Using Analysts' Forecasts for Stock Predictions - An Entropic Tilting Approach

Christoph Frey

Erasmus University Rotterdam

christoph.frey@gmail.com

October 27, 2017

Motivation

- Predicting asset returns is a popular exercise in finance
- Challenging task because of (i) estimation and model uncertainty, (ii) a substantial unpredictable component (shocks) in future stock returns and (iii) model exploitation by other market participants
- Recent evidence for predictability: model averaging Pettenuzzo and Ravazzolo (2016), time-varying-parameters Dangl and Halling (2012), stochastic volatility and predictive distributions Johannes et al. (2014)

But

- Most prediction models relate the future stock returns to past observations of asset specific variables such as the dividend yield
- Stock price is forward-looking as it reflects the expectations of market participants about the future cash flows
- In this paper, use forward-looking information (professional analyst forecasts) to make stock return predictions
- Approach similar to Metaxoglou et al. (2016), who improve equity premium forecasts by using option prices

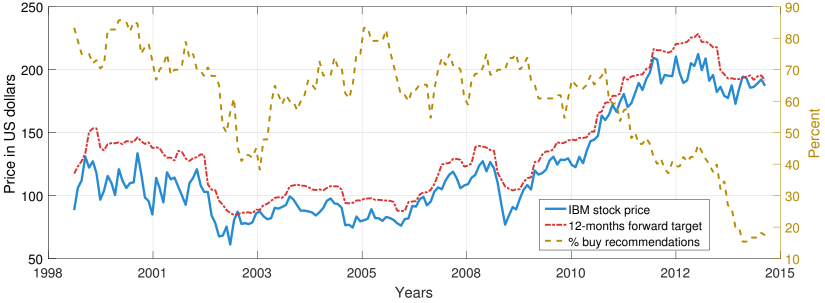
Motivation

- Predicting asset returns is a popular exercise in finance
- Challenging task because of (i) estimation and model uncertainty, (ii) a substantial unpredictable component (shocks) in future stock returns and (iii) model exploitation by other market participants
- Recent evidence for predictability: model averaging Pettenuzzo and Ravazzolo (2016), time-varying-parameters Dangl and Halling (2012), stochastic volatility and predictive distributions Johannes et al. (2014)

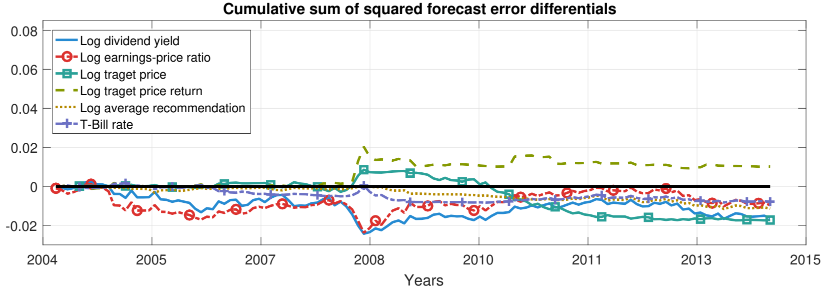
But

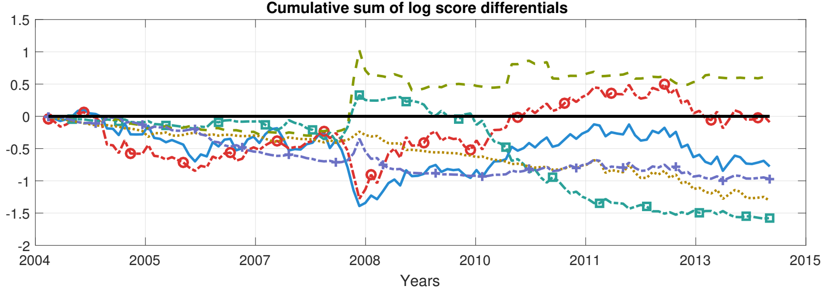
- Most prediction models relate the future stock returns to past observations of asset specific variables such as the dividend yield
- Stock price is forward-looking as it reflects the expectations of market participants about the future cash flows
- In this paper, use forward-looking information (professional analyst forecasts) to make stock return predictions
- Approach similar to Metaxoglou et al. (2016), who improve equity premium forecasts by using option prices

Motivation - IBM Stock Price



Motivation - IBM Return Prediction





Contributions

In this paper,

- we adopt the so-called **forgetting factors** approach of Koop and Korobilis (2013) which allows for all the features recently found to be important to find significant return predictability: Time-varying parameters, stochastic volatility, parameter shrinkage as well as dynamic model averaging and variable selection
- we use the (dis-)agreement in the analysts' forecast to regularize/reweight the predictive return distribution via by a method called **entropic (exponential) tilting**
- we restrict the mean and variance of the predictive distribution to coincide with the mean and the variance of monthly target price implied expected returns, i.e. simple returns between the current spot and the target price
- we find that restricting the variance of the asset returns is beneficial in terms of out-of-sample performance, as it provides a forward looking measure for (un-)certainty in the market

Contributions

In this paper,

- we adopt the so-called **forgetting factors** approach of Koop and Korobilis (2013) which allows for all the features recently found to be important to find significant return predictability: Time-varying parameters, stochastic volatility, parameter shrinkage as well as dynamic model averaging and variable selection
- we use the (dis-)agreement in the analysts' forecast to regularize/reweight the predictive return distribution via by a method called **entropic (exponential) tilting**
- we restrict the mean and variance of the predictive distribution to coincide with the mean and the variance of monthly target price implied expected returns, i.e. simple returns between the current spot and the target price
- we find that restricting the variance of the asset returns is beneficial in terms of out-of-sample performance, as it provides a forward looking measure for (un-)certainty in the market

Contributions

In this paper,

- we adopt the so-called **forgetting factors** approach of Koop and Korobilis (2013) which allows for all the features recently found to be important to find significant return predictability: Time-varying parameters, stochastic volatility, parameter shrinkage as well as dynamic model averaging and variable selection
- we use the (dis-)agreement in the analysts' forecast to regularize/reweight the predictive return distribution via by a method called **entropic (exponential) tilting**
- we restrict the mean and variance of the predictive distribution to coincide with the mean and the variance of monthly target price implied expected returns, i.e. simple returns between the current spot and the target price
- we find that restricting the variance of the asset returns is beneficial in terms of out-of-sample performance, as it provides a forward looking measure for (un-)certainty in the market

Contributions

In this paper,

- we adopt the so-called **forgetting factors** approach of Koop and Korobilis (2013) which allows for all the features recently found to be important to find significant return predictability: Time-varying parameters, stochastic volatility, parameter shrinkage as well as dynamic model averaging and variable selection
- we use the (dis-)agreement in the analysts' forecast to regularize/reweight the predictive return distribution via by a method called **entropic (exponential) tilting**
- we restrict the mean and variance of the predictive distribution to coincide with the mean and the variance of monthly target price implied expected returns, i.e. simple returns between the current spot and the target price
- we find that restricting the variance of the asset returns is beneficial in terms of out-of-sample performance, as it provides a forward looking measure for (un-)certainty in the market

Motivation - Analysts Recommendations

How informative are sell-side based analyst forecasts for return prediction?

- Ramnath et al. (2008): HUGE literature about shortcomings of analyst forecasts: Incentives and biases, information content and market inefficiencies, mostly related to accounting figures (earnings forecasts)

Analysts recommendations

- Barber et al. (2001); Green (2006); Cvitanić et al. (2006) report abnormal returns from trading strategies that go long in stocks with favorable recommendations and short in stocks with unfavorable ratings

Twelve months forward target price

- Brav and Lehavy (2003) show that there are substantial short-term market reactions in the stock price to target changes
- Da and Schaumburg (2011) find that aggregating stocks across sectors according to their twelve month forward target price implied expected return, i.e. simple return between the current and the target price, yields significant risk-adjusted abnormal returns for different long-short portfolio

Motivation - Analysts Recommendations

How informative are sell-side based analyst forecasts for return prediction?

- Ramnath et al. (2008): HUGE literature about shortcomings of analyst forecasts: Incentives and biases, information content and market inefficiencies, mostly related to accounting figures (earnings forecasts)

Analysts recommendations

- Barber et al. (2001); Green (2006); Cvitanić et al. (2006) report abnormal returns from trading strategies that go long in stocks with favorable recommendations and short in stocks with unfavorable ratings

Twelve months forward target price

- Brav and Lehavy (2003) show that there are substantial short-term market reactions in the stock price to target changes
- Da and Schaumburg (2011) find that aggregating stocks across sectors according to their twelve month forward target price implied expected return, i.e. simple return between the current and the target price, yields significant risk-adjusted abnormal returns for different long-short portfolio

Motivation - Analysts Recommendations

How informative are sell-side based analyst forecasts for return prediction?

- Ramnath et al. (2008): HUGE literature about shortcomings of analyst forecasts: Incentives and biases, information content and market inefficiencies, mostly related to accounting figures (earnings forecasts)

Analysts recommendations

- Barber et al. (2001); Green (2006); Cvitanić et al. (2006) report abnormal returns from trading strategies that go long in stocks with favorable recommendations and short in stocks with unfavorable ratings

Twelve months forward target price

- Brav and Lehavy (2003) show that there are substantial short-term market reactions in the stock price to target changes
- Da and Schaumburg (2011) find that aggregating stocks across sectors according to their twelve month forward target price implied expected return, i.e. simple return between the current and the target price, yields significant risk-adjusted abnormal returns for different long-short portfolio

Outline of the Talk

- Prediction model
- Entropic tilting (Krüger et al., 2015)
- Financial Analysts' data
- Empirical application with 20 Dow Jones constituents
- Conclusions

Prediction Model I

- Kandel and Stambaugh (1987) propose a vector autoregression formulation to jointly model the dynamics of asset returns and its predictor:

$$r_{t+1} = a_r + b_r x_t + \varepsilon_{r,t+1}, \quad (1)$$

$$x_{t+1} = a_x + b_x x_t + \varepsilon_{x,t+1}, \quad (2)$$

- the excess return and the predictor variable, only depend on their own lag
- Formulation for K predictor variables:

$$\begin{bmatrix} r_t \\ x_t \end{bmatrix} = a + \sum_{i=1}^{p} A_i \begin{bmatrix} r_{t-i} \\ x_{t-i} \end{bmatrix} + \varepsilon_t, \quad t = 1, \ldots, T, \quad (3)$$

where r_t is the excess return of a particular stock, $x_t = [x_{1,t}, \ldots, x_{K,t}]'$ is a $K \times 1$ vector of predictor variables and $\varepsilon_t \overset{iid}{\sim} \mathrm{N}(0, \Sigma)$

Prediction model II

- Restrict the system such that r_t depends on the entire x_{t-1} vector but $x_{k,t}$, $1 \leq k \leq K$, only depends on its own lag $x_{k,t-1}$. Compactly, the resulting model is of the form

$$y_t = (r_t, x_t)' = a + A_1 y_{t-1} + \varepsilon_t, \tag{4}$$

where $a = (a_r, a_{x_1}, \ldots, a_{x_K})'$ and

$$A_1 = \begin{pmatrix} 0 & A_1^{1,2} & A_1^{1,3} & \cdots & A_1^{1,K+1} \\ 0 & A_1^{2,2} & 0 & \cdots & 0 \\ \vdots & \ddots & A_1^{3,3} & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & A_1^{K+1,K+1} \end{pmatrix}$$

- To implement these restrictions on the coefficient matrix A_1 we use a variant of the Minnesota prior (Doan et al., 1984)

Prior Model specification for VAR(1) system

- Independent marginal normal priors for each set of parameters:

$$p(a) \sim \mathrm{N}\big(0, \zeta \times \mathrm{I}_{(K+1\times K+1)}\big)\,, \tag{5}$$

$$p(A_1^{1,1}) \sim \mathrm{N}(0, \varrho \times 1)\,, \tag{6}$$

$$p(A_1^{1,k}) \sim \mathrm{N}\left(0, \zeta \times \frac{\sigma_r^2}{\sigma_{x_k}^2}\right), \quad k = 1, \ldots, K \tag{7}$$

$$p(A_1^{k,1}) \sim \mathrm{N}\left(0, \varrho \times \frac{\sigma_{x_k}^2}{\sigma_r^2}\right), \quad k = 2, \ldots, K, \tag{8}$$

$$p(A_1^{k,l}) \sim \mathrm{N}\left(\underline{A}_1^{k,l}, \varrho \times \frac{\sigma_{x_k}^2}{\sigma_{x_l}^2}\right), \quad k = 2, \ldots, K, \quad l = 2, \ldots, K \tag{9}$$

with $\underline{A}_1^{k,l} = d_k$ if $k = l$, and $\underline{A}_1^{k,l} = 0$ otherwise.

- $d_k = 0$ for real, $d_k = 0.8$ for nominal variables, $\varrho = 10^{-4}$, $\zeta = 0.2$
- Prior is centered around zero implying no predictability.
- $\sigma_{x_k}^2$ $\forall k, r$ are approximated by the residual variances of AR(1) regressions
- Diffuse prior for Σ, $p(\Sigma) \propto |\Sigma|^{-2(2M+1)/2}$.

Time-varying Bayesian VAR and stochastic volatility

- Numerous examples favoring equity prediction models with time-varying parameters (TVP) (Dangl and Halling, 2012), stochastic volatility (SV) (Johannes et al., 2014) and BMA (Pettenuzzo and Ravazzolo, 2016)
- MCMC/simulations methods computational too costly to estimate VAR system with many predictor variables for recursive forecasting
- Solution: *forgetting factors* approach of Koop and Korobilis (2013)
- Consider a time-varying VAR version of (4) with stochastic volatility:

$$y_t = a_t + A_{1,t}\, y_{t-1} + \varepsilon_t, \tag{10}$$

$$A_t = \phi A_{t-1} + (1-\phi)\underline{A}_0 + u_t, \tag{11}$$

where $A_t = [a_t \;\; A_{1,t}]$, $\varepsilon_t \overset{iid}{\sim} \mathrm{N}(0, \Sigma_t)$, $u_t \overset{iid}{\sim} \mathrm{N}(0, \Omega_t)$ and $\varepsilon_t \perp u_s \forall\, t, s$

- $\phi = 1$ implies a random walk behavior, $\phi = 0$ implies a random behavior of each A_t around $\underline{A}_0$

Forgetting factors I

- Typically, the estimation of (10) - (11) involves drawing A_t conditional on Σ_t and Ω_t (e.g. through a Kalman filter), then drawing Σ_t conditional on A_t and Ω_t, the sampling Ω_t given A_t and Σ_t and eventually drawing further parameters given conditional on A_t, Σ_t, and Ω_t for all t.
- Computationally demanding as it involves simulating Σ_t, and Ω_t for every $t = 1, \ldots, T$.
- **Idea of the forgetting factors:** avoid simulating Ω_t in the Kalman filter by approximating the one-step ahead predictor variance of $A_t|y^{t-1} \sim \mathrm{N}\big(A_{t|t-1}, P_{t|t-1}\big)$, i.e. $P_{t|t-1}$, by the variance of the filtered estimator $A_{t-1}|y^{t-1} \sim \mathrm{N}\big(A_{t-1|t-1}, P_{t-1|t-1}\big)$, i.e. $P_{t-1|t-1}$, divided by a *forgetting factor* $\lambda \in [0, 1]$:

$$P_{t|t-1} = P_{t-1|t-1}/\lambda \tag{12}$$

- Then approximate Ω_t by $(\lambda^{-1} - 1)P_{t-1|t-1}$.
- Further, $\hat{\Sigma}_t = \kappa\hat{\Sigma}_{t-1} + (1-\kappa)\hat{\varepsilon}_t\hat{\varepsilon}_t'$, where $\hat{\varepsilon}_t = y_t - A_{t|t}[1\ y_{t-1}]$ is obtained in the Kalman filter

Forgetting factors I

- Typically, the estimation of (10) - (11) involves drawing A_t conditional on Σ_t and Ω_t (e.g. through a Kalman filter), then drawing Σ_t conditional on A_t and Ω_t, the sampling Ω_t given A_t and Σ_t and eventually drawing further parameters given conditional on A_t, Σ_t, and Ω_t for all t.
- Computationally demanding as it involves simulating Σ_t, and Ω_t for every $t = 1, \ldots, T$.
- **Idea of the forgetting factors:** avoid simulating Ω_t in the Kalman filter by approximating the one-step ahead predictor variance of $A_t|y^{t-1} \sim \mathrm{N}\left(A_{t|t-1}, P_{t|t-1}\right)$, i.e. $P_{t|t-1}$, by the variance of the filtered estimator $A_{t-1}|y^{t-1} \sim \mathrm{N}\left(A_{t-1|t-1}, P_{t-1|t-1}\right)$, i.e. $P_{t-1|t-1}$, divided by a *forgetting factor* $\lambda \in [0, 1]$:

$$P_{t|t-1} = P_{t-1|t-1}/\lambda \tag{12}$$

- Then approximate Ω_t by $(\lambda^{-1} - 1)P_{t-1|t-1}$.
- Further, $\hat{\Sigma}_t = \kappa\hat{\Sigma}_{t-1} + (1-\kappa)\hat{\varepsilon}_t\hat{\varepsilon}'_t$, where $\hat{\varepsilon}_t = y_t - A_{t|t}[1\ y_{t-1}]$ is obtained in the Kalman filter

Kalman filter with forgetting factors

For $t = 1, \ldots, T$ it is as follows:

I. Prediction step:

1. Set $A_{t|t-1} = \phi A_{t-1|t-1} + (1-\phi)\underline{A}_0$.
2. Set $P_{t|t-1} = \frac{1}{\lambda} P_{t-1|t-1}$
   where for $t = 1$ we set $A_{0|0} = \underline{A}_0$ and $P_{0|0} = \underline{P}_0$.

II. Update step:

1. Calculate $\tilde{\varepsilon}_t = y_t - a_{t|t-1} + A_{t|t-1}\, y_{t-1}$.
2. Calculate $\hat{\Sigma}_t = \kappa \hat{\Sigma}_{t-1} + (1-\kappa)\tilde{\varepsilon}_t' \tilde{\varepsilon}_t$ with $\hat{\Sigma}_1 = \kappa \Sigma_0$.
3. Estimate $A_{t|t} = A_{t|t-1} + P_{t|t-1}[1 y_{t-1}]' \left(\hat{\Sigma}_t + [1 y_{t-1}] P_{t|t-1} [1 y_{t-1}]'\right)^{-1} \tilde{\varepsilon}_t$.
4. Calculate
   $P_{t|t} = P_{t|t-1} + P_{t|t-1}[1 y_{t-1}]' \left(\hat{\Sigma}_t + [1 y_{t-1}] P_{t|t-1} [1 y_{t-1}]'\right)^{-1} P_{t|t-1}$.

The one-step ahead predictive density of the VAR model is

$$p(y_t|y^t) \sim \mathrm{N}\big([1\ y_{t+1}] A_{t+1|t}, \hat{\Sigma}_{t+1} + [1\ y_{t+1}] A_{t+1|t} [1\ y_{t+1}]'\big)\,. \qquad (13)$$

Kalman filter with forgetting factors

For $t = 1, \dots, T$ it is as follows:

I. Prediction step:

1. Set $A_{t|t-1} = \phi A_{t-1|t-1} + (1-\phi)\underline{A}_0$.
2. Set $P_{t|t-1} = \frac{1}{\lambda} P_{t-1|t-1}$
   where for $t = 1$ we set $A_{0|0} = \underline{A}_0$ and $P_{0|0} = \underline{P}_0$.

II. Update step:

1. Calculate $\tilde{\varepsilon}_t = y_t - a_{t|t-1} + A_{t|t-1}\, y_{t-1}$.
2. Calculate $\hat{\Sigma}_t = \kappa\hat{\Sigma}_{t-1} + (1-\kappa)\tilde{\varepsilon}_t'\tilde{\varepsilon}_t$ with $\hat{\Sigma}_1 = \kappa\Sigma_0$.
3. Estimate $A_{t|t} = A_{t|t-1} + P_{t|t-1}[1 y_{t-1}]' \left(\hat{\Sigma}_t + [1 y_{t-1}]P_{t|t-1}[1 y_{t-1}]'\right)^{-1} \tilde{\varepsilon}_t$.
4. Calculate
   $P_{t|t} = P_{t|t-1} + P_{t|t-1}[1 y_{t-1}]' \left(\hat{\Sigma}_t + [1 y_{t-1}]P_{t|t-1}[1 y_{t-1}]'\right)^{-1} P_{t|t-1}$.

The one-step ahead predictive density of the VAR model is

$$p(y_t|y^t) \sim \mathrm{N}\left([1\ y_{t+1}]A_{t+1|t}, \hat{\Sigma}_{t+1} + [1\ y_{t+1}]A_{t+1|t}[1\ y_{t+1}]'\right). \qquad (13)$$

Dynamic model averaging and selection

- Specification of the model involves a set of prior parameters, namely λ, κ and ϕ
- Here, *dynamic model selection and averaging* technique as in Raftery et al. (2010) over a grid of three parameters
- Weights for model j, i.e. j-th combination of λ, κ and ϕ, at time t using all the information up to $t-1$ are given by

$$\omega_{t|t-1,j} = \omega^{\alpha}_{t-1|t-1,j} \Big/ \sum_{j=1}^{J} \omega^{\alpha}_{t-1|t-1,j}, \text{ and} \tag{14}$$

$$\omega_{t|t,j} = \omega_{t|t-1,j} p_j(y_t|y^{t-1}) \Big/ \sum_{j=1}^{J} \omega_{t|t-1,j} p_j(y_t|y^{t-1}), \tag{15}$$

  where $p_j(y_t|y^{t-1})$ is the predictive likelihood of model j evaluated at y_t and $\alpha = 0.99$ is a decay factor governing the weighting of past obs.
- In total, 48 models for the grid $\lambda \in \{0.97, 0.98.0.99, 1\}$, $\kappa \in \{0.94, 0.96.0.98\}$ and $\phi \in \{0, 0.5, 0.75, 1\}$

Entropic tilting I

- Non-parametric method to combine time-series model forecasts with information from other origins
- At time t we want to make a forecast h periods ahead for a $N \times 1$ vector of interest r_{t+h}
- Denote by $f_{t,h} := \{r_{t+h,i}\}_{i=1}^{I}$, where $r_{t+h} \in \mathbb{R}^N$ and $N \geq 1$, a baseline sample from a predictive return distribution $p(r_{t+h}|r^t)$, i.e. a discrete sample of I (MCMC) draws of the h-step ahead forecasts
- Incorporate additional information about the return r_{t+h}, which was not used to generate the base sample, in the form of M moment conditions on the function $g(r_{t+h}) : \mathbb{R}^N \to \mathbb{R}^M$ in the following sense:

$$\mathbb{E}\left[g(r_{t+h})\right] = \bar{g}_t, \tag{16}$$

where $\bar{g}_t \in \mathbb{R}^M$ and $M, N \geq 1$

Entropic tilting II

- For example $g(r_{t+h}) = r_{t+h}$ imposes that the mean of r_{t+h} is equal to $\bar{g}_t$ and $g(r_{t+h}) = (r_{t+h} - \mathbb{E}(r_{t+h}))^2$ sets the variance equal to it.
- $\bar{g}_t$ can be formed from various origins: Giacomini and Ragusa (2014) use an Euler equation to specify $\bar{g}_t$, Altavilla et al. (2014); Krüger et al. (2015) use survey forecasts and Metaxoglou et al. (2016) adopt option-implied information for $\bar{g}_t$.
- In general under the base density $f_{t,h}$, the moments of $g(r_{t+h})$ are not equal to $\bar{g}_t$:

$$\mathbb{E}_{f_{t,h}}\left[g(r_{t+h})\right] = \int g(r_{t+h}) f_{t,h}(r_{t+h})\, dr_{t+h} \neq \bar{g}_t. \tag{17}$$

- Entropic tilting describes finding the density $\tilde{f}_{t,h}$ out of the set of densities that fulfill the moment condition in (16) that is closest to the base density in terms of the Kullback-Leibler divergence measure

Entropic tilting III

Proposition

If a solution $\tilde{f}_{t,h}(r)$ to the constrained minimization

$$\min_{\tilde{f}_{t,h}\in\mathcal{F}} \mathbb{E}_{\tilde{f}_{t,h}}\left[\log\frac{\tilde{f}_{t,h}(r)}{f_{t,h}(r)}\right] = \int \log\frac{\tilde{f}_{t,h}(r)}{f_{t,h}(r)}\tilde{f}_{t,h}(r)\,dr, \qquad (18)$$

$$\text{s.t. } \mathbb{E}_{\tilde{f}_{t,h}}[g(r)] = \int g(r)\tilde{f}_{t,h}(r)\,dr = \bar{g}_t, \qquad (19)$$

exists, then it is unique and it is given by

$$\tilde{f}^{*}_{t,h}(r) = f_{t,h}(r)\exp\left(\gamma^{*\prime}_{t,h}g(r)\right) \Big/ \int \exp\left(\gamma^{*\prime}_{t,h}g(r)\right) f_{t,h}(r)\,dr, \qquad (20)$$

$$\gamma^{*}_{t,h} = \arg\min_{\gamma_{t,h}} \int f_{t,h}(r)\exp\left(\gamma^{\prime}_{t,h}(g(r)-\bar{g}_t)\right)dr. \qquad (21)$$

- Proof is given in Giacomini and Ragusa (2014)

Entropic tilting IV

- For a sample of I draws from the base predictive density, the expectation (KLIC) in (18) is

$$\mathbb{E}_{\tilde{f}_{t,h}}\left[\log\frac{\tilde{f}_{t,h}(r)}{f_{t,h}(r)}\right] = \sum_{i=1}^{I}\tilde{\pi}_i\log\left(\frac{\tilde{\pi}_i}{\pi_i}\right) \overset{\pi_i=1/I}{=} \log I + \sum_{i=1}^{I}\tilde{\pi}_i\log\left(\tilde{\pi}_i\right), \quad (22)$$

where π_i, $i = 1, \dots, I$ are the original weights for the base density

- Following Robertson et al. (2005), imposing $\mathbb{E}_{\tilde{f}_{t,h}}[g(r)] = \sum_{i=1}^{I}\tilde{\pi}_i g(r_{t,i})$ yields the tilting solution from (20) and (21) as

$$\pi_i^* = \frac{\exp\left(\gamma_{t,h}^{*\prime} g(r_{t+h,i})\right)}{\sum_{i=1}^{I}\exp\left(\gamma_{t,h}^{*\prime} g(r_{t+h,i})\right)}, \quad (23)$$

$$\gamma_{t,h}^* = \arg\min_{\gamma_{t,h}}\sum_{i=1}^{I}\exp\left(\gamma_{t,h}^{\prime}(g(r_{t+h,i}) - \bar{g}_t)\right). \quad (24)$$

- Solution (23) easily obtained by a Lagrangian optimization
- Entropic tilting also has a shrinkage interpretation

A Gaussian Example

- Consider bivariate normal with $f(y) = N(\theta, \Sigma)$ with the restriction that the mean of the second variable y_2 is μ_2 and its variance is Ω_{22}. Let γ_1 and γ_2 be the Lagrange multiplier, then

$$\tilde{f}^*(y) = cf(y)\exp(\gamma_1 y_2 + \gamma_2 y_2^2) \tag{25}$$

- With completing the square is follows that $\tilde{f}^*(y) = N(\mu, \Omega)$ with

$$\mu_1 = \theta_1 + \Sigma_{22}^{-1}\Sigma_{12}(\mu_2 - \theta_2) \tag{26}$$

$$\Omega_{12} = \Sigma_{12}\Sigma_{22}^{-1}\Omega_{22} \tag{27}$$

$$\Omega_{11} = \Sigma_{22}^{-1}(\Sigma_{11}\Sigma_{22} - \Sigma_{21}\Sigma_{12}) + \Omega_{22}(\Sigma_{22}^{-1}\Sigma_{21}) \tag{28}$$

- If $\Omega_{22} = 0 \Rightarrow$ Conditional bivariate normal distribution.
- If $\Omega_{22} = \Sigma_{22} \Rightarrow$ tilted and un-tilted variances of y_2 are equal.
- If $\Omega_{22} < \Sigma_{22} \Rightarrow$ variance reduction for y_2.

Entropic Tilting - General Methodology IV

Pros and Cons

+ Entropic tilting is a non-parametric method
+ Easy to incorporate any kind of information
+ Solution is easily obtained

– Solution strictly fulfills moment condition
– Uncertainty around the *moment information* is not measured

Entropic Tilting - General Methodology IV

Pros and Cons

+ Entropic tilting is a non-parametric method
+ Easy to incorporate any kind of information
+ Solution is easily obtained

− Solution strictly fulfills moment condition
− Uncertainty around the *moment information* is not measured

Data and Ste-up I

- All data obtained from Thomson Reuters Datastream.
- Investigate forecast performance of various prediction models for 20 Dow Jones constituents for which the Institutional Brokers Estimate System (I/B/E/S) database provide target prices and analyst recommendations
- I/B/E/S is available from April 1999 to October 2014
- Monthly data, expanding window estimation with initial estimation window has size $h = 60$
- Predictor variables: (i) firm specific *fundamentals* such as the log dividend yield, the log earnings price ratio, the log dividend-payout ratio, and the book-to-market ratio, (ii) market and economic measures such as the 3-month T-bill rate, the yield on long-term government bonds, the market excess return and CPI inflation

I/B/E/S Overview

- Thomson Reuters I/B/E/S provides detailed and consensus estimates featuring up to 26 forecast measures for more than 70,000 companies in more than 90 countries worldwide
- **Price Targets** (detail and summary): The mean and standard deviation of the projected price level forecasted by professional analysts with a 12-month time horizons. At each point in time, we also consider the entire vector of target prices from individual analysts.
- **Recommendations summary**: The mean and standard deviation of analysts' recommendation based on a five point standardized scale (strong buy = 1, buy, hold, sell, strong sell = 5) as well as the total number of recommendations, the number of up- and downgrade revisions and the percentage of buy, hold and sell recommendations.
- Use target price returns and variance as well as the consensus analyst recommendations and their revisions as predictor variables.

Titling Moments

(i) Monthly forward target price implied expected return, i.e. simple returns between the spot and the twelve months forward target price at each point t divided by twelve

- Mean target returns provides a measure for the market direction

(ii) Monthly target price implied expected return variances: Use target prices of individual analysts and first calculate monthly forward target price implied expected returns for every individual analyst and then use the mean and variance of these returns as first and second moment restrictions

- Target return variance provides a measure for (un-)certainty in the market

Titling Moments

(i) Monthly forward target price implied expected return, i.e. simple returns between the spot and the twelve months forward target price at each point t divided by twelve

- Mean target returns provides a measure for the market direction

(ii) Monthly target price implied expected return variances: Use target prices of individual analysts and first calculate monthly forward target price implied expected returns for every individual analyst and then use the mean and variance of these returns as first and second moment restrictions

- Target return variance provides a measure for (un-)certainty in the market

Evaluation Criteria

- We consider two evaluation criteria in this study.

1. Out-of-sample R^2:

$$R^2_{OoS,j} = 1 - \frac{\sum_{i=h+1}^{T} e_{j,i}^2}{\sum_{i=h+1}^{T} e_{0,i}^2}, \tag{29}$$

   where $e_{0,i} = y_i^1 - \hat{y}_{0,i}^1 = r_i - \hat{r}_{0,i}$ denotes the forecast error of a simple mean or intercept only model ($r_i = a + \varepsilon_i$) and $e_{j,i}$ the forecast error in the returns of model j at time i

2. Average log score differential (LSD):

$$\mathrm{LSD}_{j,t} = \frac{\sum_{i=h+1}^{t} (\mathrm{LS}_{j,i} - \mathrm{LS}_{0,i})}{\sum_{i=h+1}^{t} \mathrm{LS}_{0,i}}, \tag{30}$$

   where where $\mathrm{LS}_{j,i}$ is log predictive score of model j at time i

- Values above zero indicate that model j produces lower forecasts error than the intercept only model

Descriptive statistics on the returns, target prices and recommendations for 20 Dow Jones constituents (sample: 1999 - 2015)

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
Mean log ret	-0.52	2.01	-1.79	0.26	0.41	0.43	-0.05	-0.18	-0.35	0.27
Std log return	12.00	14.13	21.28	9.52	8.95	10.13	6.03	8.27	8.74	8.22
# price tragets	14.38	25.19	13.17	16.61	16.18	13.99	13.11	11.81	13.33	18.39
Mean exp ret	1.44	1.44	2.52	1.12	1.03	1.08	0.95	1.30	1.31	1.10
Std exp ret	10.72	7.49	25.95	6.08	7.04	7.20	4.57	5.79	6.07	6.37
# RECs	18.40	32.33	20.19	21.36	23.18	20.09	17.73	16.48	17.07	25.76
Mean RECs	2.35	2.13	2.19	2.36	2.26	2.29	2.11	2.40	2.02	2.11
Std RECs	0.38	0.38	0.58	0.35	0.38	0.25	0.25	0.28	0.34	0.25
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
Mean log ret	-0.13	0.14	0.24	0.27	-0.24	-0.09	0.14	0.39	0.10	0.40
Std log return	11.57	7.76	5.26	6.54	7.88	8.90	5.78	7.25	5.76	7.91
# price tragets	28.66	17.20	14.66	14.92	15.53	23.65	13.34	14.89	17.98	20.19
Mean exp ret	1.34	0.95	0.71	1.10	0.97	1.53	0.86	0.95	1.06	1.26
Std exp ret	7.91	5.29	3.68	5.40	5.91	6.91	3.80	4.52	4.16	6.62
# RECs	39.53	23.10	23.78	20.60	24.28	32.73	18.77	19.98	26.02	27.22
Mean RECs	2.18	2.18	2.10	2.16	2.42	1.91	2.09	1.95	2.05	2.29
Std RECs	0.29	0.26	0.25	0.26	0.36	0.26	0.21	0.24	0.25	0.25

Out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) for various forecasting models

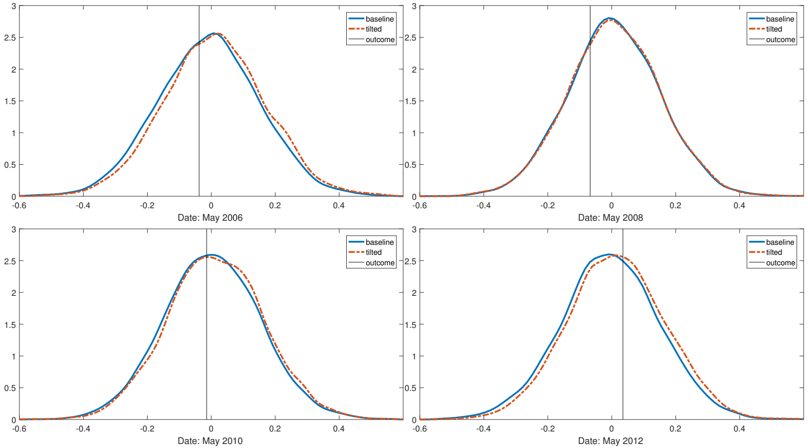
Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
AR1	-0.10	0.09	-0.07	0.33	0.06	-0.18	-0.36	-0.28	-0.11	-0.17
VAR-Full	-1.16	-1.57	-1.78	-1.25	-0.57	-1.08	-1.86	-0.50	-0.33	-0.71
VAR-Minnesota	0.74	0.67	0.88	0.39	0.65	0.11	0.47	-0.09	0.33	0.22
TVPVAR-DMA	0.85	0.67	-0.03	0.39	0.91	0.30	0.04	0.55	0.27	0.39
TVPVAR-DMS	0.62	0.37	-0.09	0.15	0.85	0.00	-0.20	0.43	0.06	0.24
TVPVAR-DMAm	0.89	0.70	-0.02	0.41	0.94	0.35	0.09	0.60	0.31	0.41
TVPVAR-DMAm/v	1.91***	0.80	0.02	0.72	1.55**	0.88	1.04*	1.15*	0.75	0.98
TVPVAR-DMSm	0.63	0.38	-0.07	0.20	0.87	0.02	-0.19	0.47	0.08	0.28
TVPVAR-DMSm/v	1.66**	1.30**	0.31	1.05*	1.77**	0.79	0.53	0.83	0.34	0.55
Bayesian lasso	0.85	0.74	0.00	0.46	0.95	0.34	0.05	0.56	0.28	0.44

Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
AR1	-0.36	-0.41	-0.64	-0.62	-0.32	-0.32	-0.13	-0.09	-0.17	0.27
VAR-Full	-1.41	-2.36	-2.06	-1.24	-0.90	-2.02	-1.96	-1.37	-0.68	0.10
VAR-Minnesota	0.47	0.26	-0.43	-0.30	-0.19	0.35	0.44	0.08	-0.02	0.75
TVPVAR-DMA	0.39	0.50	-0.29	0.06	0.70	-0.24	0.48	0.34	0.15	0.55
TVPVAR-DMS	0.15	0.39	-0.31	-0.11	0.43	-0.30	0.35	0.12	0.14	0.27
TVPVAR-DMAm	0.44	0.53	-0.28	0.08	0.74	-0.21	0.52	0.39	0.20	0.58
TVPVAR-DMAm/v	1.17*	0.52	0.59	0.22	1.23**	0.04	0.89	1.07*	0.34	0.86
TVPVAR-DMSm	0.18	0.41	-0.27	-0.08	0.45	-0.26	0.40	0.14	0.16	0.30
TVPVAR-DMSm/v	0.47	0.92	-0.28	0.89	1.28**	-0.24	0.51	0.53	0.56	0.67
Bayesian lasso	0.41	0.59	-0.21	0.16	0.75	-0.22	0.50	0.40	0.23	0.59

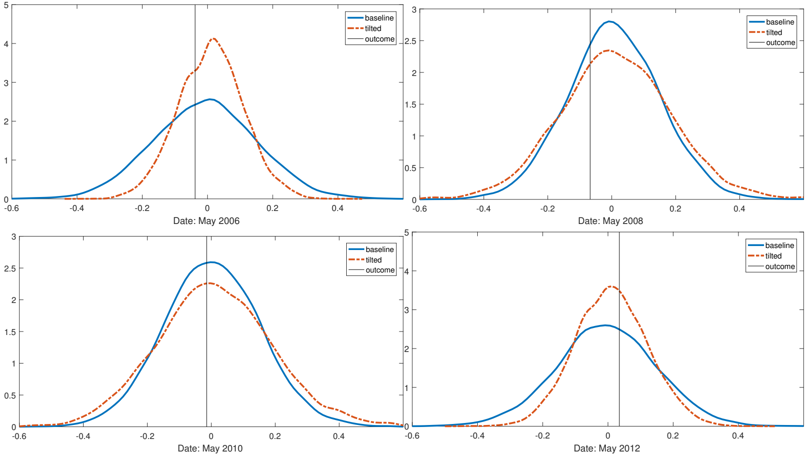
Average log predictive score differentials for 20 Dow Jones constituents (sample: 2004 - 2015) for various forecasting models

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
AR1	0.01	-0.01	-0.46	0.33	0.12	-0.09	-0.06	-0.01	0.05	-0.05
VAR-Full	-0.91	-1.29	-2.29	0.01	-1.31	-1.24	-0.93	-1.78	-0.74	-0.41
VAR-Minnesota	0.10	0.32	0.25	0.77	0.42	0.08	0.29	0.75	0.15	0.32
TVPVAR-DMA	0.21	0.74	-0.11	0.75	0.28	0.73	0.56	0.72	0.85	0.01
TVPVAR-DMS	-0.06	0.57	-0.28	0.49	0.27	0.47	0.44	0.71	0.63	-0.03
TVPVAR-DMAm	0.21	0.76	-0.07	0.76	0.32	0.77	0.60	0.77	0.87	0.03
TVPVAR-DMAm/v	0.80	1.83***	0.72	1.83***	0.54	1.31**	0.62	1.56**	1.52**	0.96
TVPVAR-DMSm	-0.02	0.60	-0.27	0.53	0.30	0.51	0.48	0.73	0.66	-0.01
TVPVAR-DMSm/v	0.88	1.55**	0.11	0.87	0.48	1.05*	0.82	1.56**	0.90	0.55
Bayesian lasso	0.23	0.84	-0.04	0.85	0.32	0.83	0.56	0.79	0.93	0.04
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
AR1	-0.09	-0.07	-0.08	-0.05	-0.09	-0.08	-0.04	-0.01	-0.01	0.18
VAR-Full	-1.35	-1.32	-0.73	-1.66	-2.09	-2.04	-0.29	-0.48	-0.06	-1.03
VAR-Minnesota	0.60	-0.03	0.48	0.73	0.13	0.03	0.51	0.72	0.48	0.90
TVPVAR-DMA	0.86	0.43	0.68	0.69	0.74	0.08	0.42	0.61	0.92	1.02*
TVPVAR-DMS	0.82	0.25	0.60	0.60	0.62	-0.04	0.30	0.42	0.87	0.96
TVPVAR-DMAm	0.89	0.45	0.72	0.73	0.77	0.09	0.45	0.62	0.95	1.04*
TVPVAR-DMAm/v	1.95***	1.45**	1.13*	0.69	1.33**	0.31	0.66	0.97	1.03*	1.84***
TVPVAR-DMSm	0.85	0.29	0.64	0.60	0.66	-0.04	0.32	0.46	0.91	0.98
TVPVAR-DMSm/v	1.01*	1.18*	0.83	1.14*	1.64**	0.40	0.46	0.80	1.28**	1.32**
Bayesian lasso	0.96	0.51	0.76	0.75	0.82	0.11	0.44	0.64	0.98	1.10*

Kernel Predictive Densities - IBM - Mean Tilting



Kernel Predictive Densities - IBM - Mean/Variance Tilting



Conclusions and Outlook

- We demonstrate that financial analyst forecasts can have predictive power for equity assets of the Dow Jones Industrial index
- Extent of predictability varies across assets but using Bayesian VARs with time-varying coefficients, stochastic volatility and model averaging and selection among priors improves return predictions across all assets
- Tilting the mean of the predictive distribution towards the target price implied expected returns does not improve forecast performance, but tilting mean and variance of the predictive return distribution improves forecasts

Possible extensions:

- Relax assumption of strict moment conditions
- Panel VAR systems for various assets to model joint predictive return distributions for finding optimal portfolios
- Apply the tilting framework to predictive portfolio weight regressions (Frey and Pohlmeier, 2015)

Conclusions and Outlook

- We demonstrate that financial analyst forecasts can have predictive power for equity assets of the Dow Jones Industrial index
- Extent of predictability varies across assets but using Bayesian VARs with time-varying coefficients, stochastic volatility and model averaging and selection among priors improves return predictions across all assets
- Tilting the mean of the predictive distribution towards the target price implied expected returns does not improve forecast performance, but tilting mean and variance of the predictive return distribution improves forecasts

Possible extensions:

- Relax assumption of strict moment conditions
- Panel VAR systems for various assets to model joint predictive return distributions for finding optimal portfolios
- Apply the tilting framework to predictive portfolio weight regressions (Frey and Pohlmeier, 2015)

References

Altavilla, C., R. Giacomini, and G. Ragusa (2014): "Anchoring the Yield Curve Using Survey Expectations," ECB Working Paper No. 1632, ECB.

Barber, B., R. Lehavy, M. McNichols, and B. Trueman (2001): "Can Investors Profit from the Prophets? Security Analyst Recommendations and Stock Returns," *The Journal of Finance*, 56, 531–563.

Brav, A. and R. Lehavy (2003): "An Empirical Analysis of Analysts' Target Prices: Short-Term Informativeness and Long-Term Dynamics," *The Journal of Finance*, 58, 1933–1967.

Cvitanić, J., A. Lazrak, L. Martellini, and F. Zapatero (2006): "Dynamic Portfolio Choice with Parameter Uncertainty and the Economic Value of Analysts' Recommendations," *Review of Financial Studies*, 19, 1113–1156.

Da, Z. and E. Schaumburg (2011): "Relative Valuation and Analyst Target Price Forecasts," *Journal of Financial Markets*, 14, 161–192.

Dangl, T. and M. Halling (2012): "Predictive regressions with time-varying coefficients," *Journal of Financial Economics*, 106, 157–181.

Doan, T., R. Litterman, and C. Sims (1984): "Forecasting and Conditional Projection using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100.

Frey, C. and W. Pohlmeier (2015): "Bayesian Regularization of Portfolio Weights," Tech. rep., Department of Economics, University of Konstanz, http://ssrn.com/abstract=2730475.

Giacomini, R. and G. Ragusa (2014): "Theory-Coherent Forecasting," *Journal of Econometrics*, 182, 145–155.

Green, T. C. (2006): "The Value of Client Access to Analyst Recommendations," *Journal of Financial and Quantitative Analysis*, 41, 1–24.

Johannes, M., A. Korteweg, and N. Polson (2014): "Sequential Learning, Predictability, and Optimal Portfolio Returns," *The Journal of Finance*, 69, 611–644.

Kandel, S. and R. F. Stambaugh (1987): "Long-horizon returns and short-horizon models," CRSP Working paper 222, University of Chicago.

Koop, G. and D. Korobilis (2013): "Large time-varying parameter VARs," *Journal of Econometrics*, 177, 185–198.

Krüger, F., T. E. Clark, and F. Ravazzolo (2015): "Using Entropic Tilting to Combine BVAR Forecasts with External Nowcasts," *Journal of Business & Economic Statistics*, 1–48.

Metaxoglou, K., D. Pettenuzzo, and A. Smith (2016): "Option-Implied Equity Premium Predictions via Entropic Tilting," *Available at SSRN 2754246*.

Pettenuzzo, D. and F. Ravazzolo (2016): "Optimal Portfolio Choice Under Decision-Based Model Combinations," *Journal of Applied Econometrics*, forthecoming.

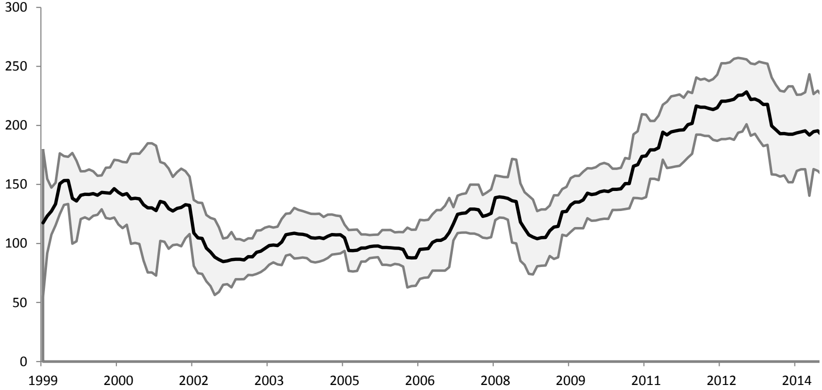
Raftery, A. E., M. Kárný, and P. Ettler (2010): "Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill," *Technometrics*, 52, 52–66.

Ramnath, S., S. Rock, and P. Shane (2008): "The financial analyst forecasting literature: A taxonomy with suggestions for further research," *International Journal of Forecasting*, 24, 34–75.

Robertson, J. C., E. W. Tallman, and C. H. Whiteman (2005): "Forecasting Using Relative Entropy," *Journal of Money, Credit and Banking*, 37, 383–401.

Motivation - IBM Stock Price II

- Mean target price $\pm$ 1.96 $\times$ S.D. of target prices



Predictor variables

- Company specific: Log DY, Log EPR, Log DPR, BMR,
- Macroeconomic: 3M Tbill rate, Longterm yield, Market return, CPI inflation
- Analyst information: Log TPR, Log TPRV, Log REC, Log REC return