Fraud Detection using Autobox’s™ Automatic Intervention Detection

International Symposium on Forecasting 2002

David P. Reilly AFS, Inc.
Paul Sheldon Foote California State at Fullerton
David Lindsay California State at Stanislaus
Annhenrie Campbell California State at Stanislaus

Introduction

- Increased Attention to Accounting Fraud
- $600 Billion per annum
- Investor Concern
- Current State of the art: Ratio analysis, data mining

Intervention Detection

- Using Box-Jenkins (B-J) time series analysis with intervention detection

- Misperceptions on need for a pre-set minimum number of observations to detect model structure and violations to that structure.

- Signal to Noise Ratio establishes the “identifiability of the model”.

Study Goal

- Perform a blinded study to identify those firms that commited fraud from those that had not
- Forecasting is not the goal…modeling is
- Use 10 years of Balance Sheet data prior to the year in which the fraud was publicly detected
- Analyze 45 Balance Sheet items and identify interventions in the most recent time period
- Use a count of interventions found in the last period for the 45 Balance Sheet items to identify which companies did commit fraud before it became public knowledge.

Methodology

- Identify 8 fraudulent companies in different industries and then identify “matched-pairs”
- Present the companies blinded, but known that one of the companies is fraudulent
- Run Autobox in batch mode for 20 companies for 45 Balance Sheet items
- Count the number of interventions in the last period for each company
- Identify the company with the most interventions in each of the 8 industries as a fraudulent company

Count of Interventions

	Number of 45 B/S Items found as interventions				
	in the year before fraud was publicly identified				
Fraud Firm		Pair Match 1		Pair Match 2	
Cendant	**26**	Advance Tobacco Products	3	Competitive Technologies	12
Con Agra	9	Sara Lee	15	Classica	10
Enron	**22**	Mercury Air Group	8	World Fuel Service	15
Grace	**21**	Great Lakes Chemical	11		
McKesson	**29**	Bergen Brunswig	16		
Rite Aid	**29**	Drug Emporium	10		
Sunbeam*	2	Decorator Industries	7		
Waste Mg	**38**	Rich Coast	4	Wastemasters	22
* - The prior year to this analysis showed unusual activity					

Results

- 6 of the 8 companies were correctly identified as fraudulent
- Cendant, Enron, Grace, McKesson, Rite Aid, Waste Management were identified
- Con Agra and Sunbeam were not identified
- Further research showed Sunbeam was found to be unusual the previous year and history has shown that Sunbeam did their best to “normalize” their Balance Sheet the next year
- WasteMasters was supposed to be a matched-pair to Waste Management and since it was a blinded study the research suggested that there were two and not one companies that committed fraud in that industry

“All Models are Wrong, but some models are useful” G.E.P. Box

With only 10 data points, you can only justify simple models :

e.g. Y(t)= Constant +I(t)

or Y(t)= Phi*Y(t-1) + Constant + I(t)

Where I(t) could be a pulse, level shift, time trend with an arbitrary starting point or some combination thereof. You need to scan the “sample space” in order to detect what is “visually obvious” or “statistically obvious” and then submit this candidate for necessity and sufficiency checking ala step-down and step-forward regression

Number Crunching to find if there is a Significant Intervention

We create an iterative computer based experiment where we establish a base case model(no intervention) and then compare the base case to models with an intervention. We then choose the model with smallest variance. If none of the intervention models has a significantly lower variance then the base model, then we keep the base case model.

Base Case

$$Y_t = BO + U_t$$

We will estimate this model using a standard regression model with only an intercept to get $\widehat{BO}$ and σ^2_U

Modeling Interventions -Pulse

We will first try $Y_t = BO + B3Z_t + U_t$

where $Z_t = 1,0,0,0,0,0,0,0,0,,,,,,,,,,0$

or $Z_t = 1 \quad t = 1$

$Z_t = 0 \quad t > 1$

We run our regression with a pulse at time period = 1.

$\sigma^2{}_U$ is an indicator of how just good our candidate intervention model is.

Modeling Interventions - Pulse

It’s clear we can create a second candidate intervention model which has Z_t = 0,1,0,0,0,0,0,0,0,0,,,,,,,,,,0

We run our regression with a pulse at time period = 2.

We can continue this path for all possible time periods.

Table of Summary Variances

(1)σ^2_U Base Case (No Pulse)

(2)σ^2_U Pulse at time period=1

(3)σ^2_U Pulse at time period=2

•

•

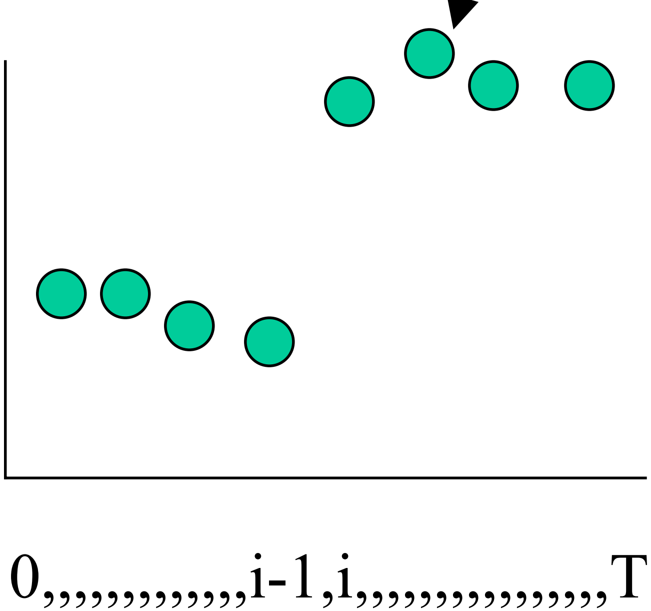
•

(60)σ^2_U Pulse at time period=T

If we had 60 observations then we would have run 61 regressions which yield 61 estimates of the variance.

Modeling Interventions - Level Shift

If there was a level shift and not a pulse then it is clear that a single pulse model would be inadequate thus $Y_t = BO + B3Z_t + U_t$



Assume the appropriate Z_t is

$Z_t = 0,0,0,0,1,1,1,1,1,1,,,,,,,T$

or $Z_t = 0 \quad t < i$

$Z_t = 1 \quad t > i-1$

Modeling Interventions -Level Shift

Similar to how we approached pulse interventions, we will try the various possible level shifts at the same time that we are also evaluating our base case and the pulse models. So our tournament of models is now up to 120; One base case model, 60 models for pulses and 59 models with level shifts.

Modeling Interventions -Level Shift

Our first level shift model would be

Z_t = 0,1,1,1,1,1,1,1,,,,,1

Z_t = 0 i = 1

Z_t = 1 i > 1

We can continue this path for all possible time periods.

Table of Summary Variances

(1)σ^2_U Base Case (No Pulse)

(2)σ^2_U Pulse at time period=1

-

(61)σ^2_U Pulse at time period=T

(62)σ^2_U Level shift starting at time period=2

-

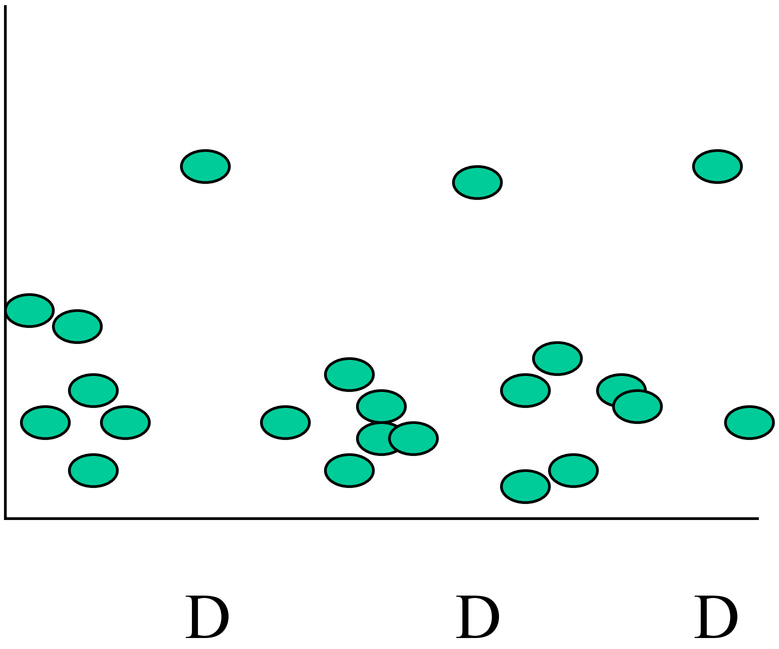
(120)σ^2_U Level shift starting at time period=T

Here are the 120 regressions which yield 120 estimates of the variance.

Modeling Interventions - Seasonal Pulses

There are other kinds of pulses that might need to be considered otherwise our model may be insufficient.

For example, December sales are high.



The data suggest this model

$Y_t = BO + B3Z_t + U_t$

$Z_t = 0$ i <>12,24,36,48,60

$Z_t = 1$ i = 12,24,36,48,60

Modeling Interventions - Seasonal Pulses

In the case of 60 monthly observations, we would have 48 candidate regressions to consider. We will try the various possible seasonal pulses at the same time that we are also evaluating our base case, pulse and level shift models. So our tournament of models is now up to 168; One base case model, 60 models for pulses and 59 models with level shifts, 48 models for seasonal pulses. The first seasonal model:

Z_t = 1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,,,,,,,,T

Modeling Interventions - Seasonal Pulses

Our second seasonal pulse model would be

Z_t = 0,1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,,,,,

Z_t = 0   i <> 2,14,26,38,50

Z_t = 1   i = 2,14,26,38,50

We can continue this path for all possible time periods.

Table of Summary Variances

(1)σ^2_U Base Case (No Pulse)

(2)σ^2_U Pulse at time period=1

(60)σ^2_U Pulse at time period=T

(61)σ^2_U Level shift starting at time period=2

(120)σ^2_U Level shift starting at time period=T

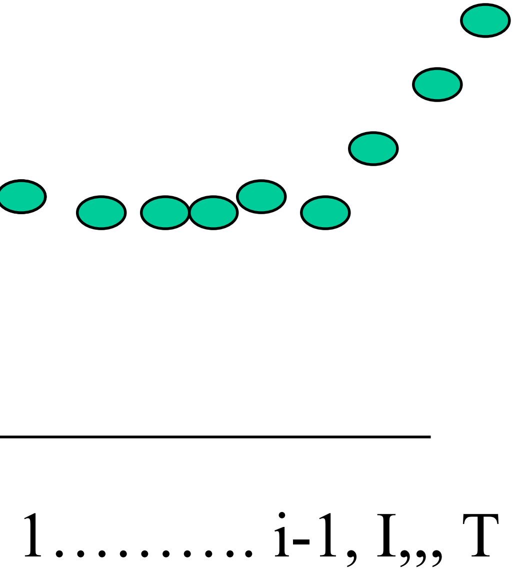
(121)σ^2_U Seasonal pulse starting at time period=1

(168)σ^2_U Seasonal pulse starting at time period=T

Here are the 168 regressions which yield 168 estimates of the variance.

Modeling Interventions - Local Time Trend

The fourth and final form of a determinstic variable is the the local time trend. For example,



The appropriate form of Z_t is

$Z_t = 0 \quad t < i$

$Z_t = 1 \quad (t-(i-1)) * 1 >= i$

$Z_t = 0,0,0,0,0,0,1,2,3,4,5,,,,,$

Modeling Interventions - Local Time Trend

Our first local time trend model is

Z_t = 1,2,3,4,5,6,7,,,,

$Z_t = Z_t + 1$ i >= 1

Our second local time trend model is

Z_t = 0,1,2,3,4,5,6,7,,,,

$Z_t = Z_t + 1$ i >= 2

We can continue this path for all possible time periods.

Table of Summary Variances

(1)σ^2_U Base Case (No Pulse)

(2)σ^2_U Pulse at time period=1

(60)σ^2_U Pulse at time period=T

(61)σ^2_U Level shift starting at time period=2

(120)σ^2_U Level shift starting at time period=T

(121)σ^2_U Seasonal pulse starting at time period=1

(168)σ^2_U Seasonal pulse starting at time period=T

(169)σ^2_U Local time trend starting at time period=1

(228)σ^2_U Local time trend starting at time period=T

Here are the 228 regressions which yield 228 estimates of the variance.

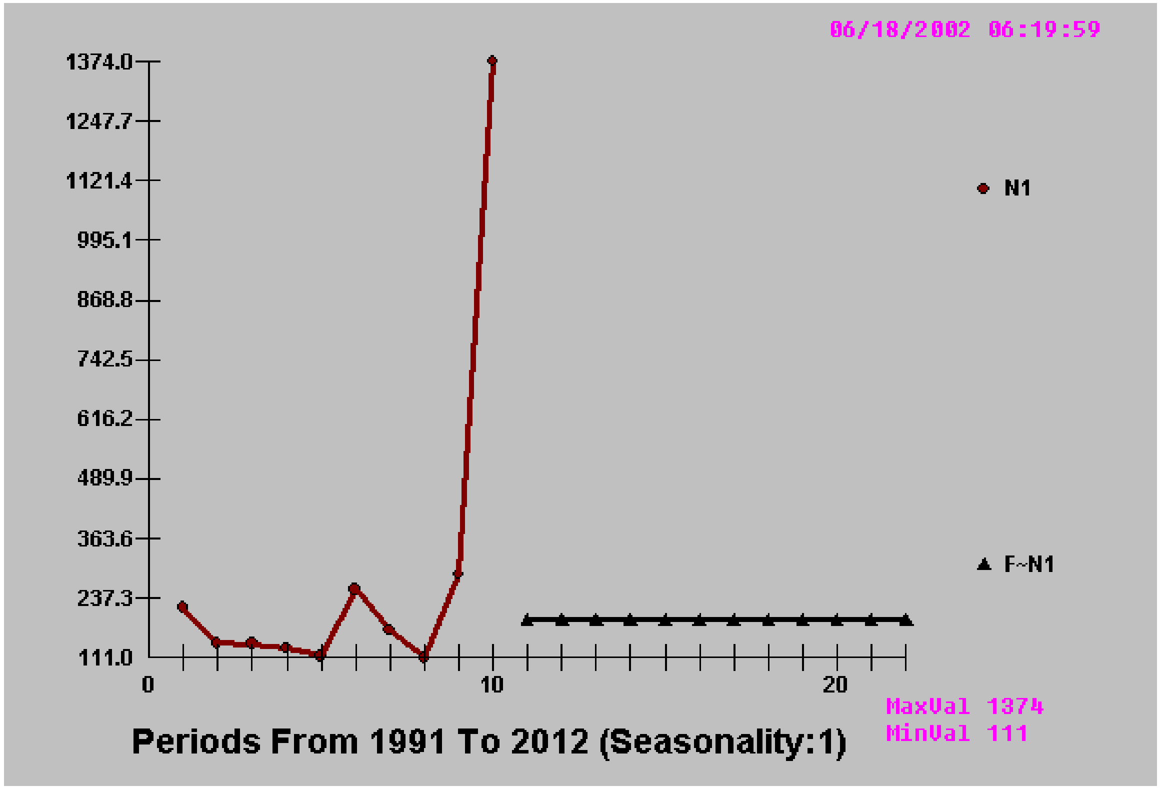
The intervention variable that generated the smallest error variance is the winner of the tournament. We now must test if this winner is statistically significant. In other words, has the winner created a reduction in the variance that is significantly different from zero?

We add the intervention variable into the model which then creates a new base case model. We can rerun the tournament and subsequent statistical testing to determine if a second intervention variable is needed. This process can be continued until no more variables are added to the base case model.

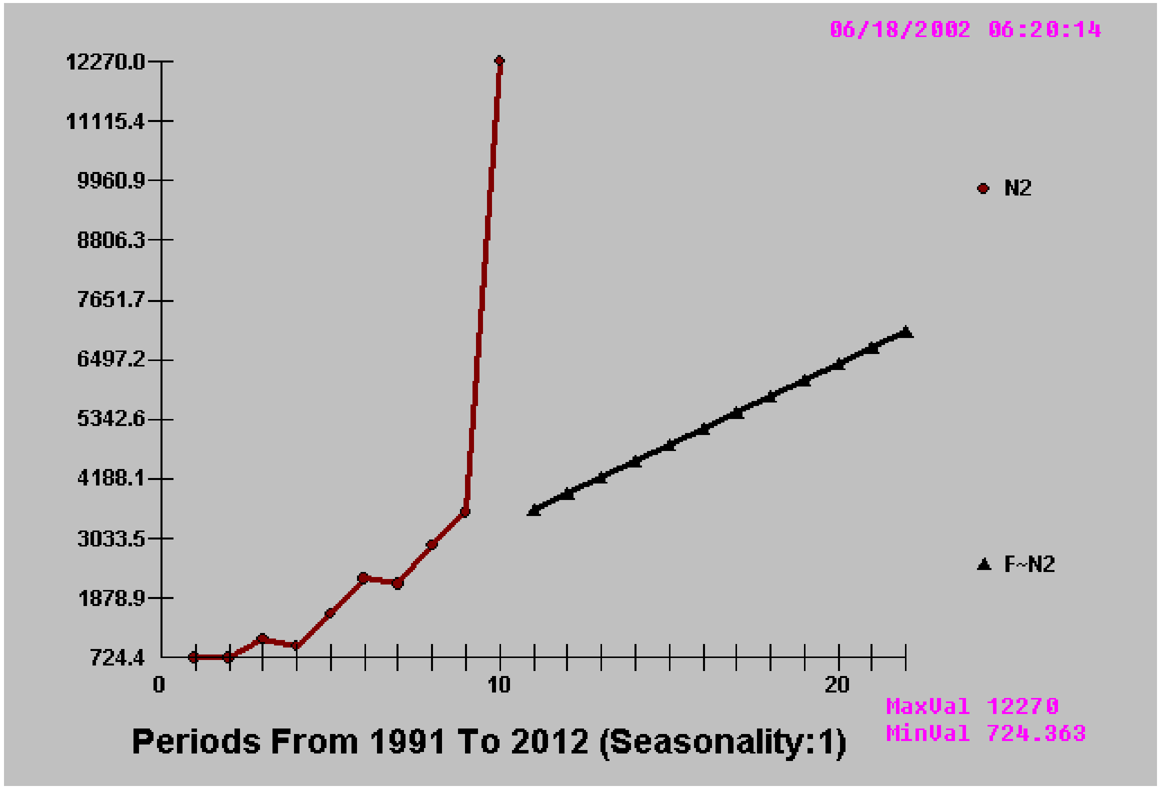
Conclusion

- Tolerance thresholds could be setup to detect fraud by industry(SIC?)

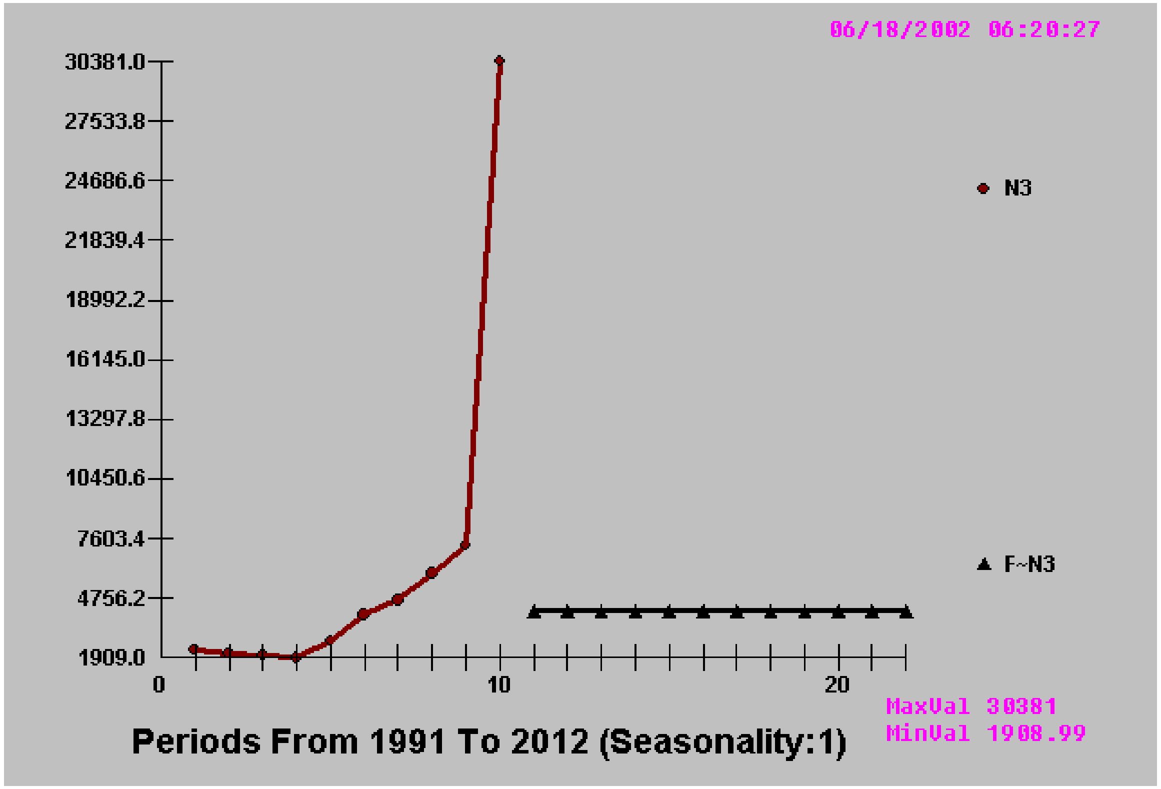
Cash & Short Term Inv



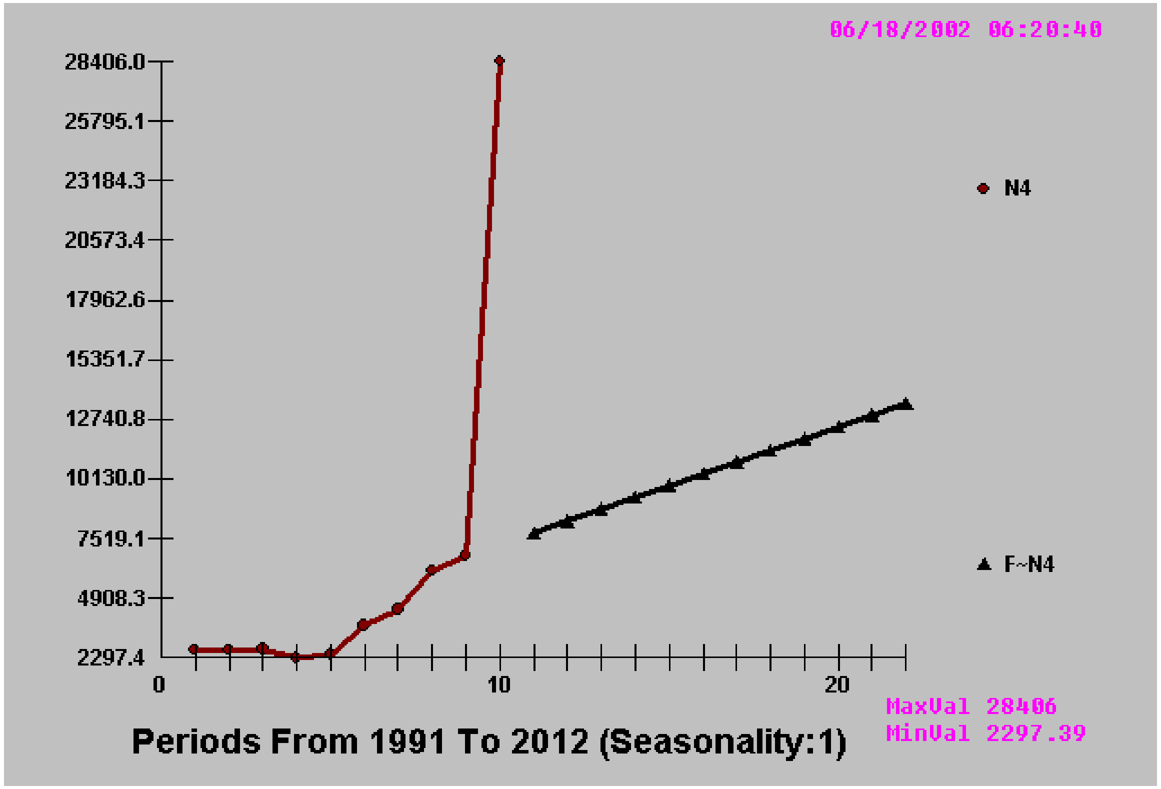
Receivables



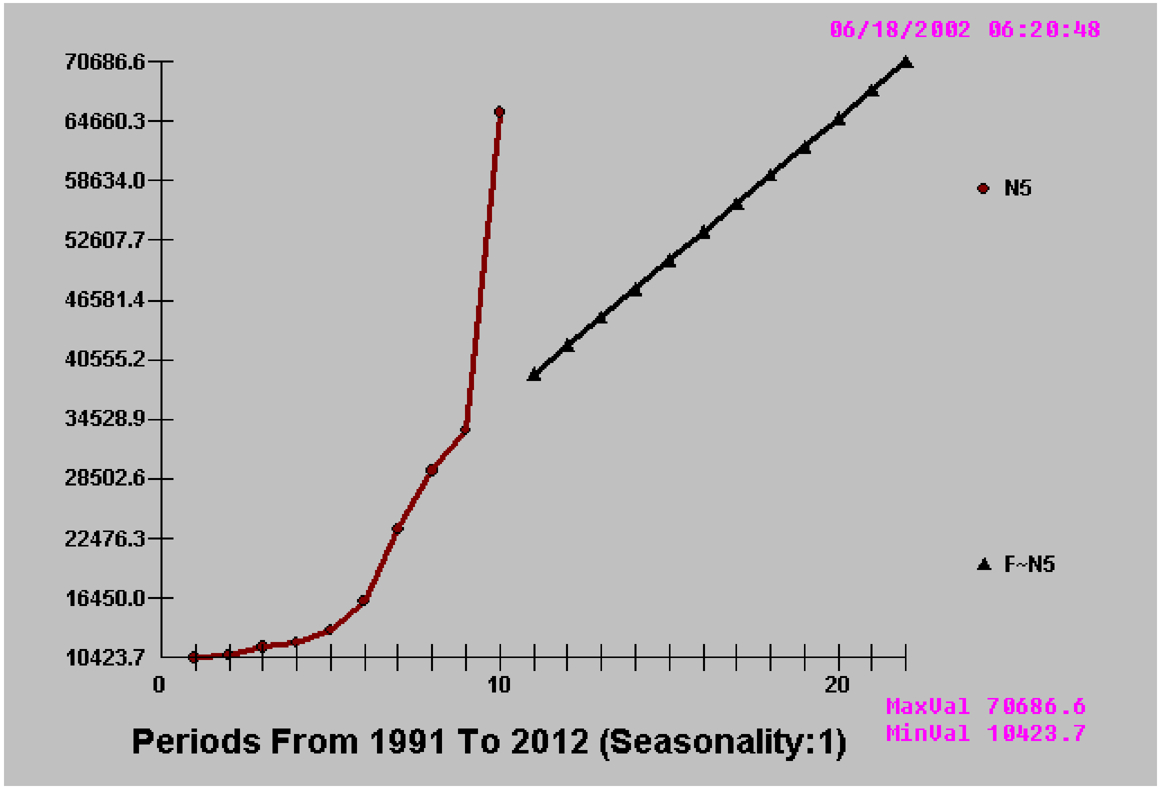
Total Current Assets



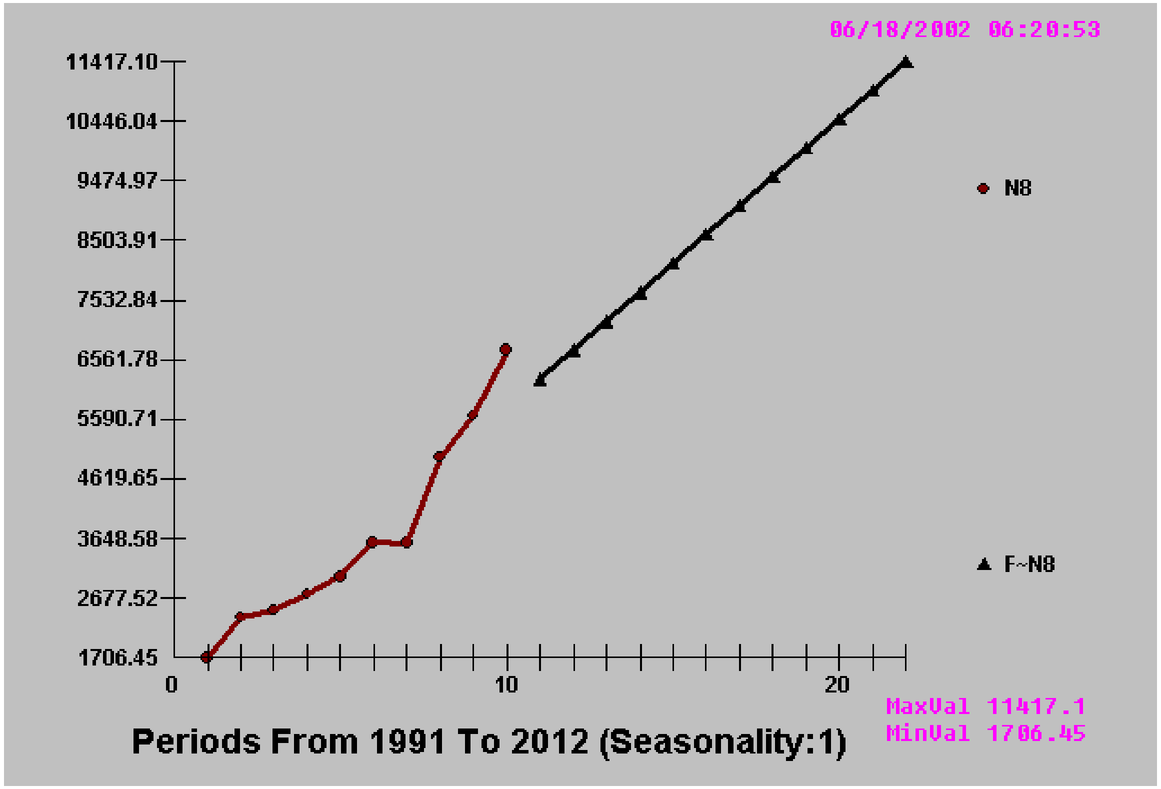
Total Current Liabilities



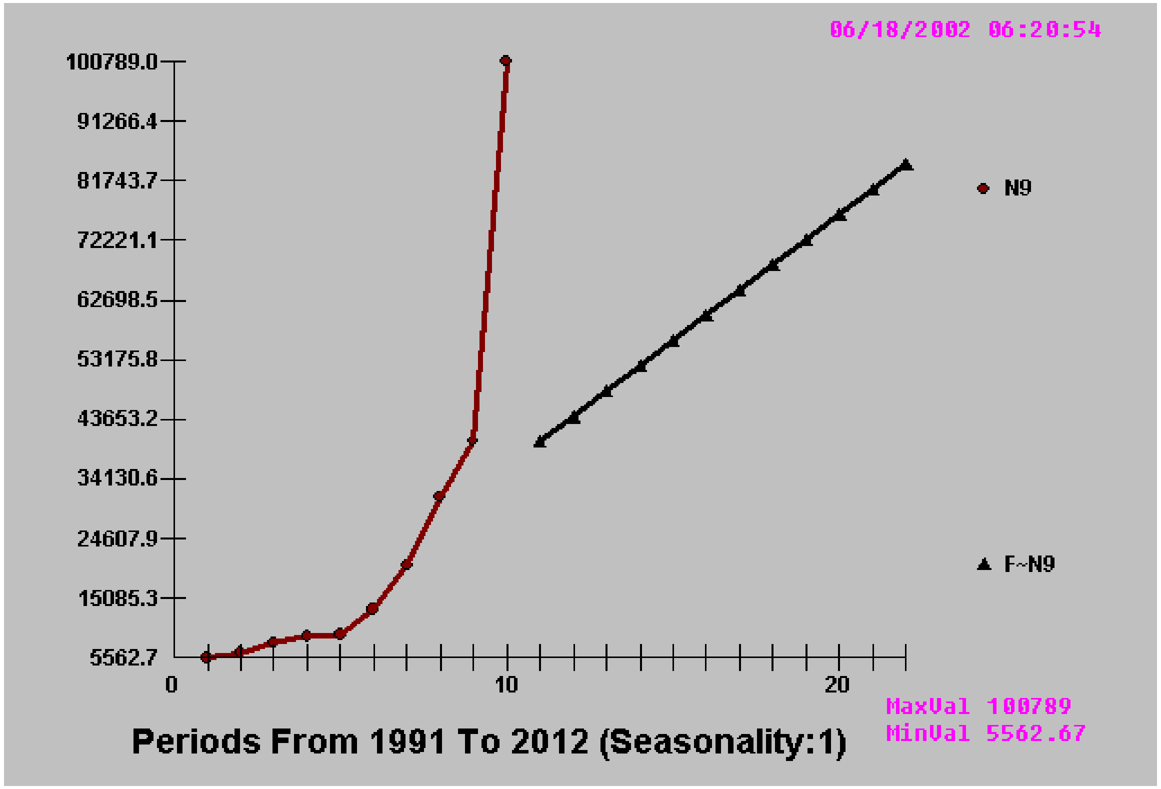
Total Assets



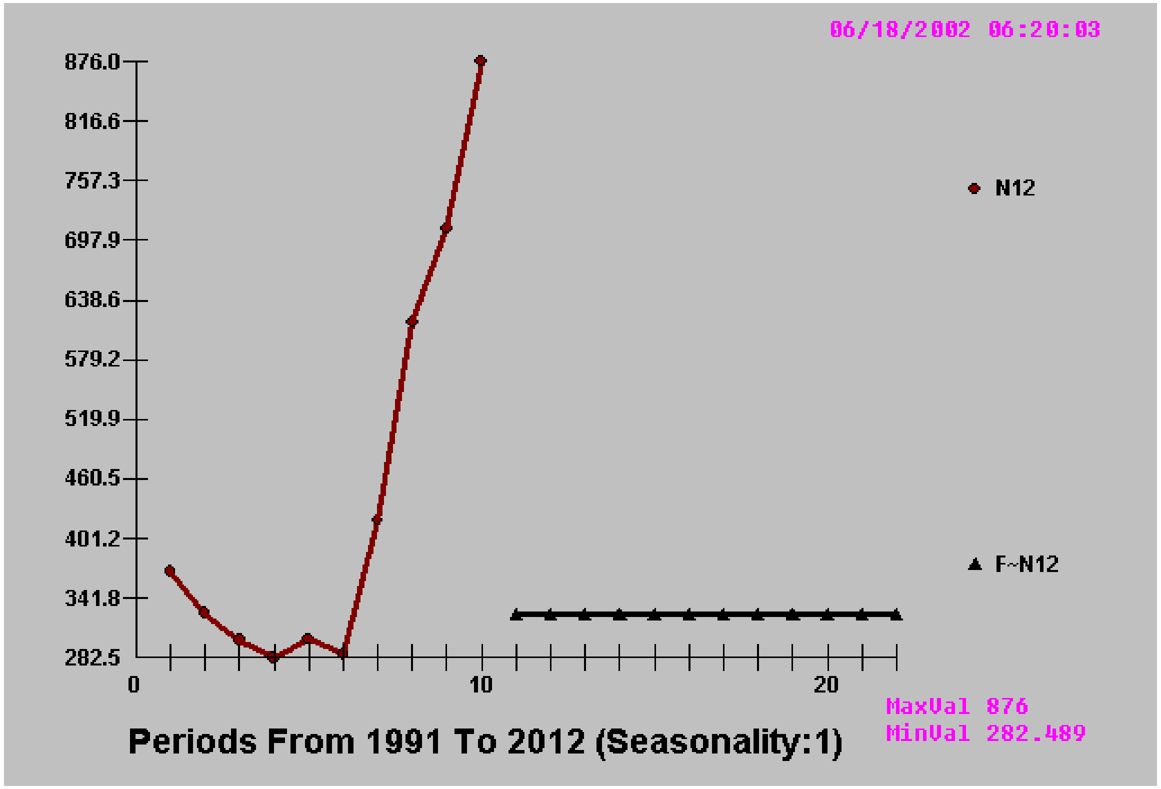
Tangible Common Equity



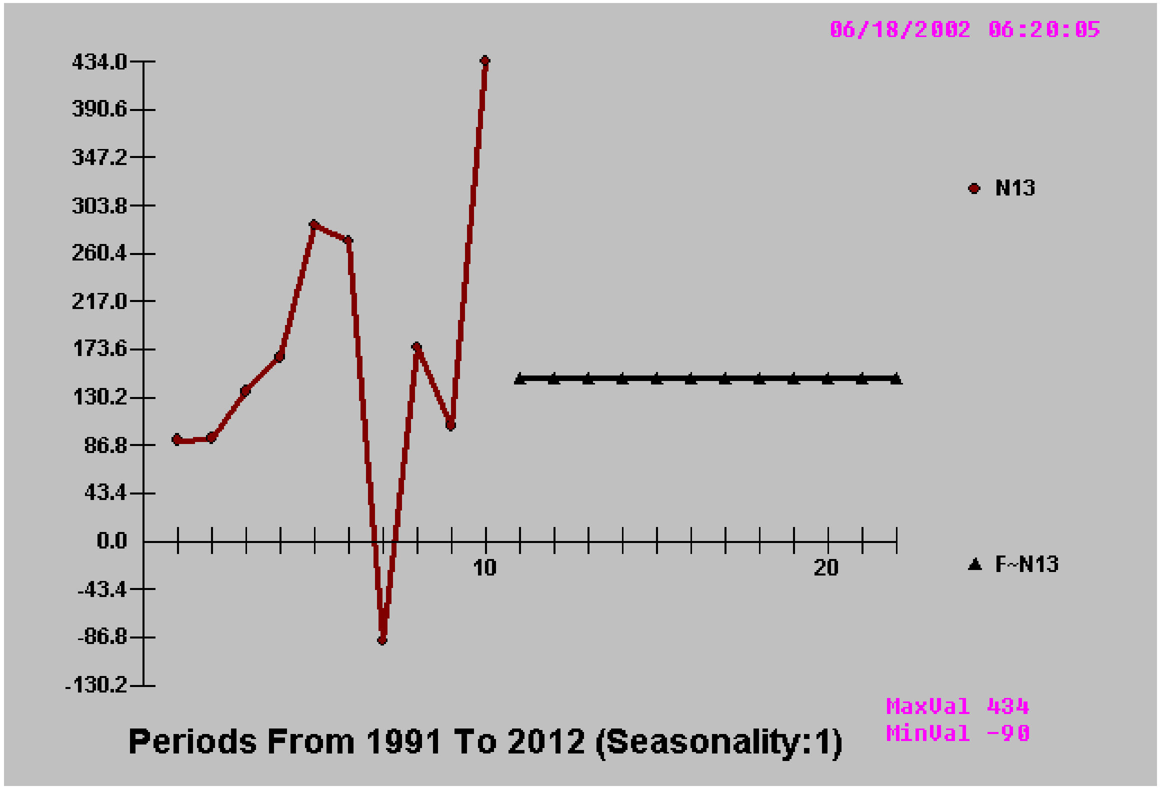
Net Sales


06/18/2002 06:20:54
100789.0
91266.4
81743.7
72221.1
62698.5
53175.8
43653.2
34130.6
24607.9
15085.3
5562.7
0
10
20
● N9
▲ F~N9
MaxVal 100789
MinVal 5562.67
Periods From 1991 To 2012 (Seasonality:1)

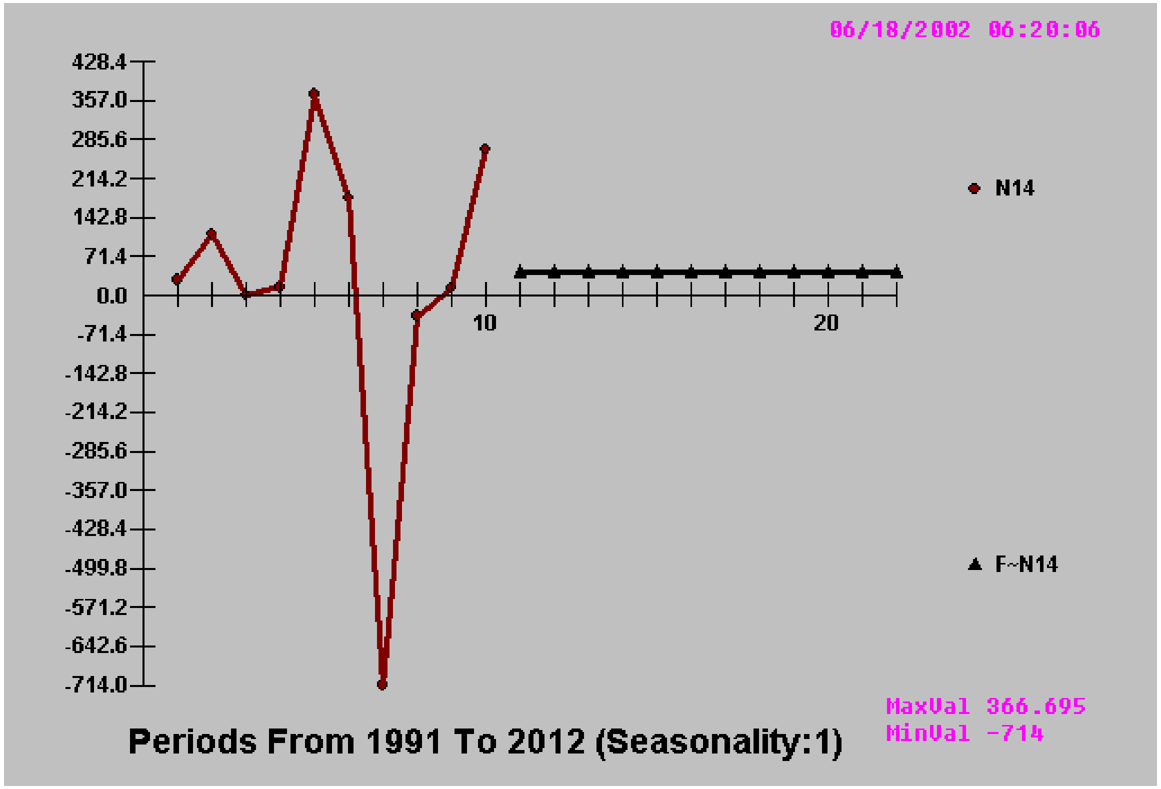
Interest Expense



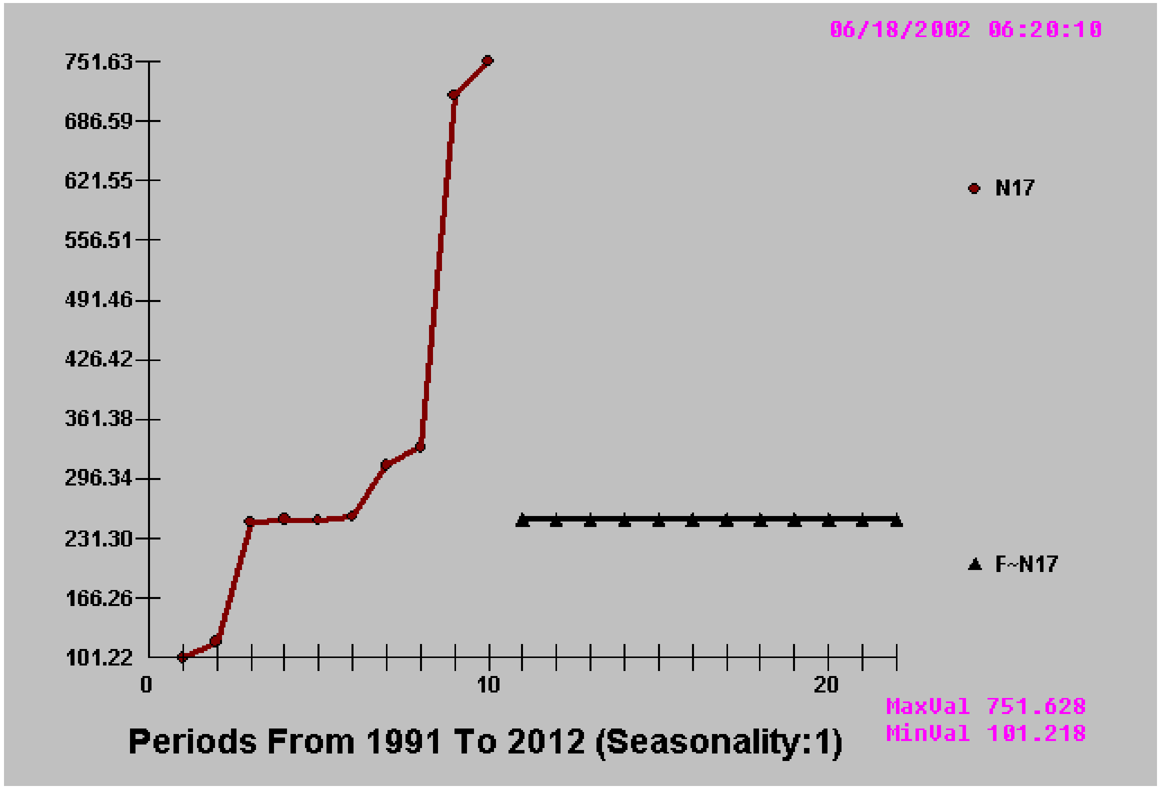
Total Income Taxes



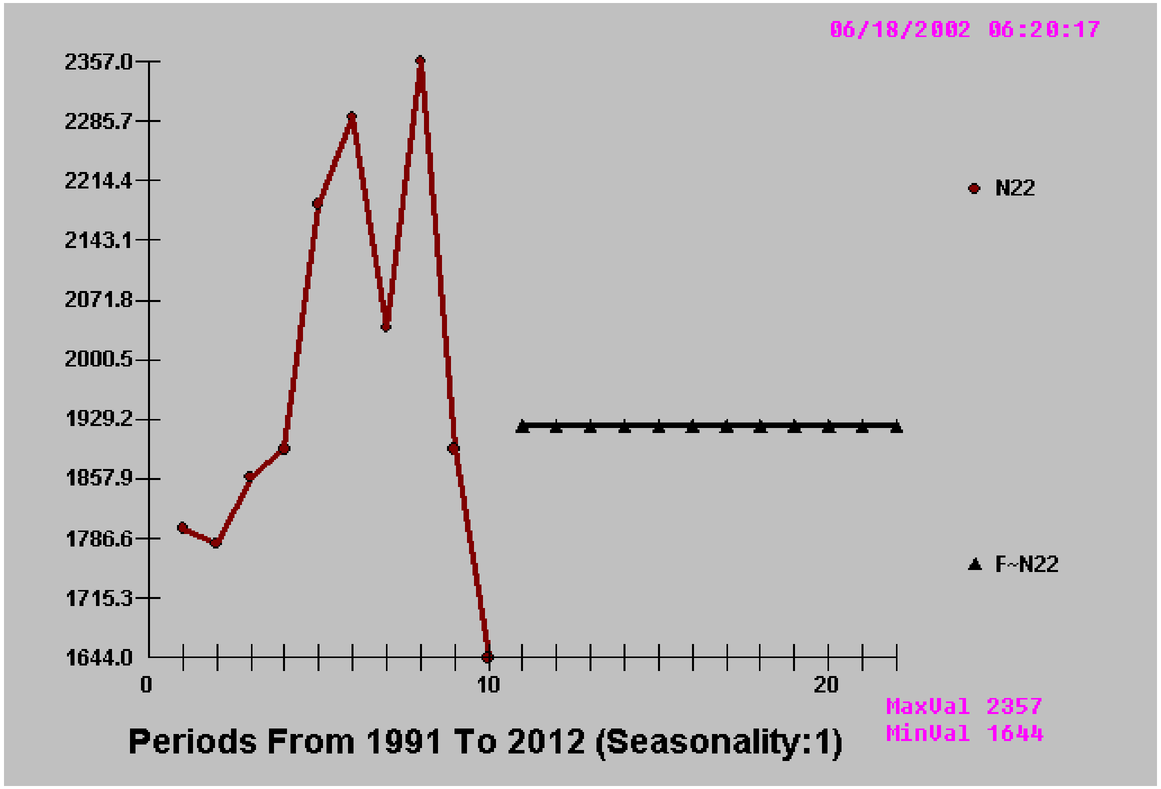
Special Items



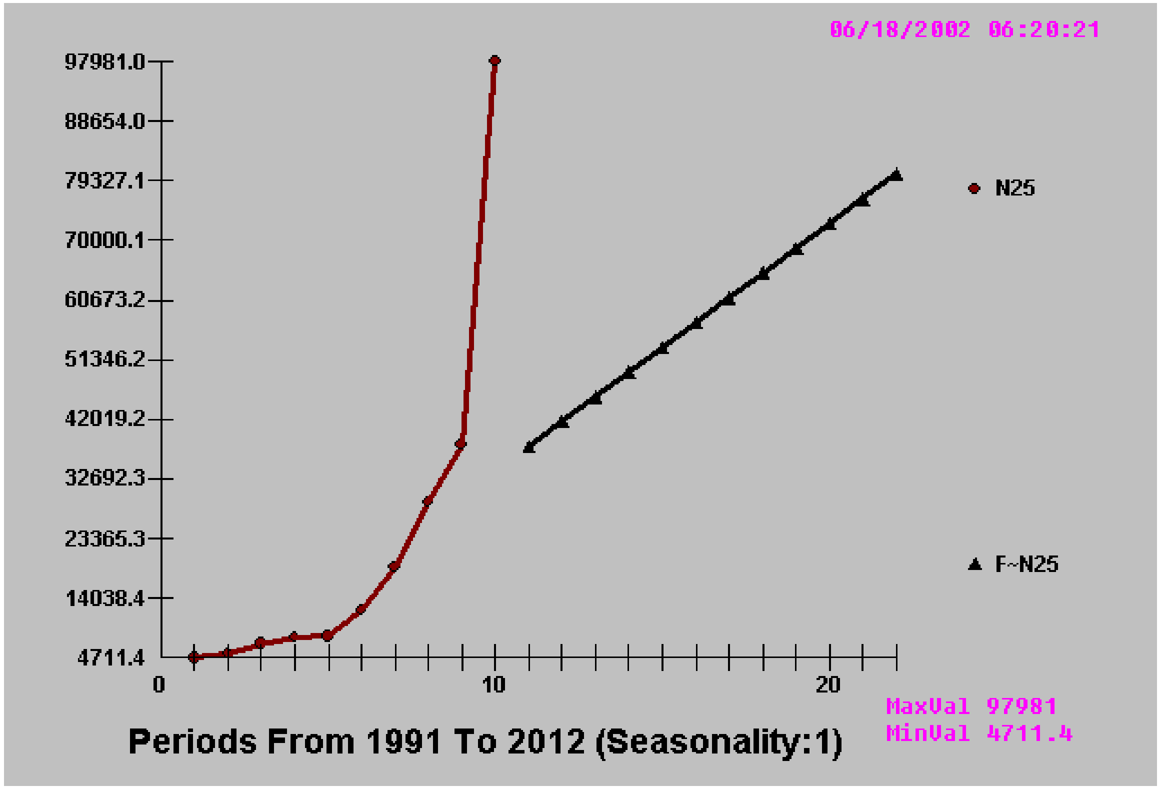
Common Shares Outstanding



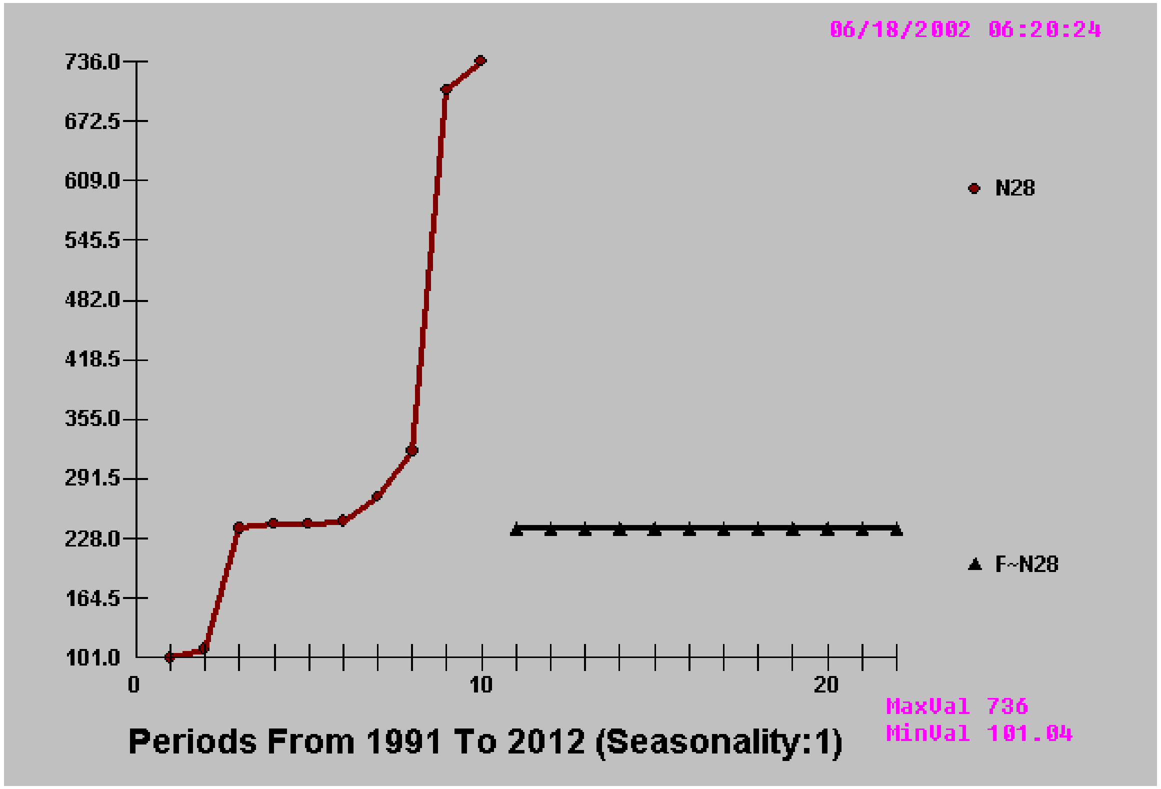
Def Taxes & Inv Credit



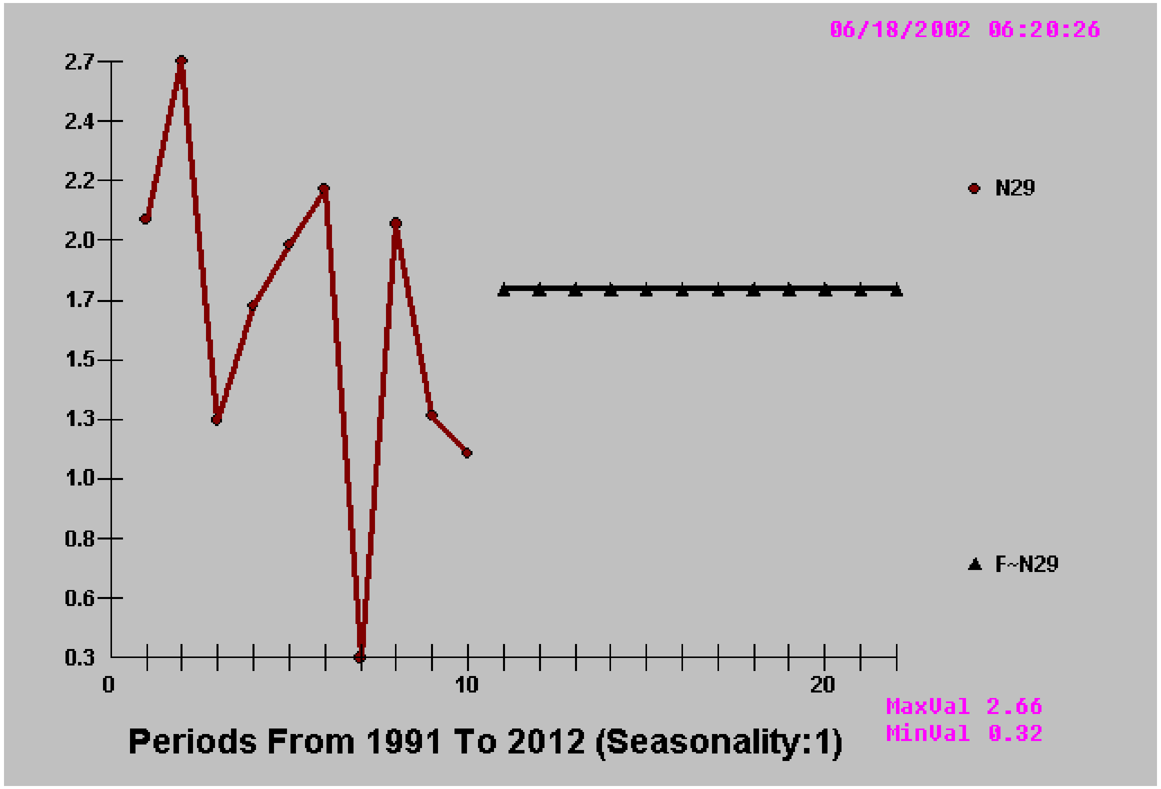
Cost of Goods Sold



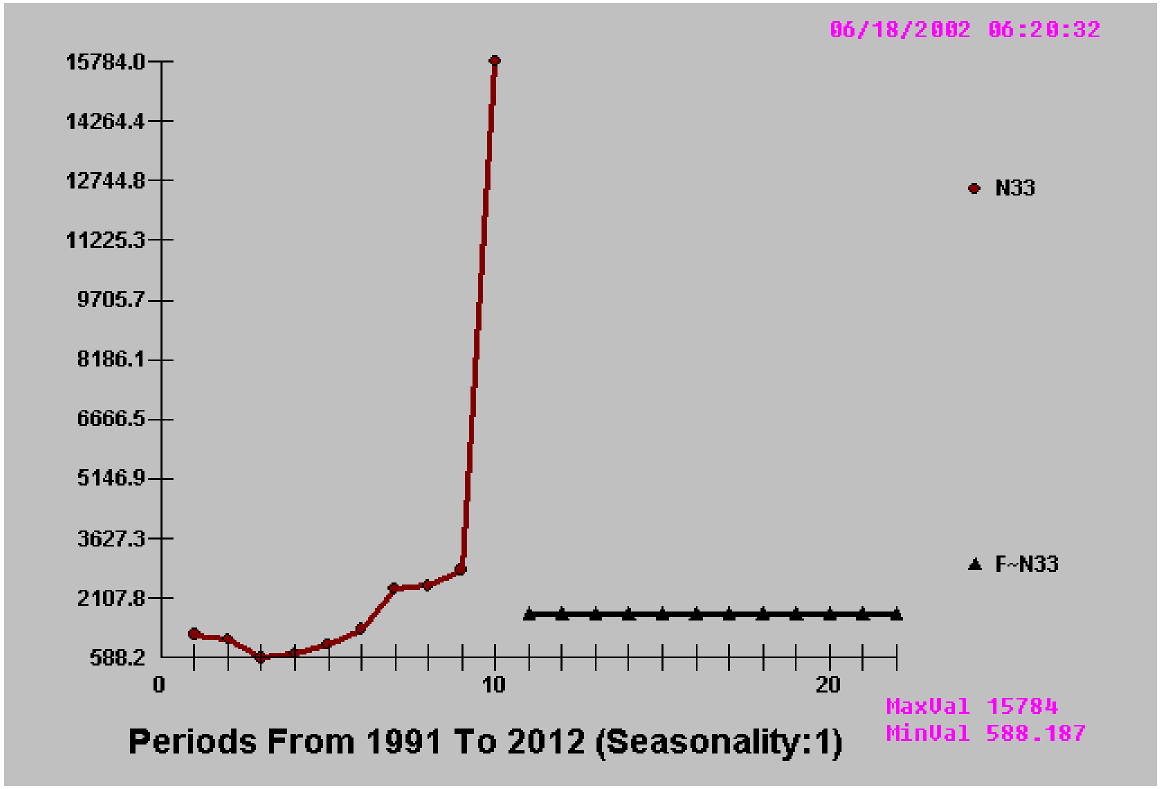
Shares Used To Compute EPS



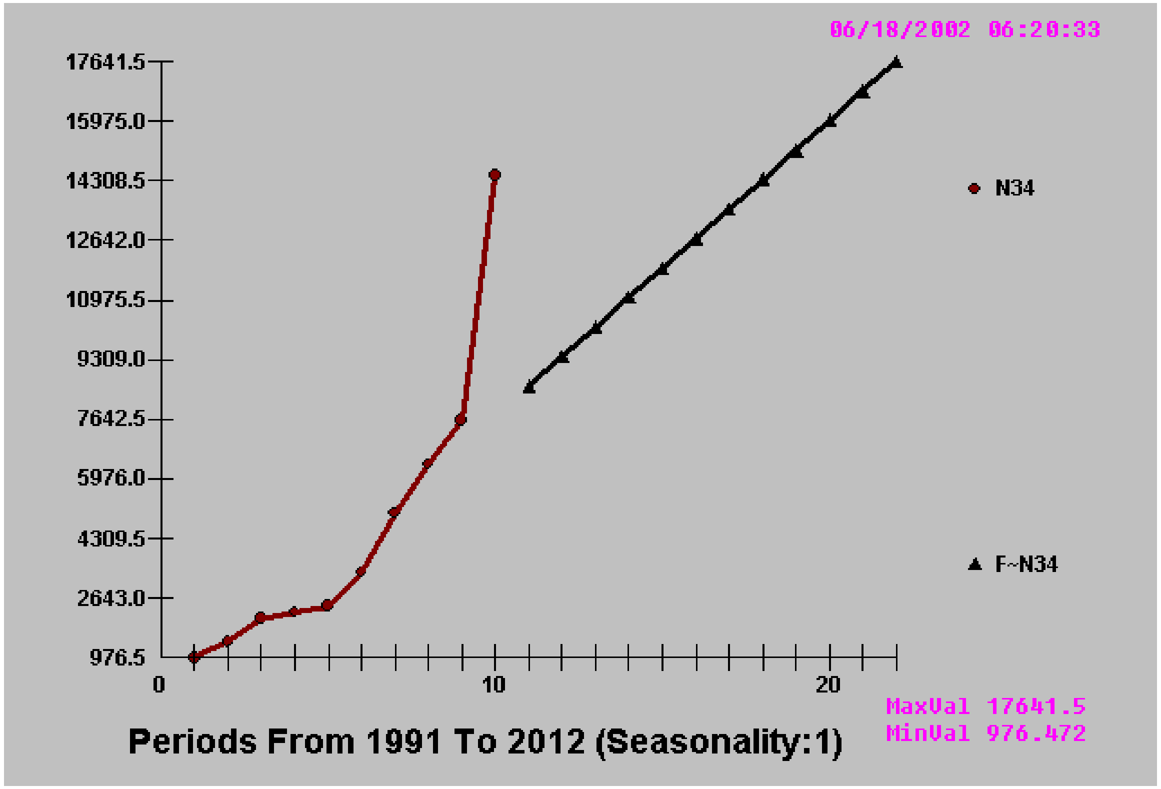
Dilluted EPS



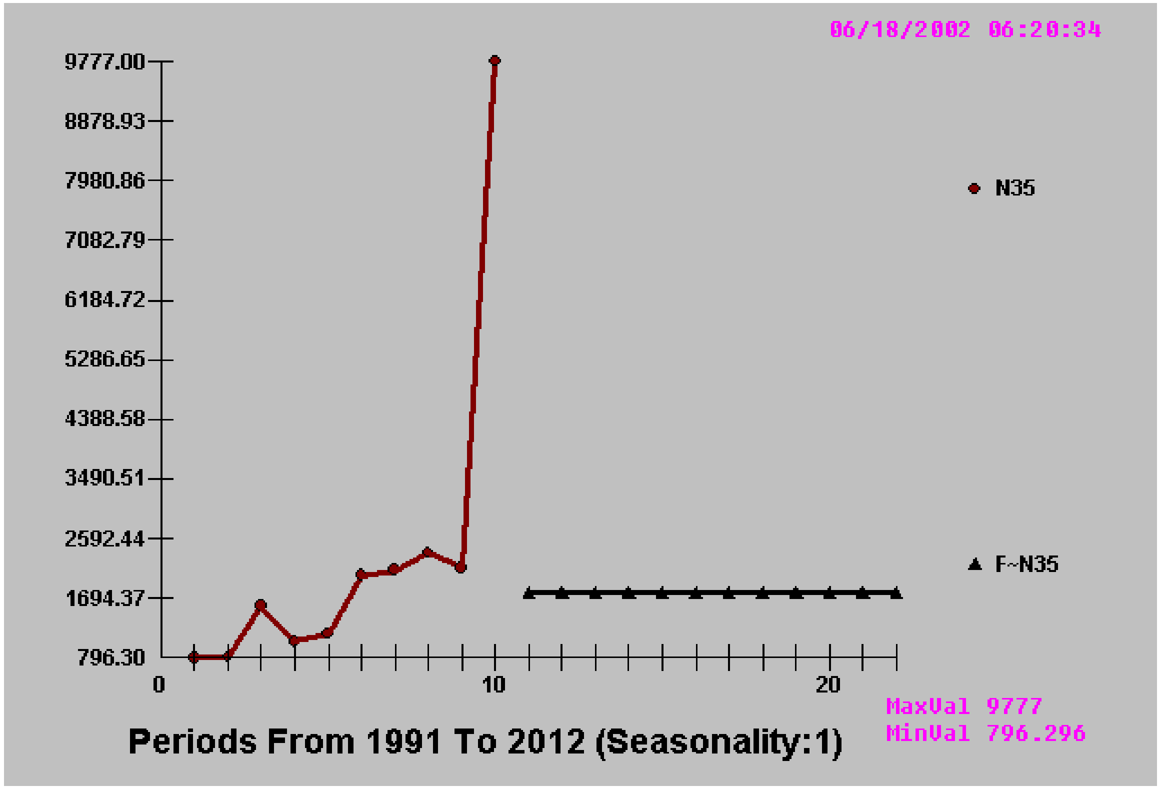
Other Current Assets



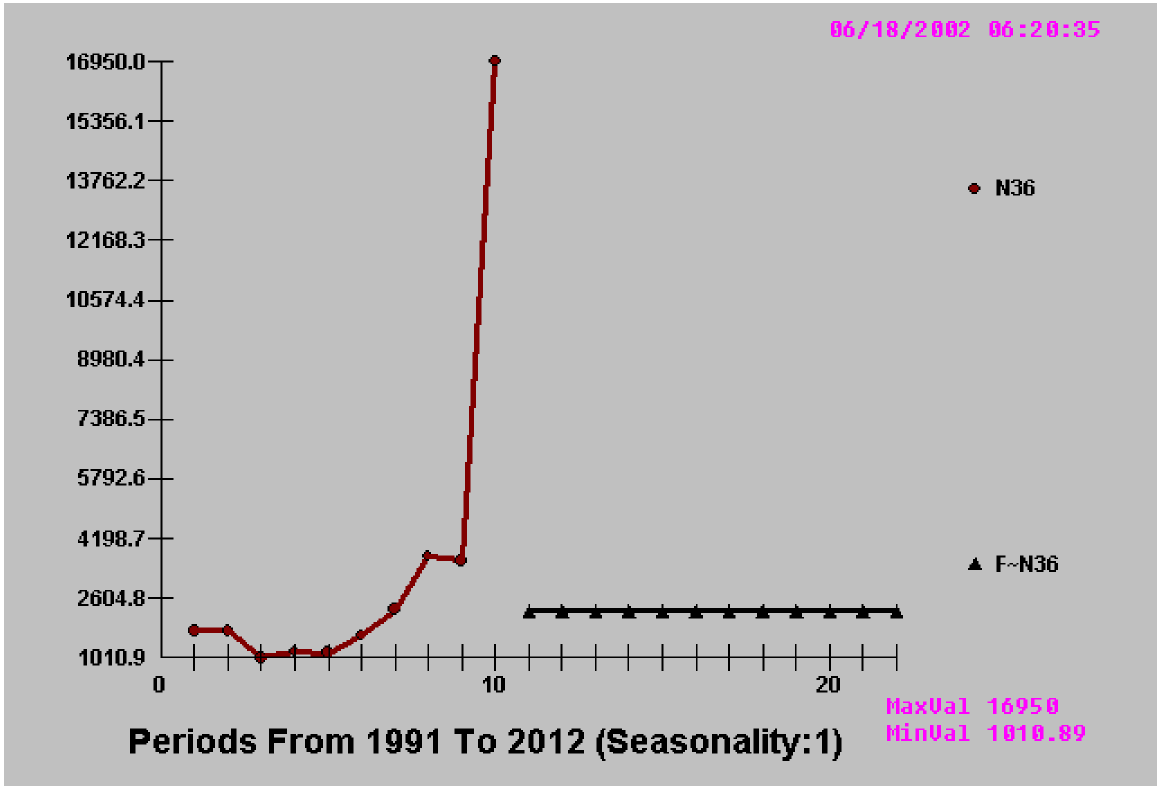
Other Assets



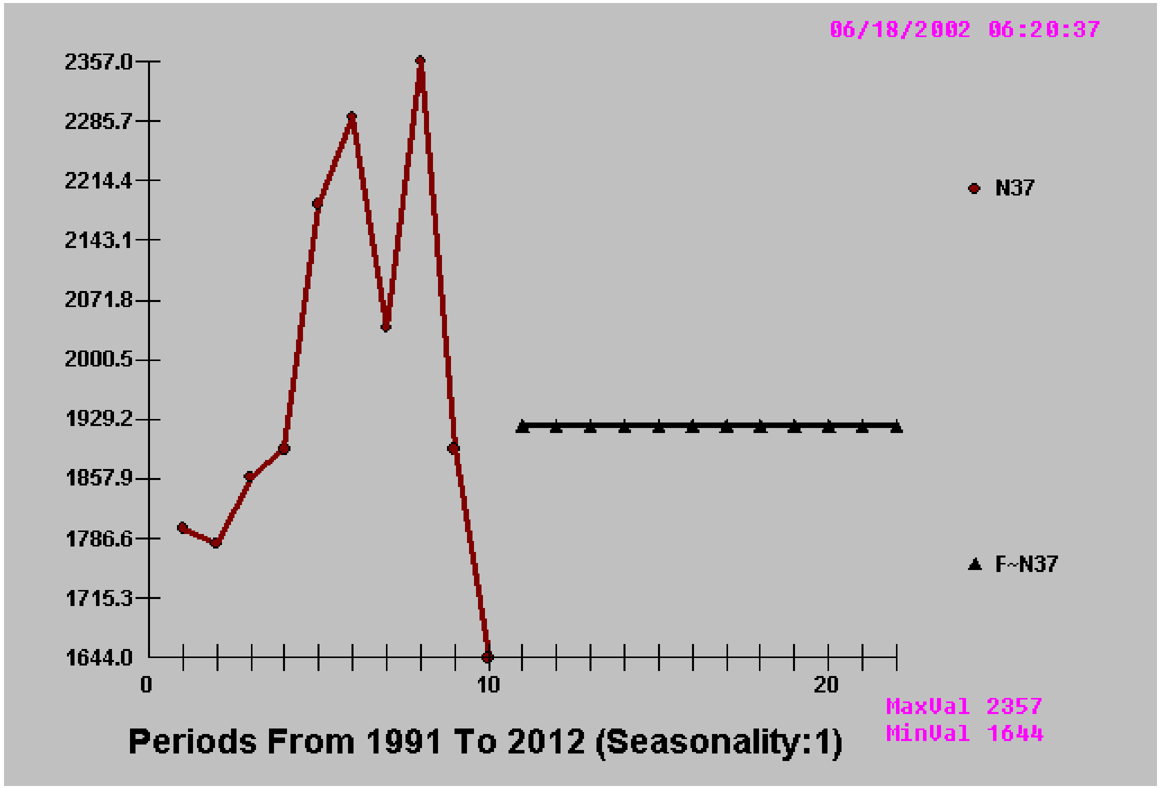
Accounts Payable



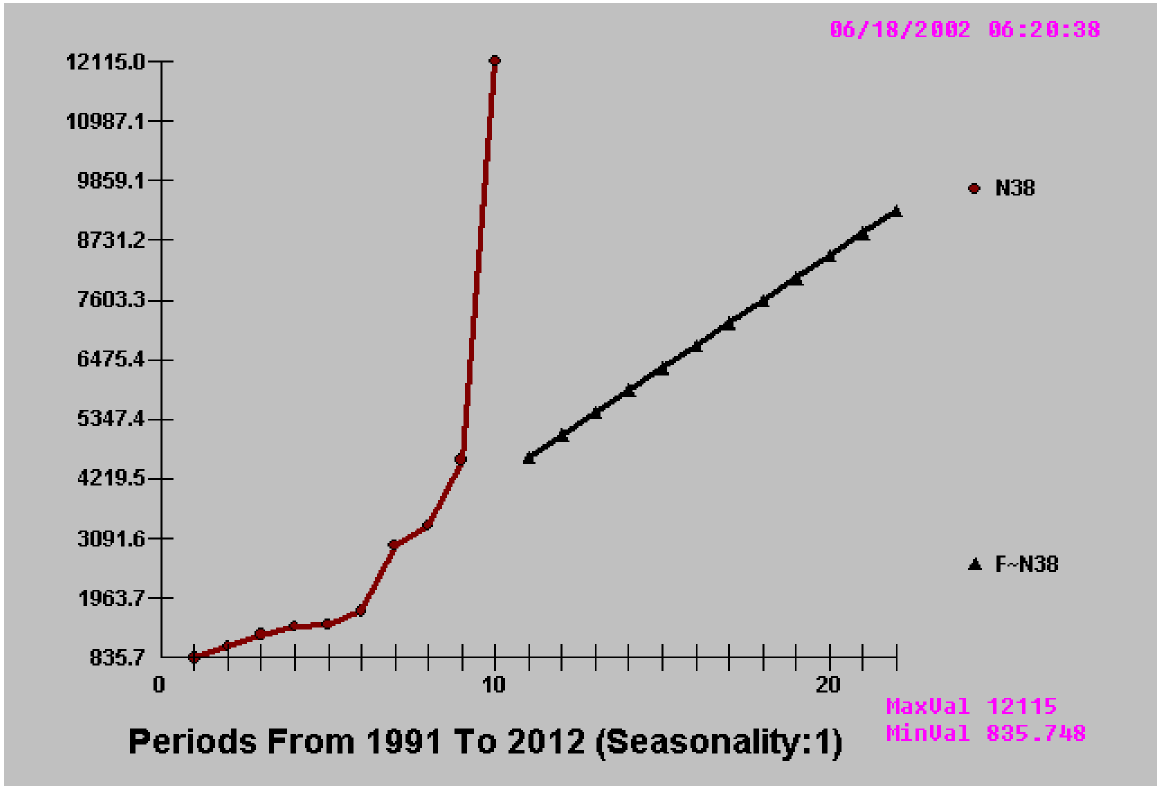
Other Current Liabilities



Deferred Taxes

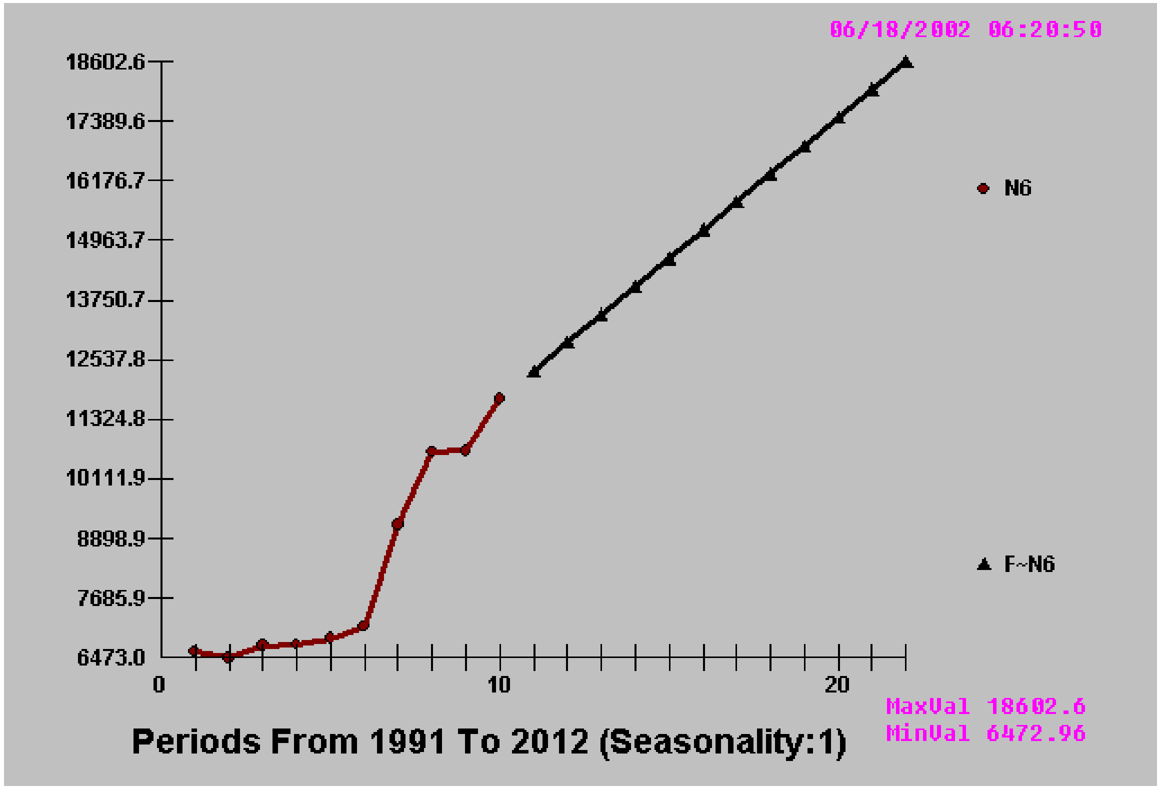


Other Liabilities

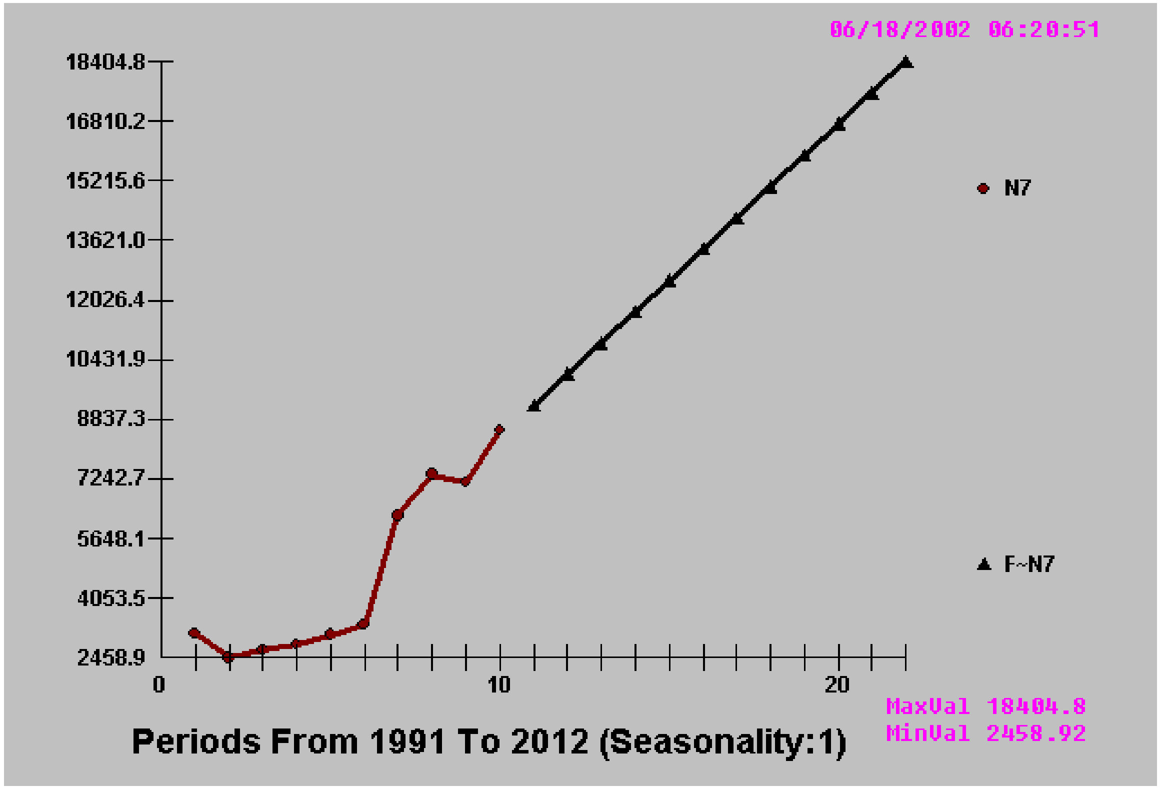


And Now The Unexceptional

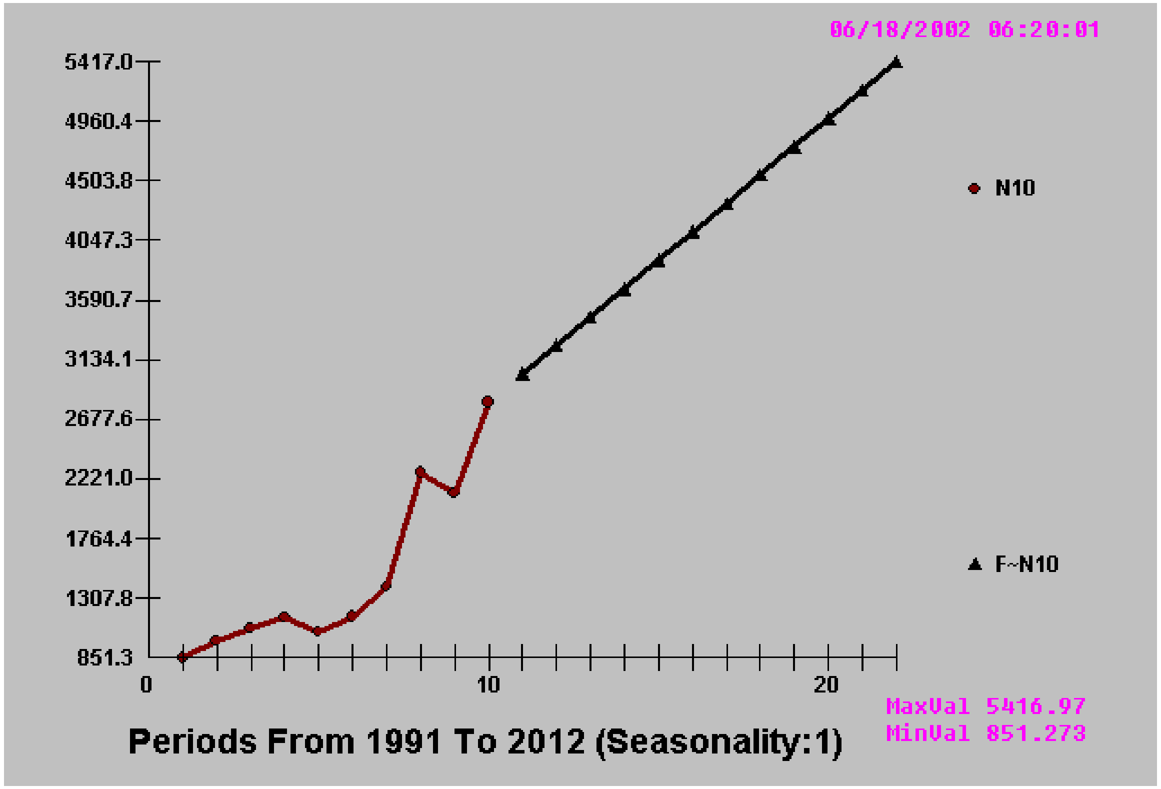
Net Plant & Equipment



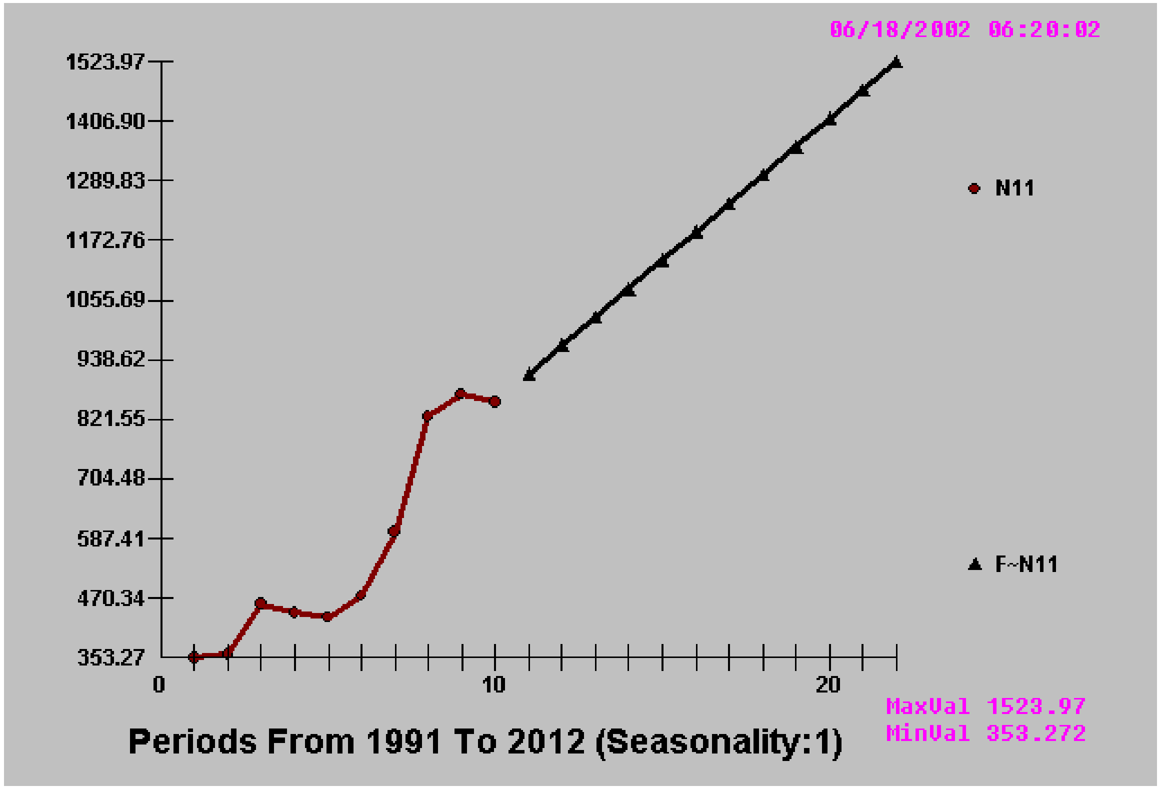
Total Long Term Debt



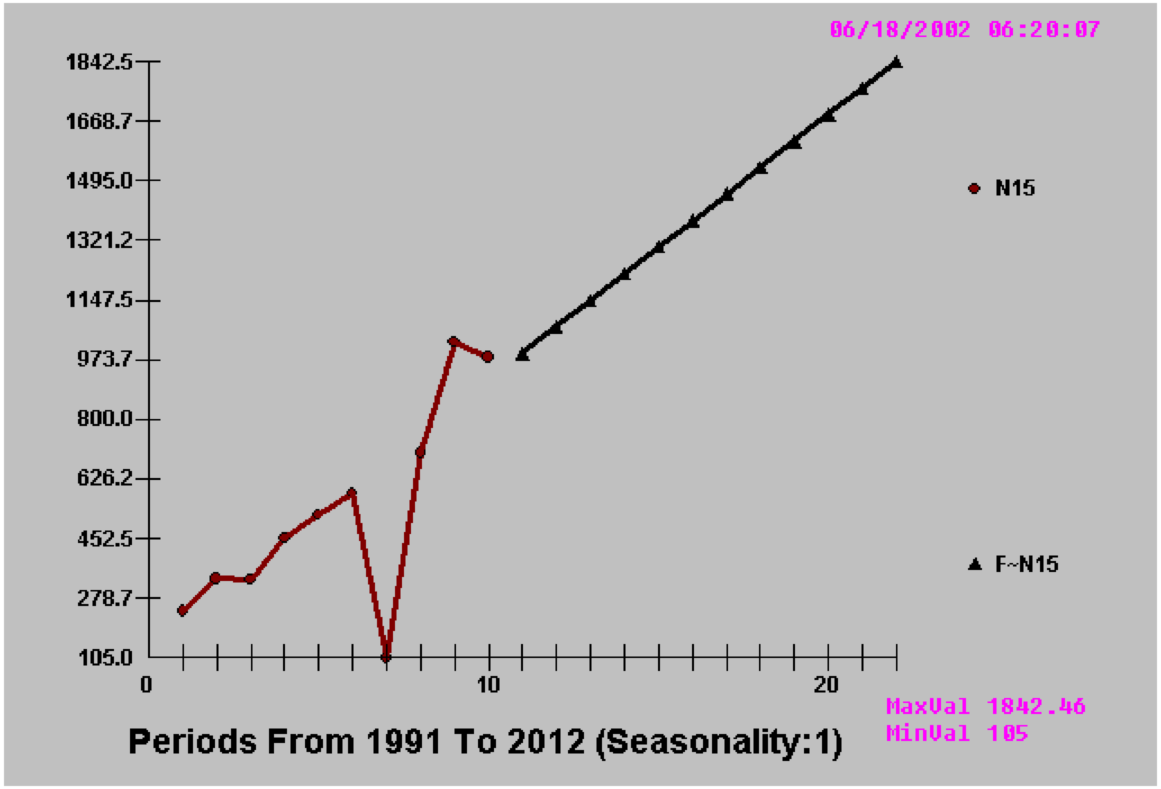
Operating Income Before Deprec



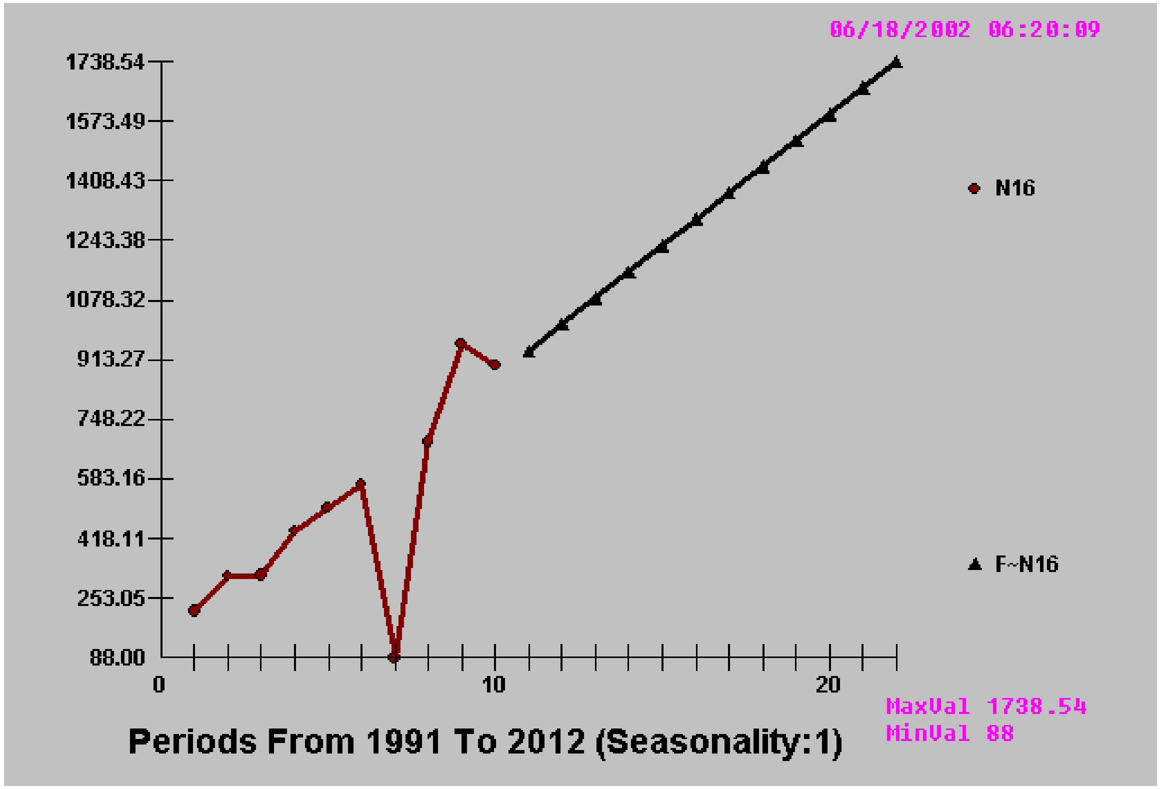
Depreciation & Amortization



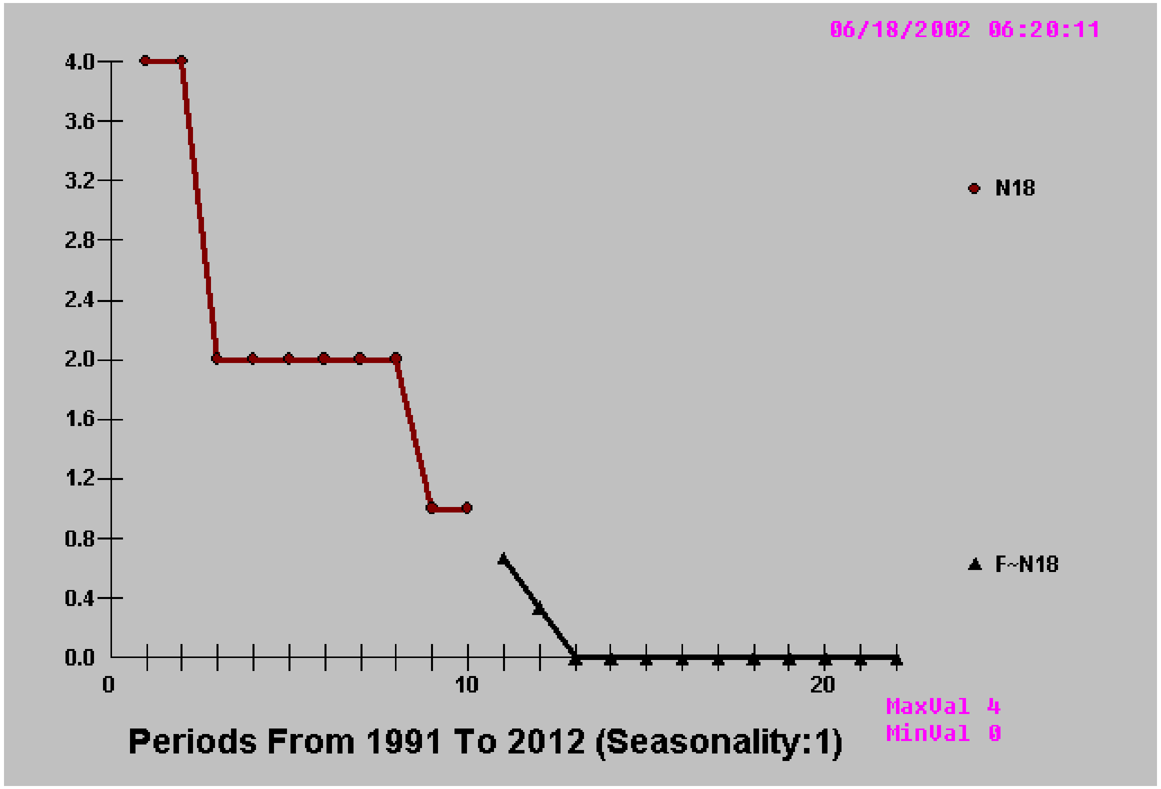
Income Before Special Items



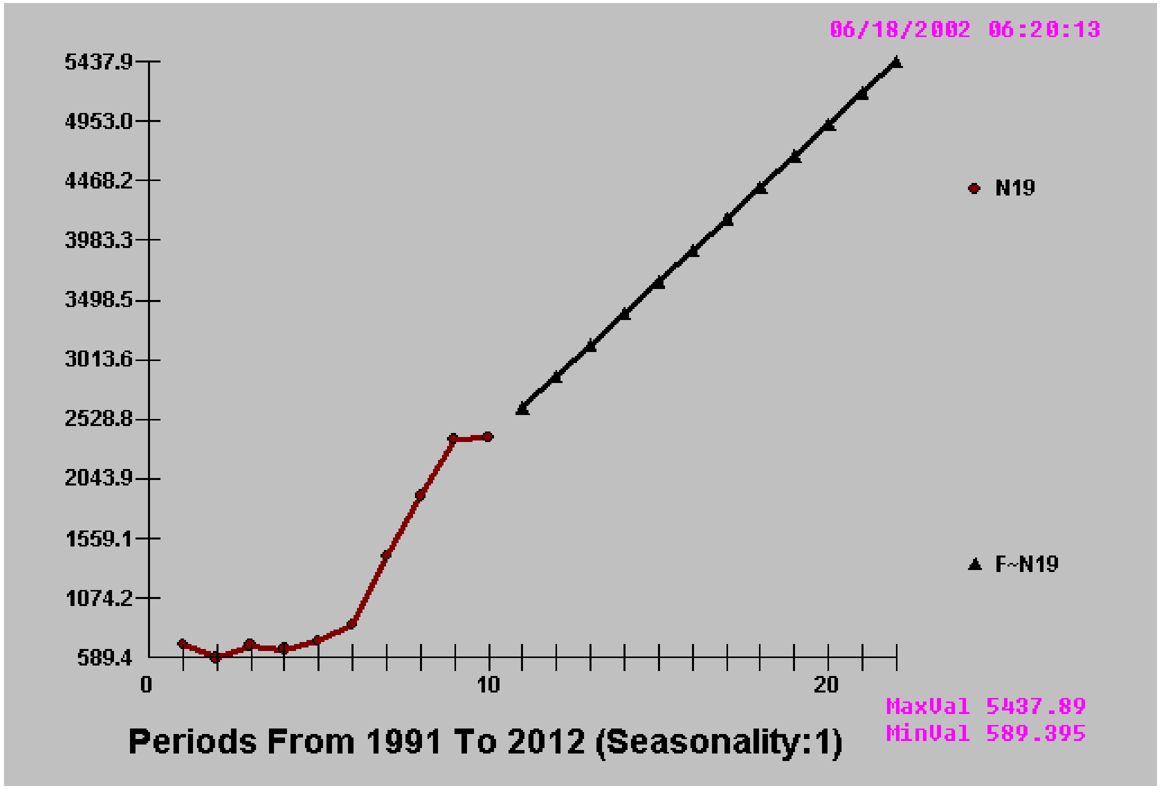
Avail For Common Shrs



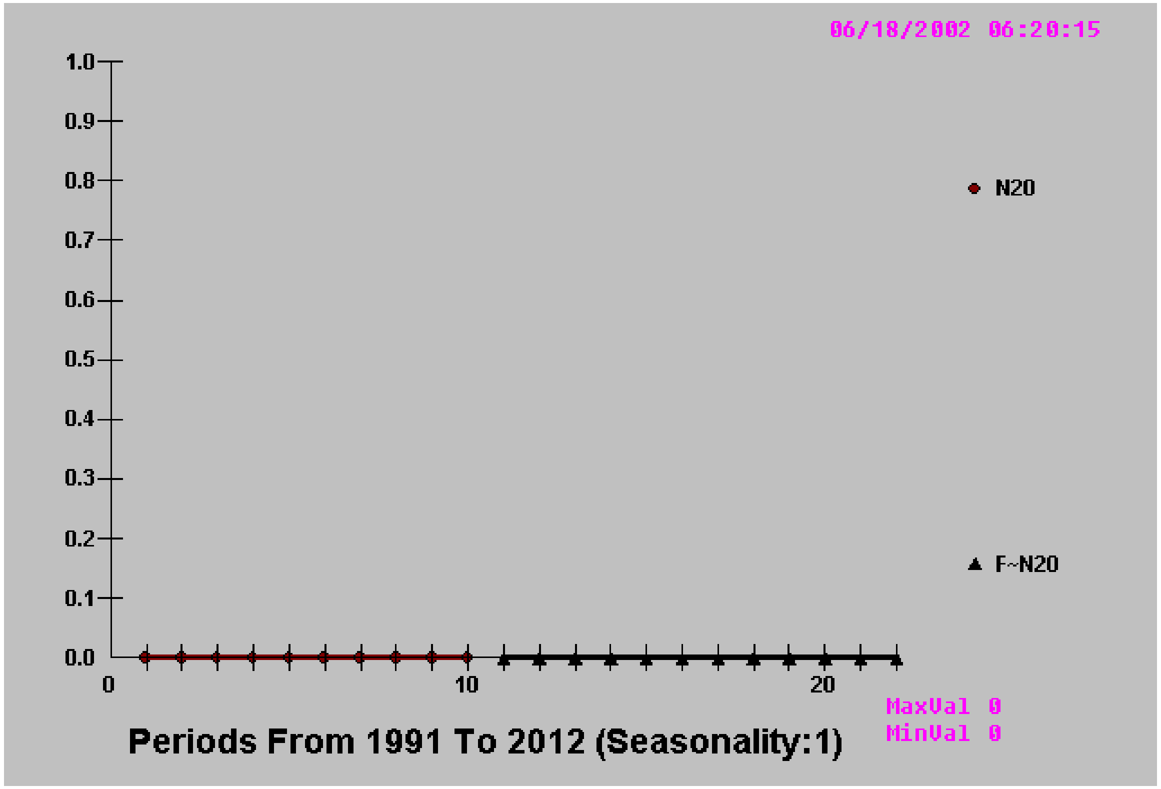
Cumulative Adjustment Factor



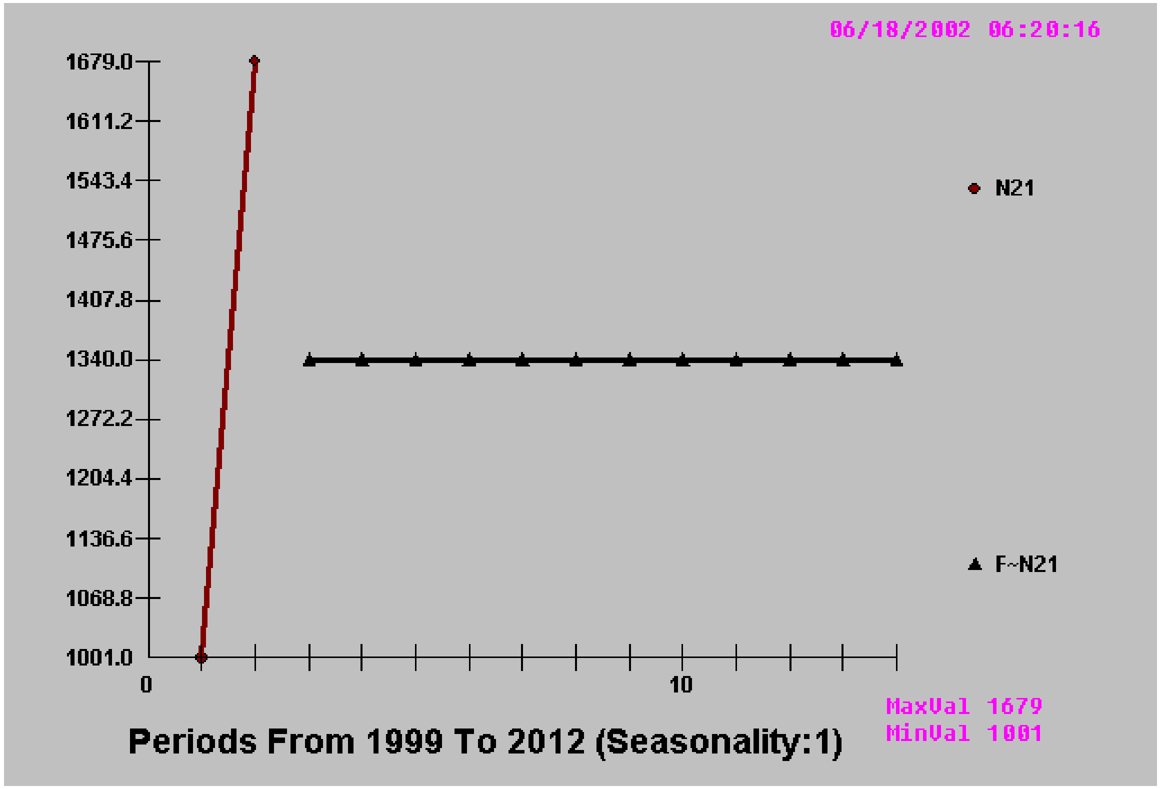
Capital Expenditures



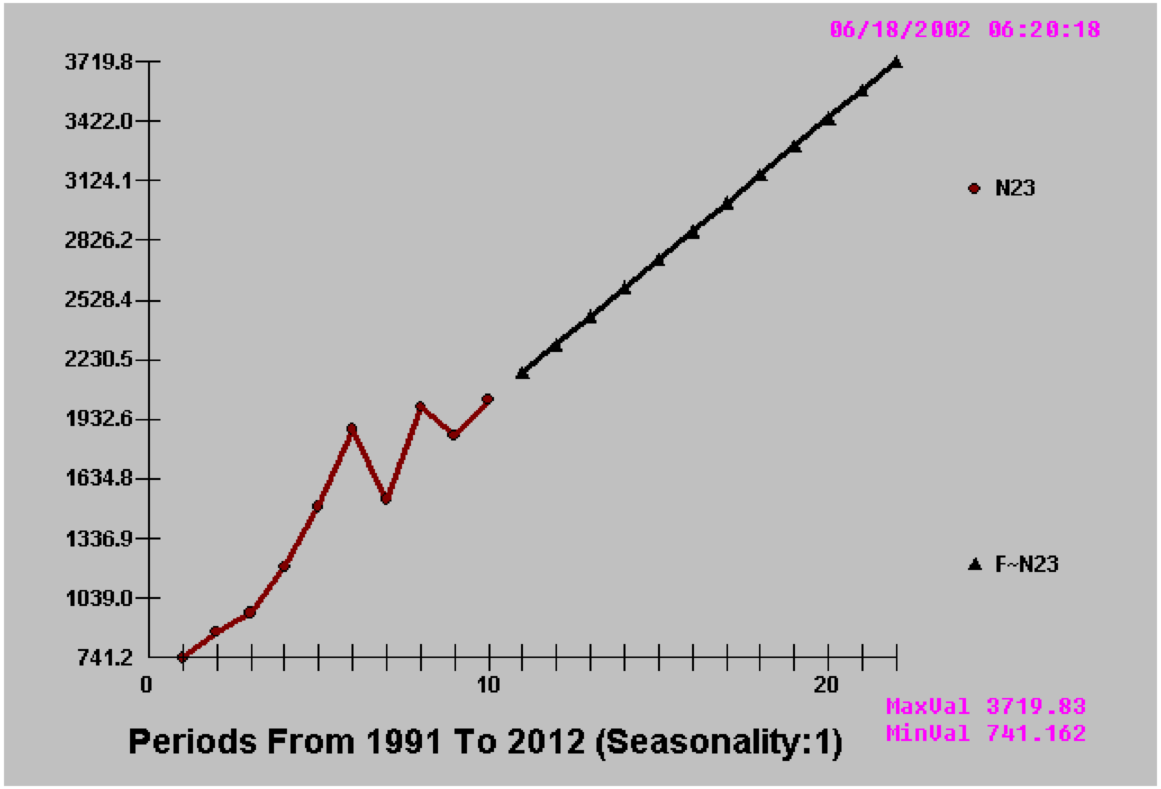
Investments In Others



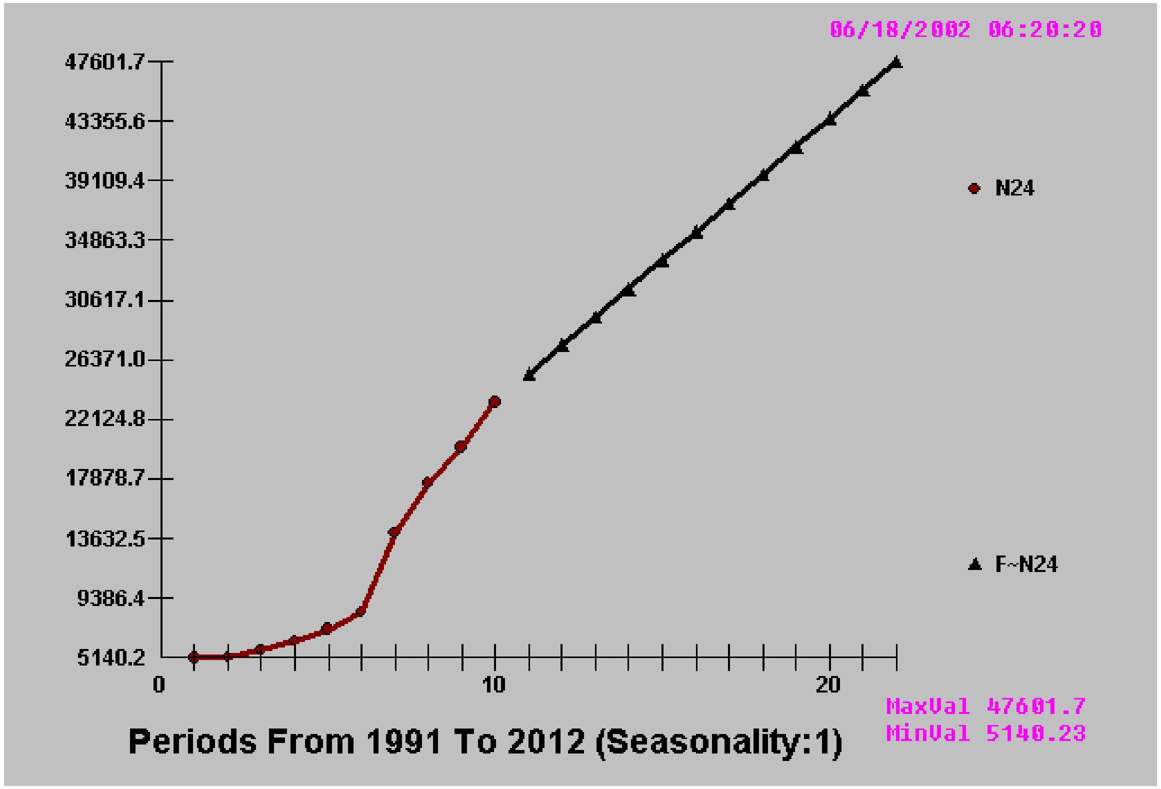
Debt In Current Liabilities



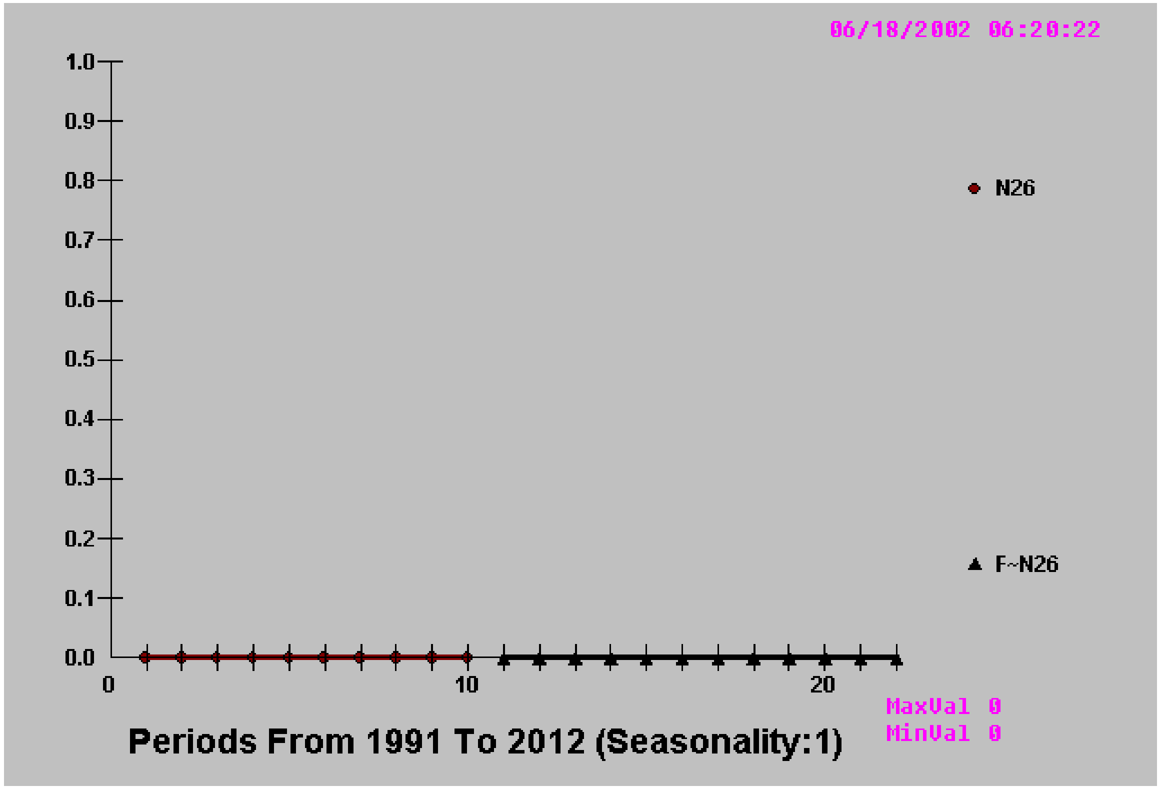
Retained Earnings



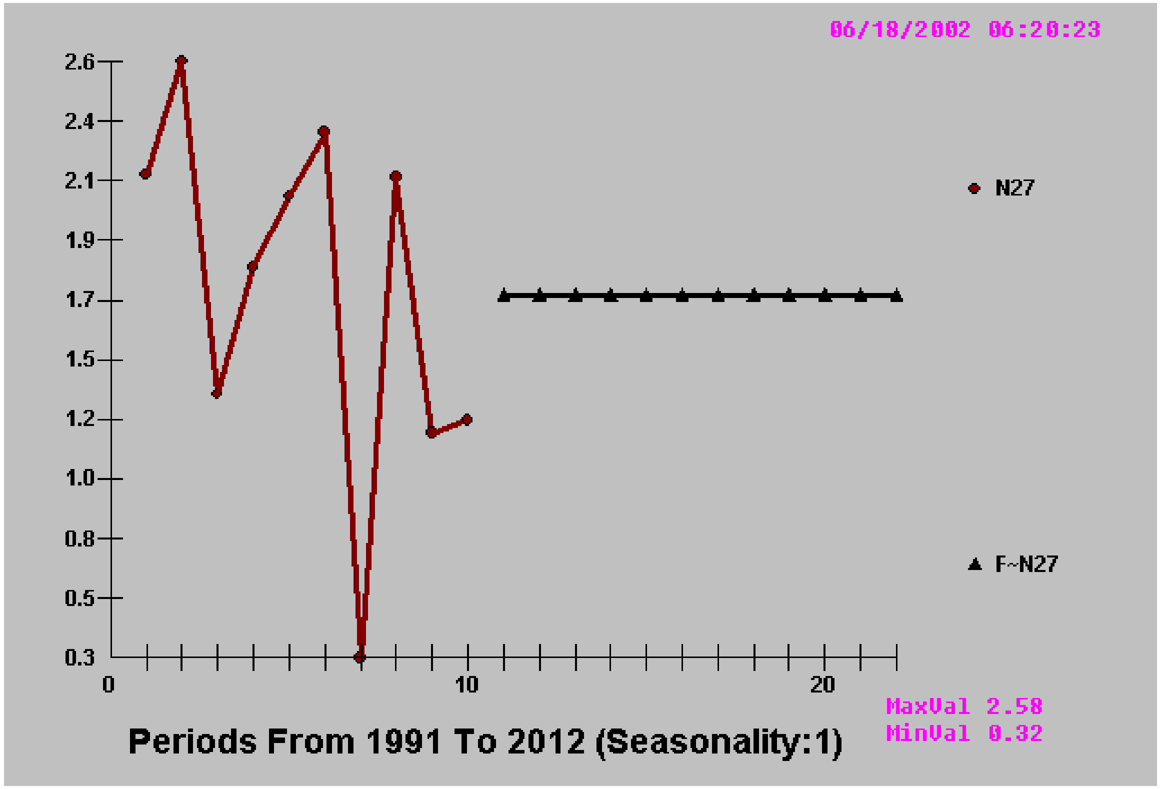
Total Invested Capital



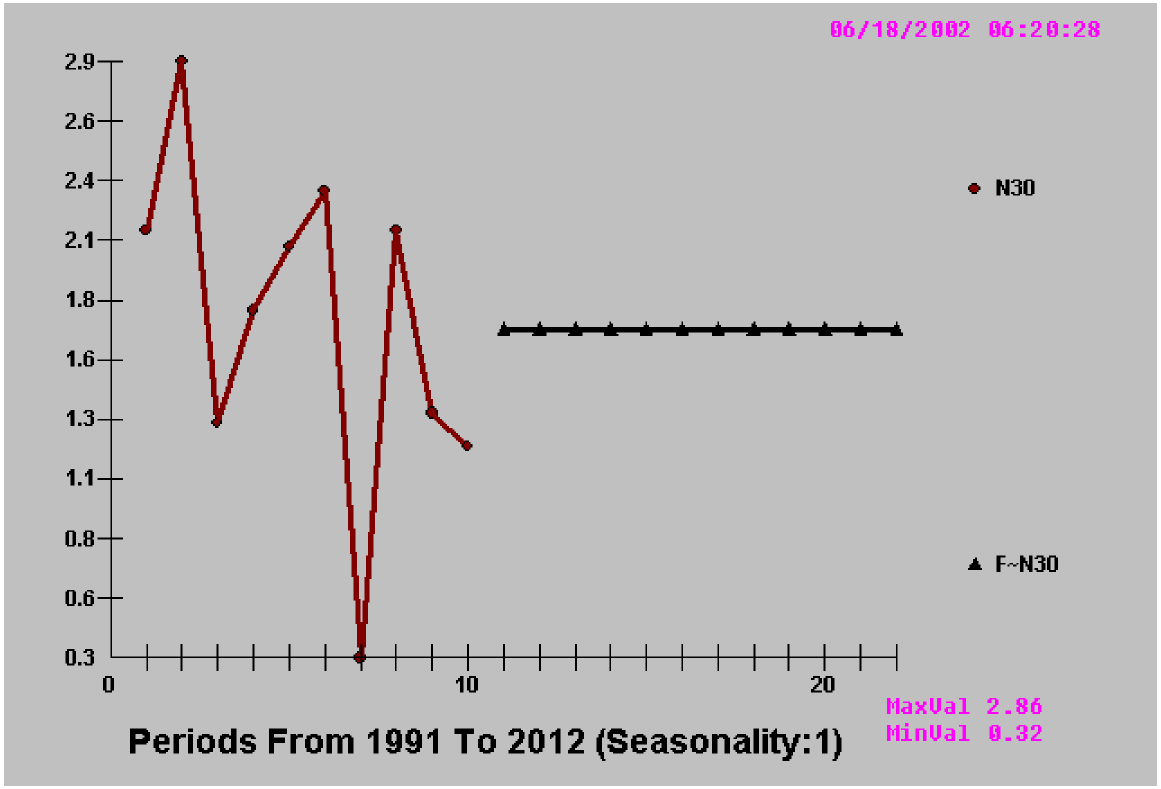
Debt Due In 1 Year



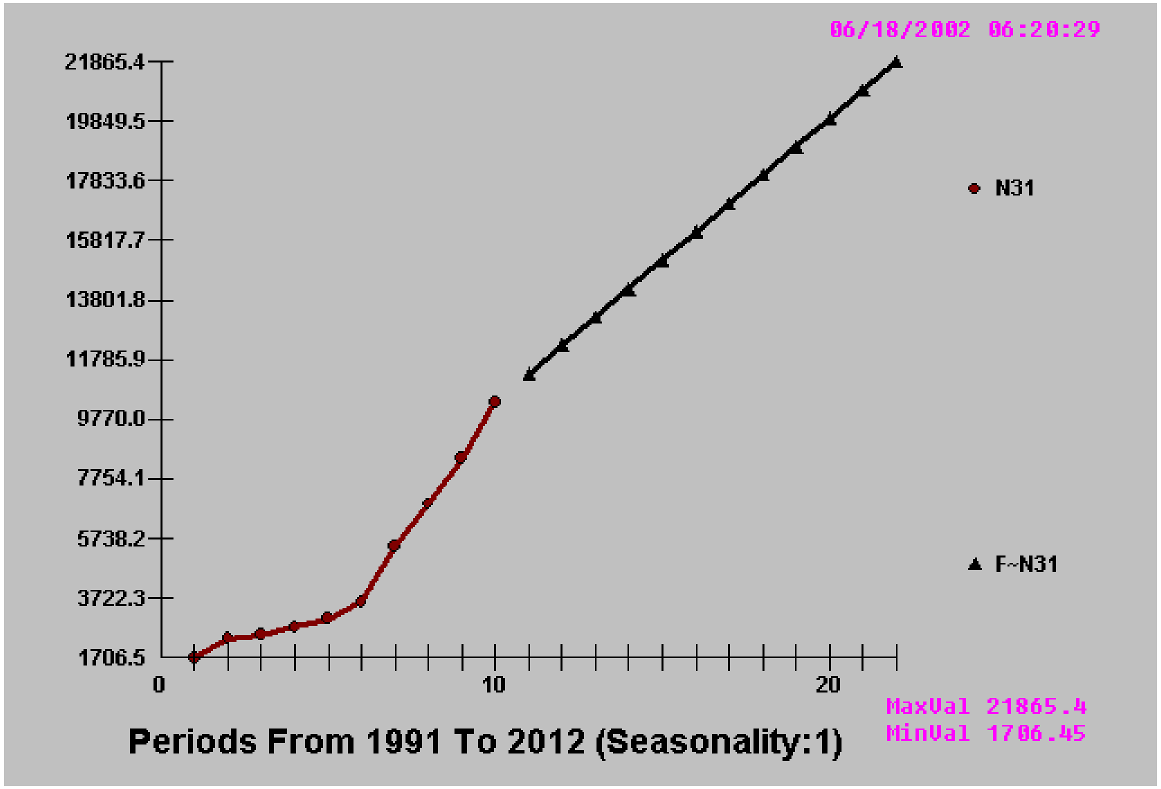
Pri EPS Including Extraord



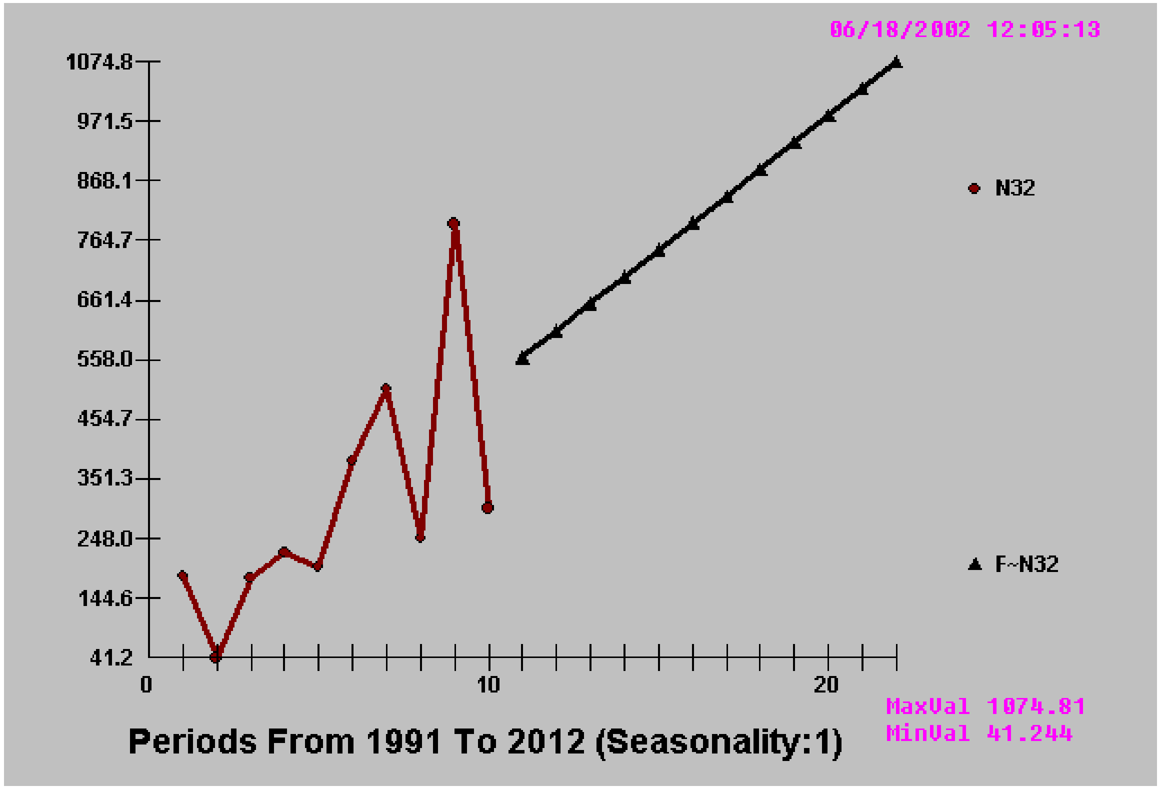
Primary EPS Ex. Extraord



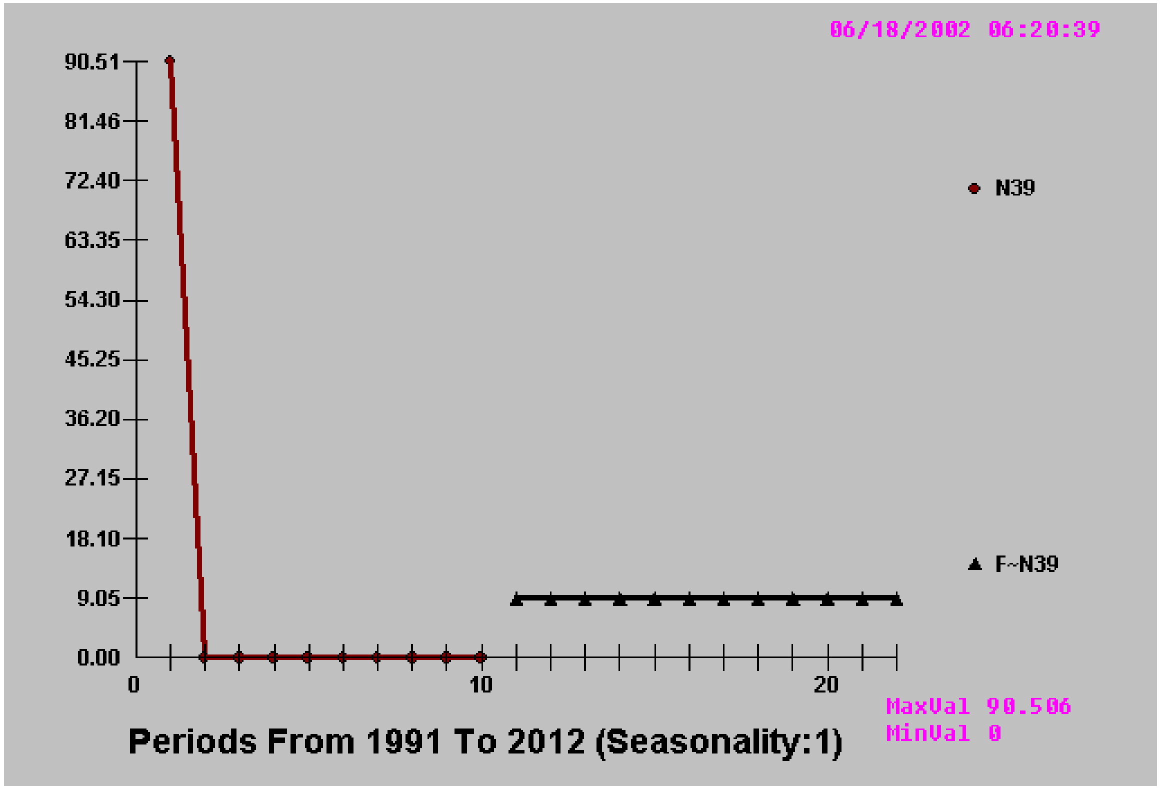
Common Equity



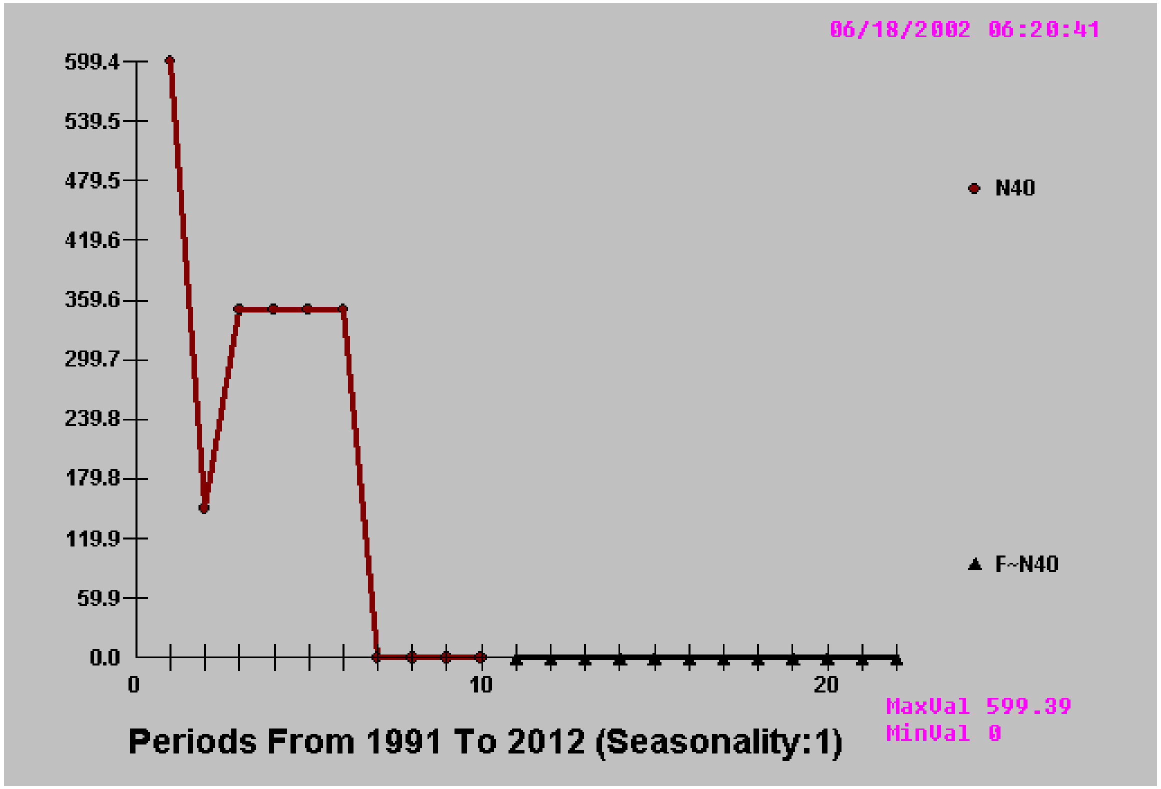
Non-Operating Income



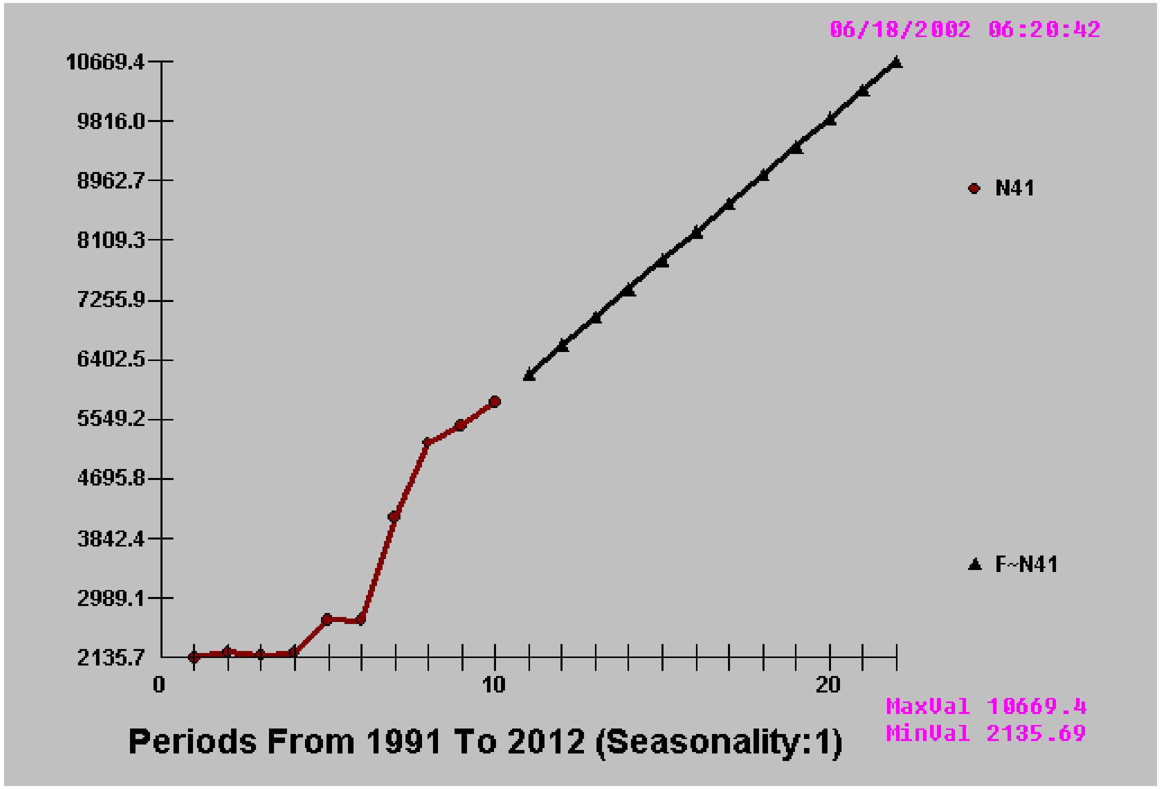
Debt (Convertible)



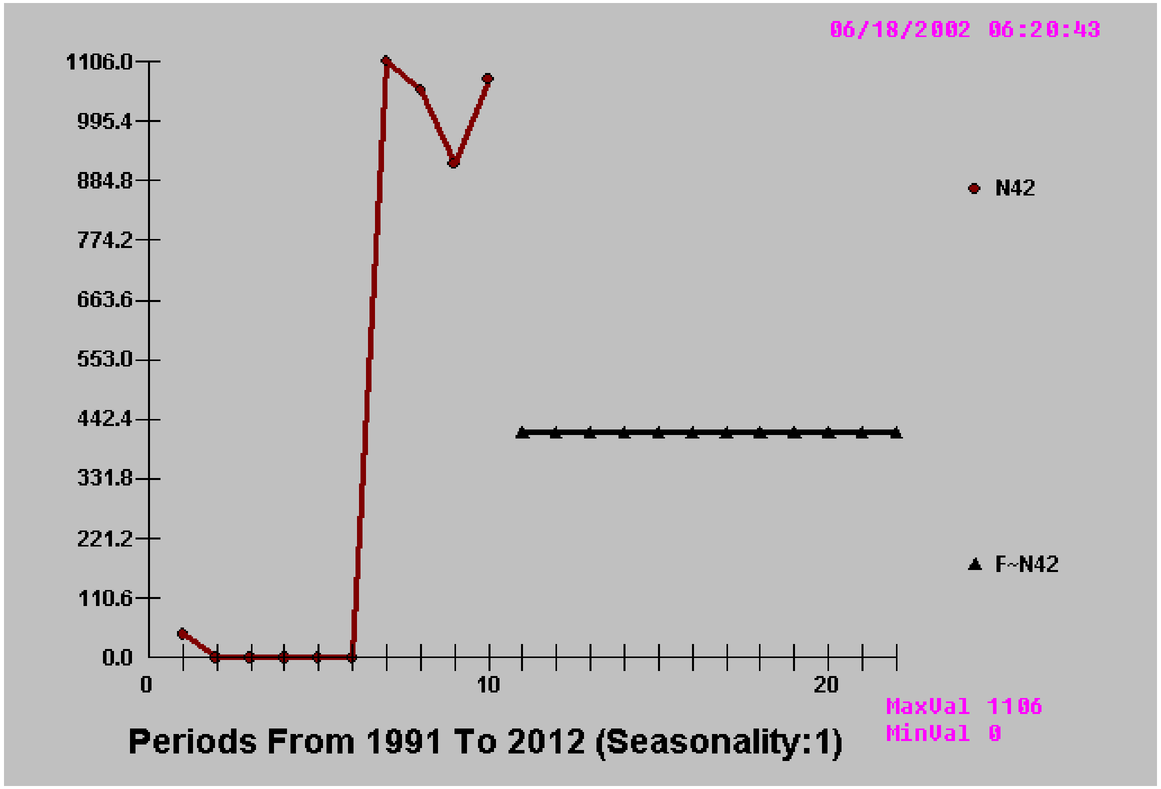
Debt (Subordinated)



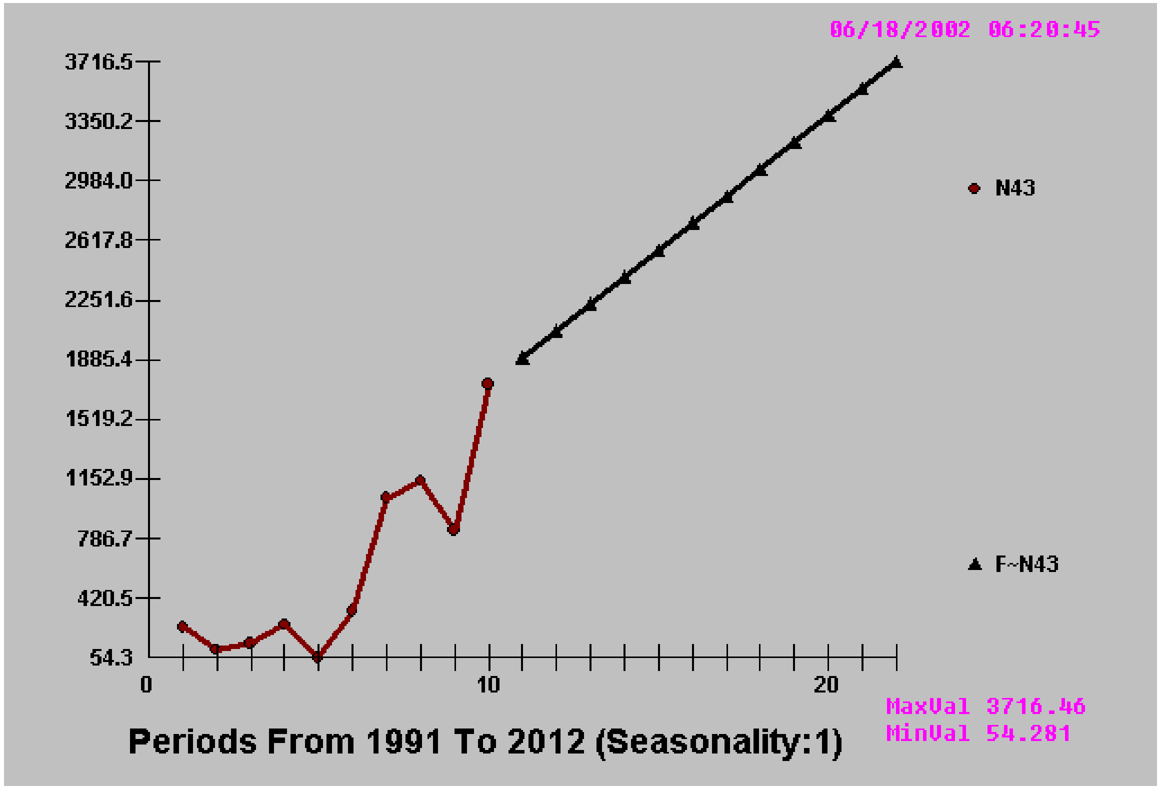
Debt (Notes)


06/18/2002 06:20:42
10669.4
9816.0
8962.7
8109.3
7255.9
6402.5
5549.2
4695.8
3842.4
2989.1
2135.7
0
10
20
● N41
▲ F~N41
Periods From 1991 To 2012 (Seasonality:1)
MaxVal 10669.4
MinVal 2135.69

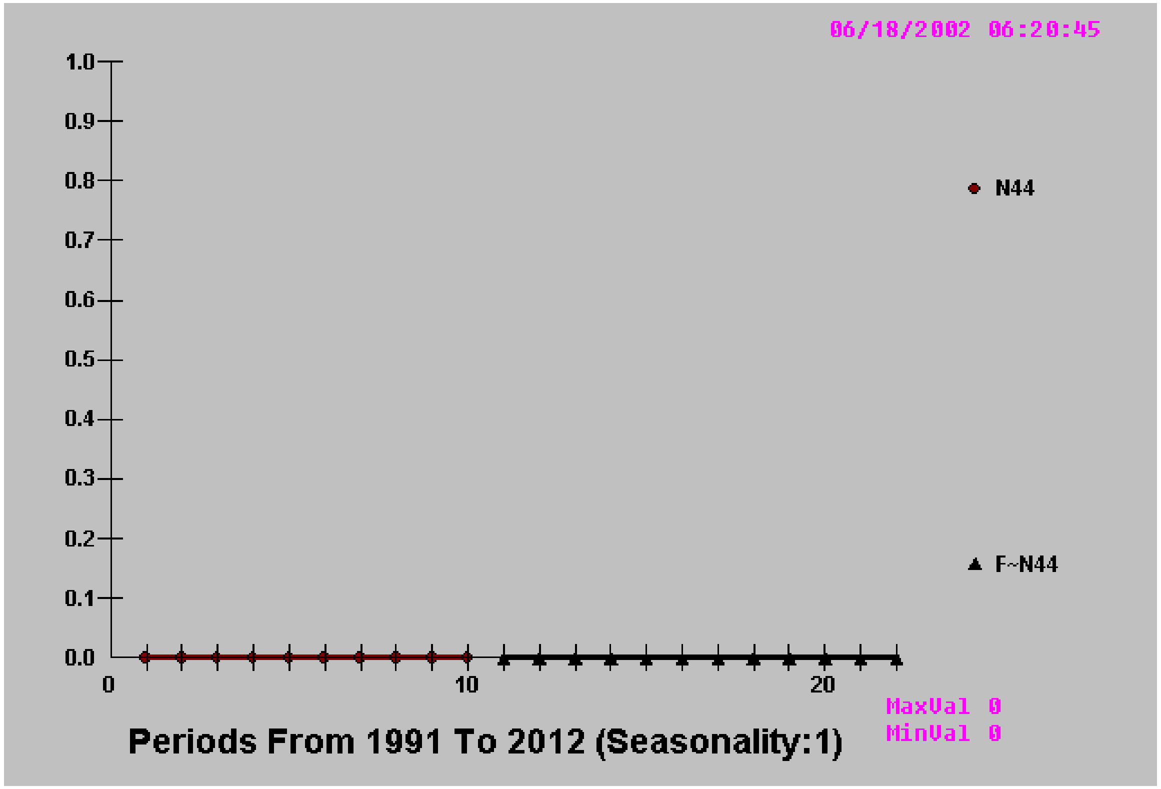
Debt (Debentures)



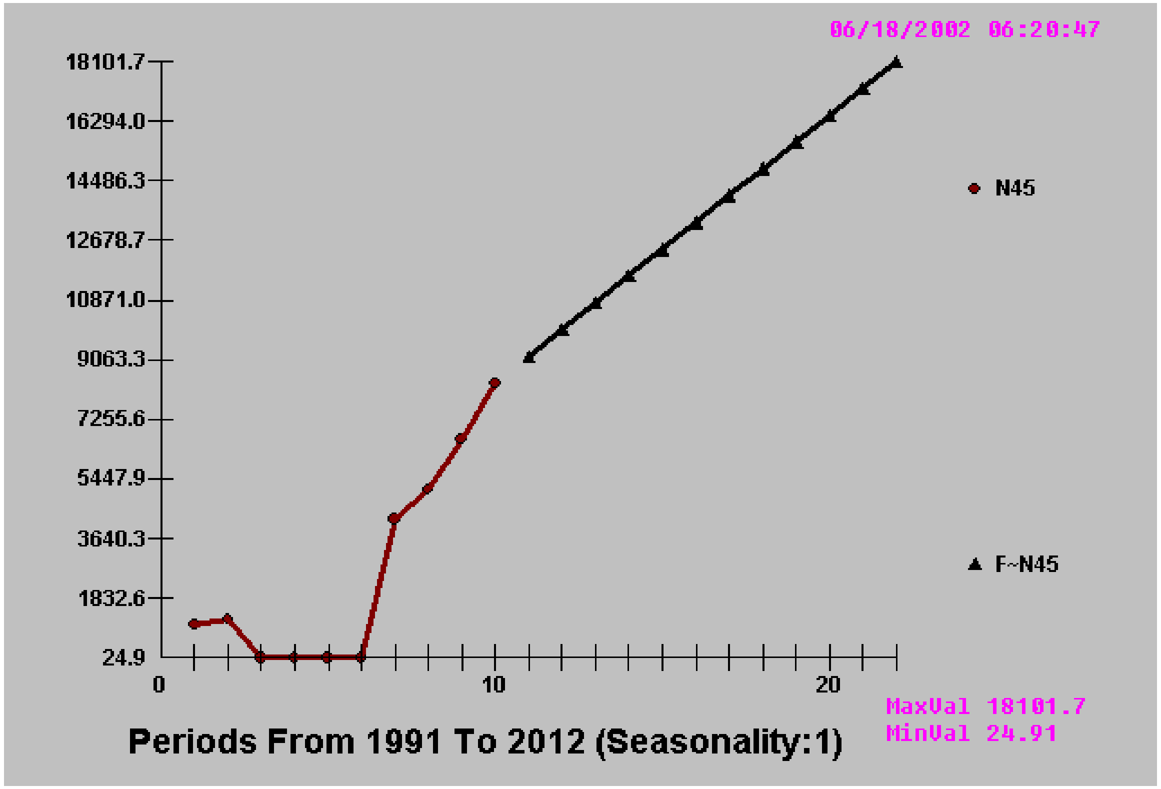
Debt (Other Long-Term)



Capitalized Lease Obligation



Common Stock


06/18/2002 06:20:47
● N45
▲ F~N45
Periods From 1991 To 2012 (Seasonality:1)
MaxVal 18101.7
MinVal 24.91