03_linear_regression

September 29, 2021

1 Stock Price Prediction using the Quantopian Trading Platform

The notebook linear_regression.ipynb contains examples for the prediction of stock prices using OLS with statsmodels and sklearn, as well as ridge and lasso models.

It is designed to run as a notebook on the Quantopian research platform and relies on the factor_library introduced in Chapter 4, Research and Evaluation of Alpha Factors.

1.1 How to run this notebook

This notebook is written for the Quantopian research environment. You can upload it after signing up and execute it on the Quantopian platform to gain access to the datasets.

1.2 Imports

[2]:
```
import pandas as pd
import numpy as np
from time import time
import talib
import re
from statsmodels.api import OLS
from sklearn.metrics import mean_squared_error
from scipy.stats import spearmanr
from sklearn.linear_model import LinearRegression, Ridge, RidgeCV, Lasso,␣
↪LassoCV, LogisticRegression
from sklearn.preprocessing import StandardScaler

from quantopian.research import run_pipeline
from quantopian.pipeline import Pipeline, factors, filters, classifiers
from quantopian.pipeline.data.builtin import USEquityPricing
from quantopian.pipeline.factors import (Latest,
                                         Returns,
                                         AverageDollarVolume,
                                         SimpleMovingAverage,
                                         EWMA,
                                         BollingerBands,
                                         CustomFactor,
                                         MarketCap,
                                        SimpleBeta)
```

```
from quantopian.pipeline.filters import QTradableStocksUS, StaticAssets
from quantopian.pipeline.data.quandl import fred_usdontd156n as libor
from empyrical import max_drawdown, sortino_ratio

import seaborn as sns
import matplotlib.pyplot as plt
```

1.3 Data Sources

[3]:
```
#################
# Fundamentals #
#################

# Morningstar fundamentals (2002 - Ongoing)
# https://www.quantopian.com/help/fundamentals
from quantopian.pipeline.data import Fundamentals

######################
# Analyst Estimates #
######################

# Earnings Surprises - Zacks (27 May 2006 - Ongoing)
# https://www.quantopian.com/data/zacks/earnings_surprises
from quantopian.pipeline.data.zacks import EarningsSurprises
from quantopian.pipeline.factors.zacks import␣
 ↪BusinessDaysSinceEarningsSurprisesAnnouncement

##########
# Events #
##########

# Buyback Announcements - EventVestor (01 Jun 2007 - Ongoing)
# https://www.quantopian.com/data/eventvestor/buyback_auth
from quantopian.pipeline.data.eventvestor import BuybackAuthorizations
from quantopian.pipeline.factors.eventvestor import BusinessDaysSinceBuybackAuth

# CEO Changes - EventVestor (01 Jan 2007 - Ongoing)
# https://www.quantopian.com/data/eventvestor/ceo_change
from quantopian.pipeline.data.eventvestor import CEOChangeAnnouncements

# Dividends - EventVestor (01 Jan 2007 - Ongoing)
# https://www.quantopian.com/data/eventvestor/dividends
from quantopian.pipeline.data.eventvestor import (
    DividendsByExDate,
    DividendsByPayDate,
    DividendsByAnnouncementDate,
)
```

```
from quantopian.pipeline.factors.eventvestor import (
    BusinessDaysSincePreviousExDate,
    BusinessDaysUntilNextExDate,
    BusinessDaysSinceDividendAnnouncement,
)

# Earnings Calendar - EventVestor (01 Jan 2007 - Ongoing)
# https://www.quantopian.com/data/eventvestor/earnings_calendar
from quantopian.pipeline.data.eventvestor import EarningsCalendar
from quantopian.pipeline.factors.eventvestor import (
    BusinessDaysUntilNextEarnings,
    BusinessDaysSincePreviousEarnings
)

# 13D Filings - EventVestor (01 Jan 2007 - Ongoing)
# https://www.quantopian.com/data/eventvestor/_13d_filings
from quantopian.pipeline.data.eventvestor import _13DFilings
from quantopian.pipeline.factors.eventvestor import␣
 ↪BusinessDaysSince13DFilingsDate

##############
# Sentiment #
##############

# News Sentiment - Sentdex Sentiment Analysis (15 Oct 2012 - Ongoing)
# https://www.quantopian.com/data/sentdex/sentiment
from quantopian.pipeline.data.sentdex import sentiment
```

1.4 Prepare the Data

We need to select a universe of equities and a time horizon, build and transform alpha factors that we will use as features, calculate forward returns that we aim to predict, and potentially clean our data.

1.4.1 Time horizon

[4]:
```
# trading days per period
MONTH = 21
YEAR = 12 * MONTH
```

[5]:
```
START = '2014-01-01'
END = '2015-12-31'
```

1.4.2 Universe

We will use equity data for the years 2014 and 2015 from a custom Q100US universe that uses built-in filters, factors, and classifiers to select the 100 stocks with the highest average dollar volume

of the last 200 trading days filtered by additional default criteria (see Quantopian docs linked on GitHub for detail). The universe dynamically updates based on the filter criteria so that, while there are 100 stocks at any given point, there may be more than 100 distinct equities in the sample:

[6]:
```
def Q100US():
    return filters.make_us_equity_universe(
        target_size=100,
        rankby=factors.AverageDollarVolume(window_length=200),
        mask=filters.default_us_equity_universe_mask(),
        groupby=classifiers.fundamentals.Sector(),
        max_group_weight=0.3,
        smoothing_func=lambda f: f.downsample('month_start'),
    )
```

[7]:
```
# UNIVERSE = StaticAssets(symbols(['MSFT', 'AAPL']))
UNIVERSE = Q100US()
```

1.4.3 Factor Transformations

[8]:
```
class AnnualizedData(CustomFactor):
    # Get the sum of the last 4 reported values
    window_length = 260

    def compute(self, today, assets, out, asof_date, values):
        for asset in range(len(assets)):
            # unique asof dates indicate availability of new figures
            _, filing_dates = np.unique(asof_date[:, asset], return_index=True)
            quarterly_values = values[filing_dates[-4:], asset]
            # ignore annual windows with <4 quarterly data points
            if len(~np.isnan(quarterly_values)) != 4:
                out[asset] = np.nan
            else:
                out[asset] = np.sum(quarterly_values)
```

[9]:
```
class AnnualAvg(CustomFactor):
    window_length = 252

    def compute(self, today, assets, out, values):
        out[:] = (values[0] + values[-1])/2
```

[10]:
```
def factor_pipeline(factors):
    start = time()
    pipe = Pipeline({k: v(mask=UNIVERSE).rank() for k, v in factors.items()},
                    screen=UNIVERSE)
    result = run_pipeline(pipe, start_date=START, end_date=END)
    return result, time() - start
```

1.5 Factor Library

1.5.1 Value Factors

[11]:
```
class ValueFactors:
    """Definitions of factors for cross-sectional trading algorithms"""

    @staticmethod
    def PriceToSalesTTM(**kwargs):
        """Last closing price divided by sales per share"""
        return Fundamentals.ps_ratio.latest

    @staticmethod
    def PriceToEarningsTTM(**kwargs):
        """Closing price divided by earnings per share (EPS)"""
        return Fundamentals.pe_ratio.latest

    @staticmethod
    def PriceToDilutedEarningsTTM(mask):
        """Closing price divided by diluted EPS"""
        last_close = USEquityPricing.close.latest
        diluted_eps = AnnualizedData(inputs = [Fundamentals.
↪diluted_eps_earnings_reports_asof_date,
                                               Fundamentals.
↪diluted_eps_earnings_reports],
                                     mask=mask)
        return last_close / diluted_eps

    @staticmethod
    def PriceToForwardEarnings(**kwargs):
        """Price to Forward Earnings"""
        return Fundamentals.forward_pe_ratio.latest

    @staticmethod
    def DividendYield(**kwargs):
        """Dividends per share divided by closing price"""
        return Fundamentals.trailing_dividend_yield.latest

    @staticmethod
    def PriceToFCF(mask):
        """Price to Free Cash Flow"""
        last_close = USEquityPricing.close.latest
        fcf_share = AnnualizedData(inputs = [Fundamentals.
↪fcf_per_share_asof_date,
                                             Fundamentals.fcf_per_share],
                                   mask=mask)
        return last_close / fcf_share
```

```
    @staticmethod
    def PriceToOperatingCashflow(mask):
        """Last Close divided by Operating Cash Flows"""
        last_close = USEquityPricing.close.latest
        cfo_per_share = AnnualizedData(inputs = [Fundamentals.
↪cfo_per_share_asof_date,
                                                 Fundamentals.cfo_per_share],
                                       mask=mask)
        return last_close / cfo_per_share

    @staticmethod
    def PriceToBook(mask):
        """Closing price divided by book value"""
        last_close = USEquityPricing.close.latest
        book_value_per_share = AnnualizedData(inputs = [Fundamentals.
↪book_value_per_share_asof_date,
                                                        Fundamentals.
↪book_value_per_share],
                                              mask=mask)
        return last_close / book_value_per_share


    @staticmethod
    def EVToFCF(mask):
        """Enterprise Value divided by Free Cash Flows"""
        fcf = AnnualizedData(inputs = [Fundamentals.free_cash_flow_asof_date,
                                       Fundamentals.free_cash_flow],
                             mask=mask)
        return Fundamentals.enterprise_value.latest / fcf

    @staticmethod
    def EVToEBITDA(mask):
        """Enterprise Value to Earnings Before Interest, Taxes, Deprecation and␣
↪Amortization (EBITDA)"""
        ebitda = AnnualizedData(inputs = [Fundamentals.ebitda_asof_date,
                                          Fundamentals.ebitda],
                                mask=mask)

        return Fundamentals.enterprise_value.latest / ebitda

    @staticmethod
    def EBITDAYield(mask):
        """EBITDA divided by latest close"""
        ebitda = AnnualizedData(inputs = [Fundamentals.ebitda_asof_date,
                                          Fundamentals.ebitda],
                                mask=mask)
        return USEquityPricing.close.latest / ebitda
```

[12]:
```
VALUE_FACTORS = {
    'DividendYield'            : ValueFactors.DividendYield,
    'EBITDAYield'              : ValueFactors.EBITDAYield,
    'EVToEBITDA'               : ValueFactors.EVToEBITDA,
    'EVToFCF'                  : ValueFactors.EVToFCF,
    'PriceToBook'              : ValueFactors.PriceToBook,
    'PriceToDilutedEarningsTTM': ValueFactors.PriceToDilutedEarningsTTM,
    'PriceToEarningsTTM'       : ValueFactors.PriceToEarningsTTM,
    'PriceToFCF'               : ValueFactors.PriceToFCF,
    'PriceToForwardEarnings'   : ValueFactors.PriceToForwardEarnings,
    'PriceToOperatingCashflow' : ValueFactors.PriceToOperatingCashflow,
    'PriceToSalesTTM'          : ValueFactors.PriceToSalesTTM,
}
```

[13]:
```
value_factors, t = factor_pipeline(VALUE_FACTORS)
print('Pipeline run time {:.2f} secs'.format(t))
value_factors.info()
```

```
/usr/local/lib/python2.7/dist-packages/numpy/lib/arraysetops.py:200:
FutureWarning: In the future, NAT != NAT will be True rather than False.
  flag = np.concatenate(([True], aux[1:] != aux[:-1]))

Pipeline run time 91.43 secs
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 50362 entries, (2014-01-02 00:00:00+00:00, Equity(24 [AAPL])) to
(2015-12-31 00:00:00+00:00, Equity(47208 [GPRO]))
Data columns (total 11 columns):
DividendYield                40772 non-null float64
EBITDAYield                  49823 non-null float64
EVToEBITDA                   49823 non-null float64
EVToFCF                      46400 non-null float64
PriceToBook                  50343 non-null float64
PriceToDilutedEarningsTTM    50215 non-null float64
PriceToEarningsTTM           48956 non-null float64
PriceToFCF                   49133 non-null float64
PriceToForwardEarnings       39607 non-null float64
PriceToOperatingCashflow     50343 non-null float64
PriceToSalesTTM              50362 non-null float64
dtypes: float64(11)
memory usage: 4.6+ MB
```

1.5.2 Momentum

[14]:
```
class MomentumFactors:
    """Custom Momentum Factors"""
    class PercentAboveLow(CustomFactor):
        """Percentage of current close above low
        in lookback window of window_length days
```

```
        """
        inputs = [USEquityPricing.close]
        window_length = 252

        def compute(self, today, assets, out, close):
            out[:] = close[-1] / np.min(close, axis=0) - 1

    class PercentBelowHigh(CustomFactor):
        """Percentage of current close below high
        in lookback window of window_length days
        """

        inputs = [USEquityPricing.close]
        window_length = 252

        def compute(self, today, assets, out, close):
            out[:] = close[-1] / np.max(close, axis=0) - 1

    @staticmethod
    def make_dx(timeperiod=14):
        class DX(CustomFactor):
            """Directional Movement Index"""
            inputs = [USEquityPricing.high,
                      USEquityPricing.low,
                      USEquityPricing.close]
            window_length = timeperiod + 1

            def compute(self, today, assets, out, high, low, close):
                out[:] = [talib.DX(high[:, i],
                                   low[:, i],
                                   close[:, i],
                                   timeperiod=timeperiod)[-1]
                          for i in range(len(assets))]
        return DX

    @staticmethod
    def make_mfi(timeperiod=14):
        class MFI(CustomFactor):
            """Money Flow Index"""
            inputs = [USEquityPricing.high,
                      USEquityPricing.low,
                      USEquityPricing.close,
                      USEquityPricing.volume]
            window_length = timeperiod + 1

            def compute(self, today, assets, out, high, low, close, vol):
                out[:] = [talib.MFI(high[:, i],
```

```
                                    low[:, i],
                                    close[:, i],
                                    vol[:, i],
                                    timeperiod=timeperiod)[-1]
                            for i in range(len(assets))]
        return MFI

    @staticmethod
    def make_oscillator(fastperiod=12, slowperiod=26, matype=0):
        class PPO(CustomFactor):
            """12/26-Day Percent Price Oscillator"""
            inputs = [USEquityPricing.close]
            window_length = slowperiod

            def compute(self, today, assets, out, close_prices):
                out[:] = [talib.PPO(close,
                                    fastperiod=fastperiod,
                                    slowperiod=slowperiod,
                                    matype=matype)[-1]
                          for close in close_prices.T]
        return PPO

    @staticmethod
    def make_stochastic_oscillator(fastk_period=5, slowk_period=3,␣
↪slowd_period=3,
                                   slowk_matype=0, slowd_matype=0):
        class StochasticOscillator(CustomFactor):
            """20-day Stochastic Oscillator """
            inputs = [USEquityPricing.high,
                      USEquityPricing.low,
                      USEquityPricing.close]
            outputs = ['slowk', 'slowd']
            window_length = fastk_period * 2

            def compute(self, today, assets, out, high, low, close):
                slowk, slowd = [talib.STOCH(high[:, i],
                                            low[:, i],
                                            close[:, i],
                                            fastk_period=fastk_period,
                                            slowk_period=slowk_period,
                                            slowk_matype=slowk_matype,
                                            slowd_period=slowd_period,
                                            slowd_matype=slowd_matype)[-1]
                                for i in range(len(assets))]

                out.slowk[:], out.slowd[:] = slowk[-1], slowd[-1]
        return StochasticOscillator
```

```
    @staticmethod
    def make_trendline(timeperiod=252):
        class Trendline(CustomFactor):
            inputs = [USEquityPricing.close]
            """52-Week Trendline"""
            window_length = timeperiod

            def compute(self, today, assets, out, close_prices):
                out[:] = [talib.LINEARREG_SLOPE(close,
                                                timeperiod=timeperiod)[-1]
                          for close in close_prices.T]
        return Trendline
```

[15]:
```
MOMENTUM_FACTORS = {
    'Percent Above Low'            : MomentumFactors.PercentAboveLow,
    'Percent Below High'           : MomentumFactors.PercentBelowHigh,
    'Price Oscillator'             : MomentumFactors.make_oscillator(),
    'Money Flow Index'             : MomentumFactors.make_mfi(),
    'Directional Movement Index'   : MomentumFactors.make_dx(),
    'Trendline'                    : MomentumFactors.make_trendline()
}
```

[16]:
```
momentum_factors, t = factor_pipeline(MOMENTUM_FACTORS)
print('Pipeline run time {:.2f} secs'.format(t))
momentum_factors.info()
```

```
Pipeline run time 20.68 secs
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 50362 entries, (2014-01-02 00:00:00+00:00, Equity(24 [AAPL])) to
(2015-12-31 00:00:00+00:00, Equity(47208 [GPRO]))
Data columns (total 6 columns):
Directional Movement Index    50362 non-null float64
Money Flow Index              50362 non-null float64
Percent Above Low             49536 non-null float64
Percent Below High            49536 non-null float64
Price Oscillator              50355 non-null float64
Trendline                     49536 non-null float64
dtypes: float64(6)
memory usage: 2.7+ MB
```

1.5.3 Efficiency Factors

[17]:
```
class EfficiencyFactors:

    @staticmethod
    def CapexToAssets(mask):
```

```
        """Capital Expenditure divided by Total Assets"""
        capex = AnnualizedData(inputs = [Fundamentals.
↪capital_expenditure_asof_date,
                                         Fundamentals.capital_expenditure],
                               mask=mask)
        assets = Fundamentals.total_assets.latest
        return - capex / assets

    @staticmethod
    def CapexToSales(mask):
        """Capital Expenditure divided by Total Revenue"""
        capex = AnnualizedData(inputs = [Fundamentals.
↪capital_expenditure_asof_date,
                                         Fundamentals.capital_expenditure],
                               mask=mask)
        revenue = AnnualizedData(inputs = [Fundamentals.total_revenue_asof_date,
                                         Fundamentals.total_revenue],
                               mask=mask)
        return - capex / revenue

    @staticmethod
    def CapexToFCF(mask):
        """Capital Expenditure divided by Free Cash Flows"""
        capex = AnnualizedData(inputs = [Fundamentals.
↪capital_expenditure_asof_date,
                                         Fundamentals.capital_expenditure],
                               mask=mask)
        free_cash_flow = AnnualizedData(inputs = [Fundamentals.
↪free_cash_flow_asof_date,
                                         Fundamentals.free_cash_flow],
                               mask=mask)
        return - capex / free_cash_flow

    @staticmethod
    def EBITToAssets(mask):
        """Earnings Before Interest and Taxes (EBIT) divided by Total Assets"""
        ebit = AnnualizedData(inputs = [Fundamentals.ebit_asof_date,
                                         Fundamentals.ebit],
                               mask=mask)
        assets = Fundamentals.total_assets.latest
        return ebit / assets

    @staticmethod
    def CFOToAssets(mask):
        """Operating Cash Flows divided by Total Assets"""
        cfo = AnnualizedData(inputs = [Fundamentals.
↪operating_cash_flow_asof_date,
```

```
                                        Fundamentals.operating_cash_flow],
                                    mask=mask)
        assets = Fundamentals.total_assets.latest
        return cfo / assets

    @staticmethod
    def RetainedEarningsToAssets(mask):
        """Retained Earnings divided by Total Assets"""
        retained_earnings = AnnualizedData(inputs = [Fundamentals.
↪retained_earnings_asof_date,
                                        Fundamentals.retained_earnings],
                                    mask=mask)
        assets = Fundamentals.total_assets.latest
        return retained_earnings / assets
```

[18]:
```
EFFICIENCY_FACTORS = {
    'CFO To Assets' :EfficiencyFactors.CFOToAssets,
    'Capex To Assets' :EfficiencyFactors.CapexToAssets,
    'Capex To FCF' :EfficiencyFactors.CapexToFCF,
    'Capex To Sales' :EfficiencyFactors.CapexToSales,
    'EBIT To Assets' :EfficiencyFactors.EBITToAssets,
    'Retained Earnings To Assets' :EfficiencyFactors.RetainedEarningsToAssets
    }
```

[19]:
```
efficiency_factors, t = factor_pipeline(EFFICIENCY_FACTORS)
print('Pipeline run time {:.2f} secs'.format(t))
efficiency_factors.info()
```

```
Pipeline run time 38.54 secs
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 50362 entries, (2014-01-02 00:00:00+00:00, Equity(24 [AAPL])) to
(2015-12-31 00:00:00+00:00, Equity(47208 [GPRO]))
Data columns (total 6 columns):
CFO To Assets                  50351 non-null float64
Capex To Assets                46997 non-null float64
Capex To FCF                   45799 non-null float64
Capex To Sales                 46997 non-null float64
EBIT To Assets                 46635 non-null float64
Retained Earnings To Assets    50349 non-null float64
dtypes: float64(6)
memory usage: 2.7+ MB
```

1.5.4 Risk Factors

[20]:
```
class RiskFactors:

    @staticmethod
    def LogMarketCap(mask):
        """Log of Market Capitalization log(Close Price * Shares Outstanding)"""
        return np.log(MarketCap(mask=mask))

    class DownsideRisk(CustomFactor):
        """Mean returns divided by std of 1yr daily losses (Sortino Ratio)"""
        inputs = [USEquityPricing.close]
        window_length = 252

        def compute(self, today, assets, out, close):
            ret = pd.DataFrame(close).pct_change()
            out[:] = ret.mean().div(ret.where(ret<0).std())

    @staticmethod
    def MarketBeta(**kwargs):
        """Slope of 1-yr regression of price returns against index returns"""
        return SimpleBeta(target=symbols('SPY'), regression_length=252)

    class DownsideBeta(CustomFactor):
        """Slope of 1yr regression of returns on negative index returns"""
        inputs = [USEquityPricing.close]
        window_length = 252

        def compute(self, today, assets, out, close):
            t = len(close)
            assets = pd.DataFrame(close).pct_change()

            start_date = (today - pd.DateOffset(years=1)).strftime('%Y-%m-%d')
            spy = get_pricing('SPY',
                              start_date=start_date,
                              end_date=today.strftime('%Y-%m-%d')).
↪reset_index(drop=True)
            spy_neg_ret = (spy
                           .close_price
                           .iloc[-t:]
                           .pct_change()
                           .pipe(lambda x: x.where(x<0)))

            out[:] = assets.apply(lambda x: x.cov(spy_neg_ret)).div(spy_neg_ret.
↪var())

    class Vol3M(CustomFactor):
```

```
        """3-month Volatility: Standard deviation of returns over 3 months"""

        inputs = [USEquityPricing.close]
        window_length = 63

        def compute(self, today, assets, out, close):
            out[:] = np.log1p(pd.DataFrame(close).pct_change()).std()
```

[21]:
```
RISK_FACTORS = {
    'Log Market Cap' : RiskFactors.LogMarketCap,
    'Downside Risk'  : RiskFactors.DownsideRisk,
    'Index Beta'     : RiskFactors.MarketBeta,
#     'Downside Beta'  : RiskFactors.DownsideBeta,
    'Volatility 3M'  : RiskFactors.Vol3M,
}
```

[22]:
```
risk_factors, t = factor_pipeline(RISK_FACTORS)
print('Pipeline run time {:.2f} secs'.format(t))
risk_factors.info()
```

```
Pipeline run time 46.76 secs
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 50362 entries, (2014-01-02 00:00:00+00:00, Equity(24 [AAPL])) to
(2015-12-31 00:00:00+00:00, Equity(47208 [GPRO]))
Data columns (total 4 columns):
Downside Risk     50362 non-null float64
Index Beta        50079 non-null float64
Log Market Cap    50362 non-null float64
Volatility 3M     50362 non-null float64
dtypes: float64(4)
memory usage: 1.9+ MB
```

1.5.5 Growth Factors

[23]:
```
def growth_pipeline():
    revenue = AnnualizedData(inputs = [Fundamentals.total_revenue_asof_date,
                                       Fundamentals.total_revenue],
                             mask=UNIVERSE)
    eps = AnnualizedData(inputs = [Fundamentals.
 ↪diluted_eps_earnings_reports_asof_date,
                                       Fundamentals.
 ↪diluted_eps_earnings_reports],
                             mask=UNIVERSE)

    return Pipeline({'Sales': revenue,
                     'EPS': eps,
                     'Total Assets': Fundamentals.total_assets.latest,
```

```
                     'Net Debt': Fundamentals.net_debt.latest},
                    screen=UNIVERSE)
```

[24]:
```
start_timer = time()
growth_factors = run_pipeline(growth_pipeline(), start_date=START, end_date=END)

for col in growth_result.columns:
    for month in [3, 12]:
        new_col = col + ' Growth {}M'.format(month)
        kwargs = {new_col: growth_factors[col].pct_change(month*MONTH).
 ↪groupby(level=1).rank()}
        growth_factors = growth_factors.assign(**kwargs)
print('Pipeline run time {:.2f} secs'.format(time() - start_timer))
growth_factors.info()
```

```
Pipeline run time 23.05 secs
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 50362 entries, (2014-01-02 00:00:00+00:00, Equity(24 [AAPL])) to
(2015-12-31 00:00:00+00:00, Equity(47208 [GPRO]))
Data columns (total 12 columns):
EPS                         50215 non-null float64
Net Debt                    47413 non-null float64
Sales                       50351 non-null float64
Total Assets                50362 non-null float64
EPS Growth 3M               50152 non-null float64
EPS Growth 12M              49963 non-null float64
Net Debt Growth 3M          47350 non-null float64
Net Debt Growth 12M         47171 non-null float64
Sales Growth 3M             50288 non-null float64
Sales Growth 12M            50099 non-null float64
Total Assets Growth 3M      50299 non-null float64
Total Assets Growth 12M     50110 non-null float64
dtypes: float64(12)
memory usage: 5.0+ MB
```

1.5.6 Quality Factors

[25]:
```
class QualityFactors:

    @staticmethod
    def AssetTurnover(mask):
        """Sales divided by average of year beginning and year end assets"""

        assets = AnnualAvg(inputs=[Fundamentals.total_assets],
                           mask=mask)
        sales = AnnualizedData([Fundamentals.total_revenue_asof_date,
                                Fundamentals.total_revenue], mask=mask)
```

```
        return sales / assets

    @staticmethod
    def CurrentRatio(mask):
        """Total current assets divided by total current liabilities"""

        assets = Fundamentals.current_assets.latest
        liabilities = Fundamentals.current_liabilities.latest
        return assets / liabilities

    @staticmethod
    def AssetToEquityRatio(mask):
        """Total current assets divided by common equity"""

        assets = Fundamentals.current_assets.latest
        equity = Fundamentals.common_stock.latest
        return assets / equity


    @staticmethod
    def InterestCoverage(mask):
        """EBIT divided by interest expense"""

        ebit = AnnualizedData(inputs = [Fundamentals.ebit_asof_date,
                                        Fundamentals.ebit], mask=mask)

        interest_expense = AnnualizedData(inputs = [Fundamentals.
↪interest_expense_asof_date,
                                        Fundamentals.interest_expense],␣
↪mask=mask)
        return ebit / interest_expense

    @staticmethod
    def DebtToAssetRatio(mask):
        """Total Debts divided by Total Assets"""

        debt = Fundamentals.total_debt.latest
        assets = Fundamentals.total_assets.latest
        return debt / assets

    @staticmethod
    def DebtToEquityRatio(mask):
        """Total Debts divided by Common Stock Equity"""

        debt = Fundamentals.total_debt.latest
        equity = Fundamentals.common_stock.latest
        return debt / equity
```

```
    @staticmethod
    def WorkingCapitalToAssets(mask):
        """Current Assets less Current liabilities (Working Capital) divided by␣
↪Assets"""

        working_capital = Fundamentals.working_capital.latest
        assets = Fundamentals.total_assets.latest
        return working_capital / assets

    @staticmethod
    def WorkingCapitalToSales(mask):
        """Current Assets less Current liabilities (Working Capital), divided␣
↪by Sales"""

        working_capital = Fundamentals.working_capital.latest
        sales = AnnualizedData([Fundamentals.total_revenue_asof_date,
                                Fundamentals.total_revenue], mask=mask)
        return working_capital / sales


    class MertonsDD(CustomFactor):
        """Merton's Distance to Default """

        inputs = [Fundamentals.total_assets,
                  Fundamentals.total_liabilities,
                  libor.value,
                  USEquityPricing.close]
        window_length = 252

        def compute(self, today, assets, out, tot_assets, tot_liabilities, r,␣
↪close):
            mertons = []

            for col_assets, col_liabilities, col_r, col_close in zip(tot_assets.
↪T, tot_liabilities.T,
                                                                     r.T, close.
↪T):
                vol_1y = np.nanstd(col_close)
                numerator = np.log(
                        col_assets[-1] / col_liabilities[-1]) + ((252 *␣
↪col_r[-1]) - ((vol_1y ** 2) / 2))
                mertons.append(numerator / vol_1y)

            out[:] = mertons
```

[26]:
```
QUALITY_FACTORS = {
    'AssetToEquityRatio'    : QualityFactors.AssetToEquityRatio,
    'AssetTurnover'         : QualityFactors.AssetTurnover,
    'CurrentRatio'          : QualityFactors.CurrentRatio,
    'DebtToAssetRatio'      : QualityFactors.DebtToAssetRatio,
    'DebtToEquityRatio'     : QualityFactors.DebtToEquityRatio,
    'InterestCoverage'      : QualityFactors.InterestCoverage,
    'MertonsDD'             : QualityFactors.MertonsDD,
    'WorkingCapitalToAssets': QualityFactors.WorkingCapitalToAssets,
    'WorkingCapitalToSales' : QualityFactors.WorkingCapitalToSales,
}
```

[27]:
```
quality_factors, t = factor_pipeline(QUALITY_FACTORS)
print('Pipeline run time {:.2f} secs'.format(t))
quality_factors.info()
```

```
Pipeline run time 37.24 secs
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 50362 entries, (2014-01-02 00:00:00+00:00, Equity(24 [AAPL])) to
(2015-12-31 00:00:00+00:00, Equity(47208 [GPRO]))
Data columns (total 9 columns):
AssetToEquityRatio        45176 non-null float64
AssetTurnover             50314 non-null float64
CurrentRatio              45680 non-null float64
DebtToAssetRatio          50080 non-null float64
DebtToEquityRatio         48492 non-null float64
InterestCoverage          35250 non-null float64
MertonsDD                 50362 non-null float64
WorkingCapitalToAssets    45680 non-null float64
WorkingCapitalToSales     45669 non-null float64
dtypes: float64(9)
memory usage: 3.8+ MB
```

1.5.7 Payout Factors

[28]:
```
class PayoutFactors:

    @staticmethod
    def DividendPayoutRatio(mask):
        """Dividends Per Share divided by Earnings Per Share"""

        dps = AnnualizedData(inputs = [Fundamentals.
 ↪dividend_per_share_earnings_reports_asof_date,
                                       Fundamentals.
 ↪dividend_per_share_earnings_reports], mask=mask)
```

```
        eps = AnnualizedData(inputs = [Fundamentals.
↪basic_eps_earnings_reports_asof_date,
                                       Fundamentals.
↪basic_eps_earnings_reports], mask=mask)
        return dps / eps

    @staticmethod
    def DividendGrowth(**kwargs):
        """Annualized percentage DPS change"""
        return Fundamentals.dps_growth.latest
```

[29]:
```
PAYOUT_FACTORS = {
    'Dividend Payout Ratio': PayoutFactors.DividendPayoutRatio,
    'Dividend Growth': PayoutFactors.DividendGrowth
}
```

[30]:
```
payout_factors, t = factor_pipeline(PAYOUT_FACTORS)
print('Pipeline run time {:.2f} secs'.format(t))
payout_factors.info()
```

```
Pipeline run time 22.46 secs
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 50362 entries, (2014-01-02 00:00:00+00:00, Equity(24 [AAPL])) to
(2015-12-31 00:00:00+00:00, Equity(47208 [GPRO]))
Data columns (total 2 columns):
Dividend Growth          40517 non-null float64
Dividend Payout Ratio    39947 non-null float64
dtypes: float64(2)
memory usage: 1.2+ MB
```

1.5.8 Profitability Factors

[31]:
```
class ProfitabilityFactors:

    @staticmethod
    def GrossProfitMargin(mask):
        """Gross Profit divided by Net Sales"""

        gross_profit = AnnualizedData([Fundamentals.gross_profit_asof_date,
                                       Fundamentals.gross_profit], mask=mask)
        sales = AnnualizedData([Fundamentals.total_revenue_asof_date,
                                Fundamentals.total_revenue], mask=mask)
        return gross_profit / sales

    @staticmethod
    def NetIncomeMargin(mask):
        """Net income divided by Net Sales"""
```

```
        net_income = AnnualizedData([Fundamentals.
↪net_income_income_statement_asof_date,
                                     Fundamentals.net_income_income_statement],␣
↪mask=mask)
        sales = AnnualizedData([Fundamentals.total_revenue_asof_date,
                                  Fundamentals.total_revenue], mask=mask)
        return net_income / sales
```

[32]:
```
PROFITABIILTY_FACTORS = {
    'Gross Profit Margin': ProfitabilityFactors.GrossProfitMargin,
    'Net Income Margin': ProfitabilityFactors.NetIncomeMargin,
    'Return on Equity': Fundamentals.roe.latest,
    'Return on Assets': Fundamentals.roa.latest,
    'Return on Invested Capital': Fundamentals.roic.latest
}
```

[33]:
```
profitability_factors, t = factor_pipeline(PAYOUT_FACTORS)
print('Pipeline run time {:.2f} secs'.format(t))
payout_factors.info()
```

```
Pipeline run time 22.56 secs
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 50362 entries, (2014-01-02 00:00:00+00:00, Equity(24 [AAPL])) to
(2015-12-31 00:00:00+00:00, Equity(47208 [GPRO]))
Data columns (total 2 columns):
Dividend Growth          40517 non-null float64
Dividend Payout Ratio    39947 non-null float64
dtypes: float64(2)
memory usage: 1.2+ MB
```

[34]:
```
# profitability_pipeline().show_graph(format='png')
```

1.6 Build Dataset

1.6.1 Get Returns

We will test predictions for various lookahead periods to identify the best holding periods that generate the best predictability, measured by the information coefficient.

More specifically, we compute returns for 1, 5, 10, and 20 days using the built-in Returns function, resulting in over 50,000 observations for the universe of 100 stocks over two years (that include approximately 252 trading days each)

[35]:
```
lookahead = [1, 5, 10, 20]
returns = run_pipeline(Pipeline({'Returns{}D'.format(i):␣
↪Returns(inputs=[USEquityPricing.close],
```

```
                                                window_length=i+1, mask=UNIVERSE) for␣
↪i in lookahead},
                                    screen=UNIVERSE),
                         start_date=START,
                         end_date=END)
return_cols = ['Returns{}D'.format(i) for i in lookahead]
returns.info()
```

```
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 50362 entries, (2014-01-02 00:00:00+00:00, Equity(24 [AAPL])) to
(2015-12-31 00:00:00+00:00, Equity(47208 [GPRO]))
Data columns (total 4 columns):
Returns10D    50362 non-null float64
Returns1D     50362 non-null float64
Returns20D    50360 non-null float64
Returns5D     50362 non-null float64
dtypes: float64(4)
memory usage: 1.9+ MB
```

We will use over 50 features that cover a broad range of factors based on market, fundamental, and alternative data. The notebook also includes custom transformations to convert fundamental data that is typically available in quarterly reporting frequency to rolling annual totals or averages to avoid excessive season fluctuations.

Once the factors have been computed through the various pipelines outlined in Chapter 4, Alpha Factors – Research and Evaluation, we combine them using pd.concat(), assign index names, and create a categorical variable that identifies the asset for each data point:

[36]:
```
data = pd.concat([returns,
                  value_factors,
                  momentum_factors,
                  quality_factors,
                  payout_factors,
                  growth_factors,
                  efficiency_factors,
                  risk_factors], axis=1).sortlevel()
data.index.names = ['date', 'asset']
```

[37]:
```
data['stock'] = data.index.get_level_values('asset').map(lambda x: x.asset_name)
```

1.7 Remove columns and rows with less than 80% of data availability

In a next step, we remove rows and columns that lack more than 20 percent of the observations, resulting in a loss of six percent of the observations and three columns:

[38]:
```
rows_before, cols_before = data.shape
data = (data
        .dropna(axis=1, thresh=int(len(data)*.8))
```

```
        .dropna(thresh=int(len(data.columns) * .8)))
data = data.fillna(data.median())
rows_after, cols_after = data.shape
print('{:,d} rows and {:,d} columns dropped'.format(rows_before-rows_after,␣
 ↪cols_before-cols_after))
```

```
2,985 rows and 3 columns dropped
```

At this point, we have 51 features and the categorical identifier of the stock:

[39]:
```
data.sort_index(1).info()
```

```
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 47377 entries, (2014-01-02 00:00:00+00:00, Equity(24 [AAPL])) to
(2015-12-31 00:00:00+00:00, Equity(47208 [GPRO]))
Data columns (total 52 columns):
AssetToEquityRatio              47377 non-null float64
AssetTurnover                   47377 non-null float64
CFO To Assets                   47377 non-null float64
Capex To Assets                 47377 non-null float64
Capex To FCF                    47377 non-null float64
Capex To Sales                  47377 non-null float64
CurrentRatio                    47377 non-null float64
DebtToAssetRatio                47377 non-null float64
DebtToEquityRatio               47377 non-null float64
Directional Movement Index      47377 non-null float64
Dividend Growth                 47377 non-null float64
DividendYield                   47377 non-null float64
Downside Risk                   47377 non-null float64
EBIT To Assets                  47377 non-null float64
EBITDAYield                     47377 non-null float64
EPS                             47377 non-null float64
EPS Growth 12M                  47377 non-null float64
EPS Growth 3M                   47377 non-null float64
EVToEBITDA                      47377 non-null float64
EVToFCF                         47377 non-null float64
Index Beta                      47377 non-null float64
Log Market Cap                  47377 non-null float64
MertonsDD                       47377 non-null float64
Money Flow Index                47377 non-null float64
Net Debt                        47377 non-null float64
Net Debt Growth 12M             47377 non-null float64
Net Debt Growth 3M              47377 non-null float64
Percent Above Low               47377 non-null float64
Percent Below High              47377 non-null float64
Price Oscillator                47377 non-null float64
PriceToBook                     47377 non-null float64
PriceToDilutedEarningsTTM       47377 non-null float64
```

```
PriceToEarningsTTM              47377 non-null float64
PriceToFCF                      47377 non-null float64
PriceToOperatingCashflow        47377 non-null float64
PriceToSalesTTM                 47377 non-null float64
Retained Earnings To Assets     47377 non-null float64
Returns10D                      47377 non-null float64
Returns1D                       47377 non-null float64
Returns20D                      47377 non-null float64
Returns5D                       47377 non-null float64
Sales                           47377 non-null float64
Sales Growth 12M                47377 non-null float64
Sales Growth 3M                 47377 non-null float64
Total Assets                    47377 non-null float64
Total Assets Growth 12M         47377 non-null float64
Total Assets Growth 3M          47377 non-null float64
Trendline                       47377 non-null float64
Volatility 3M                   47377 non-null float64
WorkingCapitalToAssets          47377 non-null float64
WorkingCapitalToSales           47377 non-null float64
stock                           47377 non-null object
dtypes: float64(51), object(1)
memory usage: 19.2+ MB
```

1.8 Data Exploration

For linear regression models, it is important to explore the correlation among the features to identify multicollinearity issues, and to check the correlation between the features and the target. The notebook contains a seaborn clustermap that shows the hierarchical structure of the feature correlation matrix. It identifies a small number of highly correlated clusters.

[ ]:
```
g = sns.clustermap(data.drop(['stock'] + return_cols, axis=1).corr())
plt.gcf().set_size_inches((14,14));
```

1.9 Dummy encoding of categorical variables

We need to convert the categorical stock variable into a numeric format so that the linear regression can process it. For this purpose, we use dummy encoding that creates individual columns for each category level and flags the presence of this level in the original categorical column with an entry of 1, and 0 otherwise. The pandas function get_dummies() automates dummy encoding. It detects and properly converts columns of type objects as illustrated next. If you need dummy variables for columns containing integers, for instance, you can identify them using the keyword columns:

[41]:
```
X = pd.get_dummies(data.drop(return_cols, axis=1))
X.info()
```

```
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 47377 entries, (2014-01-02 00:00:00+00:00, Equity(24 [AAPL])) to
(2015-12-31 00:00:00+00:00, Equity(47208 [GPRO]))
```

```
Columns: 182 entries, DividendYield to stock_YELP INC
dtypes: float64(182)
memory usage: 66.1+ MB
```

1.10 Creating forward returns

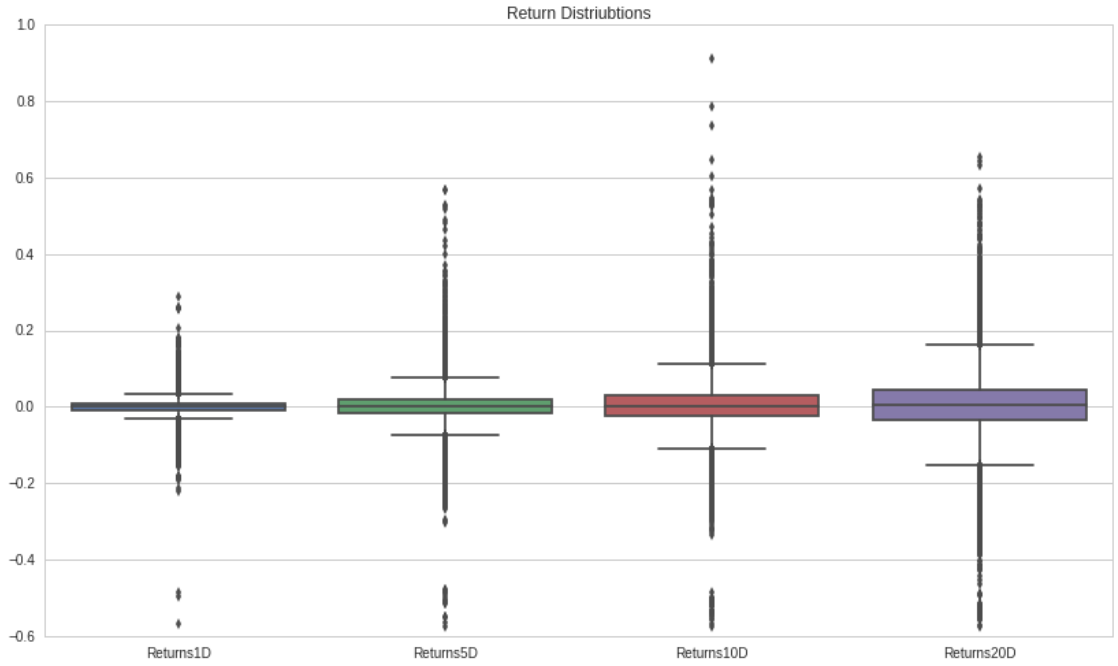
The goal is to predict returns over a given holding period. Hence, we need to align the features with return values with the corresponding return data point 1, 5, 10, or 20 days into the future for each equity. We achieve this by combining the pandas .groupby() method with the .shift() method as follows:

[42]:
```
y = data.loc[:, return_cols]
shifted_y = []
for col in y.columns:
    t = int(re.search(r'\d+', col).group(0))
    shifted_y.append(y.groupby(level='asset')['Returns{}D'.format(t)].shift(-t).
 ↪to_frame(col))
y = pd.concat(shifted_y, axis=1)
y.info()
```

```
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 47377 entries, (2014-01-02 00:00:00+00:00, Equity(24 [AAPL])) to
(2015-12-31 00:00:00+00:00, Equity(47208 [GPRO]))
Data columns (total 4 columns):
Returns1D     47242 non-null float64
Returns5D     46706 non-null float64
Returns10D    46036 non-null float64
Returns20D    44696 non-null float64
dtypes: float64(4)
memory usage: 1.8+ MB
```

[43]:
```
ax = sns.boxplot(y[return_cols])
ax.set_title('Return Distriubtions');
```

```
/usr/local/lib/python2.7/dist-packages/seaborn/categorical.py:2171: UserWarning:
The boxplot API has been changed. Attempting to adjust your arguments for the
new API (which might not work). Please update your code. See the version 0.6
release notes for more info.
  warnings.warn(msg, UserWarning)
```



1.11 Linear Regression

1.11.1 Statsmodels

We can estimate a linear regression model using OLS with statsmodels as demonstrated previously. We select a forward return, for example for a 10-day holding period, remove outliers below the 2.5% and above the 97.5% percentiles, and fit the model accordingly:

[44]:
```
target = 'Returns1D'
model_data = pd.concat([y[[target]], X], axis=1).dropna()
model_data = model_data[model_data[target].between(model_data[target].quantile(.
↪025),
                                                   model_data[target].quantile(.
↪975))]

model = OLS(endog=model_data[target], exog=model_data.drop(target, axis=1))
trained_model = model.fit()
trained_model.summary()
```

[44]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:              Returns1D   R-squared:                       0.004
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     1.084
```

```
Date:                Sun, 09 Sep 2018   Prob (F-statistic):              0.214
Time:                        21:17:58   Log-Likelihood:             1.3306e+05
No. Observations:               44878   AIC:                        -2.658e+05
Df Residuals:                   44703   BIC:                        -2.642e+05
Df Model:                         174
Covariance Type:            nonrobust
================================================================================
================================
                                              coef    std err          t
P>|t|      [95.0% Conf. Int.]
--------------------------------------------------------------------------------
--------------------------------
DividendYield                            1.249e-06    4.7e-07      2.659
0.008      3.28e-07  2.17e-06
EBITDAYield                             -5.537e-06   9.48e-06     -0.584
0.559     -2.41e-05   1.3e-05
EVToEBITDA                               4.87e-06    8.7e-06      0.560
0.575     -1.22e-05  2.19e-05
EVToFCF                                  3.621e-06   1.08e-05      0.334
0.738     -1.76e-05  2.48e-05
PriceToBook                             -5.411e-07   1.03e-05     -0.052
0.958     -2.08e-05  1.97e-05
PriceToDilutedEarningsTTM                4.561e-06   4.64e-06      0.983
0.326     -4.54e-06  1.37e-05
PriceToEarningsTTM                           6e-08   2.05e-07      0.293
0.770     -3.41e-07  4.61e-07
PriceToFCF                              -9.612e-07   9.07e-06     -0.106
0.916     -1.87e-05  1.68e-05
PriceToOperatingCashflow                -6.258e-07   6.16e-06     -0.102
0.919     -1.27e-05  1.14e-05
PriceToSalesTTM                         -2.088e-06   5.16e-07     -4.048
0.000      -3.1e-06 -1.08e-06
Directional Movement Index              -3.169e-07   2.09e-06     -0.152
0.879      -4.4e-06  3.77e-06
Money Flow Index                        -2.768e-06    2.7e-06     -1.025
0.305     -8.06e-06  2.52e-06
Percent Above Low                        2.875e-06   5.17e-06      0.556
0.578     -7.26e-06   1.3e-05
Percent Below High                      -1.786e-06   4.14e-06     -0.431
0.666      -9.9e-06  6.33e-06
Price Oscillator                         9.18e-07   2.92e-06      0.314
0.753     -4.81e-06  6.64e-06
Trendline                                2.007e-06   4.41e-06      0.455
0.649     -6.64e-06  1.07e-05
AssetToEquityRatio                       5.633e-07   3.98e-07      1.414
0.157     -2.17e-07  1.34e-06
AssetTurnover                           -1.27e-05   1.74e-05     -0.731
```

0.465 -4.68e-05 2.14e-05
CurrentRatio 3.246e-07 5.2e-07 0.625
0.532 -6.94e-07 1.34e-06
DebtToAssetRatio 5.697e-07 3.18e-07 1.792
0.073 -5.35e-08 1.19e-06
DebtToEquityRatio -5.451e-07 3.69e-07 -1.476
0.140 -1.27e-06 1.79e-07
MertonsDD -0.0001 4.64e-05 -2.661
0.008 -0.000 -3.25e-05
WorkingCapitalToAssets 1.452e-07 6.31e-07 0.230
0.818 -1.09e-06 1.38e-06
WorkingCapitalToSales -2.24e-05 2.03e-05 -1.106
0.269 -6.21e-05 1.73e-05
Dividend Growth -1.505e-08 1.96e-07 -0.077
0.939 -4e-07 3.69e-07
EPS -1.219e-07 3.1e-07 -0.393
0.694 -7.3e-07 4.86e-07
Net Debt -1.968e-14 1.12e-14 -1.760
0.078 -4.16e-14 2.23e-15
Sales 5.85e-15 1.04e-14 0.563
0.574 -1.45e-14 2.62e-14
Total Assets -1.665e-14 7.87e-15 -2.116
0.034 -3.21e-14 -1.23e-15
EPS Growth 3M -5.625e-08 4.75e-07 -0.118
0.906 -9.87e-07 8.75e-07
EPS Growth 12M 6.686e-07 4.76e-07 1.405
0.160 -2.64e-07 1.6e-06
Net Debt Growth 3M 8.433e-08 6.03e-07 0.140
0.889 -1.1e-06 1.27e-06
Net Debt Growth 12M 4.558e-07 6.14e-07 0.743
0.458 -7.47e-07 1.66e-06
Sales Growth 3M 2.469e-07 6.4e-07 0.386
0.699 -1.01e-06 1.5e-06
Sales Growth 12M -4.861e-07 6.53e-07 -0.745
0.456 -1.77e-06 7.93e-07
Total Assets Growth 3M 4.535e-07 7.16e-07 0.633
0.527 -9.5e-07 1.86e-06
Total Assets Growth 12M -3.477e-07 7.64e-07 -0.455
0.649 -1.84e-06 1.15e-06
CFO To Assets 1.731e-05 8.66e-06 1.999
0.046 3.36e-07 3.43e-05
Capex To Assets -2.097e-05 1.87e-05 -1.124
0.261 -5.75e-05 1.56e-05
Capex To FCF 1.29e-06 1.05e-05 0.123
0.902 -1.92e-05 2.18e-05
Capex To Sales 1.637e-05 2e-05 0.820
0.412 -2.28e-05 5.55e-05

```
EBIT To Assets                              7.612e-06   8.75e-06      0.870
0.385     -9.55e-06  2.48e-05
Retained Earnings To Assets                -3.547e-05   1.84e-05     -1.923
0.054     -7.16e-05  6.76e-07
Downside Risk                               4.943e-07   5.32e-06      0.093
0.926     -9.93e-06  1.09e-05
Index Beta                                 -8.585e-08   1.09e-07     -0.786
0.432        -3e-07  1.28e-07
Log Market Cap                              1.847e-05   1.86e-05      0.995
0.320     -1.79e-05  5.49e-05
Volatility 3M                              -8.138e-06   4.53e-06     -1.798
0.072      -1.7e-05  7.33e-07
stock_3D SYSTEMS CORP                          0.0157      0.004      3.600
0.000         0.007     0.024
stock_3M COMPANY                               0.0108      0.004      2.823
0.005         0.003     0.018
stock_ABBOTT LABORATORIES                      0.0069      0.003      2.315
0.021         0.001     0.013
stock_ABBVIE INC                               0.0163      0.006      2.916
0.004         0.005     0.027
stock_ALLERGAN INC                             0.0103      0.003      3.563
0.000         0.005     0.016
stock_ALLERGAN PLC                             0.0135      0.004      3.310
0.001         0.006     0.022
stock_ALTABA INC                               0.0161      0.005      3.560
0.000         0.007     0.025
stock_ALTRIA GROUP INC.                        0.0111      0.004      2.847
0.004         0.003     0.019
stock_AMAZON.COM INC                           0.0123      0.005      2.683
0.007         0.003     0.021
stock_AMERICAN AIRLINES GROUP INC              0.0143      0.005      2.720
0.007         0.004     0.025
stock_AMERICAN EXPRESS COMPANY                 0.0065      0.002      2.656
0.008         0.002     0.011
stock_AMERICAN INTL GROUP INC                  0.0118      0.004      2.850
0.004         0.004     0.020
stock_AMGEN INC                                0.0064      0.003      2.421
0.015         0.001     0.012
stock_ANADARKO PETROLEUM CORP                  0.0077      0.002      3.231
0.001         0.003     0.012
stock_APACHE CORP                              0.0045      0.004      1.262
0.207        -0.003     0.012
stock_APPLE INC                                0.0086      0.004      2.278
0.023         0.001     0.016
stock_APPLIED MATERIALS INC                    0.0075      0.003      2.572
0.010         0.002     0.013
stock_ARCONIC INC                              0.0026      0.002      1.210
```

```
0.226        -0.002     0.007
stock_AT&T INC. COM                          0.0103      0.004      2.420
0.016         0.002     0.019
stock_Alphabet Inc. Cl A                     0.0192      0.005      3.528
0.000         0.009     0.030
stock_BAKER HUGHES INC                       0.0078      0.003      2.946
0.003         0.003     0.013
stock_BANK OF AMERICA CORP                   0.0404      0.016      2.532
0.011         0.009     0.072
stock_BERKSHIRE HATHAWAY INC CL-B            0.0171      0.006      3.069
0.002         0.006     0.028
stock_BIOGEN INC                             0.0120      0.004      3.275
0.001         0.005     0.019
stock_BOEING CO                              0.0053      0.003      1.864
0.062        -0.000     0.011
stock_BOOKING HOLDINGS INC                   0.0153      0.005      3.232
0.001         0.006     0.025
stock_BRISTOL MYERS SQUIBB COMPANY           0.0097      0.003      3.001
0.003         0.003     0.016
stock_BROADCOM CORP                          0.0134      0.004      3.087
0.002         0.005     0.022
stock_BROADCOM INC                           0.0173      0.006      3.112
0.002         0.006     0.028
stock_CATERPILLAR INC                        0.0042      0.003      1.502
0.133        -0.001     0.010
stock_CELGENE CORP                           0.0091      0.003      2.905
0.004         0.003     0.015
stock_CHESAPEAKE ENERGY CORP                 0.0041      0.005      0.874
0.382        -0.005     0.013
stock_CHEVRON CORPORATION                    0.0137      0.006      2.386
0.017         0.002     0.025
stock_CISCO SYSTEMS INC                      0.0066      0.003      2.363
0.018         0.001     0.012
stock_CITIGROUP                              0.0371      0.014      2.713
0.007         0.010     0.064
stock_COCA-COLA CO                           0.0101      0.004      2.629
0.009         0.003     0.018
stock_COMCAST CORP                           0.0081      0.003      2.917
0.004         0.003     0.013
stock_CONOCOPHILLIPS                         0.0125      0.005      2.566
0.010         0.003     0.022
stock_COVIDIEN PLC                           0.0196      0.005      3.839
0.000         0.010     0.030
stock_CVS HEALTH CORP                        0.0077      0.004      2.122
0.034         0.001     0.015
stock_DEERE & CO                             0.0050      0.003      1.671
0.095        -0.001     0.011
```

```
stock_DELTA AIR LINES INC                    0.0141      0.005      2.882
0.004         0.005     0.024
stock_DIRECTV                                0.0114      0.005      2.405
0.016         0.002     0.021
stock_DOLLAR GENERAL CORP                    0.0143      0.005      2.899
0.004         0.005     0.024
stock_DOW CHEMICAL CO                        0.0055      0.003      1.960
0.050     -5.93e-07     0.011
stock_E.I. Du Pont De Nemours A              0.0060      0.003      2.109
0.035         0.000     0.012
stock_EBAY INC                               0.0166      0.005      3.281
0.001         0.007     0.027
stock_EMC CORPORATION                        0.0082      0.003      2.886
0.004         0.003     0.014
stock_EOG RESOURCES INC                      0.0099      0.003      3.577
0.000         0.004     0.015
stock_EXPRESS SCRIPTS HOLDING CO             0.0034      0.003      1.166
0.244        -0.002     0.009
stock_EXXON MOBIL CORPORATION                0.0128      0.007      1.936
0.053        -0.000     0.026
stock_FACEBOOK INC                           0.0199      0.006      3.452
0.001         0.009     0.031
stock_FEDEX CORPORATION                      0.0079      0.003      2.618
0.009         0.002     0.014
stock_FIRST SOLAR INC                        0.0180      0.005      3.411
0.001         0.008     0.028
stock_FORD MOTOR CO(NEW)                     0.0057      0.004      1.626
0.104        -0.001     0.013
stock_FREEPORT-MCMORAN INC                   0.0090      0.004      2.333
0.020         0.001     0.017
stock_GENERAL ELECTRIC CO                    0.0165      0.005      3.025
0.002         0.006     0.027
stock_GENERAL MOTORS CO                      0.0131      0.006      2.358
0.018         0.002     0.024
stock_GILEAD SCIENCES INC                    0.0108      0.003      3.141
0.002         0.004     0.017
stock_GOLDMAN SACHS GROUP INC                0.0317      0.007      4.257
0.000         0.017     0.046
stock_GOPRO INC                              0.0199      0.006      3.550
0.000         0.009     0.031
stock_HALLIBURTON CO (HOLDING CO)            0.0084      0.003      2.752
0.006         0.002     0.014
stock_HOME DEPOT INC                         0.0070      0.003      2.052
0.040         0.000     0.014
stock_HP INC                                 0.0051      0.003      1.624
0.104        -0.001     0.011
stock_INTEL CORP                             0.0087      0.003      2.535
```

```
0.011         0.002     0.015
stock_INTL BUSINESS MACHINES CORP              0.0087      0.004      2.280
0.023         0.001     0.016
stock_JOHNSON AND JOHNSON                      0.0111      0.004      2.834
0.005         0.003     0.019
stock_JPMORGAN CHASE & CO COM STK              0.0545      0.019      2.818
0.005         0.017     0.092
stock_KEURIG GREEN MOUNTAIN INC                0.0146      0.004      3.591
0.000         0.007     0.023
stock_KINDER MORGAN INC                        0.0132      0.005      2.470
0.014         0.003     0.024
stock_LAS VEGAS SANDS CORP                     0.0127      0.005      2.488
0.013         0.003     0.023
stock_LILLY ELI & CO                           0.0102      0.004      2.885
0.004         0.003     0.017
stock_LINKEDIN CORP                            0.0197      0.006      3.505
0.000         0.009     0.031
stock_LOWES COMPANIES INC                      0.0075      0.003      2.483
0.013         0.002     0.013
stock_LYONDELLBASELL INDUSTRIES NV             0.0135      0.005      2.593
0.010         0.003     0.024
stock_MARATHON PETROLEUM CORP                  0.0128      0.006      2.268
0.023         0.002     0.024
stock_MASTERCARD INCORPORATED                  0.0206      0.006      3.731
0.000         0.010     0.031
stock_MCDONALDS CORP                           0.0099      0.004      2.513
0.012         0.002     0.018
stock_MEDTRONIC PLC                            0.0119      0.004      3.390
0.001         0.005     0.019
stock_MERCK & CO INC                           0.0110      0.004      2.966
0.003         0.004     0.018
stock_METLIFE  INC                             0.0260      0.008      3.353
0.001         0.011     0.041
stock_MICHAEL KORS HOLDINGS LTD                0.0214      0.006      3.646
0.000         0.010     0.033
stock_MICRON TECHNOLOGY INC                    0.0102      0.003      3.287
0.001         0.004     0.016
stock_MICROSOFT CORP                           0.0115      0.004      3.028
0.002         0.004     0.019
stock_MONDELEZ INTERNATIONAL INC               0.0133      0.005      2.875
0.004         0.004     0.022
stock_MONSANTO COMPANY                         0.0153      0.005      3.201
0.001         0.006     0.025
stock_MORGAN STANLEY                           0.0273      0.007      3.884
0.000         0.014     0.041
stock_MYLAN NV                                 0.0111      0.003      3.260
0.001         0.004     0.018
```

```
stock_NATIONAL OILWELL VARCO  INC.              0.0131      0.005      2.642
0.008         0.003     0.023
stock_NETFLIX INC                              0.0179      0.005      3.555
0.000         0.008     0.028
stock_NEWMONT MINING CORP (HOLDING COMPANY)    0.0098      0.004      2.773
0.006         0.003     0.017
stock_NEWS CP - CL A                           0.0129      0.004      3.348
0.001         0.005     0.020
stock_NIKE INC CL-B                            0.0104      0.004      2.884
0.004         0.003     0.018
stock_OCCIDENTAL PETROLEUM CORP                0.0118      0.004      3.153
0.002         0.004     0.019
stock_ORACLE CORP                              0.0118      0.004      3.297
0.001         0.005     0.019
stock_PANDORA MEDIA INC                        0.0235      0.006      3.942
0.000         0.012     0.035
stock_PENNEY J.C. CO INC (HOLDING COMPANY)     0.0027      0.003      0.905
0.365        -0.003     0.009
stock_PEPSICO INC                              0.0104      0.004      2.533
0.011         0.002     0.018
stock_PFIZER INC                               0.0134      0.004      3.287
0.001         0.005     0.021
stock_PHILIP MORRIS INTERNATIONAL INC          0.0154      0.006      2.576
0.010         0.004     0.027
stock_PIONEER NAT RES CO                       0.0177      0.004      4.022
0.000         0.009     0.026
stock_PRECISION CASTPARTS CORP                 0.0145      0.004      3.574
0.000         0.007     0.022
stock_PROCTER & GAMBLE CO                      0.0115      0.004      2.722
0.006         0.003     0.020
stock_QUALCOMM INC                             0.0131      0.004      3.223
0.001         0.005     0.021
stock_REGENERON PHARMACEUTICALS INC            0.0117      0.005      2.380
0.017         0.002     0.021
stock_SALESFORCE.COM INC                       0.0173      0.005      3.443
0.001         0.007     0.027
stock_SALIX PHARMACEUTICALS LTD           -2.111e-18   7.34e-17     -0.029
0.977     -1.46e-16  1.42e-16
stock_SANDISK CORP                             0.0154      0.004      3.629
0.000         0.007     0.024
stock_SCHLUMBERGER LTD.                        0.0120      0.004      3.003
0.003         0.004     0.020
stock_SKYWORKS SOLUTIONS INC                   0.0191      0.005      3.796
0.000         0.009     0.029
stock_SOLARCITY CORP                           0.0208      0.006      3.549
0.000         0.009     0.032
stock_SOUTHWEST AIRLINES CO                    0.0092      0.003      2.885
```

```
0.004         0.003     0.015
stock_STARBUCKS CORPORATION                    0.0147      0.004      3.658
0.000         0.007     0.023
stock_SUNEDISON INC                            0.0183      0.006      3.237
0.001         0.007     0.029
stock_TARGET CORPORATION                       0.0104      0.004      2.349
0.019         0.002     0.019
stock_TESLA INC                                0.0196      0.006      3.520
0.000         0.009     0.031
stock_TEXAS INSTRUMENTS INC                    0.0141      0.004      3.229
0.001         0.006     0.023
stock_TIME WARNER CABLE INC                    0.0140      0.005      2.883
0.004         0.004     0.024
stock_TIME WARNER INC.                         0.0049      0.002      2.324
0.020         0.001     0.009
stock_TWITTER INC                              0.0185      0.006      3.201
0.001         0.007     0.030
stock_UNION PACIFIC CORPORATION                0.0130      0.004      3.275
0.001         0.005     0.021
stock_UNITED CONTINENTAL HOLDINGS IN           0.0097      0.005      2.019
0.044         0.000     0.019
stock_UNITED PARCEL SERVICE INC.CL B           0.0101      0.005      2.184
0.029         0.001     0.019
stock_UNITED TECHNOLOGIES CORP                 0.0113      0.004      2.830
0.005         0.003     0.019
stock_UNITEDHEALTH GROUP INC                   0.0100      0.004      2.453
0.014         0.002     0.018
stock_VALERO ENERGY CORP (NEW)                 0.0099      0.004      2.405
0.016         0.002     0.018
stock_VERIZON COMMUNICATIONS                   0.0116      0.005      2.369
0.018         0.002     0.021
stock_VISA INC                                 0.0189      0.005      3.588
0.000         0.009     0.029
stock_WALGREENS BOOTS ALLIANCE INC             0.0109      0.004      2.732
0.006         0.003     0.019
stock_WALMART INC                              0.0078      0.006      1.208
0.227        -0.005     0.020
stock_WALT DISNEY CO                           0.0102      0.003      3.309
0.001         0.004     0.016
stock_WELLS FARGO & CO(NEW)                    0.0434      0.013      3.348
0.001         0.018     0.069
stock_WILLIAMS COMPANIES                       0.0109      0.004      2.720
0.007         0.003     0.019
stock_WYNN RESORTS LTD                         0.0109      0.005      2.254
0.024         0.001     0.020
stock_YELP INC                                 0.0231      0.006      3.978
0.000         0.012     0.034
```

```
==============================================================================
Omnibus:                       41.371   Durbin-Watson:                   1.329
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               45.849
Skew:                          -0.034   Prob(JB):                     1.11e-10
Kurtosis:                       3.141   Cond. No.                     5.40e+21
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 1.4e-16. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
"""
```

The summary is available in the notebook to save some space due to the large number of variables. The diagnostic statistics show that, given the high p-value on the Jarque—Bera statistic, the hypothesis that the residuals are normally distributed cannot be rejected.

However, the Durbin—Watson statistic is low at 1.5 so we can reject the null hypothesis of no autocorrelation comfortably at the 5% level. Hence, the standard errors are likely positively correlated. If our goal were to understand which factors are significantly associated with forward returns, we would need to rerun the regression using robust standard errors (a parameter in statsmodels .fit() method), or use a different method altogether such as a panel model that allows for more complex error covariance.

[45]:
```
target = 'Returns5D'
model_data = pd.concat([y[[target]], X], axis=1).dropna()
model_data = model_data[model_data[target].between(model_data[target].quantile(.
↪025),
                                                   model_data[target].quantile(.
↪975))]

model = OLS(endog=model_data[target], exog=model_data.drop(target, axis=1))
trained_model = model.fit()
trained_model.summary()
```

[45]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:              Returns5D   R-squared:                       0.016
Model:                            OLS   Adj. R-squared:                  0.012
Method:                 Least Squares   F-statistic:                     4.124
Date:                Sun, 09 Sep 2018   Prob (F-statistic):           2.28e-66
Time:                        21:18:00   Log-Likelihood:                 95279.
No. Observations:               44370   AIC:                        -1.902e+05
Df Residuals:                   44195   BIC:                        -1.887e+05
Df Model:                         174
```

```
Covariance Type:            nonrobust
================================================================================
==============================
                                                 coef    std err          t
P>|t|      [95.0% Conf. Int.]
--------------------------------------------------------------------------------
------------------------------
DividendYield                                4.504e-06   1.08e-06      4.185
0.000      2.39e-06  6.61e-06
EBITDAYield                                  9.323e-06   2.17e-05      0.430
0.667     -3.32e-05  5.19e-05
EVToEBITDA                                   1.425e-05   1.98e-05      0.719
0.472     -2.46e-05  5.31e-05
EVToFCF                                      2.654e-05   2.47e-05      1.075
0.283     -2.19e-05  7.49e-05
PriceToBook                                  1.582e-05   2.36e-05      0.669
0.503     -3.05e-05  6.21e-05
PriceToDilutedEarningsTTM                    5.462e-07   1.06e-05      0.052
0.959     -2.01e-05  2.12e-05
PriceToEarningsTTM                           1.449e-07   4.67e-07      0.310
0.757     -7.71e-07  1.06e-06
PriceToFCF                                  -1.94e-05   2.07e-05     -0.939
0.348     -5.99e-05  2.11e-05
PriceToOperatingCashflow                    -4.155e-05   1.42e-05     -2.934
0.003     -6.93e-05 -1.38e-05
PriceToSalesTTM                             -8.975e-06   1.19e-06     -7.530
0.000     -1.13e-05 -6.64e-06
Directional Movement Index                  -4.457e-06   4.75e-06     -0.938
0.348     -1.38e-05  4.86e-06
Money Flow Index                            -5.894e-06   6.16e-06     -0.956
0.339      -1.8e-05  6.19e-06
Percent Above Low                           -1.985e-05   1.18e-05     -1.675
0.094     -4.31e-05  3.38e-06
Percent Below High                           7.651e-06   9.42e-06      0.812
0.417     -1.08e-05  2.61e-05
Price Oscillator                             3.706e-06   6.64e-06      0.558
0.577     -9.31e-06  1.67e-05
Trendline                                    1.414e-05   1.01e-05      1.402
0.161     -5.64e-06  3.39e-05
AssetToEquityRatio                          -1.226e-07   8.91e-07     -0.138
0.891     -1.87e-06  1.62e-06
AssetTurnover                                -7.3e-06   3.95e-05     -0.185
0.854     -8.48e-05  7.02e-05
CurrentRatio                                 4.11e-07   1.18e-06      0.347
0.728     -1.91e-06  2.73e-06
DebtToAssetRatio                             2.484e-06   7.25e-07      3.427
0.001      1.06e-06   3.9e-06
```

```
DebtToEquityRatio                          -5.365e-07   8.37e-07     -0.641
0.521     -2.18e-06   1.1e-06
MertonsDD                                    -0.0004      0.000     -3.508
0.000        -0.001    -0.000
WorkingCapitalToAssets                      1.341e-06   1.44e-06      0.932
0.351     -1.48e-06  4.16e-06
WorkingCapitalToSales                      -3.118e-05   4.62e-05     -0.674
0.500        -0.000  5.95e-05
Dividend Growth                            -2.629e-07    4.5e-07     -0.585
0.559     -1.14e-06  6.18e-07
EPS                                         2.901e-07   7.04e-07      0.412
0.680     -1.09e-06  1.67e-06
Net Debt                                   -8.116e-14   2.56e-14     -3.172
0.002     -1.31e-13  -3.1e-14
Sales                                      -5.373e-14    2.4e-14     -2.239
0.025     -1.01e-13 -6.69e-15
Total Assets                               -4.104e-14    1.8e-14     -2.276
0.023     -7.64e-14  -5.7e-15
EPS Growth 3M                              -2.787e-06   1.08e-06     -2.582
0.010      -4.9e-06 -6.72e-07
EPS Growth 12M                              1.257e-06   1.08e-06      1.161
0.246     -8.64e-07  3.38e-06
Net Debt Growth 3M                          2.628e-07   1.37e-06      0.191
0.848     -2.43e-06  2.95e-06
Net Debt Growth 12M                         2.592e-06    1.4e-06      1.858
0.063     -1.43e-07  5.33e-06
Sales Growth 3M                             3.824e-06   1.45e-06      2.634
0.008      9.79e-07  6.67e-06
Sales Growth 12M                            -2.38e-06   1.48e-06     -1.604
0.109     -5.29e-06  5.28e-07
Total Assets Growth 3M                     -6.292e-07   1.63e-06     -0.386
0.699     -3.82e-06  2.56e-06
Total Assets Growth 12M                     -1.46e-06   1.73e-06     -0.842
0.400     -4.86e-06  1.94e-06
CFO To Assets                               4.634e-05   1.98e-05      2.342
0.019      7.56e-06  8.51e-05
Capex To Assets                              -0.0001   4.26e-05     -3.232
0.001        -0.000 -5.42e-05
Capex To FCF                                3.673e-05   2.39e-05      1.534
0.125     -1.02e-05  8.37e-05
Capex To Sales                               0.0001   4.56e-05      2.247
0.025      1.31e-05     0.000
EBIT To Assets                              8.146e-06      2e-05      0.408
0.683      -3.1e-05  4.73e-05
Retained Earnings To Assets                  -0.0001    4.2e-05     -3.036
0.002        -0.000 -4.51e-05
Downside Risk                               1.166e-05   1.22e-05      0.960
```

```
0.337     -1.22e-05  3.55e-05
Index Beta                              -7.363e-07   2.49e-07     -2.958
0.003     -1.22e-06 -2.48e-07
Log Market Cap                           9.103e-05   4.24e-05      2.148
0.032      7.98e-06     0.000
Volatility 3M                           -5.492e-05   1.04e-05     -5.306
0.000     -7.52e-05 -3.46e-05
stock_3D SYSTEMS CORP                       0.0568      0.010      5.697
0.000         0.037     0.076
stock_3M COMPANY                            0.0426      0.009      4.865
0.000         0.025     0.060
stock_ABBOTT LABORATORIES                   0.0264      0.007      3.854
0.000         0.013     0.040
stock_ABBVIE INC                            0.0572      0.013      4.482
0.000         0.032     0.082
stock_ALLERGAN INC                          0.0487      0.007      7.349
0.000         0.036     0.062
stock_ALLERGAN PLC                          0.0543      0.009      5.812
0.000         0.036     0.073
stock_ALTABA INC                            0.0605      0.010      5.857
0.000         0.040     0.081
stock_ALTRIA GROUP INC.                     0.0453      0.009      5.074
0.000         0.028     0.063
stock_AMAZON.COM INC                        0.0513      0.010      4.905
0.000         0.031     0.072
stock_AMERICAN AIRLINES GROUP INC           0.0490      0.012      4.079
0.000         0.025     0.073
stock_AMERICAN EXPRESS COMPANY              0.0232      0.006      4.177
0.000         0.012     0.034
stock_AMERICAN INTL GROUP INC               0.0452      0.009      4.770
0.000         0.027     0.064
stock_AMGEN INC                             0.0266      0.006      4.400
0.000         0.015     0.038
stock_ANADARKO PETROLEUM CORP               0.0347      0.005      6.387
0.000         0.024     0.045
stock_APACHE CORP                           0.0130      0.009      1.487
0.137        -0.004     0.030
stock_APPLE INC                             0.0471      0.009      5.478
0.000         0.030     0.064
stock_APPLIED MATERIALS INC                 0.0333      0.007      4.965
0.000         0.020     0.046
stock_ARCONIC INC                           0.0090      0.005      1.822
0.068        -0.001     0.019
stock_AT&T INC. COM                         0.0372      0.010      3.817
0.000         0.018     0.056
stock_Alphabet Inc. Cl A                    0.0713      0.012      5.733
0.000         0.047     0.096
```

```
stock_BAKER HUGHES INC                          0.0296      0.006      4.920
0.000          0.018     0.041
stock_BANK OF AMERICA CORP                      0.1209      0.037      3.306
0.001          0.049     0.193
stock_BERKSHIRE HATHAWAY INC CL-B               0.0629      0.013      4.939
0.000          0.038     0.088
stock_BIOGEN INC                                0.0551      0.008      6.605
0.000          0.039     0.071
stock_BOEING CO                                 0.0230      0.007      3.524
0.000          0.010     0.036
stock_BOOKING HOLDINGS INC                      0.0600      0.011      5.539
0.000          0.039     0.081
stock_BRISTOL MYERS SQUIBB COMPANY              0.0405      0.007      5.456
0.000          0.026     0.055
stock_BROADCOM CORP                             0.0564      0.010      5.697
0.000          0.037     0.076
stock_BROADCOM INC                              0.0733      0.013      5.818
0.000          0.049     0.098
stock_CATERPILLAR INC                           0.0139      0.006      2.172
0.030          0.001     0.027
stock_CELGENE CORP                              0.0385      0.007      5.398
0.000          0.025     0.053
stock_CHESAPEAKE ENERGY CORP                    0.0269      0.011      2.555
0.011          0.006     0.048
stock_CHEVRON CORPORATION                       0.0562      0.013      4.283
0.000          0.030     0.082
stock_CISCO SYSTEMS INC                         0.0285      0.006      4.464
0.000          0.016     0.041
stock_CITIGROUP                                 0.1090      0.031      3.474
0.001          0.048     0.171
stock_COCA-COLA CO                              0.0384      0.009      4.352
0.000          0.021     0.056
stock_COMCAST CORP                              0.0325      0.006      5.155
0.000          0.020     0.045
stock_CONOCOPHILLIPS                            0.0456      0.011      4.089
0.000          0.024     0.067
stock_COVIDIEN PLC                              0.0743      0.012      6.354
0.000          0.051     0.097
stock_CVS HEALTH CORP                           0.0358      0.008      4.345
0.000          0.020     0.052
stock_DEERE & CO                                0.0115      0.007      1.683
0.092         -0.002     0.025
stock_DELTA AIR LINES INC                       0.0545      0.011      4.896
0.000          0.033     0.076
stock_DIRECTV                                   0.0383      0.011      3.558
0.000          0.017     0.059
stock_DOLLAR GENERAL CORP                       0.0508      0.011      4.509
```

```
0.000         0.029     0.073
stock_DOW CHEMICAL CO                          0.0234      0.006      3.686
0.000         0.011     0.036
stock_E.I. Du Pont De Nemours A                0.0204      0.007      3.137
0.002         0.008     0.033
stock_EBAY INC                                 0.0611      0.012      5.279
0.000         0.038     0.084
stock_EMC CORPORATION                          0.0341      0.007      5.245
0.000         0.021     0.047
stock_EOG RESOURCES INC                        0.0413      0.006      6.573
0.000         0.029     0.054
stock_EXPRESS SCRIPTS HOLDING CO               0.0177      0.007      2.628
0.009         0.005     0.031
stock_EXXON MOBIL CORPORATION                  0.0677      0.015      4.484
0.000         0.038     0.097
stock_FACEBOOK INC                             0.0762      0.013      5.785
0.000         0.050     0.102
stock_FEDEX CORPORATION                        0.0321      0.007      4.657
0.000         0.019     0.046
stock_FIRST SOLAR INC                          0.0484      0.012      4.004
0.000         0.025     0.072
stock_FORD MOTOR CO(NEW)                       0.0213      0.008      2.647
0.008         0.006     0.037
stock_FREEPORT-MCMORAN INC                     0.0269      0.009      3.063
0.002         0.010     0.044
stock_GENERAL ELECTRIC CO                      0.0556      0.013      4.444
0.000         0.031     0.080
stock_GENERAL MOTORS CO                        0.0485      0.013      3.829
0.000         0.024     0.073
stock_GILEAD SCIENCES INC                      0.0459      0.008      5.868
0.000         0.031     0.061
stock_GOLDMAN SACHS GROUP INC                  0.0965      0.017      5.644
0.000         0.063     0.130
stock_GOPRO INC                                0.0687      0.013      5.377
0.000         0.044     0.094
stock_HALLIBURTON CO (HOLDING CO)              0.0322      0.007      4.601
0.000         0.018     0.046
stock_HOME DEPOT INC                           0.0336      0.008      4.320
0.000         0.018     0.049
stock_HP INC                                   0.0254      0.007      3.507
0.000         0.011     0.040
stock_INTEL CORP                               0.0383      0.008      4.891
0.000         0.023     0.054
stock_INTL BUSINESS MACHINES CORP              0.0329      0.009      3.784
0.000         0.016     0.050
stock_JOHNSON AND JOHNSON                      0.0431      0.009      4.833
0.000         0.026     0.061
```

stock_JPMORGAN CHASE & CO COM STK	0.1531	0.044	3.455	
0.001	0.066	0.240		
stock_KEURIG GREEN MOUNTAIN INC	0.0569	0.009	6.135	
0.000	0.039	0.075		
stock_KINDER MORGAN INC	0.0427	0.012	3.497	
0.000	0.019	0.067		
stock_LAS VEGAS SANDS CORP	0.0494	0.012	4.235	
0.000	0.027	0.072		
stock_LILLY ELI & CO	0.0411	0.008	5.097	
0.000	0.025	0.057		
stock_LINKEDIN CORP	0.0700	0.013	5.460	
0.000	0.045	0.095		
stock_LOWES COMPANIES INC	0.0307	0.007	4.474	
0.000	0.017	0.044		
stock_LYONDELLBASELL INDUSTRIES NV	0.0478	0.012	4.043	
0.000	0.025	0.071		
stock_MARATHON PETROLEUM CORP	0.0508	0.013	3.939	
0.000	0.026	0.076		
stock_MASTERCARD INCORPORATED	0.0778	0.013	6.157	
0.000	0.053	0.103		
stock_MCDONALDS CORP	0.0380	0.009	4.236	
0.000	0.020	0.056		
stock_MEDTRONIC PLC	0.0482	0.008	6.017	
0.000	0.033	0.064		
stock_MERCK & CO INC	0.0425	0.008	5.002	
0.000	0.026	0.059		
stock_METLIFE INC	0.0738	0.018	4.162	
0.000	0.039	0.109		
stock_MICHAEL KORS HOLDINGS LTD	0.0770	0.013	5.750	
0.000	0.051	0.103		
stock_MICRON TECHNOLOGY INC	0.0363	0.007	5.112	
0.000	0.022	0.050		
stock_MICROSOFT CORP	0.0468	0.009	5.410	
0.000	0.030	0.064		
stock_MONDELEZ INTERNATIONAL INC	0.0519	0.011	4.934	
0.000	0.031	0.073		
stock_MONSANTO COMPANY	0.0556	0.011	5.111	
0.000	0.034	0.077		
stock_MORGAN STANLEY	0.0829	0.016	5.139	
0.000	0.051	0.115		
stock_MYLAN NV	0.0430	0.008	5.515	
0.000	0.028	0.058		
stock_NATIONAL OILWELL VARCO INC.	0.0397	0.011	3.525	
0.000	0.018	0.062		
stock_NETFLIX INC	0.0690	0.011	6.016	
0.000	0.047	0.092		
stock_NEWMONT MINING CORP (HOLDING COMPANY)	0.0312	0.008	3.782	

```
0.000         0.015     0.047
stock_NEWS CP - CL A                              0.0462      0.009      5.253
0.000         0.029     0.063
stock_NIKE INC CL-B                               0.0466      0.008      5.636
0.000         0.030     0.063
stock_OCCIDENTAL PETROLEUM CORP                   0.0420      0.009      4.899
0.000         0.025     0.059
stock_ORACLE CORP                                 0.0434      0.008      5.331
0.000         0.027     0.059
stock_PANDORA MEDIA INC                           0.0953      0.014      6.986
0.000         0.069     0.122
stock_PENNEY J.C. CO INC (HOLDING COMPANY)        0.0163      0.007      2.353
0.019         0.003     0.030
stock_PEPSICO INC                                 0.0387      0.009      4.139
0.000         0.020     0.057
stock_PFIZER INC                                  0.0498      0.009      5.350
0.000         0.032     0.068
stock_PHILIP MORRIS INTERNATIONAL INC             0.0577      0.014      4.236
0.000         0.031     0.084
stock_PIONEER NAT RES CO                          0.0716      0.010      7.136
0.000         0.052     0.091
stock_PRECISION CASTPARTS CORP                    0.0510      0.009      5.448
0.000         0.033     0.069
stock_PROCTER & GAMBLE CO                         0.0442      0.010      4.581
0.000         0.025     0.063
stock_QUALCOMM INC                                0.0532      0.009      5.739
0.000         0.035     0.071
stock_REGENERON PHARMACEUTICALS INC               0.0537      0.011      4.675
0.000         0.031     0.076
stock_SALESFORCE.COM INC                          0.0610      0.011      5.303
0.000         0.038     0.083
stock_SALIX PHARMACEUTICALS LTD                -4.65e-16   2.62e-15     -0.178
0.859      -5.6e-15  4.67e-15
stock_SANDISK CORP                                0.0624      0.010      6.426
0.000         0.043     0.081
stock_SCHLUMBERGER LTD.                           0.0441      0.009      4.843
0.000         0.026     0.062
stock_SKYWORKS SOLUTIONS INC                      0.0772      0.011      6.749
0.000         0.055     0.100
stock_SOLARCITY CORP                              0.0820      0.013      6.129
0.000         0.056     0.108
stock_SOUTHWEST AIRLINES CO                       0.0430      0.007      5.909
0.000         0.029     0.057
stock_STARBUCKS CORPORATION                       0.0606      0.009      6.590
0.000         0.043     0.079
stock_SUNEDISON INC                               0.0234      0.017      1.401
0.161        -0.009     0.056
```

```
stock_TARGET CORPORATION                        0.0356      0.010      3.541
0.000         0.016     0.055
stock_TESLA INC                                 0.0720      0.013      5.659
0.000         0.047     0.097
stock_TEXAS INSTRUMENTS INC                     0.0522      0.010      5.206
0.000         0.033     0.072
stock_TIME WARNER CABLE INC                     0.0468      0.011      4.218
0.000         0.025     0.069
stock_TIME WARNER INC.                          0.0227      0.005      4.748
0.000         0.013     0.032
stock_TWITTER INC                               0.0686      0.013      5.194
0.000         0.043     0.094
stock_UNION PACIFIC CORPORATION                 0.0493      0.009      5.446
0.000         0.032     0.067
stock_UNITED CONTINENTAL HOLDINGS IN            0.0379      0.011      3.469
0.001         0.016     0.059
stock_UNITED PARCEL SERVICE INC.CL B            0.0366      0.011      3.462
0.001         0.016     0.057
stock_UNITED TECHNOLOGIES CORP                  0.0390      0.009      4.277
0.000         0.021     0.057
stock_UNITEDHEALTH GROUP INC                    0.0455      0.009      4.866
0.000         0.027     0.064
stock_VALERO ENERGY CORP (NEW)                  0.0432      0.009      4.596
0.000         0.025     0.062
stock_VERIZON COMMUNICATIONS                    0.0392      0.011      3.502
0.000         0.017     0.061
stock_VISA INC                                  0.0769      0.012      6.389
0.000         0.053     0.100
stock_WALGREENS BOOTS ALLIANCE INC              0.0444      0.009      4.859
0.000         0.026     0.062
stock_WALMART INC                               0.0554      0.015      3.742
0.000         0.026     0.084
stock_WALT DISNEY CO                            0.0432      0.007      6.119
0.000         0.029     0.057
stock_WELLS FARGO & CO(NEW)                     0.1324      0.030      4.449
0.000         0.074     0.191
stock_WILLIAMS COMPANIES                        0.0472      0.009      5.143
0.000         0.029     0.065
stock_WYNN RESORTS LTD                          0.0410      0.011      3.710
0.000         0.019     0.063
stock_YELP INC                                  0.0728      0.013      5.489
0.000         0.047     0.099
==============================================================================
Omnibus:                       71.266   Durbin-Watson:                   1.433
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               77.090
Skew:                          -0.067   Prob(JB):                     1.82e-17
Kurtosis:                       3.154   Cond. No.                     3.42e+20
```

```
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 3.44e-14. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
"""
```

[46]:
```
target = 'Returns10D'
model_data = pd.concat([y[[target]], X], axis=1).dropna()
model_data = model_data[model_data[target].between(model_data[target].quantile(.
↪025),
                                                   model_data[target].quantile(.
↪975))]

model = OLS(endog=model_data[target], exog=model_data.drop(target, axis=1))
trained_model = model.fit()
trained_model.summary()
```

[46]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:             Returns10D   R-squared:                       0.035
Model:                            OLS   Adj. R-squared:                  0.031
Method:                 Least Squares   F-statistic:                     9.045
Date:                Sun, 09 Sep 2018   Prob (F-statistic):          5.42e-219
Time:                        21:18:02   Log-Likelihood:                 78861.
No. Observations:               43734   AIC:                        -1.574e+05
Df Residuals:                   43559   BIC:                        -1.559e+05
Df Model:                         174
Covariance Type:            nonrobust
==============================================================================
==============================
                                                 coef    std err          t
P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
------------------------------
DividendYield                                 6.757e-06   1.55e-06      4.370
0.000      3.73e-06  9.79e-06
EBITDAYield                                  -1.021e-05   3.11e-05     -0.329
0.743     -7.11e-05  5.07e-05
EVToEBITDA                                    9.163e-05   2.83e-05      3.233
0.001      3.61e-05     0.000
EVToFCF                                       2.372e-05   3.51e-05      0.676
0.499      -4.5e-05  9.25e-05
```

```
PriceToBook                                 8.448e-05   3.34e-05      2.533
0.011      1.91e-05     0.000
PriceToDilutedEarningsTTM                  -1.332e-05    1.5e-05     -0.889
0.374     -4.27e-05  1.61e-05
PriceToEarningsTTM                         -6.983e-07   6.67e-07     -1.047
0.295        -2e-06  6.08e-07
PriceToFCF                                 -4.525e-06   2.91e-05     -0.155
0.877     -6.16e-05  5.26e-05
PriceToOperatingCashflow                   -8.243e-05   2.02e-05     -4.071
0.000        -0.000 -4.27e-05
PriceToSalesTTM                            -2.324e-05   1.72e-06    -13.493
0.000     -2.66e-05 -1.99e-05
Directional Movement Index                 -2.235e-05   6.75e-06     -3.308
0.001     -3.56e-05 -9.11e-06
Money Flow Index                            4.299e-06   8.77e-06      0.490
0.624     -1.29e-05  2.15e-05
Percent Above Low                          -4.475e-05   1.71e-05     -2.623
0.009     -7.82e-05 -1.13e-05
Percent Below High                          1.295e-05   1.34e-05      0.965
0.335     -1.34e-05  3.93e-05
Price Oscillator                             6.06e-06   9.48e-06      0.640
0.522     -1.25e-05  2.46e-05
Trendline                                    2.48e-05   1.44e-05      1.722
0.085     -3.42e-06   5.3e-05
AssetToEquityRatio                         -2.548e-06    1.3e-06     -1.964
0.049     -5.09e-06 -5.59e-09
AssetTurnover                                -0.0001   5.63e-05     -2.054
0.040        -0.000 -5.31e-06
CurrentRatio                                3.392e-06   1.69e-06      2.007
0.045      7.96e-08   6.7e-06
DebtToAssetRatio                            2.605e-06   1.04e-06      2.500
0.012      5.63e-07  4.65e-06
DebtToEquityRatio                          -2.576e-08   1.21e-06     -0.021
0.983      -2.4e-06  2.35e-06
MertonsDD                                    -0.0010      0.000     -6.752
0.000        -0.001    -0.001
WorkingCapitalToAssets                       3.97e-06   2.04e-06      1.943
0.052     -3.38e-08  7.97e-06
WorkingCapitalToSales                        -0.0002   6.58e-05     -2.654
0.008        -0.000 -4.56e-05
Dividend Growth                            -9.485e-07   6.45e-07     -1.471
0.141     -2.21e-06  3.16e-07
EPS                                         5.254e-07      1e-06      0.525
0.599     -1.43e-06  2.48e-06
Net Debt                                    -1.53e-13   3.65e-14     -4.193
0.000     -2.25e-13 -8.15e-14
Sales                                      -1.061e-13   3.43e-14     -3.094
```

```
0.002     -1.73e-13 -3.89e-14
Total Assets                          -7.923e-14   2.58e-14     -3.068
0.002      -1.3e-13 -2.86e-14
EPS Growth 3M                         -6.126e-06   1.53e-06     -4.002
0.000     -9.13e-06 -3.13e-06
EPS Growth 12M                         2.182e-06   1.53e-06      1.423
0.155     -8.24e-07  5.19e-06
Net Debt Growth 3M                    -1.748e-06   1.95e-06     -0.897
0.370     -5.57e-06  2.07e-06
Net Debt Growth 12M                    5.675e-06   1.98e-06      2.872
0.004       1.8e-06  9.55e-06
Sales Growth 3M                         4.22e-06   2.06e-06      2.051
0.040      1.87e-07  8.25e-06
Sales Growth 12M                       -6.66e-06    2.1e-06     -3.169
0.002     -1.08e-05 -2.54e-06
Total Assets Growth 3M                -2.435e-07   2.31e-06     -0.105
0.916     -4.77e-06  4.28e-06
Total Assets Growth 12M               -2.601e-06   2.46e-06     -1.059
0.290     -7.42e-06  2.21e-06
CFO To Assets                          4.355e-05   2.81e-05      1.551
0.121     -1.15e-05  9.86e-05
Capex To Assets                          -0.0001    6.1e-05     -2.246
0.025        -0.000 -1.74e-05
Capex To FCF                           6.95e-05   3.42e-05      2.030
0.042       2.4e-06     0.000
Capex To Sales                           0.0002   6.51e-05      3.016
0.003      6.87e-05     0.000
EBIT To Assets                         1.444e-05   2.83e-05      0.511
0.610      -4.1e-05  6.99e-05
Retained Earnings To Assets              -0.0004   6.01e-05     -6.676
0.000        -0.001    -0.000
Downside Risk                          3.345e-05   1.73e-05      1.932
0.053     -4.83e-07  6.74e-05
Index Beta                            -2.09e-06   3.55e-07     -5.886
0.000     -2.79e-06 -1.39e-06
Log Market Cap                           0.0002   6.04e-05      3.147
0.002      7.17e-05     0.000
Volatility 3M                         -8.754e-05   1.47e-05     -5.939
0.000        -0.000 -5.86e-05
stock_3D SYSTEMS CORP                    0.1609      0.014     11.264
0.000         0.133     0.189
stock_3M COMPANY                         0.1423      0.012     11.416
0.000         0.118     0.167
stock_ABBOTT LABORATORIES                0.0903      0.010      9.116
0.000         0.071     0.110
stock_ABBVIE INC                         0.1833      0.018     10.070
0.000         0.148     0.219
```

stock_ALLERGAN INC			0.1434	0.009	15.113
0.000	0.125	0.162			
stock_ALLERGAN PLC			0.1683	0.013	12.648
0.000	0.142	0.194			
stock_ALTABA INC			0.1825	0.015	12.388
0.000	0.154	0.211			
stock_ALTRIA GROUP INC.			0.1533	0.013	12.015
0.000	0.128	0.178			
stock_AMAZON.COM INC			0.1494	0.015	10.035
0.000	0.120	0.179			
stock_AMERICAN AIRLINES GROUP INC			0.1433	0.017	8.373
0.000	0.110	0.177			
stock_AMERICAN EXPRESS COMPANY			0.0785	0.008	9.844
0.000	0.063	0.094			
stock_AMERICAN INTL GROUP INC			0.1021	0.014	7.505
0.000	0.075	0.129			
stock_AMGEN INC			0.0915	0.009	10.550
0.000	0.074	0.108			
stock_ANADARKO PETROLEUM CORP			0.0978	0.008	12.627
0.000	0.083	0.113			
stock_APACHE CORP			0.0568	0.014	4.038
0.000	0.029	0.084			
stock_APPLE INC			0.1309	0.012	10.689
0.000	0.107	0.155			
stock_APPLIED MATERIALS INC			0.1080	0.010	11.289
0.000	0.089	0.127			
stock_ARCONIC INC			0.0422	0.007	5.984
0.000	0.028	0.056			
stock_AT&T INC. COM			0.1122	0.014	8.068
0.000	0.085	0.139			
stock_Alphabet Inc. Cl A			0.2024	0.018	11.450
0.000	0.168	0.237			
stock_BAKER HUGHES INC			0.0945	0.009	11.076
0.000	0.078	0.111			
stock_BANK OF AMERICA CORP			0.2669	0.052	5.090
0.000	0.164	0.370			
stock_BERKSHIRE HATHAWAY INC CL-B			0.1630	0.018	8.968
0.000	0.127	0.199			
stock_BIOGEN INC			0.1688	0.012	14.185
0.000	0.145	0.192			
stock_BOEING CO			0.0769	0.009	8.299
0.000	0.059	0.095			
stock_BOOKING HOLDINGS INC			0.1898	0.015	12.281
0.000	0.159	0.220			
stock_BRISTOL MYERS SQUIBB COMPANY			0.1327	0.011	12.504
0.000	0.112	0.153			
stock_BROADCOM CORP			0.1655	0.014	11.708

```
0.000         0.138     0.193
stock_BROADCOM INC                          0.2090      0.018     11.611
0.000         0.174     0.244
stock_CATERPILLAR INC                       0.0647      0.009      7.093
0.000         0.047     0.083
stock_CELGENE CORP                          0.1326      0.010     12.974
0.000         0.113     0.153
stock_CHESAPEAKE ENERGY CORP                0.0682      0.016      4.174
0.000         0.036     0.100
stock_CHEVRON CORPORATION                   0.1637      0.019      8.770
0.000         0.127     0.200
stock_CISCO SYSTEMS INC                     0.0975      0.009     10.689
0.000         0.080     0.115
stock_CITIGROUP                             0.2358      0.045      5.236
0.000         0.148     0.324
stock_COCA-COLA CO                          0.1313      0.013     10.439
0.000         0.107     0.156
stock_COMCAST CORP                          0.0961      0.009     10.691
0.000         0.078     0.114
stock_CONOCOPHILLIPS                        0.1432      0.016      9.039
0.000         0.112     0.174
stock_COVIDIEN PLC                          0.2103      0.017     12.561
0.000         0.177     0.243
stock_CVS HEALTH CORP                       0.1147      0.012      9.823
0.000         0.092     0.138
stock_DEERE & CO                            0.0556      0.010      5.698
0.000         0.036     0.075
stock_DELTA AIR LINES INC                   0.1560      0.016      9.833
0.000         0.125     0.187
stock_DIRECTV                               0.1262      0.015      8.223
0.000         0.096     0.156
stock_DOLLAR GENERAL CORP                   0.1543      0.016      9.600
0.000         0.123     0.186
stock_DOW CHEMICAL CO                       0.0810      0.009      8.981
0.000         0.063     0.099
stock_E.I. Du Pont De Nemours A             0.0722      0.009      7.789
0.000         0.054     0.090
stock_EBAY INC                              0.1879      0.016     11.397
0.000         0.156     0.220
stock_EMC CORPORATION                       0.1141      0.009     12.320
0.000         0.096     0.132
stock_EOG RESOURCES INC                     0.1212      0.009     13.575
0.000         0.104     0.139
stock_EXPRESS SCRIPTS HOLDING CO            0.0745      0.010      7.775
0.000         0.056     0.093
stock_EXXON MOBIL CORPORATION               0.1865      0.021      8.679
0.000         0.144     0.229
```

stock_FACEBOOK INC			0.2171	0.019	11.586
0.000	0.180	0.254			
stock_FEDEX CORPORATION			0.1002	0.010	10.231
0.000	0.081	0.119			
stock_FIRST SOLAR INC			0.1560	0.017	9.046
0.000	0.122	0.190			
stock_FORD MOTOR CO(NEW)			0.0733	0.011	6.384
0.000	0.051	0.096			
stock_FREEPORT-MCMORAN INC			0.0974	0.013	7.766
0.000	0.073	0.122			
stock_GENERAL ELECTRIC CO			0.1500	0.018	8.332
0.000	0.115	0.185			
stock_GENERAL MOTORS CO			0.1448	0.018	8.032
0.000	0.109	0.180			
stock_GILEAD SCIENCES INC			0.1445	0.011	12.934
0.000	0.123	0.166			
stock_GOLDMAN SACHS GROUP INC			0.2410	0.025	9.822
0.000	0.193	0.289			
stock_GOPRO INC			0.1946	0.018	10.636
0.000	0.159	0.230			
stock_HALLIBURTON CO (HOLDING CO)			0.1050	0.010	10.582
0.000	0.086	0.124			
stock_HOME DEPOT INC			0.1160	0.011	10.522
0.000	0.094	0.138			
stock_HP INC			0.0879	0.010	8.554
0.000	0.068	0.108			
stock_INTEL CORP			0.1170	0.011	10.511
0.000	0.095	0.139			
stock_INTL BUSINESS MACHINES CORP			0.1135	0.012	9.176
0.000	0.089	0.138			
stock_JOHNSON AND JOHNSON			0.1375	0.013	10.821
0.000	0.113	0.162			
stock_JPMORGAN CHASE & CO COM STK			0.3443	0.064	5.416
0.000	0.220	0.469			
stock_KEURIG GREEN MOUNTAIN INC			0.1655	0.013	12.506
0.000	0.140	0.191			
stock_KINDER MORGAN INC			0.1372	0.017	7.888
0.000	0.103	0.171			
stock_LAS VEGAS SANDS CORP			0.1657	0.017	9.963
0.000	0.133	0.198			
stock_LILLY ELI & CO			0.1296	0.011	11.275
0.000	0.107	0.152			
stock_LINKEDIN CORP			0.1981	0.018	10.815
0.000	0.162	0.234			
stock_LOWES COMPANIES INC			0.1054	0.010	10.827
0.000	0.086	0.125			
stock_LYONDELLBASELL INDUSTRIES NV			0.1509	0.017	8.959

```
0.000         0.118     0.184
stock_MARATHON PETROLEUM CORP                  0.1477      0.018      8.006
0.000         0.112     0.184
stock_MASTERCARD INCORPORATED                  0.2324      0.018     12.916
0.000         0.197     0.268
stock_MCDONALDS CORP                           0.1322      0.013     10.363
0.000         0.107     0.157
stock_MEDTRONIC PLC                            0.1496      0.011     13.103
0.000         0.127     0.172
stock_MERCK & CO INC                           0.1372      0.012     11.332
0.000         0.113     0.161
stock_METLIFE  INC                             0.1720      0.025      6.765
0.000         0.122     0.222
stock_MICHAEL KORS HOLDINGS LTD                0.2136      0.019     11.178
0.000         0.176     0.251
stock_MICRON TECHNOLOGY INC                    0.1122      0.010     11.094
0.000         0.092     0.132
stock_MICROSOFT CORP                           0.1369      0.012     11.104
0.000         0.113     0.161
stock_MONDELEZ INTERNATIONAL INC               0.1580      0.015     10.549
0.000         0.129     0.187
stock_MONSANTO COMPANY                         0.1725      0.015     11.144
0.000         0.142     0.203
stock_MORGAN STANLEY                           0.2094      0.023      9.058
0.000         0.164     0.255
stock_MYLAN NV                                 0.1362      0.011     12.231
0.000         0.114     0.158
stock_NATIONAL OILWELL VARCO  INC.             0.1420      0.016      8.817
0.000         0.110     0.174
stock_NETFLIX INC                              0.2052      0.016     12.512
0.000         0.173     0.237
stock_NEWMONT MINING CORP (HOLDING COMPANY)    0.0949      0.012      7.982
0.000         0.072     0.118
stock_NEWS CP - CL A                           0.1394      0.013     11.131
0.000         0.115     0.164
stock_NIKE INC CL-B                            0.1440      0.012     12.236
0.000         0.121     0.167
stock_OCCIDENTAL PETROLEUM CORP                0.1303      0.012     10.690
0.000         0.106     0.154
stock_ORACLE CORP                              0.1384      0.012     11.935
0.000         0.116     0.161
stock_PANDORA MEDIA INC                        0.2348      0.020     11.961
0.000         0.196     0.273
stock_PENNEY J.C. CO INC (HOLDING COMPANY)     0.0553      0.010      5.558
0.000         0.036     0.075
stock_PEPSICO INC                              0.1325      0.013      9.962
0.000         0.106     0.159
```

```
stock_PFIZER INC                                0.1554      0.013     11.701
0.000         0.129     0.181
stock_PHILIP MORRIS INTERNATIONAL INC           0.2017      0.019     10.405
0.000         0.164     0.240
stock_PIONEER NAT RES CO                        0.1980      0.014     13.850
0.000         0.170     0.226
stock_PRECISION CASTPARTS CORP                  0.1519      0.014     11.190
0.000         0.125     0.179
stock_PROCTER & GAMBLE CO                       0.1412      0.014     10.286
0.000         0.114     0.168
stock_QUALCOMM INC                              0.1585      0.013     12.011
0.000         0.133     0.184
stock_REGENERON PHARMACEUTICALS INC             0.1549      0.018      8.846
0.000         0.121     0.189
stock_SALESFORCE.COM INC                        0.1797      0.016     10.945
0.000         0.148     0.212
stock_SALIX PHARMACEUTICALS LTD              1.482e-11   1.92e-11      0.771
0.440     -2.28e-11  5.25e-11
stock_SANDISK CORP                              0.1842      0.014     13.280
0.000         0.157     0.211
stock_SCHLUMBERGER LTD.                         0.1437      0.013     11.126
0.000         0.118     0.169
stock_SKYWORKS SOLUTIONS INC                    0.1829      0.016     11.217
0.000         0.151     0.215
stock_SOLARCITY CORP                            0.2218      0.019     11.593
0.000         0.184     0.259
stock_SOUTHWEST AIRLINES CO                     0.1205      0.010     11.615
0.000         0.100     0.141
stock_STARBUCKS CORPORATION                     0.1770      0.013     13.501
0.000         0.151     0.203
stock_SUNEDISON INC                             0.1297      0.042      3.095
0.002         0.048     0.212
stock_TARGET CORPORATION                        0.1226      0.014      8.556
0.000         0.094     0.151
stock_TESLA INC                                 0.2039      0.018     11.215
0.000         0.168     0.240
stock_TEXAS INSTRUMENTS INC                     0.1730      0.014     12.080
0.000         0.145     0.201
stock_TIME WARNER CABLE INC                     0.1446      0.016      9.138
0.000         0.114     0.176
stock_TIME WARNER INC.                          0.0757      0.007     11.026
0.000         0.062     0.089
stock_TWITTER INC                               0.1978      0.019     10.476
0.000         0.161     0.235
stock_UNION PACIFIC CORPORATION                 0.1502      0.013     11.655
0.000         0.125     0.175
stock_UNITED CONTINENTAL HOLDINGS IN            0.1245      0.016      7.991
```

```
0.000         0.094     0.155
stock_UNITED PARCEL SERVICE INC.CL B           0.1248      0.015      8.288
0.000         0.095     0.154
stock_UNITED TECHNOLOGIES CORP                 0.1288      0.013      9.937
0.000         0.103     0.154
stock_UNITEDHEALTH GROUP INC                   0.1402      0.013     10.559
0.000         0.114     0.166
stock_VALERO ENERGY CORP (NEW)                 0.1296      0.013      9.686
0.000         0.103     0.156
stock_VERIZON COMMUNICATIONS                   0.1198      0.016      7.496
0.000         0.088     0.151
stock_VISA INC                                 0.2206      0.017     12.867
0.000         0.187     0.254
stock_WALGREENS BOOTS ALLIANCE INC             0.1369      0.013     10.559
0.000         0.111     0.162
stock_WALMART INC                              0.1522      0.021      7.225
0.000         0.111     0.194
stock_WALT DISNEY CO                           0.1309      0.010     13.023
0.000         0.111     0.151
stock_WELLS FARGO & CO(NEW)                    0.2969      0.043      6.946
0.000         0.213     0.381
stock_WILLIAMS COMPANIES                       0.1325      0.013     10.104
0.000         0.107     0.158
stock_WYNN RESORTS LTD                         0.1418      0.016      8.999
0.000         0.111     0.173
stock_YELP INC                                 0.1947      0.019     10.257
0.000         0.157     0.232
==============================================================================
Omnibus:                        0.654   Durbin-Watson:                   1.503
Prob(Omnibus):                  0.721   Jarque-Bera (JB):                0.669
Skew:                          -0.007   Prob(JB):                        0.716
Kurtosis:                       2.986   Cond. No.                     6.39e+16
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 6.39e+16. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

[47]:
```
target = 'Returns20D'
model_data = pd.concat([y[[target]], X], axis=1).dropna()
model_data = model_data[model_data[target].between(model_data[target].quantile(.
↪025),
                                                   model_data[target].quantile(.
↪975))]
```

```
model = OLS(endog=model_data[target], exog=model_data.drop(target, axis=1))
trained_model = model.fit()
trained_model.summary()
```

[47]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:             Returns20D   R-squared:                       0.072
Model:                            OLS   Adj. R-squared:                  0.068
Method:                 Least Squares   F-statistic:                     19.10
Date:                Sun, 09 Sep 2018   Prob (F-statistic):               0.00
Time:                        21:18:04   Log-Likelihood:                 62705.
No. Observations:               42460   AIC:                        -1.251e+05
Df Residuals:                   42288   BIC:                        -1.236e+05
Df Model:                         171
Covariance Type:            nonrobust
==============================================================================
                                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
DividendYield                1.193e-05   2.23e-06      5.351      0.000      7.56e-06  1.63e-05
EBITDAYield                    -0.0001   4.51e-05     -3.287      0.001        -0.000 -5.98e-05
EVToEBITDA                      0.0003   4.07e-05      7.808      0.000         0.000     0.000
EVToFCF                     -3.67e-06   4.99e-05     -0.074      0.941        -0.000  9.42e-05
PriceToBook                  1.86e-05   4.73e-05      0.393      0.694     -7.41e-05     0.000
PriceToDilutedEarningsTTM   -5.406e-05   2.12e-05     -2.549      0.011     -9.56e-05 -1.25e-05
PriceToEarningsTTM           -8.22e-07   9.41e-07     -0.873      0.382     -2.67e-06  1.02e-06
PriceToFCF                   2.678e-06   4.12e-05      0.065      0.948      -7.8e-05  8.33e-05
PriceToOperatingCashflow       -0.0001   2.91e-05     -4.551      0.000        -0.000 -7.54e-05
PriceToSalesTTM             -4.302e-05    2.5e-06    -17.222      0.000     -4.79e-05 -3.81e-05
Directional Movement Index  -3.389e-05   9.51e-06     -3.566      0.000     -5.25e-05 -1.53e-05
Money Flow Index             2.884e-05   1.23e-05      2.336
```

```
0.019      4.64e-06   5.3e-05
Percent Above Low                          -7.04e-05   2.47e-05     -2.855
0.004        -0.000 -2.21e-05
Percent Below High                          4.825e-05   1.89e-05      2.559
0.010      1.13e-05  8.52e-05
Price Oscillator                           -5.428e-06   1.33e-05     -0.407
0.684     -3.15e-05  2.07e-05
Trendline                                   4.502e-05   2.04e-05      2.204
0.028      4.98e-06  8.51e-05
AssetToEquityRatio                         -8.615e-06   1.86e-06     -4.637
0.000     -1.23e-05 -4.97e-06
AssetTurnover                                -0.0002   8.11e-05     -2.720
0.007        -0.000 -6.16e-05
CurrentRatio                                1.14e-05    2.4e-06      4.756
0.000       6.7e-06  1.61e-05
DebtToAssetRatio                            2.131e-06    1.5e-06      1.421
0.155     -8.08e-07  5.07e-06
DebtToEquityRatio                           4.261e-06   1.74e-06      2.445
0.014      8.45e-07  7.68e-06
MertonsDD                                    -0.0013      0.000     -6.145
0.000        -0.002    -0.001
WorkingCapitalToAssets                       5.08e-06   2.89e-06      1.759
0.079      -5.8e-07  1.07e-05
WorkingCapitalToSales                        -0.0004   9.31e-05     -4.593
0.000        -0.001    -0.000
Dividend Growth                            -3.037e-06   9.15e-07     -3.319
0.001     -4.83e-06 -1.24e-06
EPS                                         2.228e-06   1.41e-06      1.583
0.113      -5.3e-07  4.99e-06
Net Debt                                   -2.368e-13   5.14e-14     -4.609
0.000     -3.37e-13 -1.36e-13
Sales                                      -2.182e-13   5.03e-14     -4.341
0.000     -3.17e-13  -1.2e-13
Total Assets                               -2.513e-13   3.73e-14     -6.741
0.000     -3.24e-13 -1.78e-13
EPS Growth 3M                              -1.786e-05   2.15e-06     -8.320
0.000     -2.21e-05 -1.37e-05
EPS Growth 12M                              2.263e-06   2.15e-06      1.053
0.292     -1.95e-06  6.47e-06
Net Debt Growth 3M                          -4.19e-06   2.73e-06     -1.533
0.125     -9.55e-06  1.17e-06
Net Debt Growth 12M                         1.181e-06   2.76e-06      0.428
0.669     -4.23e-06  6.59e-06
Sales Growth 3M                            -7.312e-07   2.88e-06     -0.254
0.800     -6.38e-06  4.92e-06
Sales Growth 12M                           -1.036e-05   2.94e-06     -3.525
0.000     -1.61e-05  -4.6e-06
```

```
Total Assets Growth 3M                    7.478e-06   3.24e-06      2.310
0.021      1.13e-06  1.38e-05
Total Assets Growth 12M                  -3.848e-06   3.44e-06     -1.120
0.263     -1.06e-05  2.89e-06
CFO To Assets                             9.296e-05   3.95e-05      2.352
0.019      1.55e-05     0.000
Capex To Assets                             -0.0002   8.77e-05     -1.768
0.077        -0.000  1.68e-05
Capex To FCF                                 0.0002   4.86e-05      3.199
0.001      6.03e-05     0.000
Capex To Sales                               0.0003   9.32e-05      2.992
0.003      9.62e-05     0.000
EBIT To Assets                            4.836e-06      4e-05      0.121
0.904     -7.35e-05  8.32e-05
Retained Earnings To Assets                 -0.0010   8.46e-05    -12.222
0.000        -0.001    -0.001
Downside Risk                            -2.381e-05   2.45e-05     -0.972
0.331     -7.18e-05  2.42e-05
Index Beta                               -4.429e-06   5.04e-07     -8.793
0.000     -5.42e-06 -3.44e-06
Log Market Cap                               0.0001   8.62e-05      1.329
0.184     -5.44e-05     0.000
Volatility 3M                               -0.0002   2.08e-05     -7.260
0.000        -0.000    -0.000
stock_3D SYSTEMS CORP                        0.3173      0.020     15.578
0.000         0.277     0.357
stock_3M COMPANY                             0.3182      0.018     17.973
0.000         0.284     0.353
stock_ABBOTT LABORATORIES                    0.2168      0.014     14.970
0.000         0.188     0.245
stock_ABBVIE INC                             0.3427      0.026     13.306
0.000         0.292     0.393
stock_ALLERGAN INC                           0.3183      0.014     23.540
0.000         0.292     0.345
stock_ALLERGAN PLC                           0.3267      0.019     17.361
0.000         0.290     0.364
stock_ALTABA INC                             0.3407      0.021     16.280
0.000         0.300     0.382
stock_ALTRIA GROUP INC.                      0.3366      0.018     18.421
0.000         0.301     0.372
stock_AMAZON.COM INC                         0.2840      0.021     13.457
0.000         0.243     0.325
stock_AMERICAN AIRLINES GROUP INC            0.2416      0.024     10.047
0.000         0.194     0.289
stock_AMERICAN EXPRESS COMPANY               0.1996      0.011     17.471
0.000         0.177     0.222
stock_AMERICAN INTL GROUP INC                0.2697      0.020     13.737
```

```
0.000         0.231     0.308
stock_AMGEN INC                                0.2179      0.012     17.559
0.000         0.194     0.242
stock_ANADARKO PETROLEUM CORP                  0.2158      0.011     19.604
0.000         0.194     0.237
stock_APACHE CORP                              0.1494      0.056      2.651
0.008         0.039     0.260
stock_APPLE INC                                0.3383      0.018     19.158
0.000         0.304     0.373
stock_APPLIED MATERIALS INC                    0.2390      0.014     17.579
0.000         0.212     0.266
stock_ARCONIC INC                              0.1068      0.010     10.685
0.000         0.087     0.126
stock_AT&T INC. COM                            0.2514      0.020     12.594
0.000         0.212     0.291
stock_Alphabet Inc. Cl A                       0.4051      0.025     16.157
0.000         0.356     0.454
stock_BAKER HUGHES INC                         0.2139      0.012     17.586
0.000         0.190     0.238
stock_BANK OF AMERICA CORP                     0.7417      0.076      9.797
0.000         0.593     0.890
stock_BERKSHIRE HATHAWAY INC CL-B              0.3567      0.026     13.681
0.000         0.306     0.408
stock_BIOGEN INC                               0.3593      0.017     21.207
0.000         0.326     0.392
stock_BOEING CO                                0.2107      0.013     15.879
0.000         0.185     0.237
stock_BOOKING HOLDINGS INC                     0.3602      0.022     16.449
0.000         0.317     0.403
stock_BRISTOL MYERS SQUIBB COMPANY             0.3115      0.015     20.543
0.000         0.282     0.341
stock_BROADCOM CORP                            0.2813      0.020     14.099
0.000         0.242     0.320
stock_BROADCOM INC                             0.3808      0.026     14.895
0.000         0.331     0.431
stock_CATERPILLAR INC                          0.1730      0.013     13.255
0.000         0.147     0.199
stock_CELGENE CORP                             0.2960      0.015     20.285
0.000         0.267     0.325
stock_CHESAPEAKE ENERGY CORP                -4.999e-14   8.77e-14     -0.570
0.568     -2.22e-13  1.22e-13
stock_CHEVRON CORPORATION                      0.3485      0.027     13.048
0.000         0.296     0.401
stock_CISCO SYSTEMS INC                        0.2298      0.013     17.569
0.000         0.204     0.255
stock_CITIGROUP                                0.6481      0.065      9.957
0.000         0.521     0.776
```

```
stock_COCA-COLA CO                          0.2999      0.018     16.645
0.000         0.265     0.335
stock_COMCAST CORP                          0.2309      0.013     17.938
0.000         0.206     0.256
stock_CONOCOPHILLIPS                        0.2784      0.022     12.429
0.000         0.235     0.322
stock_COVIDIEN PLC                          0.3792      0.024     15.807
0.000         0.332     0.426
stock_CVS HEALTH CORP                       0.2568      0.017     15.384
0.000         0.224     0.289
stock_DEERE & CO                            0.1513      0.014     10.763
0.000         0.124     0.179
stock_DELTA AIR LINES INC                   0.2806      0.022     12.572
0.000         0.237     0.324
stock_DIRECTV                               0.2139      0.022      9.864
0.000         0.171     0.256
stock_DOLLAR GENERAL CORP                   0.2719      0.023     11.971
0.000         0.227     0.316
stock_DOW CHEMICAL CO                       0.1912      0.013     14.832
0.000         0.166     0.216
stock_E.I. Du Pont De Nemours A             0.1633      0.013     12.300
0.000         0.137     0.189
stock_EBAY INC                              0.3558      0.023     15.256
0.000         0.310     0.402
stock_EMC CORPORATION                       0.2526      0.013     19.131
0.000         0.227     0.278
stock_EOG RESOURCES INC                     0.2644      0.013     20.854
0.000         0.240     0.289
stock_EXPRESS SCRIPTS HOLDING CO            0.1640      0.014     11.874
0.000         0.137     0.191
stock_EXXON MOBIL CORPORATION               0.4155      0.031     13.352
0.000         0.355     0.477
stock_FACEBOOK INC                          0.3934      0.026     14.853
0.000         0.341     0.445
stock_FEDEX CORPORATION                     0.2296      0.014     16.546
0.000         0.202     0.257
stock_FIRST SOLAR INC                       0.3130      0.025     12.617
0.000         0.264     0.362
stock_FORD MOTOR CO(NEW)                    0.1766      0.017     10.680
0.000         0.144     0.209
stock_FREEPORT-MCMORAN INC                  0.1663      0.018      9.357
0.000         0.131     0.201
stock_GENERAL ELECTRIC CO                   0.3810      0.026     14.621
0.000         0.330     0.432
stock_GENERAL MOTORS CO                     0.2661      0.026     10.406
0.000         0.216     0.316
stock_GILEAD SCIENCES INC                   0.3193      0.016     20.030
```

```
0.000         0.288     0.351
stock_GOLDMAN SACHS GROUP INC                0.5289      0.035     14.945
0.000         0.460     0.598
stock_GOPRO INC                              0.3186      0.026     12.297
0.000         0.268     0.369
stock_HALLIBURTON CO (HOLDING CO)            0.2384      0.014     16.933
0.000         0.211     0.266
stock_HOME DEPOT INC                         0.2761      0.016     17.535
0.000         0.245     0.307
stock_HP INC                                 0.2021      0.015     13.749
0.000         0.173     0.231
stock_INTEL CORP                             0.2633      0.016     16.526
0.000         0.232     0.294
stock_INTL BUSINESS MACHINES CORP            0.2830      0.018     15.965
0.000         0.248     0.318
stock_JOHNSON AND JOHNSON                    0.3200      0.018     17.592
0.000         0.284     0.356
stock_JPMORGAN CHASE & CO COM STK            0.8913      0.092      9.698
0.000         0.711     1.071
stock_KEURIG GREEN MOUNTAIN INC              0.3387      0.019     18.109
0.000         0.302     0.375
stock_KINDER MORGAN INC                      0.2413      0.025      9.811
0.000         0.193     0.289
stock_LAS VEGAS SANDS CORP                   0.2950      0.024     12.543
0.000         0.249     0.341
stock_LILLY ELI & CO                         0.2792      0.016     16.965
0.000         0.247     0.311
stock_LINKEDIN CORP                          0.3443      0.026     13.316
0.000         0.294     0.395
stock_LOWES COMPANIES INC                    0.2297      0.014     16.545
0.000         0.202     0.257
stock_LYONDELLBASELL INDUSTRIES NV           0.2773      0.024     11.680
0.000         0.231     0.324
stock_MARATHON PETROLEUM CORP                0.2467      0.026      9.411
0.000         0.195     0.298
stock_MASTERCARD INCORPORATED                0.4555      0.025     17.920
0.000         0.406     0.505
stock_MCDONALDS CORP                         0.2920      0.018     16.033
0.000         0.256     0.328
stock_MEDTRONIC PLC                          0.3192      0.016     19.660
0.000         0.287     0.351
stock_MERCK & CO INC                         0.3003      0.017     17.338
0.000         0.266     0.334
stock_METLIFE  INC                           0.3781      0.037     10.338
0.000         0.306     0.450
stock_MICHAEL KORS HOLDINGS LTD              0.3756      0.027     13.904
0.000         0.323     0.429
```

```
stock_MICRON TECHNOLOGY INC                    0.2171      0.014     15.211
0.000         0.189     0.245
stock_MICROSOFT CORP                           0.3047      0.018     17.242
0.000         0.270     0.339
stock_MONDELEZ INTERNATIONAL INC               0.2891      0.021     13.662
0.000         0.248     0.331
stock_MONSANTO COMPANY                         0.3312      0.022     15.138
0.000         0.288     0.374
stock_MORGAN STANLEY                           0.4563      0.033     13.722
0.000         0.391     0.521
stock_MYLAN NV                                 0.2558      0.016     16.113
0.000         0.225     0.287
stock_NATIONAL OILWELL VARCO  INC.             0.2658      0.023     11.681
0.000         0.221     0.310
stock_NETFLIX INC                              0.3875      0.023     16.682
0.000         0.342     0.433
stock_NEWMONT MINING CORP (HOLDING COMPANY)    0.1446      0.017      8.514
0.000         0.111     0.178
stock_NEWS CP - CL A                           0.2669      0.018     15.067
0.000         0.232     0.302
stock_NIKE INC CL-B                            0.3016      0.017     18.078
0.000         0.269     0.334
stock_OCCIDENTAL PETROLEUM CORP                0.2756      0.017     15.892
0.000         0.242     0.310
stock_ORACLE CORP                              0.2984      0.017     18.058
0.000         0.266     0.331
stock_PANDORA MEDIA INC                        0.3925      0.028     14.067
0.000         0.338     0.447
stock_PENNEY J.C. CO INC (HOLDING COMPANY)     0.0737      0.014      5.167
0.000         0.046     0.102
stock_PEPSICO INC                              0.2931      0.019     15.449
0.000         0.256     0.330
stock_PFIZER INC                               0.3336      0.019     17.563
0.000         0.296     0.371
stock_PHILIP MORRIS INTERNATIONAL INC          0.3844      0.028     13.959
0.000         0.330     0.438
stock_PIONEER NAT RES CO                       0.3579      0.020     17.747
0.000         0.318     0.397
stock_PRECISION CASTPARTS CORP                 0.3008      0.020     14.774
0.000         0.261     0.341
stock_PROCTER & GAMBLE CO                      0.3094      0.020     15.745
0.000         0.271     0.348
stock_QUALCOMM INC                             0.3232      0.019     17.225
0.000         0.286     0.360
stock_REGENERON PHARMACEUTICALS INC            0.3153      0.043      7.267
0.000         0.230     0.400
stock_SALESFORCE.COM INC                       0.3156      0.023     13.571
```

```
0.000         0.270     0.361
stock_SALIX PHARMACEUTICALS LTD           -5.165e-15   7.41e-15     -0.697
0.486     -1.97e-14  9.37e-15
stock_SANDISK CORP                            0.3350      0.020     17.064
0.000         0.296     0.373
stock_SCHLUMBERGER LTD.                       0.3011      0.018     16.396
0.000         0.265     0.337
stock_SKYWORKS SOLUTIONS INC                  0.3422      0.023     14.843
0.000         0.297     0.387
stock_SOLARCITY CORP                          0.3803      0.027     14.070
0.000         0.327     0.433
stock_SOUTHWEST AIRLINES CO                   0.2513      0.015     16.796
0.000         0.222     0.281
stock_STARBUCKS CORPORATION                   0.3558      0.019     19.139
0.000         0.319     0.392
stock_SUNEDISON INC                         1.713e-16   2.85e-16      0.601
0.548     -3.87e-16   7.3e-16
stock_TARGET CORPORATION                      0.2298      0.020     11.314
0.000         0.190     0.270
stock_TESLA INC                               0.3540      0.026     13.789
0.000         0.304     0.404
stock_TEXAS INSTRUMENTS INC                   0.3573      0.021     17.423
0.000         0.317     0.397
stock_TIME WARNER CABLE INC                   0.2637      0.022     11.821
0.000         0.220     0.307
stock_TIME WARNER INC.                        0.1773      0.010     18.052
0.000         0.158     0.197
stock_TWITTER INC                             0.3333      0.027     12.463
0.000         0.281     0.386
stock_UNION PACIFIC CORPORATION               0.3164      0.018     17.283
0.000         0.281     0.352
stock_UNITED CONTINENTAL HOLDINGS IN          0.1975      0.022      8.977
0.000         0.154     0.241
stock_UNITED PARCEL SERVICE INC.CL B          0.2391      0.021     11.199
0.000         0.197     0.281
stock_UNITED TECHNOLOGIES CORP                0.2750      0.018     14.928
0.000         0.239     0.311
stock_UNITEDHEALTH GROUP INC                  0.2970      0.019     15.679
0.000         0.260     0.334
stock_VALERO ENERGY CORP (NEW)                0.2539      0.019     13.356
0.000         0.217     0.291
stock_VERIZON COMMUNICATIONS                  0.2472      0.023     10.825
0.000         0.202     0.292
stock_VISA INC                                0.4192      0.024     17.309
0.000         0.372     0.467
stock_WALGREENS BOOTS ALLIANCE INC            0.2793      0.018     15.194
0.000         0.243     0.315
```

```
stock_WALMART INC                          0.3251      0.031     10.594
0.000         0.265     0.385
stock_WALT DISNEY CO                       0.3131      0.014     21.827
0.000         0.285     0.341
stock_WELLS FARGO & CO(NEW)                0.7175      0.062     11.586
0.000         0.596     0.839
stock_WILLIAMS COMPANIES                   0.2522      0.019     13.501
0.000         0.216     0.289
stock_WYNN RESORTS LTD                     0.2461      0.022     11.047
0.000         0.202     0.290
stock_YELP INC                             0.3195      0.027     11.882
0.000         0.267     0.372
==============================================================================
Omnibus:                       26.157   Durbin-Watson:                   1.605
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               26.187
Skew:                           0.059   Prob(JB):                     2.06e-06
Kurtosis:                       2.970   Cond. No.                     1.95e+19
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 9.91e-12. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
"""
```

1.12 Linear Models for Prediction: sklearn

Since sklearn is tailored towards prediction, we will evaluate the linear regression model based on its predictive performance using cross-validation.

1.12.1 Custom Time Series Cross-Validation

Our data consists of grouped time series data that requires a custom cross-validation function to provide the train and test indices that ensure that the test data immediately follows the training data for each equity and we do not inadvertently create a look-ahead bias or leakage.

We can achieve this using the following function that returns a generator yielding pairs of train and test dates. The set of train dates that ensure a minimum length of the training periods. The number of pairs depends on the parameter nfolds. The distinct test periods do not overlap and are located at the end of the period available in the data. After a test period is used, it becomes part of the training data that grow in size accordingly:

[158]:
```python
def time_series_split(d=model_data, nfolds=5, min_train=21):
    """Generate train/test dates for nfolds
    with at least min_train train obs
    """
    train_dates = d[:min_train].tolist()
```

```
    n = int(len(dates)/(nfolds + 1)) + 1
    test_folds = [d[i:i + n] for i in range(min_train, len(d), n)]
    for test_dates in test_folds:
        if len(train_dates) > min_train:
            yield train_dates, test_dates
        train_dates.extend(test_dates)
```

1.12.2 Select Features and Target

We need to select the appropriate return series (we will again use a 10-day holding period) and remove outliers. We will also convert returns to log returns as follows:

[49]:
```
target = 'Returns10D'
outliers = .01
model_data = pd.concat([y[[target]], X], axis=1).dropna().reset_index('asset',␣
 ↪drop=True)
model_data = model_data[model_data[target].between(*model_data[target].
 ↪quantile([outliers, 1-outliers]).values)]

model_data[target] = np.log1p(model_data[target])
features = model_data.drop(target, axis=1).columns
dates = model_data.index.unique()

print(model_data.info())
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 45114 entries, 2014-01-02 to 2015-12-16
Columns: 183 entries, Returns10D to stock_YELP INC
dtypes: float64(183)
memory usage: 63.3 MB
None
```

[50]:
```
model_data[target].describe()
```

[50]:
```
count    45114.000000
mean         0.001159
std          0.045740
min         -0.157448
25%         -0.025013
50%          0.002817
75%          0.028880
max          0.146139
Name: Returns10D, dtype: float64
```

[51]:
```
idx = pd.IndexSlice
```

1.13 OLS Linear Regression

We will use 250 folds to generally predict about 2 days of forward returns following the historical training data that will gradually increase in length.

Each iteration obtains the appropriate training and test dates from our custom cross-validation function, selects the corresponding features and targets, and then trains and predicts accordingly.

We capture the root mean squared error as well as the Spearman rank correlation between actual and predicted values:

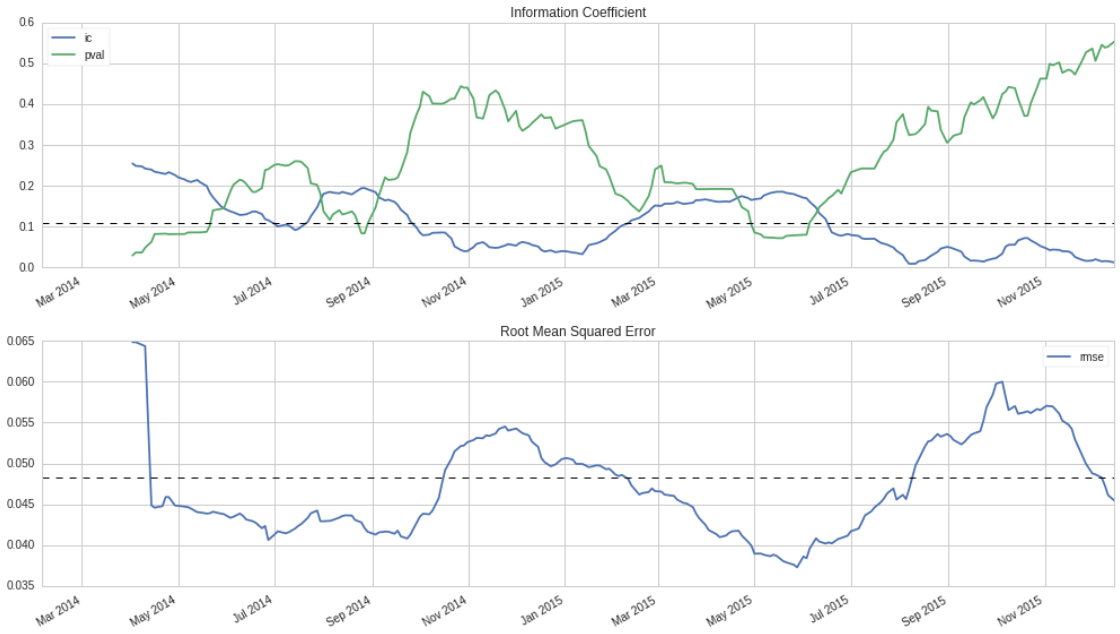
[52]:
```
nfolds = 250
lr = LinearRegression()

test_results, result_idx, preds = [], [], pd.DataFrame()
for train_dates, test_dates in time_series_split(dates, nfolds=nfolds):

    X_train = model_data.loc[idx[train_dates], features]
    y_train = model_data.loc[idx[train_dates], target]
    lr.fit(X=X_train, y=y_train)

    X_test = model_data.loc[idx[test_dates], features]
    y_test = model_data.loc[idx[test_dates], target]
    y_pred = lr.predict(X_test)

    rmse = np.sqrt(mean_squared_error(y_pred=y_pred, y_true=y_test))
    ic, pval = spearmanr(y_pred, y_test)

    test_results.append([rmse, ic, pval])
    preds = preds.append(y_test.to_frame('actuals').assign(predicted=y_pred))
    result_idx.append(train_dates[-1])
```

[53]:
```
test_result = pd.DataFrame(test_results, columns=['rmse', 'ic', 'pval'],␣
↪index=result_idx)
```

1.13.1 Results

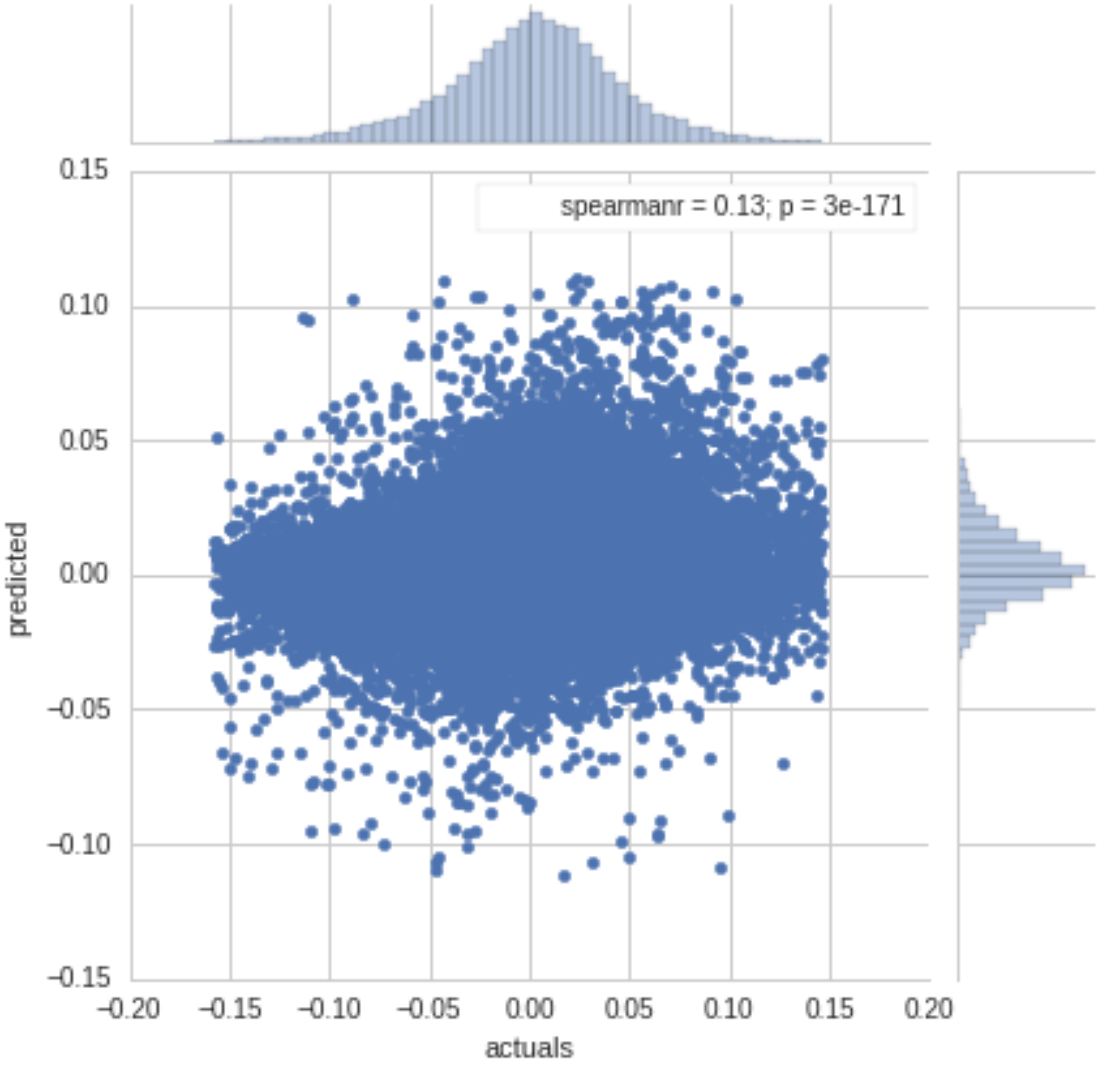
We have captured the test predictions from the 250 folds and can compute both the overall and a 21-day rolling average:

[54]:
```
fig, axes = plt.subplots(nrows=2)
rolling_result = test_result.rolling(21).mean()
rolling_result[['ic', 'pval']].plot(ax=axes[0], title='Information Coefficient')
axes[0].axhline(test_result.ic.mean(), lw=1, ls='--', color='k')
rolling_result[['rmse']].plot(ax=axes[1], title='Root Mean Squared Error')
axes[1].axhline(test_result.rmse.mean(), lw=1, ls='--', color='k')
plt.tight_layout();
```



For the entire period, we see that the Information Coefficient measured by the rank correlation of actual and predicted returns is weakly positive and statistically significant:

[55]:
```
preds_cleaned = preds[(preds.predicted.between(*preds.predicted.quantile([.001,␣
↪.999]).values))]
sns.jointplot(x='actuals', y='predicted', data=preds_cleaned,␣
↪stat_func=spearmanr);
```



1.14 Regularization

For the ridge regression, we need to tune the regularization parameter with the keyword alpha that corresponds to the we used previously. We will try 21 values from 10-5 to 105 in logarithmic steps.

1.14.1 Ridge Regression: L2 Penalty

The scale sensitivity of the ridge penalty requires us to standardize the inputs using the Standard-Scaler. Note that we always learn the mean and the standard deviation from the training set using the .fit_transform() method and then apply these learned parameters to the test set using the .transform() method.

[56]:
```
nfolds = 250
alphas = np.logspace(-5, 5, 11)
```

```
scaler = StandardScaler()

ridge_result, ridge_coeffs = pd.DataFrame(), pd.DataFrame()
for i, alpha in enumerate(alphas):
    print i,
    coeffs, test_results = [], []
    lr_ridge = Ridge(alpha=alpha)
    for train_dates, test_dates in time_series_split(dates, nfolds=nfolds):

        X_train = model_data.loc[idx[train_dates], features]
        y_train = model_data.loc[idx[train_dates], target]
        lr_ridge.fit(X=scaler.fit_transform(X_train), y=y_train)
        coeffs.append(lr_ridge.coef_)

        X_test = model_data.loc[idx[test_dates], features]
        y_test = model_data.loc[idx[test_dates], target]
        y_pred = lr_ridge.predict(scaler.transform(X_test))

        rmse = np.sqrt(mean_squared_error(y_pred=y_pred, y_true=y_test))
        ic, pval = spearmanr(y_pred, y_test)

        test_results.append([train_dates[-1], rmse, ic, pval, alpha])
    test_results = pd.DataFrame(test_results, columns=['date', 'rmse', 'ic',␣
↪'pval', 'alpha'])
    ridge_result = ridge_result.append(test_results)
    ridge_coeffs[alpha] = np.mean(coeffs, axis=0)
```

```
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
```

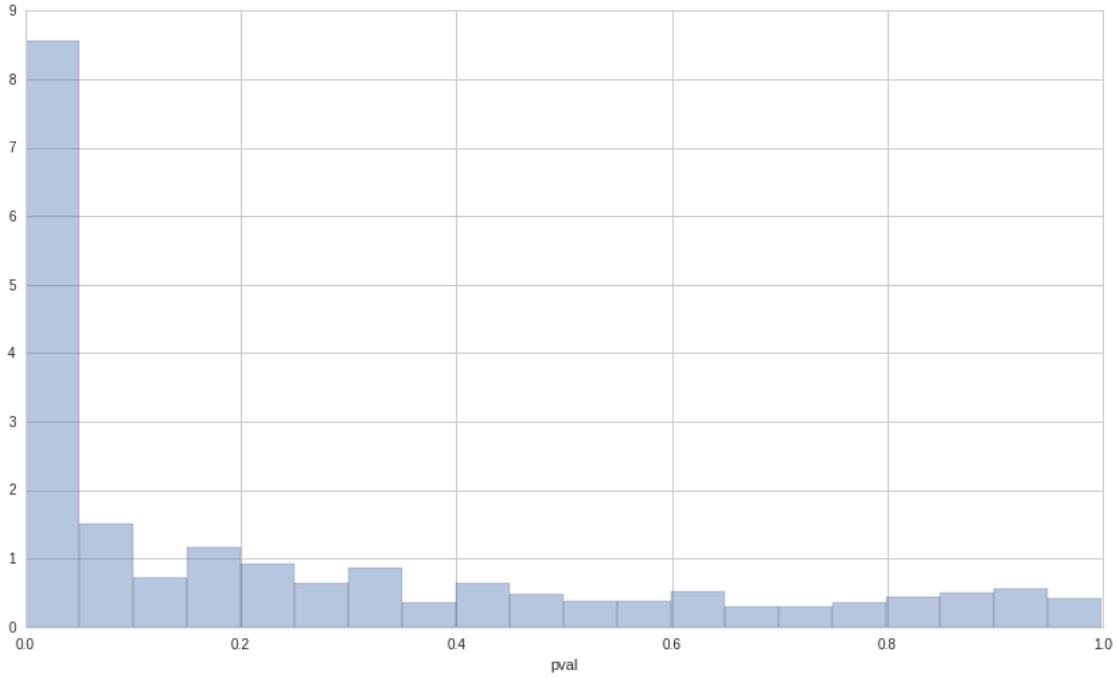
[82]:
```
ridge_result.describe()
```

[82]:

	rmse	ic	pval	alpha
count	4956.000000	4956.000000	4.956000e+03	4.956000e+03
mean	0.046201	0.095743	2.462496e-01	5.291005e+08
std	0.018699	0.148136	2.990978e-01	2.128608e+09
min	0.028501	-0.422446	1.374229e-16	1.000000e-10
25%	0.038269	-0.005162	4.278807e-03	1.000000e-05
50%	0.043257	0.095800	9.506434e-02	1.000000e+00
75%	0.050236	0.201505	4.138598e-01	1.000000e+05
max	0.466332	0.576016	9.991168e-01	1.000000e+10

1.14.2 Significance of Information Coefficients - p-value Distribution

[91]:
```
sns.distplot(ridge_result.pval, bins=20, norm_hist=True, kde=False);
```



[109]:
```
ridge_result_sig = ridge_result[(ridge_result.pval < .05) & (ridge_result.alpha.
↪between(10**-5, 10**5))]
ridge_result_sig_alpha = ridge_result_sig.groupby('alpha')
```

[115]:
```
ridge_coeffs_main = ridge_coeffs.filter(ridge_result_sig.alpha.unique())
```

1.14.3 Ridge Path

We can now plot the information coefficient obtained for each hyperparameter value and also visualize how the coefficient values evolve as the regularization increases. The results show that we get the highest IC value for a value of =10. For this level of regularization, the right-hand panel reveals that the coefficients have been already significantly shrunk compared to the (almost) unconstrained model with =10-5:

[172]:
```
ridge_result.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4956 entries, 0 to 235
Data columns (total 5 columns):
date     4956 non-null datetime64[ns, UTC]
rmse     4956 non-null float64
ic       4956 non-null float64
pval     4956 non-null float64
alpha    4956 non-null float64
```

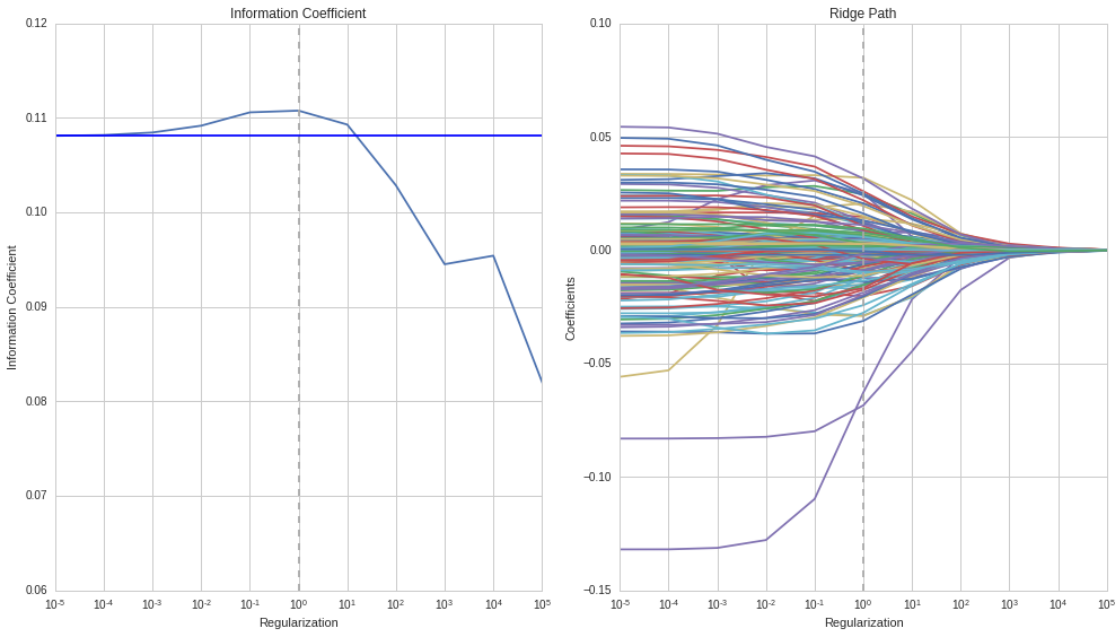
```
dtypes: datetime64[ns, UTC](1), float64(4)
memory usage: 232.3 KB
```

[103]:
```
best_ic = ridge_result_sig_alpha['ic'].mean().max()
best_alpha = ridge_result_sig_alpha['ic'].mean().idxmax()
```

[176]:
```

```

[176]:
```
1.0
```

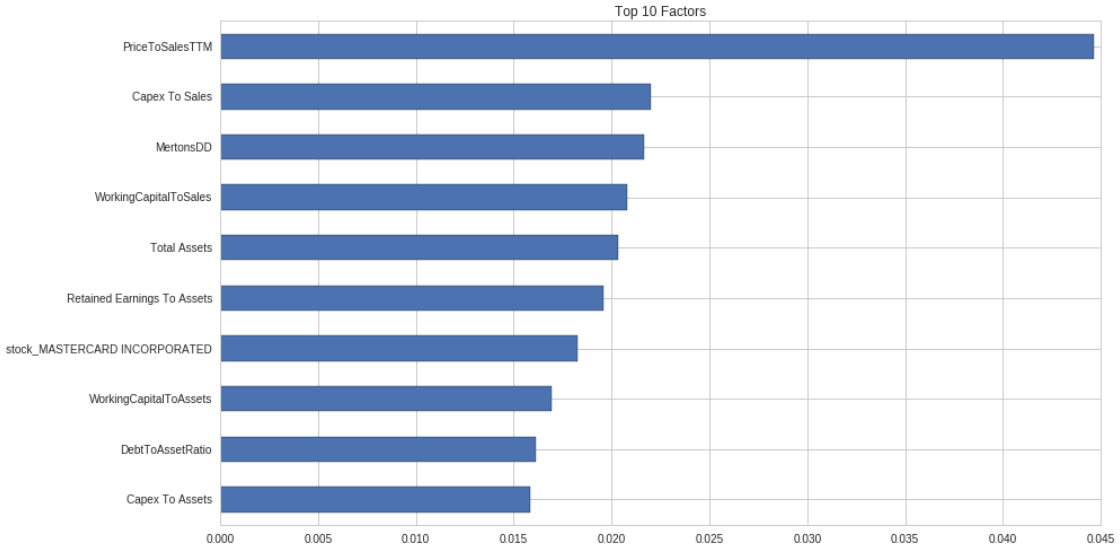
[178]:
```
fig, axes = plt.subplots(ncols=2, sharex=True)

ridge_result.groupby('alpha')['ic'].mean().plot(logx=True, title='Information␣
↪Coefficient', ax=axes[0])
axes[0].axhline(ridge_result.groupby('alpha').ic.mean().median())
axes[0].axvline(x=ridge_result.groupby('alpha').ic.mean().idxmax(),␣
↪c='darkgrey', ls='--')
axes[0].set_xlabel('Regularization')
axes[0].set_ylabel('Information Coefficient')

ridge_coeffs_main.T.plot(legend=False, logx=True, title='Ridge Path',␣
↪ax=axes[1])
axes[1].set_xlabel('Regularization')
axes[1].set_ylabel('Coefficients')
axes[1].axvline(x=ridge_result.groupby('alpha').ic.mean().idxmax(),␣
↪c='darkgrey', ls='--')
fig.tight_layout();
```



1.14.4 Top 10 Coefficients

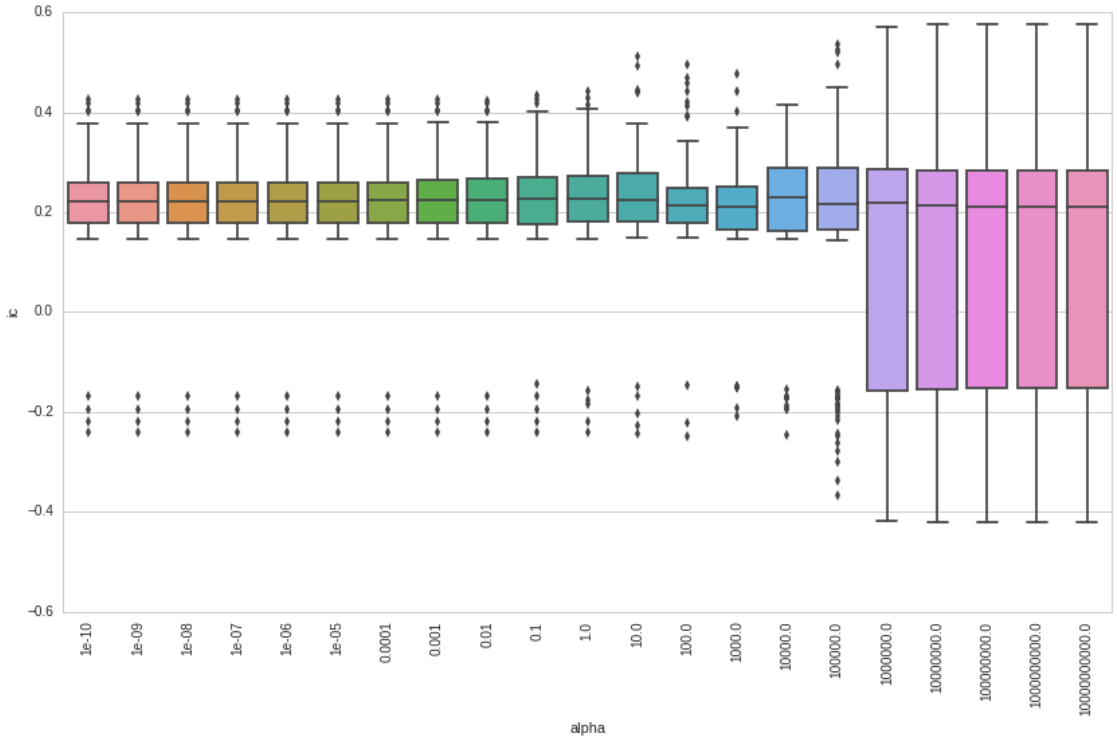
The standardization of the coefficients allows us to draw conclusions about their relative importance by comparing their absolute magnitude. The 10 most relevant coefficients are:

[130]:
```
model_coeffs = ridge_coeffs_main.loc[:, best_alpha]
model_coeffs.index = features
model_coeffs.abs().sort_values().tail(10).plot.barh(title='Top 10 Factors');
```



1.14.5 CV Result Distribution

[105]:
```
ax = sns.boxplot(y='ic', x='alpha', data=ridge_result_sig)
plt.xticks(rotation=90);
```



1.15 Lasso Regression

The lasso implementation looks very similar to the ridge model we just ran. The main difference is that lasso needs to arrive at a solution using iterative coordinate descent whereas ridge can rely on a closed-form solution:

[163]:
```
nfolds = 250
alphas = np.logspace(-8, -2, 13)
scaler = StandardScaler()

lasso_results, lasso_coeffs = pd.DataFrame(), pd.DataFrame()
for i, alpha in enumerate(alphas):
    print i,
    coeffs, test_results = [], []
    lr_lasso = Lasso(alpha=alpha)
    for i, (train_dates, test_dates) in enumerate(time_series_split(dates,␣
 ↪nfolds=nfolds)):
        X_train = model_data.loc[idx[train_dates], features]
        y_train = model_data.loc[idx[train_dates], target]
        lr_lasso.fit(X=scaler.fit_transform(X_train), y=y_train)

        X_test = model_data.loc[idx[test_dates], features]
        y_test = model_data.loc[idx[test_dates], target]
```

```
        y_pred = lr_lasso.predict(scaler.transform(X_test))

        rmse = np.sqrt(mean_squared_error(y_pred=y_pred, y_true=y_test))
        ic, pval = spearmanr(y_pred, y_test)

        coeffs.append(lr_lasso.coef_)
        test_results.append([train_dates[-1], rmse, ic, pval, alpha])
    test_results = pd.DataFrame(test_results, columns=['date', 'rmse', 'ic',␣
↪'pval', 'alpha'])
    lasso_results = lasso_results.append(test_results)
    lasso_coeffs[alpha] = np.mean(coeffs, axis=0)
```

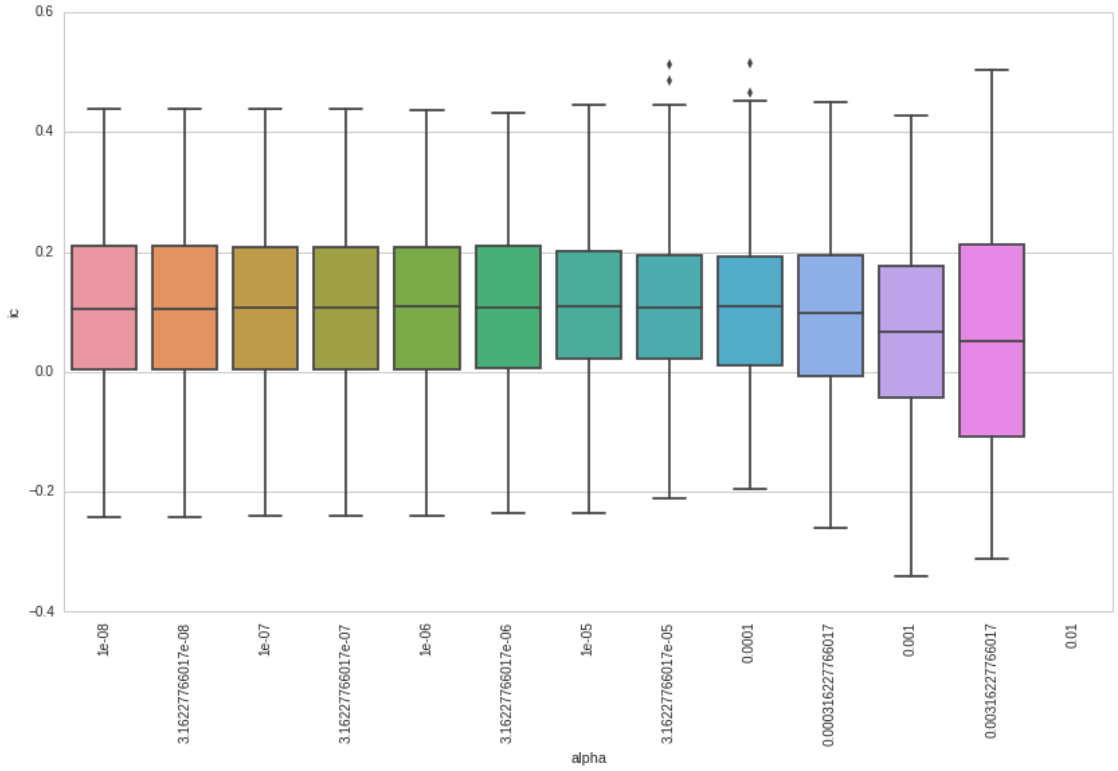
```
0 1 2 3 4 5 6 7 8 9 10 11 12
```

[164]:
```
lasso_results.groupby('alpha').mean()
```

[164]:

alpha	rmse	ic	pval
1.000000e-08	0.045714	0.108370	0.255438
3.162278e-08	0.045713	0.108384	0.255490
1.000000e-07	0.045710	0.108429	0.255493
3.162278e-07	0.045699	0.108550	0.255804
1.000000e-06	0.045667	0.108794	0.255666
3.162278e-06	0.045572	0.109276	0.254775
1.000000e-05	0.045365	0.110997	0.247775
3.162278e-05	0.045343	0.110751	0.244619
1.000000e-04	0.044766	0.103843	0.248817
3.162278e-04	0.044462	0.095700	0.238646
1.000000e-03	0.044479	0.059093	0.251200
3.162278e-03	0.044529	0.050669	0.190781
1.000000e-02	0.044537	NaN	NaN

[165]:
```
ax = sns.boxplot(y='ic', x='alpha', data=lasso_results)
plt.xticks(rotation=90);
```

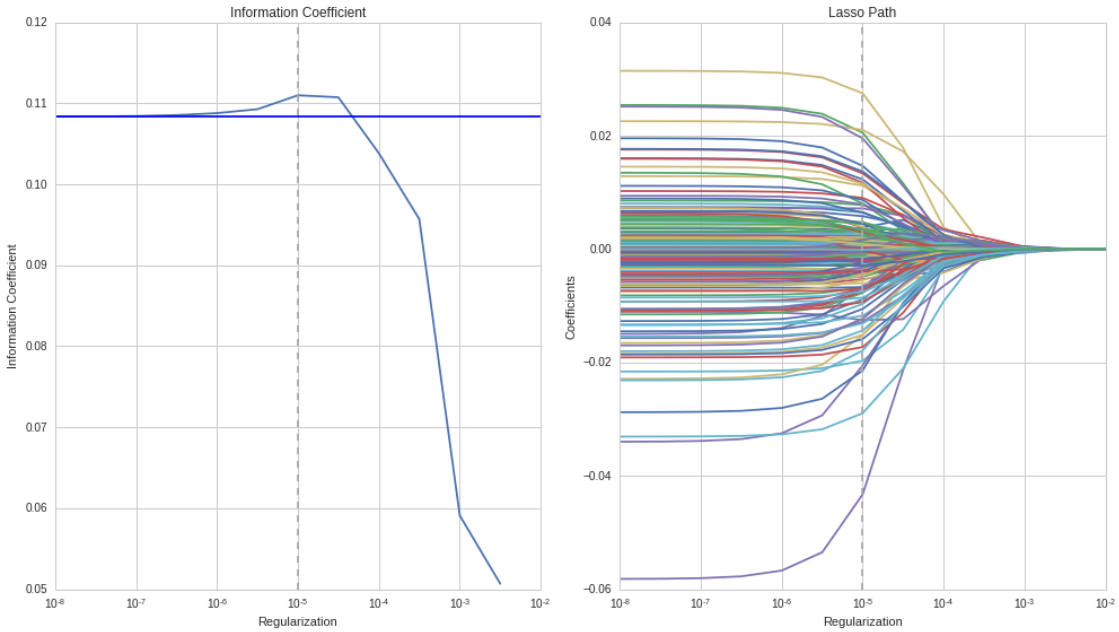


1.15.1 Cross-validated information coefficient and Lasso Path

As before, we can plot the average information coefficient for all test sets used during cross-validation. We see again that regularization improves the IC over the unconstrained model, delivering the best out-of-sample result at a level of =10-5. The optimal regularization value is quite different from ridge regression because the penalty consists of the sum of the absolute, not the squared values of the relatively small coefficient values. We can also see that for this regularization level, the coefficients have been similarly shrunk, as in the ridge regression case:

[170]:
```
fig, axes = plt.subplots(ncols=2, sharex=True)

lasso_results.groupby('alpha')['ic'].mean().plot(logx=True, title='Information␣
↪Coefficient', ax=axes[0])
axes[0].axhline(lasso_results.groupby('alpha')['ic'].mean().median())
axes[0].axvline(x=lasso_results.groupby('alpha')['ic'].mean().idxmax(),␣
↪c='darkgrey', ls='--')
axes[0].set_xlabel('Regularization')
axes[0].set_ylabel('Information Coefficient')

lasso_coeffs.T.plot(legend=False, logx=True, title='Lasso Path', ax=axes[1])
axes[1].set_xlabel('Regularization')
axes[1].set_ylabel('Coefficients')
```

```
axes[1].axvline(x=lasso_results.groupby('alpha')['ic'].mean().idxmax(),␣
 ↪c='darkgrey', ls='--')
fig.tight_layout();
```



In sum, ridge and lasso will produce similar results. Ridge often computes faster, but lasso also yields continuous features subset selection by gradually reducing coefficients to zero, hence eliminating features.

[ ]: