Data: Small-Cap Value and Momentum

MGMT 767: Data-Driven Investments Lab

Kerry Back and Kevin Crotty, Rice University

In [4]:

```
from sqlalchemy import create_engine

server = 'fs.rice.edu'
database = 'stocks'
username = 'stocks'
password = '6LAZH1'
driver1 = 'SQL+Server'
driver2 = 'ODBC+Driver+17+for+SQL+Server'
string1 = f"mssql+pyodbc://{username}:{password}@{server}/{database}?driver={
string2 = f"mssql+pyodbc://{username}:{password}@{server}/{database}?driver={
try:
    conn = create_engine(string1).connect()
except:
    conn = create_engine(string2).connect()
```

In [6]:

```python
import pandas as pd
prices = pd.read_sql(
    """
    select date, ticker, closeadj, closeunadj, lastupdated from sep_weekly
    where date>='2010-01-01'
    order by ticker, date, lastupdated
    """,
    conn,
)
prices = prices.groupby(["ticker", "date"]).last()
```

In [7]:

```
rets = prices.groupby(
    "ticker",
    group_keys=False
).closeadj.pct_change()

rets_annual = prices.groupby(
    "ticker",
    group_keys=False
).closeadj.pct_change(52)

rets_monthly = prices.groupby(
    "ticker",
    group_keys=False
).closeadj.pct_change(4)

mom = (1 + rets_annual) / (1 + rets_monthly) - 1
```

In [9]:

```
df = pd.read_sql(
    """
    select date, ticker, pb, marketcap, lastupdated from weekly
    where date>='2010-01-01'
    order by ticker, date, lastupdated
    """,
    conn,
)
df = df.groupby(["ticker", "date"]).last()
```

In [ ]:

```python
df["close"] = prices.closeunadj
df["ret"] = rets
df["mom"] = mom
for col in ["marketcap", "close", "mom", "pb"]:
    df[col] = df.groupby("ticker", group_keys=False)[col].shift()
df = df.dropna()
```

In [ ]:

```
size_rank = df.groupby(
    "date",
    group_keys=False
).marketcap.rank(ascending=False)

df = df[size_rank>1000]
df = df[df.close > 5]
```

In [ ]:

```
df.to_csv("02_data.csv")
```