining *fps*. Because of the different fundamentals which determine *fps* for a given commodity, there is likely to be considerable diversity in the generating process for *fps* changes in different commodities.

Gold spreads combine a number of features useful for empirical analysis. For example, due to the nature of the cash and carry arbitrage for gold, the gold futures price structure is not subject to discontinuities. More importantly, because the physical commodity does not earn a carry return, the relationship between prices for gold futures contracts with different maturities is affected only by one interest rate process: the implied carry return. These factors significantly simplify the problem of identifying the relative contribution of the component distributions in determining the *fps* distribution. This study seeks to identify which component variable plays the greatest role in determining *fps* by comparing distributional test results for different specifications of the component and *fps* variables.

The following definitions are used:

$F(t, S)$ = the gold futures price at time t for delivery at time $S \in \{N,T\}$, T is the date for deferred delivery, N is the date for nearby delivery ($T > N$).

$ic(t,T\text{-}N)$ = the carry return for the period T-N implied in gold futures prices at time t ($t < N$).

These can be used to specify:

$$F(t,T) = (1 + ic(t,T\text{-}N))F(t,N) \tag{1(a)}$$

$$fps(t) = F(t,T) - F(t,N) = ic(t,T\text{-}N)F(t,N) \tag{1(b)}$$

$$ic(t,T\text{-}N) = (F(t,T) - F(t,N))/F(t,N) \tag{1(c)}$$

$$\Delta F(t,N) = F(t+1,N) - F(t,N) \tag{1(d)}$$

$$\Delta ic(t,T\text{-}N) = ic(t+1,T\text{-}N) - ic(t,T\text{-}N) \tag{1(e)}$$

(1(b)) shows that *fps* levels are not likely to be stationary because futures price levels are not stationary. In effect, the *fps* level is proportional to the price level $F(t,N)$. With this in mind, the change in *fps* is given:

$$\begin{aligned}\Delta fps(t,t+1) &= (F(t+1,T) - F(t+1,N)) - (F(t,T) - F(t,N)) \\ &= ic(t+1,T\text{-}N)\Delta F(t,N) + F(t,N)\Delta ic(t,T\text{-}N)\end{aligned} \tag{2}$$

As illustrated in (2), the change in *fps* is a function of two processes which have been shown in a number of studies to be stationary (or near stationary).[3]

From (2) it is apparent that the payoff from a *fps* is dependent on the change in two variables: implied carry costs and futures price levels. To eliminate the impact of price level changes, some spread traders "tail" the spread trade by adjusting the spread so that, when initiated, there is equal dollar value in both legs of the spread.[4] The size of the tail is usually determined by taking $(F(t,T)/F(t,N))$ contracts of the

[3]See Poitras (1985) for a bibliography. The problems that occur when the process is near stationarity are not examined here.

[4]Yano (1989) provides a theoretical discussion of the general problem of determining tails for various types of spreads.

nearby contract for every deferred contract.[5] In other words, the tailed futures spread (*tfps*) is specified:

$$tfps(t) = F(t,T) - (F(t,T)/F(t,N))F(t,N)$$

The significance of tailing the spread is apparent when the change in the *tfps* is considered:

$$\begin{aligned}\Delta tfps(t,t+1) &= (F(t+1,T) - (F(t,T)/F(t,N))\,F(t+1,N)) \\ &\quad - (F(t,T) - (F(t,T)/F(t,N))\,F(t,N)) \\ &= F(t+1,N)\Delta ic(t) \qquad (3)\end{aligned}$$

In other words, the payoff on a tailed spread is a function of the change in implied carry costs alone and is not a function of the change in futures prices. However, because (3) is scaled by $F(t+1,N)$, the $\Delta tfps$ distribution differs empirically from that of Δic.

Identification of the distributional process generating observations of a random variable usually involves an application of transformations. For example, in many cases the price level data must be transformed by first differencing to achieve stationarity. In addition, use of parametric normality tests to identify specific distributional alternatives involves transforming the data to a form consistent with the null hypothesis for the test statistics. In this study, the distributional tests reported are: the standardized sample moments of skewness ($\sqrt{\beta_1}$) and kurtosis (β_2), the studentized range and an asymptotic chi-squared test combining the third and fourth sample moments. These specific statistics are selected for reasons of power and practicality. In particular, there is considerable evidence (e.g., Pearson, D'Agostino, and Bowman (1977), White and Macdonald (1980), Bera and Mackenzie (1986), and Poitras (1990)) demonstrating that the power of these tests compare favorably with other available parametric and nonparametric alternatives.

Regarding the statistical properties of the distributional tests used, the third and fourth standardized sample moments are desirable tests in identifying specific departures from normality. However, neither statistic performs as well in detecting departures from normality when distributions are of unknown shape. Further, the standardized sample moments may be sensitive to outliers. To identify outliers, the studentized range (SR) is provided where: $SR = (\max(X_t) - \min(X_t))/\sigma_x$. Fama and Roll (1971) recommend the SR test for detecting symmetric stable distributions. Results are also reported for a chi-squared test based on the asymptotic normality of the distributions for $\sqrt{\beta_1}$ and β_2, i.e.: chi-squared (2) $= T((\beta_1/6) + (\beta_2)^2/24)$ where T is the number of observations in the sample.[6] The primary difficulties with combining the two statistics in this manner are that the sampling distributions are not independent when the sample size is small and there is evidence that the speed of convergence to the asymptotic distribution is somewhat slow. To correct for this, appropriately adjusted confidence regions for calculating the significance of the chi-squared tests provided by Bera and Mackenzie (1986) are used.

[5]Because contracts take only integer values, tailed spread positions usually involve a sizable number of contracts. For example, if $ic = 10\%$, then a tailed spread would establish 11 nearby contracts for every 10 deferred contracts. This alters the margin requirements somewhat because, for margin purposes, this trade would be treated as 10 spreads and 1 open position.

[6]This formulation of the chi-squared test assumes that β_2 is centered about its expected value for the normal population (i.e., $\beta_2 = 0$ under the null hypothesis of normality, not its expected value of 3).

Given this statistical background, normality is tested in two ways: by estimating the relevant distributional parameters, skewness, kurtosis, and standard deviation, and conducting the relevant parametric normality tests on the appropriately transformed variables; and, by examining the behavior of the estimated distributional parameters when the differencing interval is lengthened. In the latter case, both daily and weekly differencing intervals are examined to assess whether convergence to normality is taking place. To formulate the null hypotheses which are to be tested, the following transformations for *fps* are examined at various points:

$$\text{DFPS} = fps(t + 1) - fps(t)$$
$$\text{RFPS} = (\text{DFPS}/fps(t))$$
$$\text{DTL} = tfps(t + 1) - tfps(t)$$

Similar transformations for the underlying futures prices and implied carry returns are examined also:

$$\text{DIC} = ic(t + 1) - ic(t)$$
$$\text{RIC} = (\text{DIC}/ic(t))$$
$$\text{DPR} = \text{F}(t + 1,\text{T}) - \text{F}(t,\text{T})$$
$$\text{RPR} = (\text{DPR}/\text{F}(t,\text{T}))$$

The implied mnemonics are D for changes and R for rate of gain. Each of these transformations plays a different role in testing for the distributions of the underlying variable.

Lognormality

Though early empirical research on the distribution of speculative price changes focusses on the lognormal (e.g., Osborne (1959)), the importance of the assumption of lognormality is largely theoretical. Information on the goodness-of-fit of the lognormal is useful in interpreting the empirical validity of theoretical models based on the lognormal assumption (e.g., the Black-Scholes option pricing formula). The tests for lognormality considered here can be motivated by assuming that futures prices and implied carry costs are lognormally distributed. In continuous time, this can be represented:

$$d\text{F}(t,\text{N})/\text{F}(t,\text{N}) = f\,dt + \sigma d\text{W}_f$$
$$dic(t,\text{T-N})/ic(t,\text{T-N}) = n\,dt + \sigma_n d\text{W}_n$$

where f, n are the means of the respective process, σ_f,σ_n are the respective standard deviations, and dW_f, dW_n are the respective standardized Wiener variates.

Combining these results with (2) and manipulating:

$$\begin{aligned} d(fps)/fps &= d\text{F}(t,\text{N})/\text{F}(t,\text{N}) + dic(t,\text{T-N})/ic(t,\text{T-N}) \\ &= (f + n)\,dt + \sigma_f d\text{W}_f + \sigma_n d\text{W}_n \\ &= \mu dt + \sigma d\text{W} \end{aligned} \tag{4}$$

The implications are immediate. If futures prices and implied carry costs are lognormally distributed, then *fps* is lognormal also. This result should not be affected by temporal aggregation. To transform *fps* into a form consistent with the statistical null hypothesis of normality, empirical tests are based on the rate of gain (RFPS).[7] From (3), the *tfps* case is different. Assuming that scaling by prices does not significantly affect the resulting distribution, then the process for $\Delta tfps$ is directly related to the process for Δic. As a result, statistical tests for lognormality of *tfps* are based on the first difference of *tfps*.

Stability

The class of stable distributions is important in statistical theory because it contains all distributions which qualify as weak limits of sums of independent random variables (Zolotarev (1986)). However, decades of empirical work on the applicability of stable distributions to financial data have produced mixed results (e.g., Mandelbrot (1963), Fama (1965), Teichmoeller (1971), Fama and Roll (1971), Officer (1972), Blatberg and Gonedes (1974), Hsu, Miller, and Wichern (1974), Hagerman (1978), and Fielitz and Roselle (1983)). Recent empirical studies provide support for the *iid* stable Paretian hypothesis by introducing correction for nonstationarity. Using daily foreign exchange price changes for various major currencies, McFarland, Pettit, and Sung (1982) and So (1987) find evidence in favor of the nonnormal stable. However, these studies and a number of others (e.g., Hsieh (1988)) detect various forms of nonstationarity in means or variances. It is possible to generalize the symmetric stable hypothesis by observing that in addition to the characteristic exponent and measures of scale and location, stable distributions are characterized by a skewness parameter. (Early studies assume that the relevant stable distribution is symmetric.) Fielitz and Roselle (1983) incorporate skewness and changing scale parameters and find that nonnormal stable distributions with changing scale are descriptive of the underlying distribution if a nontrivial amount of skewness is present.

While many studies test for stability by estimating the characteristic exponent, it is possible to test for stability by using parametric distributional tests based on the standardized sample moments. In particular, if the sample size is 50 or greater, estimated kurtosis can provide an effective goodness of fit test for identifying stability (Saniga and Hayya (1977), and Bera and Mackenzie (1986)). In addition, a further heuristic test for stability can be based on examining the variation in the estimates of the standard deviation, skewness, and kurtosis across the subsamples. Erratic behavior across samples in the estimated moments is expected if the data has originated from a stable distribution with relatively small (e.g., 1.5) characteristic exponent (Officer (1972)).

The convergence behavior of estimated distribution parameters under temporal aggregation provides additional information. More precisely, if each term in a sum of independent random variables is stable with characteristic exponent α, then the sum is stable with characteristic exponent α. A similar result holds for the skewness parameter. Hence, identical, independent stable distributions observed for different differencing intervals differ only by a scale factor. This stable property is tested by examining the behavior of the estimated moments for the various transformations under temporal aggregation. In the *fps* case, if either futures prices or implied carry costs are stable then *fps* should be stable also. Similarly, for *tfps*, if *ic* is stable then *tfps* should be stable. Evidence of convergence in the estimated parameters under temporal aggregation is contrary to the hypothesis that the relevant variable is *iid* stable.[8]

[7]RFPS is equivalent to using the log difference transformations of *fps*.

[8]Only the *iid* stable hypothesis is being tested. More general forms of the stable distribution which, for example, incorporate parameter evolution are not being directly examined.

Heteroskedastic Normal

The final set of distributional tests considered here are for heteroskedastic normal processes. The attraction of this particular distributional alternative is based on the theoretical and statistical importance of normality. In particular, it is possible to explain peakedness and leptokurticity within the framework provided by the normal distribution by introducing some form of heteroskedasticity. The general problem can be heuristically specified by assuming that some appropriate transformation of the prices, say y, has a conditional normal distribution, i.e.:

$$F[y|\sigma^2] = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{\frac{-y^2}{2\pi\sigma^2 t}\right\}$$

In introducing heteroskedasticity, the variance is specified to evolve according to a probability law, $G[\sigma^2]$. For any unit time interval ($\Delta t = 1$), it follows that the unconditional distribution for y is:[9]

$$H[y] = \int_0^\infty F[y|\sigma^2]G[\sigma^2]d\sigma^2$$

$H[y]$ is the distribution that is observed in calendar time. It follows, trivially, that if $G[\sigma^2]$ is a constant then $H[y]$ will be homoskedastic normal.

Theoretically, a number of different motivations have been proposed to justify the use of the heteroskedastic model. Following Praetz (1972), Hsu (1984), and Poitras (1985) this may be associated with periods of price volatility and price tranquility, i.e., $G[\sigma^2]$ is a step or jump function producing observed distributions which are mixtures of normals with different variances. In this case, periods of constant variance are usually of significant length with, in the normal case, leptokurticity a result of adding two distributions with different variances. However, it is not necessary that one or even both mixing distributions be normal. If the price change distributions are *iid* within the different periods, then this will be reflected in both the distribution tests and temporal aggregation behavior for the relevant subsamples.

Other explanations for heteroskedasticity introduce more variation into $G[\sigma^2]$.[10] However, it can be shown that if $G[\sigma^2]$ is a well-behaved function, then widening the differencing interval should reduce the dispersion of $G[\sigma^2]$ thereby producing some evidence of convergence. Unfortunately, the general nature of the problem

[9]The limits of integration are defined by σ^2 which must be positive and finite. In addition, because the random variables involved here are time series, time is also an argument in the distributions. However, the time subscript is suppressed for convenience.

[10]For example, it is possible that the individual, independent information events which add together to produce a daily price change do not evolve according to calendar time (e.g., Mandelbrot and Taylor (1967), and Clark (1973)). The occurrence of information events could be directed by transactions volume, transaction time, open interest (for futures) or some other suitable variable. Epps and Epps (1976) and others propose a complementary approach based on revisions of trader's beliefs. The resulting unconditional process for the price change is usually referred to as a subordinated process. When the conditional process is normal, the unconditional process is referred to as subordinated normal. Strong empirical evidence is presented in favor of specific versions of both the subordinated process approach (e.g., Clark (1973), Praetz (1972), Blatberg and Gonedes (1974), and Kon (1984)) and the related transactions volume approach (e.g., Tauchen and Pitts (1983) and Harris (1984)). In practice, the properties of the distribution for a subordinated process under temporal aggregation depends on the assumptions made about the conditional price change process and the directing process.

dictates that further theoretical structure for $G[\sigma^2]$ and $F[\bullet]$ is required to identify specific convergence implications. For example, a recent development of the heteroskedastic model assumes that variance evolution is generated by some form of ARMA process, i.e., $G[\sigma^2]$ follows an ARCH or GARCH process (e.g., Engle (1982), Engle and Bollerslev (1986), and Diebold (1987)).[11] Fortunately, because of the specific structure of $G[\sigma^2]$, it is possible to derive distributional convergence properties under temporal aggregation with conditional normality. Specifically, ARCH models and other representations of m-dependent error processes constitute a set of models where convergence is expected (Diebold (1987)).[12] In these cases, because $G[\sigma^2]$ follows some form of "regular" time series process, as the differencing interval increases beyond the length of the assumed time series process for the variance, then the process will converge to the normal.

EMPIRICAL RESULTS[13]

The distributional tests for *fps*, presented in this section, provide the benchmarks required to determine the relative contributions of the component distributions for futures prices and carry returns in determining the *fps* distribution. The results for futures prices and carry returns are given in the next section. To contain the number of reported empirical results, it is necessary to systematically restrict the variables under consideration. For example, a key element in defining *fps* is the length of time between the delivery periods for the two spread legs. There are numerous possible combinations of the available contract maturities for defining *fps*. For example, two month spreads could be composed by combining contracts for: February–April, April–June, June–August, August–October, October–December, and December–February Because it is not feasible to present all possible *fps* results, only a representative subset of all available contract combinations is presented. Specifically, three spread lengths are examined, centered on the December delivery contract: 1 year (Dec–Dec), 6 months (Dec–June), and 2 months (Feb–Dec).[14]

Since there is only one December delivery per year, the focus on the December contract means that for a given spread length there is only one set of results to report per calendar year. Each sample starts when the deferred contract registers a continuous amount of volume and open interest and the sample ends two weeks prior to the first delivery date on the nearby contract. By construction, the results presented have overlapping samples because the relevant *fps* are outstanding for more than one year. For example, consider the subsamples for *fps* constructed using all the December and June contracts. Overlapping occurs because the sample for

[11]While it is possible to interpret the ARCH model as a type of subordinated process, the focus of the ARCH approach is concerned with modelling volatility rather than with identifying the nature of the directing process.

[12]Unfortunately, Diebold (1987) only considers the asymptotic result where the differencing interval goes to infinity. While the result that the ARCH model converges to normality in this case is interesting theoretically, the result provides little information about the speed of convergence.

[13]The gold futures price data used are the daily COMEX settlement prices provided by the Center for the Study of Futures Markets. The full sample runs from January 4, 1980 to March 12, 1985. The sample length results in the 1981 subsamples being about half the size of the subsamples for other years. The *fps* results reported here rely on settlement price quotes and not the spread quotes originating from the spread trader in the pit. While this discrepancy may create problems for intra-day spread studies, for present purposes deviations between closing price quotes for separate delivery months and the spread trader's closing spread quote for those same months will be effectively zero.

[14]These results are representative of those for *fps* defined using different contract months.

the Dec 84/June 84 spread begins at the end of August 1983, when the Dec 84 contract first began continuous trading. This subsample overlaps the subsample for the Dec 83/June 83 spread which ceases at the end of November 1983. Overlapping samples are useful in identifying certain types of jumps in the generating process.[15]

Table I presents the parametric distributional test results for convergence under temporal aggregation. Three different *fps* transformations, DFPS, RFPS, and DTL, are examined. Even though only the 6 month Dec/June spread results are presented, the results are similar for all spread lengths. Widening the differencing interval from daily to weekly effectively eliminates the preponderance of zero and small changes, producing more "normal" distributions. This result is important for determining the relative contributions of Δic and ΔF in determining Δfps. In addition to the temporal aggregation effect, there is evidence for two forms of nonstationarity in both the daily and weekly results. In particular, the combination of significantly negative skewness combined with fat-tails for '81 and '84 subsamples indicates discontinuity in the underlying generating process during those periods. In addition, the reduction in σ over the full, '81–'85 sample is further evidence of nonstationarity.

Table I
DISTRIBUTION FOR DEC–JUNE FPS AND TFPS UNDER TEMPORAL AGGREGATION

Transformation	$\sqrt{\beta_1}$	β_2	LM	SR	σ
Sample: '85 Nearby (June) Delivery					
Frequency: DAILY NOB = 337					
DFPS	−.057	0.997*	14.2*	6.86*	2.62
RFPS	−.137	.792*	9.87*	6.75*	.0118
DTL	−.101	.484*	3.87	6.35	2.34
Frequency: WEEKLY (Friday) NOB = 65					
DFPS	.252	.298	0.93	4.83	7.25
RFPS	.283	.030	0.87	4.48	.0316
DTL	.104	−.123	0.16	5.12	5.79
Sample: '84 Nearby (June) Delivery					
Frequency: DAILY NOB = 334					
DFPS	−1.13*	15.6*	3462*	13.5*	4.58
RFPS	−1.14*	13.3*	2520*	12.6*	.0185
DTL	−.706*	7.95*	912*	9.81*	4.42
Frequency: WEEKLY (Friday) NOB = 62					
DFPS	−1.83*	7.46*	178*	7.18*	12.0
RFPS	−1.42*	4.75*	79.2*	6.21*	.0457
DTL	−.91*	5.40*	84.1*	7.54*	10.5

*Indicates significance at the 10% level. $\beta_2 = 0$ under the null hypothesis of normality, i.e., β_2 is centered about the origin. LM is the asymptotic chi-squared test which combines skewness and kurtosis. Significance levels have been calculated using Bera and Mackenzie adjusted critical values. SR is the studentized range. σ is standard deviation.

$$\text{DFPS} = fps(t+1) - fps(t) \qquad \text{RFPS} = (\text{DFPS}/fps(t)) \qquad \text{DTL} = tfps(t+1) - tfps(t)$$

[15]The use of subsamples rather than a continuous series also partially mitigates the effects of mean nonstationarity. This follows because a different mean value is calculated for each subsample.

Table I
DISTRIBUTION FOR DEC–JUNE FPS AND TFPS UNDER TEMPORAL AGGREGATION continued

Transformation	$\sqrt{\beta_1}$	β_2	LM	SR	σ
Sample: '83 Nearby (June) Delivery					
Frequency: DAILY NOB = 333					
DFPS	.181	4.91*	336*	9.93*	6.75
RFPS	.261*	10.6*	1551*	12.4*	.0275
DTL	−.122	6.14*	523*	10.8*	5.90
Frequency: WEEKLY (Friday) NOB = 62					
DFPS	.006	−.240	0.15	4.32	14.3
RFPS	.247	.012	0.63	4.65	.0517
DTL	.440	.366	2.35	4.90	13.1
Sample: '82 Nearby (June) Delivery					
Frequency: DAILY NOB = 330					
DFPS	−.172	2.16*	66.2*	7.66*	9.53
RFPS	−.230*	1.53*	34.9*	6.90*	.0232
DTL	−.315*	2.49*	91.0*	7.64*	9.03
Frequency: WEEKLY (Friday) NOB = 62					
DFPS	.184	−.146	0.41	4.74	25.7
RFPS	.056	.075	0.05	4.96	.0615
DTL	−.019	−.012	0.004	4.96	19.8
Sample: '81 Nearby (June) Delivery					
Frequency: DAILY NOB = 167					
DFPS	−7.96*	91.3*	59757*	14.9*	37.6
RFPS	−3.66*	49.9*	17725*	15.2*	.0602
DTL	−4.38*	50.4*	18228*	14.4*	31.2
Frequency: WEEKLY (Friday) NOB = 29					
DFPS	−3.15*	12.2*	229.1*	5.77*	76.3
RFPS	−1.99*	6.18*	65.2*	5.30	.1259
DTL	−3.24*	12.6*	243.3*	5.79*	57.5*

Regarding the distributional affect of spread length, *a priori* it is expected that the shorter the spread length, the more likely that daily results will be dominated by zeroes and, hence, appear to be peaked and fat-tailed. Examining Table II which contains the weekly results for the Feb–Dec (2 month) and Dec–Dec (1 year) spread lengths this expectation is not supported. Combined with the information from Table I on the 6 month spread, it is apparent that there is little difference in *fps* distributions across spread lengths. The most significant difference identified is that *fps* volatility is found to increase directly with spread length. This result corroborates previous studies (e.g., Castelino and Vora (1984)). However, as indicated by the results for the RFPS transformation, the spread length effect does not apply when the *fps* is appropriately scaled.

In defining a weekly difference, it is necessary to select a specific day of the week over which the difference is to be taken. To verify that the weekly results are not contingent on the specific day selected, Table III provides the results for *fps* first differences taken over each day of the week. The Dec–Dec case reported in

Table III is generally representative of the results for other spread lengths. Examination of the distributional test results reveals there is little evidence of a consistently significant weekend effect across the full sample. However, there are a number of anomalous instances where the results for either a Monday or Friday result differed significantly from the results for the other days of the week. For example, the nonstationarity reported for the '84 sample is much more pronounced on Friday than for the other days of that subsample.

Table II
DISTRIBUTIONS FOR FEB–DEC, DEC–JUNE AND DEC–DEC SPREADS: WEEKLY (FRIDAY)

Transformation	$\sqrt{\beta_1}$	β_2	LM	SR	σ
Sample: '85 December Delivery					
Spread: DEC–DEC	NOB = 38				
DFPS	−.028	.632	0.64	5.10*	16.3
RFPS	.077	−.137	0.07	4.56	.0329
DTL	.035	−.914	1.33	3.95	13.9
Spread: FEB–DEC	NOB = 74				
DFPS	−.031	1.02*	3.25	5.81*	2.41
RFPS	.065	.925*	2.69	5.76*	.0322
DTL	.318	.608	2.39	5.50*	1.94
Sample: '84 December Delivery					
Spread: DEC–DEC	NOB = 38				
DFPS	−1.82*	6.99*	98.2*	6.22*	18.0
RFPS	−1.42*	5.07*	53.4*	5.98*	.0379
DTL	−.837	1.58*	8.38*	5.02*	13.1
Spread: FEB–DEC	NOB = 76				
DFPS	−.221	1.67*	9.40*	5.84*	4.28
RFPS	−.053	1.07*	3.67	5.82*	.0525
DTL	−.106	4.05*	66.3*	7.07*	3.16
Sample: '83 December Delivery					
Spread: DEC–DEC	NOB = 38				
DFPS	0.215	.255	0.40	4.75	26.7
RFPS	.385	.367	1.15	4.94	.0540
DTL	.440	.694	1.99	4.84	25.7
Spread: FEB–DEC	NOB = 77				
DFPS	−.164	.364	0.77	4.80	5.62
RFPS	−.009	−.017	.002	4.87	.0574
DTL	.221	.156	0.71	5.18	4.29
Sample: '82 December Delivery					
Spread: DEC–DEC	NOB = 37				
DFPS	.414	−.309	1.20	4.46	41.5
RFPS	.251	.096	0.40	4.83	.0588
DTL	.323	.526	1.07	4.98	39.1

*Indicates significance at the 10% level. $\beta_2 = 0$ under the null hypothesis of normality, i.e., β_2 is centered about the origin. LM is the asymptotic chi-squared test which combines skewness and kurtosis. Significance levels have been calculated using Bera and Mackenzie adjusted critical values. SR is the studentized range. σ is standard deviation.

DFPS = $fps(t+1) - fps(t)$ RFPS = (DFPS/$fps(t)$) DTL = $tfps(t+1) - tfps(t)$

Table II
DISTRIBUTIONS FOR FEB–DEC, DEC–JUNE AND DEC–DEC SPREADS: WEEKLY (FRIDAY) continued

Transformation	$\sqrt{\beta_1}$	β_2	LM	SR	σ
Spread: FEB–DEC NOB = 70					
DFPS	−.183	−.012	0.39	4.99	9.81
RFPS	−.40	.010	1.86	4.55	.0748
DTL	.052	−.493	0.74	4.46	7.72
Sample: '81 December Delivery					
Spread: DEC–DEC NOB = 25					
DFPS	−3.18*	11.3*	175*	5.29*	158.5
RFPS	−2.05*	5.79*	52.3*	5.02*	.1302
DTL	−3.30*	11.9*	192*	5.34*	117.9
Spread: FEB–DEC NOB = 34					
DFPS	−2.63*	9.35*	134.5*	5.84*	29.1
RFPS	−1.52*	4.10*	30.4*	5.36*	.1484
DTL	−2.47*	8.10*	105.0*	5.45*	20.0

Table III
DISTRIBUTION FOR RFPS TRANSFORMATION FOR THE DEC–DEC SPREAD: BY DAY OF THE WEEK

Day of the Week	$\sqrt{\beta_1}$	β_2	LM	SR	σ
Monday	−.217	.011	0.28	4.79	.0324
Tuesday	−.608*	.065	2.41	4.54	.0335
Wednesday	−.232	−.098	0.36	4.77	.0331
Thursday	−.940*	.474	5.95*	4.36	.0347
Friday	.077	−.137	0.07	4.56	.0329
Sample: '84 December Delivery					
Monday	−.277	.348	0.62	4.69	.0317
Tuesday	−.090	1.34*	2.97	5.40*	.0263
Wednesday	−.009	1.75*	4.99*	5.78*	.0284
Thursday	−.071	−.434	0.32	4.59	.0264
Friday	−1.42*	5.07*	53.4*	5.98*	.0379
Sample: '83 December Delivery					
Monday	.335	−.433	0.93	3.94	.0595
Tuesday	.841*	.326	4.65	4.20*	.0635
Wednesday	.560*	.897	3.26	5.03*	.0580
Thursday	.730*	2.44*	12.5*	5.75	.0650
Friday	.385	.367	1.15	4.93	.0540
Sample: '82 December Delivery					
Monday	−.081	−.529	0.46	4.15	.0561
Tuesday	.134	−.263	0.22	4.67	.0535
Wednesday	.129	−.044	0.11	4.60	.0584
Thursday	.289	.298	0.67	4.94	.0611
Friday	.251	.096	0.40	4.82	.0588

*See notes to Table II.

Table III
DISTRIBUTION FOR RFPS TRANSFORMATION FOR THE DEC–DEC SPREAD: BY DAY OF THE WEEK continued

Day of the Week	$\sqrt{\beta_1}$	β_2	LM	SR	σ
Sample: '81 December Delivery					
Monday	.053	−.308	0.11	4.35	.0932
Tuesday	−2.26*	6.54*	65.7*	5.27*	.1386
Wednesday	−2.74*	9.62*	138*	5.47*	.1302
Thursday	−2.18*	6.60*	67.9*	5.19*	.1302
Friday	−2.04*	5.79*	52.3*	5.02*	.1302

RESULTS FOR CHANGES IN CARRY RETURNS AND PRICES

Given the benchmark information on *fps*, the relative contributions of the component distributions for futures prices and carry returns in determining the *fps* distribution can be addressed now by examining certain empirical properties of these distributions. Temporal aggregation results for the first difference and rate of gain transformations of implied carry returns and futures prices are given in Tables IV–V. To make these results compatible with the specific contract configurations given in Tables I–III, carry returns have been derived using the Dec–June contracts. The futures price results are for the December contract which is used in deriving the *fps* and carry returns. As in the *fps* case, overlapping samples are used. In addition, the focus on the December delivery means that there is only one set of empirical results presented per calendar year.

Examining Tables IV–V reveals significant differences in the results of the distributional tests for implied carry and futures prices. For implied carry, there is definite evidence of convergence to normality when the differencing interval is widened from daily to weekly. For three of the five samples, there is complete convergence to normality. In the other two samples (81 and 84), a substantial reduction in test values is observed, even though these samples indicate the same type of nonstationarity which appears in the *fps* case. These results do not change significantly when a different spread length is used to define the implied carry. In addition, the distributional tests do not appear to be significantly affected by the day of the week selected to calculate the weekly results or by the transformation selected. Finally, as in the *fps* case, there is significant reduction in volatility throughout the full sample indicating variance nonstationarity over the 1981–85 period.

Unlike the implied carry case, the futures prices results provide little evidence of consistent convergence behavior for the estimated parameters under temporal aggregation.[16] While, in some samples, there is a noticeable reduction in the value of the LM statistic, there are samples where the LM value increases. With the excep-

[16]The use of futures rather than spot gold prices raises the "declining maturity" problem. Specifically, because gold is a full carrying charge market, futures prices are generally greater than the spot price by the amount of carry, i.e., $F(t,T) = S(t)\ (1 + ic(t,T\text{-}t))$. As a result, the change in the futures prices is dependent on both spot price changes and changes in implied carry, i.e., $\Delta F(t,T) = (1 + ic(t,T\text{-}t))\Delta S + S(t + 1)\ \Delta ic$. This result indicates that the $ic(t,T\text{-}t)$ term declines as the contract approaches maturity biasing ΔF downward at the end of the sample. This effect distorts the correspondence between results for distributional tests based on spot and futures. Given this difference, futures prices are selected instead of spot prices because the objective is to study the behavior of *fps* not gold price levels. Inspection of (2) reveals that the distribution for *fps* depends on changes in futures price levels.

Table IV
DISTRIBUTIONS FOR IMPLIED CARRY DEFINED WITH DEC–JUNE SPREADS UNDER TEMPORAL AGGREGATION

Transformation	$\sqrt{\beta_1}$	β_2	LM	SR	σ
Sample: '85 Nearby (June) Delivery					
Frequency: DAILY NOB = 337					
DIC	−.241*	.695*	10.1*	6.19	.0006
RIC	−.188	.549*	6.2*	6.33	.0106
Frequency: WEEKLY (Friday) NOB = 65					
DIC	−.132	−.170	0.27	5.00	.0015
RIC	−.005	−.157	0.07	5.16	.0261
Sample: '84 Nearby (June) Delivery					
Frequency: DAILY NOB = 334					
DIC	−.952*	8.68*	1099*	10.1*	.0009
RIC	−.808*	7.71*	865*	9.83*	.0174
Frequency: WEEKLY (Friday) NOB = 62					
DIC	−1.10*	5.31*	85.2*	7.39*	.0022
RIC	−.683*	3.15*	30.4*	6.76*	.0398
Sample: '83 Nearby (June) Delivery					
Frequency: DAILY NOB = 333					
DIC	−.179	5.73*	457*	11.1*	.0014
RIC	.272*	17.5*	4243*	14.8*	.0236
Frequency: WEEKLY (Friday) NOB = 62					
DIC	.405	.073	1.71	4.92	.0031
RIC	.591*	.878	5.60*	5.48*	.0490
Sample: '82 Nearby (June) Delivery					
Frequency: DAILY NOB = 330					
DIC	−.297*	1.50*	36.0*	7.28*	.0015
RIC	−.173*	1.25*	23.1*	6.67*	.0209
Frequency: WEEKLY (Friday) NOB = 62					
DIC	−.152	−.019	0.24	4.93	.0035
RIC	−.025	−.069	1.24	5.42*	.0504
Sample: '81 Nearby (June) Delivery					
Frequency: DAILY NOB = 167					
DIC	−5.99*	66.6*	3186*	14.6*	.0048
RIC	−1.23*	32.9*	7561*	14.6*	.0598
Frequency: WEEKLY (Friday) NOB = 29					
DIC	−3.38*	13.3*	270*	5.79*	.0088
RIC	−2.61*	9.03*	131*	5.49*	.1008

*Indicates significance at the 10% level. $\beta_2 = 0$ under the null hypothesis of normality, i.e., β_2 is centered about the origin. LM is the asymptotic chi-squared test which combines skewness and kurtosis. Significance levels have been calculated using Bera and Mackenzie adjusted critical values. SR is the studentized range. σ is standard deviation.

$$\text{DIC} = ic(t+1) - ic(t) \qquad \text{RIC} = (\text{DIC}/ic(t))$$

Table V
DISTRIBUTIONS FOR FUTURES PRICE CHANGES UNDER TEMPORAL AGGREGATION

Transformation	$\sqrt{\beta_1}$	β_2	LM	SR	σ
Sample: '85 December Delivery					
Frequency: DAILY NOB = 337					
DPR	−.169	2.82*	113*	8.64*	43.5
RPR	−.139	2.51*	89.8*	8.57*	.0103
Frequency: WEEKLY (Friday) NOB = 65					
DPR	−.118	1.37*	5.21*	6.18*	90.4
RPR	−0.14	1.39*	5.43*	6.20*	.0212
Sample: '84 December Delivery					
Frequency: DAILY NOB = 334					
DPR	−.151	1.66*	39.5*	5.84	85.6
RPR	−.027	1.42*	27.9*	6.01	.0173
Frequency: WEEKLY (Friday) NOB = 62					
DPR	−.541*	5.14*	71.4*	7.62*	158.2
RPR	−.164	3.94*	40.4*	7.30*	.0310
Sample: '83 December Delivery					
Frequency: DAILY NOB = 333					
DPR	.218	.670*	8.87*	5.55	90.1
RPR	.320*	.667*	11.8*	6.00	.0200
Frequency: WEEKLY (Friday) NOB = 62					
DPR	.831*	1.04*	9.96*	4.95	176.8
RPR	1.09*	1.89*	21.4*	5.30	.0418
Sample: '82 December Delivery					
Frequency: DAILY NOB = 330					
DPR	−.130	.180	1.37	4.78*	104.6
RPR	−.052	.001	0.15	5.66	.0166
Frequency: WEEKLY (Friday) NOB = 62					
DPR	−.178	.365	0.67	5.26	231.7
RPR	−.009	−.210	0.11	5.08	.0361
Sample: '81 December Delivery					
Frequency: DAILY NOB = 167					
DPR	−2.25*	16.2*	1971*	9.65*	217.5
RPR	−2.22*	15.8*	1864*	9.62*	.0276
Frequency: WEEKLY (Friday) NOB = 29					
DPR	−1.03*	1.71*	8.73*	4.88	376.7
RPR	−.777*	1.10*	4.38*	4.86	.0501

*Indicates significance at the 10% level. $\beta_2 = 0$ under the null hypothesis of normality, i.e., β_2 is centered about the origin. LM is the asymptotic chi-squared test which combines skewness and kurtosis. Significance levels have been calculated using Bera and Mackenzie adjusted critical values. SR is the studentized range. σ is standard deviation.

$$\text{DPR} = F(t+1,T) - F(t,T) \qquad \text{RPR} = (\text{DPR}/F(t,T))$$

tion of the 1982 sample, the null hypothesis of normality is rejected in all cases. This result is consistent across the two transformations examined. In addition, the relative performance of the two transformations examined is not markedly different. When weekly results are compared using different days to define the weekly differencing interval (not reported), the results indicate somewhat more volatility on Mondays and somewhat less volatility on Fridays. However, there appears to be less evidence of day-of-the-week effects than has been reported elsewhere (e.g., Ball, Tourous, and Tschoegl (1982)). Finally, while there is continuous reduction in volatility over time in DPR but not in RPR, both transformations exhibited significant variance nonstationarity over the full 1981–85 sample.

Turning to the nature of the distribution, the *fps* results are similar to the *ic* results. Across all samples some degree of convergence is noticeable with complete convergence achieved in three subsamples with evidence of partial convergence in the same other two samples. Hence, in (2), the Δic term dominates the ΔF term. The tailed spread results reinforce the key role that implied carry plays in determining *fps* distributions. By construction, the tailed spread eliminates the effect of changes in price levels on the spread. The resulting randomness is due to the (scaled) effect of implied carry. Comparing the standard deviations of the tailed and untailed first differences reveals a relatively small (10–25%) reduction in volatility when the spread is tailed. Hence, the bulk of *fps* volatility arises from changes in implied carry. In addition, tailing does not significantly alter the distributional shape, for either the daily or weekly samples.

While the distributional tests provide strong evidence on the role of *ic* and the lack of *iid* behavior, conclusions about the specific form of the generating process for *fps* (and *ic*) are not so readily identified. Of the distributions under consideration, only the variance evolution models are consistent with convergence under temporal aggregation.[17] However, the continuous reduction in volatility over the 1981–85 sample and the lack of complete convergence for two of the subsamples argues against constancy in the generating process over the full sample. In other words, the generating process is characterized by stationary periods of price tranquility and nonstationary turbulent periods. The variance evolution process for tranquil periods is decidedly "more normal" than for turbulent periods. Applying the mixtures of distribution approach to the variance evolution models has promise. However, specific structure must be imposed to explain the reduction in variance throughout the full sample.

SUMMARY

Despite playing important roles in a number of markets, the distributions of traded price combinations are not well studied. One case in point is the futures spread: the traded difference between futures prices for different delivery months. This study demonstrates that while the distribution of gold futures spreads is theoretically dependent on the distributions for futures prices and implied carry returns, empirical evidence indicates that the *fps* distribution exhibits properties similar to the *ic* distribution alone. By examining the behavior of parametric distribution tests under temporal aggregation, the hypothesis of *iid* behavior for either *fps* or *ic* is rejected. While a specific generating process is not identified, variance nonstationarity in the form of a reduction in variance over the full 1981–85 sample is observed indicating

[17]This does not rule out the possibility of non-*iid* versions of, say, the stable distribution being the most feasible return generating process.

that the variance evolution process may be affected by price levels. Finally, only weak evidence is provided for a weekend effect in the distribution of *fps*.

Bibliography

Ball, C., Tourous, W., and Tschoegl, A. (1982): "Gold and the Weekend Effect," *Journal of Futures Markets,* 175–82.

Bera, A., and MacKenzie, C. (1986, March): "Tests for Normality with Stable Alternatives," *Journal of Statistical Computation and Simulation,* 25:37–52.

Blattberg, R. C., and Gonedes, N. (1984, April): "A Comparison of the Stable-Paretian and Student Distribution as Statistical Models for Prices," *Journal of Business,* 244–280.

Castelino, M., and Vora, A. (1984, Spring): "Spread Volatility in Commodity Futures: The Length Effect," *Journal of Futures Markets,* 4:39–46.

Cornew, R., Town, D. and Crowson, L. (1984): "Stable Distributions, Futures Prices and the Measurement of Trading Performance," *Journal of Futures Markets.*

Diebold, F. (1987, August): "Temporal Aggregation of ARCH Processes and the Distribution of Security Returns," unpublished Board of Governors discussion article.

Engle, R. (1982): "Autoregressive Conditional Heteroskedasticity with estimates of UK Inflation," *Econometrica,* 50:987–1008.

Engle, R., and Bollerslev, T. (1986): "Modelling the Persistence of Conditional Variance," *Econometric Reviews,* 5:1–50.

Epps, Thomas W., and Epps, Mary Lee. (1976, March): "The Stochastic Dependence of Security Price Changes and Transaction Volume: Implications for the Mixture of Distributions Hypothesis," *Econometrica,* 305–321.

Fama, E. F. (1965): "The Behaviour of Stock Market Prices," *Journal of Business,* 38:34–105.

Fama, E. F., and Roll, R. (1971, June): "Parameter Estimates for Symmetric Stable Distributions," *Journal of the American Statistical Association,* 66:331–338.

Fielitz, B. D., and Rozelle, J. P. (1983, March): "Stable Distributions and the Mixtures of Distributions Hypotheses for Common Stock Returns," *Journal of the American Statistical Association,* 78:28–36.

Hagerman, R. L. (1978, September): "More Evidence on the Distribution of Security Returns," *Journal of Finance,* 28:1214–1221.

Harris, L. (1986, March): "Cross-Security Tests of the Mixture of Distributions Hypothesis," *J.F.Q.A.,* 21:39–46.

Hsieh, D. (1988, February): "The Statistical Properties of Daily Foreign Exchange Rates: 1974–83," *Journal of International Ecomonics,* 24:129–45.

Hsu, D. A. (1984, March): "The Behaviour of Stock Returns: Is it Stationary or Evolutionary?," *Journal of Financial and Quantitative Analysis,* 19:11–28.

Hsu, D. A., Miller, R. B., and Wichern, D.W. (1974, March): "On the Stable Paretian Behaviour of Stock Market Prices," *Journal of the American Statistical Association,* 69:108–113.

Jones, F. (1981): "Tails, Turtles and All That," *Journal of Futures Markets.*

Knight, J. (1986, April): "Non-Normal Errors and the Distribution of OLS and 2SLS Structural Estimators," *Econometric Theory,* 2:75–105.

Kon, S. (1984, March): "Models of Stock Returns—A Comparison," *Journal of Finance,* 39:147–166.

Mandelbrot, B. (1963, October): "The Variation of Certain Speculative Prices," *Journal of Business,* 36:394–419.

Mandelbrot, B., and Taylor, H. (1967): "On the Distribution of Stock Price Differences," *Operations Research,* 15:1057–1062.

McFarland, J., Pettit, R., and Sung, S. (1982): "The Distribution of Foreign Exchange Price Changes," *Journal of Finance,* 37:693–715.

Osborne, M. F. M. (1959, March): "Brownian Motion in the Stock Market," *Operations Research,* 7:145–73.

Pearson, E., D'Agostino, R., and Bowman, K. (1977): "Tests for Departures from Normality: Comparison of Powers," *Biometrika,* 64:321–346.

Poitras, G. (1985): "A Study of Gold Futures Price Spreads," unpublished Ph.D. thesis, Columbia University.

Poitras, G. (1990): "Testing Regression Disturbances for Normality with Stable Alternatives: Further Monte Carlo Evidence," *Journal of Statistical Computation and Simulation,* forthcoming.

Rentzler, J. (1986): "Trading Treasury Bond Spreads Against Treasury Bill Futures--A Model and Empirical Test of the Turtle Trade," *Journal of Futures Markets,* 6:41–61.

Saniga, E., and Hayya, J. (1977, June): "Simple Goodness-of-Fit Tests for Symmetric Stable Distributions," *J.F.Q.A.,* 276–289.

So, J. (1987, February): "The Sub-Gaussian Distribution of Currency Futures: Stable Paretian or Nonstationary," *Review of Economics and Statistics,* 100–107.

Stram, D., and Wei, W. (1986): "Temporal Aggregation in the ARIMA Process," *Journal of Time Series Analysis,* 7:279–292.

Tauchen, G., and Pitts, M. (1983, March): "The Price-Variability-Volume Relationship on Speculative Markets," *Econometrica,* 51:485–505.

Teichmoeller, J. (1971, June): "A Note on the Distribution of Stock Price Changes," *Journal of the American Statistical Association,* 66:282–284.

Weiss, A. (1984): "Systematic Sampling and Temporal Aggregation," *Journal of Econometrics,* 26:271–281.

White, H., and Macdonald, G. (1980, March): "Some Large-Sample Tests for Nonnormality in the Linear Regression Model," *Journal of the American Statistical Association,* 75: 16–27.

Yano, A. (1989, June): "Configurations for Arbitrage using Financial Futures," *Journal of Futures Markets.*

Zolotarev, V. (1986): *One Dimensional Stable Distributions,* Providence, R.I., American Mathematical Society.