Prediction of stocks:
A new way to look at it

Jens Perch Nielsen and Stefan Sperlich

Cass Business School
Université de Genève

December 8, 2017

The Problem

1. Long term investors have contradicting aims of minimizing risk and maximizing return over the long run.
2. professional financial advisers say that expected returns in financial markets vary over time with a significant predictable component
3. e.g. dividend-price ratio, earning-price ratio, have some predictive power when looking at R^2
4. Therefore, time periods exist where long term investors might choose to sell stocks and buy bonds
5. Campbell, Lo, and MacKinlay (1997) & Wilkie (1993) argued that predictable component is increasing with time horizon as R^2 increases rapidly with it
6. The R^2 is in-sample measure but standard time-series prediction checks only work with long series (many data)

The Basic Model and Data

Traditional equation for value P_t of stock is based on unknown quantities like discount rate, constant growth of dividend yields, etc

but also dividend D_t : Campbell and Shiller (1988) referred to the model as the "dividend-ratio" in absence of uncertainty.

Danish stock market data, Lund & Engsted 1996, extended to 1922 – 2001

$$W_t = (S_t, d_t, I_t, r_t),$$

S_t stock return, I_t inflation, r_t short-term interest rate, $d_t = D_t / P_t$

Stock index is based on a value weighted portfolio of individual stocks chosen to obtain maximum coverage of the marked index of the Copenhagen Stock Exchange (CBS). Notice that CBS was open during the second world war.

Our model for prediction

Real excess stock return is

$$S_t = \log\{(P_t + D_t)/P_{t-1}\} - r_{t-1}\ , \qquad r_t = \log(1 + R_t/100)$$

Average of excess stock returns are 2.5% for 1922 – 2001 and 3.4% for the after war.

Consider $Y_t = \sum_{i=0}^{T-1} S_{t+i}$, i.e. excess over next T years.

Approximate by model

$$Y_t = g(W_{t-1}) + \epsilon_t, \qquad t \in \{K_1, \dots, K_2\}\,, \tag{1}$$

Due to definition of Y_t, time period (K_1, K_2) depends is $(T_{first}, T_{last} - T + 1)$ with e.g. $T_{first} = 1923$ and $T_{last} = 2001$, etc.

For $g(\cdot)$ we can imagine any feasible estimator : *your prior knowledge?*

Our framework for evaluating prediction

Define loss of estimator $\widehat{g}_h$ (h for smoothness / complexity) as

$$Q(\widehat{g}_h) = \sum_{t=K_1}^{K_2} \left\{ g(W_{t-1}) - \widehat{g}_h(W_{t-1}) \right\}^2$$

which can be estimated by (leave- ? -out cross validation)

$$\widehat{Q}(\widehat{g}_h) = \sum_{t=K_1}^{K_2} \left\{ Y_t - \widehat{g}_h^{(t)}(W_{t-1}) \right\}^2$$

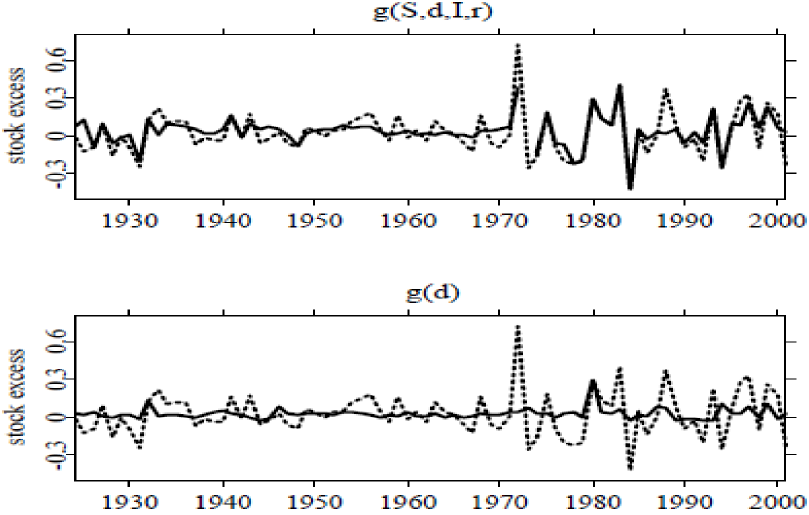
i.e. predict $g(W_{t-1})$ without information contained in Y_t.

$Q(\widehat{g}_h)$ is not estimated well by goodness-of-fit measure

$$\overline{Q}(\widehat{g}_h) = \sum_{t=K_1}^{K_2} \left\{ Y_t - \widehat{g}_h(W_{t-1}) \right\}^2$$

this measure always will be in favor of most complex model

Illustration: better fit gives better prediction?



The prediction power measure

While predicting, optimal prediction scheme is to minimize $\widehat{Q}(\widehat{g}_h)$

over all (feasible) h — discuss ...

and other model selection choices – discuss ...

Let h_0 correspond to the trivial prediction strategy

$$Y_t = \mu + \epsilon_t \; , \tag{2}$$

where μ is estimated by $\widehat{\mu} = (K_2 - K_1 + 1)^{-1} \sum_{t=K_1}^{K_2} Y_t$.

Define out new R^2 value, $R^2_{V,h}$, as

$$R^2_{V,h} = 1 - \frac{\widehat{Q}(\widehat{g}_h)}{\widehat{Q}(\widehat{g}_{h_0})}$$

in the following without h in notation

More on the R_V^2

- It measures how well a given model and estimation principle h predicts compared to simple principle $h_0 = no - model$
- If positive then modeling and estimation principle *h predicts*
- otherwise it *does not predict*
- Note that $R_V^2 \in (-\infty, 1]$ – opening range of classic R^2
- but else, interpretation is similar, as the classical

$$R^2 = 1 - \frac{\overline{Q}(\widehat{g}_h)}{\overline{Q}(\widehat{g}_{h_0})}$$

- i.e. the 'reference model' has changed
- discuss the need and sense (or not) of an *adjusted* R_V^2

R_V^2 in action: simple (log-)linear models

Consider two versions of regression

$$Y_t = S_{t+1} + ... + S_{t+T} = \alpha + \beta\delta_t + \epsilon_{t+T}, \quad (3)$$

where $\delta_t = d_t$ (left-hand) and $\delta_t = \ln(d_t)$ respectively (right-hand)

horizon	$\delta_t = d_t$		$\delta_t = \ln(d_t)$	
T	1923-1996	1949-1996	1923-1996	1949-1996
1	-0.2%	1.4%	-1.1%	-0.3%
2	4.9%	8.2%	2.2%	3.0%
3	7.8%	14.2%	4.6%	7.7%
4	10.3%	16.0%	7.4%	9.4%
5	10.3%	9.5%	6.5%	0.5%
6	6.9%	-4.6%	5.2%	-19.5%

note that predictions for 1949-1996 can be improved a lot if using all data

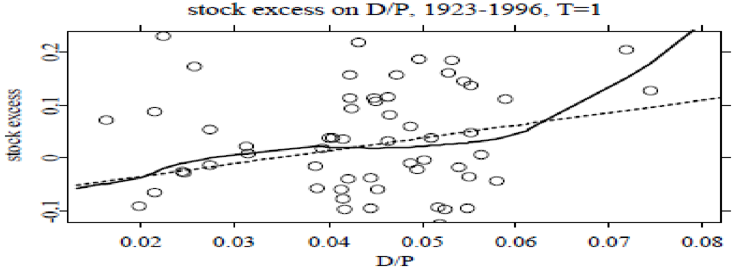
The classical R^2 in comparison

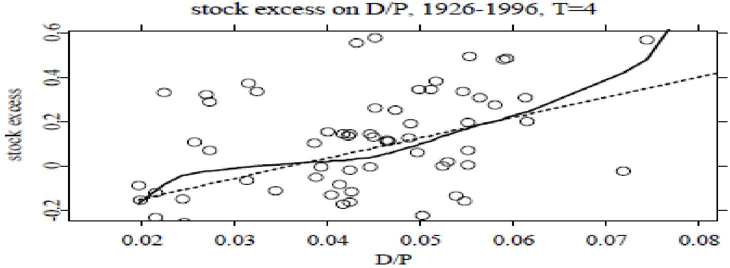
Campbell, Lo and Mackinlay (1997, p.269) arrived at conclusion that longer horizons are easier to predict.

horizon	$\delta_t = d_t$		$\delta_t = \ln(d_t)$	
T	1923-1996	1949-1996	1923-1996	1949-1996
1	3.8%	7.3%	3.2%	5.9%
2	8.8%	14.9%	6.6%	11.5%
3	13.0%	21.1%	10.5%	17.1%
4	17.5%	25.8%	14.2%	21.0%
5	18.7%	24.2%	15.7%	20.6%
6	16.4%	25.0%	15.5%	23.5%

Can be shown that is inherent to time series data.

Extension to nonparametric $g(\cdot)$





R_V^2 in action : variable selection

- Investigate the potential advantages that one can obtain by including other variables for prediction.
- restrict our investigation to a time horizon of one year

Consider time series regression problem of following form

$$S_t = g(S_{t-1}, d_{t-1}, l_{t-1}, r_{t-1}) + \epsilon_t$$

using same data as before.

For variable selection you have 15 choices:

$$1 + 4 + 6 + 4 = 15$$

not counting the constant

when only looking at the fully nonparametric ones

R_V^2 in action : non/semi-model selection

- A function $g(\cdot)$ without any parametric assumptions
- nor assumptions of structure such as additivity or multiplicativity.
- is most often too complex for both to visualize and/or to predict well
- lack of prediction due to estimation error rather than insufficient model
- So may impose some structure on $g(\cdot)$ for prediction

Concentrating only at the additive models adds

$$1 + 4 + 6 = 11$$

additive models, including the (log-) linear ones

therefore recommended to local linear estimators

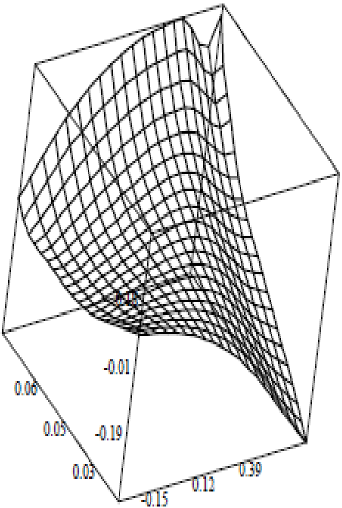
Main findings from (nonparametric) model selection

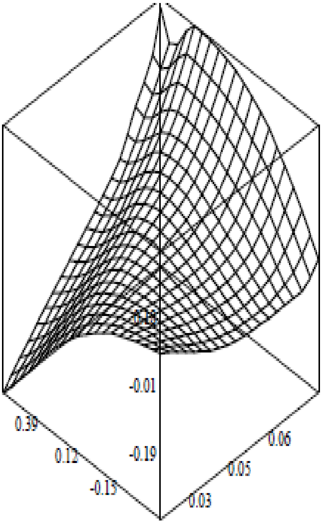
- fully non-parametric models always did better (interactions)
- only linear model that does better than simple constant is d_{t-1} for 1948 – 2001
- fully nonparametric 2-dimensional with d_{t-1} and S_{t-1}
- has $R_V^2 = 5.5\%$ for 1923 – 1996 and 9.1% for 1948 – 1996
- clearly, all this with optimal prediction bandwidth

However, results change with

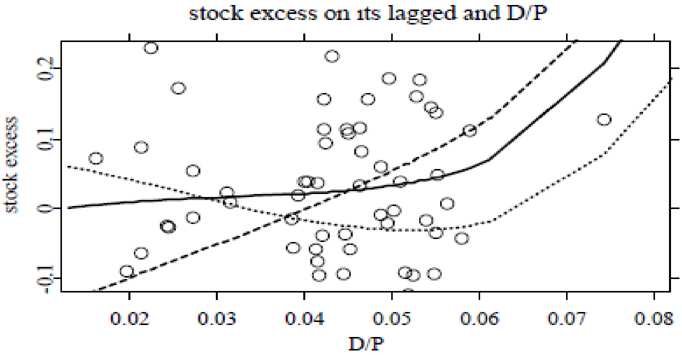
- amount of information
- predicted time period
- time horizon considered

The optimal model graphically





Looking at slices



Nonparametric regression fits of stock excess on D/P with stock excess lagged fixed at –25% (dotted, starting above zero), at 1% (solid), and at 30% (dashed) for the period 1923 – 1996.

Example for a conclusion

... this graph does show that Danish investors should have kept away for new investments in stocks in 2001, since they were just about to finish a magnificent year with a general Danish excess return on stocks above 30% *resulting in a historical low dividend-price ratio of around* 1.5%

Remark from a talk given to the Danish Actuarial Society in december 2000 under the title "Be careful : the Danish stocks are too expensive".

Extending period to 2001

- statistical evidence does not change curves and variables much
- main statements and findings still hold
- but estimated predictive power leaves a much less optimistic impression of possibility of predicting stock returns
- perhaps not surprising for followers of the stock market that the last five years, 1997-2001, have been unusual.
- all considered linear models break down in contradiction to Fama and French (1988), Wilkie (1993) and others
- optimal R_V^2 is reached for $T = 4$ only including $d = D/P$
- For $T = 1$ best model for 1922 – 2001 only uses S_{t-1}
- but including d_{t-1} gives almost same result [careful to only use S_{t-1}]