Forward Price Simulation

Hao Wu

Risk Management Group
Constellation Energy &
Department of Biostatistics
Johns Hopkins School of Public Health

August 24, 2007

Overview

Motivation

Long term price simulation is useful for:

- Potential Exposure
- Asset valuation
- Portfolio optimization

Challenges

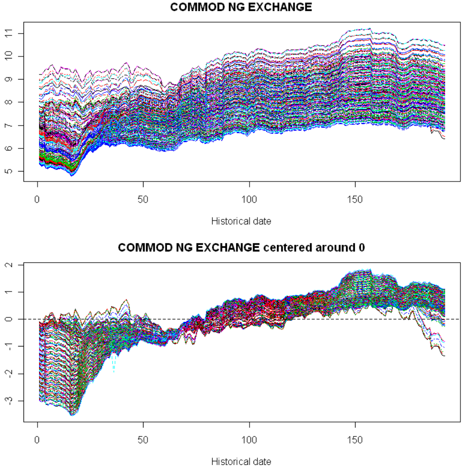
- Simulation of a few correlated time series is trivial.
- Main challenge in this project is the scale of the data.
  - Around 650 curves (commodities)
  - 24 to 60 contract month for each curve
  - Correlations among 30,000 curve/month must be kept

Data Exploration

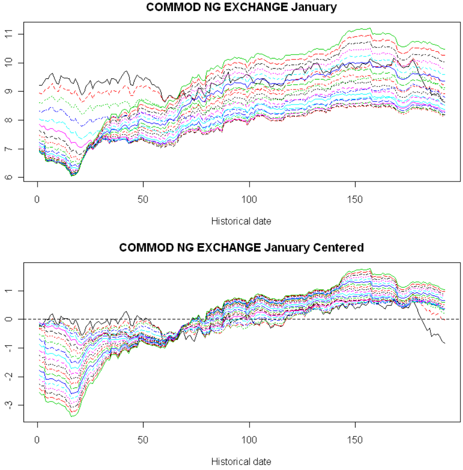
Understand the historical data is crucial in forward simulation.

- Natural Gas, Coal, Oil
- Electricity and its dependency on fuels
- Volatility

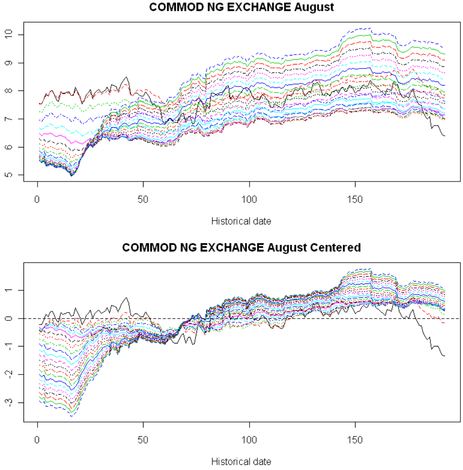
NG EXCHANGE historical prices



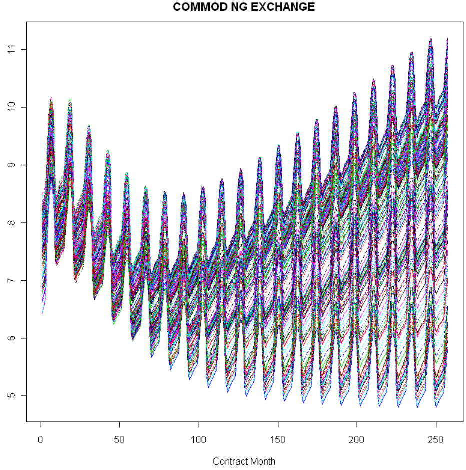
NG EXCHANGE prices - January contracts



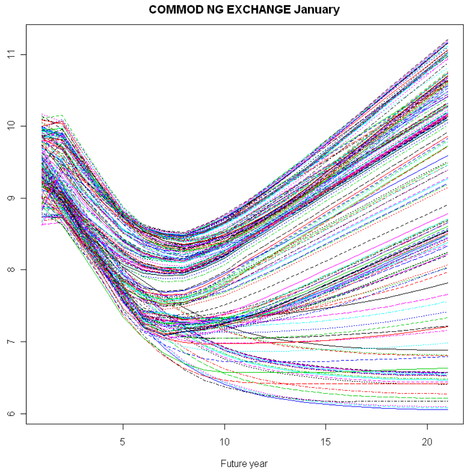
NG EXCHANGE prices - August contracts



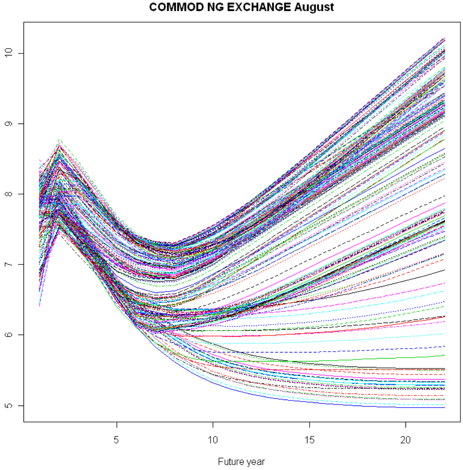
NG EXCHANGE historical prices by contract month



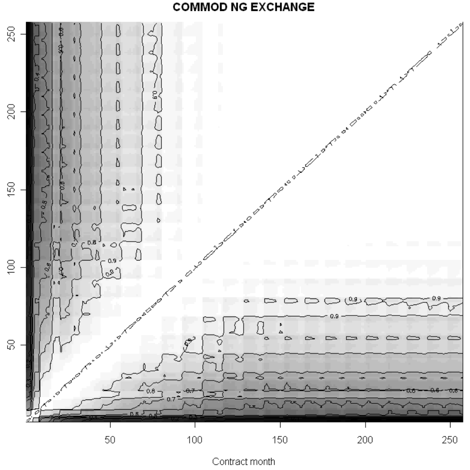
NG EXCHANGE prices by contract month - January contracts



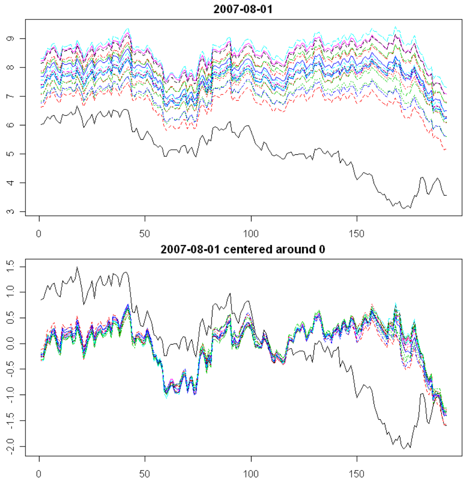
NG EXCHANGE prices by contract month - August contracts



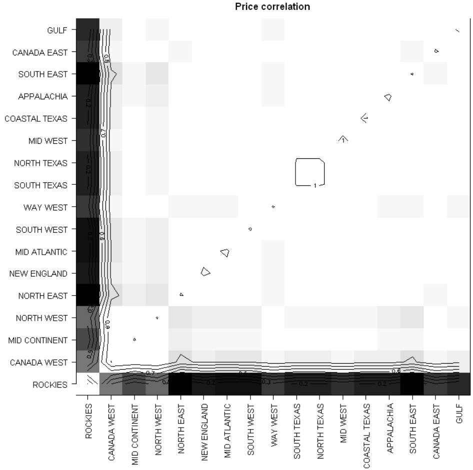
NG EXCHANGE: Correlations among prices for different contracts



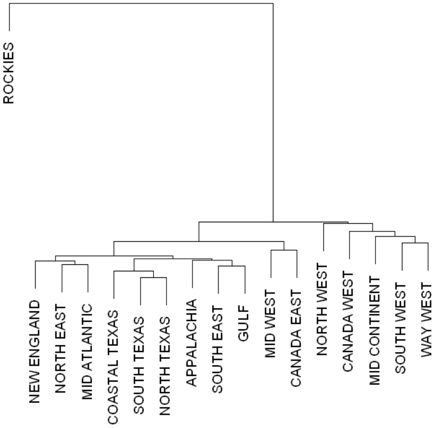
August 07 contracts, 17 regional reference curves



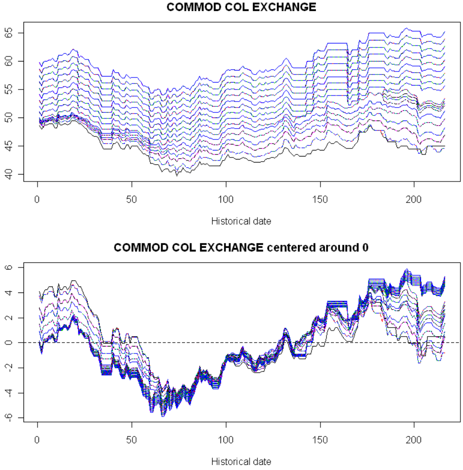
Price correlation among 17 regional reference curves



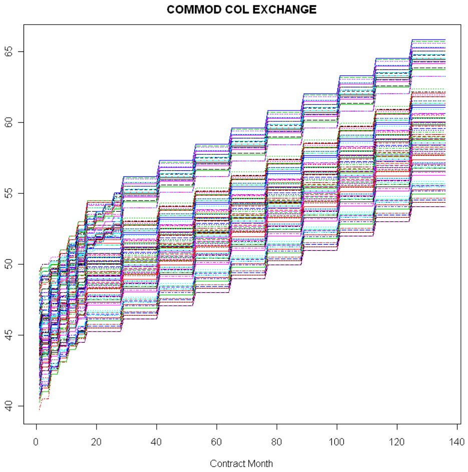
Hierarchical clustering for regional curves



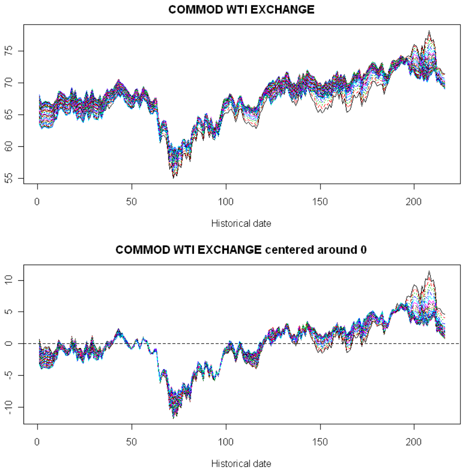
COL EXCHANGE historical prices



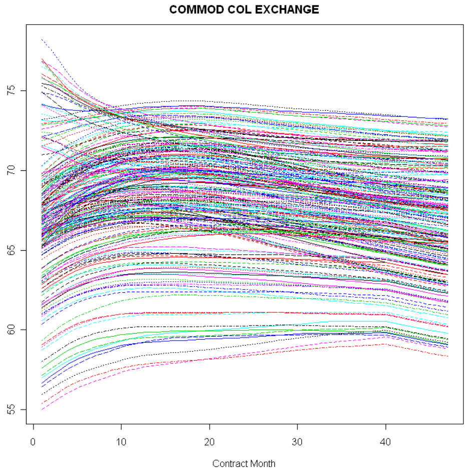
COL EXCHANGE prices by contract month



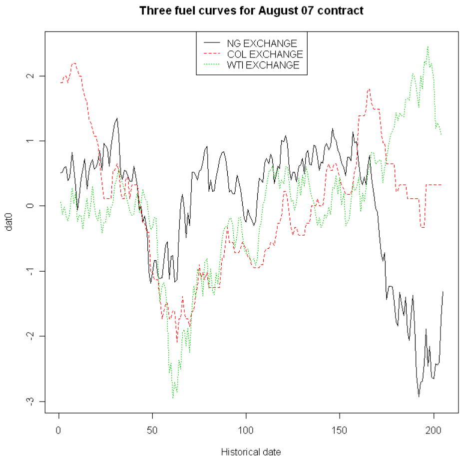
WTI EXCHANGE historical prices



WTI EXCHANGE prices by contract month



Historical prices for three fuels



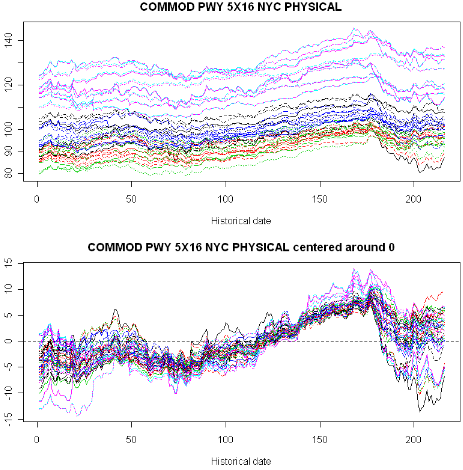
Findings

- Curves are very similar.
- Seasonality in NG, not COL/WTI
- Three types of fuels are mildly correlated.

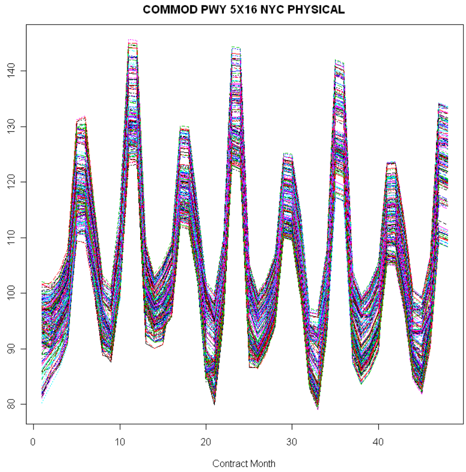
PWY 5X16 NYC historical prices

Forward Price Simulation

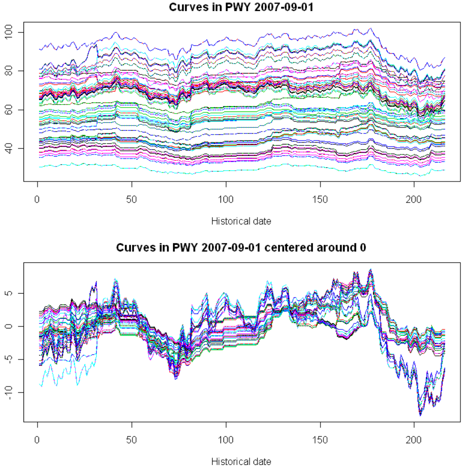
Hao Wu



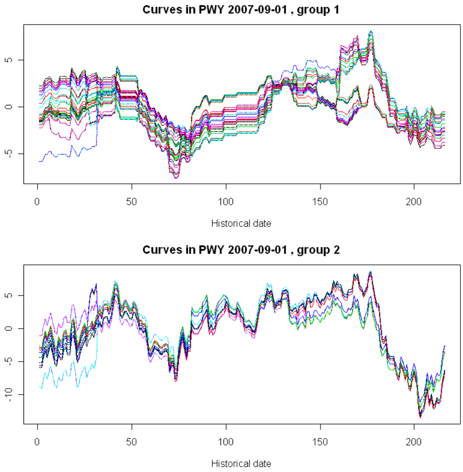
PWY 5X16 NYC prices by contract month



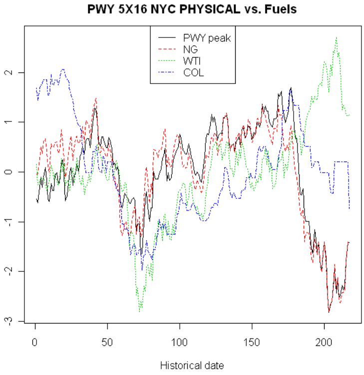
All PWY curves for Sep 07 contracts

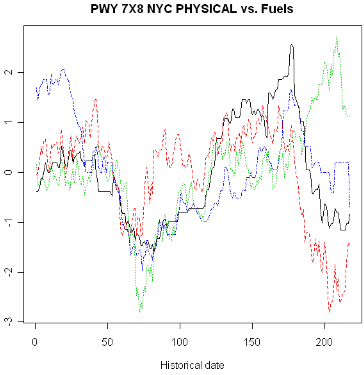


K-means clustering on PWY curves, Sep 07

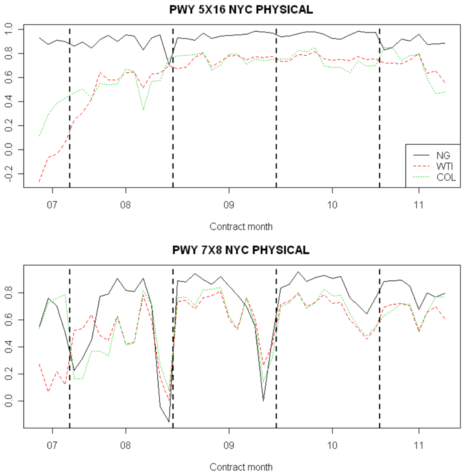


PWY curve vs. Fuels





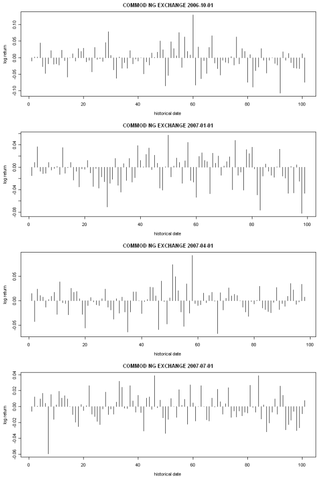
Correlation between PWY curves and Fuels, by contract month



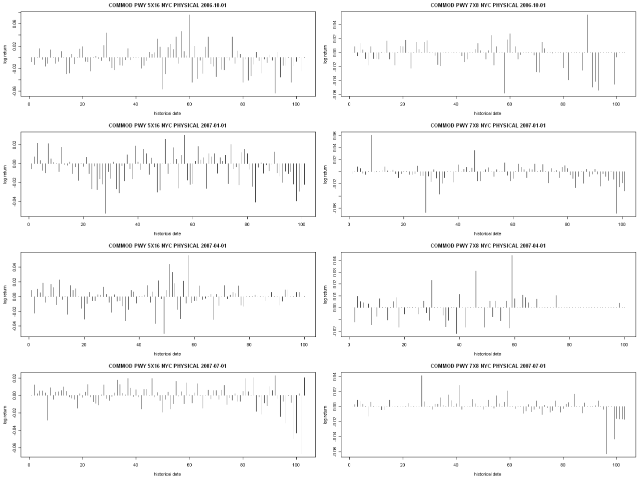
Findings in electricity curves

- For the same curve name, all contracts are similar.
- For the same contract, curves are in groups.
- Show Seasonality.
- Correlated with fuels in different ways.

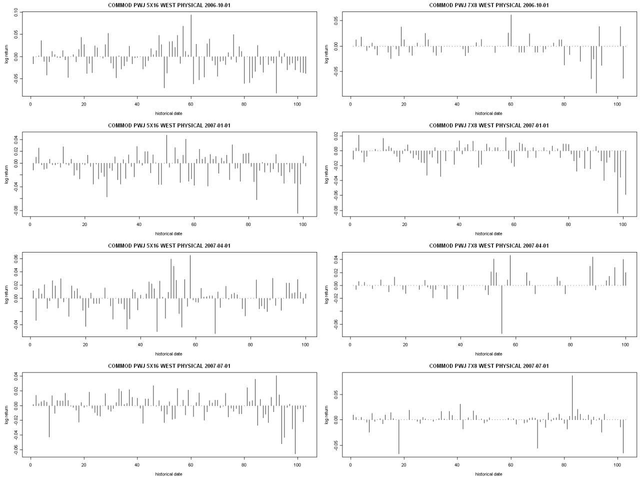
Volatility - NG EXCHANGE



Volatility - PWY NYC



Volatility - PWJ WEST



Findings

- Volatilities for NG didn't change approaching maturity
- Volatilities for power curves seem to increase for contracts in peak season??
- Need more work to validate the findings.

Goals

Our goal is to simulate forward prices for all curves in portfolio (around 30,000).

- To simulate 30,000 correlated time series, dimension reduction (PCA) is needed.
- The result should be consistent with recent history.

Why not do PCA on everything?

- PCA needs to work on the var/cov matrix (30,000 by 30,000).
- Correlations are high among same types of commodities, relatively low between different types of commodities. We will need many PCs to reasonably capture the variance.

Our approach

1. Took a hierarchical view on the data and built a curve pedigree. There are parents and children curves.
2. Use PCA for dimension reduction.
3. Simulate parents correlatedly based on OU process assumption.
4. Simulated children based on simulated parents.

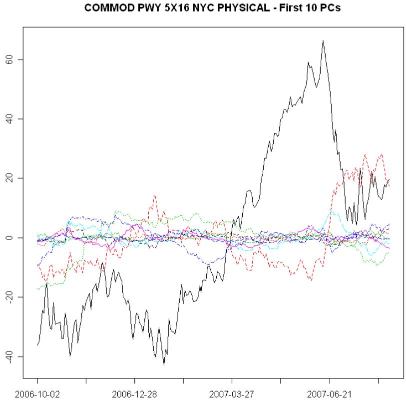
Principal Component Analysis (PCA)

PCA can do:

- Orthogonalization: PCs are independent.
- Dimension reduction: Most of the variances were captures by a few PCs.

PCA result on PWY 5X16 NYC

The first 10 PCs of the historical prices (for 48 contracts and 200 historical days):



PCA result on PWY 5X16 NYC (cont.)

Variance explained by the first 10 PCs:

	PC 1	PC 2	PC 3	PC 4	PC 5
% var	0.7962	0.1240	0.0424	0.0168	0.0083
Cumulative % var	0.7962	0.9202	0.9625	0.9794	0.9877
	PC 6	**PC 7**	**PC 8**	**PC 9**	**PC 10**
% var	0.0032	0.0018	0.0012	0.0010	0.0007
Cumulative % var	0.9909	0.9928	0.9939	0.9949	0.9956

Curve hierarchy

- The hierarchy was made manually based on fundamentals.
- Saved as a flat table in an Excel file.
- Program will read the file and do simulation.
- User will have to maintain this until integration with other IT projects.

Curve hierarchy - Fuels

- One commodity reference curve for each fuel type: NG EXCHANGE, WTI EXCHANGE, COL NYMEX PHYSICAL.
- NG curves were divided into 17 groups by region.
- Each region has a regional reference curve. They are children of NG commodity reference curve.
- Rest of the NG curves are children of regional reference curves respectively.
- COL/WTI curves are children of their commodity reference curves.

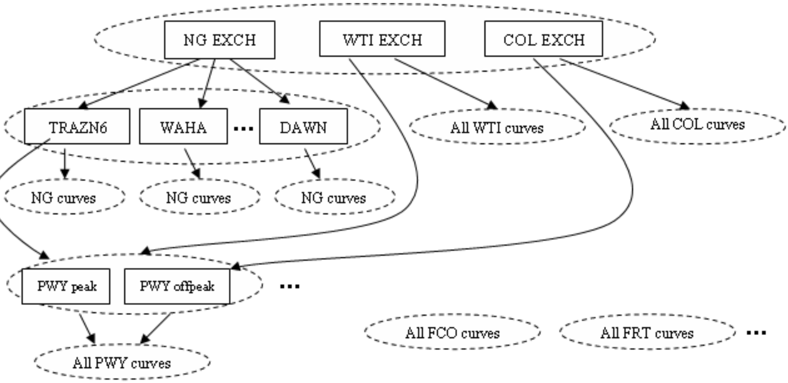
Curve hierarchy - Electricity

- Two reference curves (peak and offpeak) for each electricity market. They are children of WTI EXCHANGE, COL NYMEX PHYSICAL and the corresponding NG regional reference curves.
- Rest of the electricity curves are children of their corresponding market reference curves.

Curve hierarchy - Other market

Other markets includes freight, emission, etc. At this time We didn't put any structure and these markets are simulated independently. The hierarchy can be added easily by modifying the Excel file.

Curve hierarchy illustration



Simulating Parent Curves

One crucial assumption is that log price follows OU process. For several curves their log prices follow correlated OU processes. Simulation steps are:

- Do two rounds of PCA to break the correlations among contracts and curves. Resulting PCs (linear transformation of log prices) are uncorrelated OU processes.
- Estimate OU parameters from historical then simulate forward PCs.
- Transform back to original scale.

Simulating Parent Curves (cont.)

What if OU assumption fails?

Currently we make up some numbers for parameters. This isn't too bad.

- Results are approximately GBM.
- Weak force pulling price back to historical median.

Potentially we can do:

- Use HMM to divide historical into "up", "down", "flat" periods. Estimate transition/emission probability.
- Remove trend and model residuals as OU.
- Simulate both trend and residuals.

Simulating Children Curves

Children curves were correlated to their parents through regression.

Notations for next a few slides are:

- $y(t)$: log historical prices of the child curve for day t,
- $\Delta y(t)$: log return of the child curve on day t.
- $x_p(t)$: log historical prices of the p^{th} parent curve for day t, $p = 1, 2, \ldots, P$.
- $\Delta x_p(t)$: log return of the p^{th} parent for day t,

The following several models were considered.

Regression of log prices (Model 1)

Regress child's log price on parents' log prices:

$$y(t) = \beta_0 + \sum_{p=1}^{P} \beta_p x_p(t) + \epsilon(t). \qquad (1)$$

Problems are:

- "Spurious" regression so R^2 is not reliable
- Assuming children prices change with parents on a daily basis, which isn't necessarily true.
- $\epsilon(t)$ auto-correlated, hard to simulate.

Regression of log returns (Model 2)

Regress child's log return on parents' log return:

$$\Delta y(t) = \beta_0 + \sum_{p=1}^{P} \beta_p \Delta x_p(t) + \epsilon(t). \tag{2}$$

Problems are:

- Fitting is poor with R^2 dropped to below 0.5.
- Child and parents are co-integrated processes: prices roughly track each other, but not on daily basis.

Another regression model (Model 3)

Model **??** can be written as:

$$y(t) = \beta_0 + y(t-1) + \sum_{p=1}^{P} \beta_p[x_p(t) - x_p(t-1)] + \epsilon(t).$$

It regresses child's today's price on its previous day's price and parents' today and previous day's prices, with constrains on the coefficients. We removed the constrains and came up with:

$$y(t) = \beta_0 + \beta_1 y(t-1) + \sum_{p=1}^{P} [\beta_{p1} x_p(t) + \beta_{p2} x_p(t-1)] + \epsilon(t). \quad (3)$$

Model 3 (cont.)

Connection between Model **??** and regression on log prices:

- After regression on log price, do another regress on residuals: $\epsilon(t) \sim \epsilon(t-1)$. The residuals show little AC.
- The residual regression is similar to our model (with 1 less degree of freedom).

Model 3 (cont.)

There's connection between Model **??** and Error Correction Model (ECM) for co-integrated processes.

- From our model, we can get:

$$\Delta y(t) = \beta_1 \Delta y(t-1) + \sum_{p=1}^{P} [\beta_{p1} \Delta x_p(t) + \beta_{p2} \Delta x_p(t-1)] + \Delta \epsilon(t).$$

- ECM doesn't have $\Delta x_p(t)$ because they don't assume to know it beforehand.
- Our model is more liberal.

Final model

Note that the input of children curve simulation are PCs so they were centered around 0. We remove the intercept in regression and came up with our final model:

$$y(t) = \beta_1 y(t-1) + \sum_{p=1}^{P}[\beta_{p1}x_p(t) + \beta_{p2}x_p(t-1)] + \epsilon(t). \tag{4}$$

- Model selection procedure was used to select parents.
- Fitting is good with high R^2
- No auto-correlation in residuals

Some details

There are some details in simulating children curves:

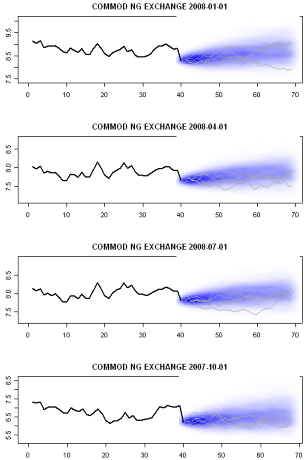
1. Children curves in the same group (market) are simulated together to keep the correlation in residual processes $\epsilon(t)$. What's the correlation among $\epsilon(t)$?
    - Correlation among estimated residuals underestimate the truth because children are co-integrated.
    - I'm using the correlation in log prices $y(t)$.
2. For different forward time step, should residuals $\epsilon(t)$ form the same distribution?
    - Yes. $\epsilon(t)$ can be (kind of) viewed as the co-integrated process of the child and parent curves, so it should be stationary.

Simulation settings

- Use 300 historical days to predict future. Exclude the weekends and holidays there are around 210 historical pricing days.
- Use 5 principal components to represent both the curves and the months. This will capture over 95% of the variance in data most of the time.
- Simulate daily until the end of next month, then do monthly for another 46 months. Totally there are around 80 forward time points.
- Simulation was done for contracts of 48 future months.
- Generate 1000 independent simulations.

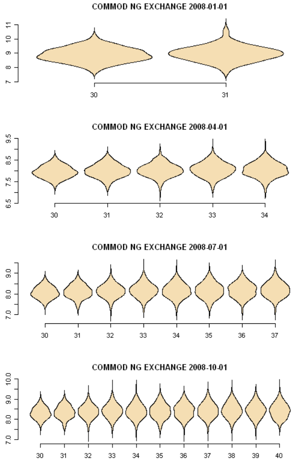
Simulation result for NG EXCHANGE

Daily forward prices for four future contracts:



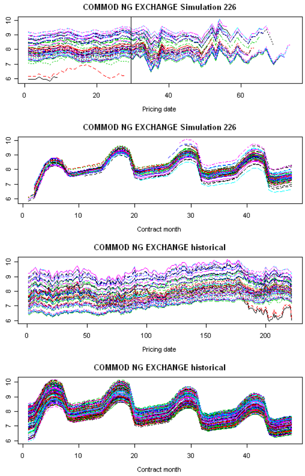
Simulation result for NG EXCHANGE (cont.)

Monthly forward prices for four future contracts:



Simulation result for NG EXCHANGE (cont.)

All contracts, one simulation:



Volatility

Historical vs. simulated volatility for NG EXCHANGE.

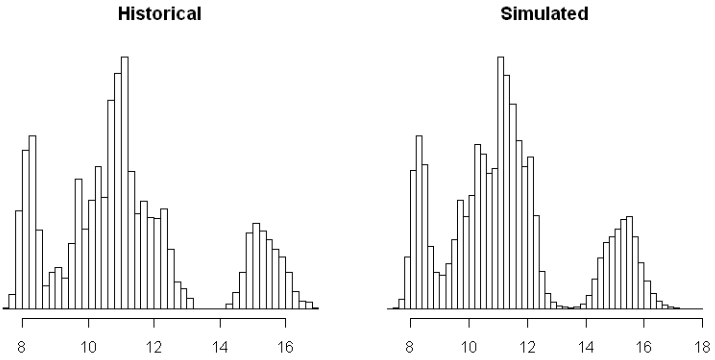
		9/07	10/07	11/07	12/07	1/08
Daily	Hist	0.0228	0.0185	0.0154	0.0144	0.0142
	Sim	0.0245	0.0194	0.0154	0.0142	0.0139
Weekly	Hist	0.0423	0.0347	0.0291	0.0275	0.0271
	Sim	0.0519	0.0390	0.0303	0.0273	0.0267
Biweekly	Hist	0.0557	0.0452	0.0378	0.0362	0.0358
	Sim	0.0409	0.0356	0.0284	0.0258	0.0254

Volatility (cont.)

- Daily volatilities are usually accurate.
- Long term volatilities are very curve dependent.
- When it fails, it's likely that the OU assumption was incorrect, or unstable. Other process assumption worth trying.

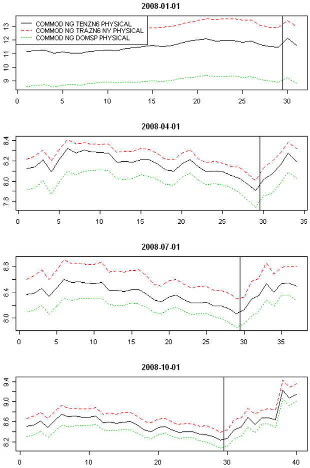
Simulated vs. historical heat content

Heat content for PWY 5X16 NYC PHYSICAL



Simulation results for multiple curves

One simulation for three NG curves:



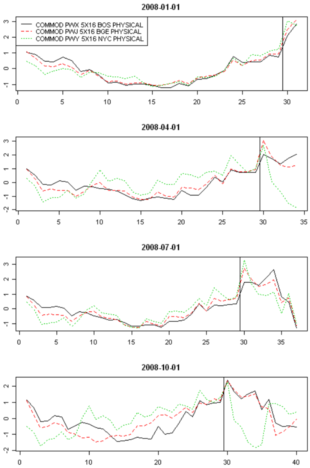
Correlations in simulated curves

Historical vs. simulated correlations among three NG curves:

	Historical		
	TENZN6	TRAZN6 NY	DOMSP
TENZN6	1.0000	0.9988	0.9983
TRAZN6 NY	0.9988	1.0000	0.9976
DOMSP	0.9983	0.9976	1.0000
	Simulated		
TENZN6	1.0000	0.9737	0.9640
TRAZN6	0.9737	1.0000	0.9763
DOMSP	0.9640	0.9763	1.0000

Simulation results for multiple curves

One simulation for three electricity curves in different markets (prices were shifted and scaled):



Correlations among simulated curves

Historical vs. simulated correlations among three electricity curves:

	Historical		
	5X16 BOS	5X16 BGE	5X16 NYC
5X16 BOS	1.0000	0.8643	0.9293
5X16 BGE	0.8643	1.0000	0.9414
5X16 NYC	0.9293	0.9414	1.0000
	Simulated		
	5X16 BOS	5X16 BGE	5X16 NYC
5X16 BOS	1.0000	0.8052	0.8205
5X16 BGE	0.8052	1.0000	0.8813
5X16 NYC	0.8205	0.8813	1.0000

The correlations among simulated curves in different market were underestimated. Why?? Because we correlate them through their parents.

To show this, let X, Y, Z be random variables with $cor(X, Y) = \rho_1$, $cor(Y, Z) = \rho_2$. What's the minimum value of $\rho = cor(X, Z)$ when ρ_1 and ρ_2 are sufficiently large, say, 0.8?

Correlation is projection. So $\rho \geq cos[cos^{-1}(\rho_1) + cos^{-1}(\rho_2)]$.

- $\rho_1 = \rho_2 = 0.99 \Rightarrow \rho \geq 0.96$, this is the case for NG curves.
- $\rho_1 = \rho_2 = 0.90 \Rightarrow \rho \geq 0.62$, this is the case for electricity curves.

Potential solutions are:

- Further divide regions into smaller subregions.
- Use more parent curves.

Implementation

- Software was written in open source statistical programming language `R`.
- One run takes 3 hours on a single PC (2.8GHz, 2G RAM).
- Parallelly distributed using `Condor` (John).

Future Works

1. Stochastic process assumption.
   - Currently use OU process but other models can be plugged in easily.
   - Use HMM to simulate trend?
2. Curve dependency: talk to the traders.
3. Nonlinear dimension reduction (LLE): use a few components to capture nonlinearity in the data.
   - Nonlinearity in data is not obvious.
   - Involves many other calculations. The computational gain is doubtful.
4. Modeling volatility: this is difficult.
5. Econometrics models?

Acknowledgment

- Adrian
- Aram
- John
- Haibin
- Sailesh
- Leyu, Pramod and Aristide
- Karl, Rafa and other professors in Hopkins
- Shijie
- The audience