A Generalized Earning-Based Stock Valuation Model with Learning[*]

Gady Jacoby[†]

I.H. Asper School of Business

The University of Manitoba

Alexander Paseka[‡]

I.H. Asper School of Business

The University of Manitoba

Yan Wang[±]

I.H. Asper School of Business

The University of Manitoba

[*] JEL classification: G12. Key words: Asset pricing, incomplete information, earnings growth, price-earnings ratio. Jacoby thanks the Social Sciences and Humanities Research Council of Canada for its financial support. Wang would like to acknowledge the University of Manitoba and the Asper School of Business for its financial support.

[†] Dept. of Accounting and Finance, I.H. Asper School of Business, University of Manitoba, Winnipeg, MB, Canada, R3T 5V4. Tel: (204) 474 9331, Fax: 474 7545. E-mail: jacobyg@ms.umanitoba.ca.

[‡] Dept. of Accounting and Finance, I.H. Asper School of Business, University of Manitoba, Winnipeg, MB, Canada, R3T 5V4. Tel: (204) 474 8353, Fax: 474 7545. E-mail: paseka@cc.umanitoba.ca.

[±] Corresponding author. Dept. of Accounting and Finance, I.H. Asper School of Business, University of Manitoba, Winnipeg, MB, Canada, R3T 5V4. Tel: (204) 474 6985, Fax: 474 7545. E-mail: umwangy6@cc.umanitoba.ca.

A Generalized Earning-Based Stock Valuation Model with Learning

Abstract

This paper extends a recent generalized complete information stock valuation model with incomplete information environment. In practice, mean earnings-per-share growth rate (MEGR) is random and unobservable. Therefore, asset prices should reflect how investors learn about the unobserved state variable. In our model investors learn about MEGR in continuous time. Firm characteristics, such as stronger mean reversion and lower volatility of MEGR, make learning faster and easier. As a result, the magnitude of risk premium due to uncertainty about MEGR declines over learning horizon and converges to a long-term steady level. Due to the stochastic nature of the unobserved state variable, complete learning is impossible (except for cases with perfect correlation between earnings and MEGR). As a result, the risk premium is non-zero at all times reflecting a persistent uncertainty that investors hold in an incomplete information environment.

I. Introduction

This paper extends the earnings-based stock valuation model of Bakshi and Chen (2005) (BC hereafter) by relaxing the complete information assumption and allowing for a market with incomplete information. To this end, we assume as in the BC model that earnings growth is observed by investors. However, they do not observe the instantaneous mean of earnings growth rate (thereafter, MEGR). The MEGR is an additional state variable, and we model it as a mean-reverting process. Our model allows for continuous learning about the unobserved state variable, and asset prices reflect this learning process. We investigate the effects of firm characteristics, such as mean-reversion speed and volatility of earnings growth, on differences in asset pricing between our incomplete-information and the BC complete-information models as well.

Our results indicate that the faster the earnings-growth mean reverts to its long-term value, the smaller the mispricing attributed to information incompleteness. This effect results from the fact that the higher speed of reversion towards the constant long-term mean leads to a faster exponential decay of any initial deviation from this mean and, therefore, faster learning. Ceteris paribus, the higher volatility of the unobservable MEGR results in larger mispricing. This result is more pronounced for younger firms with shorter learning horizons for which, naturally, there is a short history of data available for learning. This finding is consistent with Pastor and Veronesi (2003), who predict that M/B declines over a typical firm's lifetime, and younger firms should have higher M/B ratios than otherwise identical older firms since uncertainty about younger firms' average profitability is greater.

In our model the mean squared error of MEGR estimate, a measure of the degree of learning, persists and remains especially large for short learning horizons. The persistent uncertainty of the MEGR estimate generates an extra risk premium beyond what is accounted for in the complete information model. Over time both the uncertainty about MEGR estimate and extra risk premium decline to equilibrium levels as more information becomes available. In a perfect learning environment (e.g., unobservable MEGR is perfectly correlated with earnings), the extra risk premium on MEGR declines and converges to zero in the long run. At the same time, the variance of the estimate of MEGR decreases over learning horizon and converges to zero.[1] Perfect correlation implies that investors eventually have complete knowledge of the true process of the mean growth rate.

However, in non-perfect learning environment, the extra risk premium on MEGR never vanishes regardless of learning horizon. This long run risk premium reflects a persistent uncertainty that investors hold in an incomplete information environment.

For comparison, we compute the risk premiums based on our incomplete-information model and the complete-information model of BC. First, MEGR risk premium in incomplete information case is always bigger than that under complete information environment. They are the same only if the correlation between earnings and MEGR is perfect. Second, The difference in MEGR risk premiums declines with learning horizon faster for firms with larger correlation between earnings and underlying MEGR. Third, for 20 technology stocks used in Bakshi and Chen (2005), we find that the difference in risk premiums can be as high as 40%-50% for short learning horizons of several months. Given BC parameter values the difference declines to a steady state level after 6-11 months. Finally, the level of incomplete information premium can reach up to 7 percent for firms with short learning horizons and weaker mean reversion even if their earnings are perfectly correlated with MEGR.

The equilibrium stock prices computed based on our model have patterns similar to those of risk premiums. With perfect correlation between earnings growth and MEGR, investors perfectly learn about MEGR within ~ 11 months (based on 20 technology stock data of Bakshi and Chen, 2005). By this time there is no longer any difference in prices between BC model and our model. Further, average price differential between our model and BC model ranges from 0 percent for perfect learning case (the correlation between

[1] When the correlation between earnings and their latent MEGR is perfectly negative, this result holds as long as the speed of mean reversion is not too small relative to the volatility of MEGR. This condition is the consequence of measuring the long-term uncertainty of MEGR by the ratio of the earnings volatility to the speed of mean reversion. See Proposition 1 below.

earnings and MEGR is perfect) to -15.5% for zero-learning case (the correlation between earnings and MEGR is zero), with incomplete information price being lower on average. The lower stock price based on our incomplete-information model is corresponding to the extra risk premium on MEGR that investors demand implying that investors' uncertainty about MEGR should be compensated.

We find that the price differential between our model and that of BC, defined as pricing error, can persist for years even under perfect learning conditions. The more volatile MEGR is, the longer the persistence. We also show that fast mean-reversion speed of MEGR facilitates learning in that pricing errors are small in magnitude even after short learning process; while with low mean-reversion speed of MEGR, pricing errors are reduced substantially only after long learning process. Holding MEGR's volatility and mean-reversion speed constant, we find that there is a negative association between long-term pricing errors and degree of incompleteness of information environment as reflected by correlation between earnings and MEGR (in absolute value). For an extreme incomplete-information environment, such as one with zero correlation between earnings and MEGR, investors basically learn nothing about state variable MEGR from earnings. In this case, pricing errors are largest on average. Finally, we show that pricing errors still exist after long learning horizon (e.g., eight years) with precisely estimated MEGR as long as the information environment is incomplete. The non-vanishing pricing errors reflect residual risk premium (not present in the complete information model) due to investors' imperfect forecasts of the underlying state variable.

The remainder of the paper is organized as follows. The next section discusses related literature. Section 3 extends the complete information stock valuation model by modeling investors' inference about an unobserved state variable. Section 4 compares risk premiums and prices in the incomplete and complete information models. Section 5 concludes the paper.

2. Related Literature

Prior studies, such as Grossman and Shiller (1981), have found that the volatility of stock return is too high relative to the volatility of its underlying dividends and consumption.[2] The discrepancy between the high volatility of stock return and low volatility of dividends and consumption is viewed as the basic reason for the equity premium puzzle in recent work such as Campbell (1996) and Brennan and Xia (2001). To reconcile the discrepancy, learning about an unobservable state variable, such as the dividend growth rate, has been introduced to stock valuation (see, for example, Timmermann, 1993; Brennan, 1998; Brennan and Xia, 2001; Veronesi, 1999 and 2001, and Lewellen and Shanken, 2002).

Most of traditional stock valuation models neglect the learning process and implicitly assume that state variables for return predictability are known to investors (see, for example, Merton, 1971, and 1973; Samuelson, 1969, Breedon, 1979, and Bakshi and Chen, 2005). However there is substantial evidence indicating that market information is incomplete (see, for example, Faust, Rogers, and Wright, 2000; and Shapiro and Wicox, 1996). With an incomplete information set, investors may face an estimation risk because they are unable to observe many of state variables characterizing financial markets. This limitation is recognized by recent studies, (see, for example, Williams, 1977; Dothan and Feldman, 1986; Detemple, 1986; Gennotte, 1986; Timmerman, 1993; Brennan, 1997; and Feldman, 2007), which examine the role of learning with incomplete information in equilibrium.

For example, Timmermann (1993) provides a simple learning model, in which average dividend growth is unknown, to account for the fact that agents may not observe the true data-generating process for dividends. The model of Timmermann (1993) shows that dividend surprise affects stock price not only through current dividends but also through the effect on expected dividend growth rate, which also changes expected future dividends. The latter effect also explains why return volatility is much higher than that of dividend growth.

[2] Among others, Brennan and Xia (2001) state that the standard deviation of real annual continuously compounded stock returns in the U.S. was 17.4 % from 1871 to 1996, while the standard deviation for dividend growth was only 12.9 %, and 3.44 % for consumption growth. Pastor and Veronesi (2009) document that the postwar volatility of market returns was 17% per year while volatility of dividend growth was 5%.

Instead of using price-to-dividend ratio (P/D), Pastor and Veronesi (2003) assume that M/B is the only observed state variable but its long term mean (a constant) is not. Their learning model predicts that the uncertainty of the estimate declines to zero hyperbolically. In the end, the case is identical to complete information. In a later study, Pastor and Veronesi (2006) calibrate their 2003 model to value stocks at the peak of the Nasdaq "bubble" in March 2000. They find a positive link between uncertainty about average dividend growth and the level and variance of stock prices. Pastor and Veronesi (2006) argue that the observed Nasdaq bubble is associated with the time-varying nature of uncertainty about technology firms' future productivity, and can be explained by learning model. Pastor and Veronesi (2009) extend Timmermann (1993) and show the positive association between the volatility of stock returns and its sensitivity to the uncertainty of average dividend growth.

The calibration of Pastor and Veronesi (2003) model to annual data from the CRSP/COMPUSTAT database shows that it takes about 10 years with learning to revert to complete information case under their parameter values. Further, once their model reverts back to complete information case, eventually there is no risk premium associated with uncertainty about latent state variable (mean of dividend growth rate). This result is the artifact of the long term mean being a constant (although unknown). In contrast, MEGR in our model is an additional state variable. Complete learning is impossible (except for perfect correlation cases) and therefore risk premium is non-zero at all times. The non-vanishing risk premium in our model reflects a persistent uncertainty that investors hold in an incomplete information environment. The greater risk premium on MEGR results in lower stock price as a compensation to investors for remaining uncertainty about the state variable.

In a more sophisticated framework, Brennan and Xia (2001) provide a dynamic equilibrium model of stock prices in which representative agents learn about time-varying mean of dividend growth rate. They claim that the non-observability of expected dividend growth demands a learning process which increases the volatility of stock prices. The calibration of their model matches the observed aggregate dividend and consumption data for the U.S. capital market. Unlike us, they assume a constant risk-less interest rate in their dynamic model. Similarly, Pastor and Veronesi (2003) do not model risk free rate as random. In contrast, our model incorporates a stochastic interest rate into a pricing-kernel process to discount future risky payoff. The dynamic interest rate is consistent with a single-factor Vasicek (1977) interest-rate process which makes the model arbitrage-free as in Harrison and Kreps (1979).

Bakshi and Chen (2005) derive an earnings-based stock valuation model which is directly related to our paper. The model of Bakshi and Chen (2005) makes a more realistic assumption about the stochastic nature of risk-free interest rate. They adopt a stochastic pricing kernel process together with a mean-reverting process of earnings. Based on a sample of stocks and S&P 500 index, they show that the empirical performance of their model produces significantly lower pricing errors than existing models. [3]

In contrast to Bakshi and Chen (2005), in our model we recognize that the state variable, MEGR, is uncertain and subject to learning. In our model investors estimate MEGR based on earnings growth observations. Our incomplete-information model shows that the uncertainty about MEGR declines exponentially over time. Complete information case of Bakshi and Chen (2005) is a special case of our model with perfect correlation between MEGR and earnings growth in the limit of very long learning horizons. In addition, in our model estimates of state variable are imprecise resulting in an incremental risk premium not present in complete information models.

3. A Generalized Earnings-Based Model with Incomplete-information

[3] However, the applicability of Bakshi and Chen (2005) model is limited to stocks with zero or negative earnings. To address this issue, Dong and Hirshleifer (2004) introduce an alternative earnings adjustment parameter to the earnings process of BC model. The models of both Bakshi and Chen (2005) and Dong and Hirshleifer (2004) implicitly assume that information is complete about the mean of earnings growth rate. However, they do not recognize that the state variable, mean of earnings growth rate, is unobservable and has to be learned by observing realized earnings data.

In this section, we introduce an incomplete-information stock valuation model, in which investors estimate the latent state variable, MEGR. We retain several desirable features in the BC model.

Assumption 1: *The basic building block for pricing is earnings rather than dividends.* $D(\tau)d\tau$ *is dividend-per-share paid out over a time period* $d\tau$*, and it is assumed to be equal, on average, to a fraction of the firm's earning-per-share (EPS) with white noise that is uncorrelated with the pricing kernel,*

$$D(t)dt = \delta Y(t)dt + dw_d(t), \tag{1}$$

where $0 \le \delta \le 1$*, which is a constant dividend-payout ratio, and* $dw_d(t)$ *is the increment to a standard Wiener process that is orthogonal to everything else.* [4]

The constant dividend-payout-ratio assumption is widely used in equity literature (eg. Lee et al. 1999; and Bakshi and Chen, 2005). [5] Consistent with Bakshi and Chen (2005), the inclusion of $dw_d(t)$ allows firm's paid dividend to randomly deviate from a fixed percentage of earnings. In practice, many firms do not pay cash dividends and therefore the implementation of dividend-based valuation model is limited (e.g., Gordon model and its variants). [6] To avoid this problem, the specification in equation (1) allows us to value stocks based on firm's earnings, instead of cash dividends directly.

Assumption 2: *As in BC model, earnings growth in our model follows arithmetic Brownian motion. EPS, denoted by Y, follows an Itô process:*

---

[4] The white noise process of $dw_d(t)$ is uncorrelated with other variables, (eg., earnings growth, MEGR, risk-less interest rate, and pricing kernel), and therefore not a priced risk factor.

[5] In practice, many aspects are exogenous (eg. firm's production plan, operating revenues and expenses, target dividend-payout-ratio) to net earnings process and any deviation from the fixed exogenous structure will affect the earnings process. To simplify the valuation of cash flow, Bakshi and Chen (2005) assume that the earnings process indirectly incorporates these aspects reflecting firm's investment policy and growth opportunities.

[6] Fama and French (2001) find that, in recent years, many firms (especially technology firms) repurchase outstanding shares or reinvest in new projects with earnings, instead of paying cash dividends. As shown in the bottom panel of Figure 7 of their paper, the fraction of firms that pay no dividend rises from 27 percent in 1963 drastically to 68 percent in 2000. Similarly, while only 31 percent of firms neither pay dividends nor repurchase shares in 1971 (when repurchase data is available), the fraction grows to 52 in 2000.

$$\frac{dY(t)}{Y(t)} = G(t)dt + \sigma_y dw_y(t). \tag{2}$$

MEGR, denoted by G(t), follows an Ornstein-Uhlenbeck mean-reverting process:

$$\begin{aligned} dG(t) &= k_g(\mu_g^0 - G(t))dt + \sigma_g d\omega_g(t) \\ &= k_g(\mu_g^0 - G(t))dt + \sigma_g(\rho_{gy} dw_y(t) + \sqrt{1-\rho_{gy}^2}\, dw_0(t)), \end{aligned} \tag{3}$$

where $k_g, \mu_g^0, \sigma_y, \text{and}\, \sigma_g$ *are constants, and* $dw_y(t)$ *and* $d\omega_g(t)$ *are increments to standard Wiener processes. Shocks to G(t), the MEGR, are correlated with shocks to EPS growth with an instantaneous correlation coefficient* ρ_{gy}. *The orthogonal part of* $d\omega_g(t)$ *is denoted by* $dw_0(t)$. *The long-term mean for* $G(t)$, *under the actual probability measure, is* μ_g^0, *and the speed at which* $G(t)$ *reverts to* μ_g^0 *is governed by* k_g.

The specification in equation (2) provides a link between actual EPS growth and expected EPS growth. Both EPS growth (actual and expected), as Bakshi and Chen (2005) analyze that, could be positive or negative reflecting firm's transition stages in its growth cycle. The mean-reverting process for expected EPS growth *G(t)* in equation (3) implies that any deviations of *G(t)* from its long-term mean μ_g^0 decline exponentially over time.

Assumption 3: *The pricing kernel follows a geometric Brownian motion, which makes the model arbitrage-free as in Harrison and Kreps (1979):*

$$\frac{dM(t)}{M(t)} = -R(t)dt - \sigma_m dw_m(t),$$

where σ_m *is a constant, and* $R(t)$ *is the instantaneous riskless interest rate.*

Assumption 4: *The instantaneous riskless interest rate,* $R(t)$, *follows an Ornstein-Uhlenbeck mean-reverting process:*

$$dR(t) = k_r(\mu_r^0 - R(t))dt + \sigma_r dw_r(t),$$

where k_r, μ_r^0 *and* σ_r *are constants. This process is consistent with a single-factor Vasicek (1977) interest-rate process.*

Shocks to earnings growth, denoted by $w_y(t)$ in equation (2), is correlated with systematic shocks $w_m(t)$ and interest rate shocks $w_r(t)$ with their respective correlation coefficients, denoted by ρ_{my} and ρ_{yr}. In addition, $w_g(t)$ is correlated with $w_m(t)$ and $w_r(t)$ with correlation coefficients ρ_{mg} and ρ_{gr}, respectively. Consistent with BC, both actual and expected EPS growth shocks are priced risk factors.

Following the BC model we consider a continuous-time, infinite-horizon economy with an exogenously specified pricing kernel, $M(t)$. For a firm in this economy, its shareholders receive infinite dividend stream $\{D(t): t \geq 0\}$ as specified in equation (1). The per-share price of firm's equity, P_t, for each time $t \geq 0$, is determined by the sum of expected present value of all future dividends, as given by

$$P_t = \int_t^{\infty} E_t[\frac{M(\tau)}{M(t)} D(\tau)] d\tau , \tag{4}$$

where $E_t(\cdot)$ is the time-*t* conditional expectation operator with respect to the objective probability measure.

Following assumptions 1 to 4, the equilibrium stock price at time *t* is determined by three state variables: *Y(t)*, *G(t)*, and *R(t)*. Note that, EPS and risk-less interest rate, *Y(t)* and *R(t)*, are observable at time *t*. However, the mean EPS growth, *G(t)*, is unobservable in any point of time in practice. Bakshi and Chen (2005) use analyst estimates as unobserved *G(t)* to implement their valuation formula, in which the uncertainty about estimates is neglected, and the associated risk premium is missing in asset prices. In contrast, we recognize the fact that investors cannot observe *G(t)* and have to learn it by observing available relevant information, such as earnings. The learning process in our model affects risk premium and equilibrium prices reflecting investors' uncertainty about estimates of *G(t)*. In the next subsection, we describe the dynamic learning process for the unobserved MEGR. The time-varying nature of uncertainty about estimates is explored as well.

3.1 Learning about unobserved MEGR

In practice analysts use past observations of EPS growth to build their forecasts of MEGR into the future. To be consistent with this observation we model the best (in the mean square sense) estimate of the unobserved MEGR as an expectation conditional on previous observations on earnings growth. Due to the Markovian nature of the model a representative agent takes as given the estimate of MEGR (Genotte, 1986; and Dothan and Feldman, 1986) when pricing assets.

Theorem 1: *Following standard results from one-dimensional linear filtering (see, for example, Liptser and Shiryaev, 1977 and 1978), the processes for* $Y(t)$ *and the MEGR estimate,* $\hat{G}(t)$, *based on the information set available to the agents, are given by*

$$\frac{dY(t)}{Y(t)} = \hat{G}(t)dt + \sigma_y dw_y^*,$$

$$d\hat{G}(t) = k_g(\mu_g^0 - \hat{G}(t))dt + \Sigma_t dw_y^*, \tag{5}$$

where $\Sigma_t = \dfrac{S(t)+\sigma_{gy}}{\sigma_y}$, $\sigma_{gy} = \rho_{gy}\sigma_y\sigma_g$, and $dw_y^* = \dfrac{1}{\sigma_y}\left(\dfrac{dY(t)}{Y(t)} - \hat{G}(t)dt\right)$. $S(t)$ *is the posterior variance of the agent's estimate of G(t) given earnings information accumulated until time t, which is defined as,* $S(t) \equiv E[(G(t)-\hat{G}(t))^2 \mid Y(t)]$. *If an initial forecast error variance is* $S(0)$, *S(t) is given by,*

$$S(t) = S_2 + \frac{S_1 - S_2}{1 - Ce^{\gamma(S_1 - S_2)t}}, \qquad \text{when } S(0) \in [S_1, \infty), \tag{6}$$

where $S_1 = \frac{-\eta}{2} + \sqrt{\frac{\eta^2}{4} - \alpha}$, $S_2 = \frac{-\eta}{2} - \sqrt{\frac{\eta^2}{4} - \alpha}$, $C = \dfrac{S(0)-S_1}{S(0)-S_2}$, $\alpha = -\sigma_g^2\sigma_y^2(1-\rho_{gy}^2)$,

$\eta = 2\sigma_y^2\left(\dfrac{\sigma_{gy}}{\sigma_y^2} + k_g\right)$, *and* $\gamma = -\dfrac{1}{\sigma_y^2}$.

Proof. See Appendix A.

The term dw_y^* represents an increment of the standard Wiener process given earnings information available to investors. σ_{gy} is an instantaneous covariance between the innovations in MEGR and earnings. *S(t)* quantifies the forecast error of $\hat{G}(t)$ reflecting the degree of information incompleteness. For example, *S(t)* of zero implies perfect knowledge of the underlying state variable.

Note that $\gamma < 0$ and $S_1 > S_2$. Hence, equation (6) implies that in the long run as more information becomes available, $S(t)$ declines and eventually converges to S_1, which is always nonnegative. In addition to S_1, another bound for $S(t)$ is denoted by S_2, which is always non-positive and lower than S_1. Therefore, S_2 is irrelevant to our analysis of the long-term value of $S(t)$. Nevertheless, S_2 is one of the parameters determining the speed of convergence of $S(t)$ to S_1.

Next, we change the parameters in SDE (5) to reflect the agent's information set:

$$d\hat{G}(t) = \left(k_g + \beta + \frac{S(t)}{\sigma_y^2}\right)\left[\hat{\mu}_g^0 - \hat{G}(t)\right]dt + \left(\frac{S(t)}{\sigma_y^2} + \beta\right)\frac{dY(t)}{Y(t)}, \tag{7}$$

where $\beta = \frac{\sigma_{gy}}{\sigma_y^2}$. Note that, under this representation of the process for the MEGR estimate, the speed of mean reversion is governed by $\left(k_g + \beta + \frac{S(t)}{\sigma_y^2}\right)$ and its long-term mean is given by $\hat{\mu}_g^0 = \frac{k_g}{k_g + \beta + \frac{S(t)}{\sigma_y^2}}\mu_g^0$. Since in the long run $S(t)$ converges to S_1, we define the long-run speed of mean reversion, k_g^*, as $k_g^* = \left(k_g + \beta + \frac{S_1}{\sigma_y^2}\right)$. Substituting for S_1 and rearranging the terms we get the following expression for the long-run speed of mean reversion: $k_g^* = \sqrt{(\beta + k_g)^2 + \frac{\sigma_g^2}{\sigma_y^2}(1-\rho_{gy}^2)}$. The last expression for k_g^* is intuitive. In our model, investors learn about the true MEGR from historical changes in EPS. Specifically, investors update the latent mean growth rate based on an OLS-type relation between the "explanatory variable", $\frac{dY(t)}{Y(t)}$, and the "dependent variable", $d\hat{G}(t)$. This is very similar to the case of hedging a short position in an underlying asset with futures contracts. In both cases, the hedge ratio is the OLS slope coefficient, or β. In our model, β is the sensitivity of MEGR to the percentage change in EPS.

Note that β is an imperfect "hedge ratio" due to the less than perfect correlation in general between EPS and latent MEGR. Analogous to the case of hedging with futures, in our model this imperfect correlation translates into "basis risk" measured as $\frac{\sigma_g^2}{\sigma_y^2}(1-\rho_{gy}^2)$, and serves as an adjustment for an imperfect forecast $\hat{G}(t)$. Another adjustment for the latent MEGR comes from parameter k_g, the strength of latent mean growth rate reversion towards its long-term mean. In the following propositions we consider two special cases for the correlation, ρ_{gy}, between EPS and the mean of earnings growth rate, MEGR.

Proposition 1.a: *When the correlation, ρ_{gy}, between EPS growth and MEGR is perfectly positive, the posterior error variance of MEGR estimate, S(t), declines with time and converges to zero, which suggests that complete learning is obtained eventually in this case.*

Proof: see Appendix A.

Proposition 1.b: *When the correlation, ρ_{gy}, between EPS growth and MEGR is perfectly negative, the posterior variance of the MEGR estimate, S(t), converges to S_1. S_1 could be either positive or zero, depending on the sign of $(k_g + \beta)$, which is the long-run speed of mean reversion for the latent MEGR in this case.*

Proof: see Appendix A.

The intuition behind Proposition 1 is that a perfect and positive correlation between earnings and MEGR eventually allows investors to estimate the true mean growth rate with perfect accuracy, which implies perfect learning. When the correlation is perfect negative, the learning is perfect as long as the speed of mean reversion of the true process for the mean growth rate, k_g, is not too small relative to the absolute value of β, which measures the relative variability of MEGR and EPS growth. [7] In other words, learning is perfect in this perfect-negative-correlation case as long as the long-run speed of mean reversion for the process of MEGR, k_g^*, is positive. We can think of this situation as interplay of two effects. First, absent uncertainty, mean reversion represented by k_g, implies an exponential decay of any initial forecast error facilitating learning in this case. The second effect, representing the inverse of the signal-to-noise ratio, $\frac{\sigma_y}{\sigma_g}$, counteracts learning due to noise in the latent variable. The signal is the volatility of EPS growth, and the noise is the standard deviation of MEGR. In this case, the signal is too weak (β is large in absolute value), and complete learning is not possible in the long run despite the perfect negative correlation. The long-run result is determined by relative magnitudes of k_g and β.

To illustrate Proposition 1, we demonstrate the evolution of the learning process for MEGR estimate, $\hat{G}(t)$, in an incomplete-information environment. By using Euler approximation, we discretize the continuous processes for EPS growth rate, Y, its true mean, $G(t)$, and its mean estimate, $\hat{G}(t)$, which are given by:

$$Y(t) = Y(t-1)\left(1 + G(t-1)\Delta t + \sigma_y \sqrt{\Delta t}\varepsilon_y\right),$$

$$G(t) = G(t-1) + k_g(\mu_g^0 - G(t-1))\Delta t + \sigma_g\left(\rho_{gy}\varepsilon_y + \sqrt{1-\rho_{gy}^2}\varepsilon_0\right)\sqrt{\Delta t},$$

$$\hat{G}(t) = \hat{G}(t-1) + k_g(\mu_g^0 - \hat{G}(t-1))\Delta t + \left(\frac{S(t)}{\sigma_y^2} + \beta\right)\left(\frac{Y(t)-Y(t-1)}{Y(t-1)} - \hat{G}(t-1)\Delta t\right),$$

where Δt is discrete time interval, which is set to be 1/12 for monthly observations. Parameters ε_y and ε_0 are independent random variables following standard normal distribution.

[7] In this case, $\beta = -\frac{\sigma_g}{\sigma_y}$ for $\rho_{gy} = -1$.

The base case parameter values are chosen to closely match the corresponding values of 20 technology stocks analyzed in Bakshi and Chen (2005).[8] In particular, we assume the following annualized initial values: $Y(0)=2$; $G(0)=0.5$;[9] $\hat{G}(0)=0.2$; and $S(0)=0.5$. Further, base case parameter values are: $\delta = 4\%$, $k_g = 3$; $\mu_g^0 = 0.3$; $\sigma_y = 0.5$; $\sigma_g = 0.5$.[10] To examine a perfect learning case, we assume that EPS and its unobservable MEGR are negatively but perfectly correlated, that is $\rho_{gy} = -1$. In this case, $\beta = \frac{\sigma_{gy}}{\sigma_y^2} = -1$ and $k_g^* = (k_g + \beta) = 2$, corresponding to the case of Proposition 1.b. Based on these values, the lower bound for $S(t)$ is $S_1=0$ suggesting perfect learning in the long run.

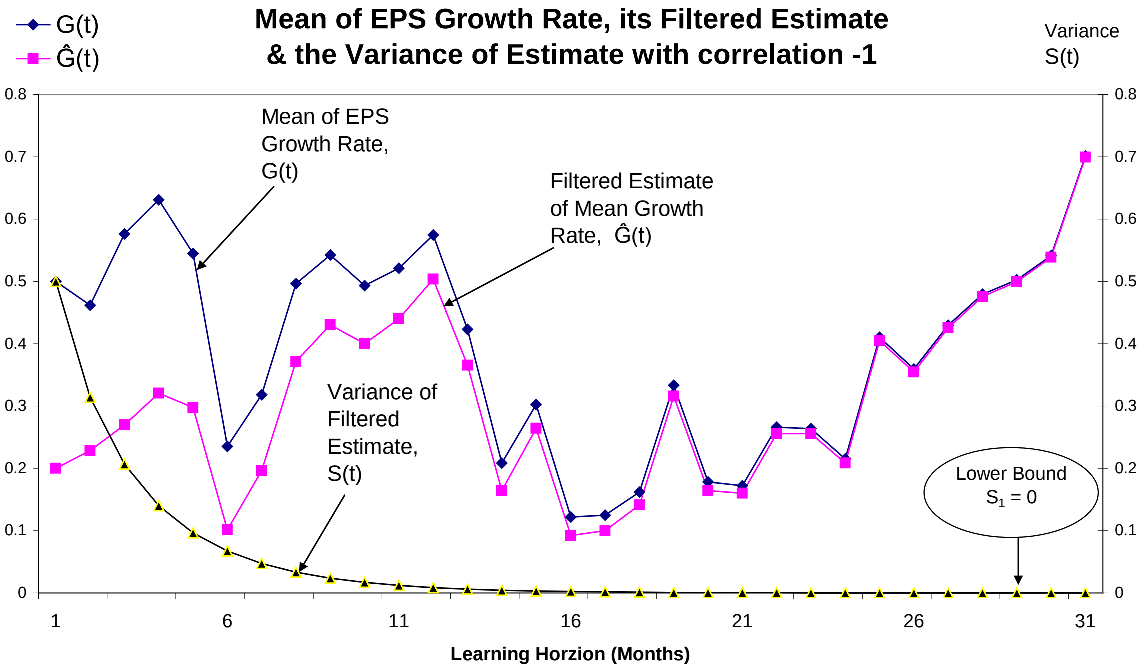
Based on the base parameter values, we plot three processes in Figure 1: the process for the true MEGR, $G(t)$, the process for the MEGR estimate, $\hat{G}(t)$, and the process for the posterior variance of the estimate, $S(t)$. As time progresses, the MEGR estimate, $\hat{G}(t)$, converges to the true MEGR, $G(t)$, as expected in the complete learning case. At the same time, the forecast error variance of the estimate, $S(t)$, converges to its lower bound of $S_1=0$. Thus, all uncertainty about the MEGR estimate is eventually eliminated by learning.

[8] The 20 technology stocks used in Bakshi and Chen (2005) includes firms under ticker ADBE, ALTR, AMAT, CMPQ, COMS, CSC, CSCO, DELL, INTC, KEAN, MOT, MSFT, NNCX, NT, ORCL, QNTM, STK, SUNW, TXN and WDC.

[9] Consistent with Table 1 of Bakshi and Chen (2005), in which the expected earnings growth (*G(t)*) is reported to be 0.4923 for 20 technology stocks.

[10] BC estimates the parameter values under the objective probability measure, which are given below for reference: $k_g = 2.688\,(0.485)$; $\mu_g^0 = 0.296\,(0.044)$; $\sigma_g = 0.425\,(0.083)$; $\rho_{yr} = -0.02\,(0.02)$; and $\delta = 4\%$. The market-implied estimate of σ_y is reported to be 0.345. The values in parentheses are cross-sectional standard errors. δ is obtained by regressing dividend yield on the earnings yield (without a constant). Average dividend divided by average net-earnings per share yields a similar δ. Note that throughout the empirical exercise, BC fixes two parameters to be that $\rho_{gy} = 1$, and $\rho_{gr} = \rho_{yr}$ to reduce estimation burden.

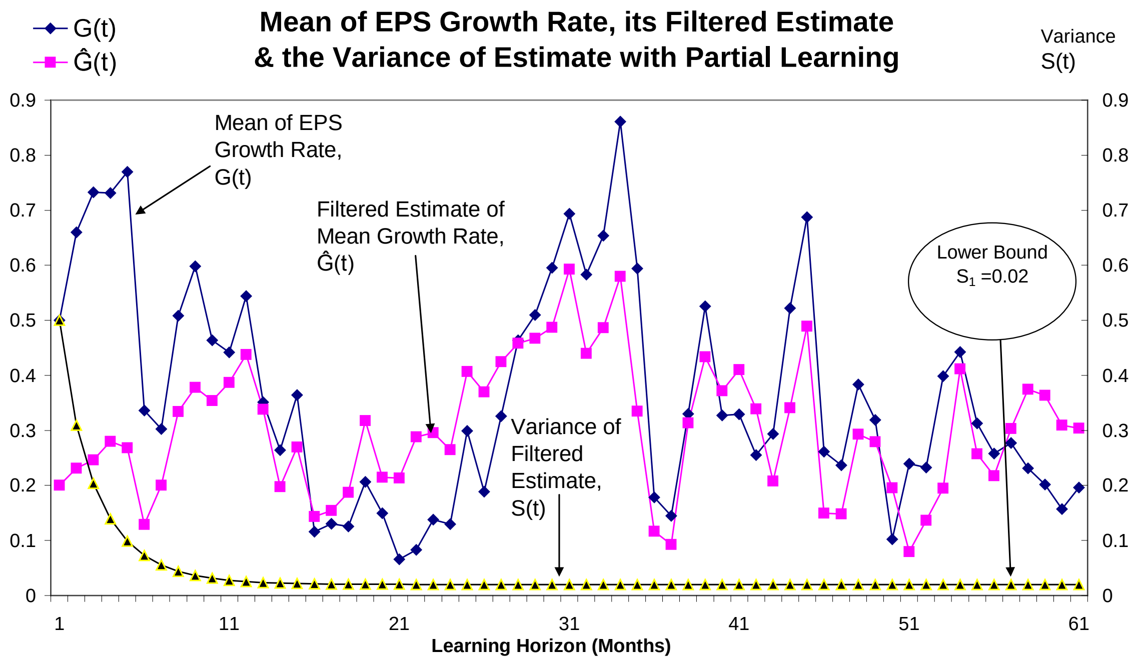
Figure 1



In this figure we plot three processes: the process for the true MEGR, $G(t)$; the process for the MEGR estimate, $\hat{G}(t)$; and the process for the posterior variance of the estimate, $S(t)$. To generate the figure we assume the following initial values: $Y(0)=2$; $G(0)=0.5$; $\hat{G}(0)=0.2$; and $S(0)=0.5$. Parameters values for the assumed stochastic processes take the following values: $k_g = 3; \mu_g^0 = 0.3; \sigma_y = 0.5; \sigma_g = 0.5;$ and $\rho_{gy} = -1.$ Based on these values, the lower bound for $S(t)$ is $S_1=0$, which suggests that complete learning is obtained eventually.

Next, we consider the case of imperfect correlation. We assume that $\rho_{gy} = -0.8$, while maintaining all other parameters at the same base case level as used in Figure 1. Figure 2 shows that although the MEGR estimate, $\hat{G}(t)$, does not converge to the true mean growth rate, $G(t)$, the difference between the two decreases with time. At the same time, the forecast error variance of the estimate, $S(t)$, converges to its positive lower bound of S_1 = 0.02008.[11] Thus, investors can only partially learn about the true mean growth rate.

Figure 2



In this figure we plot three processes: the process for the true MEGR, $G(t)$; the process for the estimated MEGR, $\hat{G}(t)$; and the process for the posterior variance of the filtered estimate, $S(t)$. To generate the figure we assume the following initial values: $Y(0)=2$; $G(0)=0.5$; $\hat{G}(0)=0.2$; and $S(0)=0.5$. Parameters values for the assumed stochastic processes take the following values: $k_g = 3; \mu_g^0 = 0.3; \sigma_y = 0.5; \sigma_g = 0.5;$ and $\rho_{gy} = -0.8$. Based on these values, the lower bound for $S(t)$: S_1= 0.02008.

The learning speed at which $S(t)$ converges to its long-run value S_1 is affected by the speed of mean reversion of MEGR, the volatilities of MEGR and EPS growth, and the correlation between them. From the solution for $S(t)$ in equation (6), the speed of its convergence, which we denote by K, is given by:

[11] Using $\rho_{gy} = -0.8$ along with the base parameter values in the formula $S_1 = \frac{-\eta}{2} + \sqrt{\frac{\eta^2}{4} - \alpha}$, where $\eta = 2\sigma_y^2(\frac{\sigma_{gy}}{\sigma_y^2} + k_g)$ and $\alpha = -\sigma_g^2\sigma_y^2(1-\rho_{gy}^2)$, we obtain that S_1 = 0.02008.

$$K = \left|\gamma(S_1 - S_2)\right| = 2k_g^*. \qquad (8)$$

Recall that $k_g^* = \sqrt{(\beta + k_g)^2 + \frac{\sigma_g^2}{\sigma_y^2}(1 - \rho_{gy}^2)}$. Note that β is a function of parameters σ_g, σ_y, and ρ_{gy}. In the following propositions, we examine the impact of these parameters on the speed of learning.

Proposition 2: *The learning speed at which the posterior forecast error variance $S(t)$ converges to its lower bound, S_1, increases in ρ_{gy}, the correlation between EPS growth and MEGR.*

Proof: see Appendix A.

The intuition behind Proposition 2 is that the information from EPS growth receives smaller weight if the correlation between EPS growth and its unobservable MEGR is smaller. In such case, learning the true MEGR from EPS data is slower.

Proposition 3: *The learning speed at which the posterior forecast error variance $S(t)$ converges to its lower bound, S_1, increases in k_g if $(k_g + \beta)$ is positive, where $\beta = \frac{\sigma_{gy}}{\sigma_y^2}$.*

Proof: see Appendix A.

Information about the true MEGR, $G(t)$, comes from two sources: (i) mean-reverting nature of the unobservable mean process; and (ii) continuous observations on change in EPS, $\frac{dY(t)}{Y(t)}$. Even in the absence of observations on earnings growth we know from equations (3) and (5) that regardless of the initial value of $\hat{G}(t = 0)$, in the long term $\hat{G}(t)$ converges to the true MEGR, $G(t)$. The speed of this convergence is governed by k_g. A higher value of k_g means that $\hat{G}(t)$ will be close to its mean more often, making it easier to learn the value of the latter. However, investors' learning by observing actual EPS growth, $\frac{dY(t)}{Y(t)}$ can increase or decrease the speed of this convergence depending on the correlation between MEGR and earnings growth. If correlation between $\frac{dY(t)}{Y(t)}$ and $G(t)$ is negative and large enough in absolute value, learning may become slower simply because the updates of $\hat{G}(t)$ become less sensitive to new information, $\left(\frac{dY(t)}{Y(t)} - \hat{G}(t)dt\right)$.

3.2 The Valuation Equation

In this section we derive share price using standard SDE arguments based on stochastic discount factor (SDF) approach (see, e.g., Chochrane, 2005). The implicit assumption here is that any shock responsible for the difference between dividend $D(t)$ and $\delta Y(t)$ is not priced:

$$E_t^*\left[d\left(MP\right)\right] + M\delta Ydt = 0,$$

where operator E_t^* represents an expectation with respect to investors' information set.

Under standard assumptions (see Dothan and Feldman, 1986; Detemple, 1986; Gennotte, 1986; and Feldman, 2007), the equilibrium price at time t is given in the following form:

$$P(t, Y, \hat{G}(t), R) = \delta YZ(t, \hat{G}(t), R), \quad \text{subject to } P(t) < \infty, \tag{9}$$

where δY represents dividends-per-share. The time-t price-dividend ratio, $Z(t, \hat{G}, R)$, is given below,

$$Z(t, s, \hat{G}, R) = \int_t^\infty \exp^{[\varphi(t,s)+\psi(t,s)\hat{G}(t)-\nu(t,s)R(t)]}\, ds, \tag{10}$$

which represents the expected present value of a continuous stream of future dividends arriving at a unit rate. The functions under the integral $Z(\hat{G}, R, t)$ have the following form (see Appendix B for details of derivation):

$$\varphi(t,s)=\frac{1-e^{-k_g(s-t)}}{k_g},\ \upsilon(t,s)=\frac{1-e^{-k_r(s-t)}}{k_r}, \text{ and}$$

$$\psi(t,s)=-\lambda_y\tau+\int_t^s\left(\frac{\Sigma_t^2}{2}\varphi^2+k_g\mu_g^*\varphi+\frac{\sigma_r^2}{2}\upsilon^2-\rho_{gr}\sigma_r\Sigma_t\varphi\upsilon-k_r\mu_r^*\upsilon\right)du, \tag{11}$$

where $\Sigma_t=\dfrac{S(t)+\sigma_{gy}}{\sigma_y}$, $\lambda_y\equiv\rho_{my}\sigma_m\sigma_y$ representing the risk premium for firm's earnings shocks, $\mu_r^*\equiv\mu_r^0-\dfrac{\rho_{mr}\sigma_m\sigma_r-\rho_{yr}\sigma_y\sigma_r}{k_r}$ and $\mu_g^*\equiv\mu_g^0-\dfrac{(\rho_{my}\sigma_m-\sigma_y)\Sigma_t}{k_g}$ are, respectively, the long-term means of $\hat{G}(t)$ and *R(t)* under the risk-neutral probability measure defined by the pricing kernel *M(t)*. We denote $\lambda_g\equiv(\rho_{my}\sigma_m-\sigma_y)\Sigma_t$ as the risk premium for $\hat{G}(t)$ in our incomplete-information model.

For the integral in equation (10) to exist, the integrand should be declining with time *s* sufficiently fast. Since functions $\varphi(t,s)$ and $\upsilon(t,s)$ in equations (11) are bounded, this requirement implies that function $\psi(t,s)$ should be negative and unbounded at large time *s*. The latter restriction implies certain constraint on model parameters, called a transversality condition as given below (see Appendix B for proof):

$$-\lambda_y+\frac{\sigma_r^2}{2k_r^2}-\mu_r^*+\mu_g^0+\frac{(2\sigma_{gy}-\eta)S_1+\sigma_{gy}^2-\alpha}{2(k_g\sigma_y)^2}+((\sigma_y-\rho_{my}\sigma_m)k_r-\rho_{gr}\sigma_r)\frac{(S_1+\sigma_{gy})}{k_rk_g\sigma_y}<0.$$

(12)

In the following proposition we show that the risk premium on MEGR based on BC full-information model is only a special case of our model.

Proposition 4: *Following BC, we define* $\lambda_g^{BC}=\rho_{mg}\sigma_g\sigma_m-\sigma_{gy}$, *as the risk premium on MEGR under BC complete-information model. The magnitude of difference in risk premium on MEGR between our incomplete-information model and BC model is given by* $\Delta\lambda_g=\lambda_g-\lambda_g^{BC}=(\rho_{my}\sigma_m-\sigma_y)\dfrac{S(t)}{\sigma_y}+\sigma_g\sigma_m(\rho_{my}\rho_{gy}-\rho_{mg})$ *at time t. The difference in risk premiums declines with learning and converges to a long-run level* *equal to* $(\rho_{my}\sigma_m - \sigma_y)\dfrac{S_1}{\sigma_y} + \sigma_g\sigma_m(\rho_{my}\rho_{gy} - \rho_{mg})$. *When EPS growth and MEGR are perfectly correlated, the long-run difference in risk premium vanishes. Similarly, the risk-neutral long-term mean of MEGR, defined as* μ_g^* *in our model, converges to that of the complete-information (BC) model.*

Proof: see Appendix B.

A higher value of posterior variance $S(t)$ results in less precise pricing. As a result, stocks with higher $S(t)$ are considered relatively risky in the market. As $S(t)$ is reduced by learning, risk premium due to information incompleteness is reduced as well. The lower bound of posterior variance, S_1, determines the minimum level of information risk premium investors demand to compensate for the uncertainty in an incomplete-information environment.

In figure 4, we demonstrate this result. We plot information-related risk premium on MEGR for firms with varying levels of correlation between EPS growth and MEGR: $\rho_{gy} = -1, \rho_{gy} = 0, \rho_{gy} = 0.5, \rho_{gy} = 1$. Holding the other parameters constant, according to proposition 4, the only two special cases in which information-related risk premium on MEGR is zero in the long run are the cases of perfect correlation, $\rho_{gy} = 1$ and $\rho_{gy} = -1$. These are the instances in which complete learning is possible. The only difference between the two cases is that the curve of information-related risk premium for $\rho_{gy} = 1$ is much steeper than that for $\rho_{gy} = -1$ reflecting a quicker learning process. Note that the case of $\rho_{gy} = 0$ has the largest long-run risk premium. In fact, zero correlation implies that learning about MEGR is most difficult because the unobservable state variable is independent of available earnings observations. As a result, investors will demand the highest information-related risk premium on MEGR in the zero-correlation case among all cases with varying correlations.

4. Comparison of the Incomplete and Complete Information Models

In this section we examine the differences between our learning-based model and the complete-information (BC) model. The purpose of this section is to investigate the properties of our estimates of latent mean growth rate, examine how different firm characteristics affect the learning process, and compare the time series of price differentials in our incomplete-information model to those in complete-information model.

To simplify discussion, we assume deterministic risk-less interest rate, i.e., $\sigma_r = k_r = 0,\ \mu_r = r_f$ for both models. To understand the major differences between the two models, we focus on the difference in risk premium on MEGR and price difference in equilibrium which are functions of the parameter vector, $\Omega = \{k_g, \sigma_g, \rho_{gy}\}$ and learning horizons. The difference in risk premium is computed following Proposition 4. The per-share price in equilibrium with incomplete-information is computed following equations (9) to (11). The stock price with complete-information is computed based on the price formula in Bakshi and Chen (2005). Lastly, the pricing error in equilibrium between two models is defined as (Price with complete-information - BC price)/BC price, in percentage format.

Two issues are explored in this section. First, we examine the time series behaviour of risk premium difference based on varying parameter values. Next, we examine the dynamic change of percentage price errors observed at different learning horizons, such as short-term (4 months), intermediate-term (10 months), and long-term horizons (25 months), respectively, for varying parameter values.

In figure 3, we plot three processes: the process for the risk premium based on true mean EPS growth rate, $G(t)$; the process for the risk premium based on the filtered mean growth rate, $\hat{G}(t)$; and the process for the variance of the filtered estimate, $S(t)$. To generate the figure we use similar base parameter values as used in figure 1 and figure 2 with minor adjustment, that is $\rho_{gy} = 1$. With perfect correlation between EPS and its

MEGR, the lower bound for $S(t)$, given by S_1, is equal to 0. While complete information risk premium is flat at 4%, the risk premium based on MEGR estimate, $\hat{G}(t)$, is substantially higher than 4% during the initial period. As posterior variance of estimate $S(t)$ reaches its minimum (in this figure, the minimum bound S_1=0), the risk premium based on MEGR estimate, $\hat{G}(t)$, drops over time and reaches 4% in the long term limit. This figure suggests that the investors demand an extra risk premium to compensate their estimation risk due to incomplete-information. As learning progresses, the extra risk premium declines over time.

Figure 4 demonstrates the impact of change in the correlation between EPS and its MEGR on $\Delta\lambda_g$, the risk premium difference between our incomplete-information model and the complete-information model (BC). Holding other parameters constant, we change the correlation coefficient to be: $\rho_{gy} = -1$, $\rho_{gy} = 0$, $\rho_{gy} = 0.5$, and $\rho_{gy} = 1$, respectively. Based on these values, we compute the lower bound for $S(t)$ as given below: when $\rho_{gy} = -1$ or $\rho_{gy} = 1$, S_1= 0; when $\rho_{gy} = 0.5$, S_1=2.63%; and when $\rho_{gy} = 0$, S_1=4.05%. Following Proposition 4, we compute the difference of risk premium on MEGR based on our incomplete-information model and BC model. Figure 4 shows that when $\rho_{gy} = -1$ or $\rho_{gy} = 1$, both $\Delta\lambda_g$ decline and eventually converge to zero in agreement with propositions 1.a, 1.b, and 4. The minor difference between the two perfect learning cases ($\rho_{gy} = 1$ and $\rho_{gy} = -1$) is in the speed at which $\Delta\lambda_g$ converges to zero. As demonstrated in Figure 4, for perfect positive correlation ($\rho_{gy} = 1$), $\Delta\lambda_g$ declines much faster and converges to zero after nine months, while for $\rho_{gy} = -1$, it takes around sixteen months for $\Delta\lambda_g$ to converge to zero. This finding implies that with the same degree of learning (ρ_{gy} equals one in absolute value), extra risk premium for positive ρ_{gy} case diminishes much faster than that for negative ρ_{gy} case as corresponding posterior variance $S(t)$ declines faster. Slower learning in the case of negative correlation reflects the conflict between the mean-reverting nature of the MEGR process and new information coming from earnings growth as described in Proposition 3. For partial learning case, we find that when $\rho_{gy} = 0.5,$ risk premium difference $\Delta\lambda_g$ declines at a medium speed which is faster than that for $\rho_{gy} = -1,$ but slower than that for $\rho_{gy} = 1,$ in support of Proposition 2.

Note that in Figure 4, for $\rho_{gy} = 0.5,$ $\Delta\lambda_g$ converges to 4.2%, which is not equal to zero any more, implying that partial learning process results in compensation for the fact that the posterior variance of estimate *S(t)* cannot be eliminated completely even for long-term learning horizons ($S_1 > 0$). For the case of $\rho_{gy} = 0,$ $\Delta\lambda_g$ converges to 6.4%, which is the highest one among all of the risk premium differences in Figure 4. Note that the highest long-term $\Delta\lambda_g$ in this case is corresponding to its posterior variance of MEGR estimate equal to S_1=4.05% for $\rho_{gy} = 0,$ which is largest among all of that in Figure 4 (S_1= 0 for both $\rho_{gy} = 1$ and -1; and S_1=2.63% for $\rho_{gy} = 0.5$). The presence of S_1 affects the risk-neutral drift of $\hat{G}(t)$ process and stock price in equilibrium reflecting the systematic nature of uncertainty about MEGR estimate. Consistent with Proposition 4, the magnitude of S_1 positively affects the long-term magnitude of extra risk premium demanded by learning process. Recall that in Proposition 4, the long-term risk premium difference is parameterized to be,

$$\Delta\lambda_g = \lambda_g - \lambda_g^{BC} = (\rho_{my}\sigma_m - \sigma_y)\frac{S_1}{\sigma_y} + \sigma_g\sigma_m(\rho_{my}\rho_{gy} - \rho_{mg}).$$

We further find that the additional risk premium on MEGR, $\Delta\lambda_g$, declines faster with learning for firms with higher k_g, which governs mean-reversion speed. This result is demonstrated in Figure 5. Holding parameters at base case levels and $\rho_{gy} = 1$, we let the mean-reversion speed take three different values: $k_g = 2,$ $k_g = 3,$ and $k_g = 4,$ respectively. For BC model, the risk premium on *G(t)* remains flat at 4% level regardless of mean-reversion speed. While for incomplete-information model, risk premium curves for each k_g start with different magnitude and declines at varying speed, but eventually converge to complete-information premium of 4% due to perfect learning.

We see that before converging to its long run level, the risk premium on $\hat{G}(t)$ is highest for the case with the smallest speed of mean-reversion ($k_g = 2$), while lowest for the case with the largest speed of mean-reversion ($k_g = 4$). This phenomenon is in line with Proposition 3. In this case with $\rho_{gy} = 1$, learning speed, *K*, is positively correlated with k_g, implying that the uncertainty *S(t)* declines faster if MEGR reverts to its long-term mean at a larger speed. At the same time, the faster decline of *S(t)* is associated with a lower risk premium at the same point in time during learning process.

In addition to examining the impact of k_g on risk premium, we examine its impact on stock price in equilibrium as well. In figure 6, we plot time series of pricing errors between our model and BC model in percentage terms with respect to, respectively, low speed, medium speed, and high speed of k_g. The mean-reversion speed of MEGR is assumed to be $k_g = 2$; $k_g = 3$; to $k_g = 4$, respectively, for each time series. We find that pricing errors are most volatile for low speed k_g, but small in magnitude and stable for high speed. This is consistent with our proposition 3, because the higher speed k_g implies larger learning speed *K*. For example, in Figure 6 when $k_g = 2$, the percentage pricing errors decline slowly until below 1% after 37 months of learning; when $k_g = 3$, the percentage pricing errors decline relatively fast until below 1% after 15 months of learning; while when $k_g = 4$, the percentage pricing errors decline faster to reach 1% only after 5 months of learning.

In Figure 7, we further examine whether the pricing errors decline faster with learning for firms with lower σ_g which implies a less noisy MEGR process. For comparison, we plot three time series of percentage pricing errors with respect to relatively low uncertainty $(\sigma_g = 0.5)$, medium uncertainty $(\sigma_g = 0.65)$, and high uncertainty ($\sigma_g = 0.8$). We find that the magnitude of pricing errors is reduced more when MEGR is less volatile during the same learning horizon (e.g. 15 months). That is, the less uncertainty about MEGR, the smaller magnitude the percentage pricing error will decline to. This result follows from our proposition 3, in which we show that the learning speed K is inversely related to the level of σ_g. Results in figures 6 and 7 reveal that parameters σ_g and k_g have opposing effects on learning.

We also find that the effect of k_g on pricing errors is stronger for a young firm. Young firm is interpreted as a firm with short history of observations on earnings implying short learning horizon. Similarly, we find that prices are much less sensitive to learning horizon when k_g is large. These results are demonstrated in Figure 8 which presents the paths of pricing errors for k_g varying from a low level of 1.8 to a high level of 5.8, for short learning horizon (t=4 months), intermediate learning horizon (t=10 months), and long learning horizon (t=25 months), respectively. For relatively low k_g ranging from 1.8 to 3.0, pricing errors are most sensitive to learning horizon. For example, on average, pricing error for short learning horizon is around -8%, which is most volatile; pricing error for medium-learning-horizon is around -5%; and pricing error for long-learning-horizon is around -2%, which is lowest in absolute value but non-zero. For medium k_g ranging from 3.0 to 4.6, pricing errors for long learning horizon converge to zero, and pricing errors for the other two learning horizons are substantially lower than those with low k_g. For high k_g ranging from 4.6 to 5.8, pricing errors for both long and medium learning horizons are zero, on average, while producing pricing errors of -1% for short learning horizon. This phenomenon observed in Figure 8 reveals that large mean-reversion speed of MEGR facilitates learning in that pricing error is small in magnitude even after short learning process; while with low mean-reversion speed of MEGR, pricing errors are reduced substantially only after long learning process.

In Figure 9, we examine the impact of precision of MEGR $(1/\sigma_g)$ on pricing errors at different observation times. We make σ_g range from 0.80 to 0.48 in the direction of improving precision of MEGR process. Similar to Figure 8, we choose three observation times (learning horizons) for comparison, which are: t = 4 Months; t = 10

Months; and t = 25 Months. We find that for all three horizons the pricing errors decrease as σ_g declines in general. With a relatively low precision of MEGR (high σ_g ranging from 0.80 to 0.66), the pricing error for the long learning horizon varies around zero but does not vanish; the average pricing error for the medium learning horizon is -4%; and the pricing error for the short learning horizon varies widely and averages at -7%. In comparison, with a relatively high precision of MEGR (low σ_g ranging from 0.64 to 0.48), the pricing errors for the long learning horizon converge to zero, those for the medium learning horizon vary around zero, and decline substantially and approach zero for the short learning horizon. The pattern in Figure 9 suggests that high precision level of MEGR makes learning easier in that it facilitates in reducing pricing errors even in the short learning horizon case. Increasing precision of the MEGR process is equivalent to increasing its mean-reversion speed, k_g .

In Figure 10, we examine the impact of parameter ρ_{gy} on the long-term level of pricing errors with incomplete information. We assume that the estimated $\hat{G}(t)$ and the true $G(t)$ are the same to examine whether pricing error still exists in an incomplete information environment (e.g., $|\rho_{gy}| \neq 1$). Parameter ρ_{gy} determines how well investors can eventually learn about the state variable, MEGR. To see price variation as a function of learning environment we let the correlation take four different values: $\rho_{gy} = 0$; $\rho_{gy} = 0.5$; $\rho_{gy} = 0.9$; $\rho_{gy} = 1$. The sample period covers eight years (96 months). We find that for perfect correlation such as $\rho_{gy} = 1$, the pricing errors are largely around -10% at the beginning of learning horizon, and then converge at zero over fourteen-month learning period. For non-perfect learning cases, the magnitude of long-term pricing errors for $\rho_{gy} = 0.9$ is 1.21%, increasing to 7.05% for $\rho_{gy} = 0.5$, and finally to 15.48% for $\rho_{gy} = 0$ (all numbers are in absolute value). These findings suggest two implications. First, there is a negative association between long-term pricing errors and degree of incompleteness of information environment as reflected by absolute value of ρ_{gy}.

Secondly, pricing errors still exist after long learning horizon (e.g., eight years) with precisely estimated $\hat{G}(t)$ as long as the information environment is incomplete.

Since long-term pricing errors never vanish in an imperfect learning environment, we examine whether faster learning affects the magnitude of long-term pricing errors. Following Proposition 3, faster learning can be achieved at higher mean-reversion speed, k_g. Figure 11 presents the relation between long-term pricing errors and mean-reversion speed k_g in an imperfect learning environment. To generate the figure, we assume that correlation $\rho_{gy} = 0.9$, and the mean-reversion speed k_g takes on the following values: $k_g = 2, k_g = 3$, and $k_g = 4$, respectively. The magnitude (absolute value) of long-term pricing error is 3.42% for $k_g = 2$, decreasing to 1.32% for $k_g = 3$, and again decreasing to 1.14% for $k_g = 4$. This result implies that larger speed of mean-reversion leads to a reduction in the magnitude of long-term pricing errors, holding the other parameters constant. As before, the long run pricing errors are not zero. Similar to the intuition suggested by Figure 10, the non-vanishing pricing errors reflect residual risk premium (not present in the complete information model) due to investors' imperfect forecasts of the underlying state variable.

5. **Conclusions**

This paper develops a dynamic framework for valuing stocks which allows for learning about a stochastic but unobservable MEGR (mean of earnings growth rate) in an incomplete-information environment. The instantaneous MEGR is a state variable in our model, and investors can learn about it from continuously released earnings information.

We have shown in this paper that the posterior variance of MEGR estimate generates extra risk premium on MEGR beyond what is accounted for in the complete information model. We further show that the time-varying nature of posterior variance of MEGR leads to a dynamic change in risk premium and more volatile stock prices. As learning reduces the posterior variance of estimate, extra risk premium declines to an equilibrium level over time. We parameterize the risk premium on MEGR and find that the magnitude of risk premium is not only affected by posterior error variance of estimate but also affected by firm characteristics, such as volatility of earnings, volatility of MEGR, mean-reversion speed of earnings, and correlation between earnings and latent MEGR.

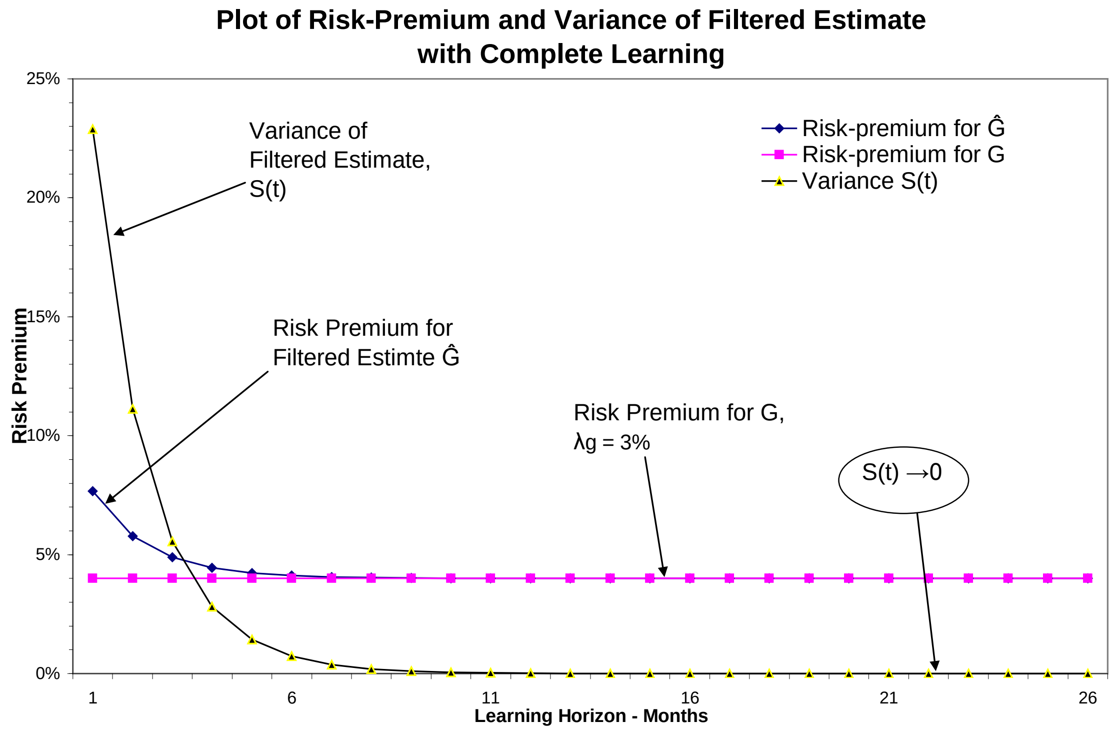
Our results indicate that the faster the MEGR reverts to its long-term value, the smaller the magnitude of risk premium attributed to information incompleteness. This effect results from the fact that the higher speed of reversion towards the constant long-term mean leads to a faster exponential decay of any initial deviation from this mean and, therefore, faster learning. With a lower mean-reversion speed, risk premium on MEGR and posterior variance of MEGR estimate decline slowly but essentially constant over time if learning horizon is long enough. We also find that the effect of mean-reversion speed on pricing errors is stronger for a young firm with short history of information. By increasing the speed of mean-reversion, pricing errors due to information-incompleteness can be reduced substantially and quickly even learning horizon is short.

Lower volatility on MEGR has similar effect of higher effective speed of mean-reversion process of latent variable on learning. Both facilitate faster learning process about the true unobservable state variable, which is shown by the fast reduction in the posterior variance of MEGR estimate.

We have also shown that higher correlation between earnings and latent MEGR leads to more complete learning about the true unobservable variable. With a perfect correlation (1 or -1), complete learning is achievable which leads to the same magnitude of risk premium and equilibrium prices in the long run as those in complete-information environment. In such case, the extra risk premium due to information-incompleteness vanishes eventually. In contrast, with an imperfect correlation (between -1 and 1), complete learning is impossible and therefore extra risk premium is non-zero at all times. The non-vanishing risk premium in our model reflects a persistent uncertainty that investors hold in an incomplete information environment. The additional long-term risk premium on MEGR results in lower equilibrium price as a compensation to investors for remaining uncertainty about the state variable.

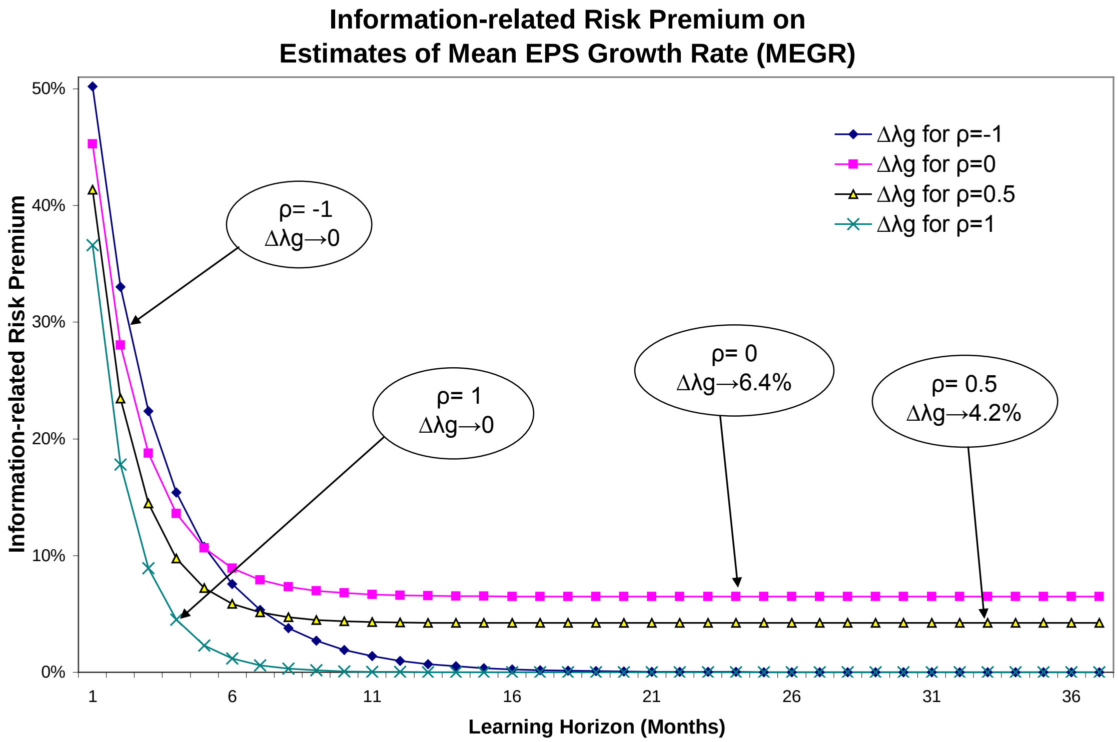
Our finding is consistent with that learning can generate higher equity premium when investors are ambiguity averse (e.g., Cagetti et al. 2002; Leippold et al. 2008; and Epstein and Schneider 2008). As Pastor and Veronesi (2009) predict that when investors are cautious of model misspecification in incomplete-information environment, model uncertainty is penalized and risk premium rises as compensation.

Figure 3



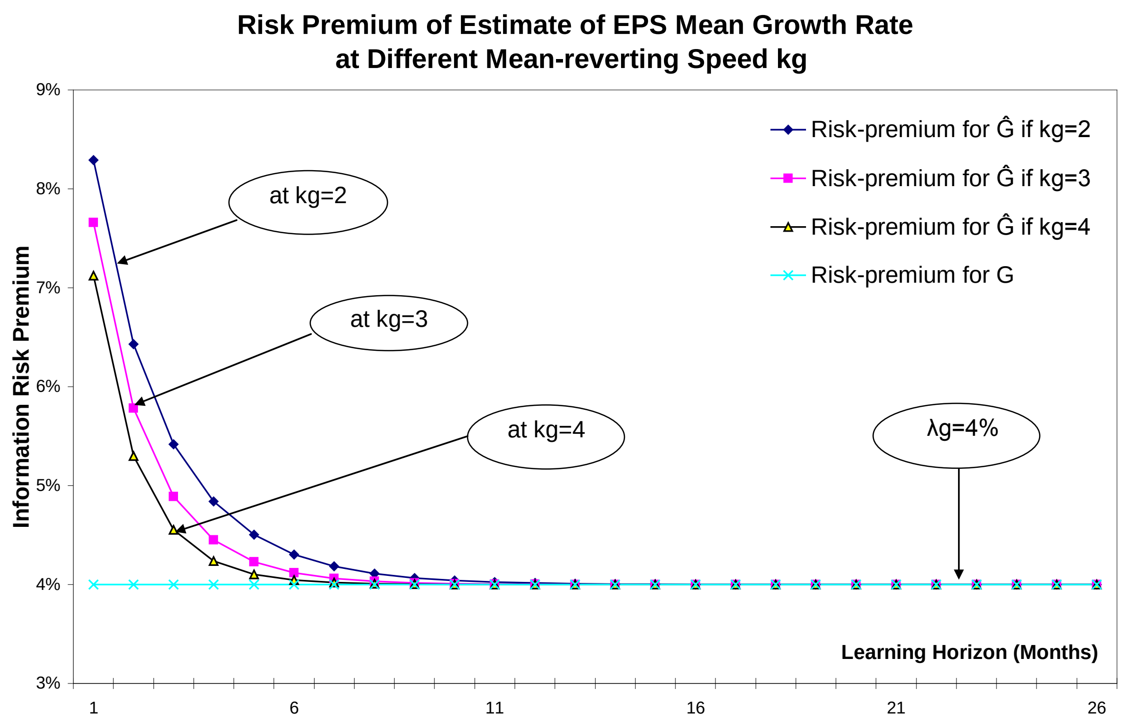
In this figure we plot three processes: the process for the risk premium based on true MEGR, $G(t)$; the process for the risk premium based on the estimated MEGR, $\hat{G}(t)$; and the process for the variance of the filtered estimate, $S(t)$. To generate the figure we assume the following initial values: $Y(0)=2$; $G(0)=0.5$; $\hat{G}(0)=0.2$; and $S(0)=0.5$. Parameters values for the assumed stochastic processes are given by: $k_g = 3$; $\mu_g^0 = 0.3$; $\sigma_y = 0.5$; $\sigma_g = 0.5$; $\sigma_m = 0.8$; $\rho_{gy} = 1$; $\rho_{my} = 0.1$; and $\rho_{mg} = 0.1$. Based on these values, we get the following lower bound for $S(t)$: $S_1 = 0$.

Figure 4



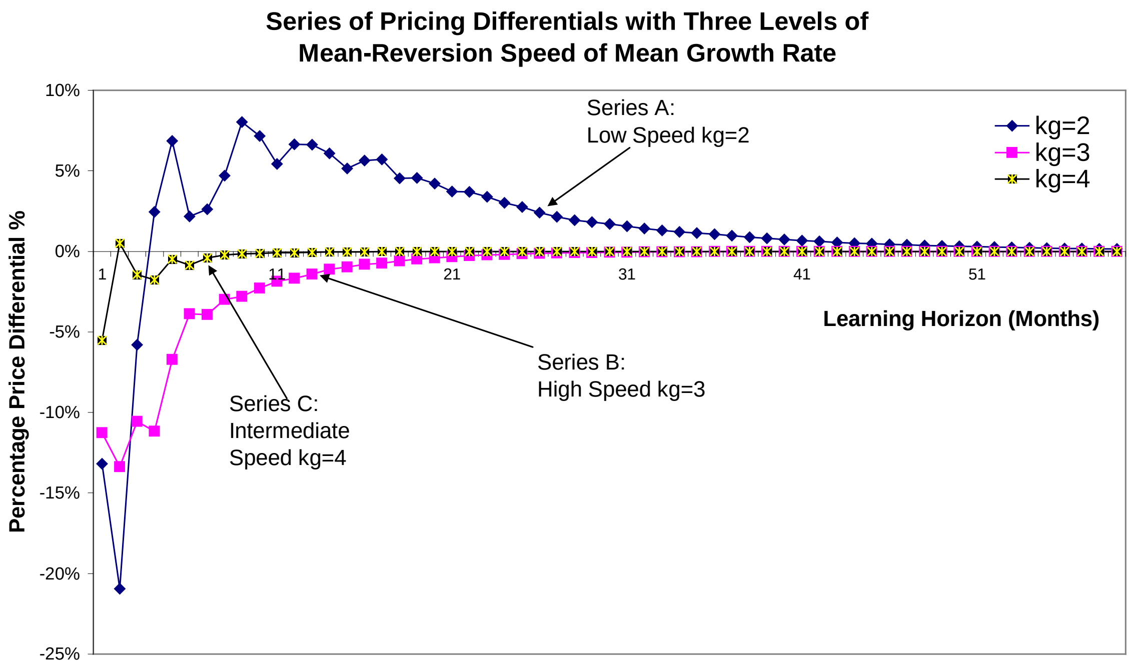
This figure demonstrates the curves of noise-related risk premium for four firms with different degree of correlation between EPS and its MEGR, holding the other parameters constant. The correlation is assumed to be: $\rho_{gy} = -1$, $\rho_{gy} = 0$, $\rho_{gy} = 0.5$, and $\rho_{gy} = 1$, respectively. To generate the figure we assume the following initial values for each firm: $Y(0)$=2; $G(0)$=0.5; $\hat{G}(0)$=0.2; and $S(0)$=0.5. Parameters values for the assumed stochastic processes take the following values: $k_g = 3$; $\mu_g^0 = 0.3$; $\sigma_y = 0.5$; $\sigma_g = 0.5$; $\sigma_m = 0.8$; $\rho_{my} = 1$; $\rho_{mg} = \rho_{my}\rho_{gy}$; r = 3%; and $\delta = 4\%$. Based on these values, we obtain the following lower bounds for $S(t)$: when $\rho_{gy} = 1$ or -1, S_1= 0; when $\rho_{gy} = 0.5$, S_1=2.63.%; and when $\rho_{gy} = 0$, S_1=4.05%. Let $\Delta\lambda_g$ denote the information-related risk premium on MEGR, we obtain the convergence level of noise-related risk premium for each firm: $\Delta\lambda_g(\rho_{gy} = -1 \text{ or } \rho_{gy} = 1) \to 0$; $\Delta\lambda_g(\rho_{gy} = 0.5) \to 4.2\%$; and $\Delta\lambda_g(\rho_{gy} = 0) \to 6.4\%$, respectively.

Figure 5



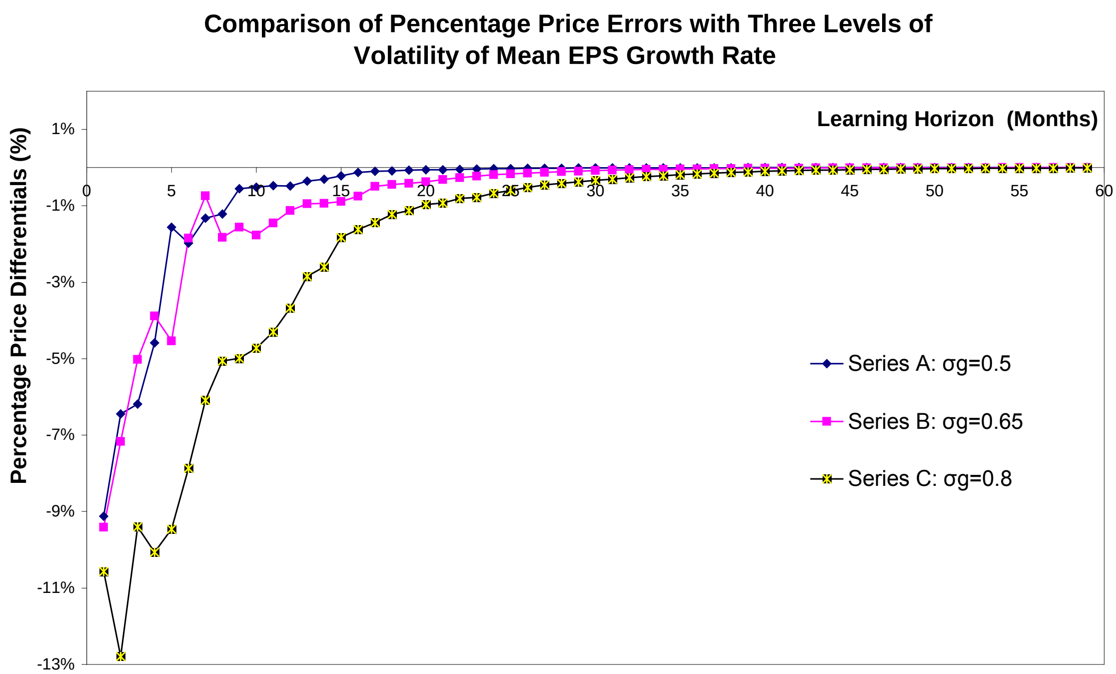
In this figure, we examine the impact of change in mean-reversion speed (k_g) of MEGR on the risk premium under our incomplete-information model and the complete-information model. To generate the figure we assume the following initial values: $Y(0)=2$; $G(0)=0.5$; $\hat{G}(0)=0.2$; and $S(0)=0.5$. Parameters values for the assumed stochastic processes take the following values: $\mu_g^0 = 0.3$; $\sigma_y = 0.5$; $\sigma_g = 0.5$; $\sigma_m = 0.8$; $\rho_{gy} = 1$; $\rho_{my} = 0.1$; and $\rho_{my} = 0.1$. The speed of mean-reversion of MEGR is assumed to be, $k_g = 2$, $k_g = 3$, and $k_g = 4$, respectively. Based on these values, we obtain the following lower bound for $S(t)$: $S_1 = 0$. The constant risk premium on $G(t)$ under complete-information model is 4 per cent.

Figure 6



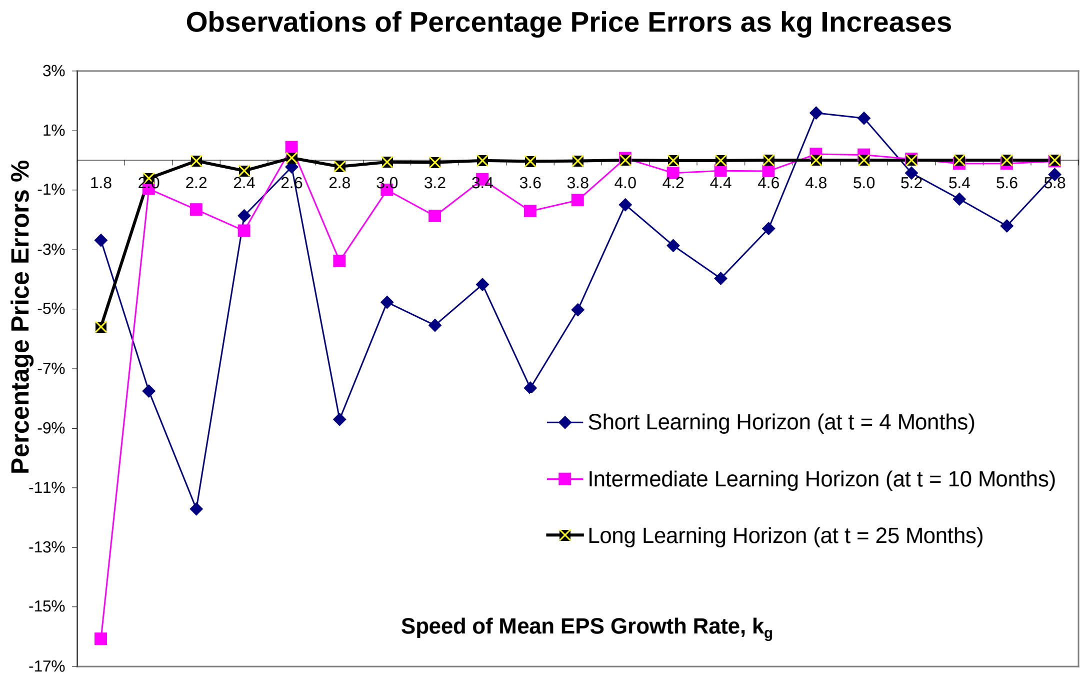
In this figure, we examine the impact of change in mean-reversion speed (k_g) of MEGR on the pricing performance based on our incomplete-information model and the complete-information model. At each time during the sample period for each level of speed, k_g, prices are computed by the learning model respectively. Percentage pricing error is defined as the ratio of (Incomplete-Information model price – BC Complete-Information model price)/ BC Complete-Information model price**.** This chart show the time series of pricing errors for each level of speed, k_g. To generate the figure we assume the following initial values: Y(0)=2; G(0)=0.5; $\hat{G}$(0)=0.2; and S(0)=0.5. The mean-reversion speed of MEGR for each series is assumed to be, $k_g = 2$, $k_g = 3$, and $k_g = 4$, respectively. The other parameters for the assumed stochastic processes take the following values: $\mu_g^0 = 0.3$; $\sigma_y = 0.5$; $\sigma_g = 0.5$; $\sigma_m = 0.8$; $\rho_{gy} = -1$; $\rho_{my} = 1$; $\rho_{mg} = -1$; r = 3%; and $\delta = 4\%$.

Figure 7



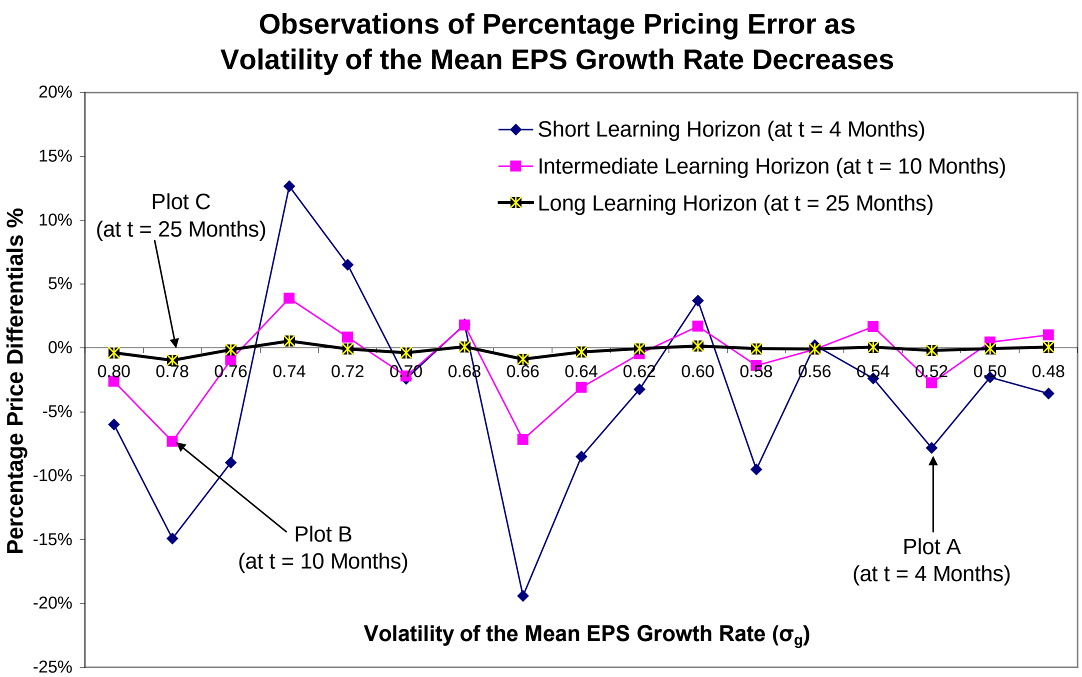
In this figure, we examine how the percentage pricing errors are influenced by the precision level of the MEGR. To generate the figure we assume the following initial values: $Y(0)$=2; $G(0)$=0.5; $\hat{G}(0)$=0.2; and $S(0)$=0.5. Parameters for the assumed stochastic processes take the following values: $k_g = 3;\ \mu_g^0 = 0.3;\ \sigma_y = 0.5;\ \sigma_g = 0.5;\ \sigma_m = 0.8; \rho_{gy} = -1;\ \rho_{my} = 1; \rho_{mg} = -1;$ r = 3%; and $\delta = 4\%$. We assume that the volatility of the mean EPS growth rate for each series are: $\sigma_y = 0.5, \sigma_y = 0.65$, and $\sigma_y = 0.8$, respectively. With each level of volatility, σ_y, percentage pricing errors are computed for each month during the sample period respectively. Percentage pricing error is defined as the ratio of (Incomplete-Information model price – BC Complete-Information model price)/ BC Complete-Information model price**.**

Figure 8



In this figure, we observe the pricing errors (percentage) by increasing the speed of MEGR. We focus on the pricing errors at three observation times (learning horizon), which are: t = 4 Months; t = 10 Months; and t = 25 Months, respectively. Percentage pricing error is defined as the ratio of (Incomplete-Information model price – BC Complete-Information model price)/ BC Complete-Information model price**.** To generate the figure we assume the following initial values: $Y(0)=2$; $G(0)=0.5$; $\hat{G}(0)=0.2$; and $S(0)=0.5$. Parameters for the assumed stochastic processes take the following values: $\mu_g^0 = 0.3$; $\sigma_y = 0.5$; $\sigma_g = 0.5$; $\sigma_m = 0.8$; $\rho_{gy} = -1$; $\rho_{my} = 1$; $\rho_{mg} = -1$; $r = 3\%$; and $\delta = 4\%$. The value of speed k_g increases from 1.8 to 5.8 gradually.

Figure 9

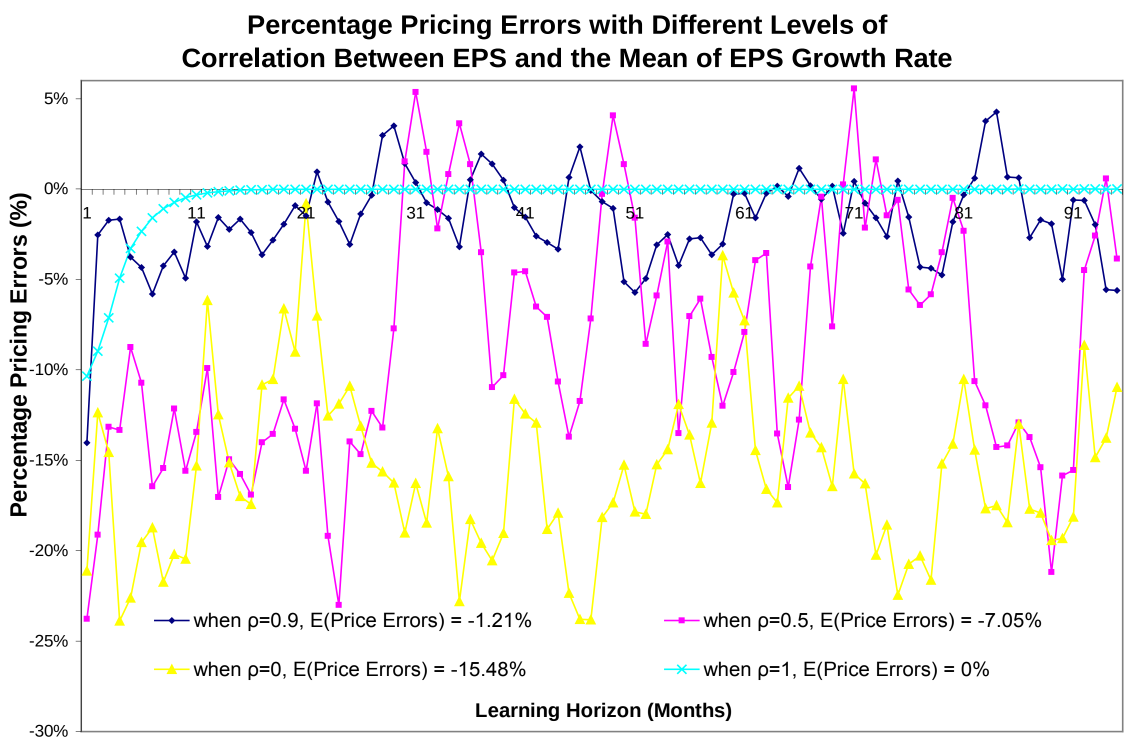


In this figure, we observe the percentage pricing errors by decreasing the volatility of MEGR. We focus on the pricing errors at three observation times (learning horizon), which are: t = 4 Months; t = 10 Months; and t = 25 Months, respectively. Percentage pricing error is defined as the ratio of (Incomplete-Information model price – BC Complete-Information model price)/ BC Complete-Information model price**.** To generate the figure we assume the following initial values: $Y(0)=2$; $G(0)=0.5$; $\hat{G}(0)=0.2$; and $S(0)=0.5$. Parameters for the assumed stochastic processes take the following values: $k_g = 3;\ \mu_g^0 = 0.3;\ \sigma_y = 0.5;\ \sigma_m = 0.8;\ \rho_{gy} = -1;\ \rho_{my} = 1;\ \rho_{mg} = -1;\ \mathrm{r} = 3\%;$ and $\delta = 4\%$. The volatility of MEGR decreases from 0.80 to 0.48 gradually.

.

Figure 10

In this figure, we examine that in an incomplete-information environment, how pricing errors are influenced by a parameter, ρ, the correlation between EPS and its MEGR. The value of parameter ρ determines the degree to which learning on the true MEGR can be achieved by using available data on EPS. We assume the correlation parameter ρ_{gy} to take the following four different levels: $\rho_{gy} = 0; \rho_{gy} = 0.5; \rho_{gy} = 0.9; \rho_{gy} = 1,$ respectively. The sample period covers eight years (96 months). Percentage pricing error is defined as the ratio of (Incomplete-Information model price – BC Complete-Information model price)/ BC Complete-Information model price**.** To generate the figure we assume the following initial values: $Y(0)$=2; $G(0)$=0.5; $\hat{G}(0)$=0.2; and $S(0)$=0.5. Parameters for the assumed stochastic processes take the following values: $k_g = 3; \mu_g^0 = 0.3;\ \sigma_y = 0.5; \sigma_g = 0.5;\ \sigma_m = 0.8;\ \rho_{my} = 1; \rho_{mg} = \rho_{my}\rho_{gy};$ r = 3%; and $\delta = 4\%$.



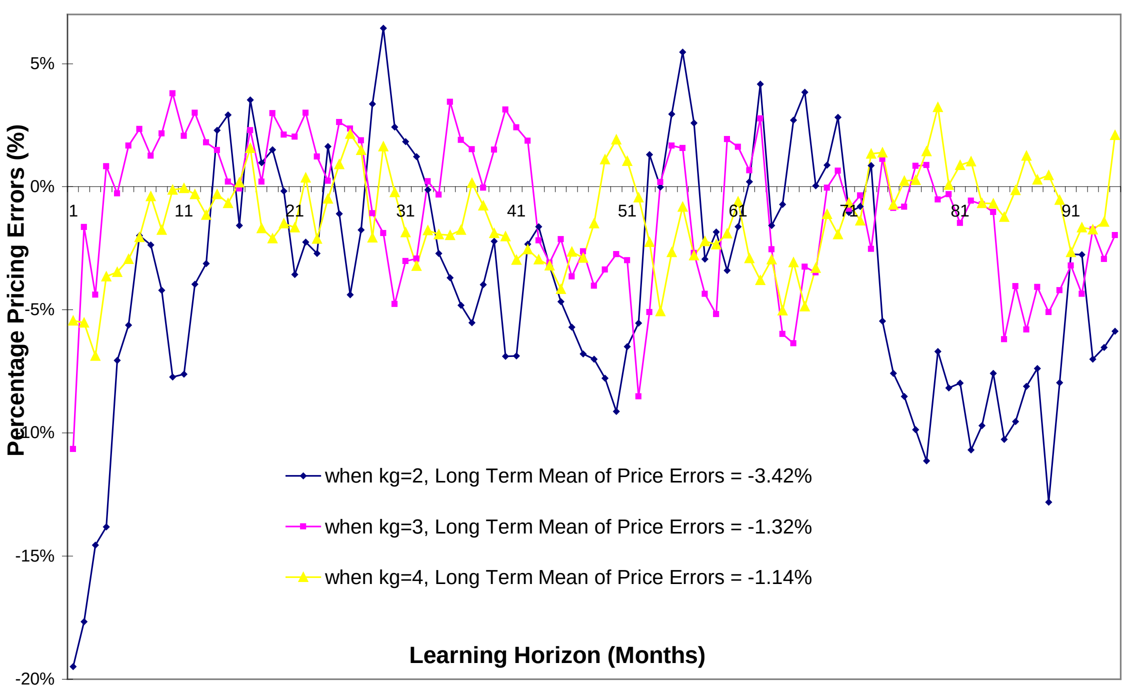
Appendix: Sample of Data for Figure 10

Long Term Mean of Pricing Errors		Percentage Pricing Errors over Learning Horizon (Months)									
		t=3	t=6	t=9	t=12	t=15	t=18	t=21	t=24	t=27	t=30
ρ_{gy}=1	**0%**	-0.071	-0.023	-0.007	-0.002	-0.001	0	0	0	0	0
ρ_{gy}=0.9	**-1.21%**	-0.017	-0.043	-0.035	-0.032	-0.017	-0.028	-0.015	-0.018	-0.003	0.014
ρ_{gy}=0.5	**-7.05%**	-0.131	-0.107	-0.121	-0.099	-0.158	-0.135	-0.156	-0.230	-0.123	0.015
ρ_{gy}=0	**-15.48%**	-0.145	-0.195	-0.202	-0.062	-0.170	-0.105	-0.008	-0.119	-0.151	-0.19

Figure 11

In this figure, we examine that in an imperfect learning environment (eg., $\rho_{gy}=0.9$), how long-term steady level of pricing errors are affected by boosting the speed of MEGR, k_g, to a higher level. Due to imperfect learning, pricing errors will decrease over learning horizon but never converge to zero. But the long-term mean of pricing errors would sustain at a relatively lower level (at absolute value) with a higher speed k_g. Percentage pricing error is defined as the ratio of (Incomplete-Information model price – BC Complete-Information model price)/ BC Complete-Information model price**.** To generate the figure we assume the following initial values: $Y(0)=2$; $G(0)=0.5$; $\hat{G}(0)=0.2$; and $S(0)=0.5$. The other parameters for the assumed stochastic processes take the following values: $\mu_g^0=0.3$; $\sigma_y=0.5$; $\sigma_g=0.5$; $\sigma_m=0.8$; $\rho_{my}=1$; $\rho_{mg}=0.9$; r = 3%; and $\delta=4\%$. The value of mean-reversion speed k_g is assumed to be: $k_g=2$, $k_g=3$, and $k_g=4$, respectively. The sample period covers eight years (96 months).

Percentage Pricing Errors with Different Levels of Speed Kg



Appendix: Sample of Data for Figure 11

Long Term Mean of Pricing Errors		Percentage Pricing Errors over Learning Horizon (Months) t=1	t=11	t=21	t=31	t=41	t=51	t=61	t=71	t=81	t=91
$k_g=2$	**-3.42%**	-0.195	-0.076	-0.036	0.018	-0.069	-0.065	-0.016	-0.010	-0.080	-0.027
$k_g=3$	**-1.32%**	-0.107	0.021	0.020	-0.030	0.024	-0.030	0.016	-0.009	-0.015	-0.032
$k_g=4$	**-1.14%**	-0.054	-0.001	-0.017	-0.018	-0.030	0.010	-0.006	-0.007	0.009	-0.027

Appendix A

Proof of Theorem 1:

EPS, denoted by Y, follows an Ito processes:

$$\frac{dY(t)}{Y(t)} = G(t)dt + \sigma_y dw_y \ . \tag{A1}$$

The mean of EPS follows an Ornstein-Uhlenbeck mean-reverting process:

$$dG(t) = k_g(\mu_g^0 - G(t))dt + \sigma_g d\omega_g \, . \tag{A2}$$

According to standard results from one-dimensional linear filtering (see, for example, Liptser and Shiryaev, 1977 and 1978), the solution for the filtered estimate of mean growth rate $\hat{G}(t)$, specialized in equations (A1) and (A2), is given by the following stochastic differential equation (SDE):

$$d\hat{G}(t) = k_g(\mu_g^0 - \hat{G}(t))dt + E_t[G(t)\frac{(G(t)-\hat{G}(t))Y(t)}{\sigma_y^2 Y(t)^2} + \rho_{gy}\frac{\sigma_g}{\sigma_y Y(t)}](dY(t) - \hat{G}(t)Y(t)dt),$$

that is,

$$d\hat{G}(t) = k_g(\mu_g^0 - \hat{G}(t))dt + \Sigma_t dw_y^*, \tag{A3}$$

where $\Sigma_t = \dfrac{S(t)+\sigma_{gy}}{\sigma_y}$, $\sigma_{gy} = \rho_{gy}\sigma_y\sigma_g$, and $dw_y^* = \dfrac{1}{\sigma_y}\left(\dfrac{dY(t)}{Y(t)} - \hat{G}(t)dt\right)$. The posterior variance of the agent's estimate of $G(t)$, $S(t) \equiv E[(G(t)-\hat{G}(t))^2 \mid Y(t)]$, satisfies the following Riccati ordinary differential equation (ODE):

$$\frac{dS(t)}{dt} = \gamma(S(t)^2 + \eta S(t) + \alpha), \tag{A4}$$

where $\alpha = -\sigma_g^2\sigma_y^2(1-\rho_{gy}^2)$, $\eta = 2\sigma_y^2(\dfrac{\rho_{gy}\sigma_g}{\sigma_y} + k_g)$,, and $\gamma = -\dfrac{1}{\sigma_y^2}$.

Equation (A4) is equivalent to the following,

$$\frac{dS(t)}{(S(t)-S_1)(S(t)-S_2)} = \gamma dt \, . \tag{A5}$$

Re-arranging equation (A5), we obtain the following ODE:

$$\frac{1}{(S_1 - S_2)}\left(\frac{dS(t)}{S(t)-S_1} - \frac{dS(t)}{S(t)-S_2}\right) = \gamma dt \, . \tag{A6}$$

Taking integral with respect to time t on both sides of equation (A6), we get,

$$\ln\left(\frac{S(t)-S_1}{S(t)-S_2}\right)=\gamma(S_1-S_2)t+c\,, \tag{A7}$$

where *c* denotes a constant. We can think of *S(t)* as a variance of forecast error based on all relevant information up to time *t*. If an initial forecast error variance is $S(0)$, then solving equation (A7), we obtain:

$$S(t)=S_2+\frac{S_1-S_2}{1-Ce^{\gamma(S_1-S_2)t}}\,, \qquad \text{when } S(0)\in[S_1,\infty), \tag{A8}$$

where $S_1=\frac{-\eta}{2}+\sqrt{\frac{\eta^2}{4}-\alpha}$, $S_2=\frac{-\eta}{2}-\sqrt{\frac{\eta^2}{4}-\alpha}$, $C=\dfrac{S(0)-S_1}{S(0)-S_2}$.

Note that $\gamma<0$ and $S_1>S_2$. Hence, equation (A8) implies that in the long run as more information becomes available $S(t)$ converges to S_1, which is always nonnegative. Another bound for $S(t)$ is denoted by S_2, which is always non-positive and lower than S_1. Therefore, S_2 is not relevant to our analysis of the long-term value of $S(t)$. Nevertheless, S_2 is one of the parameters determining the speed of convergence of $S(t)$ to S_1.

Proof of Proposition 1.a:

When $\rho_{gy}=1$ we have $\beta=\dfrac{\rho_{gy}\sigma_y\sigma_g}{\sigma_y^2}=\dfrac{\sigma_g}{\sigma_y}>0$ and $\eta=2\sigma_y^2(\beta+k_g)>0$. Therefore, ODE (4) becomes

$$\frac{dS(t)}{S(t)[S(t)+\eta]}=\gamma dt\,,$$

Thus the two bounds for $S(t)$ are $S_1=0$ and $S_2=-\eta<0$. Further, the solution for $S(t)$ is given by $S(t)=\dfrac{\eta Ce^{\gamma\eta t}}{1-Ce^{\gamma\eta t}}$. In the limit, S(t) converges to $S_1=0$ as $t\to\infty$.

Q.E.D.

Proof of Proposition 1.b:

When $\rho_{gy} = -1$, then $\beta = \frac{\rho_{gy}\sigma_y\sigma_g}{\sigma_y^2} = -\frac{\sigma_g}{\sigma_y} < 0$, $\alpha = 0$, $k_g^* = (\beta + k_g)$, and $\eta = 2\sigma_y^2(\beta + k_g)$. We consider three cases:

(i) when $k_g > |\beta|$, the effective speed of mean reversion for the process of the latent mean growth rate, k_g^* is positive and we have: $\eta > 0$. Similar to Proposition 1.a, the two bounds for $S(t)$ become $S_1 = 0$ and $S_2 = -\eta < 0$. Therefore, the solution for $S(t)$ is $S(t) = \frac{\eta C e^{\gamma\eta t}}{1 - Ce^{\gamma\eta t}} \to 0$ as $t \to \infty$.

(ii) when $k_g < |\beta|$, then $\eta < 0$, $S_1 = -\eta > 0$, $S_2 = 0$, and $k_g^* < 0$. If the initial value of $S(t)$ satisfies $S(0) > S_1 = |\eta|$, then $S(t)$ is given by $S(t) = \frac{-\eta C e^{\gamma\eta t}}{1 - Ce^{\gamma\eta t}}$, where $C = \frac{S(0) - |\eta|}{S(0)}$. In this case, $S(t)$ will decrease over time and will converge to $|\eta|$.

If $S(0) < S_1 = |\eta|$, then $S(t) = \frac{-\eta}{1 + Ce^{\gamma|\eta|t}}$, where $C = \frac{|\eta|}{S(0)} - 1$. Therefore, $S(t)$ will increase over time and will converge to $|\eta|$ from below.

Finally, if $S(0) = |\eta|$, then $S(t)$ will remain at $|\eta|$ for every *t*.

(iii) when $k_g = |\beta|$, we have: $\eta = 0$, $S_1 = 0$, $S_2 = 0$, and $k_g^* = 0$. This means that ODE (4) becomes:

$$\frac{dS(t)}{S(t)^2} = \gamma dt,$$

and the solution to this ODE is:

$$S(t) = \frac{S(0)}{1 + S(0)|\gamma|t}.$$

Therefore, $S(t)$ will approach zero hyperbolically as $t \to \infty$, and thus slower than in case (i), in which learning is exponential.

Q.E.D.

Proof of Proposition 2:

We differentiate *K* with respect to ρ_{gy}. Note that k_g^* is positive, therefore:

$$\frac{\partial K}{\partial \rho_{gy}} = \frac{\partial \left| \gamma (S_1 - S_2) \right|}{\partial \rho_{gy}} = \frac{\partial 2k_g^*}{\partial \rho_{gy}} > 0 .$$

Specifically, since $\frac{\partial K}{\partial \rho_{gy}} = \frac{2k_g \sigma_g}{\sigma_y} [(\beta + k_g)^2 + \frac{\sigma_g^2}{\sigma_y^2}(1 - \rho_{gy}^2)]^{-1/2} = \frac{2k_g \sigma_g}{\sigma_y k_g^*}$ is positive, therefore the speed of convergence is positively related to ρ_{gy}.

Q.E.D.

Proof of Proposition 3:

We differentiate *K* with respect to k_g and get:

$$\frac{\partial K}{\partial k_g} = \frac{\partial (2k_g^*)}{\partial k_g} = \frac{2}{k_g^*}(k_g + \beta) .$$

Since k_g^* is positive we have: $\frac{\partial K}{\partial k_g} > 0,$ if $(k_g + \beta) > 0$. Otherwise, if $(k_g + \beta) \le 0$, then $\frac{\partial K}{\partial k_g} \le 0.$

Q.E.D.

Appendix B

Derivation of the Asset Price:

Our model of learning unobserved state variables is consistent with evidence that analysts use past observations of EPS growth to build their forecasts. Due to the Markovian nature of the model the valuation procedure by a representative agent takes as given the filtered estimate of the mean EPS growth (Genotte, Dothan and Feldman) when pricing assets.

Given the information set available to the agents, the processes for $Y(t)$ and the MEGR estimate are given in Theorem 1,

$$\frac{dY(t)}{Y(t)} = \hat{G}(t)dt + \sigma_y dw_y^*$$

$$d\hat{G}(t) = k_g(\mu_g^0 - \hat{G}(t))dt + \Sigma_t dw_y^* \tag{B1}$$

where $\Sigma_t = \dfrac{S(t)+\sigma_{gy}}{\sigma_y}$ and $\sigma_{gy} = \rho_{gy}\sigma_y\sigma_g$.

Now we derive share price using standard SDE arguments based on stochastic discount factor (SDF) approach (see, e.g., Cochrane, 2005). The implicit assumption here is that any shock responsible for the difference between D_t and δY is not priced:

$$E_t^*\left[d\left(MP\right)\right] + M\delta Y dt = 0 \tag{B2}$$

Evaluating the differential and dividing through by $M\delta Y$ we obtain:

$$E_t^*\left(\frac{dM}{M}\frac{P}{\delta Y} + \frac{dP}{\delta Y} + \frac{dM}{M}\frac{dP}{\delta Y}\right) + dt = 0 \tag{B3}$$

We now guess a solution for the price in the following form:

$$\text{P}(\text{t, Y, } \hat{G}(t)\text{, R}) = \delta\text{YZ}(\text{t, } \hat{G}(t)\text{, R}) \tag{B4}$$

where δY represents dividends-per-share. Operator E_t^* represents an expectation with respect to dw_y^*, investors' information set. $Z(t,\hat{G},R)$ is the time-t price-dividend ratio.

The second and the third terms under the expectation in (B3) follow from a simple application of the Itô rule:

$$\frac{dP}{\delta Y} = dZ + Z\frac{dY}{Y} + dZ\frac{dY}{Y} = Z_t dt + Z_{\hat{G}} d\hat{G} + Z_R dR + \frac{1}{2}\left(Z_{\hat{G}\hat{G}} d\hat{G}^2 + 2Z_{\hat{G}R} d\hat{G} dR + Z_{RR} dR^2\right) +$$
$$Z\left(\hat{G}dt + \sigma_y dw_y^*\right) + \left(\sigma_y Z_{\hat{G}} \Sigma_t + Z_R \sigma_{yr}\right)dt, \text{ where } \sigma_{yr} = \rho_{yr}\sigma_y\sigma_r,$$
$$\frac{dM}{M}\frac{dP}{\delta Y} = -\left(\rho_{my}\sigma_m Z_{\hat{G}} \Sigma_t + \lambda_y Z + \lambda_r Z_R\right)dt, \text{ where } \lambda_y = \rho_{my}\sigma_m\sigma_y \text{ and } \lambda_r = \rho_{mr}\sigma_m\sigma_r.$$

Collecting all the terms in (B3), taking the expectation, and dividing through by *dt*, we obtain the PDE for the share price:

$$Z_t + k_g(\mu_g^* - \hat{G}(t))Z_{\hat{G}} + k_r(\mu_r^* - R)Z_R + \frac{\Sigma_t^2}{2} Z_{\hat{G}\hat{G}} + \frac{\sigma_r^2}{2} Z_{RR} + \rho_{gr}\sigma_r\Sigma_t Z_{\hat{G}R} - (R + \lambda_y - \hat{G})Z + 1 = 0,$$
$$\text{where } \mu_g^* = \mu_g^0 + \frac{(\sigma_y - \rho_{my}\sigma_m)\Sigma_t}{k_g}, \text{ and } \mu_r^* = \mu_r^0 + \frac{\sigma_{yr} - \lambda_r}{k_r}.$$

(B5)

The above PDE satisfies Feynman-Kac conditions, and therefore allows us to write the solution which can be written as follows:

$$Z = \int_t^\infty E_t^* \exp\left(\int_t^s (\hat{G}(u) - R(u) - \lambda_y)du\right)ds.$$

The integrand solves the same equation as Z with the free term 1 deleted from the equation. We look for an integrand solution as

$$\exp(\varphi(t,s) + \psi(t,s)\hat{G}(t) - \upsilon(t,s)R(t)).$$

Equivalently, we are looking for price-dividend ratio in the form:

$$Z(t,s,\hat{G},R) = \int_t^\infty \exp^{[\varphi(t,s) + \psi(t,s)\hat{G}(t) - \nu(t,s)R(t)]} ds. \qquad \text{(B6)}$$

Inserting the proposed expression for the integrand into its PDE and recognizing that the resulting ordinary differential equation (ODE) must hold for arbitrary values of $\hat{G}(t)$ and *R(t)*, we arrive at the following ODEs for functions $\varphi(t,s)$, $\psi(t,s)$, and $\upsilon(t,s)$ (prime denotes $\partial/\partial t$ derivative):

$$\psi' - k_g\psi + 1 = 0$$
$$\upsilon' - k_r\upsilon + 1 = 0 \qquad \text{(B7)}$$
$$\varphi' + \frac{\Sigma_t^2}{2}\psi^2 + k_g\mu_g^*\psi + \frac{\sigma_r^2}{2}\upsilon^2 - \rho_{gr}\sigma_r\Sigma_t\psi\upsilon - k_r\mu_r^*\upsilon + \lambda_y = 0$$

When $s=t$ (or $\tau \equiv s-t = 0$), the integrand is equal to zero. Therefore, we have the following initial conditions for functions $\varphi(t,s)$, $\psi(t,s)$, and $\upsilon(t,s)$:

$$\varphi(s,s)=\psi(s,s)=\upsilon(s,s)=0.$$

Subject to these initial conditions, the solution to the decoupled system (B7) is:

$$\psi(t,s)=\frac{1-e^{-k_g(s-t)}}{k_g},\quad \upsilon(t,s)=\frac{1-e^{-k_r(s-t)}}{k_r},\ \text{and}$$

$$\varphi(t,s)=-\lambda_y\tau+\int_t^s\left(\frac{\Sigma_t^2}{2}\psi^2+k_g\mu_g^*\psi+\frac{\sigma_r^2}{2}\upsilon^2-\rho_{gr}\sigma_r\Sigma_t\psi\upsilon-k_r\mu_r^*\upsilon\right)du. \tag{B8}$$

The requirement that the integral (B6) exist places certain restrictions on function φ(t,s). For the integral in (B6) to exist, the integrand should be declining with *s* sufficiently fast. Since functions ψ and υ are bounded, this requirement implies that function φ should be negative and unbounded at large *s*. The latter restriction implies certain constraint on model parameters (a transversality condition), which we now derive. We need three auxiliary results to complete the derivation of the transversality condition:

a. For any bounded positive function *f(u)* and positive constant *k*:

$$\sup_{u\in[0,\tau]} f(u)=M<\infty,\int_0^\tau f(u)e^{-k(\tau-u)}du\le M\int_0^\tau e^{-k(\tau-u)}du=\frac{M}{k}\left(1-e^{-k\tau}\right)\le\frac{M}{k} \tag{B9}$$

In what follows, we ignore non-growing integrals such as (B9) and keep only the leading terms that are unbounded in τ.

b. In equation (6) for the posterior variance of the MEGR estimate, *S(t)*, we assume that *S(0) > S₁*. This condition also implies that constant *C < 1*. Therefore, the solution for variance *S(t)* is

$$S(t)=S_2+\frac{S_1-S_2}{1-Ce^{\gamma(S_1-S_2)t}}$$

$$\int_t^s S_u du \overset{\tau=s-t}{=} \int_0^\tau\left(S_2+\frac{S_1-S_2}{1-Ce^{\gamma(S_1-S_2)u}}\right)du=S_1\tau-\frac{1}{\gamma}\ln\frac{1-Ce^{\gamma(S_1-S_2)\tau}}{1-C}\xrightarrow{\text{leading terms}}S_1\tau$$

(B10)

c. Using equation (6) for the posterior variance *S(t)* we have:

$$\int_t^s S_u^2du=\int_t^s\left(\frac{1}{\gamma}S_u^{'}-\eta S_u-\alpha\right)du=\frac{1}{\gamma}\big(S(s)-S(t)\big)-\eta\int_t^s S_u du-\alpha(s-t)\xrightarrow{\text{leading terms}}-\left(\eta S_1+\alpha\right)\tau$$

(B11)

Using results (B10) and (B11) as well as (B8) to eliminate non-growing integrals we obtain the following leading terms in function φ:

$$\varphi(t,s) = -\lambda_y \tau + \int_t^s \left(\frac{\Sigma_t^2}{2}\psi^2 + k_g \mu_g^* \psi + \frac{\sigma_r^2}{2}\upsilon^2 - \rho_{gr}\sigma_r \Sigma_t \psi \upsilon - k_r \mu_r^* \upsilon \right) du \xrightarrow{\text{leading terms}}$$

$$\xrightarrow{=} -\lambda_y \tau + \frac{1}{2(k_g \sigma_y)^2} \int_t^s \left(S_u^2 + 2\sigma_{gy} S_u + \sigma_{gy}^2 \right) du + \int_t^s \left(\mu_g^0 + \frac{\sigma_y - \rho_{my}\sigma_m}{k_g} \frac{S_u + \sigma_{gy}}{\sigma_y} \right) du +$$

$$+ \frac{\sigma_r^2}{2k_r^2}\tau - \frac{\rho_{gr}\sigma_r}{k_r k_g \sigma_y} \int_t^s \left(S_u + \sigma_{gy} \right) du - \mu_r^* \tau \xrightarrow{\text{leading terms}}$$

$$\left(-\lambda_y + \frac{\sigma_r^2}{2k_r^2} - \mu_r^* \right)\tau + \frac{\tau}{2(k_g \sigma_y)^2}\left(-\eta S_1 - \alpha + 2\sigma_{gy} S_1 + \sigma_{gy}^2 \right) +$$

$$\left(\mu_g^0 + \frac{\sigma_y - \rho_{my}\sigma_m}{k_g \sigma_y}\left(S_1 + \sigma_{gy} \right) \right)\tau - \frac{\rho_{gr}\sigma_r}{k_r k_g \sigma_y}\left(S_1 + \sigma_{gy} \right)\tau$$

$$= \left[-\lambda_y + \frac{\sigma_r^2}{2k_r^2} - \mu_r^* + \mu_g^0 + \frac{-\eta S_1 - \alpha + 2\sigma_{gy} S_1 + \sigma_{gy}^2}{2(k_g \sigma_y)^2} + \left(\frac{\sigma_y - \rho_{my}\sigma_m}{k_g \sigma_y} - \frac{\rho_{gr}\sigma_r}{k_r k_g \sigma_y} \right)\left(S_1 + \sigma_{gy} \right) \right]\tau$$

Finally, the transversality condition states that the leading terms must be negative for the price integral to exist:

$$-\lambda_y + \frac{\sigma_r^2}{2k_r^2} - \mu_r^* + \mu_g^0 + \frac{(2\sigma_{gy} - \eta)S_1 + \sigma_{gy}^2 - \alpha}{2(k_g \sigma_y)^2} + \left((\sigma_y - \rho_{my}\sigma_m)k_r - \rho_{gr}\sigma_r \right)\frac{(S_1 + \sigma_{gy})}{k_r k_g \sigma_y} < 0$$

(B12)

Applying Itô's lemma to $P(t)$, we obtain a stochastic differential equation (SDE) for $P(t)$, subject to $P(t) < \infty$:

$$\frac{dP(t)}{P(t)} = [\frac{1}{Z}Z_t + \hat{G} + k_g(\mu_g^0 - \hat{G})\frac{1}{Z}Z_{\hat{G}} + \frac{\Sigma_t^2}{2Z}\frac{1}{Z}Z_{\hat{G}\hat{G}} + \Sigma_t \sigma_y \frac{1}{Z}Z_{\hat{G}} + \rho_{yr}\sigma_r \sigma_y \frac{1}{Z}Z_R +$$

$$\rho_{gr}\sigma_r \Sigma_t \sigma_y \frac{1}{Z}Z_{\hat{G}R} + k_r(\mu_r^0 - R)\frac{1}{Z}Z_R + \frac{\sigma_r^2}{2Z}Z_{RR}]dt + [\sigma_y + \Sigma_t \frac{1}{Z}Z_{\hat{G}}]dw_y^* + \sigma_r \frac{1}{Z}Z_R dw_r.$$

Plugging the SDE for $\dfrac{dP(t)}{P(t)}$ into equation (B3), we get the following risk-neutral drift of stock return in an incomplete-information environment,

$$E_t^*\left(\frac{dP}{P} \right) + \frac{dY}{Y}dt = Rdt + \sigma_m\left(\sigma_1 \rho_{my} + \sigma_2 \rho_{mr} \right)dt,$$

where $\sigma_1 = \sigma_y + \Sigma_t \frac{1}{Z} Z_{\hat{G}}$, and $\sigma_2 = \sigma_r \frac{1}{Z} Z_R$.

QED.

Proof of Proposition 4:

The risk premium for *G(t)* based on complete-information BC model is defined by λ_g^{BC}, given below,

$$\lambda_g^{BC} = \rho_{mg}\sigma_g\sigma_m - \sigma_{gy},$$

The risk premium for $\hat{G}(t)$ based on our incomplete-information model is defined by λ_g, given below,

$$\lambda_g = (\rho_{my}\sigma_m - \sigma_y)\Sigma_t = (\rho_{my}\sigma_m - \sigma_y)\left(\frac{S(t) + \sigma_{gy}}{\sigma_y}\right)$$

$$= (\rho_{my}\sigma_m - \sigma_y)\frac{S(t)}{\sigma_y} + (\rho_{my}\rho_{gy}\sigma_g\sigma_m - \sigma_{gy}),$$

Therefore, the difference in risk premiums between our incomplete-information model and complete-information model (BC) is given by,

$$\Delta\lambda_g = \lambda_g - \lambda_g^{BC} = (\rho_{my}\sigma_m - \sigma_y)\frac{S(t)}{\sigma_y} + \sigma_g\sigma_m(\rho_{my}\rho_{gy} - \rho_{mg}).$$

In the long-run limit as $S(t) \to S_1$=0, we obtain the following result:

$$\Delta\lambda_g \to \sigma_g\sigma_m(\rho_{my}\rho_{gy} - \rho_{mg}) = \sigma_g\sigma_m\rho_{mg}\left(\frac{\rho_{my}\rho_{gy}}{\rho_{mg}} - 1\right).$$

For special cases such as $\rho_{gy} = 1$ or $\rho_{gy} = -1$, we get $\frac{\rho_{my}\rho_{gy}}{\rho_{mg}} = |\rho_{gy}| = 1$. Therefore, the term in parentheses above disappears, which implies a zero difference in risk premiums. That is $\Delta\lambda_g \to 0$.

Let μ_g^{BC} denote the risk-neutral long-term mean of earnings growth under BC model. Following BC model, $\mu_g^{BC} = \mu_g^0 + \frac{\sigma_{gy} - \rho_{mg}\sigma_g\sigma_m}{k_g}$. The parameter μ_g^* denotes the risk-neutral long-term mean of EPS growth rate under our incomplete-information model, as given below:

$$\mu_g^* = \mu_g^0 - \frac{\lambda_g}{k_g} = \mu_g^0 + \frac{\sigma_y - \rho_{my}\sigma_m}{k_g}\left(\frac{S(t) + \sigma_{gy}}{\sigma_y}\right).$$

According to Proposition (1), if $\rho_{gy} = 1$ or $\rho_{gy} = -1$, $S(t)$ declines over time and converges to zero.[12] Therefore, as $S(t) \rightarrow 0$

$$\mu_g^* \equiv \mu_g^0 - \frac{\lambda_g}{k_g} \xrightarrow{S(t)\rightarrow 0} \mu_g^0 + \frac{\sigma_{gy} - \rho_{my}\rho_{gy}\sigma_g\sigma_m}{k_g} = \mu_g^0 + \frac{\sigma_{gy} - \rho_{mg}\sigma_g\sigma_m}{k_g} = \mu_g^{BC}.$$

QED.

[12] The only exception is the case of $k_g < |\beta|$.

References

Bakshi, G. and Chen, Z. (2005) “Stock Valuation in Dynamic Economies”, *Journal of Financial Markets*. No. 2, 115-151.

Breeden, D.T. (1979) “An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities”, *Journal of Financial Economics*, 7, 265-296.

Brennan, M.J. (1998) “The Role of Learning in Dynamic Portfolio Decisions”, *European Finance Review*, 1, 295-396.

Brennan, M.J. and Xia, Y. (2001) “Stock Return Volatility and Equity Premium”, *Journal of Monetary Economics*, Vol. 47:249-83.

Cagetti, M., Lars Peter Hansen, T. S., and Williams, N. (2002) “Robustness and Pricing with Uncertain Growth”, *Review of Financial Studies* 15, 363–404.

Campbell, J.Y. (1996) “Consumption and the Stock Market: Interpreting International Experience”, *Swedish Economic Policy Review*, 3, pp. 251–299.

Cochrane, J.H. (2005) “Asset Pricing” (Revised Edition), *Princeton University Press*, Princeton, NJ.

Detemple, J.B. (1986) “Asset Pricing in a Production Economy with Incomplete Information”, Journal of Finance, 41, pp. 383-91.

Dothan, M.U. and Feldman, D. (1986) “Equilibrium Interest Rates and Multiperiod Bonds in a Partially Observable Economy”, Journal of Finance, 41, pp.369-82.

Dong, M. and Hirshleifer, D. (2004), “A Generalized Earnings-Based Stock Valuation Model” unpublished working paper, York University and Ohio State University.

Epstein, L., and Schneider, M. (2008) “Ambiguity, Information Quality and Asset Prices”, *Journal of Finance* 63, 197–228.

Fama, E.F. and French, K.R. (2001) “Disappearing Dividends: Changing Characteristics or Lower Propensity to Pay”, Journal of Financial Economics, 60, pp.3-43.

Faust, J., J.H. Rogers, and J. H. Wright (2000), “News and Noise in G-7 GDP Announcements”, Federal Reserve Board.

Feldman, D. (2007) “Incomplete Information Equilibria: Separation Theorems and Other Myths”, *Annals of Operations Research*, *Special Issue on Financial Modeling*, Vol. 151, pp. 119-149.

Gennotte, G. (1986) “Optimal Portfolio Choice under Incomplete Information”, *Journal of Finance*, 41, pp.733-46.

Gordon, M. (1962). “The Investment, Financing and Valuation of the Corporation” Home-wood, IL: Irwin.

Grossman, S.J. and Shiller, R.J. (1981) “The Determinants of the Variability of Stock Market Prices”, *American Economic Review*, 71, pp.222-227.

Harrison, M. and Kreps, D. (1979) “Martingales and Arbitrage in Multiperiod Security Markets”, *Journal of Economic Theory*, Vol. 20, pp. 381-408.

Lee, C., Myers, J. and Swaminathan, B. (1999) “What is the intrinsic value of the Dow?” Journal of Finance, 54, 1693–1741.

Leippold, M., Trojani, F., and Vanini, P. (2008) “Learning and Asset Prices under Ambiguous Information”, *Review of Financial Studies* 21, 2565–2597.

Liptser, R. S. and Shiryayev, A. N. (1977) *Statistics of Random Processes I: General Theory*. New York: Springer-Verlag.

Liptser, R. S. and Shiryayev, A. N. (1978) *Statistics of Random Processes II: Applications*. New York: Springer-Verlag.

Pastor, L. and Veronesi, P. (2003) “Stock Valuation and Learning about Profitability”, *Journal of Finance*, 58: 1749-89.

Pastor, L. and Veronesi, P. (2006) “Was There a Nasdaq Bubble in the Late 1990s?” *Journal of Financial Economics*, Vol. 81: 61-100.

Pastor, L. and Veronesi, P. (2009) “Technological Revolutions and Stock Prices”, The Annual Review of Financial Economics, 99: 1451-83.

Merton, R. C. (1971) “Optimum Consumption and Portfolio Rules in a Continuous Time Model”, *Journal of Economic Theory*, 3: 373-413.

Merton, R. C. (1973) “An Intertemporal Capital Asset Pricing Model”, *Econometrica*, Vol. 41, No. 5, September.

Timmermann, AG. (1993). “How Learning in Financial Markets Generates Excess volatility and Predictability of Stock Returns”, *Quarterly Journal of Economics*, Vol. 108: 1135-45.

Samuelson, PA. (1969) “Lifetime Portfolio Selection by Dynamic Stochastic Programming”, *Review of Economics and Statistics*, 51, 239-246.

Shapiro, M. D., and D. W. Wilcox, (1996) “Mismeasurement in the Consumer Price Index: an Evaluation,” *NBER Macroeconomic Annual*.

Vasicek, O. (1977). “An Equilibrium Characterization of the Term Structure”, *Journal of Financial Economics*, Vol. 5, pp. 177-188.

Williams, J.T. (1977) “Capital Asset Prices with Heterogeneous Beliefs”, *Journal of*` *Financial Economics*, 5: 219-39.