Reading the Tea Leaves: Model Uncertainty, Robust Forecasts, and the Autocorrelation of Analysts' Forecast Errors

J.T. Linnainmaa, W. Torous and J.Yae

December 1, 2016

Table of Contents

Autocorrelation Puzzle

- For a one-period forecast, if analysts know the process and seek to minimize mean squared-error, forcast errors will have mean zero and be serially uncorrelated.
  - Empirical evidence that forecaster errors tend to be positive and auto-correlated.
  - This would imply that analysts do not learn from past mistakes. Why?

Motivating Example

- **Parameter uncertainty:**
  - $x_t = (a + u_t)x_{t-1} + \varepsilon_t$ where $u_t \sim N(0, \sigma^2)$
  - The error dissipates as analysts learn.
- **Model (Knightian) uncertainty**:
  - Analyst does not know the underlying model. They have an **approximating model.**

Robust Forecasting

- In a **robust forecast**, analysts overestimate
- Authors find that **variation in mean forecast errors** contributes to one-fifth of the measured autocorrelation
- Estimation errors of earnings growth shocks contributes another one-fifth of the measured autocorrelation
- Finally, model uncertainty contributes to 60% of the measured autocorrelation.

Why is this Important?

- Contributes to the literature on analyst behavior and asset pricing anomalies.
- Important regarding the question of efficient distribution of information and welfare.

Relation to other literature

Other literature

Uppal and Wang (2003), Maenhout (2004), and Epstein and Schneider (2008) suggest that model uncertainty is of **first-order importance** for portfolio choice and asset pricing.
Hilary and Hsu (2013) find analyst *consistency* rather than *accuracy* determines their ranking.

Earning and Signals Processes

Earnings Process

$$y_{t+1} = \mu + x_{t+1} + a_{t+1}$$

where y_{t+1} is the reported earnings, x_{t+1} is the persistent (**permanent)** component of the earnings process and a_{t+1} is noise.

Signal Process (Private Signal)

$$s_t = e_{t+1} + n_t$$

where e_{t+1} is the permanent earnings shock and $n_t \sim N(0, \sigma_n^2)$. All shocks have zero cross-correlations, autocorrelations and cross-autocorrelations.

Earning and Signals Processes

- The analyst's objective in period t is estimate y_{t+1} given the history of earnings and signals

$$E[y_{t+1}|s_t, s_{t-1}, ..., s_1, y_t, y_{t-1}, ..., y_1] = E[y_{t+1}|\mathscr{F}_t]$$

- This is the **linear** part of the model

The Uncertainty Environment

$$a_t^w = \kappa_0 + \kappa_1 a_t^*$$

where a^w is the worse case realization and a^* is the analyst's approximating model

- Analysts do not know the distribution but we assume they approximate this noise as $a_t \sim i.i.d. N(0, \hat{\sigma}_{a*}^2)$. The author's assume the approximated variance, $\hat{\sigma}_a^2$, is equal to the real variance σ_a^2 in order to ensure the approximating model is good.
- The actual realization is $a_t^w \sim N(\kappa_0, \kappa_1^2 \sigma_a^2)$, where κ_0 is a real number and κ_1 is a non-negative number. The realization is a function of a random draw from this distribution.

The Robust Forecasting Problem

$$\min_{\hat{y}_{t|t-1}} \max_{(\kappa_0,\kappa_1)} E[\{y_t^w - \hat{y}_{t|t-1}|\mathscr{F}_{t-1}]$$

subject to

$$E[\{\underbrace{(a_t^w - a_t^*)}_{Deviation} + \underbrace{(\hat{x}_{t|t-1}^w - \hat{x}_{t|t-1})}_{Perceived\ Bias}\}^2|\mathscr{F}_{t-1}] \leq \eta^2\sigma_a^2$$

where y^w is the worst *ex ante* outcome; $\hat{y}_{t|t-1}$ is the analyst's optimal forecast given information hitherto; $\hat{x}_{t|t-1}^w$ is the optimal forecast of x_t (using a Kalman filter) under the worst case; and $\hat{x}_{t|t-1}$ is the optimal forecast forecast of x_t given the analyst's expectations of the "evil agent's" choice of κ_0 and κ_1. Finally, a_t^w is the worse case realization of a_t, whereas a^* is the approximating estimate.

Direct and Indirect Effects

- $(a_t^w - a^*)$ is the direct effect. This expresses the amount of distortion induced by the "evil agent."
- $(\hat{x}_{t||t-1}^w - \hat{x}_{t|t-1})$ is the indirect effect from the analysts relying on inaccurate historical information in their future estimations.
- η measures the agent's concern for model misspecification and σ_a^2 the variance of the noise induced by the "evil agent." Thus, $\eta\sigma_a^2$ is the degree of robustness in the model. As $\eta \to \infty$, the entropy becomes so great that it becomes impossible for the analyst to distinguish models. When $\eta = 0$, we have a standard Rational Expectations model.

Minimax Optimization

- The analyst solves a static optimization problem: the forecasts are independent from her last forecasts and the same solution applies at every date t.
- The analyst knows the parameters of the true earnings process completely determine her current estimate ($\hat{y}_{t|t-1}$). In other words, after choosing ($\hat{\kappa}_0, \hat{\kappa}_1$), their estimate of the "evil agent's" noise process, the analyst obtains an optimal forecast using a Kalman filter.

Intuition behind Lemma 2.1

- The forecast is a function of the previous forecast $\hat{y}_t$, the forecast error $(y_t - \hat{y}_t)$ and the additional signal s_t.
- The Kalman gain K captures how much the analyst uses previous forecast errors to revise estimates of x_t
- The weight w measures how much the analyst uses the extra-signal s_t to estimate e_{t+1}, the permanent growth shock.

Intuition behind Prop. 2

- If $\hat{\theta} = \theta$- that is the analyst predicts the **true values** of the model - autocorrelation of forecast errors goes to zero.
- With robust forecasting, analyst knows everything but the distribution of the noise a_t. The first term goes to zero but the second two terms are strictly positive.

Intuition behind Robust Forecasting

- Analysts concerned about *model misspecification* will issue forecasts that perform well under the worst cast (highest variance).
- The analyst will overestimate the amount of noise in reported earnings (y_t) in order to achieve better accuracy than expected. Why?
  - The noisier the reported earnings, the less accurate the analyst's forecast will be.
  - The analyst's inference of x_t will be farther away, on average, from the actual state.
- The analyst underreacts to historical earnings. As a result, we find **positive autocorrelation in forecast errors.**

Robustness in asset pricing versus forecasting

- In asset pricing literature, it is the **investor's preferences,** the structure of their utility function, which determine the worst-case scenario.
- In the forecasting problem, the decision maker has a preference for **accuracy**.

Parameter vs Model uncertainty

- Collin-Dufresne, Johannes, and Lochstoer (2013, 2015) show that if investors have **recursive preferences**, rational parameter learning generates *subjective*, long-run risks. Why?
  - The investors **can learn or know** the true model. They face **parameter uncertainty.**
    - The shocks are therefore **permanent** and affect all future periods of consumption
- With a robust decision maker, the analyst accepts model misspecification as a permanent state of affairs. **They focus on robust controls.**

Data Sources

- Combine data from I/B/E/S, Compustat and the the Center for Research in Securities Prices (CRSP).
- Use data from January 1984 to December 2013.
- Match firms against Compustat and CRSP: firms must be listed on NYSE, Nasdaq or AMEX.
- Sample Selection rules (to control for outliers):
    - Delete observations with beginning of the quarter stock price below $5.
    - Delete observations where the forecasted year-to-year change in quarterly earnings per share is greater than $10 in absolute value
    - Trim extreme values (1% and 99%) for earnings, forecasts and forecast errors.
    - Require a firm to have at least 20 observations of actual earnings and forecasts.

Table 1
Descriptive statistics, 1984–2013

This table reports the distributions of year-to-year quarterly earnings growth (y_t), forecasted earnings growth ($\hat{y}_t$), and forecast errors. The data combine IBES, Compustat, and CRSP data from 1984 through 2013. See the text for details on sample construction. Autocorrelations are estimated from AR(1) regressions. The pooled estimate uses data on all firms and firm-specific estimates are the average estimates from firm-specific regressions. Short-lived firms ($N = 3{,}349$) are firms with fewer than 20 quarterly observations. Long-lived firms ($N = 3{,}804$) are firms with at least 20 quarterly observations. Except for these firm-specific autocorrelation estimates, the short-lived firms are not part of the main sample. The main sample therefore has 185,420 firm-quarter observations on 3,804 firms that survive for at least five years. The bias-adjusted autocorrelation estimates $\hat{\rho}_{\text{bias-adjusted}}$ correct raw estimates $\hat{\rho}$ for Kendall's (1954) small-sample bias, $\hat{\rho}_{\text{bias-adjusted}} = \frac{\hat{\rho}(T-1)+1}{T-4}$, where T is the number of observations.

Statistic	Year-to-year earnings growth, y_t	Forecasted growth, $\hat{y}_t$	Forecast error, $y_t - \hat{y}_t$: Estimate	Forecast error, $y_t - \hat{y}_t$: t-value
Mean	0.004	0.000	0.004	2.07
SD	1.113	0.992	0.501	
Percentiles				
5%	−1.615	−1.398	−0.607	
25%	−0.250	−0.246	−0.037	
50%	0.030	0.015	0.020	13.37
75%	0.340	0.289	0.119	
95%	1.442	1.294	0.563	
Negative	46.0%	47.7%	32.4%	
Zero	0.2%	0.1%	10.3%	
Positive	53.8%	52.2%	57.3%	
Pooled autocorrelation	0.429	0.434	0.216	28.87
Firm-specific autocorrelations				
All firms			0.102	20.12
All firms (bias-adjusted)			0.244	23.36
Short-lived firms			0.042	4.49
Short-lived firms (bias-adjusted)			0.299	13.79
Long-lived firms			0.153	31.80
Long-lived firms (bias-adjusted)			0.196	36.88

AR(1)-plus-noise

$$FE_{i,t+1} = \alpha + \rho FE_{i,t} + \varepsilon_{i,t+1}$$

The pooled estimate of the autocorrelation in forecast errors, 0.216, is significant with a heteroscedasticity and autocorrelation consistent t-value of 28.87.

A Joint Model of earnings and forecasts

This whole system described above can be estimated as a VARMA(1,1):

$$Y_{t+1} = A + BY_t = C\varepsilon_{t+1} + D\varepsilon_t$$

where

$$Y_t = \begin{bmatrix} y_t \\ \hat{y}_t \end{bmatrix}, \quad \varepsilon_t = \begin{bmatrix} a_t \\ e_t \\ n_{t-1} \end{bmatrix}, \quad cov(\varepsilon_t) = \begin{bmatrix} \sigma_a^2 & 0 & 0 \\ 0 & \sigma_a^2 & 0 \\ 0 & 0 & \sigma_a^2 \end{bmatrix}$$

$A = \begin{bmatrix} \mu(1-\phi) \\ \hat{\mu}(1-\phi) \end{bmatrix}, \quad B = \begin{bmatrix} \phi & 0 \\ \hat{\phi}\hat{K} & \hat{\phi}(1-\hat{K}) \end{bmatrix}$ where μ is the long-term mean of y_t.

$$C = \begin{bmatrix} 1 & 1 & 0 \\ 0 & \hat{w} & \hat{w} \end{bmatrix}, \quad D = \begin{bmatrix} -\phi & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Estimation Procedure

1. Estimate the parameters of the AR-plus-noise using maximum likelihood.
2. Holding the parameter values fixed, estimate the rest of the VARMA model using conditional maximum likelihood.
   1. Use a block bootstrapping procedure, resampling at the *firm* level to preserve time series properties.
      1. Why block bootstrap? The errors are correlated, so simple residual resampling will fail. Rather, resample blocks of data.

Table 2
Estimates of a VARMA(1,1) model of earnings and analyst forecasts

This table presents parameter estimates from a VARMA(1,1) model that describes the evolution of firms' earnings and analyst forecasts. The data are analysts' earnings forecasts and actual earnings per share from 1984 through 2013 from IBES. We estimate the VARMA model in two steps. First, we fit the earnings dynamics with an AR(1)-plus-noise model using maximum likelihood estimation and a Kalman filter. Second, we use these estimates of the earnings process within the VARMA model to estimate the remaining parameters using maximum likelihood estimation and a Kalman filter. We report bootstrapped standard errors that draw firms as blocks with replacement. Rows labeled True report the estimated parameters of the earnings process; Implied are the parameters used by the analysts, as implied by their forecasts. The bottom part reports R^2s for the AR(1)-plus-noise model and for analysts' forecasts. The latter compares the variance of forecast errors to the variance of earnings growth, $R^2 = 1 - \text{var}(y_{t+1} - \hat{y}_{t+1})/\text{var}(y_{t+1})$.

Parameter	Estimate	SE
Persistence of the earnings-growth shocks		
True, ρ	0.472	0.006
Implied, $\hat{\rho}$	0.470	0.005
SD(noise term) / SD(earnings growth shock), σ_α/σ_e	0.106	0.073
SD(additional signal) / SD(earnings growth shock), σ_n/σ_e	0.538	0.022
Kalman gain		
True, K	0.953	0.059
Implied, $\hat{K}$	0.414	0.049
Weight placed on the additional signal		
True, w	0.775	0.014
Implied, $\hat{w}$	0.783	0.014
R^2s for predicting earnings growth		
AR(1)-plus-noise model	14.0%	
Analysts' median forecast (IBES)	79.8%	

Reliance on historical data

- Pseudo-R^2 of analyst forecasts $= 1 - \frac{var(y_{t+1} - \hat{y}_{t+1})}{var(y_{t+1})}$
- If analysts use only historical data, the precision of their forecasts should be comparable to the R^2 of the ARMA(1,1). Using the pseudo-R^2, the authors find the R^2 of the analyst's forecasts is 79.8%.

Which "mistakes" drive the autocorrelation in forecast errors?

- The optimal forecast is: $\hat{y}_t = (1-\hat{\phi})\hat{\mu} + \hat{\phi}\{\hat{y}_t + \hat{K}(FE_t)\} + \hat{w}s_t$
  - $\hat{\phi}$ is the belief about the persistence of earnings growth shocks
  - $\hat{K}$ summarizes the belief in the informativeness of the reported earnings growth
  - $\hat{w}$ is how much the analyst weighs the additional signal's informativeness.
- An overconfident analyst weighs their private information more ($\hat{w} \gg w$), while an analyst who herds weighs their information less ($\hat{w} \ll w$).

Why is the Kalman Gain underestimated?

- Table 2 shows that it is the *underestimation* of K driving the autocorrelation; $\hat{\phi}$ and $\hat{w}$ are quite close to their true values.
  - This suggests the analysts have correct beliefs about the precision of the extra signals and the variance of the shocks to the persistent component of earnings growth.
- The Kalman gain used by the analysts ($\hat{K} = 0.414$) vs. actual Kalman gain ($K = 0.953$)
- $\hat{w} - w \approx 0$, meaning analysts do not overestimate the precision of the additional signal ($\hat{\sigma}_n^2$) or the permanent growth shocks ($\hat{\sigma}_e^2$).
  - The only explanation (in this model) for this is that $\hat{\sigma}_a^2 \gg \sigma_a^2$ or an overestimation **of the noise of the reported earnings.**

Detection Error Probabilities

- In order to estimate the amount of robustness, we estimate the "distance" between the approximating and worst case model.

Definitions

Detection Error Probability

$$p(\eta) = \frac{1}{2}(Pr(mistake|A) + Pr(mistake|W))$$

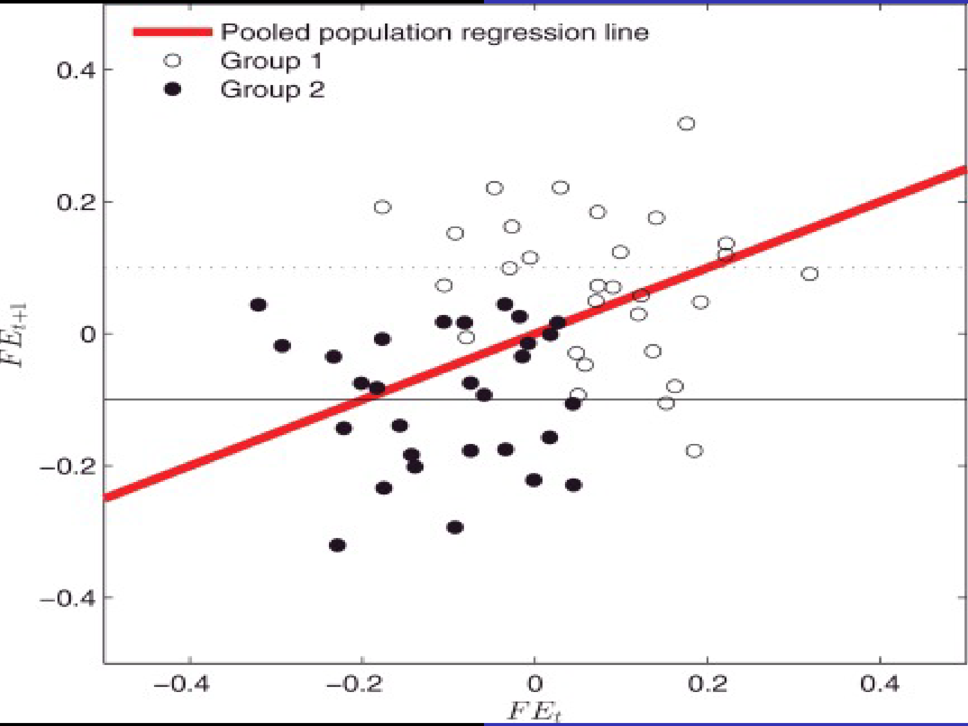
where "A" is the approximating model and "W" is the worst case model

Table 3

- Insert Table 3 here

Autocorrelation from the variation in the mean forecast error

- Intuition: A positive error is likely to be followed by a positive error. The sample is drawn from "groups" (firms, time periods or combination of both). If the mean forecasts differ between groups - then the error terms will
  - $$cor(FE_{t+1}, FE_t) = \frac{var(b_m)}{var(FE_t)} = \frac{var(\mu - \hat{\mu})}{var(FE_t)}$$
  - For examples, analysts could be accurate but issue systematically too low or too high forecasts for some firms. This will lead to an upward bias in the forecasts.



Pooled population regression line
Group 1
Group 2
0.4
0.2
0
−0.2
−0.4
FE_{t+1}
−0.4
−0.2
0
0.2
0.4
FE_t

Autocorrelation from estimation errors in the persistence of the earnings growth shocks

- This effect comes from the first term in proposition 2. Variation in the estimation of the persistent component of the earnings growth.
- $$\frac{cov(\hat{K}y_t + (1-\hat{K})\hat{y}_t, FE_t)}{var(FE_t)}(\phi - \hat{\phi})$$

Table 4
Decomposing the autocorrelation of forecast errors

This table decomposes the autocorrelation of forecast errors into three main components: (1) autocorrelation due to variation in mean forecast errors, (2) autocorrelation due to estimation errors in $\hat{\phi}$, and (3) autocorrelation due to analysts' concerns for model misspecification. These three components add up to the total autocorrelation of forecast errors estimated from a pooled regression. The first two components are further decomposed by the source of heterogeneity. Mean forecast errors, for example, vary as a function of calendar time (year), firm age, and firm, and this table reports how much the variation in each dimension contributes to the autocorrelation of forecast errors. Standard errors associated with the variation-in-mean forecast errors channel are are heteroskedasticity and autocorrelation consistent Newey and West (1987) with the number of lags selected using Newey and West (1994). Standard errors associated with the estimation errors-in-$\hat{\phi}$ channel are computed using a parametric bootstrap.

Autocorrelation estimate	Estimate	SE
Total autocorrelation of forecast errors, (1) + (2) + (3)	**0.216**	**0.008**
(1) Autocorrelation due to variation in mean forecast errors	**0.043**	**0.008**
Calendar-time variation	0.012	0.009
Age variation	0.000	0.008
Variation across firms	0.030	0.009
(2) Autocorrelation due to estimation errors in $\hat{\phi}$	**0.047**	**0.006**
Calendar-time variation	0.001	0.002
Age variation	0.000	0.003
Variation across firms	0.046	0.010
(3) Autocorrelation due to analysts' concerns for model misspecification	**0.125**	

Ideas for further research

- If autocorrelation is present in analysts' forecasts, then it should have an effect on how investors learn about analysts ability and objectives.
- Following *Chen et al.* (2005), we have a simple estimation of Bayesian learning:

$$M_{i,t} = \alpha_0 + a_1 NEWS_{i,t} + a_2 w(N_{i,t}) \cdot NEWS_{i,t} + a_3 w(N_{i,t}) \cdot ACC(N_{i,t}) \cdot NEW$$

1. $M_{i,t}$ is a measure of market impact
2. N is performance signals
3. $NEWS$ is difference between forecast and consensus
4. $ACC(N)$ is accuracy (average absolute forecast error)
5. $w(N)$ is a weight increasing in observations of N