Literature Review Summary:

Deep Learning and the Cross-Section of Expected Returns

Deep feedforward **neural networks** based on a set of 68 firm characteristics (FC) to predict the US cross-section of stock returns.
Results are robust to size, weighting schemes and portfolio points.
Price related FC (short-term reversal and the 12-month momentum) are among the main drivers of the return predictions.
Method: Deep feedforward neural network – objective function, SGD, mini-batch, regularization, early stopping, dropout, L1&L2 regularization
Data: only CRSP stocks w/ share code 10 and 11 which are traded either on the NYSE, AMEX or NASDAQ. Stocks with missing mkt capitalization data and/or where book values are unavailable are dropped.
Performed with a standard lag of 6 months of the fiscal year end date.
Extreme FC observations are controlled for by using relative cross-section ranks at each point in the time; missing data are replaced by the median value of 0.5 at each point in time.
Return are de-meaned for each period

Shrinking the Cross section

Constructed a robust stochastic discount factor (SDF) summarizes the joint explanatory power of a large number of cross-sectional stock return predictors.
Use Bayesian approach with a novel specification of prior beliefs. For factor selection, augmented the estimation criterion w/ an additional penalty on the sum of absolute SDF coeffs.
Results show that to perform well, SDF needs to load on. A. large. Number of characteristics-based factors.
In high-dimensional setting, shrinkage of estimated sdf coeffs to 0 is critical to find an sdf. Representation that performs well out-of-sample. **Ridge** work well.

Dissecting Characteristics Nonparametrically

To find which characteristics provide incremental information for the cross-section of expected returns.
Used **adaptive group Lasso** to select characteristics and to estimate how they affect expected returns nonparametrically.
First estimated the model on 62 characteristics on a sample period from 7/1965 to 6/2014.
Second compare out-of-sample performance w/ linear model.
Third study whether the predictive power of characteristics for expected returns varies over time.

Empirical Asset Pricing via Machine Learning

Perform machine learning for measuring asset risk premia. Identified best methods: tree and neural networks, and traced the predictive gains to allowance of nonlinear predictor interactions that are missed by their methods.
Basic objective of minimizing MSE; regularization (adding parameterization penalties and robustification against outliers)

(This paper includes the comparison of different linear and nonlinear methods)
“Shallow” learning outperforms “deep” learning; ML methods are most valuable for forecasting larger and more liquid stock returns and portfolios; all methods agree on a fairly small set of dominant predictive signals, the most powerful predictors are measures of stock liquidity, stock volatility and valuation ratios.

Deep Learning for Predicting Asset Returns
(Asset pricing studies using multi-layer deep learners such as rectified linear units (ReLU) or long-short-term-memory (LSTM) for time-series effects.
State-of-the-art algorithms including SGD, TensorFlow and dropout design.)
This paper found the existence of nonlinear factors which explain predictability of returns, especially at the extremes of the characteristic space.
By varying the number of hidden layers and the number of neurons within each layer, very flexible predictors can be training and out-of-sample cv provides a technique to avoid overfitting. LSTM models are alternatives to traditional state space modeling.

Forecasting ETFs with Machine Learning Algorithms
Applied ML algorithms in predicting returns (focus on predicting the direction of several liquid ETFs, not the magnitude of price changes). Employed approximately 5-yrs of historical daily data from 01/2011 to 01/2016.
Method: Deep Neural Network, random forest SVM.
Algorithms work well over 10- to 60-day horizon; documented the importance of cross-sectional and intertemporal volume as a powerful info set; many features are needed for predictability as each feature provides very small contributions. ETFs can be predicted with ML algorithms but practitioners should incorporate prior knowledge of markets and intuition on asset behavior.

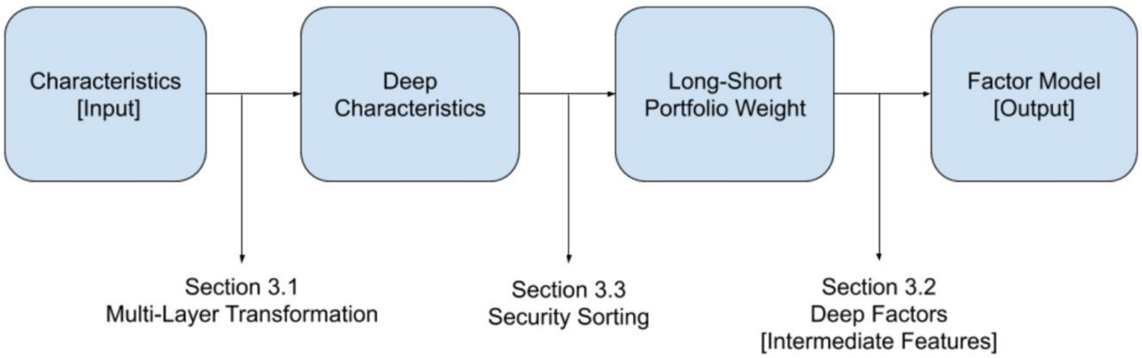
Machine Learning in empirical asset pricing
This is a literature review of the ML methods in financial area.
(Can use this one to find other references.)

Big Data and AI Strategies
General introduction of the applications of different ML methods.

Bond Risk Premia with Machine Learning
Using extreme trees and neural networks (NNs) to provide strong statistical evidence in favor of **bond return** prediction. Methodologically, partial least squares, penalized linear regressions, boosted regression trees, random forests, extremely randomized regression trees, and shallow and deep neural networks (NNs) are used.
Data comes from the CRSP Treasuries Time Series. The sample is from June 1961 to December 2018 15 in monthly frequency

Deep Learning in Characteristics-Sorted Factor Models

Kozak et al. (2019) provide a **shrinkage** approach to model fitness, and Feng et al. (2019) test new factors through model selections. In dimension reduction through principal components (PCA), Kelly et al. (2019) employ characteristics as instruments. The paper follows their research directions and provide a deep learning framework of the SDF model with dimension reduction. The model acts like PCA but can be applied to nonlinear characteristics to reduce dimension and provide new factors



Taming the Factor Zoo

Our factor library contains 150 risk factors at the monthly frequency for the period from July 1976 to December 2017, obtained from multiple sources.

To construct these factors, we include only stocks for companies listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11.

We start from a linear specification for the SDF:

$$m_t := \gamma_0^{-1} - \gamma_0^{-1}\lambda_v^\intercal v_t := \gamma_0^{-1}(1 - \lambda_g^\intercal g_t - \lambda_h^\intercal h_t), \tag{1}$$

To guard against omitted variable biases due to selection mistakes, they therefore adopt a double-selection strategy in the same spirit as what Belloni et al. (2014b) propose for estimating the treat- ment e↲ect.

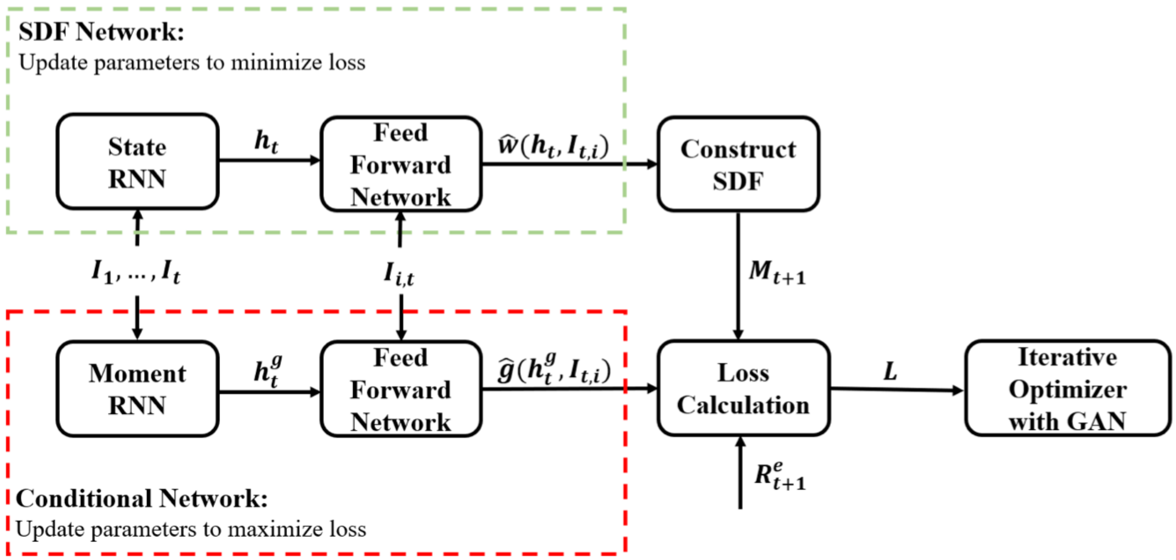
The first selection (basically, standard LASSO) searches for factors in ht whose covariances with returns are useful for explaining the cross section of expected returns. A second selection (also using LASSO) is then added to search for factors in ht potentially missed from the first step.

Deep Learning in Asset Pricing

The macroeconomic information dynamics are summarized by macroeconomic state variables ht which are obtained by a Recurrent Neural Network (RNN) with Long-Short-Term-Memory units.

GAN (Generative Adversarial Network) are used.

Our empirical analysis is based on a data set of all available U.S. stocks from CRSP with monthly returns from 1967 to 2016 combined with 46 time-varying firm-specific characteristics and 178 macroeconomic time series. It includes the most relevant pricing anomalies and forecasting variables for the equity risk premium.



Alpha Go Everywhere: ML and International Stock Returns

In this paper, linear models include: OLS-3, OLS-3 with Huber loss, OLS, OLS with Huber loss, LASSO, RIDGE, and ENET are used.

Data get from DataStream, only use data with CRSP and CSMAR in U.S and China.

They first use 12 predictors to estimate U.S stock returns, then use U.S model to test international stock, finally usd country-specific model.

Forest through the Trees: Building Cross-Sections of Stock Returns

Our method is rooted in the idea of decision trees and builds up on the appeal of standard double and triple sorts.

Generally using decision trees.