Time Series Modelling of Pulp Prices

Mônica Barros
Reinaldo Castro Souza

DEE, PUC-RIO

August 1997

Objectives and Overview

- To analyze quarterly (and monthly) data on pulp prices.
- Pulp prices: highly volatile series - large drops in prices have occurred on the past years.
- On recent years, pulp prices had suffered from an erratic behavior, which can be related to global demand for paper and the level of inventories.

We present 2 distinct approaches to the problem:

- 1- Dynamic regression model to forecast pulp prices and,
- 2- Simulation approach based on returns calculated from pulp prices.

Characteristics of Pulp Market and Industry

- Capital intensive
- International (over 80 % of Bleached Kraft Pulp is exported from country of production and there are about 30 producing countries)
- Scattered production - even the largest supplier *cannot* control prices (has less than 6 % of global market share)

Characteristics of Pulp Market and Industry

- Highly Integrated - most companies are both pulp and paper producers.
- Pulp Price increases are not necessarily a major problem for paper producers, since they also increase their prices.
- However, there is some upper bound for paper prices - above a certain level demand tends to decrease.

Characteristics of Pulp Market and Industry

- New pulp plants under construction in Asia and Latin America.
- Supply tends to increase substantially for the next 5 years.
- Pulp prices directly affect profitability of paper mills.
- *Thus, pulp and paper prices tend to move in the same direction.*

Characteristics of Pulp Market and Industry

- Production costs tend to be lower in emerging countries, but technological changes in plants have lowered considerably the costs of production in the USA, Canada and Scandinavian countries, which were historically the largest pulp producers.

The Available Data

- Here we deal with *quoted prices* (in US$) - individual contract prices may vary.
- Quoted prices serve as a guideline for contracts (usually an upper bound)
- NORSCAN (North America + Scandinavian Countries) inventories are available on a monthly basis and may serve as an explanatory variable for pulp prices.

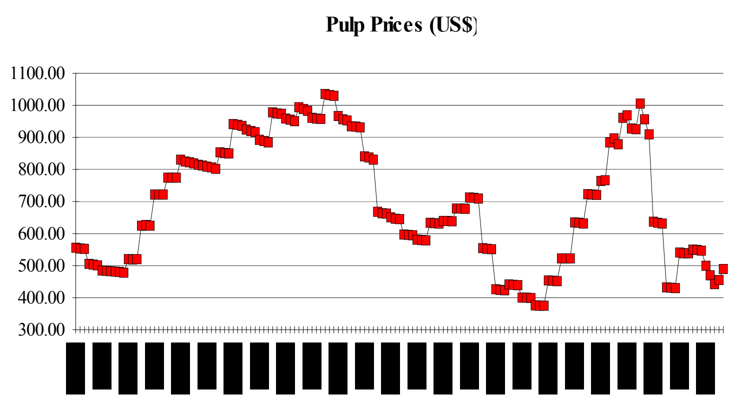
The Available Data

- Quarterly Prices: available since the 1st quarter of 1978.
- The price series used refers to constant prices (base = December 1996) whose values in US dollars were inflated using the Consumer Price Index (CPI) in the USA.
- Monthly prices are obtained by interpolation (via CPI) quarterly prices.

The Available Data

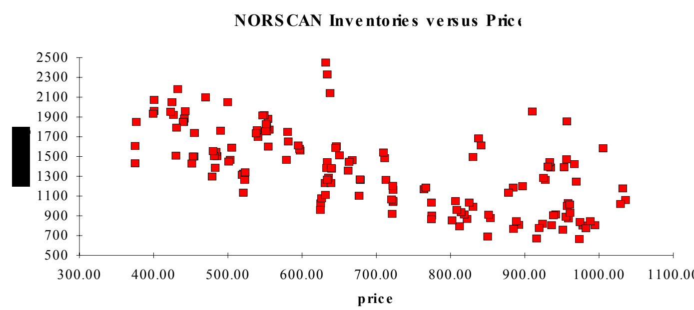
- NORSCAN inventories available monthly since January 1985.
- To match NORSCAN inventories, we will use price data from January 1985 on.
- The idea of using an artificial monthly price series is due to the fact that we intend to use monthly inventories as a leading indicator of price changes.

Pulp Prices



NORSCAN Inventories versus Pulp Prices

- Low prices usually correspond to high inventories, the converse is also true !!!



Dynamic Regression Models

- 1- Basic Model Structure
  - $price_t = b_0 + b_1.(trend_t) + b_2.(inv_{t-1}) + b_3.(price_{t-1}) + b_4.(price_{t-3}) + e_t$
- where :
  - $price_t$ = pulp price at time t
  - $trend_t$ = linear trend
  - inv_{t-1} = NORSCAN inventory at time t-1
  - e_t = error term

Basic Dynamic Regression Model

- All variables are significant, no further lags of Y or inv are necessary, residuals do not show serial correlation BUT:
- The model is unable to identify radical price changes!
- The usefulness of the model is restricted to "normal" periods, that is, whenever substantial price increases or drops do not occur.

Basic Dynamic Regression Model

- For example, the 6 months ahead forecasts obtained by fitting the model until June 1995 are:

Period	Lower 2.5 %	Forecast	Upper 97.5 %	Real Value
1995.07	897	**965**	1033	**969**
1995.08	856.	**939**	1022	**927**
1995.09	845	**933**	1022	**924**
1995.10	833	**928**	1023	**1005**
1995.11	787	**887**	987	**956**
1995.12	725	**828**	931	**910**

Basic Dynamic Regression Model

- All forecasts seem reasonable and real values fall within the confidence limits.
- A totally different situation occurs when we consider 6 months ahead forecasts produced by the model fitted until December 1995, as shown next:

Basic Dynamic Regression Model

- The 6 months ahead forecasts obtained by fitting the model until December 1995 are:

Period	Lower 2.5 %	Forecast	Upper 97.5 %	Real Value
1996-01	786	**854**	922	**638**
1996-02	704	**787**	870	**634**
1996-03	623	**712**	802	**632**
1996-04	543	**639**	734	**433**
1996-05	504	**605**	706	**431**
1996-06	506	**611**	716	**430**

Basic Dynamic Regression Model

- Now the forecasts generated by the model are clearly inappropriate, since the model fails to capture the substantial price fall occurring on the second quarter of 1996.

What to do then ?????

- Inventories seem to be a reasonable leading indicator for prices.
- Also, NORSCAN inventories are available monthly, and this might give us an idea, within one quarter, of the price movements on the next quarter.

What to do then ?????

- Moreover, the inventory series is more “predictable” than the price series, so more accurate and reasonable forecasts can be generated for NORSCAN inventories.
- The important question is: when are the inventories “atypical”, and can this information be useful in the prediction of pulp prices?

What to do then ?????

- To access whether inventory levels are "typical" or not we compare observed NORSCAN inventories with forecasts generated by Moving Averages (MA) using 2, 3, 6 and 12 months.
- The discrepancies between each MA prediction and the actual inventory is measured, and we define dummy variables for the corresponding periods.

Dummy Variable Creation

- The Moving Averages computed with 6 and 12 months of data are much "smoother" estimators of inventories than 2 and 3 months moving averages.
- We would expect forecasting errors to be large when using the moving averages with 6 and 12 points as predictors.

Dummy Variable Creation

- The dummy variables constructed from the forecasting errors of the moving average models indicate whether inventories are within "reasonable" bounds.

Dummy Variable Creation

- Why use Moving Averages to construct the dummies?
- The basic answer is: simplicity! The procedure can be easily implemented in a worksheet without any knowledge of Time Series methods.
- Any other automatic method would also work, but the client would need access to a statistical software.

Dummy Variable Creation

- Define the percent discrepancy between the observed NORSCAN inventory at time t and its moving average prediction as:

$$DIF_t = 100.\frac{(MA_t - inv_t)}{MA_t}$$

Dummy Variable Creation

- where MA_t denotes the moving average prediction for the inventory at time t and inv_t is the real value of NORSCAN inventory at t.
- If $DIF_t > 0$, inventories were overestimated by the moving averages, otherwise they were underestimated.

Dummy Variable Creation

- We next compute the DIF variables based on 2, 3 , 6 and 12 months Moving Averages in the period January/1986 to February/1997 and look at their descriptive statistics. We observe that:

Dummy Variable Creation

- Percentage errors (in absolute value) tend are larger when we estimate inventories using the “smoother” estimators (6 and 12 months moving averages).
- For example, the largest (in absolute value) forecast error is -64.1%, indicating that the inventory was severely underestimated by a moving average with 12 points.

Dummy Variable Creation

- Negative errors tend to be larger (in absolute value) than positive errors.
- Thus, moving averages estimators tend to grossly underestimate inventories.
- The moving average estimators are very bad when inventory level rises.

Dummy Variable Creation

- The 25% percentile for all forecast error series (except that generated from the 12 point moving average) is around -9.
- The 75% percentile for all error series is larger than 7.
- We propose the use of the percentiles of the forecast error distributions as bounds used to indicate whether the inventory level at a given month is "atypical".

Dummy Variable Creation

- Let DIFMM2, DIFMM3, DIFMM6, DIFMM12 denote, respectively, the forecast errors generated by predictions based on 2, 3, 6 and 12 months moving averages.
- The next table shows the percentiles for the forecast errors based on 2, 3, 6 and 12 months moving averages.

Dummy Variable Creation

Percentile	DIFMM2	DIFMM3	DIFMM6	DIFMM12
10 %	-14.3	-14.8	-21.4	-34.4
20 %	-11.4	-10.8	-11.5	-16.3
30 %	-6.3	-6.9	-7.4	-9.0
40 %	-3.7	-3.4	-3.0	-3.9
50 % (median)	-0.95	+0.25	+0.40	+1.95
60 %	2.1	3.3	3.6	5.9
70 %	4.9	5.9	7.4	9.8
80 %	8.7	8.3	11.4	17.1
90 %	13.8	12.2	15.5	22.0

Dummy Variable Creation

- Based on these values we define as "abnormal" situations those when the forecast error is outside the [p30% , p70%] interval.
- We round these percentiles when creating the dummy variables, and we obtain the following decision rule:

Dummy Variable Creation

- An "abnormal" inventory level is detected when :
  - DIFMM2 [-6, 5] or
  - DIFMM3 [-7, 6] or
  - DIFMM6 [-7 , 7] or
  - DIFMM12 [-9, 10]

Dummy Variable Creation

- We next create dummy variables based on this decision rule. The dummies are constructed as:

$$I = \begin{cases} -1 & \text{if DIFMM below the lower limit of the interval} \\ 0 & \text{if DIFMM is inside the interval} \\ 1 & \text{if DIFMM is above the interval's upper limit} \end{cases}$$

Two Fundamental Questions Now Arise

- 1- Do the dummy variables have any explanatory power over pulp prices?
- This can be verified through one way ANOVA models using the dummies as factors.
- 2- Do we improve forecasts of pulp prices when dummies are included in the dynamic regression model?

Do Dummies have explanatory power over Pulp Prices ?

- We fit one way ANOVA models for Pulp Prices using the dummies as factors.
- All ANOVA models are significant at 90% level.
- Thus, different levels of dummies affect pulp prices.
- We next insert dummies into the basic regression model to access whether any gains in forecasting are accomplished.

Regression Model including dummies

- We add dummy variables constructed from 2 and 12 month MA inventory estimators to the basic model.
- The new model structure is:

$price_t = b_0 + b_1.(trend_t) + b_2.(inv_{t-1}) + b_3.(price_{t-1}) + b_4.(price_{t-3}) + b_5.(dummy2_t) + b_6.(dummy12_t) + e_t$

Regression Model including dummies

- Where dummy2 and dummy12 are constructed from the prediction errors using 2 and 12 month moving averages to predict inventories.

- These 2 variables are not “strictly” significant, their significance level is about 85 %.

Regression Model including dummies

- Dummies based on 3 and 6 months MA were not included in the model, because they’re highly insignificant.

Forecast Comparison

- The next table shows the forecasts produced from 2 models, that exclude and include dummy variables

		ONE STEP AHEAD FORECASTS			
	actual price	forecast (with dummies)	% error	forecast (without dummies)	% error %
Jan/96	**638**	**850**	**-33.23**	**854**	**-33.86**
Feb/96	634	596	5.99	611	3.63
Mar/96	632	570	9.81	582	7.91
Apr/96	**433**	**538**	**-24.25**	**527**	**-21.71**
May/96	431	431	0.00	426	1.16
Jun/96	430	480	-11.63	478	-11.16
Jul/96	**541**	**481**	**11.09**	**475**	**12.20**
Aug/96	539	518	3.90	524	2.78
Sep/96	538	517	3.90	515	4.28
Oct/96	**551**	**541**	**1.81**	**539**	**2.18**
Nov/96	549	538	2.00	545	0.73
Dec/96	547	521	4.75	523	4.39
Jan/97	**500**	**516**	**-3.20**	**525**	**-5.00**
Feb/97	470	473	-0.64	475	-1.06
Mar/97	443	446	-0.79	449	-1.47
Apr/97	**455**	**451**	**0.88**	**442**	**2.86**
May/97	490	480	2.04	476	2.86
		mean % error :	-1.62		-1.72
		min. % error :	-33.23		-33.86
		max. % error :	11.09		12.20

Forecast Comparison

	ABSOLUTE VALUES OF % ERRORS			
		model with dummies		model without dummies
Jan/96		33.23		33.86
Feb/96		5.99		3.63
Mar/96		9.81		7.91
Apr/96		24.25		21.71
May/96		0.00		1.16
Jun/96		11.63		11.16
Jul/96		11.09		12.20
Aug/96		3.90		2.78
Sep/96		3.90		4.28
Oct/96		1.81		2.18
Nov/96		2.00		0.73
Dec/96		4.75		4.39
Jan/97		3.20		5.00
Feb/97		0.64		1.06
Mar/97		0.79		1.47
Apr/97		0.88		2.86
May/97		2.04		2.86
	mean :	7.05		7.01
	max :	33.23		33.86
	min :	0.00		0.73

Forecast Comparison

	ABSOLUTE VALUES OF % ERRORS "QUARTERLY" FORECASTS			
		model with dummies		model without dummies
Jan/96		33.23		33.86
Apr/96		24.25		21.71
Jul/96		11.09		12.20
Oct/96		1.81		2.18
Jan/97		3.20		5.00
Apr/97		0.88		2.86
	mean :	12.41		12.97
	max :	33.23		33.86
	min :	0.88		2.18

Forecast Comparison

- Some gain was obtained by including the dummies, but it is marginal!!!
- The comparison of forecasts was done in a “worst case” scenario.
- The price drop from 1996/Q1 to 1996/Q2 was roughly 32%.

Forecast Comparison

- One should attempt to include other explanatory variables, but there aren't many available candidates!
- Shipment/Inventory ratio is one of the candidates, but shipments series is short (since 1994).

Forecasts until December 1997

- Basic scenario uses automatic forecasts for NORSCAN inventories.
- These projections correspond to dummies with value 0 until the end of 1997.

Forecasts until December 1997

	model with dummies	model without dummies	model with dummies +ARCH error
Jun/97	495	492	497
Jul/97	514	507	518
Aug/97	510	504	513
Sep/97	498	493	501
Oct/97	511	504	515
Nov/97	512	504	515
Dec/97	499	493	502

Alternative Approach: Simulation

- We now adopt a totally different approach.
- We will work with returns constructed from quarterly prices, and attempt to model and forecast these returns.

Alternative Approach: Simulation

- Let P_t and P_{t-1} denote pulp prices on quarters t and t-1.
- The geometric returns are defined as:

$$y_t = \log\left(\frac{P_t}{P_{t-1}}\right)$$

- Where log denotes the natural logarithm.

We propose the following model for the returns:

$$y_t = \log\left(\frac{P_t}{P_{t-1}}\right) = \mu_t + \sigma_t Z_t \quad \text{where } Z_t \text{ are iid N(0,1)}$$

- where both μ_t and σ_t^2 are time varying according to the equations:

$$\mu_t = (1-\lambda_1).y_t + \lambda_1.\mu_{t-1}$$

and

$$\sigma_t^2 = (1-\lambda_2).y_t^2 + \lambda_2.\sigma_{t-1}^2$$

Rationale behind this structure

- Both μ_t and σ_t^2 are time varying, and their updating equations are easy to implement.
- The equation for σ_t^2 resembles that of an integrated GARCH model.
- If the lambdas are large (> 0.7 or 0.8), information decay is slow;
- Otherwise, the latest observations will have a large weight, and the memory is short.

Drawbacks

- Choice of smoothing constants λ_1 and λ_2
- If we perform a grid search we find out that the mimimum mean squared error constants are obtained when λ_1 and λ_2 tend to zero.
- This is due to the fact that we are attempting to optimize both constants for the whole sample !

Drawbacks

- We need to make an arbitrary choice.
- Intuitively, the smoothing constant for σ_t^2 should be larger than that of μ_t .

Implementation

- One step ahead prediction (corresponding to 1997/Q3 can be easily obtained by simulating a large Normal sample and applying the last estimated values for μ_t and σ_t^2 .
- Besides the point estimates of pulp prices, we can use additional information from the generated probability distribution, that can lead, for example, to Value at Risk (VaR) estimates.

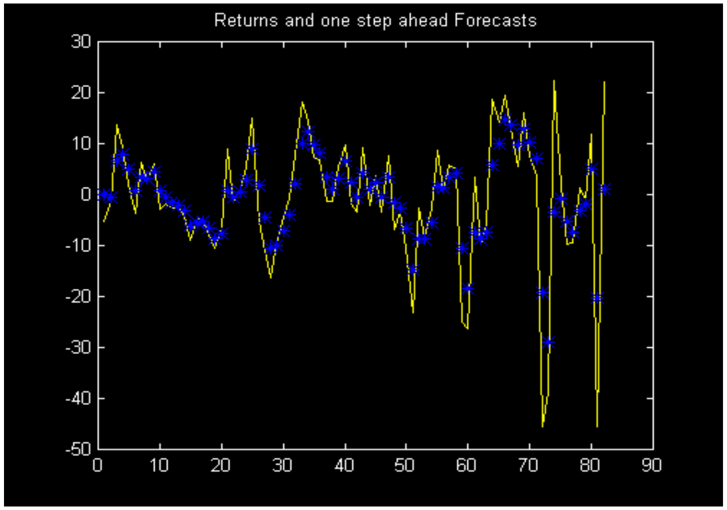
Implementation

- k step ahead forecast could be obtained in an analogous fashion, by incorporating at each step the last prediction, re-estimating μ_t and σ_t^2 by the same procedure and generating another sample of iid Normal variables.

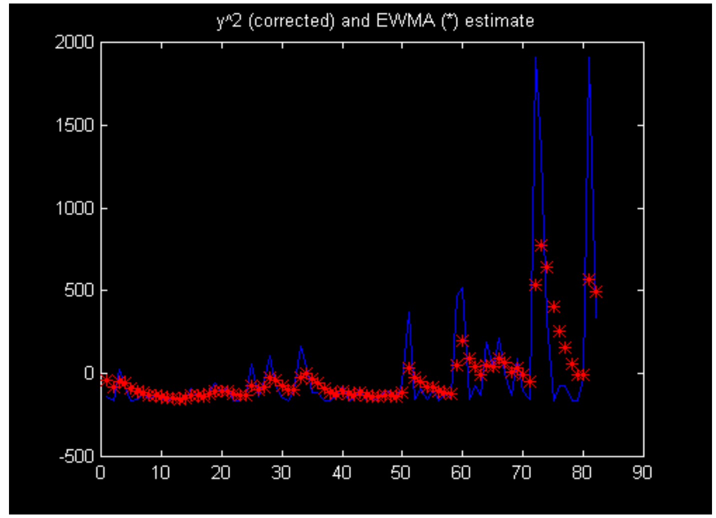
Example

- We next show the result of one implementation where:
- the smoothing constants are λ_1 = 0.5 and λ_2 = 0.7,
- the last observed price is US$ 490 (2nd quarter 1997),
- 5000 iid Normal variables are generated.

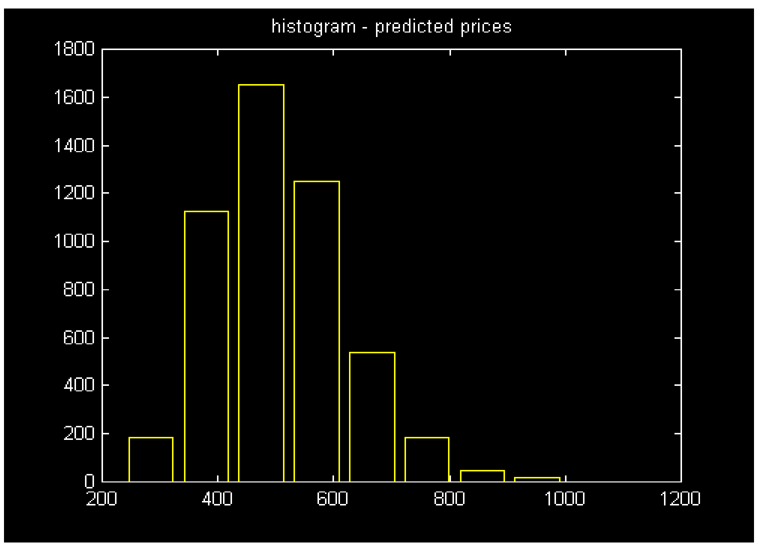
Example



Example



Example



Example

- The last estimates of μ_t and σ_t are, respectively, 1.01% and 22.27%
- The simulated price distribution is:

Distribution of Predicted Prices			
3rd Quarter 1997			
mean :	507		
minimum :	237		
maximum :	1190		
std.dev. :	115		
percentiles	value		
5%	342		
10%	373		
25%	424		
50%	494		
75%	577		
90%	656		
95%	714		

Comparison of Estimates

- For comparison, the estimated values for Prices in July 1997 are:
  - model with dummies : 514
  - model without dummies : 507
  - model with dummies and ARCH errors : 518
  - median of simulated prices : 494
  - mean of simulated prices : 507

Conclusion

- The alternative approach is completely heuristic, but may serve as an approximation when the true data generating process is hard to identify, or when dynamic regression models do not seem to provide a completely satisfactory answer.