

04_statistical_inference_of_stock_returns_with_statsmodels

September 29, 2021

1 Statistical inference of stock returns with linear regression

1.1 Imports & Settings

```
[1]: import warnings
      warnings.filterwarnings('ignore')
```

```
[2]: %matplotlib inline

      import pandas as pd

      from statsmodels.api import OLS, add_constant, graphics
      from statsmodels.graphics.tsaplots import plot_acf
      from scipy.stats import norm

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

```
[3]: sns.set_style('whitegrid')
      idx = pd.IndexSlice
```

1.2 Load Data

```
[4]: with pd.HDFStore('data.h5') as store:
      data = (store['model_data']
              .dropna()
              .drop(['open', 'close', 'low', 'high'], axis=1))
```

1.2.1 Select Investment Universe

```
[5]: data = data[data.dollar_vol_rank<100]
```

```
[6]: data.info(null_counts=True)
```

```
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 109675 entries, ('AAL', Timestamp('2013-07-25 00:00:00')) to ('ZTS',
Timestamp('2014-12-04 00:00:00'))
Data columns (total 65 columns):
```

#	Column	Non-Null Count	Dtype
----	-----	-----	-----
0	volume	109675 non-null	float64
1	dollar_vol	109675 non-null	float64
2	dollar_vol_1m	109675 non-null	float64
3	dollar_vol_rank	109675 non-null	float64
4	rsi	109675 non-null	float64
5	bb_high	109675 non-null	float64
6	bb_low	109675 non-null	float64
7	atr	109675 non-null	float64
8	macd	109675 non-null	float64
9	return_1d	109675 non-null	float64
10	return_5d	109675 non-null	float64
11	return_10d	109675 non-null	float64
12	return_21d	109675 non-null	float64
13	return_42d	109675 non-null	float64
14	return_63d	109675 non-null	float64
15	return_1d_lag1	109675 non-null	float64
16	return_5d_lag1	109675 non-null	float64
17	return_10d_lag1	109675 non-null	float64
18	return_21d_lag1	109675 non-null	float64
19	return_1d_lag2	109675 non-null	float64
20	return_5d_lag2	109675 non-null	float64
21	return_10d_lag2	109675 non-null	float64
22	return_21d_lag2	109675 non-null	float64
23	return_1d_lag3	109675 non-null	float64
24	return_5d_lag3	109675 non-null	float64
25	return_10d_lag3	109675 non-null	float64
26	return_21d_lag3	109675 non-null	float64
27	return_1d_lag4	109675 non-null	float64
28	return_5d_lag4	109675 non-null	float64
29	return_10d_lag4	109675 non-null	float64
30	return_21d_lag4	109675 non-null	float64
31	return_1d_lag5	109675 non-null	float64
32	return_5d_lag5	109675 non-null	float64
33	return_10d_lag5	109675 non-null	float64
34	return_21d_lag5	109675 non-null	float64
35	target_1d	109675 non-null	float64
36	target_5d	109675 non-null	float64
37	target_10d	109675 non-null	float64
38	target_21d	109675 non-null	float64
39	year_2014	109675 non-null	uint8
40	year_2015	109675 non-null	uint8
41	year_2016	109675 non-null	uint8
42	year_2017	109675 non-null	uint8
43	month_2	109675 non-null	uint8
44	month_3	109675 non-null	uint8
45	month_4	109675 non-null	uint8

```

46 month_5                109675 non-null  uint8
47 month_6                109675 non-null  uint8
48 month_7                109675 non-null  uint8
49 month_8                109675 non-null  uint8
50 month_9                109675 non-null  uint8
51 month_10               109675 non-null  uint8
52 month_11               109675 non-null  uint8
53 month_12               109675 non-null  uint8
54 capital_goods          109675 non-null  uint8
55 consumer_durables      109675 non-null  uint8
56 consumer_non-durables  109675 non-null  uint8
57 consumer_services      109675 non-null  uint8
58 energy                 109675 non-null  uint8
59 finance                 109675 non-null  uint8
60 health_care            109675 non-null  uint8
61 miscellaneous          109675 non-null  uint8
62 public_utilities       109675 non-null  uint8
63 technology             109675 non-null  uint8
64 transportation        109675 non-null  uint8
dtypes: float64(39), uint8(26)
memory usage: 36.5+ MB

```

1.2.2 Create Model Data

```

[7]: y = data.filter(like='target')
     X = data.drop(y.columns, axis=1)
     X = X.drop(['dollar_vol', 'dollar_vol_rank', 'volume', 'consumer_durables'],
               ↪axis=1)

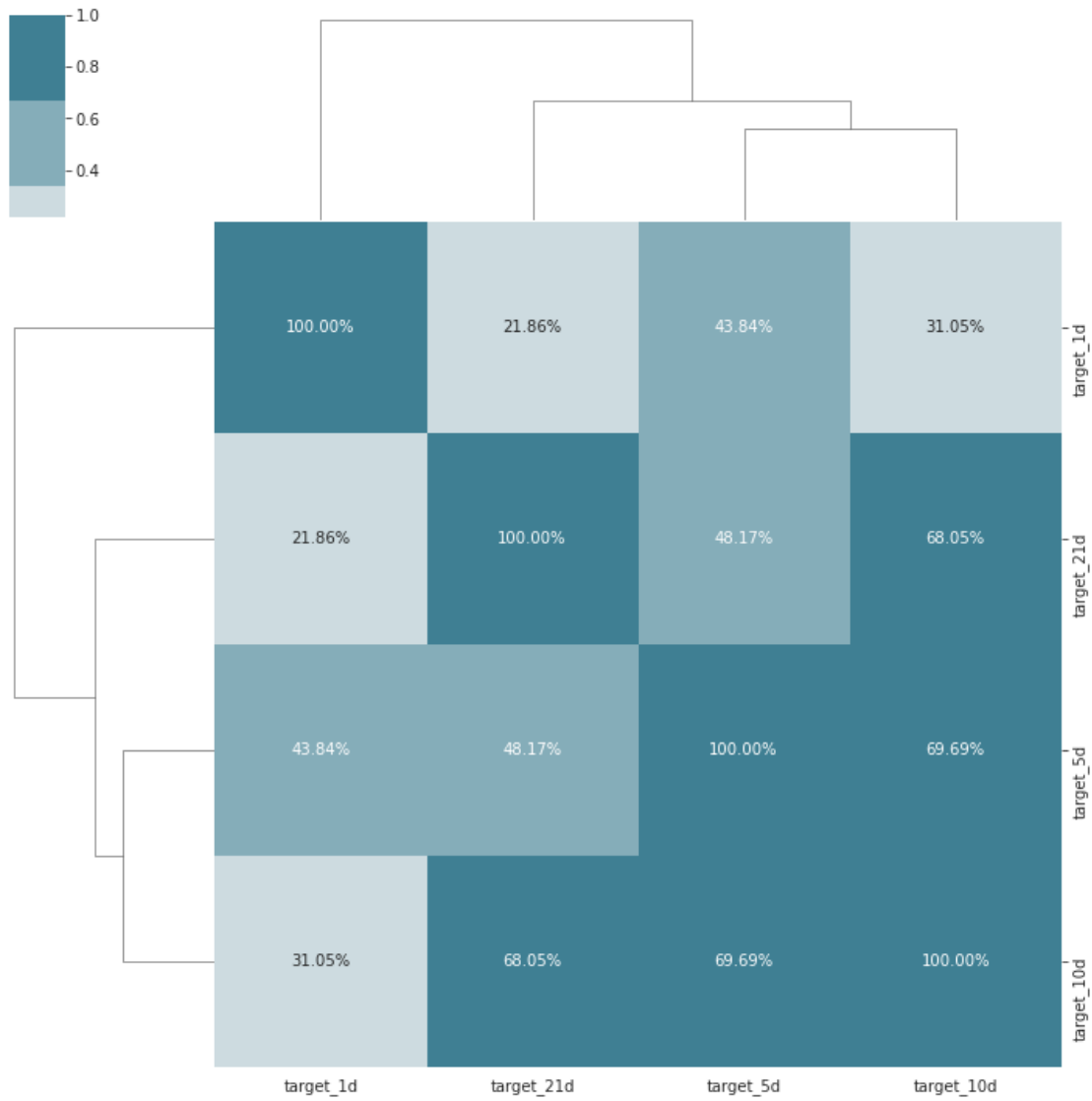
```

1.3 Explore Data

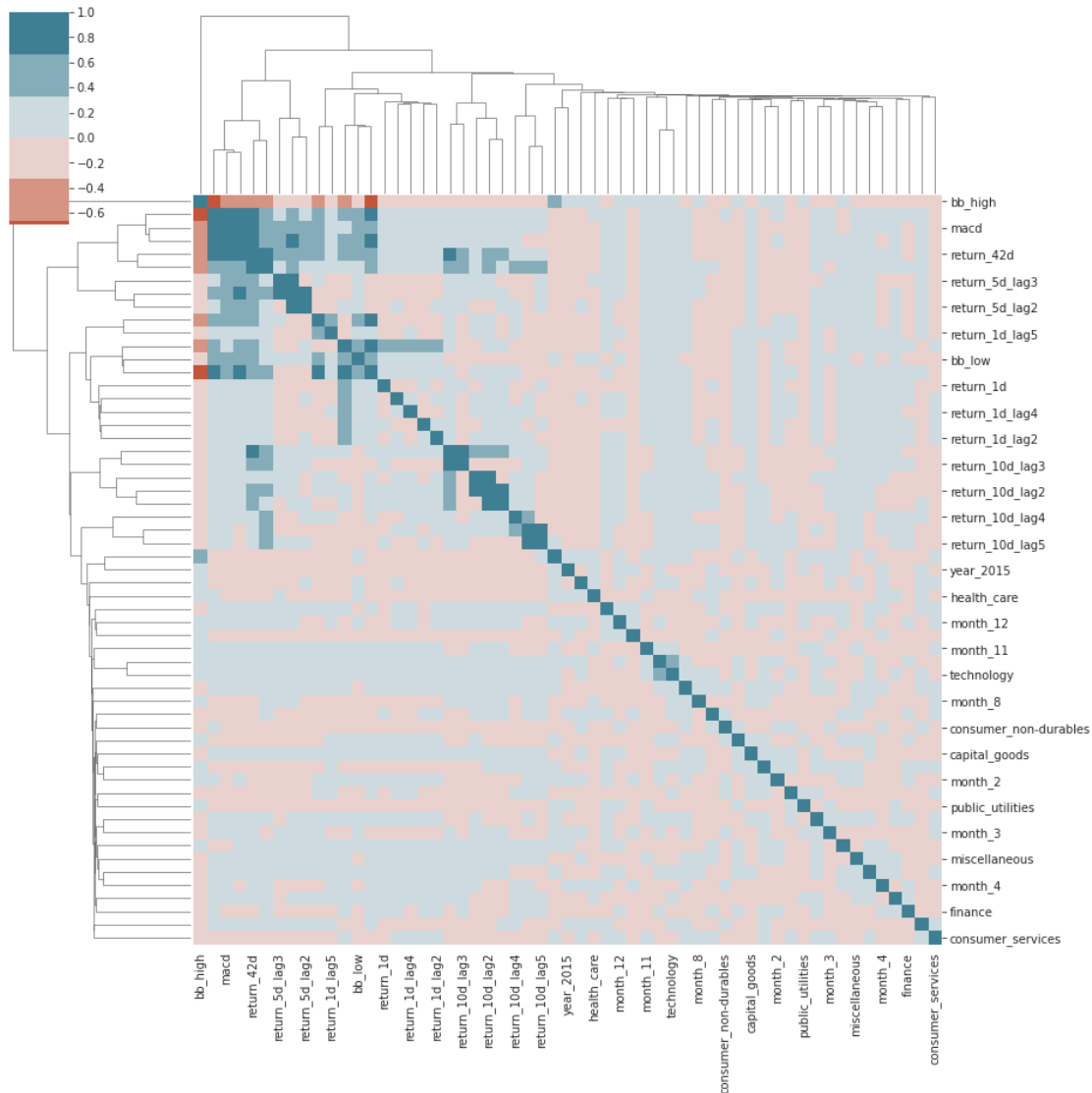
```

[8]: sns.clustermap(y.corr(), cmap=sns.diverging_palette(h_neg=20, h_pos=220),
               ↪center=0, annot=True, fmt='.2%');

```



```
[9]: sns.clustermap(X.corr(), cmap=sns.diverging_palette(h_neg=20, h_pos=220),
↪ center=0);
plt.gcf().set_size_inches((14, 14))
```



```
[10]: corr_mat = X.corr().stack().reset_index()
corr_mat.columns=['var1', 'var2', 'corr']
corr_mat = corr_mat[corr_mat.var1!=corr_mat.var2].sort_values(by='corr',
↪ascending=False)
```

```
[11]: corr_mat.head().append(corr_mat.tail())
```

```
[11]:
```

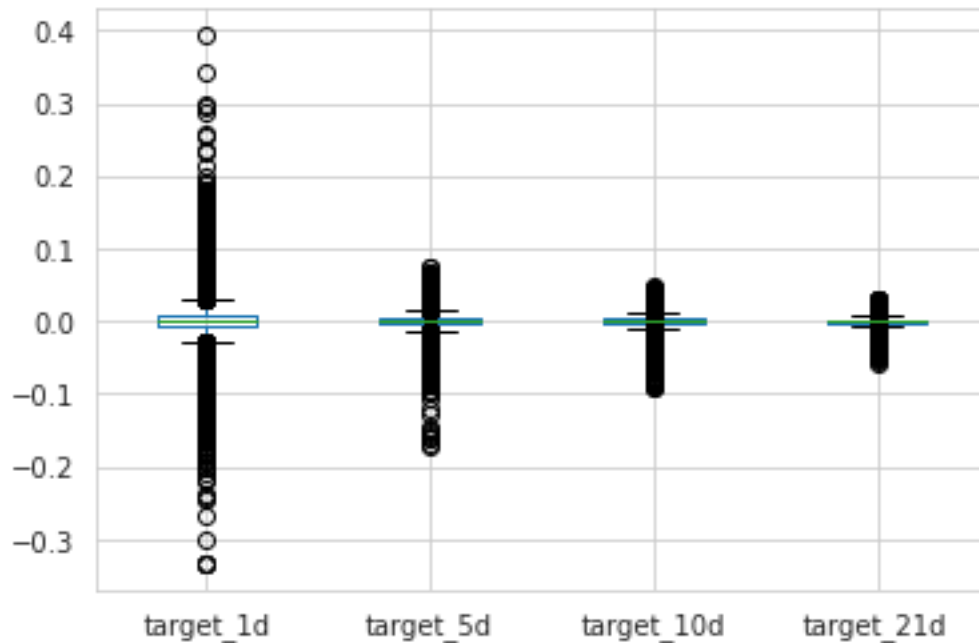
	var1	var2	corr
581	return_42d	return_63d	0.835634
637	return_63d	return_42d	0.835634
286	macd	rsi	0.817113
62	rsi	macd	0.817113
518	return_21d	macd	0.793893

```

515 return_21d      bb_high -0.632777
122   bb_high      return_10d -0.693640
458 return_10d      bb_high -0.693640
59     rsi          bb_high -0.696555
115   bb_high          rsi -0.696555

```

```
[12]: y.boxplot();
```



1.4 Linear Regression for Statistical Inference: OLS with statsmodels

1.4.1 Ticker-wise standardization

`statsmodels` warns of high design matrix condition numbers. This can arise when the variables are not standardized and the Eigenvalues differ due to scaling. The following step avoids this warning.

```

[13]: sectors = X.iloc[:, -10:]
X = (X.drop(sectors.columns, axis=1)
     .groupby(level='ticker')
     .transform(lambda x: (x - x.mean()) / x.std())
     .join(sectors)
     .fillna(0))

```

1.4.2 1-Day Returns

```
[14]: target = 'target_1d'
model = OLS(endog=y[target], exog=add_constant(X))
trained_model = model.fit()
print(trained_model.summary())
```

```

                                OLS Regression Results
=====
Dep. Variable:                  target_1d    R-squared:                0.010
Model:                        OLS          Adj. R-squared:           0.009
Method:                       Least Squares  F-statistic:              19.03
Date:                        Thu, 15 Apr 2021  Prob (F-statistic):    9.43e-189
Time:                        15:03:16       Log-Likelihood:          2.8852e+05
No. Observations:              109675      AIC:                    -5.769e+05
Df Residuals:                  109617      BIC:                    -5.764e+05
Df Model:                      57
Covariance Type:               nonrobust
=====
=====
                                coef      std err          t      P>|t|      [0.025
0.975]
-----
-----
const                -0.0002      0.000      -0.793      0.428      -0.001
0.000
dollar_vol_1m        -0.0004      6.88e-05     -5.214      0.000      -0.000
-0.000
rsi                   0.0002      0.000       0.978      0.328      -0.000
0.001
bb_high              0.0002      0.000       0.927      0.354      -0.000
0.001
bb_low               0.0006      0.000       2.940      0.003       0.000
0.001
atr                 -3.752e-05   7.44e-05     -0.504      0.614      -0.000
0.000
macd                 -0.0004      0.000      -1.835      0.066      -0.001
2.96e-05
return_1d            0.0029      0.000       9.684      0.000       0.002
0.003
return_5d           -0.0019      0.001      -2.015      0.044      -0.004
-5.08e-05
return_10d          -0.0064      0.001      -6.443      0.000      -0.008
-0.004
return_21d           0.0028      0.000       6.242      0.000       0.002
0.004
return_42d          -0.0036      0.001      -5.995      0.000      -0.005
-0.002

```

return_63d	-0.0019	0.000	-4.241	0.000	-0.003
-0.001					
return_1d_lag1	0.0027	0.000	8.998	0.000	0.002
0.003					
return_5d_lag1	0.0048	0.001	7.033	0.000	0.003
0.006					
return_10d_lag1	-0.0010	0.001	-0.896	0.370	-0.003
0.001					
return_21d_lag1	0.0029	0.000	7.667	0.000	0.002
0.004					
return_1d_lag2	0.0027	0.000	9.177	0.000	0.002
0.003					
return_5d_lag2	0.0009	0.001	1.141	0.254	-0.001
0.002					
return_10d_lag2	8.065e-05	0.001	0.097	0.923	-0.002
0.002					
return_21d_lag2	0.0003	0.000	0.857	0.392	-0.000
0.001					
return_1d_lag3	0.0027	0.000	8.973	0.000	0.002
0.003					
return_5d_lag3	0.0015	0.001	1.919	0.055	-3.15e-05
0.003					
return_10d_lag3	0.0005	0.000	3.457	0.001	0.000
0.001					
return_21d_lag3	-0.0002	5.55e-05	-3.739	0.000	-0.000
-9.88e-05					
return_1d_lag4	0.0030	0.000	9.976	0.000	0.002
0.004					
return_5d_lag4	0.0006	0.001	1.057	0.290	-0.001
0.002					
return_10d_lag4	0.0004	0.000	3.845	0.000	0.000
0.001					
return_21d_lag4	4.323e-05	5.5e-05	0.786	0.432	-6.46e-05
0.000					
return_1d_lag5	-8.924e-05	6.19e-05	-1.441	0.149	-0.000
3.21e-05					
return_5d_lag5	0.0001	0.001	0.241	0.810	-0.001
0.001					
return_10d_lag5	0.0007	9.95e-05	7.235	0.000	0.001
0.001					
return_21d_lag5	9.403e-05	5.48e-05	1.716	0.086	-1.34e-05
0.000					
year_2014	-0.0004	8.38e-05	-4.430	0.000	-0.001
-0.000					
year_2015	-0.0006	9.15e-05	-6.804	0.000	-0.001
-0.000					
year_2016	-0.0005	8.95e-05	-5.113	0.000	-0.001
-0.000					

year_2017	-0.0002	8.83e-05	-2.024	0.043	-0.000
-5.66e-06					
month_2	0.0010	7.21e-05	14.079	0.000	0.001
0.001					
month_3	0.0003	7.44e-05	4.637	0.000	0.000
0.000					
month_4	0.0005	7.36e-05	7.053	0.000	0.000
0.001					
month_5	0.0005	7.25e-05	6.741	0.000	0.000
0.001					
month_6	0.0004	7.34e-05	5.696	0.000	0.000
0.001					
month_7	0.0006	7.63e-05	8.415	0.000	0.000
0.001					
month_8	6.821e-05	7.7e-05	0.886	0.375	-8.26e-05
0.000					
month_9	0.0004	7.59e-05	5.017	0.000	0.000
0.001					
month_10	0.0007	7.8e-05	8.524	0.000	0.001
0.001					
month_11	0.0006	7.6e-05	7.697	0.000	0.000
0.001					
month_12	0.0004	7.42e-05	5.153	0.000	0.000
0.001					
capital_goods	0.0010	0.000	2.714	0.007	0.000
0.002					
consumer_non-durables	0.0007	0.000	1.793	0.073	-6.33e-05
0.001					
consumer_services	0.0008	0.000	2.234	0.025	9.29e-05
0.001					
energy	0.0003	0.000	0.913	0.361	-0.000
0.001					
finance	0.0009	0.000	2.533	0.011	0.000
0.002					
health_care	0.0007	0.000	2.065	0.039	3.51e-05
0.001					
miscellaneous	0.0011	0.000	2.565	0.010	0.000
0.002					
public_utilities	0.0003	0.000	0.749	0.454	-0.001
0.001					
technology	0.0011	0.000	3.285	0.001	0.000
0.002					
transportation	0.0010	0.000	2.426	0.015	0.000
0.002					

```

=====
Omnibus:                29139.970    Durbin-Watson:                2.010
Prob(Omnibus):          0.000    Jarque-Bera (JB):            2800989.218
Skew:                   -0.105    Prob(JB):                     0.00

```

Kurtosis: 27.757 Cond. No. 80.7

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

1.4.3 5-Day Returns

```
[15]: target = 'target_5d'
model = OLS(endog=y[target], exog=add_constant(X))
trained_model = model.fit()
print(trained_model.summary())
```

OLS Regression Results

```
=====
Dep. Variable:          target_5d    R-squared:                0.031
Model:                  OLS          Adj. R-squared:           0.031
Method:                 Least Squares  F-statistic:              61.86
Date:                   Thu, 15 Apr 2021  Prob (F-statistic):        0.00
Time:                   15:03:16      Log-Likelihood:           3.7883e+05
No. Observations:       109675       AIC:                     -7.575e+05
Df Residuals:           109617       BIC:                     -7.570e+05
Df Model:                57
Covariance Type:        nonrobust
=====
```

```
=====
               coef      std err          t      P>|t|      [0.025
0.975]
-----
const          -0.0005      0.000      -3.412      0.001      -0.001
-0.000
dollar_vol_1m   -0.0003     3.02e-05     -8.790      0.000      -0.000
-0.000
rsi              0.0005     8.43e-05      5.477      0.000      0.000
0.001
bb_high          0.0010     9.88e-05     10.353      0.000      0.001
0.001
bb_low          -0.0004     9.46e-05     -4.550      0.000      -0.001
-0.000
atr             -9.8e-05     3.27e-05     -2.999      0.003      -0.000
-3.4e-05
macd            -0.0004      0.000      -3.880      0.000      -0.001
-0.000
return_1d        0.0010      0.000       7.468      0.000      0.001
0.001
return_5d        0.0012      0.000       2.947      0.003      0.000
=====
```

0.002					
return_10d	-0.0038	0.000	-8.639	0.000	-0.005
-0.003					
return_21d	0.0021	0.000	10.714	0.000	0.002
0.003					
return_42d	-0.0024	0.000	-9.124	0.000	-0.003
-0.002					
return_63d	-0.0017	0.000	-8.800	0.000	-0.002
-0.001					
return_1d_lag1	0.0009	0.000	7.281	0.000	0.001
0.001					
return_5d_lag1	0.0035	0.000	11.632	0.000	0.003
0.004					
return_10d_lag1	0.0007	0.000	1.546	0.122	-0.000
0.002					
return_21d_lag1	0.0020	0.000	11.876	0.000	0.002
0.002					
return_1d_lag2	0.0010	0.000	7.493	0.000	0.001
0.001					
return_5d_lag2	-6.79e-06	0.000	-0.020	0.984	-0.001
0.001					
return_10d_lag2	-0.0003	0.000	-0.918	0.359	-0.001
0.000					
return_21d_lag2	0.0006	0.000	4.309	0.000	0.000
0.001					
return_1d_lag3	0.0010	0.000	7.844	0.000	0.001
0.001					
return_5d_lag3	0.0001	0.000	0.397	0.691	-0.001
0.001					
return_10d_lag3	0.0005	6.4e-05	8.171	0.000	0.000
0.001					
return_21d_lag3	-0.0002	2.44e-05	-7.933	0.000	-0.000
-0.000					
return_1d_lag4	0.0010	0.000	7.887	0.000	0.001
0.001					
return_5d_lag4	0.0007	0.000	2.443	0.015	0.000
0.001					
return_10d_lag4	4.558e-05	4.51e-05	1.012	0.312	-4.27e-05
0.000					
return_21d_lag4	8.989e-05	2.41e-05	3.724	0.000	4.26e-05
0.000					
return_1d_lag5	1.138e-05	2.72e-05	0.419	0.675	-4.19e-05
6.46e-05					
return_5d_lag5	0.0005	0.000	1.848	0.065	-2.91e-05
0.001					
return_10d_lag5	0.0004	4.37e-05	8.446	0.000	0.000
0.000					
return_21d_lag5	0.0002	2.4e-05	7.811	0.000	0.000

0.000					
year_2014	-0.0003	3.68e-05	-9.407	0.000	-0.000
-0.000					
year_2015	-0.0005	4.02e-05	-13.632	0.000	-0.001
-0.000					
year_2016	-0.0003	3.93e-05	-8.157	0.000	-0.000
-0.000					
year_2017	-0.0002	3.88e-05	-4.454	0.000	-0.000
-9.67e-05					
month_2	0.0009	3.16e-05	27.541	0.000	0.001
0.001					
month_3	0.0001	3.27e-05	3.949	0.000	6.5e-05
0.000					
month_4	0.0003	3.23e-05	9.721	0.000	0.000
0.000					
month_5	0.0005	3.18e-05	14.481	0.000	0.000
0.001					
month_6	0.0002	3.22e-05	7.201	0.000	0.000
0.000					
month_7	0.0004	3.35e-05	13.321	0.000	0.000
0.001					
month_8	3.574e-05	3.38e-05	1.058	0.290	-3.05e-05
0.000					
month_9	0.0003	3.33e-05	7.745	0.000	0.000
0.000					
month_10	0.0004	3.42e-05	12.425	0.000	0.000
0.000					
month_11	0.0005	3.33e-05	13.788	0.000	0.000
0.001					
month_12	0.0001	3.26e-05	4.388	0.000	7.91e-05
0.000					
capital_goods	0.0011	0.000	6.699	0.000	0.001
0.001					
consumer_non-durables	0.0009	0.000	5.179	0.000	0.001
0.001					
consumer_services	0.0009	0.000	6.038	0.000	0.001
0.001					
energy	0.0004	0.000	2.488	0.013	8.11e-05
0.001					
finance	0.0010	0.000	6.510	0.000	0.001
0.001					
health_care	0.0008	0.000	5.240	0.000	0.000
0.001					
miscellaneous	0.0012	0.000	6.354	0.000	0.001
0.002					
public_utilities	0.0003	0.000	1.864	0.062	-1.77e-05
0.001					
technology	0.0011	0.000	7.760	0.000	0.001

```

0.001
transportation          0.0010      0.000      5.920      0.000      0.001
0.001
=====
Omnibus:                43640.257      Durbin-Watson:                0.436
Prob(Omnibus):          0.000      Jarque-Bera (JB):            2362450.641
Skew:                   -1.138      Prob(JB):                    0.00
Kurtosis:               25.623      Cond. No.                    80.7
=====

```

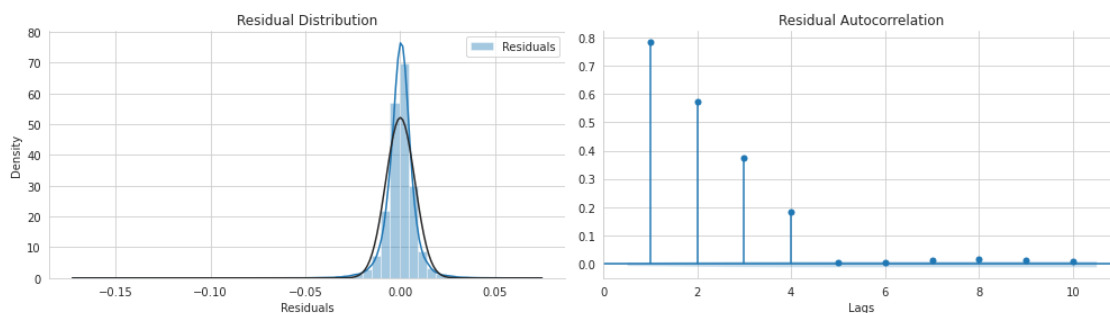
Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Obtain the residuals

```
[16]: preds = trained_model.predict(add_constant(X))
      residuals = y[target] - preds
```

```
[17]: fig, axes = plt.subplots(ncols=2, figsize=(14,4))
      sns.distplot(residuals, fit=norm, ax=axes[0], axlabel='Residuals',
      ↪label='Residuals')
      axes[0].set_title('Residual Distribution')
      axes[0].legend()
      plot_acf(residuals, lags=10, zero=False, ax=axes[1], title='Residual
      ↪Autocorrelation')
      axes[1].set_xlabel('Lags')
      sns.despine()
      fig.tight_layout();
```



1.4.4 10-Day Returns

```
[18]: target = 'target_10d'
      model = OLS(endog=y[target], exog=add_constant(X))
      trained_model = model.fit()
      print(trained_model.summary())
```

OLS Regression Results

```

=====
Dep. Variable:          target_10d    R-squared:                0.043
Model:                  OLS          Adj. R-squared:           0.042
Method:                 Least Squares  F-statistic:              85.38
Date:                  Thu, 15 Apr 2021  Prob (F-statistic):        0.00
Time:                  15:03:18      Log-Likelihood:           4.1892e+05
No. Observations:      109675       AIC:                     -8.377e+05
Df Residuals:          109617       BIC:                     -8.372e+05
Df Model:               57
Covariance Type:       nonrobust
=====

```

```

=====
                                coef      std err          t      P>|t|      [0.025
0.975]
-----
const                -0.0004      9.5e-05      -3.922      0.000      -0.001
-0.000
dollar_vol_1m        -0.0001      2.1e-05      -6.800      0.000      -0.000
-0.000
rsi                   0.0002      5.85e-05      2.759      0.006      4.67e-05
0.000
bb_high              0.0006      6.85e-05      9.273      0.000      0.001
0.001
bb_low               -0.0002      6.56e-05      -2.847      0.004      -0.000
-5.82e-05
atr                  -0.0002      2.27e-05      -7.328      0.000      -0.000
-0.000
macd                 -0.0001      7.22e-05      -1.908      0.056      -0.000
3.76e-06
return_1d            0.0005      9.01e-05      6.046      0.000      0.000
0.001
return_5d            5.98e-05      0.000      0.212      0.832      -0.000
0.001
return_10d           -0.0008      0.000      -2.741      0.006      -0.001
-0.000
return_21d           0.0017      0.000      11.902      0.000      0.001
0.002
return_42d           -0.0025      0.000     -13.475      0.000      -0.003
-0.002
return_63d           -0.0014      0.000     -10.322      0.000      -0.002
-0.001
return_1d_lag1        0.0005      8.98e-05      5.980      0.000      0.000
0.001
return_5d_lag1        0.0013      0.000      6.090      0.000      0.001
0.002
return_10d_lag1       0.0009      0.000      2.615      0.009      0.000

```

0.002					
return_21d_lag1	0.0017	0.000	14.947	0.000	0.002
0.002					
return_1d_lag2	0.0005	8.99e-05	5.956	0.000	0.000
0.001					
return_5d_lag2	6.813e-05	0.000	0.294	0.769	-0.000
0.001					
return_10d_lag2	0.0002	0.000	0.690	0.490	-0.000
0.001					
return_21d_lag2	0.0002	9.01e-05	2.768	0.006	7.28e-05
0.000					
return_1d_lag3	0.0005	9.03e-05	6.066	0.000	0.000
0.001					
return_5d_lag3	9.299e-06	0.000	0.040	0.968	-0.000
0.000					
return_10d_lag3	0.0005	4.44e-05	12.279	0.000	0.000
0.001					
return_21d_lag3	-0.0001	1.69e-05	-6.699	0.000	-0.000
-8.02e-05					
return_1d_lag4	0.0006	9.08e-05	6.278	0.000	0.000
0.001					
return_5d_lag4	0.0002	0.000	1.250	0.211	-0.000
0.001					
return_10d_lag4	0.0002	3.13e-05	7.472	0.000	0.000
0.000					
return_21d_lag4	9.467e-05	1.67e-05	5.653	0.000	6.18e-05
0.000					
return_1d_lag5	-1.792e-05	1.89e-05	-0.950	0.342	-5.49e-05
1.9e-05					
return_5d_lag5	0.0002	0.000	1.064	0.287	-0.000
0.001					
return_10d_lag5	0.0003	3.03e-05	10.693	0.000	0.000
0.000					
return_21d_lag5	0.0001	1.67e-05	8.453	0.000	0.000
0.000					
year_2014	-0.0004	2.55e-05	-13.817	0.000	-0.000
-0.000					
year_2015	-0.0006	2.79e-05	-20.441	0.000	-0.001
-0.001					
year_2016	-0.0002	2.73e-05	-9.022	0.000	-0.000
-0.000					
year_2017	-0.0002	2.69e-05	-7.333	0.000	-0.000
-0.000					
month_2	0.0006	2.2e-05	28.394	0.000	0.001
0.001					
month_3	4.204e-05	2.27e-05	1.855	0.064	-2.37e-06
8.65e-05					
month_4	0.0002	2.24e-05	8.025	0.000	0.000

0.000					
month_5	0.0003	2.21e-05	14.300	0.000	0.000
0.000					
month_6	0.0002	2.24e-05	7.894	0.000	0.000
0.000					
month_7	0.0002	2.32e-05	8.109	0.000	0.000
0.000					
month_8	-6.35e-05	2.34e-05	-2.709	0.007	-0.000
-1.76e-05					
month_9	4.795e-05	2.31e-05	2.075	0.038	2.66e-06
9.32e-05					
month_10	0.0003	2.37e-05	14.479	0.000	0.000
0.000					
month_11	0.0002	2.31e-05	10.264	0.000	0.000
0.000					
month_12	-3.354e-05	2.26e-05	-1.485	0.138	-7.78e-05
1.07e-05					
capital_goods	0.0010	0.000	8.686	0.000	0.001
0.001					
consumer_non-durables	0.0007	0.000	6.399	0.000	0.001
0.001					
consumer_services	0.0007	0.000	7.209	0.000	0.001
0.001					
energy	0.0003	0.000	2.736	0.006	8.26e-05
0.001					
finance	0.0009	0.000	8.318	0.000	0.001
0.001					
health_care	0.0006	0.000	6.168	0.000	0.000
0.001					
miscellaneous	0.0012	0.000	8.854	0.000	0.001
0.001					
public_utilities	0.0002	0.000	1.618	0.106	-4.39e-05
0.000					
technology	0.0010	0.000	9.984	0.000	0.001
0.001					
transportation	0.0010	0.000	7.928	0.000	0.001
0.001					

```

=====
Omnibus:                38454.440    Durbin-Watson:                0.233
Prob(Omnibus):          0.000    Jarque-Bera (JB):            1111784.823
Skew:                   -1.087    Prob(JB):                    0.00
Kurtosis:               18.446    Cond. No.                    80.7
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

1.4.5 Monthly Returns

```
[19]: target = 'target_21d'
model = OLS(endog=y[target], exog=add_constant(X))
trained_model = model.fit()
print(trained_model.summary())
```

```

                                OLS Regression Results
=====
Dep. Variable:                target_21d    R-squared:                0.059
Model:                        OLS          Adj. R-squared:         0.058
Method:                       Least Squares  F-statistic:           119.7
Date:                         Thu, 15 Apr 2021  Prob (F-statistic):      0.00
Time:                         15:03:18      Log-Likelihood:        4.6155e+05
No. Observations:             109675       AIC:                  -9.230e+05
Df Residuals:                 109617       BIC:                  -9.224e+05
Df Model:                      57
Covariance Type:              nonrobust
=====
=====
                                coef      std err          t      P>|t|      [0.025
0.975]
-----
-----
const                -0.0003    6.44e-05    -4.358    0.000    -0.000
-0.000
dollar_vol_1m        -4.008e-05  1.42e-05    -2.820    0.005   -6.79e-05
-1.22e-05
rsi                   2.026e-05  3.97e-05     0.511    0.609   -5.75e-05
9.8e-05
bb_high               0.0002    4.65e-05     3.684    0.000    8.01e-05
0.000
bb_low               0.0002    4.45e-05     4.367    0.000     0.000
0.000
atr                  -0.0002    1.54e-05   -12.330    0.000    -0.000
-0.000
macd                 -0.0003    4.89e-05    -5.137    0.000    -0.000
-0.000
return_1d            0.0003    6.11e-05     4.853    0.000     0.000
0.000
return_5d            -0.0005     0.000     -2.575    0.010    -0.001
-0.000
return_10d           -0.0003     0.000     -1.332    0.183    -0.001
0.000
return_21d           0.0018    9.41e-05    18.701    0.000     0.002
0.002
return_42d           -0.0018     0.000   -14.314    0.000    -0.002
-0.002

```

return_63d	-0.0012	9.31e-05	-12.676	0.000	-0.001
-0.001					
return_1d_lag1	0.0003	6.09e-05	5.014	0.000	0.000
0.000					
return_5d_lag1	0.0003	0.000	2.387	0.017	6.01e-05
0.001					
return_10d_lag1	9.985e-05	0.000	0.445	0.656	-0.000
0.001					
return_21d_lag1	0.0017	7.93e-05	21.339	0.000	0.002
0.002					
return_1d_lag2	0.0003	6.09e-05	5.494	0.000	0.000
0.000					
return_5d_lag2	7.005e-05	0.000	0.446	0.655	-0.000
0.000					
return_10d_lag2	0.0001	0.000	0.716	0.474	-0.000
0.000					
return_21d_lag2	0.0005	6.11e-05	7.894	0.000	0.000
0.001					
return_1d_lag3	0.0004	6.12e-05	5.956	0.000	0.000
0.000					
return_5d_lag3	0.0001	0.000	0.696	0.486	-0.000
0.000					
return_10d_lag3	0.0001	3.01e-05	4.171	0.000	6.65e-05
0.000					
return_21d_lag3	-8.503e-05	1.15e-05	-7.415	0.000	-0.000
-6.26e-05					
return_1d_lag4	0.0004	6.16e-05	6.187	0.000	0.000
0.001					
return_5d_lag4	5.046e-05	0.000	0.401	0.688	-0.000
0.000					
return_10d_lag4	3.895e-05	2.12e-05	1.838	0.066	-2.59e-06
8.05e-05					
return_21d_lag4	0.0001	1.14e-05	13.100	0.000	0.000
0.000					
return_1d_lag5	3.351e-06	1.28e-05	0.262	0.793	-2.17e-05
2.84e-05					
return_5d_lag5	-9.804e-06	0.000	-0.080	0.936	-0.000
0.000					
return_10d_lag5	6.256e-06	2.05e-05	0.305	0.761	-3.4e-05
4.65e-05					
return_21d_lag5	-3.545e-05	1.13e-05	-3.134	0.002	-5.76e-05
-1.33e-05					
year_2014	-0.0003	1.73e-05	-17.088	0.000	-0.000
-0.000					
year_2015	-0.0006	1.89e-05	-29.827	0.000	-0.001
-0.001					
year_2016	-0.0002	1.85e-05	-9.707	0.000	-0.000
-0.000					

year_2017	-0.0002	1.82e-05	-10.028	0.000	-0.000
-0.000					
month_2	8.855e-05	1.49e-05	5.950	0.000	5.94e-05
0.000					
month_3	-0.0001	1.54e-05	-9.594	0.000	-0.000
-0.000					
month_4	-0.0001	1.52e-05	-7.960	0.000	-0.000
-9.11e-05					
month_5	2.356e-05	1.5e-05	1.574	0.116	-5.78e-06
5.29e-05					
month_6	-1.92e-05	1.52e-05	-1.266	0.206	-4.89e-05
1.05e-05					
month_7	-0.0003	1.57e-05	-18.369	0.000	-0.000
-0.000					
month_8	-0.0002	1.59e-05	-14.877	0.000	-0.000
-0.000					
month_9	-0.0002	1.57e-05	-11.661	0.000	-0.000
-0.000					
month_10	6.676e-05	1.61e-05	4.147	0.000	3.52e-05
9.83e-05					
month_11	-0.0001	1.57e-05	-7.393	0.000	-0.000
-8.52e-05					
month_12	-0.0004	1.53e-05	-25.897	0.000	-0.000
-0.000					
capital_goods	0.0008	7.57e-05	11.198	0.000	0.001
0.001					
consumer_non-durables	0.0007	7.84e-05	8.618	0.000	0.001
0.001					
consumer_services	0.0006	7.01e-05	8.933	0.000	0.000
0.001					
energy	0.0002	7.22e-05	2.652	0.008	4.99e-05
0.000					
finance	0.0008	7.37e-05	10.794	0.000	0.001
0.001					
health_care	0.0005	6.89e-05	7.493	0.000	0.000
0.001					
miscellaneous	0.0012	9.02e-05	12.969	0.000	0.001
0.001					
public_utilities	0.0001	8.71e-05	1.241	0.215	-6.26e-05
0.000					
technology	0.0009	6.95e-05	13.417	0.000	0.001
0.001					
transportation	0.0009	8.24e-05	11.165	0.000	0.001
0.001					

```

=====
Omnibus:                40182.050    Durbin-Watson:                0.129
Prob(Omnibus):          0.000    Jarque-Bera (JB):            848232.085
Skew:                   -1.254    Prob(JB):                    0.00

```

Kurtosis:	16.391	Cond. No.	80.7
=====			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.