# Modeling Locational Spreads in Natural Gas Markets

*By: David Soronow and Cheryl Morgan, Financial Engineering Associates, Inc.*

Spread Options based on the price differential between two commodities are some of the most common types of derivatives found in energy markets. Spreads can be categorized into four primary categories:

- Margin or Refining Spreads
- Quality Spreads
- Calendar or time spreads
- Basis or locational spreads

The techniques used to model each type of spread are fairly similar. In this article we will illustrate those techniques with reference to locational spreads in natural gas markets. In a follow-up article next issue we will show how those same techniques can be applied to other types of spreads.

Options based on the locational spread between two prices are some of the most common types of derivatives found in the natural gas markets. These locational or basis spreads mimic the price differential of natural gas between two different delivery points. For example, basis spreads are traded on the price differential between most of the active hubs in the US such as Henry Hub and Katy Hub. But how, precisely, should those spreads be modeled?

## One-Factor versus Two-Factor Approach

When modeling any random process we must decide on the number of independent sources of uncertainty to incorporate into the model. For spread options, the process under scrutiny is the evolution of the spread through time. Traditionally, models of the spread have taken either a one-factor or a two-factor approach.

The one-factor approach explicitly models the spread by treating the value of the spread itself as a single stochastic factor. In contrast, the two-factor approach models the spread indirectly and assumes each price used in constructing the spread represents an independent (although usually correlated) source of risk. The two approaches have a variety of advantages and disadvantages. For example, the single factor approach is simpler and requires collection of only one set of price and volatility data. In addition it does not require us to understand how the two prices that create the spread are correlated.

On the other hand, the two-factor approach can produce more useful diagnostics. Since the one-factor approach models the spread directly, the model produces a delta with respect to the spread itself. Conversely, the two-factor approach models the spread indirectly and produces two deltas: a delta with respect to each individual price. The two-factor approach is clearly preferably if the only available hedging vehicle is a position in each factor and the deltas of the two components are likely to be very different.

## Lognormal Distribution versus Normal Distribution

In addition to the number of separate factors that we use, we must also decide how to model those factors. The lognormal model assumes that the factor under consideration follows a lognormal distribution. In other words, continuously compounded price returns follow a normal distribution. This distribution is asymmetric, positively skewed, and most significantly assumes that the factor being modeled cannot take negative values. Most practitioners agree that the lognormal model is a better approximation of the behavior of energy prices, but does that also hold good for a spread?

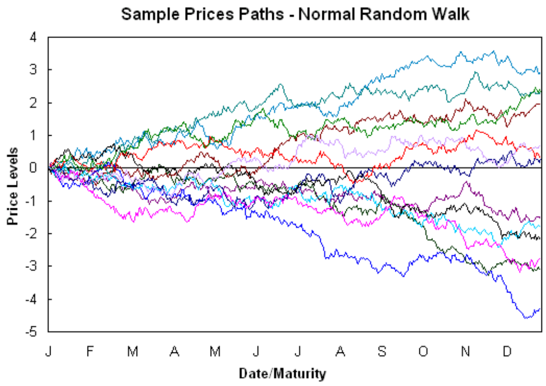
Clearly not, because if we assume that the spread itself is lognormally distributed then we have implied that the spread cannot become negative. For some spreads, we have no means of knowing which of the two prices in the spread is going to be the greater at any time, and therefore we have to allow for negative spreads. The lognormal assumption would therefore impose a serious limitation on the one-factor model. In the two-factor model, on the other hand, each factor represents a price. The assumption that natural gas prices cannot take negative values is much more reasonable, and the spread can be positive or negative depending on which price is greater.

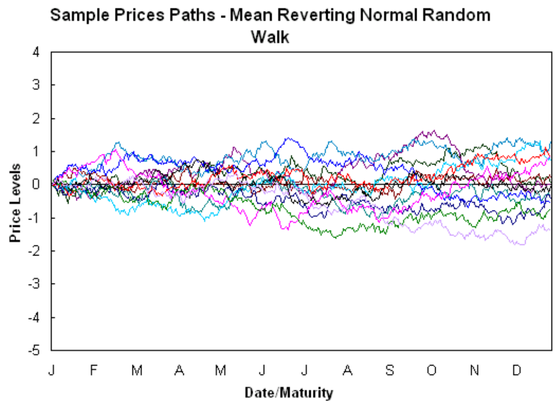
The normal model assumes that the factor being modeled follows a normal distribution between any two points in time. Unlike the lognormal distribution, the normal distribution is symmetric and allows for negative values. For example, with a mean of zero a normal distribution assumes that negative values are just as likely as positive ones. This is clearly the sort of distribution that we are looking for with a spread. In the two-factor model of the spread, however, using a normal distribution for each of the prices would be inappropriate. Gas price distributions are neither symmetric, nor are they capable of negative values.

## Pure Random Walk versus Mean Reverting Random Walk

Most Brownian motion types of models used by practitioners assume that the variance of the distribution (either normal or lognormal) of the stochastic factor grows linearly with time. In other words, the further out in time we gaze, the greater is our uncertainty about the value the factor will take. If the factor being modeled is the spread, then this assumption implies that the distribution of the spread grows larger with time. For the two-factor case, this implies that the distribution of each price grows larger with time.

Is this a valid assumption? Most practitioners would argue that the distribution of the spread follows a mean reverting process. In other words, if the spread grows larger or smaller it will eventually gravitates back towards its mean equilibrium level. The same can be said about natural gas prices, that is, prices gravitate toward their mean levels over time. For both one-factor and two-factor approaches we should therefore use a mean reverting model for our stochastic processes.




## Gas Transportation as a Locational Spread

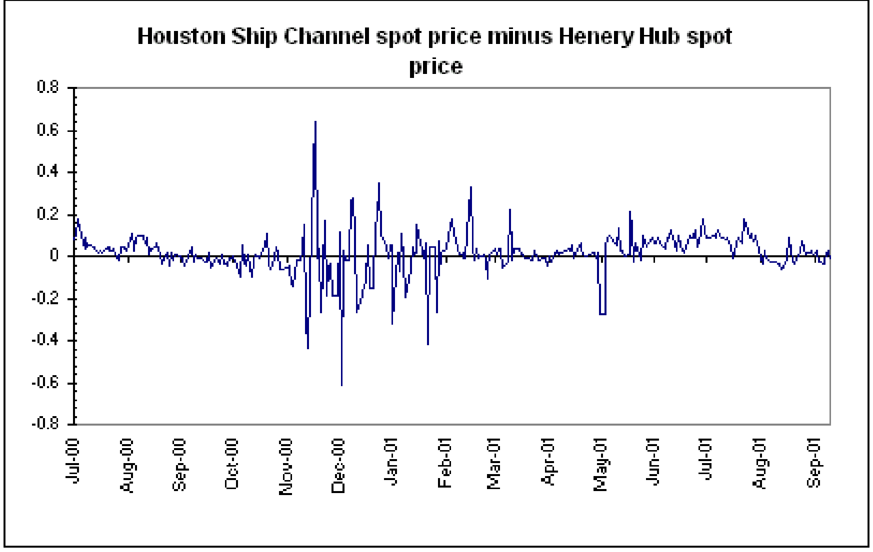
By the very nature of the physical transportation constraints of the commodity, natural gas has a different price for delivery in each of the many locations throughout the United States. The most commonly quoted location is Henry Hub, which interconnects with numerous pipelines and is the delivery point for the NYMEX natural gas futures contract.

What is the price relationship between say Henry Hub and Houston Ship Channel? Is the two-factor approach suitable? Should we use a normal or a lognormal model?

One could argue that the one-factor approach is more suitable given that a strong arbitrage relationship exists between Henry Hub and Houston Ship Channel due to the close geographical proximity and ample pipeline capacity between the two locations. The ability for market participants to transport gas from one location to the other ensures that prices do not wander too far out of line with one another.

For example, if the price at Henry Hub is greater than the price at Houston Ship Channel plus the transportation costs between the two locations, then market participants will buy at Houston Ship Channel, sell at Henry Hub, and fulfill the sell obligation by transporting gas from one location to the other. This will act to increase supply at Henry Hub and increase demand at Houston Ship Channel, thereby placing downward pressure on the price at Henry Hub and upward pressure on the price at Houston Ship Channel. If enough pipeline capacity exists between the two locations, then this arbitrage activity will cause the spread to gravitate towards an equilibrium arbitrage-free level.

The accompanying figure shows that this argument is borne out in practice. The variation in the spread is very small, with a mean close to zero and an amplitude that only rarely exceeds 20c.




The fact that the two prices generally move in step with each other allows us to use the much simpler one-factor model with confidence. As we can see from the figure, the spread between two locations often becomes negative so we employ the normal distribution assumption for modeling it.

With the single factor approach the change in the spread between any two points in time consists of a drift term plus a random term, which represents "noise". Earlier we argued that the spread gravitates towards an "equilibrium" level governed by the cost of transportation. Given this tendency for the spread to gravitate towards an equilibrium level, a mean reverting process is the most suitable.

If we were to simulate the spread value each day into the future, we would see that the mean reverting assumption causes the spread to drift towards the equilibrium level. A higher mean reversion rate implies quicker reversion to the mean level. At any point in time, the spread value is a function of the equilibrium spread value, plus the undecayed distance between the prior day's spread value and the equilibrium spread value, plus the normally distributed random disturbance for that day.

**Simulating the Mean Reverting Normal spread**

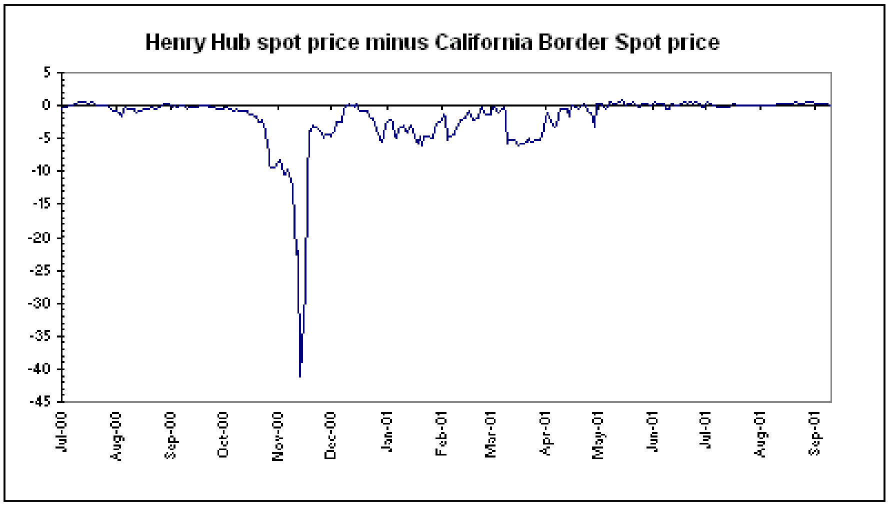
$$S_T = e^{-a(T-t)}[S_t - \bar{S}] + \bar{S} + \sigma\sqrt{T-t}\varepsilon$$

Where:
$S_T$ is the spread at time T
$S_t$ is the spread at time t
$\bar{S}$ is the equilibrium spread value
$a$ is the mean reversion rate
$\sigma$ is the dollar volatility of the spread
$\varepsilon$ is a drawing from a standard normal distribution

## Locational Spreads: Life is Not Always Simple

The one-factor mean reverting normal model is most reasonable for natural gas hubs that have a high degree of integration into the pipeline system and have significant pipeline capacity between the two locations. As we have demonstrated, the prices at each location will not only depend on the demand at each hub, but also on the available pipeline transportation capacity between the two hubs. The one-factor mean reverting normal model ensures that the equilibrium spread value being assumed by the model stays within an arbitrage bound governed by the cost of gas transportation.

But what happens if we look at a spread between hubs that are located in geographically disparate regions of the country with limited pipeline transportation capacity? Does the spread follow a similar process, and can we use the same model?




The degree to which the spreads are predictable depends on the degree to which the spread is governed by transportation arbitrage. This relationship is time dependent and the spread may "blow up" when demand at one location exceeds a threshold pipeline capacity level. Of course, the lower the available pipeline capacity, the lower the threshold level and the greater the chances of the spread "blowing up". Our second figure shows what happened to the spread between Henry Hub and the California Border during the time of the California Energy Crisis. At this time natural gas was in great demand in California and pipeline capacity was not available to meet that demand. As the figure shows, the spread spiked alarmingly, reaching a magnitude of over $40 at one point. For several months the magnitude of the spread was around $5. This is much larger than the 20c we observed between Henry Hub and Houston Ship Channel.

For region pairs where limitations on pipeline capacity cause prices to diverge even under regular conditions, (in other words, the threshold level is often surpassed) the two-factor model is more practical since it is reasonable to assume that the price in each region has its own source of uncertainty. Obviously, changes in the price at each location will have a high degree of correlation since gas prices in general tend to be influenced by similar nationwide issues. But they will also diverge according to local conditions and, as a matter for practical hedging, we will be interested to know the delta of our spread option with respect to both individual prices, not just to the spread.

For the two-factor model, the lognormal distribution is preferred since it prevents gas prices from becoming negative and also assumes that changes in the price are larger when the price is higher. This is consistent with observed behavior of gas prices. Also, a mean reversion model for each price is usually the best approach given that prices gravitate towards equilibrium levels dictated by the cost of production and level of demand.

In this article we have shown how different approaches to modeling locational spreads are appropriate dependent on the precise nature of the spread in question. Next month we will show how the same arguments apply to other types of spread options.

*David Soronow and Cheryl Morgan are derivatives specialists in the Global Support and Professional Services Group at FEA. Soronow and Morgan thank Dr. Angelo Barbieri and Dr. Carlos Blanco for advice and assistance. E-mail Soronow at davids@fea.com and Morgan at cheryl@fea.com.*

.