Logistic
Research & Trading Inc.

Gold: Volatility and Options,
A Reconciliation

Joel C. Gibbons
President
Logistic Research & Trading Co.
Elmhurst, Illinois, 60126

and

Stuart School of Business
The Illinois Institute of Technology
Chicago, Illinois, 60661

January 14, 1998

Preliminary Draft: Not to
be Cited Without Permission

Abstract

Gold, as a financial asset, recommends itself to the student of option pricing for several compelling reasons. Unlike almost all financial instruments, the definition of the asset is objectively verifiable and has remained essentially unchanged for centuries, and as a result, there is an abundance of data against which we can test competing theories. It is also free of distracting cash flows to either long or short positions, and with futures the cash flows simply accrue at the risk free rate of interest. In this article we will investigate the statistical properties of gold in an attempt to reconcile them with the valuation of listed options on gold. We find that in some important respects option valuations seem to be inconsistent with the actual cash market for the metal. In other respects they are not clearly inconsistent, but they suggest some interesting properties of the gold market which would be necessary to reconcile the prices of gold and of options on gold.

Two distinct questions arise in regard to the pricing of options on financial assets. One question is normative, and deals with how to price options efficiently given the actual stochastic – and deterministic, if appropriate – behavior of the underlying asset. The other question is inherently empirical, addressing as it does the correlation between actual option pricing and the price behavior of the underlying asset. This paper having its roots in the empirical issues, attempts to reconcile actual option pricing with observed asset prices. Obviously, we can not tackle so broad a task for a wide range of assets, and in fact we will limit our attention to the gold market.

Gold is unique among financial assets in a couple of respects. First, it is by far the oldest of all financial assets. Since ancient times it has constituted a unique transnational currency which provided irreplaceable services in an age before international electronic banking. Even today there are a few persons in the world who pack up their business every night in a pickup truck or on the back of a dromedary, and who must carry all their worldly wealth with them wherever they go. Gold and gem stones are the only commodities which have a high enough value per ounce to make that possible. And gold is preferable even to gems, because its valuable characteristics are easily verified by even comparatively unsophisticated parties. Only an expert jeweler, by contrast, can detect flaws which make the difference between a unique gem and an interesting rock. Despite the long history of monetary uses, the usable actual pricing history in dollars is much more limited, because the price of gold was fixed arbitrarily in dollars from 1932 into 1971. The price explosion which followed the relaxation of restrictions is convincing testimony to the effectiveness of price controls, and resulted in a long and ultimately painful period of adjustment in the 1970's. Through the 1970's the rate of return of gold was very large, but since early 1980 its nominal rate of return has been negative. The price data at our disposal covers the quarter century from January, 1973 to May, 1997, amounting to about 6100 observations.

The relaxation of restrictions on owning and trading gold is only one of several fundamental changes in the gold market over the last twenty-five years. A second change has been the popularization of gold as an asset for small retail investors. The conventional unit of gold bullion is a bar weighing 1200 grams, worth about $120,000 today. This is a physical unit of very limited appeal. The minting of ceremonial gold coins of one ounce weight and the introduction of trading in gold futures on the Comex exchange have brought to a mass market the opportunity to participate. Two competing trends weigh heavily against the appeal of gold as an asset class. One of them is the spread of sophisticated modern banking practices which have brought competitive market yields to the small investor. At the same time, modern central banks have adopted a policy of strict monetarism, thereby greatly reducing the risk of run away inflation. These are only a few of important but the slow moving trends which have had an impact on the gold market in recent times.

It would far exceed the scope of this paper to try to summarize all the large, permanent changes which have had an impact on the supply or demand for gold. More pointedly, it would be inappropriate to do so, because we want to adopt the point of view that trading in gold, and especially the market clearing price of gold, reflect all those forces. We want to defer to the prices to characterize the history of supply and demand. Our casual summary of the history does, however, reveal supply and demand shocks which do not obviously bespeak a stationary process. The fact that gold has been quoted in a open auction market for that time presents us with a first rate opportunity to view a market adapting to structural change.

The Gold Data Set.

Our data on the price of gold consists of daily closing levels of spot gold over the period from January 2, 1973 to May 6, 1997. Until 1994, the data comes from the London afternoon fixing; after the start of 1994 it is the Handy and Harman daily quotation. While these two sources do not have to agree perfectly, in practice the discrepancy is insignificant, and we simply treat the entire history as if it came from the same source.[1] Since a price is quoted only on regular London business days, the time interval between successive observations is a random variable. That fact by itself does not introduce any insurmountable obstacle, although it would be more convenient to have observations at regular intervals. Weekends and holidays present a more substantive problem, however, because it is not at all obvious that the variance of price shocks should be the same on weekends and holidays as it is on business days. In variance terms, if the twenty-four hours from 9:20 a.m. Monday to 9:20 a.m. Tuesday is one unit of time, how many units is the interval from Friday morning to the next observation, on Monday morning? We will have to approach this question empirically, by testing for a weekend effect in the variance.

The size of the sample, which is rarely found in any financial asset, makes it possible to construct subsamples for estimation purposes, and we will take full advantage of this ability. There is hardly any other asset which admits even a run of twenty-five years without significant change in definition. The universe of stock issues which have been continuously quoted for twenty-five years is a very special and unrepresentative universe, as so many of the original compatriots have fallen by the wayside and new ones have emerged over time. Even equity issues which have carried the same name for that length of time are often attached to corporate entities which have drastically redefined themselves in order to survive or to prosper. The problem of comparability is partially solved by resorting to aggregates, like the S&P 500 index. While Standard and Poors Corporation undoubtedly take pains to minimize the disruption caused by dropping old names from the list and adding newcomers, the evolution of the universe is bound to have an impact on the time series properties of the index. Whether the universe is actually more stable for the purposes of the investing public – whether it represents fairly homogeneous shares in a single commodity called "corporate America" – is an open question. Questions like this do not arise for gold, because the definition of the asset has not changed at all in recent history.

The Gold Option Data Set.

Our data on options on gold is much more limited in coverage. The data is also daily, starting in August, 1995 and continuing to May, 1997. In the case of options, each day we have not a single market, but a whole family of them, because the market prices a wide range of put and call options to different strikes and dates of expiration.

While the options contract and the underlying futures on gold are related to essentially the same asset as the spot price, there are two potentially significant sources of non-comparability. One of them comes from the timing of the price observations. Gold trades around the world in something which closely approximates a continuous, twenty-four hours trading day. The recognized daily spot series – whether it is referred to as the London p.m. fixing or the Handy and Harman spot price – is observed at the time of the close of the London market, which happens shortly before 10:30 a.m. New York time. The futures and options on futures that we have available to us are the ones traded on the Comex exchange in New York, and the price associated with a given day is for the close of the Comex, at about 2:30 p.m. New York time.

In the intervening four hour period, the Comex is the principal site of trading, and so the closing price of futures incorporates a lot of trading which has happened after the Handy and Harman spot price was fixed.

The other source of slippage between spot and futures is the basis, which reflects the convenience that comes from the financing of a futures position. We do not need to indulge in any lengthy speculation on the economics of the basis. It suffices to observe that the basis is a variable over time, in part as the result of observable factors like short term interest rates, and partly because of hidden factors like convenience yield.

For the purposes of this study, it is not necessary to explain either of these phenomena. Our only interest in them focuses on the net effect on option pricing. In other words, is there any systematic reason to believe that either spot or futures would be more volatile than the other? Sound arguments can be made on either side of this question. On the one hand, the timing mismatch probably causes futures to be the more volatile. The reason is that the Comex market clears only a very small fraction of all transactions. The London exchange clears about ten times as many ounces of gold, and the over-the-counter market is probably equal in size to the London exchange. Thus throughout its whole trading day, some of which occurs before the London close, the Comex is clearing perhaps five per cent of all transactions. The part of the Comex day from 10:30 a.m. to 2:30 p.m. should, moreover, probably be thought of as after hours trading, in a market which has been notoriously subject to manipulation.

Weighing in on the opposite side is the belief that futures trading smooths volatility, by removing purely transient swings. Even though the basis introduces some variability on its own, the belief in the smoothing properties of futures presumably relates to gross rather than net prices, because otherwise it is not at all clear what is the economic function of futures markets. We will be able to address these questions empirically rather than a priori, because we have data on both futures and spot, although our time series on futures does not cover nearly as long an historical period.

Characterization of the Spot Price Series.

There are potentially two kinds of interesting, non-random behavior in a series of asset prices. One of them relates to the stationarity of the series. Are there identifiable subperiods in which different rules apply, where by “rules” I mean any model of the transition rules from one observation, or string of observations, to the next one. The other non-randomness relates to the set of rules which, assuming stationarity, apply to the whole data set. As we proceed, we want to leave open either or both kinds of non-randomness, and to apply methods capable of finding either one. Other things being equal, however, this commission would leave us with a hopelessly broad research mandate. We are able to rein in the mandate appreciably because in the end we are only interested in non-randomness which has implications for option pricing. Thus, for instance, a tendency for certain patterns, once initiated, to complete themselves is not pertinent to thus study, although it would be of vital interest in other areas.

We have looked at our data on spot prices through the double lens of two statistical models, which we will identify as the Holding Period Variance Model (henceforth "Period Model") and the Mixture of Distributions Model (henceforth "Mixture of Distributions Model").

The Period Model.

This model posits that the cumulative variance of the differenced price of gold is a function of the length of the interval of observation. If the price series is a Weiner process the variance is simply proportional to the length of the interval, because of the independence of the periodic increments.[2] We want to allow simultaneously for two kinds of non-random behavior of a fairly intuitive kind, which are that the series contains both short term transient components and long term trended components. In terms of the differenced series, these imply that over a range of short holding periods, variance grows most slowly that the holding period, while over a sample of long holding periods, variance grows more than proportionately with the length of the holding period. Specifically, we estimated the parameters of a model of the following form.

Let p(t) be the price at time t, measured in calendar days. Define

1. $V(t,n) = [\ln(p(t + n)) - \ln(p(t)) - g * n]^2 / n$,

where g is the growth rate of p over a much longer time interval. In practice, we use the growth rate over the whole sample of 6079 observations, and we accordingly assume that g does not covary with the price change from time t to t + n. Under this assumption, V(t,n) is both unbiased and the best estimate of the variance of price change over this interval. If price followed a Weiner process, V would be independent of t and would be proportionate to n. This specification corresponds to a regression model which relates the logarithm of V to the log of n, in which ln(n) has coefficient equal to unity, viz.

2. Ln (V(t,n)) = α + ln(n) + u(t).

Against this Weiner null hypothesis, we estimate a quadratic model of the form

3. Ln (V(t,n)) = $\alpha + \beta_1$ ln(n) + β_2 [ln(n)]^2 + u(t,n).

The Weiner hypothesis implies that β_1 equals 1.0 and β_2 equal s zero. It implies, further, that the residuals are independent as long as the time intervals [t,t + n] do not overlap. Against this prior we postulate that variance first grows more slowly than n, and then grows faster than n, and we further postulate that the variance of the price series changes slowly over time, so that u is positively autocorrelated.

Formally,

4. H_0 : β_1 = 1.0, β_2 = 0.0, and Corr(u(t),u(s)) = 0 if the intervals starting at t and s, respectively, do not overlap.

Versus

5. H_A : β_1 < 1.0, β_2 > 0.0, and Corr(u(t),u(s)) > 0 if the intervals starting at t and s, respectively, are close together (i.e. if t and s are nearly equal).

The data needed to estimate these parameters and to test these hypotheses against each other are based on the sample of spot prices described above, but they are different in one important respect. In order to build as estimation sample we need to first create the V variable by sampling from the raw, daily data. The method of sampling, moreover, must be designed to rule out overlapping observations. The method for doing this is somewhat complicated, but can be explained in the following way. First, for simplicity suppose that we had a price observation every day, so that day count and observation count would be the same. Then construction of the data set would start with a random sample of positive integers, $\{n_j \mid j = 1,2,\ldots,K\}$. This sequence can be accumulated to give a strictly increasing sequence of integers $\{m_j \mid j = 1,2,\ldots,K\}$, satisfying

6. $m_j = n_j + m_{j-1}$.

The associated sample of prices at these dates, $\{p(m_j) \mid j=1,2,\ldots K\}$ provides the series we need for constructing V(t,n).

Now in practice there are many missing observations, corresponding to days of the weekend and to holidays, which requires an adjustment to this method. We start off in the same way, by drawing a sample of increments $\{n_j\}$, and accumulating them into the m series. This identifies the business days on which we observe the price, but it does not give the correct length of the time interval from one observation to the next. Each n is the number of rows we move down the column of data, and m is the row numbers we sample. We need a new counter, call it m', which is equal to the actual calendar dates (i.e. the day counts) on the rows we are sampling, and the associated intervals, n, are the differences of the m' series, not of the original m series.

It is possible to construct many different samples from the initial universe of daily observations,[3] and indeed we can in practice examine only a tiny fraction of them. There is good reason, as the sample statistics demonstrate, to look at more than one such sample. Each of these samples corresponds to an independently generated set of random integers $\{n_j\}$. For each specification of the regression model, equation 3 above, we generated between 125 and 160 samples from the raw data. We report the statistics based on these sets of regression, rather for any one regression equation. The specifications correspond to three implementations of the test. The first specification uses the entire universe of days, and allows for horizons (i.e. n_j's} of up to ten business days. As explained previously, this could give rise to horizons of up to about sixteen calendar days. The average sample of this sort has about 1100 observations, because on average the span from one observation to the next is a little less than six business days. The second specification differs in that we sample for longer return intervals. In this case, the interval was allowed to extend to forty business days, or about two calendar months. As a result, there are far fewer observation in each sample. The third specification is exactly the same as the first, except that we limited the universe of gold prices by starting at the 2500th day, which falls early in 1982. Thus, the third specification is a repetition of the first specification, but covers only the last fifteen years. The reason for looking at this specification is that there is some reason, a priori, to imagine that the structure of the price series might have changed after the explosive rally of the late 1970's. As we shall see, there is no evidence of a structural shift in any of the characteristics that interest us. There is, however, a dramatic shift in the trend of the price series.

The variance measure which appears on the right hand side of the model in equation 3 demands an exogenous estimate of the price trend. For this purpose, we simply used the trend which obtained over the appropriate historical period. This, for the first and second specifications we used the average trend from January, 1973 to May, 1997. For the third specification we used the trend which obtained over roughly the last fifteen years. In each case, this trend estimate was fixed across all repetitions. On the assumption that this trend is essentially independent of any single price return, the variable on the right hand side of equation 3 is the logarithm of a $\chi^2(1)$ variate which is an unbiased estimate of the variance of the price change.

The parameter estimates are summarized in the following table, Table 1. The three replications of this research plan described above are identified as the three columns of the table. Sample A is the full historical sample, using intervals between one and ten business days. The second column, Sample B, skips the first 2500 observations, which therefore starts around January, 1983. These two samples are alike in that the intervals of observation range from one to ten days. The third sample, identified as C, uses the full historical sample and draws intervals ranging from one to forty business days, or roughly two months. The sampling process can itself be replicated by simply starting at a different seed on the pseudo-random numbers. For the first two samples we repeated the drawing one hundred twenty-five times, and for Sample C we repeated it one hundred sixty times. One of the interesting features of our empirical results is the distribution of regression parameters across these replications.

One hundred twenty-five replications of Sample A produced the following model, where n_i is the elapsed time in calendar days between t_i and t_{i+1}.

7. $$\mathrm{Ln}\,(\hat{V}(t_i, n_i)) = \alpha + .531 \ln(n_i) + .148\,[\ln(n_i)]^2 ,\ R^2 = .115.$$

It is obvious that the linear coefficient is much less than unity, and that the quadratic coefficient is reasonably large. To put these in perspective, it is useful to calculate the implied cross-over points, at which the total derivative of the left-hand-side with respect to ln(n) becomes equal to unity. Clearly, this happens

at an interval of length z, where

8. $$\ln(z) = (1 - .531) / (2 * .148) = \ln(4.88 \text{ days}).$$

Table 1.

	Sample A	Sample B	Sample C
Linear Coefficient			
Average Coefficient	0.531	0.646	0.820
Average St. Error	0.139	0.184	0.279
St. Deviation	0.343	0.493	0.718
S.E.	0.059	0.060	0.061
Quadratic Coefficient			
Average Coefficient	0.148	0.117	0.059
Average St. Error	0.046	0.061	0.054
St. Deviation	0.115	0.160	0.136
S.E.	0.020	0.020	0.012
Residual Autocorrelation			
Average Coefficient	0.195	0.105	0.189
Mean R-squared	0.115	0.096	0.174
Observations per regression	1100.	646.	293.
Degrees of freedom correction	5.524	9.408	20.735
Number of replications	125	125	160

Notes: The origin of the samples is explained in the text. Average Standard Error is equal to the average of the estimated parameter standard errors, averaging over the replications of each sample. The Standard Deviation associated with a parameter is the standard deviation of the regression parameters across the pool of replications.

The cross-over from a return-reversal regime to a return-momentum regime occurs at about five calendar days. Since this estimate is at roughly the mid point of the sample of the independent variable, ln(n), we can be fairly confident that the sample contains many actual observations of both return reversals and of return momentum. If z had been much larger than ten, all that we could say is that the tendency for return reversals decays with the length of the interval of observation. We could not have said that we had actually observed a transition to a regime of return momentum.

We have deliberately avoided attaching any measures of precision, or significance, to these estimates, because there is in fact no accepted way to do that. We actually have a plethora of candidates. Each of the 125 replications of the sample produces regression coefficients and associated standard errors. On average, these standard errors are based on about 1100 data points each; they assume about 1100 degrees of freedom. The origin of this statistic is simply that when you sample from 6079 days, at random intervals which average 5.5 days, you have selected about 1100 days. In Table 1 we report the average, across all replications of these standard errors. The average also implicitly assumes that we have 1100 degrees of freedom to work with. In truth, however, we have 6079 distinct price observations, so it seems more sensible to adjust the standard error to reflect this number of degrees of freedom. The item labeled "degrees of freedom correction" in Table 1 is simply the ratio of the total sample, 6079, to the average regression degrees of freedom (i.e. 1100 in the case of Sample A). We recommend that all standard errors be divided by the square root of this number. The correction for regression degrees of freedom is inconsequential because these numbers are so large, and we have simply suppressed it. Accepting this adjustment, the standard error of the linear coefficient becomes .059 (= .139 / sqrt(5.524)).

Whether or not we adopt this adjustment, the linear coefficient is significantly smaller than 1.0 at all conventional significance levels. The t-ratios are equal to either 3.37 (= [1 - .531] / .139) unadjusted, or 7.9 (= [1 - .531] / .059) adjusted. The quadratic coefficients are also significantly positive. The t-ratios are either 3.22 (= .148 / .046) unadjusted, or 7.56 (= sqrt(5.524) * .148 / .046) adjusted. As these calculations suggest, we can in either case reject the null hypothesis whether or not we adjust the degrees of freedom. The whole sample strongly implies that the process that has generated spot gold prices exhibits a tendency for return reversals over a very short horizon – a few days or so – and a tendency for return momentum over longer horizons – greater than five days, it seems.[4]

The other samples lead to the same conclusion. We have compiled the appropriate t-ratios into All of the coefficients for the first two samples – the ones based on a ten day interval of observation – are significant at conventional levels, and the adjusted ratios are enormously significant. In the case of the sample of longer observation intervals, only the adjusted t-ratios are significant. The reason, quite plainly, is

Table 2: t-Ratios of Regression Coefficients

	Sample A	Sample B	Sample C
Linear Coef.			
Unadjusted	3.37	1.92	0.65
Adjusted	7.90	5.90	2.94
Quadratic Coef.			

Unadjusted	3.22	1.92	1.09
Adjusted	7.56	5.88	4.98

that allowing longer intervals sharply reduces the number of observations – the number of such intervals – contained in the roughly twenty-five years of history at our disposal. On closer inspection we see that on average the R^2 coefficients from this sample are the largest of the three samples, not the smallest. The quadratic model, therefore, fits the dependent variable pretty, at least as well as it does in the first two samples. The variables $\ln(n_i)$ and $\ln(n_i)^2$, sampling at intervals of up to forty days, are quite highly correlated in this sample, however, and the resulting collinearity diminishes the efficiency of this test.

Our estimation methods affords us a second, rather different measure which is comparable to the foregoing standard errors of estimate, and which sheds a revealing light on this data. Under conventional OLS assumptions, the 125 replications of Sample A would be drawings from the same joint distribution, of which the mean is the pair of true regression coefficients. Since the estimated, unadjusted standard errors of these coefficients are estimates of the standard deviations of the corresponding sample coefficients, presumably the standard deviations across replications would be related to the OLS standard errors. In this case, they would not be equal, because of interdependence across the replications. That is to say, if the 125 replications were drawn from 125 actually different histories, the coefficient estimates would be independent, and the standard deviation of a sample of them has the same expected value as their standard errors. Our 125 regression coefficients are not independent, because were have only 6079 observations.[5] The coefficients must be positively correlated, which causes the dispersion of estimates to be narrower that it would otherwise be. What we find in practice is, however, that the simple standard deviation across the set of replications is much larger than, rather than smaller than, the mean of the unadjusted standard errors. At first blush this present something of a paradox.

There is no way to avoid the conclusion that the underlying data violates the assumptions of ordinary least squares.[6] The problem, quite simply, is that the statistical distribution theory implied by the Gauss-Markov Theorem significantly underestimates the variance of the sample coefficients. Our guess is that the culprit is heterscedasticity, caused by heteroscedasticity of the underlying gold price series.

The dispersion of estimate coefficients across the set of replications is quite remarkable. The number of replications for which the linear coefficient is negative is especially revealing. The linear coefficient would be negative if the return variance over a pair of successive days is on average smaller than the one day return variance. This implies a degree of return reversing behavior which we do not in general observe in this data – on average, the linear coefficient is significantly positive in most replications, and is highly significant across the set of replications – but for six of the 125 replications of sample A we do obtain a negative linear coefficient. For sample C, seventeen of the 160 replications produced a negative linear coefficient. Both of these rates of occurrence exceed the expected number. Since we are, in effect, holding constant the average rate of one day return reversals – i.e. since we are holding constant the average linear coefficient – the high number of negative cannot be due to a correspondingly high propensity for reversals. Our inference is that the replications which produce a negative linear coefficient are excessively influenced by a few outliers. We would expect some negatives coefficient, because even homoscedastic data will have outliers. The issue here is obviously the excessive frequency of this phenomenon, as indicated by the fact that the sample standard deviation of the linear coefficients is much larger than the theoretical standard deviation (i.e. the standard error) of the linear coefficients.

The tendency of return shocks to reverse in the short run, i.e. within a couple of days, is matched by the tendency of the remainder of the shock to persist at least for several weeks. On average, return reversals and return persistence are quantitatively of equal importance in the gold market. When we estimated a simple linear regression of the logarithm of variance against the logarithm of the observation interval, the coefficient was almost exactly equal to unity, and was far from being significantly different from unity. Thus, if we had to characterize this data by a model with only one parameter, the exponent of time in the variance of price, we would conclude that the Weiner process, with exponent equal to unity, was the best fitting model.

Autocorrelation of the Residuals.

The residuals from an OLS estimation of equation 3 exhibit high positive autocorrelation, for which reason we used Corchrane-Orcutt estimates instead. The mean, across all replications, of the first order autocorrelations appears in the table of parameter estimates, Table 1. For the models which apply to the full history the average autocorrelation is about 20%. While we do not report the standard errors there, they are typically around 3%, so all autocorrelation coefficients are highly significant.

The presence of autocorrelation in the model residuals demands some explanation, first as to how it should be interpreted, and secondly as to why it is present in this data. As far as the interpretation is concerned, positive autocorrelation of the residuals implies that when the variance of price changes is above the fitted regression point, the variance will tend to stay above the regression model for some time. In constructing any given replication, we sampled randomly from the history of daily gold prices. This was done in a way which preserves the chronological ordering of the observations, even though the length of time which elapses between successive observations is a variable. In the notation of the model in equation 3, if t_i and t_{i+1} are successive observations, of length n_i and n_{i+1} respectively, the $i+1^{st}$ period, designated t_{i+1}, starts immediately at the end of the i^{th} observation. Autocorrelation thus has the usual interpretation in terms of a relationship over time.

The economic significance of positive autocorrelation is that the price of gold experiences persistent regimes of high and low price variance. This finding is not especially surprising. It has been observed not only in gold but in most markets. One thing which we can add here is that the rate of return to the fitted volatility has been considerably faster since 1982 than it was before that time. The residual autocorrelation since 1982 was about 10%, while for the whole history it averaged around 20%. The period that preceded 1982 contains two exceptional events, the market adjustment to the decision to allow the price of gold to float and the shocking rally and crash which occurred early in 1980. Although our sample starts more than a year after the float, the early years still were a period of adjustment in many ways. Most significantly, American citizens were not allowed to own monetary gold (bars and coins) until 1976. Prior to that time they were able to participate in the gold market through futures contracts, but only a very small portion of the American public was inclined to trade in commodity futures in those years.[7]

The other epoch of exceptional volatility was a three month period from the middle of December, 1979 to the middle of March, 1980. Between mid December and mid January to price of gold nearly doubled. Reference prices are: December 14, 1979 – $456.80 – and January 21, 1980 – $850.00. By the middle of March, the price returned briefly to where it started in December. For reference we have March 17, 1980 – $484.00. From this low, the price bounced again into the $500 range, but the subsequent high was nowhere near the January peak. The highest daily close since early 1980 was $711, which occurred on September 23, 1980. Since the subsample results exclude this period of history, it is not surprising that parameters like the autocorrelation of price variance would be different.

There is a second anomaly in the residual autocorrelations, one which is masked by a coincidence of numbers. The first order autocorrelation of the residuals from Samples A and C happen to be approximately equal. Both of the samples spans the whole quarter century of price data, so in that respect it is not surprising that the autocorrelation coefficients would be more-or-less equal. Actually, however, it should be surprising for the following reasons. The degree of autocorrelation of the model residuals – i.e. the degree of inertia in the variance of daily prices – depends upon number of days between successive observations in the sample. The length of calendar time which elapses between successive observations in any given sample is a function of the range of lags which the sampling method permits. In the case of Samples A and B, the lags ranged between one and ten business days. For Sample C, however, the lags range between one and forty business days. Thus, in Sample A, successive observations are on average about five and a half days apart – this is the average interval from the middle of one interval of observation to the middle of the next one – while from Sample C the interval is twenty and a half days on average, or about four times as long.

If the autocorrelation of price variance was the same on the base of equal historical intervals, day-to-day autocorrelation, for instance, the autocorrelation found in Sample C should be about equal to the Sample A autocorrelation raised to the fourth power! More concretely, if the autocorrelation of variance is 20% over an interval of five and a half business days, it would have to be about 75% day-to-day, based on business days only.[8] By contrast, an observed autocorrelation of 20% in Sample C implies a day-to-day autocorrelation of 92% on a business day basis. The other difference between these samples is of course that the model, when applied to Sample C, forces a simple quadratic relationship between variance in the observation interval over a much wider range of intervals. The fact that the residuals from Sample C are much more highly autocorrelated than those from Sample A suggests that departures from the simple quadratic model – the presence of exceptional wiggles in the relationship – have occurred and when the did occur, that have been quite persistent.

While they differ in important details, differences which suggest that the structure of volatility has evolved over time, the three samples agree on the broad characterization of price innovations in gold. Over short holding periods, up to a few days, the cumulative volatility of price innovations grows less than linearly with time, implying that the innovations contain a large transient component which disappears in a couple of days. Over longer holding periods, the cumulative variance of innovations grows faster, and beyond a week or so, it begins to grow more than proportionately with time. Besides the transient component of price innovations, therefore, there is a persistent component which introduces a trend into the price series, when it is observed at relatively long intervals. Before we leave this topic, it makes sense to attempt to estimate the degree of short run price reversals – i.e. transient price shocks – directly. This is not easy to do with high precision because of complications in the price data. Autocorrelations at different lags are, however, both simple to estimate and revealing. Note that the question at hand is not whether the data contains short run price reversals and long run trend – the models we discussed above document those phenomena. The agenda for computing the autocorrelations is simply to estimate the magnitudes.

The resulting statistics are summarized in the following tables, 3.1 and 3.2. In order to carry out these calculations we had to fix an interval of observation, and we did so by repeating the calculations three times, for a one day, one week, and one month interval of observation. These terms are, of course, only approximate because weekends and holidays intervene. For our purposes, “daily” means successive business days, “weekly” refers to the interval from the first business day of a week until the first business day of the next week, and “monthly” designates the interval from the last business day of a month until the last business day of the following month. The estimated two distinct measures of autocorrelation: the usual parametric autocorrelation coefficient, and the difference between the conditional probabilities of “up” and “down” moves. These two methods can yield quite different indications, because of the presence of heteroscedasticity in the data, but in practice the parametric and non-parametric approaches yield quite consistent indications. Prices changes are highly negatively autocorrelated over short intervals, and are less negatively autocorrelated or are actually positively autocorrelated over longer periods. The contrast is sharpest between the daily and monthly autocorrelations of the full historical sample. Using a daily interval, the first order autocorrelation is negative, about −5%, with a t-ratio which exceeds 3.5 in absolute value. Over a monthly interval, the first order autocorrelation is positive, +25%, with a t-ratio which is much larger than 4.0. The transition probabilities produce the same contrasting results. On a day-to-day basis, the probability that a down day is followed by an up day – the probability of a sign reversal starting from a

Table 3.1 Full Sample

	Daily	Weekly	Monthly
First Order Autocorrelation	-.046	-.088	.252
No. of Obs.	6069	1276	293
t-Ratio	-3.58	-3.16	4.45
Transition Probabilities			
Pr(+ | +)	44.0%	49.1%	55.2%
Pr(+ | -)	54.9%	51.7%	43.9%
t-Ratio	8.5	0.9	1.95
First Order Autocorrelation			

Table 3.2 Post – 1982 Sample

	Daily	Weekly	Monthly
First Order Autocorrelation	-.089	-.093	-.02
No. of Obs.	3572	753	185
t-Ratio	-5.32	-2.55	-0.33
Transition Probabilities			
Pr(+ | +)	42.8%	43.3%	44.8%
Pr(+ | -)	55.9%	50.9%	49.0%
t-Ratio	7.83	2.11	0.7
First Order Autocorrelation			

Notes: Autocorrelation coefficients are first order autocorrelations of first differences of the logarithm of price. T-ratios: for autocorrelation coefficients are estimated in the usual way. The t-ratio reported of the difference of conditional probabilities uses a standard error which assumes that all conditional probabilities are equal to 50%.

down day – is about 55%. The probability that an up day follows an up day, however, as 44%. Thus the direction of daily price changes is highly negatively autocorrelated. The sign of month-to-month price changes is, by contrast, highly positively autocorrelated.

The statistics reported in this table reveal some unexpected results in the details, however. Most importantly, price changes have been much more negatively autocorrelated (less positively autocorrelated) after 1982. The first order autocorrelation of daily price changes falls from −5% to −9%. More striking still, even month-to-month price changes are slightly negatively autocorrelated after 1982. We should emphasize here that the weekly and monthly statistics are not a reliable indication of the degree of autocorrelation of the momentum component of price movements, because they contain the highly negative daily effect as well. It is nonetheless remarkable that the highly positive monthly momentum of prices is found only in the first ten years of this data.[9]

This completes the first part of this study, which dealt with the autocorrelation structure of price changes of spot gold. We find compelling evidence that prices innovations have both transient and trended, or persistent, components. The relative magnitudes of the variances of the transient and persistent components of price appear to have changed over the course of recent history. The first ten years, running from 1973 through 1982, were characterized by a very high degree of persistence of price innovations and a comparatively smaller contribution of transient shocks. From 1983 through 1989, the transient component appears to have dominated the magnitude of innovations, and since 1989, there seems to have been a more even balance between them.

In this section we also presented evidence on two other points which we will take up again in the next section. First, the variance of innovations had varied over time, and that while varying, the variance has had a great deal of inertia. Secondly, the behavior of our estimators implies that the price data is itself rather heteroscedastic.

Further Evidence on Heteroscedasticity.

The foregoing tests of the variance of price innovations produced direct evidence on two points which we will return to in this section. First, it produced strong evidence that the variance has varied over time, but that it exhibits inertia. It would be appropriate to speak of regimes of high variance and low variance. Secondly, it produced indirect indications that the price innovations are heteroscedastic. The basis for this inference is that repeated replications of the estimation method gave rise to a much broader distribution of parameter estimates than OLS would have predicted. This happened, moreover, in defiance of the fact that the dispersion of parameters should actually have been smaller than OLS would have predicted. In this section we will estimate the parameters of a simple mixture of normal distribution model which allows for two different variances, and following that we will estimate the day-to-day persistence of the variance parameter.

The model has four parameters, the mean daily price change μ, two variances σ_H^2 and σ_L^2, and a mixing parameter π. The interpretation is quite simple. Let g(t) denote the one day percentage price change observed at time t. Since we observe prices only on business days, t is in effect an index of business days. With probability π the price change was drawn from a high volatility distribution $n(g(t) \mid \mu, \sigma_H^2)$, and otherwise it is drawn from the lower volatility distribution $n(g(t) \mid \mu, \sigma_L^2)$. The since time series g(t) can be viewed as the resultant of two time series, a weiner process z(t) which has constant variance equal to 1, and a Bernouilli variable $\pi(t)$ which selects a variance to apply to z(t). We will assume for the purposes of this model that z(t) is truly white noise, although as reported in the preceding section price changes of gold exhibit some serial correlation. The only role of this assumption for this purpose is to avoid questions about the number of degrees of freedom in our sample of daily data. The implied sample of the process z(t) is not strictly a random sample, and a simple count of the observations overstates the available degrees of freedom, but in practice degrees of freedom are not a problem, and especially so since the serial correlation of the z series is small in absolute value.

These remarks do not, however, lay to rest all questions relating to autocorrelation, because the indicator series, $\pi(t)$, may also be serially correlated. The autocorrelation of variance, of which we found evidence in the previous section, is both large in magnitude, i.e. in the neighborhood of 80% in daily data, and highly significant. It follows that on this score also, the time series of price returns, g(t), is not serially independent. Serial dependence of the π's does not introduce autocorrelation into the g's, because the conditional expected value of g(t) is not a function of $\pi(t)$, but it does introduce a nonlinear serial dependence, with the result that the original time series data is not, strictly speaking, a random sample. We will return to the matter of estimating the autocorrelation of the indicator variable $\pi(t)$. For the present we will simply note that the loss of degrees of freedom need not be critical because of the size of the data set.

To summarize, we assume that the series of percentage price changes, g(t), takes the form

9. $g(t) = \mu + \pi(t) * z(t)$, where $\pi(t)$ is a Bernouilli random variable,

9a. $\Pr(\pi(t) = \sigma_H) = \pi$, $\Pr(\pi(t) = \sigma_L) = 1 - \pi$ and

9b. z(t) is i.i.d. standard normal.

Leaving aside the question of serial correlation of $\pi(t)$, we can estimate the parameters of this model by conventional maximum likelihood methods, to which we now turn.

Estimation of the Mixture of Distributions model.

Before we take up a detailed presentation of the methods and results of parameter estimation, we need to return to the issue of non-trading days. The data on gold prices does not give any evidence of a weekend effect in expected price change; the average price change in out sample was essentially zero both on week days and on weekends. There is almost sure to be a weekend effect in variance, however, because the holding period variance of gold or any other price is highly dependent on the length of the time interval. It is possible that the weekend is just like one trading day, no matter how many calendar days it comprehends, but that is very unlikely. It is much more likely that the weekend in equivalent in variance to more than one trading day. Not to adjust weekends would, therefore, build in precisely the sort of heteroscedasticity which we have set out to find. The first step in estimating the parameters of the model is to estimate the ratio of the daily variances within a business week and over a weekend.

Weekend Effects

There are potentially many sorts of weekend effects in data like this, but we will limit our attention to the two most general ones: the potential for a difference in mean price changes over a weekend or holiday, and the potential for a difference in the variance of price change. The data at our disposal reject convincingly the proposition that the mean price change in different on non-trading days than it is on open trading days. In fact, over the whole sample, the mean price change in not significantly different from zero for either kind of day. Thus we are left with one question: whether the day-to-day variance of price change is different over non-trading days than it is over open trading days. We parameterize this test in the following way. Our sample contains gaps for non-trading days, which are counted by a companion variable called "gap." We would adjust for different variances, as a function of the gap, by dividing the price change over a period of non-trading days by the gap raised to some power. The natural scale for w is one in which the exponent is equal to $-w/2$. The parameter w then has the interpretation of the effective number of open trading days per non-trading day. If prices are less volatile on non-trading days, as we would expect them to be, w should be less than one; in terms of variance a non-trading day is less than one full open trading day.

It would in theory be preferable to estimate weekend effects and the parameter of the mixture of distributions model jointly, but there are good reasons for not doing so. The chief reason is that we would have no restriction in the model to force the mixture of distributions part to ignore the weekend effect in variances. If the weekend effect is very strong, the model would tend simply to accept the mixture of distributions hypothesis, using the distribution of first trading day of a week as one distribution and all other days as the other. In formal terms, estimation errors of weekend parameters and of other mixture of distributions parameters are highly correlated.

The only course which these facts leave open to us is to estimate first the weekend effect, and then to estimate the parameters of a mixture of distributions model conditional on the indicated weekend adjustment. This approach carries with it some risks, because as we noted, estimation errors related to these two effects are correlated. We will attempt to deal with this by some simple sensitivity analysis. Specifically, since we will be able to fix a standard error to the estimated weekend effect, we can sample from the whole distribution of parameters.

Two other technical points demand some comment. First, we do not exactly have weekend variances, because all our data is from the daily close. Thus, while we refer to a "weekend," we in fact mean the period from the close, Friday, to the close, Monday. This is three days, not two days, and contains one open trading day. This is far from an insurmountable problem, requiring a small modification of the methodology.[10] The other point is to recognize that our estimate of the weekend effect is inconsistent to the extent that it ignores the heteroscedasticity which the mixture of distributions model is supposed to uncover. This is just a consequence of the fact that sequential estimation of a joint hypothesis is inconsistent. We will return to this point in a moment. While we accept that it is important in principle, we would assert in this context that it does not invalidate the findings we will report on the mixture of distributions model.

The method we employed to estimate the variance ratio parameter, w, is derived from a simple F-test which compares the variances of price changes over consecutive days against the variance over periods which span more than one day. Our point estimate of the parameter w is simply the value which equates the F-ratio to unity. For reasons explained above we wished to do extensive sensitivity analysis, and for this purpose we estimated one, two, and three standard deviation confidence intervals for w. These interval estimates were obtained by using the known standard error of the F-ratio. That is to say, our estimate w + S.E.(w) is simply the scaling parameter which equates the F-ratio to 1 + S.E.(F). The empirical estimates are as follows:

w = .383

1 S.E Interval: [.339, .426]

2 S.E. Interval: [.295, .47]

3 S.E. Interval: [.25, .51]

In words, a non-trading days is about 3/8^{ths} of an open trading day, and in any case is very unlikely to be more than half, or less the one quarter, of a trading day. The inconsistency referred to above has in practice to do with the confidence levels that attach to these intervals. In computing the F-ratio we have simply counted observations: 2809 observations on consecutive trading days and 771 observations which involve some gap in time. This implicitly assumes that the data is homoscedastic except for the weekend effect, but of course that is precisely what we set out to question. If the data contains other sorts of heterscedasticity, as we sincerely expect it will, we have overestimated the degrees of freedom of the F-ratio. In this case, our estimate of the standard error of F is too low, and our intervals are too narrow (or equivalently, our confidence levels too high). It seems nonetheless very improbable that the true factor is outside the range from .25 to .5. We will accordingly use these values for sensitivity analysis.

Parameter estimates of a simple mixture of distributions model appear in Table 4 below. The columns of the table correspond to assumptions about the weekend effect, using .25 .383, and .5 as the three test values, in that order. The parameters reported are first of all the two standard deviation estimates and the estimated fraction of the sample drawn from the high volatility distribution. Below these, we report the on a test of significance of the model, against the null hypothesis of a single normal distribution. Specifically, the statistic we report is the conventional χ^2 statistic, equal to two times the log likelihood ratio. Under the null hypothesis, it has two degrees of freedom. Quite obviously, we can reject the null hypothesis with a very high degree of confidence. The entry in the table, labeled "p," will be explained below.

Table 4. Parameters of the Mixture of Distributions Model

	.25	*.383*	*.50*
High Sigma, σ_H	.0171	.0169	.0168
S.E.	.00025	.00025	.00025
Low Sigma, σ_L	.0055	.0055	.00545
S.E.	.00006	.00006	.00006
Prob(σ=σ_H), π	.206	.207	.207
S.E.	.0126	.0127	.0127
p	.827	.775	.796
$\chi^2(2)$	898	887	880

Notes: Estimates of the parameters of a Mixture of Distributions model allowing for two volatility states. The parameter called π is the probability of being in the high variance state. The parameter p is the conditional probability of staying in the high variance state. Based on all daily observations since January 4, 1983; 3581 observations.

Regardless of how we correct for the weekend effect, within the range of corrections we propose, the ratio of high sigma to low sigma is about 3, and the high sigma state occurs about 20% of the time. Actually, the individual estimates are all very insensitive to the weekend correction. The high sigma is equal to about 1.7% per day, or equivalently an annual volatility of 29%, using the mid range estimate of the correction for non-trading days.[11] Using more conventional scaling factors, 1.7% per trading day grosses up to a little less than 27% volatility per year. Thus, in the high state we would expect annualized volatilities in the high twenties. In the low volatility state we should expect to see annualized volatilities a little under 10%. As we will see presently, these estimates contrast sharply with observed implied volatilities on traded options. The other parameter, called "p," is an estimate of the conditional probability of staying in the high volatility state. The estimator is derived in detail in an appendix, where we obtain the following formula.

Referring back to the model expressed in equation 6, we have parameters σ_H, σ_L, and π. Let the letter p denote the conditional probability of continuing in the high variance state, i.e. p = Prob($\sigma(t+1) = \sigma_H \mid \sigma(t) = \sigma_H$). Since the two subdistributions have the same mean, by assumption, the density functions cross at two points, equal to $\mu + B$ and $\mu - B$. Any observation which falls between these limits is more likely to have come from the low variance distribution, and any outlier is more likely to have come from the high variance one. We define parameters P_1 and P_2 to be the following probabilities:

10. $P_1 = \text{Prob}(|x - \mu| > B \mid \sigma = \sigma_L)$ and $P_2 = \text{Prob}(|x - \mu| > B \mid \sigma = \sigma_H)$.

If we accept for the moment that a day when the price of gold changes by more than B per cent (recall the μ is essentially zero) is a "big event" day, then the P's are the odds of having a big event. We emphasize, in passing, that the notion of a big event day is very different from that of a high variance day, even though they are related. Only a fraction of all high variance days turn out to be big event days, because most of the normal distribution lies close to the mean. Conversely, some small fraction of all low variance days actually turn out to be big event days. Thus P_1 is not equal to zero, and P_2 is probably not much larger then .5.

These parameters are estimated from the parameters of the static distributions, without reference to the time series nature of the data. The parameter p refers, however, a time series concept. To estimate p we need to relate it to the frequency of successive big event days. Let A denote the frequency of successive big event days in the sample. Then we show in the appendix that

11. $$p = \{A - [2 P_2\pi + P_1 - 2 P_1\pi] P_1\} / \pi (P_2 - P_1)^2$$

is a consistent estimator of the probability we are interested in. We actually show that if A is the true probability of successive big event days and if the static parameters are also exact estimates, then p is the true conditional probability. We do not have an exact formula for the standard error of p, but for present purposes we can estimate it as follows. The static parameters were all estimated with exceptionally high precision, as indicated by the standard error reported in Table 4. Estimation error in p is therefore dominated by sample error in A.

12. $$\text{S.E.}(p) \cong \text{S.E.}(A) / \pi (P_2 - P_1)^2$$

We obtain estimates of typically around .14 for A, which yields a standard error of A equal to .0058. π is about .2 and $P_2 - P_1$ is about .45. Thus the standard error of p is something like

13. $$\text{S.E.}(p) = .0058 / (.2 * .45\text{^}2) = .14$$

Considering how small are the standard errors of the static parameters, this is a rather large number,[12] but it is much smaller than the estimates we obtain for the probability p itself. Our estimates are typically around 80%. More to the point, we estimate that the unconditional probability of the high variance is about 20%.

The difference between the conditional and unconditional probabilities is accordingly about .6, and is more than four time the standard error of p.

It is well at this point to summarize out findings to this point.

A. The day-to-day percentage change of spot gold exhibits two identifiable components, one of which is negatively autocorrelated and the other of which is positively autocorrelated. Part of the innovation each day is thus a transient shock, which will disappear in a few days, and the other is an expression of a trend which will be realized gradually over a period of perhaps a month.

B. The scale of price variance varies over time, but it exhibits a high degree of inertia.

C. Price variance is much greater on open trading days than on non-trading days. In terms of variance, a non-trading day is only the equivalent of about 3/8ths of an open trading day.

D. Following up on point B above, there is compelling evidence that the distribution of day-to-day price changes has fat tails. While days of high variance tend to occur in runs, the variance each day is drawn from a conditional distribution which assigns significant probability to both high and low variance outcomes. Put differently, while high and low variance days tend to occur together, they are interspersed throughout the entire history.

This characterization highlights the presence in this data of two distinctly different, but equally interesting, phenomena: the presence of transient price shocks and the presence of days of exceptional variance. We have not thus far addressed the next natural question, which is whether price spikes – the result of high underlying variance – tend to be transient shocks or, on the contrary, evidence of trend. We will take up that question in the next section.

The Distribution of Prices Following a Price Spike.

This section deals with a question which is of critical importance both to the student of financial markets and to the would-be gold trader. Do price spikes have any predictive value, and if so, do they lead to price reversal – i.e. are they transient shocks – or are they precursors of a trend? The methods to be employed in this section, in parallel fashion, begin to look like so-called "technical analysis." For the more fastidious student of modern finance, we might suggest skipping direction to the following sections, which deal with option pricing. The more adventurous students will probably want to stay around to see the results of a test of technical trading rules. We pose the original question without preconceptions, and we are perfectly prepared to deal with negative results – i.e. that there is not particular pattern and no predictive value of any kind in price spikes.

The relevant test is very simple to define and to implement. We define a threshold which defines a price spike, and relate the subsequent path of prices to the originating event. For this purpose we considered two samples, the full history from 1973 onward and the restricted sample of data starting in 1983. We considered two alternative definitions of a spike. The simplest definition is a price change which in percentage terms exceed some threshold, in absolute value. The alternative definition recognizes that volatility drifts gradually over time, and adopts a moving window of reference. Specifically, the condition is that the price change exceeds some fixed multiple of a trailing standard deviation of price changes. For this purpose, we used a forty-five day – roughly speaking a two month – window for computing means and standard deviations. In both cases, whether the measure of a spike was in terms of the absolute or the moving average criterion, we investigated five different thresholds which selected on a subsample of days. Rather than attempting to predetermine the magnitude of each threshold, we simply sorted the data so that the thresholds captured 25%, 20%, 15%, 10%, and 5% of the whole sample. The resulting estimates are summarized in the following tables, Table 5.a and 5.b.

The parameters reported in the table are the slope coefficients from bivariate regression equations of the following form. Let y(t) denote a price shock observed at time t, and z(h,k,t) denote the percentage price change over the following days t+h to t+k We estimate the following regression equation,

14. $z(h,k,t) = \beta_0(h,k) + \beta_1(h,k)\, y(t) + e(h,k,t).$

The initiating event, the spike y(t), can be either positive or negative, and the variable y retains the sign of the event. Any tendency for price changes to be subsequently reversed will appear as a negative coefficient of y, and any tendency for the initial spike to trigger a continuing trend in the same direction will show up as a positive coefficient of y. The rows of the table are labeled by fractions of the sample which exceed the thresholds. Standard errors and t-ratios are shown beneath the beta coefficients.

These statistics exhibit a consistent pattern of price reversals, implying that on average large price changes contain a significant transitory component. Three days after a given spike, the price has on average given back around 10% of the initial change. The significance of these statistics is obscured a bit by the presence of overlap in the holding periods on the left hand side of equation 11. There is a tendency, caused by the persistence of high variance of price shocks, for spikes to be followed be other spikes. We did not

Table 5.a: Coefficients Based on a Fixed Threshold.

	0 - 1	*0 - 2*	*0 - 3*	*0 - 5*	*0 - 22*	*3 - 22*
25%	-.078	-.107	-.117	-.093	-.073	.044
	(.023)	(.030)	(.035)	(.043)	(.088)	(.083)
	[-3.39]	[-3.61]	[-3.31]	[-2.13]	[-.83]	[0.53]
20%	-.081	-.106	-.114	-.092	-.082	.032
	(.025)	(.032)	(.038)	(.046)	(.094)	(.088)
	[-3.27]	[-3.33]	[-3.04]	[-2.01]	[-.87]	[0.36]
15%	-.076	-.102	-.110	-.086	-.092	.018
	(.027)	(.032)	(.038)	(.048)	(.098)	(.093)
	[-2.81]	[-3.20]	[-2.89]	[-1.77]	[-.94]	[.19]
10%	-.071	-.112	-.099	-.080	-.147	-.048
	(.032)	(.035)	(.042)	(.055)	(.109)	(.101)
	[-2.24]	[-3.19]	[-2.32]	[-1.47]	[-1.34]	[-.47]
5%	-.062	-.118	-.082	-.049	-.099	-.017
	(.044)	(.047)	(.057)	(.067)	(.122)	(.109)
	[-1.40]	[-2.52]	[-1.43]	[-.73]	[-.81]	[-.15]

Table 5.b: Coefficients Based on a 45 Day Moving Average.

	0 - 1	*0 - 2*	*0 - 3*	*0 - 5*	*0 - 22*	*3 - 22*
25%	-.055	-.076	-.083	-.063	-.034	.049
	(.020)	(.026)	(.030)	(.037)	(.082)	(.076)
	[-2.76]	[-2.98]	[-2.71]	[-1.69]	[-.41]	[0.64]
20%	-.053	-.069	-.073	-.050	-.036	.037
	(.022)	(.027)	(.032)	(.040)	(.086)	(.078)
	[-2.41]	[-2.50]	[-2.26]	[-1.25]	[-.41]	[0.47]
15%	-.059	-.091	-.096	-.094	-.022	.074
	(.023)	(.027)	(.032)	(.039)	(.091)	(.083)
	[-2.62]	[-3.35]	[-2.98]	[-2.42]	[-.24]	[.90]
10%	-.054	-.074	-.076	-.071	.016	.092
	(.027)	(.030)	(.036)	(.043)	(.100)	(.091)
	[-2.00]	[-2.46]	[-2.12]	[-1.66]	[.16]	[1.02]
5%	-.077	-.128	-.115	-.088	.026	.141
	(.035)	(.035)	(.042)	(.050)	(.123)	(.113)
	[-2.19]	[-3.62]	[-2.72]	[-1.77]	[0.21]	[1.24]

filter the sample to remove observations which overlap in time. To do so would have "solved" the problem of overlapping observations, but only by introducing a far more worrisome problem of selection bias. We would have had to exclude from the sample all spikes which closely follow other spikes. Our data would then apply not to spikes, as such, but to spikes which are not preceded by other spikes. While that may be an interesting category of events, it is not the one which we set out to study.

A clear pattern emerges from these statistics. Price spike do contain a significant transitory component. The price of gold consistently reverses direction in the days following a price spike. After three days, on average, the price has given back about ten per cent of the initial spike. This behavior is more evident when we use an unchanging measure of price spikes, which is calculated from the actual prices only. Using a relative measure – measuring spikes relative to a moving window of past prices –gives essentially the same result, although the price reversal apparently does not begin as vigorously.

One of the most surprising findings is how little the statistics differ from each other when we select on higher thresholds. The 5% threshold, for instance, limits the sample to only the 20% of the sample which has the lowest – i.e. the easiest – threshold, but the parameter estimates are insignificantly different. The ten per cent reversal appears to be a relatively homogeneous phenomenon, regardless of the magnitude of the initial spike.

There is further evidence of this homogeneity, which comes from statistics reported in the second Appendix. In addition to the regression mode in equation 11, we estimated a regression in which the right hand side variable was a simple dummy variable which is equal to the sign of the price change, y(t). That is to say, the dummy is +1 when y(t) is greater than the positive threshold, and is −1 when y(t) lies to the left of the negative threshold. The coefficient of this dummy is equivalent to the conventional t-test for the difference of means. What we find is that while the coefficients are consistently negative, and are generally significant, they are never as large as the coefficients in Table 5 above. Thus a large part of the significance of the price spike, y(t), as a predicter of future price changes lies in the actual magnitude of y, and not simply in its sign.

The statistics reported in Table 5 address a second question, which is whether price spikes signal the onset of trend in the same direction. The evidence on this point is very consistently negative. After the initial rebound of price, there is no further predictable follow up to a price spike, and this is true both whether we use a fixed or a relative measure of price spikes. Focusing on the last column of the table, the cumulative price change over the period which starts three days after a spike and extends out to twenty-two days is not related to the magnitude of the spike, and it is similarly not related to the dummy variable described in the preceding paragraph.

While the apparent magnitude of the transitory component of large price spikes is statistically significant, it is surprisingly small. After three days, about ninety per cent of the spike has bee impounded permanently into the price of gold. The spikes could not, therefore, be in any sense identified with the transitory shocks. Other factors, including no doubt factors which are fundamental to the industry supply and demand, would be needed to more precisely decompose daily shocks into their permanent and transitory components, although the magnitude of the shock alone has some discriminating power as our findings make clear.

We engaged in a second test designed to characterize large price spikes. We investigated the period which precedes a spike. The methodology of this test is identical, except that the dependent variable in each regression is the cumulative price change over a holding period before the spike rather than after it. The outcome of this test was consistently negative. We could not characterize shocks in terms of the net price change over the recent past. In the version of this test which uses the actual price spike, no regression coefficient was significant. In the version which uses the dummy variable which captures just the sign of the spike, some coefficients were negative and significant, implying that shocks tended to reverse a recent price trend. The strongest findings are those which related the spike dummy to the price trend over the previous twenty-two day period, which is the longest holding period we considered. The fact that even in these cases the magnitudes of the spikes are not related to the magnitude of the recent price trend calls into question, however, the interpretation that spikes result from so-called market corrections. If they are corrections, then the magnitude of a correction is not apparently not related to the magnitude of the trend.

This concludes our characterization of price spikes, and more broadly of our survey of the stochastic behavior of the price of gold. In the section which follows we will turn our attention to drawing implications expressly for option pricing, which will serve as testable hypotheses bearing on the connection between option pricing, on the one hand, and the behavior of gold on the other. After characterizing the pricing of options, we will address the fundamental issue, which is a reconciliation of option pricing with the behavior of the underlying price series.

Implications for Option Pricing.

Up to this point we have characterized the behavior of gold prices without regard to options on gold. Indeed, there is as yet no reason to believe that any of the systematic behaviors which we have found in the data have any special relevance to options whatever. Now we want to be more pointed, and to frame so-called "stylized facts" against which we can directly test a sample of option data. We will confine our discussion at this point to merely stating the stylized facts without much explanation. We leave the explanation to later, when we actually relate them to option prices.

Stylized Fact #1.

The day-to-day volatility of gold itself varies over time. Over the period running from the start of 1983 to the present, a two-volatility-state model gives annualized standard deviations of 10%, in the low vol. state, and 30% in the high vol. state. Over the whole sample, the standard deviation averages about 13% annualized. One rather obvious implication, then, is that options which are priced with reference to this experience should have implied vols. Somewhere in the range of 10% to 30%, and observations at the high end should be rare.

Stylized Fact #2.

The same two-state model produces evidence of a high degree of persistence of volatility, which is seconded by the autocorrelation of errors from the regressions reported in the Table 1. We estimate the conditional probability that a high vol. day is followed by another high vol. day at about 80%, whereas the unconditional probability of finding a high vol. day is only about 20%. The significance of this fact is that options are quoted for extended expirations, and not just one-day-ahead. They are presumably priced on the basis of the average degree of volatility over that time, and not to the one day vol.

Stylized Fact #3.

Price innovations in spot gold contain a significant transitory component which has a decay rate measured in days. Net of this transient component, prices seem to exhibit price momentum. Over short horizons – horizons of a few days – the decay of the transient element dominates the momentum terms but over horizons exceeding one week price changes exhibit positive momentum. In terms of price volatility, the volatility of price changes is an increasing function of the term over which the volatility is measured, or projected, in the case of option pricing. Referring to the statistics presented in Table 1, we find that one day volatility is only about 75% of the vol. one would expect from a random walk model.[13] Over a holding period of 30 days, however, the variance is almost four times as large as a random walk implies.[14] We should not simply extrapolate this estimate to holding periods which greatly exceed the ones in the data set. It is obvious that, notwithstanding our finding of a trending component of price, that trend has not increasingly dominated the price history. Quite the contrary, looking over long holding periods we find not trend at all. The longest holding period used to estimate the statistics found in Table 1 was about one month, so we can conclude with some confidence that the trending component of price does on average prevail over holding periods which extend for several weeks, in any case.

Stylized Fact #4.

Referring back to our model which allows for two distinct volatilities, large price shocks are not especially persistent, and may actually be less persistent than other shocks. We found that on average, the price of gold recovers about 10% of the magnitude of a shock in the three days which follow it, and that thereafter there is no evidence that we found of any persistent effect of the large shock. It follows that the daily variance in the high volatility state – a standard deviation of roughly 30% annualized, or 1 ½ % for the day – overstates the contribution to variance over longer holding periods contributed by high vol. periods. This statement should not be confused with the observation made above that the volatility of price innovations is highly autocorrelated; high volatility is persistent. It does imply however that the volatility associated with these high vol. periods is effectively overstated by about 10%.

It is especially worthy of note that large price spikes do not apparently initiate a trend in the same direction. The simple calculations of cumulative volatility associated with trending, which were discussed in connection with Fact 2 above, show how rapidly variance can grow over a holding period if price innovations contain an appreciable trending component. If large price spikes tended to initiate a trend, the cumulative impact on price volatility could easily grow to a very large magnitude, relative to the average one day variance of price innovations. That is, however, not the case in this sample.

A Survey of the Recent History of Options on Gold.

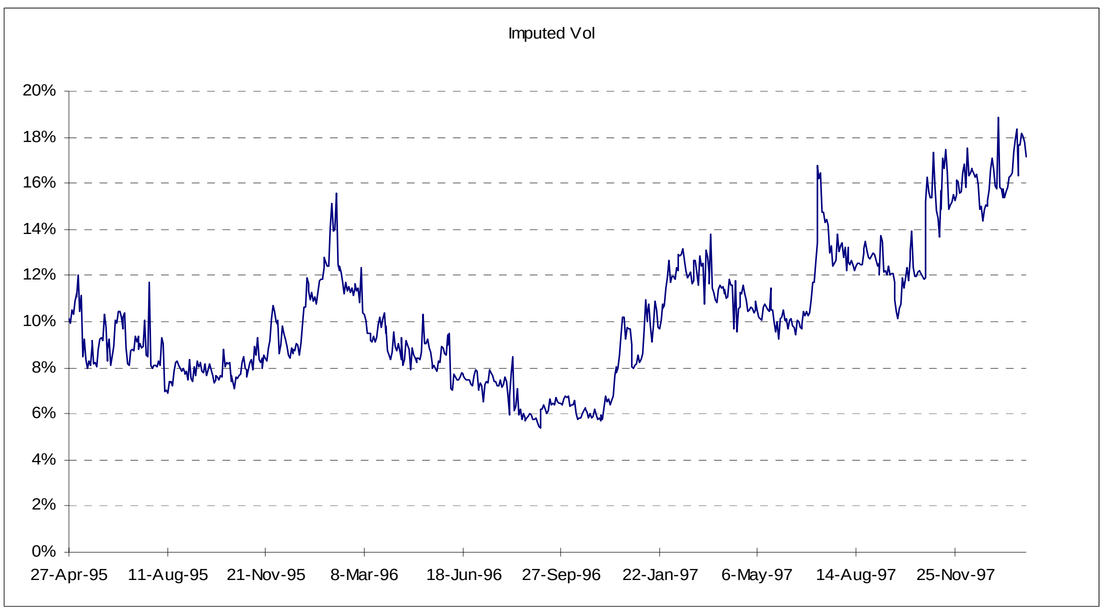
Before we delve into a direct application of these stylized facts to actual option pricing, is may be well first to simply summarize the history of options which is available to us. First of all, we note that the available history is a lot less extensive than the history we have on gold itself. Our data on options extends for about two and a half years, from April 27, 1995 to the beginning of February, 1998. It is daily data, so we have about 650 observations. This count dramatically understates the amount of data involved in this data set, however, because each day we have prices on anywhere from ten to fifteen call options and an equal number of put options. In all, we have about 18000 option prices to work with. This data comes from the Wall Street Journal. Because the earliest option data comes from 1995, when we relate it back to the gold prices themselves we will focus on the behavior of gold since 1982. While the exceptional spike which occurred in January, 1980 is seared indelibly into the memory of most metals traders, its idiosyncrasies should not be priced into options after so many years.[15] Only behaviors which have continued to make themselves felt in the intervening years should impact options priced today.

The term “option pricing” refers essentially to a calculation of the volatility attributed to the underlying series which explains or justifies the quoted option price. While there are many models of varying complexity for this purpose, we have limited our attention to the conventional Black-Scholes model of a European option. We have done so not because we believe that the Black-Scholes assumptions accurately describe the behavior of gold, but because we do not wish to try to anticipate what we will find in terms of discrepancies. Our assumption on this point is that the implied volatilities which we back out of the option model are indicative of the market consensus of price volatility appropriate to that particular option contract. The mere fact that at one point in time the market prices different options to different volatilities is, of course, rather conclusive proof that market participants do not accept the Black-Scholes assumptions. If they did, only one implied vol. would be possible. It follows, further, that any given implied volatility is not simply a forecast of what volatility will be; these forecasts can not be formed in isolation from the forecasts which are embedded into other option contracts. The implied volatilities which come out of the Black-Scholes model are nonetheless the best available common denominator of option prices, adjusting them for difference in expiration, strike, the price of the deliverable, and so on. As a summary measure of option values, they are about as good for that purpose as yield to maturity is in regard to bonds.

The structure of implied volatilities – a structure which allows for systematic relationships between implied vols. on the one hand and explanatory factors – can be summarized in a functional relationship, which usually goes by the name of the option surface. The dependent variable is the implied volatility. The conventional explanatory variables are the length of time to expiration, the difference between the strike and the current value of the deliverable asset, and the square of this difference. The quadratic term relates to a phenomenon called the "smile." We fitted an option surface to our sample of call options, and it is the parameters of that function which are the focus of attention in what follows. Our model of the option surface takes the following form:

15. $$\text{ImpVol} = \gamma_0 + \gamma_1 * \text{sqrt(term)} + \gamma_2 * [\text{strike / futures} - 1] + \gamma_3 * [\text{strike / futures} - 1]^2 + u$$

Here, "term" is the term to expiration, in days, "strike" is the option strike, and "futures" is the price of

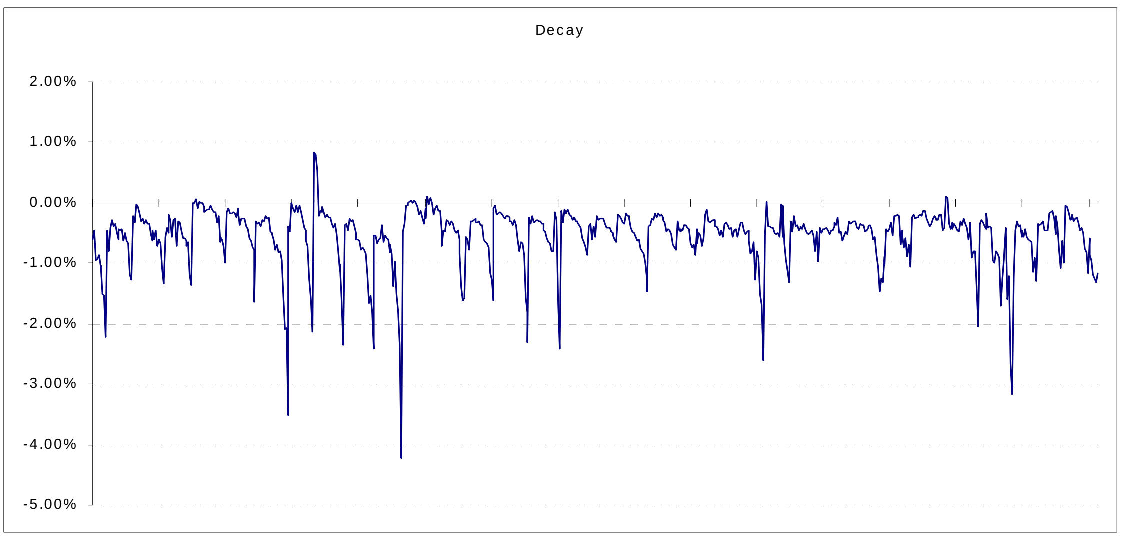


the contract deliverable against the option. The implied volatility of an at-the-money option depends only on the term to expiration. We define a concept called "Imputed Volatility," which equals the fitted volatility at a term equal to 30 days.

16. $$\text{Imputed Vol.} = \gamma_0 + \gamma_1 * \text{sqrt}(30).$$

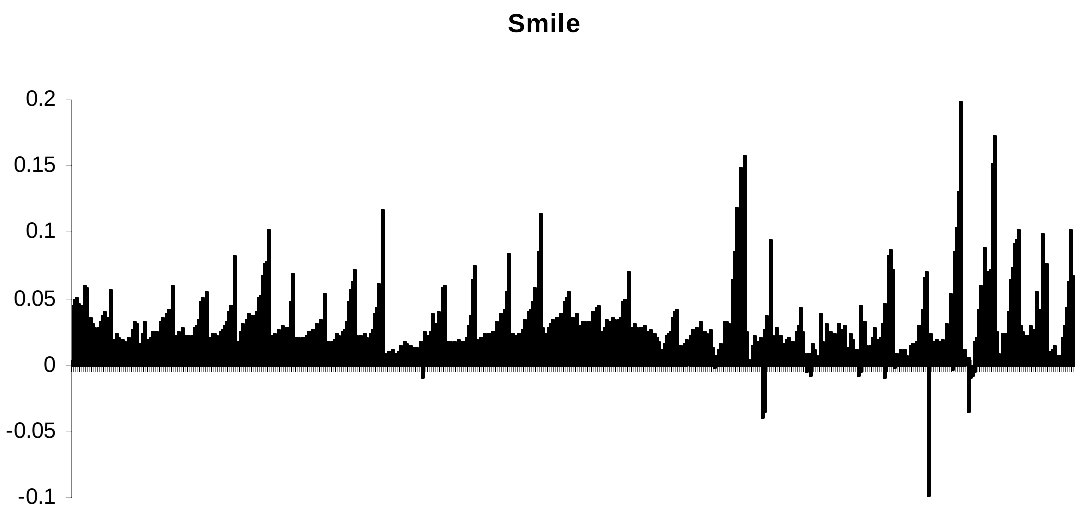
Imputed Volatility is simply intended to be a representative level of implied vol. The time path of Imputed Vol. from our sample is shown in the following figure.

We will not attempt here to give an exhaustive analysis of this data, but will content ourselves with a few general observations. Most important among them is the degree of consistency between this sample of implied vols and the estimates of historical standard deviation. We reported above a finding that spot gold could be characterized by two standard deviations. Our best estimates are 10% and 30%, with about 80% of the sample being drawn from the lower standard deviation. Even though there was one stretch of time when implied vols were priced below 10%, the sample seems quite consistent with the model of spot gold. In the last six months of the sample period imputed vol has risen steadily, as the price of gold has fallen below $300 per oz. Taken by itself, this could call into question the log-normality of the price, because in a log-normal distribution percentage price changes are independent of the level of the price. Even the highest imputed volatilities are entirely consistent with the historical vols, however, and so there is really no compelling reason to question any assumptions.



The time path of the decay coefficient, γ_1, is shown in the graph above. It is interesting that except for a few days, it is really a rate of decay, in the sense that fitted volatility is a decreasing function of the term to expiration. Some days the rate of decay is quite amazing. A decay rate of –2% means that in option expiring forty-nine days in the future would have a fitted volatility 12 percentage points lower than an option with only one day to expiration. Twelve per cent is actually the mean of historical vol and of imputed vol. On the other hand, a term of forty-nine days to expiration is far from exceptional. Options trade, in practice, out to about seventy days to expiration, so this range of implied vols is actually observed rather frequently.

The other notable feature of the fitted decay rate are the spikes which occur periodically. These are associated with expiration days. When options have only a few days to expiration, their implied vols can be astonishingly high. Over that term, implied vols. of around 20% are not at all rare. Since these vols. are for the expiring options only, the rate of decays out to the other options which are being priced on that day is very high in absolute value.



Lastly, we show the corresponding time path of quadratic coefficients, the γ_3 parameter, referred to the "smile." Since the units of this parameters are obscure at best, we simply omitted the scale from the plot. A couple of observations will convey the significance of the scale. First, the horizontal axis is at zero; almost all of the smile parameters are positive, indicating a true smile. Frowns are rare but not unheard-of. Second, most of the smile parameters are significantly greater than zero by statistical measures.

The smile exhibits the same strong seasonality as the decay does: one associated with option expiration. The smile phenomenon is almost exclusively limited to options expiring is a week or two. The option surface beyond that term is essentially flat. Immediately after option expiration day, therefore, the shortest option has about one month to expiration, and no smile is evident. As we approach expiration day, however, the option volatilities of the nearest option begin to curl up, and that imparts the appearance of a smile to the option surface as a whole. The smile and decay parameters also exhibit a strong negative cross correlation in our sample: positive peaks of the smile coincide with negative peaks of the decay rate. The simple contemporaneous correlation is $-.68$, which has an associated t-ratio of around -20. In both cases, the seasonality is characterized by extreme values which approximately coincide with the dates of option expiration.

It is not hard to see why the decay rate might peak close to expiration of the first contract; or at any rate it is not hard to appreciate what that fact implies. Implied volatilities of short term options tend to be very high, even for at-the-money options. At the same times, the smile, which is just a measure of the difference between the implied vols of out-of-the-money options to those of at-the-money options, also peak. Putting these facts together we can conclude that what happens as expiration approaches is that the implied vol. of out-of-the-money options increases dramatically, both in absolute terms – vols. can easily exceed 80% – and relative to at-the-money vols. At the same time, we might reasonably question whether the prices of short term out-of-the-money options actually reflect market clearing prices. We have no record of which options actually traded, or of trading volume in general, so we can not verify that these are actual prices. The rules of the exchange require the options pit to quote a settlement price each day for all options outstanding, whether or not they traded on that day. When the time value of an option is close to zero, it is obviously very easy to overestimate it, and there may in effect be no market process to correct the error.

Both of these parameters, decay and smile, seem to exhibit some progressive change within our sample. For decay, the change is that the peaks on days of option expiration have become much less severe. The maximum (in absolute value) roughly coincides with the peak in imputed volatility. In recent months, by contrast, there is little expiration effect in the sample. This may, however, only reflect a change in the method of data collection which occurred in the sample. In the early month we retained all data on options right up to expiration day, but more recently we have dropped options from the sample when they get within a few days of expiration. If this is what caused the peaks of decay to disappear, it suggests that the rate of decay is small and roughly constant for all options except those which are within a week of expiration. Even then, however, the decay rate is not entirely incidental. At a decay rate of -1, the difference between implied vols. at sixteen days and at sixty-four days is expected to be about four percentage points.

In the case of the smile, the noticeable secular changes are that there seems to be a slight negative trend in the mean, but also a secular gain in the volatility of the smile parameter. The recent, rather chaotic period started at the end of February of this year, and extended through March. Any secular change is, however, dwarfed by the periodic behavior which is associated with option expiration.

Trying to Reconcile Gold Pricing and Option Pricing.

We are at last in a position to formulate some testable hypotheses regarding option pricing and to test them against the option data. We can summarize our findings very easily right at the start: we find that options prices are highly inconsistent with the underlying gold price series. Our conclusion is that this sample of option prices can not be explained by anything we have found about how the price of gold actually behaves.

This proposition requires an important qualification, however, where the options being priced are American, rather than European, options. If volatility is expected to decline over time, early exercise may be warranted after a particularly large price shock which leaves the option in the money. Anticipating this fact, the purchaser of an American option should value it not at the expected average volatility over the stated period to expiration, but to something which also reflects that highest volatility which could prevail out to any intervening point in time.

This is an attribute of option pricing which is exactly analogous to Convexity in the world of bonds, and should probably be referred to by some term like "American" or "Early Exercise" Convexity of an option. It is worthwhile to pursue the analogy a bit further. The significance of convexity for bond valuation is that the stated yield to maturity does not take into account that interest rates will actually vary over the period until maturity. Since the price of the bond is a nonlinear function of its yield, fluctuations over time in the yield, even if the yields average out to the initial yield to maturity, produces fluctuations in the price of the bond which do not average out to zero The expected average price of the bond is always an increasing function of its convexity and of the variance of interest rates. Bond convexity is, moreover, an increasing function of the initial term to maturity (<u>ex</u> any embedded prepayment options). In the same way, the convexity value of an American option is an increasing function of the term to expiration. Since our calculations of implied volatility are based on a model of European options, we should find that, other things being equal, implied volatilities rise with the length of the term to expiration, or equivalently, that longer options appear to be more valuable than near term options.

What we actually find in practice, however, is the opposite phenomenon. While the early exercise option would justify a positive rate of "decay," i.e. growing implied vol, what we actually find is negative decay. This raises two questions. First of all, we still need an explanation of why the decay is negative, a topic which we will address presently. But we also need to understand why the decay is not positive. The value of early exercise convexity depends, among other things, on the degree of leptokurtosis in the price returns. The fatter are the tails of the return distribution, the more valuable the early exercise feature is, because it is in effect an option on volatility. We looked directly for evidence of leptokurtosis in the underlying price data, and came away with compelling evidence of it. The probability that the data sample we looked at, from which we had intentionally excluded the year 1980, came from a single normal distribution, against the alternative of a mixture of normals, is literally less that one in a trillion. The early exercise option, what we have called early exercise convexity, is very valuable.

It is valuable for another reason, which has to do with the autocorrelation of price returns. We found evidence that price returns are negatively autocorrelated in the short run and positively autocorrelated in the long run. The transition to positive autocorrelation occurs at a horizon of between one and two weeks. That is to say, if the price of gold is higher than it was two weeks ago, it will probably be still higher two weeks hence. This is a violation of the random walk assumption on which the Black Scholes model rests, and its effect is to increase the value of long term options relative to short term options, because it in effect increases the variance of long period price returns relative to short term price returns. For this reason also, the "decay" rate would tend to be positive, rather than negative. Since the decay is actually negative, we clearly have some explaining to do because it is actually much more negative relative to a positive benchmark.

The presence of a transitory component of price helps to explain a negative decay. Over a short horizon the variance of price returns is itself more variable than it is over longer horizons. The reason is that even though volatile periods tend to bunch together, there is always a large element of randomness which dictates the volatility of any future day. Over a short horizon we have in effect only a small sample from the distribution of volatilities, and so the resulting average volatility will itself be quite variable, but over long horizons that average daily volatility will be more tightly constrained by the mean of the distribution. The early exercise option diminishes the force of this argument, to some extent, because it implies that the value of an option is not simply based on the average volatility over the whole period to expiration. Nonetheless, it is surely true that the greater predictability of daily vol over longer time intervals must diminish the value to long term options.

This fact can not by itself account for the negative decay for two reasons. First, the decay seems to be much too large, compared with the observed incidence of transitory price shocks. We would have to observe quite frequently very large shocks which were observed to be almost entirely transitory. While there are instances of such shocks within our data, we know from the statistical analysis of prices that they are relatively rate. Even among the large shocks, on average only a small part of the shock is reversed immediately. The other problem is that the spikes in implied vol occur only in the few days which precede option expiration. If it is the presence of transitory price spikes that explains the decay of option vol, we would have to add to that explanation the proviso that large transitory spikes occur only at the expiration of options on the gold contract. Looked at in that light, the thesis has more to do with the event of expiration of the options than it has to do with transience of price shocks. In fact, the conclusion seems unavoidable that the explanation has to be sought for in the event of expiration itself.

As was noted above, our statistics highlight two characteristics of option expiration. We have already discussed the spike in the decay of implied vol. The other phenomenon is the spike in the smile. As we observed, the only evidence of a smile in gold options is to be found in options which are within a week or so of expiration. In that short period, however, the smile can become very intense. These observations suggest one possible explanation, having to do with market manipulation around expiration. Suppose that local traders on the Comex – the people who write options on demand – fear that some of their customers can manipulate the price of gold over a fairly wide range, for a few days. This is not a particular problem in regard to options which are far from expiration, because he can hedge his position or simply weather the storm. For options which are about to expire, however, he can not weather the storm, because the game will close on expiration day. He will have to price in a component of volatility which accounts for the effect of price manipulation.

Interestingly, it is not necessary that we frequently observe large transitory price spikes at expiration. We have to observe some of them, in order to keep the threat of manipulation real and tangible. But it is the threat, and not the actual eventuality, of price manipulation which drives the pricing of the options. This is especially so in the case of out-of-the-money options. How should a local trader react to a request for an offering of options which are deeply opt of the money, with only a few days to expiration? Does the very fact of receiving this request for an offer signal that the bidder is plotting to take his options into the money as soon as he gets them? It is perhaps interesting that in much deeper markets, like Treasury bond futures, option pricing exhibits essentially no smile at all at any time.

We can directly test whether the extraordinarily high decay parameter close to expiration accurately prices the risk of large transitory shocks, because we have data on prices of gold around expiration dates. We will take up this issue in appendix C which follows. The smile, however, can only be explained as a response to a threat – as perhaps a deterrent response – and not to a rational pricing of options in relation to observed price volatility. The reason is that there can be only one best forecast of price volatility over a period of time. Even of the implied vol of at-the-money options equals that estimate, the implied vols out on the ends of the smile must be much larger than this estimate. The smile prices a threat rather than an expectation of price volatility. The threat can be understood as a kind of option. It is an option in the hands of a trader, call him Trader A, to manipulate the price of gold if A succeeds in amassing a large enough portfolio of the right kind of options on the gold contract.

The foregoing discussion has focused on the evidence that options price in not only a rational forecast of price volatility, but a threat of manipulated price spikes as well. The volatility data contains another curious anomaly, which is the level of implied volatilities even in ordinary times. The data is summarized as a time series in the graph above which shows the history of Imputed Vol since September, 1995. We have already expressed amazement at the magnitude of the peaks. Equally remarkable is the height of the troughs. Only a very small part of the period lies below a 30% vol. The results of our estimates of the mixture of volatility model, by contrast, give 30% as the high vol state, which occurs on average only about 20% of the time. The rest of the time, the historical volatility of gold price is 10%! The disparity is not explained by an increase in historical vols. The volatility of gold was very high around 1980, but the next most volatile epoch was that around the price peak in December, 1987, which the price reach $500 per ounce. This episode is within the sample we used to estimate historical volatility, but it is not contained in the sample of implied vols.

135 Addison Avenue, Suite 230
Elmhurst, Illinois, 60126
Tel: 630-833-5658   Fax: 630-833-5686

Appendix A: Notes on Volatility Modeling

These note are devoted to a simple but important issue: how could we develop a model of a stochastic process which not only allows the individual observations, X(t), to have different variances, but also allows the variance to continue into the future. More concretely, which not only allows X(t) to have a different variance than X(t-1), but also allows the variance at time t to influence future variances. This kind of model will accommodate a situation where the process has persistent variance regimes, i.e. long runs of higher or lower variance than normal.

For the sake of brevity, I will limit these notes to the formal development of the heteroscedastic model. We will assume for the purposes of discussion that the stochastic process {X(t)} is drawn from a distribution which is the combination of two normal distributions which have the same mean but different variances. We will use the letter π to denote the probability of drawing from the high volatility subdistribution. $1 - \pi$ is the probability of drawing from the low vol. subdistribution. Formally, we have

A.1 $$X(t) \sim f(x \mid \mu,\sigma_L,\sigma_H,\pi) = \pi * n(x \mid \mu,\sigma_H) + (1 - \pi) * n(x \mid \mu,\sigma_L).$$

In this formula, $n(x \mid \mu,\sigma)$ is the normal distribution with mean μ and standard deviation σ. In one important sense this distribution does not really describe the process by which a value of X(t) is determined. What is really going on can be described as a two-step process, in which nature first chooses a variance, σ_H or σ_L, and then selects an observation from the normal distribution which has that variance. I will assume that at second stage -- choosing X(t) given $\sigma(t)$ -- the choice of X is independent of all previous X's. If the σ's are serially dependent, however, the X's are also. Denote the conditional probability of high variance at time t+1, given high variance at time t by p. Formally,

A.2 $$p = \Pr[\sigma(t+1) = \sigma_H \mid \sigma(t) = \sigma_H] = \Pr[\sigma(t+1) = \sigma_H \text{ and } \sigma(t) = \sigma_H] / \Pr[\sigma(t) = \sigma_H]$$

$$= \Pr[\sigma(t+1) = \sigma_H \text{ and } \sigma(t) = \sigma_H] / \pi.$$

From the laws of conditional probability we obtain the following joint probabilities relating to the probability of different successions:

A.3
$$\Pr[\sigma(t+1) = \sigma_H \text{ and } \sigma(t) = \sigma_H] = p\pi$$
$$\Pr[\sigma(t+1) = \sigma_H \text{ and } \sigma(t) = \sigma_L] = (1 - p)\pi$$
$$\Pr[\sigma(t+1) = \sigma_L \text{ and } \sigma(t) = \sigma_H] = (1 - p)\pi$$
$$\Pr[\sigma(t+1) = \sigma_L \text{ and } \sigma(t) = \sigma_L] = 1 - 2\pi + p\,\pi.$$

I have expressed the basic rules of this model in terms of facts about the variances, because that is the natural way to proceed at a conceptual level. In practice, the situation is very different because we have no way of observing the variance of any random variable directly. All we actually see are the X's, and we have to use them to draw conclusions about the σ's. The way we will do it is to extract from the forgoing assumptions testable conclusions about the X's, which we can observe and test. The method we will use is to relate probabilities of observable events from the parameters of the distribution of the stochastic process. Let B be a positive constant, and define parameters P_1 and P_2 as follows.

A.4a $P_1 = \Pr[X(t) < \mu - B \text{ or } X(t) > \mu + B| \sigma(t) = \sigma_L]$, and

A.4b $P_2 = \Pr[X(t) < \mu - B \text{ or } X(t) > \mu + B| \sigma(t) = \sigma_H]$.

In words, P_1 is the probability that we observe a "large" event from the low vol. distribution and P_2 is the probability of an equally large event coming from the high vol. distribution. Start by computing the marginal probability of a large event.

A.5 $\Pr[| X(t) - \mu | > B] = \Pr[| X(t) - \mu | > B \text{ and } \sigma(t) = \sigma_L] + \Pr[|X(t) - \mu | > B \text{ and } \sigma(t) = \sigma_H]$.

From the definition of conditional probability we obtain

A.6 $\Pr[| X(t) - \mu | > B] = P_2\, \pi + P_1(1 - \pi)$.

We want next to compute $\Pr[\sigma(t) = \sigma_H \,|\, | X(t) - \mu | > B]$.

A.7a $\Pr[\sigma(t) = \sigma_H \,|\, | X(t) - \mu | > B] = P_2\, \pi / (P_2\, \pi + P_1(1 - \pi))$.

This formula come directly from the definition of conditional probability. By similar reasoning, we have

A.7b $\Pr[\sigma(t) = \sigma_L \,|\, | X(t) - \mu | > B] = P_1\, (1-\pi) / (P_2\, \pi + P_1(1 - \pi))$.

As we would expect, these two probabilities add up to 1; the variance has to be either high or low, there are no other possibilities.

The next stage is to draw conclusions about future variances, in terms of $\Pr[\sigma(t+1) = \sigma_H \,|\, | X(t) - \mu | > B]$.

A.8 $\Pr[\sigma(t+1) = \sigma_H \,|\, | X(t) - \mu | > B] = \Pr[\sigma(t+1) = \sigma_H \text{ and } | X(t) - \mu | > B] / \Pr[| X(t) - \mu | > B]$.

The key to computing these probabilities is to expand the joint probability on the right hand side.

A.9 $\Pr[\sigma(t+1) = \sigma_L \text{ and } | X(t) - \mu | > B] = \Pr[\sigma(t+1) = \sigma_H \text{ and } \sigma(t) = \sigma_H \text{ and } | X(t) - \mu | > B] + \Pr[\sigma(t+1) = \sigma_H \text{ and } \sigma(t) = \sigma_L \text{ and } | X(t) - \mu | > B]$.

A.10 $\Pr[\sigma(t+1) = \sigma_H \text{ and } \sigma(t) = \sigma_H \text{ and } | X(t) - \mu | > B] =$

$\Pr[| X(t) - \mu | > B \mid \sigma(t+1) = \sigma_H \text{ and } \sigma(t) = \sigma_H] * \Pr[\sigma(t+1) = \sigma_H \text{ and } \sigma(t) = \sigma_H].$

A.11a $\Pr[\sigma(t+1) = \sigma_H \text{ and } \sigma(t) = \sigma_H \text{ and } | X(t) - \mu | > B] = P_2\, p\, \pi.$

Similarly,

A.11b $\Pr[\sigma(t+1) = \sigma_H \text{ and } \sigma(t) = \sigma_L \text{ and } | X(t) - \mu | > B] = P_1 (1 - p)\, \pi.$

Putting these back into equation 9,

A.12a $\Pr[\sigma(t+1) = \sigma_H \mid | X(t) - \mu | > B] = [P_2\, p\, \pi + P_1 (1 - p)\, \pi] / [P_2\, \pi + P_1 (1 - \pi)].$

In like fashion,

A.12b $\Pr[\sigma(t+1) = \sigma_L \mid | X(t) - \mu | > B] = [P_2 (1-p)\, \pi + P_1 (1 - 2\pi + p\pi)] /$

$[P_2\, \pi + P_1 (1 - \pi)].$

These formulae have a bearing on option pricing, because they relate future variances to historical volatility. They are not operational for our purposes, however, because they still relate to the unobservable condition of being in the high or low variance state.

The last step to develop the model is to get all references to unobservable events out of the model. To be precise, we want to compute $\Pr[| X(t+1) - \mu | > B \mid | X(t) - \mu | > B]$.

A.13 $\Pr[| X(t+1) - \mu | > B \mid | X(t) - \mu | > B] =$

$\Pr[| X(t+1) - \mu | > B \mid \sigma(t+1) = \sigma_H] * \Pr[\sigma(t+1) = \sigma_H \mid | X(t) - \mu | > B] +$

$\Pr[| X(t+1) - \mu | > B \mid \sigma(t+1) = \sigma_L] * \Pr[\sigma(t+1) = \sigma_L \mid | X(t) - \mu | > B].$

Every part of the right hand side has been computed previously; all we need to do is to substitute.

A.14 $\Pr[| X(t+1) - \mu | > B \mid | X(t) - \mu | > B] = P_2 [P_2\, p\, \pi + P_1 (1 - p)\, \pi] / [P_2\, \pi + P_1 (1 - \pi)] +$

$P_1 [P_2 (1-p)\, \pi + P_1 (1 - 2\pi + p\pi)] / [P_2\, \pi + P_1 (1 - \pi)]$

This formula is the one we needed. The probability on the left hand side can be estimated from a sample of X's, because it is a readily observable event which occurs with some frequency. If we had values for this probability and for P_1, P_2, and π, we could solve this equation for p, to obtain an estimate of the probability that the high vol. state continues. Naturally, we can only estimate p with some error, which arises from the fact that we don't know any of the other constants exactly either, but any estimate is better than no estimate at all.

Up to now, we were allowed to pick the threshold B arbitrarily, as long as we could compute the corresponding probabilities P_1 and P_2. Not all choices of B are equally good, however. The best choice is the one which achieves the maximum difference between P_1 and P_2. The right B for that purpose is the point where the two distributions cross. Once we have estimates of the variances, we can use the following formula for B:

A.15 Let $H = \sigma_H / \sigma_L$.

$$B = H\,[\,\ln(H) / (H^2 - 1)\,]^{1/2}\,\sigma_L.$$

Since this threshold makes the difference between the P's a maximum, it will give the most reliable estimate of p.

Equation 14 expresses the parameter p in terms of parameters which you have in hand. All that is needed is to solve the equation for p. Along the way, we can simplify terms somewhat. The right hand side of 14 is divided by $[P_2\,\pi + P_1\,(1 - \pi)]$, which is simply the probability of the event $|\,X(t) - \mu\,| > B$. When we multiply the left hand side of the equation by this probability, we simply get the joint probability of two successive events: $\Pr[|\,X(t) - \mu\,| > B| \text{ and } |X(t+1) - \mu\,| > B]$. One can estimate this probability directly from any set of data, by counting the number of successive events. To simplify the following equations, I will simply refer to this probability as A. Then equation 14 can be rewritten as follows:

A.16 $$A = P_2\,[P_2\,p\,\pi + P_1\,(1 - p)\,\pi] + P_1\,[P_2\,(1-p)\,\pi + P_1\,(1 - 2\pi + p\pi)]$$

$$= p\,\pi\,(P_2 - P_1)^2 + [2\,P_2\pi + P_1 - 2\,P_1\pi]\,P_1.$$

All that remains is to solve 16 for p.

A.17 $$p = \{A - [2\,P_2\pi + P_1 - 2\,P_1\pi]\,P_1\} / \pi\,(P_2 - P_1)^2.$$

Appendix B: Further Results Relating to Equation 14.

Table B.1.a: Coefficients Based on a Fixed Threshold and the Dummy Variable

	0 - 1	*0 - 2*	*0 - 3*	*0 - 5*	*0 - 22*
25%	-.0011	-.0012	-.0018	-.0016	-.0004
	[-2.81]	[-2.32]	[-2.85]	[-2.04]	[-.27]
20%	-.0014	-.0012	-.0018	-.0017	-.0007
	[-2.89]	[-2.00]	[-2.54]	[-1.97]	[-.41]
15%	-.0013	-.0011	-.0019	-.0017	-.0011
	[-2.28]	[-1.55]	[-2.29]	[-1.67]	[-.52]
10%	-.0013	-.0015	-.0016	-.0019	-…37
	[-1.64]	[-1.72]	[-1.55]	[-1.44]	[-1.40]
5%	-.0009	-.0016	-.0005	-.0004	-.0020
	[-.65]	[-1.09]	[-.29]	[-.22]	[-.55]

Table B.1.b: Coefficients Based on a 45 Day Moving Average and the Dummy Variable

	0 - 1	*0 - 2*	*0 - 3*	*0 - 5*	*0 - 22*
25%	-.0005	-.0005	-.0007	-.0004	.0007
	[-1.60]	[-1.24]	[-1.53]	[-.67]	[.50]
20%	-.0005	-.0003	-.0004	-.0001	.0007
	[-1.28]	[-.71]	[-1.03]	[-.10]	[.45]
15%	-.0007	-.0007	-.0012	-.0011	.0012
	[-1.54]	[-1.79]	[-1.95]	[-1.53]	[.69]
10%	-.0006	-.0006	-.0009	-.0007	.0023
	[-1.11]	[-.98]	[-1.12]	[-.80]	[1.06]
5%	-.0013	-.0022	-.0016	-.0013	-.0014
	[-1.39]	[-2.36]	[-1.45]	[-.97]	[1.18]

Appendix C: Price Volatility Around Option Expiration

A key issue raised in this paper is the behavior of historical volatility in the days before option expiration. Our finding relating to the extreme spikes in option volatility at those times naturally leads us to question to what extent actual price volatility also spiked at that time. As we will explain in this note, the evidence is very decidedly to the contrary. Price volatility is by no means usual around expiration day.

The data we used for this investigation is the daily Handy and Harman spot prices, and we have limited the sample to the period which starts at the beginning of 1982. Options on gold futures expire on the second Friday of a month. The month in question is the month preceding the stated month. That is to say, April, 1998 options expire on the second Friday of March, 1998. Gold futures themselves are quoted for six delivery months; those being the six alternate months during the year. Options expire every month, however. Options on February futures expire in December and January, those on April futures expire in February and March, and so on. Most of the open interest in the options is found in the options which expire immediately previous to the delivery month. Thus, open interest in January and March options, which expire in December and February respectively, is very small. Almost all the open interest in is the February and April options, in this example. We have therefore restricted the sample of expiration dates to the six Fridays per year on which most options expire.

We had also to fix a measure of what we mean by “near” option expiration. We used a period of five trading days ending on the Monday following expiration. In an ordinary week, this means that we used the six spot prices from the Monday preceding expiration to the Monday following it. We did not want to stop at the Friday itself, because as was explained previously, spot prices are recorded early in the day, by New York time. Thus most of the volatility of price which actually occurs on Friday in New York would be recognized in the price change from Friday to Monday. Each year there are thirty trading days which by our reckoning fall “near “ expiration dates. The remaining 222 trading days define the benchmark against which we compare volatility. Given the sixteen years from the start of 1982 to the end of 1997, we had a total of 485 days which we counted as being near expirations dates, and 3518 days in the benchmark subsample.

Calculating the volatility estimates and testing for their equality were quite routine. The standard deviation of price change around expiration we estimated to be 1.03%, while the standard deviation over the remaining days was about 1.06%. As usual, the square of the ratio of these statistics is an F statistic, F(484,3517). In this sample, we obtain a sample F = .94. The expected value of F is 1.0, so we find no evidence to contradict that hypothesis. It is extremely unlikely that the price of gold is more volatile around option expiration.

---

[1] Most importantly, the London and Handy and Harman quotations are essentially contemporaneous. Handy and Harman obtain their quote from the Comex spot market, which is in New York, to coincide in time with the announcement of the London afternoon fixing.

[2] Strictly speaking this fact also depends on the homoscedasticity of the increments. In practice, however, it is permissible for the variance of a one period increment to vary over the sample as long as it returns to some sort of average level reasonably quickly. In that case, the average variance over a holding period reverts to the single average.

[3] The potential number of samples, given the lags that we used between points to sample, is on the order of 10 raised to the 3000th power.

[4] Since the Null and Alternate hypotheses concern both the linear and quadratic coefficients, a single F-test as needed, rather than individual t tests. The appropriate F test rejects the Null hypothesis resoundingly even without adjusting the degrees of freedom.

[5] To form 125 independent samples of 1100 observations each, we would need 137,500 days of data (i.e. 125 X 1100). The is two hundred fifty years of business days! This simple calculation is a pretty sobering reminder of how quickly degrees of freedom disappear in time series data.

[6] Actually, we have estimated a model with first order autocorrelated errors, rather than a model in the conventional OLS mold. This fact only deepens the mystery, however, because it implies that the conflict with OLS is not confined to autocorrelation of the regression residuals.

[7] While the public has warmed significantly to the commodity trading business, even at this writing, 1997, only a small fraction of Americans have ever had an open position in any commodity market.

[8] Equivalently, about 81% day-to-day, on the basis of calendar days. The calculation uses the fact that the day-to-day autocorrelation decays exponentially over a period of days.

[9] We have separately estimated the same statistics for the subsample of observations starting at the beginning of 1990. This part of the historical record exhibits a moderately high degree of momentum in both weekly and month price changes, though the degree of momentum, i.e. of positive autocorrelation, is not even close to what we observe in the 1970's and early 80's.

[10] The essential parameter to estimate is the ratio of variance of a non-trading day to the variance of an open trading day. Call this ratio α. Then a period which contains N_1 open trading days and N_2 non-trading days has variance proportional to $N_1 + \alpha N_2$. Strictly speaking this formula assumes that daily innovations are independent, which we already know to be false. The autocorrelations reported in the preceding section are small enough, though, that the resulting error is not material.

[11] A year is equivalent to 250 trading days plus 115 non-trading days, and each of the latter is equivalent to .383 days. A year is therefore 250 + .383 * 115 trading days.

[12] The reason is of course that we can use only a small part of the sample, the so-called big event days, as a basis for estimating A and p. The size of the standard error of p is a sobering reminder of the difficulty of estimating more complicated mixture of distributions models. If we allowed for three, rather than two, variances, our statistical estimates would be based on the relative frequency of "medium event" days and truly "big event" days, in contrast to the number predicted by a single normal distribution. While we would probably find evidence of a third variance, the very small differences between expected and actual number of occurrences would sharply limit the achievable accuracy of any estimates.

[13] Specifically, this fraction is the sum of the linear and quadratic coefficients, which for Sample B are .646 and .117, respectively.

[14] Again using Sample B, this estimate is .646 + .117 * days. As explained in the text, the "days" in this formula are not simply calendar days because non-trading days contribute less variance than do open trading days. A Calendar month is, by our estimate, equal to about twenty-four trading days.

[15] The frantic rally and crash in 1980 was arguably associated with an attempt to corner the markets which failed disastrously (for the would-be aggressors) and will not soon be repeated. It was also in part the climactic event in the long recovery of the market from the period of price controls. It is quite sobering to observe how lengthy and painful is the adjustment of an asset market after a long period of price control.