In [1]:

```
%matplotlib inline

import matplotlib
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
import glob
import pandas as pd
import scipy.stats as stats
import pylab as pl
import pymysql
import dill

from datetime import timedelta
from functools import partial
from sqlalchemy import create_engine
from multiprocessing import Pool, cpu_count
from statsmodels.regression.linear_model import GLS

engine = create_engine('mysql+pymysql://root:maxsonic@localhost:3306/gta_data?charset=utf8')
```

In [2]:

```
def applyParallel(dfGrouped, func, pool_size=3):
    p = Pool(pool_size)
    ret_list = p.map(func, [group for name, group in dfGrouped])
    p.close()
    p.join()
    return pd.concat(ret_list)
```

In [3]:

```
l = [pd.read_csv(filename, dtype={"Symbol": str}) for filename in glob.glob("./original_data/combine
all_data = pd.concat(l, axis=0)
```

In [4]:

```
all_data = all_data[all_data["Filling"] == 0]
all_data.drop("UPDATEID", axis=1, inplace=True)
all_data.drop("BUSINESSTIME", axis=1, inplace=True)
all_data.drop("INDUSTRYNAME", axis=1, inplace=True)
all_data.drop("UTSID", axis=1, inplace=True)
all_data.drop("UPDATESTATE", axis=1, inplace=True)
all_data.drop("UPDATETIME", axis=1, inplace=True)
all_data.drop("PreClosePrice", axis=1, inplace=True)
all_data.drop("OpenPrice", axis=1, inplace=True)
all_data.drop("HighPrice", axis=1, inplace=True)
all_data.drop("LowPrice", axis=1, inplace=True)
all_data.drop("Amount", axis=1, inplace=True)
all_data.drop("Distance", axis=1, inplace=True)
all_data.drop("latestTradingDate", axis=1, inplace=True)
all_data.drop("LatestClosePrice", axis=1, inplace=True)
all_data.drop("StateCode", axis=1, inplace=True)
all_data.drop("AvgPrice", axis=1, inplace=True)
all_data.drop("Change", axis=1, inplace=True)
all_data.drop("ChangeRatio", axis=1, inplace=True)
all_data.drop("TotalShare", axis=1, inplace=True)
all_data.drop("CirculatedShare", axis=1, inplace=True)
all_data.drop("TurnoverRate1", axis=1, inplace=True)
all_data.drop("TurnoverRate2", axis=1, inplace=True)
all_data.drop("CirculatedMarketValue", axis=1, inplace=True)
all_data.drop("Amplitude", axis=1, inplace=True)
all_data.drop("RelativeIPOChange", axis=1, inplace=True)
all_data.drop("RelativeIPOChangeRatio", axis=1, inplace=True)
all_data.drop("MinTickSize", axis=1, inplace=True)
all_data.drop("LimitDown", axis=1, inplace=True)
all_data.drop("LimitUp", axis=1, inplace=True)
all_data.drop("CAT_CHANGEDATE", axis=1, inplace=True)
all_data.drop("Volume", axis=1, inplace=True)
all_data.drop("Filling", axis=1, inplace=True)
all_data.drop("SecurityID", axis=1, inplace=True)
```

In [5]:

```
all_data["pb_ratio_adjust"] = 1 / all_data["pb_ratio"]
all_data["MarketValue_adjust"] = 0 - all_data["MarketValue"]
```

In [6]:

```
stock_info = pd.read_sql_query("select * from STK_STOCKINFO", engine)
idx_quotation = pd.read_sql_query("select TRADINGDATE from IDX_MKT_QUOTATION where SYMBOL='000001'",
susp = pd.read_sql_query("select * from STK_SUSPENTIONINFO", engine)
```

```
/home/maxsonic/anaconda2/lib/python2.7/site-packages/pymysql/cursors.py:166: Warnin
g: (1681, u"'@@SESSION.GTID_EXECUTED' is deprecated and will be removed in a future
 release.")
  result = self._query(query)
```

In [7]:

```
res = idx_quotation[idx_quotation['TRADINGDATE']=='1990-12-20']
idx_quotation.index.values.size
```

Out[7]:

6918

In [8]:

```
def remove_stocks(df):
    # no ST and PT
    df = df[~df["ShortName"].str.contains("ST") | ~df["ShortName"].str.contains("PT")]

    # no suspention or resumption in such a range
    #  date - 3 <= supsention_date <= date + 3
    #  date - 3 <= resumption_date <= date + 3
    max_date_idx = idx_quotation.last_valid_index()
    for date in df.TradingDate.unique():
        if idx_quotation[idx_quotation['TRADINGDATE']==date].index.values.size == 0:
            continue
        date_idx = idx_quotation[idx_quotation['TRADINGDATE']==date].index.values[0]
        three_day_before_idx = date_idx - 3 if date_idx - 3 > 0 else 0
        three_day_after_idx = date_idx + 3 if date_idx + 3 <=  max_date_idx else max_date_idx
        susp_list = susp[(susp["SUSPENTIONDATE"] >= idx_quotation.iloc[three_day_before_idx]["TRADIN
        susp_list = susp_list[(susp_list["SUSPENTIONDATE"] <= idx_quotation.iloc[three_day_after_idx
        susp_list = susp_list[(susp_list["RESUMPTIONDATE"] >= idx_quotation.iloc[three_day_before_id
        susp_list = susp_list[(susp_list["RESUMPTIONDATE"] >= idx_quotation.iloc[three_day_after_idx

        df = df[df["Symbol"].isin(susp_list["SYMBOL"]) == False]

    # no stock that is on market for less than 1 year
    for date in df.TradingDate.unique():
        if idx_quotation[idx_quotation['TRADINGDATE']==date].index.values.size == 0:
            continue
        a_year_before_idx = date_idx - 244 if date_idx - 244 > 0 else 0

        not_a_year_old_stock = stock_info[stock_info["LISTEDDATE"] > idx_quotation.iloc[a_year_befor

        df = df[df["Symbol"].isin(not_a_year_old_stock["SYMBOL"]) == False]
    return df

def rank_fun(df, names):
    factor_column_name, target_rank_name, target_in_name, target_zscore_name = names
    df = remove_stocks(df)
    df[factor_column_name].rank(ascending=True) / (df.shape[0] + 1)
    df[target_rank_name] = df[factor_column_name].rank(ascending=True) / (df.shape[0] + 1)
    df["MarketValue_rank"] = df["MarketValue_adjust"].rank(ascending=True) / (df.shape[0] + 1)
    df[target_in_name] = stats.norm.ppf(df[target_rank_name])
    df["MarketValue_rank_inverse_normal"] = stats.norm.ppf(df["MarketValue_rank"])


    df[target_zscore_name] = (df[target_in_name] - df[target_in_name].mean())/df[target_in_name].std
    df["MarketValue_zscore"] = (df["MarketValue_rank_inverse_normal"] - df["MarketValue_rank_inverse
    return df

def get_factor_zscore_value(all_data, factor_column_name):
    '''
    factor_column_name: the factor we want to test
    all_data: the dataframe that contains all the data
    '''
    target_rank_name = factor_column_name + "_rank"
    target_in_name = factor_column_name + "_rank_inverse_normal"
    target_zscore_name = factor_column_name + "_zscore"

    partial_rank_fun = partial(rank_fun,
                               names=[factor_column_name,
                                      target_rank_name,
                                      target_in_name,
```

```
                                    target_zscore_name])

    all_data = applyParallel(all_data.groupby("TradingDate"), partial_rank_fun, 5)
    all_data = all_data[(all_data[target_zscore_name] <= 3) & (all_data[target_zscore_name] >= -3)]
    all_data = all_data[(all_data["MarketValue_zscore"] <= 3) & (all_data["MarketValue_zscore"] >= -
    all_data = all_data.reset_index()
    return all_data
```

In [9]:

```
all_data = get_factor_zscore_value(all_data, "pb_ratio_adjust")
```

```
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in greater
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in less
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in greater
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in less
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in greater
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in less
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in greater
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in less
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in greater
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in less
  cond1 = (0 < q) & (q < 1)
```

In [10]:

```
def return_rank_fun(df):
    df = remove_stocks(df)
    df["close_price_rank"] = df["ClosePrice"].rank(ascending=True) / (df.shape[0] + 1)
    df["close_price_inverse_normal"] = stats.norm.ppf(df["close_price_rank"])

    df["close_price_rank_zscore"] = (df["close_price_inverse_normal"] - df["close_price_inverse_norm

    return df

def calc_pct_return(df, period=5):
    df = df.reset_index()
    idx = df["index"]
    df = df.drop("index", axis=1)
    df = df.set_index(["TradingDate", "Symbol", "ShortName"])
    forward_return = df.pct_change(period).shift(-period)
    return forward_return.reset_index().set_index(idx)

def get_forward_return(all_data, period=5):
    partial_calc_pct_return = partial(calc_pct_return, period=period)
    forward_return = applyParallel(all_data[["TradingDate", "Symbol", "ClosePrice", "ShortName"]].gr
                                   partial_calc_pct_return,
                                   8)
    forward_return = applyParallel(forward_return.groupby("TradingDate"), return_rank_fun, 8)
    forward_return = forward_return[(forward_return["close_price_rank_zscore"] <= 3) & (forward_retu
    return forward_return

forward_return = get_forward_return(all_data)
forward_return.head()
```

```
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in greater
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in less
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in greater
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in less
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in greater
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in less
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in greater
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in less
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in greater
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in less
```

```
  cond1 = (0 < q) & (q < 1)

/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in greater
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in less
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in greater
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in less
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in greater
  cond1 = (0 < q) & (q < 1)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1901: RuntimeWarning: invalid value encountered in less
  cond1 = (0 < q) & (q < 1)
```

Out[10]:

index	TradingDate	Symbol	ShortName	ClosePrice	close_price_rank	close_price_inverse_norr
0	1990-12-31	600601	方正科技	0.040199	0.200000	-0.8416
2	1990-12-31	600602	仪电电子	0.040792	0.400000	-0.2533
1	1990-12-31	600651	飞乐音响	0.040867	0.800000	0.8416
3	1990-12-31	600654	飞乐股份	0.040827	0.600000	0.2533
5	1991-01-02	600601	方正科技	0.035289	0.166667	-0.9674

In [11]:

```
# Factor Testing
def make_x_y_for_test(all_data, factor_column_name):
    factor_zscore_name = factor_column_name + "_zscore"
    X_before_dummy = all_data[["TradingDate", factor_zscore_name,
                               "INDUSTRYCODE", "MarketValue_zscore"]]
    # Hack here to make the index of X_dummy align with return factor
    X_before_dummy = X_before_dummy[X_before_dummy.index.isin(forward_return.index)]
    X_before_dummy = X_before_dummy.dropna()
    X_dummy = pd.concat([X_before_dummy.drop("INDUSTRYCODE",  axis=1,),
                         pd.get_dummies(X_before_dummy["INDUSTRYCODE"])],
                        axis=1)
    X_dummy = pd.concat([X_dummy.drop("MarketValue_zscore",axis=1),
                         X_dummy["MarketValue_zscore"]], axis=1)

    Y = forward_return[["TradingDate", "close_price_rank_zscore"]]
    Y = Y[Y.index.isin(X_dummy.index)]
    X_dummy = X_dummy[X_dummy.index.isin(Y.index)]
    return X_dummy, Y
```

In [12]:

```
def get_factor_test(X_dummy, Y):
    res = X_dummy.merge(Y.drop("TradingDate", axis=1), left_index=True, right_index=True)

    return applyParallel(res.groupby("TradingDate"), _get_factor_test, 8)

def _get_factor_test(df):

    x_test = df.drop("TradingDate", axis=1).drop("close_price_rank_zscore", axis=1)
    y_test = df[["close_price_rank_zscore"]]

    model = GLS(y_test, x_test)
    fit_res = model.fit()
    return pd.DataFrame({"Date": df["TradingDate"],
                                "t": fit_res.tvalues,
                                "p": fit_res.pvalues})

res = get_factor_test(*make_x_y_for_test(all_data, "pb_ratio_adjust"))
```

```
/home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/regression/linear_m
odel.py:1353: RuntimeWarning: divide by zero encountered in double_scalars
  return np.dot(wresid, wresid) / self.df_resid
/home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/base/model.py:1118:
RuntimeWarning: invalid value encountered in multiply
  cov_p = self.normalized_cov_params * scale
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:879: RuntimeWarning: invalid value encountered in greater
  return (self.a < x) & (x < self.b)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:879: RuntimeWarning: invalid value encountered in less
  return (self.a < x) & (x < self.b)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1818: RuntimeWarning: invalid value encountered in less_equal
  cond2 = cond0 & (x <= self.a)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/regression/linear_m
odel.py:1353: RuntimeWarning: divide by zero encountered in double_scalars
  return np.dot(wresid, wresid) / self.df_resid
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:879: RuntimeWarning: invalid value encountered in greater
  return (self.a < x) & (x < self.b)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/base/model.py:1118:
RuntimeWarning: invalid value encountered in multiply
  cov_p = self.normalized_cov_params * scale
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:879: RuntimeWarning: invalid value encountered in less
  return (self.a < x) & (x < self.b)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1818: RuntimeWarning: invalid value encountered in less_equal
  cond2 = cond0 & (x <= self.a)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/regression/linear_m
odel.py:1353: RuntimeWarning: invalid value encountered in double_scalars
  return np.dot(wresid, wresid) / self.df_resid
/home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/base/model.py:1036:
RuntimeWarning: invalid value encountered in divide
  return self.params / self.bse
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:879: RuntimeWarning: invalid value encountered in greater
  return (self.a < x) & (x < self.b)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:879: RuntimeWarning: invalid value encountered in less
```

```
    return (self.a < x) & (x < self.b)

  /home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
  re.py:1818: RuntimeWarning: invalid value encountered in less_equal
    cond2 = cond0 & (x <= self.a)
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/base/model.py:1036:
  RuntimeWarning: invalid value encountered in divide
    return self.params / self.bse
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
  re.py:1818: RuntimeWarning: invalid value encountered in less_equal
    cond2 = cond0 & (x <= self.a)
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
  re.py:879: RuntimeWarning: invalid value encountered in greater
    return (self.a < x) & (x < self.b)
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
  re.py:879: RuntimeWarning: invalid value encountered in less
    return (self.a < x) & (x < self.b)
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/base/model.py:1036:
  RuntimeWarning: invalid value encountered in divide
    return self.params / self.bse
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
  re.py:879: RuntimeWarning: invalid value encountered in less
    return (self.a < x) & (x < self.b)
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
  re.py:879: RuntimeWarning: invalid value encountered in greater
    return (self.a < x) & (x < self.b)
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
  re.py:1818: RuntimeWarning: invalid value encountered in less_equal
    cond2 = cond0 & (x <= self.a)
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/regression/linear_m
  odel.py:1353: RuntimeWarning: divide by zero encountered in double_scalars
    return np.dot(wresid, wresid) / self.df_resid
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/base/model.py:1118:
  RuntimeWarning: invalid value encountered in multiply
    cov_p = self.normalized_cov_params * scale
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/base/model.py:1036:
  RuntimeWarning: invalid value encountered in divide
    return self.params / self.bse
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
  re.py:879: RuntimeWarning: invalid value encountered in greater
    return (self.a < x) & (x < self.b)
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
  re.py:879: RuntimeWarning: invalid value encountered in less
    return (self.a < x) & (x < self.b)
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
  re.py:1818: RuntimeWarning: invalid value encountered in less_equal
    cond2 = cond0 & (x <= self.a)
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/base/model.py:1036:
  RuntimeWarning: invalid value encountered in divide
    return self.params / self.bse
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
  re.py:879: RuntimeWarning: invalid value encountered in greater
    return (self.a < x) & (x < self.b)
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
  re.py:879: RuntimeWarning: invalid value encountered in less
    return (self.a < x) & (x < self.b)
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
  re.py:1818: RuntimeWarning: invalid value encountered in less_equal
    cond2 = cond0 & (x <= self.a)
  /home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/base/model.py:1036:
  RuntimeWarning: invalid value encountered in divide
    return self.params / self.bse
```

```
/home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/base/model.py:1036:
RuntimeWarning: invalid value encountered in divide
  return self.params / self.bse
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:879: RuntimeWarning: invalid value encountered in greater
  return (self.a < x) & (x < self.b)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:879: RuntimeWarning: invalid value encountered in less
  return (self.a < x) & (x < self.b)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/scipy/stats/_distn_infrastructu
re.py:1818: RuntimeWarning: invalid value encountered in less_equal
  cond2 = cond0 & (x <= self.a)
/home/maxsonic/anaconda2/lib/python2.7/site-packages/statsmodels/base/model.py:1036:
RuntimeWarning: invalid value encountered in divide
  return self.params / self.bse
```

In [13]:

```
# Perform T test
# y_predict = np.dot(x_test, model_ransac.estimator_.coef_.T)
# res = stats.ttest_ind(y_test, y_predict)
# t = res[0]
# p = res[1]
# print "t is %s, p is %s" % res
```

In [14]:

```
# Get IC value
```

In [15]:

```
def get_x_y_for_ic(all_data, factor_column_name):

    target_zscore_name = factor_column_name + "_zscore"
    Y_ = all_data[["TradingDate", target_zscore_name]]
    Y_ = Y_[Y_.index.isin(forward_return.index)]

    X__ = all_data[["TradingDate", "INDUSTRYCODE", "MarketValue_zscore"]]
    # Hack here to make the index of X_dummy align with return factor
    X__ = X__[X__.index.isin(forward_return.index)]
    X__ = X__.dropna()
    X_ = pd.concat([X__.drop("INDUSTRYCODE",  axis=1,), pd.get_dummies(X__["INDUSTRYCODE"])],
                       axis=1)
    X_ = pd.concat([X_.drop("MarketValue_zscore",axis=1),
                    X_["MarketValue_zscore"]], axis=1)
    return X_, Y_
```

In [16]:

```
def get_ics(X_, Y_, forward_return, factor_column_name):
    target_zscore_name = factor_column_name + "_zscore"
    res = X_.merge(Y_.drop("TradingDate", axis=1), left_index=True, right_index=True)
    clean_forward_return = forward_return[["close_price_rank_zscore"]]
    res = res.merge(clean_forward_return, left_index=True, right_index=True)
    partial__get_ic = partial(_get_ic, target_zscore_name=target_zscore_name)
    return applyParallel(res.groupby("TradingDate"), partial__get_ic, 3)

def _get_ic(df, target_zscore_name):
    x_test = df.drop("TradingDate", axis=1).drop(target_zscore_name, axis=1).drop("close_price_rank_
    y_test = df[[target_zscore_name]]
    return_on_day = df[["close_price_rank_zscore"]]
    model = GLS(y_test, x_test)
    res = model.fit()
    # return_on_day_with_factor = return_on_day.merge(res.resid.to_frame(), left_index=True, right_i
    redid_rank = res.resid.to_frame().rank(ascending=True) / (res.resid.to_frame().shape[0] + 1)
    redid_rank.columns = ["resid_rank"]

    redid_rank["resid_inverse_normal"] = stats.norm.ppf(redid_rank["resid_rank"])
    redid_rank["resid_rank_zscore"] = (redid_rank["resid_inverse_normal"] - redid_rank["resid_invers
    ic = return_on_day["close_price_rank_zscore"].corr(redid_rank["resid_rank_zscore"], method='spea
    return pd.DataFrame({"Date": [df["TradingDate"].unique()[0]], "ic": [ic]})

X_, Y_ = get_x_y_for_ic(all_data, "pb_ratio_adjust")
ics = get_ics(X_, Y_, forward_return, "pb_ratio_adjust")
```

In [17]:

```
ics.head()
```

Out[17]:

	Date	ic
0	1990-12-31	-0.500000
0	1991-01-02	0.737865
0	1991-01-03	-0.200000
0	1991-01-04	-0.400000
0	1991-01-07	0.500000

In [18]:

```
ics.tail()
```

Out[18]:

	Date	ic
0	2016-05-05	0.186535
0	2016-05-06	0.169359
0	2016-05-09	0.069745
0	2016-05-10	-0.033411
0	2016-05-11	-0.023889

In [19]:

```
ics.to_csv("./pb_factor_ic.csv", index=False)
```

In [23]:

```
ics[ics["Date"] > "2016-01-01"].plot(x='Date', y='ic', figsize=(30, 20), kind="bar")
```

Out[23]:

<matplotlib.axes._subplots.AxesSubplot at 0x7fa1a6789190>