Valuing Natural Gas Storage
Using Seasonal Principal Component Analysis

By Carlos Blanco, Ph.D., Director of Global Support, FEA
and Paul Stefiszyn, Derivatives Specialist, FEA

In our previous column, we outlined the development of a multi-factor model of the evolution through time of the forward price curve of a commodity using Principal Component Analysis (PCA), and we briefly introduced the idea of Seasonal PCA. This month, we demonstrate a practical application of the technique.

Certain instruments or contracts in commodity markets have a value that depends on more than one forward price. To value these contracts, it is necessary to develop a realistic model of the simultaneous evolution in time of all the forward prices; that is, a model of the forward price curve. Through the use of these models, we seek to accurately capture the joint behavior of forward prices, and to know how they move together, rather than how they move in isolation.

Modeling Natural Gas Storage Value

As an example of the importance of modeling the joint behavior of forward prices, consider a contract to store natural gas. Storage of natural gas (unlike many other assets) requires a specialized facility. Demand for natural gas is generally seasonal in nature: more natural gas is required in the winter than in the summer, and therefore prices in winter are generally higher. Operators of natural gas storage facilities have the opportunity to arbitrage these seasonal price differences.

Suppose that you are the operator of a natural gas storage facility. You have the ability to go long contracts for natural gas delivery with the lowest forward price, and short contracts with the highest forward price. Your strategy is to buy low, store the gas, and sell high. As long as you ensure that the rate at which gas is delivered does not exceed your capacity to inject it into your storage facility, and that your total inventory of gas will never exceed the capacity of your facility (or drop below zero), the strategy will make money. This locked-in profit is known with certainty on the first day of trading.

However, you may not be satisfied with such a simple trading strategy. Over time, the calendar spread on pairs of forward contracts will change, and may reveal additional trading opportunities. In the case that a different combination of contracts would be more profitable than our existing positions, we unwind these positions and put on the new combination. Continuing this strategy over time, we can ratchet-up our locked-in profit (net of successive unwind and transactions costs).

So now suppose you are preparing to bid on natural gas storage capacity. What is your potential revenue from such a "rolling" trading strategy? The answer can be estimated with a Monte Carlo simulation of the evolution of the forward price curve for natural gas. However, we must be careful to model the entire forward curve for natural gas in such a way that realistically accounts for changing spreads.

How can we do this? One possible approach to this simulation problem is to directly apply the variance-covariance (V-C) matrix of the forward prices, but this is cumbersome and inefficient.

Natural gas prices for delivery at different future times are highly correlated but not perfectly correlated, indicating that there are fewer sources of randomness (we can think of them as "drivers" or "factors") than there are prices. In such a situation, a more elegant solution to the simulation problem is to apply Principal Component Analysis.

Principal Component Analysis

A simulation model using PCA is more efficient than other multi-factor approaches, and is more realistic than single-factor models. It reduces the dimensionality of the data while maintaining most of the variance and possibilities for joint changes in the forward prices.

To review, PCA is a statistical technique which can identify the main independent components (sources of risk or information) in data (in our case, typically historical prices of natural gas forward contracts). There will generally be as many components as there are forward contracts in the analysis (so if we are analyzing monthly contracts up to 12 months forward, the analysis would reveal 12 components). In data which is highly correlated (such as natural gas forward prices), typically only 2 or 3 of them are significant, accounting for nearly all the variation or "movement" in the data set.

PCA explicitly reveals the importance of each principal component (its "factor score"), which is an expression of the contribution of that source of risk to the volatility of the forward price. Also associated with each principal component is a set of "factor loadings" which define how the price of each forward contract will change in response to a shock to the component.

For the forward price curve of natural gas, the first principal component generally corresponds to a parallel shift in prices, while subsequent principal components correspond to relative price changes (i.e. a change in calendar spreads). We will examine this idea more thoroughly below in the discussion of Seasonal PCA.

Applying PCA to Storage

So let's see how this works in practice. Suppose we want to value a 1 year storage contract. We proceed as follows:

1. Generally we start with historical prices of futures contracts: observation date, expiry date, and futures price.

2. Calculate the V-C matrix from the historical data (it is also possible to apply PCA directly to the historical data without this intermediate step).
3. Adjust the V-C matrix, if desired, to account for market implied volatility.
4. From the V-C matrix, calculate the principal components.
5. Simulate a possible evolution of the forward price curve for natural gas for each successive day from today until the end of the storage period, and apply the trading strategy, as follows:
    a. Start with the observed curve today. Determine the optimal portfolio of long and short forward contracts and calculate the guaranteed payoff.
    b. Draw a random sample from a standard normal distribution for each principal component in the analysis. For example, if our PCA indicates that the first 3 principal components are sufficient to describe the variability in the historical data, we would need 3 random drawings.
    c. Calculate the price shock attributable to each principal component by multiplying each normal drawing by the factor score of the corresponding principal component. Then multiply each set of factor loadings by the corresponding price shock to give the price returns of each forward price. Finally, add together the returns from each of the 3 principal components (we can do this because the principal components are uncorrelated) to get the total simulated return for each forward contract.
    d. Apply these returns to the original forward prices to get the new, simulated forward curve. Determine if the new optimal portfolio is superior to the existing positions, taking into account unwind and transaction costs, and update the locked-in profit.
    e. The new forward curve becomes our starting point for the simulation of the next day's evolution. We repeat steps (b) to (d) for each remaining day until the end of the storage period.
6. The simulation gives us our profit, according to the trading strategy, resulting from a single simulated evolution of the forward curve. By repeated application of step 5, we can calculate the average profit for all the trials, and build a histogram to analyze the distribution.

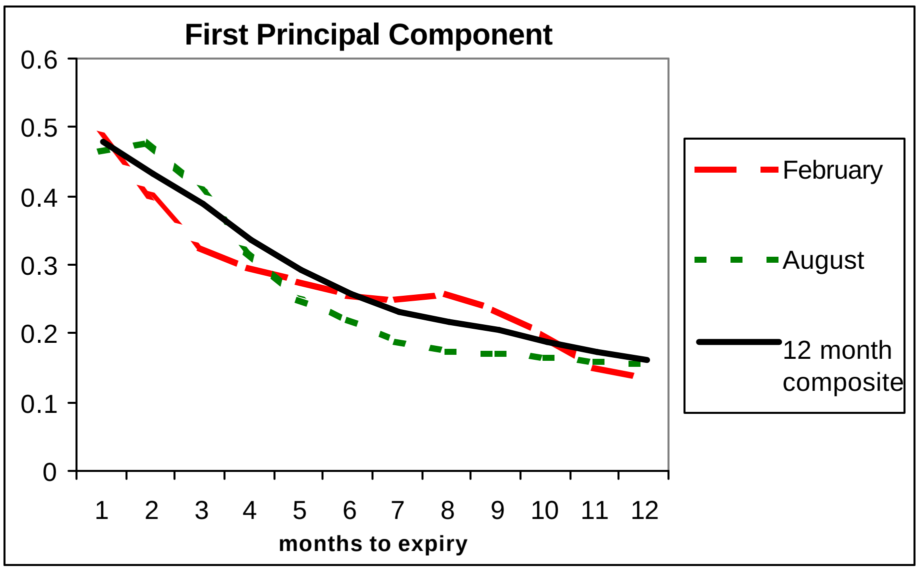
Why should we use Seasonal PCA?

At first glance, the method described above for simulating the evolution of the natural gas forward curve seems reasonable. But how realistically does it capture the changes through time of forward prices for natural gas? The answer is: not very well.

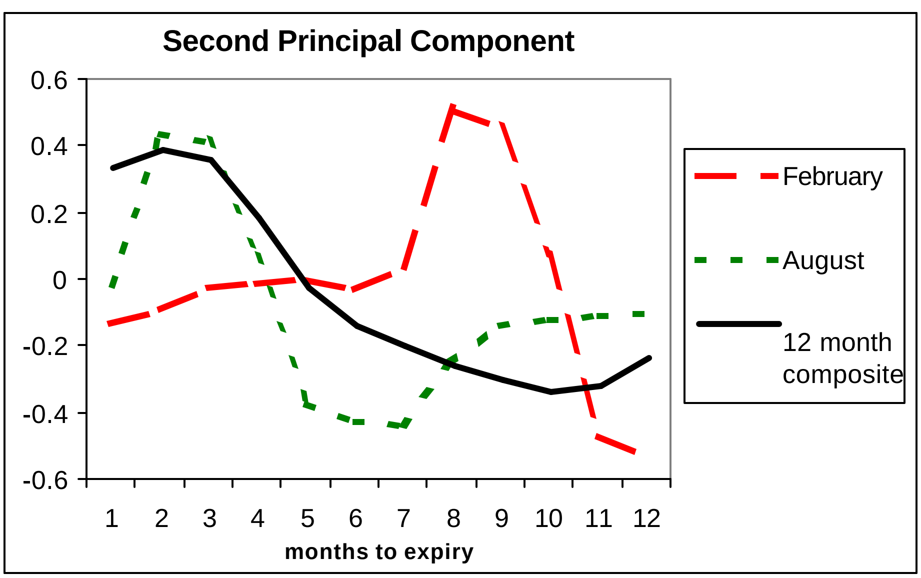
As we said, natural gas prices are seasonal, and the behavior and evolution of the forward prices are seasonal as well. By considering the full set of historical data when applying the PCA, we produce results which capture the average variability of the forward price curve throughout the year. Using these results, we can then model price behavior on a "typical" day. The problem is, the typical day never really exists because the behavior of forward prices in the summer is different from the behavior in the winter.

Our solution is to break up the entire set of historical data into monthly groups (i.e., all the forward price observations made in January, all observations made in February, etc.). We then perform 12 separate principal component analyses, to determine the behavior of forward prices in each of the 12 months of the year. During the simulation of the forward price curve evolution, we model each month turn by turn, applying the principal components which are specific to the month in question.

Let's examine this in more detail. Suppose we have extracted the principal components for each of the 12 "seasonal" buckets, and also for the entire data set.



Comparing the first principal components, we see that the "composite" first principal component derived from the entire data set would be a fair approximation to any of the seasonal first principal components (February and August are given as examples).



However, comparing the second principal components, we see that the second principal component derived from the entire data set is entirely different from the seasonal second principal components, which are themselves very different from each other. In particular, we see that a shock to the second principal components will always load onto those contracts which expire just prior to winter delivery, regardless of whether they are short-dated or long-dated contracts. In other words, the forward price volatility associated with the second principal component depends not on whether the contract is maturing soon, but whether the contract is for winter delivery.

Furthermore, for a typical sample data set which we analysed, the factor score for the composite second principal component was only 6% of the total variance, while the factor scores for those derived from the seasonal analysis ranged from 8% in the winter months to 19% in the summer months. Using the composite analysis would therefore underestimate the contribution of the second principal component to the variability of the data.

Recall that the second principal component models primarily the changes in calendar spreads between futures prices. These changes in spreads are a significant source of value for the natural gas trading strategy explained above. Any modeling strategy that fails to realistically model the seasonal differences between the second principal components may seriously mis-price a natural gas storage contact.

Carlos Blanco directs the Global Support and Professional Services Group at FEA. Paul Stefiszyn is a Derivatives Specialist in the same group. Blanco and Stefiszyn thank Cheryl Morgan, Dr. King Wang, and Dr. Angelo Barbieri for advice and assistance. Email Blanco at carlos@fea.com and Stefiszyn at pauls@fea.com.