

# A Theory of Income Smoothing: Online Appendices

## (RFS MS 20130779)

### 1 Appendix A: Extensions

In this appendix we develop two extensions to the paper's model by considering the effects of 1) stock-based compensation and 2) sales-based compensation.

#### 1.1 Stock-based compensation

Stein (1989) argues that stock-based compensation induces insiders to inflate income. How does stock-based compensation affect insiders' production incentives in our setting where market pressures apply not only with respect to the current stock price but also with respect to future payout? To explore this question we now consider the scenario where insiders get each period a fraction  $\delta$  of the existing *outside* equity. Insiders get the shares *cum* dividend and must sell them in the market upon receipt (in contrast to their existing stockholding  $1 - \varphi$  which they are not allowed to sell).<sup>1</sup> Outsiders know that their equityholding will be diluted each period by a fraction  $1 - \delta$ , and take this into account when pricing the outside equity,  $S_t$ . Managers' optimization problem is now given by:

$$M_t = \max_{q_{t+j}; j=0..\infty} E_t \left[ \sum_{j=0}^{\infty} \beta^j (\pi(q_{t+j}) - \theta E_{S,t+j} [\pi(q_{t+j})] + \delta S_{t+j}) \right] \quad (1)$$

Solving this problem gives the following proposition:

**Proposition 1** *If insiders get each period the cash equivalent of a fraction  $\delta$  of the outside equity then their optimal production decision is given by  $q_t = Hx_t$  where  $H$  is the solution to:*

$$H = 1 - \frac{hK\theta \left( 1 - \frac{\delta}{1 - \beta(1 - \delta)A} \right)}{1 - \beta\lambda A} \quad (2)$$

*The value of the outside equity (cum dividend) at time  $t$  is:*

$$S_t = \theta E_{S,t} \left[ \sum_{i=0}^{\infty} \beta^i (1 - \delta)^i \pi(q_{t+i}) \right] = \frac{\theta h}{1 - \beta A(1 - \delta)} \left( \hat{x}_t + \frac{B\beta(1 - \delta)}{1 - \beta(1 - \delta)} \right) \quad (3)$$

---

<sup>1</sup>It is not crucial for the analysis that shares are sold immediately. The key restriction is that insiders do not have discretion regarding the timing of the sale, as this would introduce an adverse selection and an optimal stopping problem.

*Stock based compensation mitigates, but does not eliminate the underinvestment problem except if outsiders in effect own 100% of the firm (i.e.  $\delta = 1$ ).*

Equation (2) shows that increasing stock-based compensation is similar to reducing  $\theta$ , outsiders' stake in the firm. Therefore stock-based compensation unambiguously improves efficiency and mitigates the under-investment problem.<sup>2</sup> From (2) it is clear that  $H = 1$  if  $\delta = 1$ , i.e., the efficient outcome is achieved if outside equityholders get 100% diluted each period.

Unlike Stein (1989) insiders do not have an incentive to inflate income in the presence of stock compensation because market pressures do not only apply to the current stock price but also to future payout. By inflating income insiders not only inflate the current stock price, but also outsiders' expectations regarding future dividend payout. Therefore insiders' immediate gain with respect to their stock-based compensation is more than offset by the loss from paying higher future dividends (unless insiders own 100% of the firm). This argument becomes clear when one realizes that insiders' objective function (1) is decreasing in  $\hat{x}_t$  (because  $(1 - \delta)/(1 - \beta A(1 - \delta))$  is always positive). Consequently, insiders' payoff is also decreasing in outsiders' beliefs about income and therefore insiders do not wish to inflate income. Still, stock compensation mitigates the underinvestment problem as it reduces outsiders' effective stake in the firm and therefore also the importance of outsiders' beliefs for insiders' production decision.

How then can incentives to inflate income arise? High powered compensation mechanisms (such as stock options, or other contracts that are convex in income) that lever up the effect of income changes may be a possible explanation. Giving insiders a tenure of limited duration may also encourage them to inflate income because they escape the market discipline with respect to future dividend payout once they are retired and they leave it to their successors to meet the raised expectations. Similarly, incentives to inflate income may arise in the run-up to an anticipated cash offer that allows insiders to cash in their shares and flee.

---

<sup>2</sup>Stock-based compensation may, however, introduce other (dis)incentives not considered in this paper. E.g. Benmelech, Kandel, and Veronesi (2010) show that stock-based compensation not only induces managers to exert costly effort, but also induces them to conceal bad news about future growth options and to choose suboptimal investment policies to support the pretense.

## 1.2 Sales-based compensation

Given that stock-based compensation does not fully eliminate the under-production problem, a natural question to ask is whether there exists a contractual compensation scheme that leads to the efficient outcome. We show below that introducing a bonus for insiders that is proportional to the observable sales can induce full efficiency.

In particular, suppose that insiders get each period a cash bonus equal to  $b s_{t+j}$ , then managers' optimization problem is now given by:

$$M_t = \max_{q_{t+j}; j=0..∞} E_t \left[ \sum_{j=0}^{\infty} \beta^j (\pi(q_{t+j}) - \theta E_{S,t+j} [\pi(q_{t+j})] + b s_{t+j}) \right] \quad (4)$$

Solving this problem gives the following proposition:

**Proposition 2** *Insiders implement the first-best production decision  $q_t = x_t$  if they get each period a cash bonus equal to  $b s_{t+j}$  where the constant  $b$  is given by:*

$$b = \frac{(1 - \beta)\theta P}{2[P + R(1 - \beta A)]} \quad (5)$$

The value for the optimal level of  $b$  turns out to be surprisingly small. E.g. if the discount factor  $\beta$  equals, say, 0.95 then  $b$  must be below 0.025 (since  $\theta < 1$ ). A relatively small cash bonus linked to sales can therefore eliminate the underinvestment, even if insiders have little or no stake in the firm. For example, for the parameter values used in the figures the optimal value for  $b$  equals 0.0197 or almost 2% of sales. Note that the optimal value for  $b$  increases with outsiders' ownership stake ( $\theta$ ) and the variance of outsiders' income estimate (as determined by  $P$ ).

Although sales-based compensation results in the first-best outcome, there are a number of potential issues that make this type of contract problematic in practice. The relevant decision variable in the model is output. Output is, however, not observable, and instead the contract is written on sales, an observable but noisy measure of output. The noise component in sales (e.g. due measurement error) makes it, however, problematic to verify this variable in court, which in turn makes it difficult to enforce the contract. One can show that a myopic outsider will always want to renege on the sales compensation. Even an outsider with a long term perspective may find it optimal to renege if the measurement error is sufficiently positive. Since  $s_t = q_t + \epsilon_t$ , each dollar of measurement error in sales costs outsiders  $b$  dollars in sales

compensation. Therefore, if the measurement error is too large, outsiders will refuse to pay and try to prove in court that the sales figure is unreliable. These issues might explain why sales-based compensation is less prevalent than stock compensation.<sup>3</sup>

**Proof of Proposition 1:** The valuation equation for  $S_t$  follows immediately from proposition 6 in the main text (by substituting  $\beta$  by  $\beta(1 - \delta)$ ). The derivation of the equilibrium value for  $H$  is as given in the proof to proposition 4 in the main text but with  $\theta$  replaced by  $\theta \left(1 - \frac{\delta}{1 - \beta(1 - \delta)A}\right)$ .

**Proof of Proposition 2:** The derivation of the equilibrium value for  $H$  is analogous as in the proof to proposition 4 in the main text, and leads to the following equilibrium condition for  $H$  (for a given value of  $b$ ):

$$H = 1 - \frac{\theta H^2 P \left(1 - \frac{H}{2}\right)}{H^2 P + R(1 - \beta A)} + \frac{b}{1 - \beta} \quad (6)$$

Setting  $H = 1$  in the above equation, and solving for  $b$  gives the expression in proposition 2

## 2 Appendix B: Robustness

In this appendix we demonstrate the robustness of our results. Instead of using sales as the observable we first consider the case in which output ( $q_t$ ) is the observable. Next, we depart from the quadratic income specification by introducing a Cobb Douglas production function with labor as input. We then treat the input measure (labor) as the observable. One can easily verify that properties (1) to (6) remain valid. More generally, changing the income specification or the variable that is being observed does not alter the results in any essential way.

### A) Latent marginal revenues with observable output

---

<sup>3</sup>Even though pure stock compensation appears somewhat less effective in mitigating the underinvestment problem, it is easier to implement and enforce. First, the stock price is not only observable but also verifiable in court. Therefore, stock based compensation should be enforceable for public firms (provided the penalty from renegeing on the stock compensation is sufficiently large). Second, the stock price (and therefore the stock compensation component) is less sensitive to the noise term than sales-based compensation. The reason is that the stock market partially filters out the transitory effect of measurement error.

Consider the following income function:

$$\pi_t = q_t x_t - \frac{q_t^2}{2} \text{ with } x_t > 0 \quad (7)$$

Outsiders observe output (i.e.  $s_t = q_t$ ) and conjecture that income is quadratic in output, i.e.  $\hat{\pi}_t = a_0 q_t^2 + b_0$ . The first and second order conditions are  $q_t = x_t / (1 + 2a_0\theta)$  and  $1 + 2a_0\theta > 0$ , respectively. Outsiders' conjecture is verified if  $a_0 = 0.5 / (1 - 2\theta)$  and  $b_0 = 0$ . This gives the following proposition:

**Proposition 3** *If the firm's income function is given by  $\pi(q_t) = q_t x_t - \frac{q_t^2}{2}$  then the first-best production policy is  $q_t^o = x_t$  and the realized income under the first-best policy is  $\pi_t^o = \frac{x_t^2}{2}$ . If outsiders can observe the firm's output ( $q_t$ ), but not the latent variable  $x_t$ , then the production policy adopted by insiders is:*

$$q_t = \frac{(1 - 2\theta)x_t}{(1 - \theta)} < q_t^o \quad (8)$$

The realized income is given by:

$$\pi_t = \frac{(1 - 2\theta)x_t^2}{2(1 - \theta)^2} < \pi_t^o \quad (9)$$

The payout to outsiders is  $d_t = \theta \hat{\pi}_t = \theta \pi_t$ . Insiders wish to produce and run the firm if and only if  $\theta < 0.5$ .

The results for latent marginal revenues with observable output are similar to the ones we obtained in previous section. In particular, one can easily verify that properties 1 to 6 are valid. Note that for the particular income function employed here the underproduction problem can become so severe that the firm stops producing altogether. In particular, once insiders lose majority control ( $\theta \geq 0.5$ ), insiders stop producing and no longer wish to run the firm.

## B) Cobb-Douglas production function with observable input

Consider next the following income function:

$$\pi(L_t; x_t) = x_t L_t^\alpha - w L_t \text{ with } w, x_t > 0 \quad (10)$$

Outsiders observe the labor input  $L_t$ , but not  $x_t$ , the price per unit of output. Outsiders conjecture that  $\hat{\pi}_t = a_0 L_t + b_0$ . Substituting this conjecture into insiders' objective function,

the first and second order conditions are:

$$\frac{\partial M_t}{\partial L_t} = x_t \alpha L_t^{\alpha-1} - w - \theta a_0 = 0 \quad (11)$$

$$\frac{\partial^2 M_t}{\partial L_t^2} = x_t \alpha (\alpha - 1) L_t^{\alpha-2} < 0 \quad (12)$$

Outsiders' conjecture is verified if and only if:

$$\hat{\pi}_t = a_0 \left[ \frac{\alpha x_t}{w + \theta a_0} \right]^{\frac{1}{1-\alpha}} + b \quad (13)$$

$$= x_t \left[ \frac{\alpha x_t}{w + \theta a_0} \right]^{\frac{1}{1-\alpha}} - w \left[ \frac{\alpha x_t}{w + \theta a_0} \right]^{\frac{1}{1-\alpha}} = \pi_t \quad (14)$$

$$\iff a_0 \left[ \frac{\alpha}{w + \theta a_0} \right]^{\frac{1}{1-\alpha}} x_t^{\frac{1}{1-\alpha}} + b = \left[ \frac{\alpha}{w + \theta a_0} \right]^{\frac{1}{1-\alpha}} x_t^{\frac{1}{1-\alpha}} - w \left[ \frac{\alpha}{w + \theta a_0} \right]^{\frac{1}{1-\alpha}} x_t^{\frac{1}{1-\alpha}} \quad (15)$$

$$\iff a_0 = \frac{w(1-\alpha)}{\alpha - \theta} \quad \text{and} \quad b = 0 \quad (16)$$

Input is positive if and only if  $L_t > 0 \iff w + \theta a_0 > 0 \iff \theta < \alpha$ . One can immediately verify that the second order condition is satisfied if  $0 < \alpha < 1$  and  $x_t > 0$ . Solving for  $L_t$  and  $\pi_t$  gives equations (17) and (18), in the below proposition, respectively.

**Proposition 4** *If the firm's income function is given by  $\pi_t = x_t L_t^\alpha - w L_t$  and labor  $L_t$  (but not  $x_t$ ) is observable to outsiders, then the insiders' (first-best) input policy  $L_t$  ( $L_t^o$ ) and the firm's (first-best) income  $\pi_t$  ( $\pi_t^o$ ) are given by:*

$$L_t = \left[ \frac{\alpha x_t}{w + \theta a_0} \right]^{\frac{1}{1-\alpha}} = \left[ \frac{\alpha - \theta}{w(1-\theta)} \right]^{\frac{1}{1-\alpha}} x_t^{\frac{1}{1-\alpha}} < \left[ \frac{\alpha x_t}{w} \right]^{\frac{1}{1-\alpha}} \equiv L_t^o \quad (17)$$

$$\pi_t = \frac{w(1-\alpha)}{\alpha - \theta} \left[ \frac{\alpha - \theta}{w(1-\theta)} \right]^{\frac{1}{1-\alpha}} x_t^{\frac{1}{1-\alpha}} < (1-\alpha) \left[ \frac{\alpha}{w} \right]^{\frac{1}{1-\alpha}} x_t^{\frac{1}{1-\alpha}} \equiv \pi_t^o \quad (18)$$

One can easily verify that properties 1 to 6 are still valid.

### 3 Appendix C: Additional proofs

#### 3.1 Uniqueness proof for $H$ (proposition 4)

In the appendix (see proof of proposition 4 in the main text) we proved there exists an equilibrium. We now prove that this equilibrium is unique.

Recall that  $H$  is the solution to:

$$f(H) \equiv 1 - H - \frac{\theta H^2 P (1 - \frac{H}{2})}{H^2 P + R(1 - \beta A)} \equiv 1 - H - g(H) = 0 \quad (19)$$

To prove that the above condition has a unique positive root, we must prove that  $f(H)$  is a decreasing function, i.e. that  $g'(H) > -1$ .

Let  $v(H) = \frac{1}{g(H)}$ . We have that  $v'(H) = -g'(H)v(H)^2$ .

We will prove that  $v'(H) < v(H)^2$ . Note that, if we prove this, it follows that  $g'(H) > -1$  and that  $f'(H) < 0$ .

Note that  $v(H) = \frac{1}{\theta(1-\frac{H}{2})} + \frac{R(1-\beta A)}{\theta} \frac{1}{H^2 P(1-\frac{H}{2})}$ . Therefore, we have that:

$$v'(H) = \frac{1}{2\theta(1-\frac{H}{2})^2} - \frac{R(1-\beta A)}{\theta} \frac{1}{(H^2 P(1-\frac{H}{2}))^2} \frac{d}{dH}(H^2 P(1-\frac{H}{2}))$$

Since  $P$  is the positive root of the quadratic equation (23), we have that

$$H^2 P = \frac{QH^2 - R(1-A^2)}{2} + \frac{\sqrt{4H^2 QR + (QH^2 - R(1-A^2))^2}}{2}$$

Therefore,

$$\frac{d}{dH}(H^2 P) = QH + \frac{1}{4\sqrt{X}}(8HQR + 2(QH^2 - R(1-A^2))2QH) \quad (20)$$

$$= QH(1 + \frac{1}{\sqrt{X}}(2R + QH^2 - R(1-A^2))) \equiv Y \quad (21)$$

where  $X \equiv 4H^2 QR + (QH^2 - R(1-A^2))^2$ . Denote the above derivative by  $Y$ . Note that  $Y > 0$ . Therefore, we have that  $\frac{d}{dH}(H^2 P(1-\frac{H}{2})) = (1-\frac{H}{2})Y - \frac{H^2 P}{2}$ . Substituting into  $v'(H)$ , we get that

$$v'(H) = \frac{1}{2\theta(1-\frac{H}{2})^2} - \frac{R(1-\beta A)}{\theta} \frac{1}{(H^2 P(1-\frac{H}{2}))^2} ((1-\frac{H}{2})Y - \frac{H^2 P}{2})$$

Therefore,  $v'(H) < v(H)^2$  reduces to the inequality

$$\frac{1}{2} - R(1-\beta A) \frac{1}{(H^2 P)^2} ((1-\frac{H}{2})Y - \frac{H^2 P}{2}) < \frac{1}{\theta} (1 + R(1-\beta A) \frac{1}{H^2 P})^2$$

Let us prove the above by setting  $\theta$  equal to 1 (the proof then automatically holds for  $\theta < 1$ ).

Rearranging terms, we get that we must prove

$$-\frac{1}{2} - \frac{3}{2}R(1-\beta A) \frac{1}{2(H^2 P)^2} < R(1-\beta A) \frac{1}{(H^2 P)^2} (1-\frac{H}{2})Y + (R(1-\beta A) \frac{1}{H^2 P})^2$$

Since  $Y$  is positive, the RHS is positive, while the LHS is negative. Therefore, the above inequality is true, and retracing our steps, we see that  $f(H)$  is a decreasing function.

Therefore, there is at most one root. Noting that  $f(0) = 1 > 0$  and  $f(1) = -\frac{\theta P}{2(P+R(1-\beta A))} < 0$ , it follows that there exists a unique  $H \in ]0, 1[$  for which  $f(H) = 0$ .

### 3.2 Uniqueness proof for $P$ (proposition 4)

In the appendix (see proof of proposition 4 in the main text) we claimed that  $P$  is the positive root of the equation:

$$P = A^2P - \frac{A^2H^2P^2}{H^2P + R} + Q. \quad (22)$$

We now prove that equation (22) has indeed one positive and one negative root. Simplifying, equation (22) can be rewritten as:

$$H^2P^2 + P[R(1 - A^2) - QH^2] - QR = 0 \quad (23)$$

This is a quadratic equation in  $P$  which has 2 roots. The product of the roots is negative and given by  $-QR/H^2 < 0$ . Consequently, one root is positive and the other negative. Since  $P$  represents a variance we need the positive root.

### 3.3 Insiders' participation constraint

Under symmetric information insiders' participation constraint is always satisfied. This is no longer the case under asymmetric information.

Insiders' payout policy guarantees that the capital market constraint is satisfied at all times, i.e.,  $S_t \geq \theta E_t[V_t|I_t]$ . But will insiders be willing to adhere to this payout policy under all circumstances? Insiders' participation constraint is satisfied if they are better off paying out than triggering collective action. Collective action implies that stockholders "open up" the firm and uncover its true value ( $V_t$ ). It is reasonable (although not necessary) to assume that collective action also imposes a cost upon insiders. Without loss of generality assume that these costs are proportional to the firm value and given by  $C_t = cV_t$ .

"Forced disclosure" pricks the bubble that has been building up over time and brings outsiders' beliefs about the firm value back to reality, i.e.,  $E_t[V_t|I_t] = V_t$ . A sufficient (but not necessary) condition for insiders to keep paying out according to outsiders' expectations is:

$$M_t = V_t - \theta E_t[V_t|I_t] \geq V_t - \theta V_t - cV_t \iff V_t \geq \frac{\theta}{\theta+c} E_t[V_t|I_t] \quad (24)$$

Insiders have an incentive to trigger collective action if the firm's actual value ( $V_t$ ) drops

sufficiently below what outsiders believe the firm to be worth ( $E_t[V_t|I_t]$ ).<sup>4</sup> This situation arises if outsiders' beliefs about the latent cost variable (as reflected by  $\hat{x}_t$ ) are overoptimistic due to measurement errors.

How can one reduce the likelihood of costly forced disclosure? One obvious solution is to reduce the outside ownership stake  $\theta$  as this relaxes insiders' participation constraint. Unfortunately, this also reduces the firm's capacity to raise outside equity. Therefore, firms that rely heavily on outside equity (e.g. public firms) adopt more efficient (in terms of cost and speed) disclosure mechanisms such as voluntary audited disclosure. While "big baths" do occur in reality, they rarely result from a very costly forced disclosure process but they are much more likely to happen through the process of regular voluntary audited disclosures. An earlier version of this paper models and discusses the role of audited disclosure. The results are available from the authors upon request.

## 4 Appendix D: On the optimality of the Kalman Filter

In this Appendix we first derive the optimal filter through Bayesian inference. We then compare the performance of our Kalman filter with the Bayesian filter. Finally, we run some diagnostic tests to assess the overall optimality of our Kalman filter.

### 4.1 Optimal Bayesian filter

We start off by deriving the optimal filter through recursive Bayesian inference. The recursion starts from a prior distribution  $p(x_0)$ . We assume that  $x_t$  is observable to all parties at  $t = 0$ . Hence:  $E[x_0] = x_0$  and  $var(x_0) = P_0 = 0$ . From Bayes' rule it follows that:

$$p(x_1|s_1) = \frac{p(s_1|x_1)p(x_1)}{p(s_1)} \quad (25)$$

where  $p(s_1)$  is the normalization constant defined as  $p(s_1) = \int p(s_1|x_1)p(x_1)dx_1$ .

We know that:

$$x_1 = Ax_0 + B^* + w_0 \text{ where } w_0 \sim N(0, Q^*; -B^*) \quad (26)$$

$$s_1 = Hx_1 + \epsilon_1 \text{ where } \epsilon \sim N(0, R) \quad (27)$$

---

<sup>4</sup>Calculating the exact condition under which insiders optimally exercise their *option* to trigger collective action by outsiders is beyond the scope of this paper.

Hence,  $x_1 \in (Ax_0, +\infty)$ . The prior estimate for  $x_1$ ,  $x_1^-$ , has a truncated normal distribution  $x_1^- \sim N(Ax_0 + B^*, Q^*; Ax_0) \equiv N(\hat{x}_1^-, P_1^-; Ax_0)$ , while  $s_1 \sim N(Hx_1, R)$ . The correlation between  $x_1$  and  $s_1$  is given by

$$\rho = \left( \frac{H^2 P_1^-}{H^2 P_1^- + R} \right)^{1/2}$$

Hence,  $p(x_1|s_1)$  is a truncated normal distribution, i.e.  $p(x_1|s_1) \sim N(\frac{b_1}{a_1}, \frac{1}{a_1}; Ax_0)$  where:

$$\frac{b_1}{a_1} \equiv \hat{x}_1^- + (s_1 - H\hat{x}_1^-) \frac{HP_1^-}{(H^2 P_1^- + R)} \quad (28)$$

$$\frac{1}{a_1} \equiv \left( 1 - H \frac{HP_1^-}{(H^2 P_1^- + R)} \right) P_1^- \equiv (1 - HK_1) P_1^- \quad (29)$$

The density is given by:

$$p(x_1|s_1) \propto \frac{1}{\sqrt{2\pi(1-\rho^2)P_1^-} Z_1} e^{-\frac{1}{2} \left[ \frac{(x_1 - \hat{x}_1^- - (P_1^-/R) \rho (s_1 - H\hat{x}_1^-))^2}{(1-\rho^2) P_1^-} \right]}$$

where  $Z_1 \equiv 1 - N(\alpha_1)$  and

$$\alpha_1 \equiv \left( Ax_0 - \frac{b_1}{a_1} \right) \sqrt{a_1} \quad (30)$$

It follows from the properties of truncated normal distributions that the mean and variance of  $p(x_1|s_1)$  are given by:

$$\hat{x}_{1bf} = \frac{b_1}{a_1} + \frac{n(\alpha_1)}{Z_1 \sqrt{a_1}} \quad (31)$$

$$P_{1bf} = \frac{1}{a_1} \left[ 1 + \frac{\alpha_1 n(\alpha_1)}{Z_1} - \left( \frac{n(\alpha_1)}{Z_1} \right)^2 \right] \quad (32)$$

where  $n(\cdot)$  and  $N(\cdot)$  are, respectively, the standard normal density and cumulative distribution. Contrary to the linear Kalman filter, the optimal Bayesian estimator  $\hat{x}_{1bf}$  includes a non-linear term  $\frac{n(\alpha_1)}{Z_1 \sqrt{a_1}}$ . The larger the distance between the unconditional (i.e. untruncated) mean  $\frac{b_1}{a_1}$  and the truncation point  $Ax_0$ , the smaller the non-linear term.

We know from proposition 4 in the main text that the linear Kalman Filter is given by:

$$\hat{x}_{kf} = Ax_0 + B^* + m + (s_1 - H(Ax_0 + B^* + m)) K_{kf} \quad (33)$$

where  $K_{kf} \equiv \frac{HP_{1kf}^-}{H^2 P_{1kf}^- + R} = \frac{HQ}{H^2 Q + R}$  and  $m \equiv \frac{n(-\frac{B^*}{\sqrt{Q^*}}) \sqrt{Q^*}}{1 - N(\frac{-B^*}{\sqrt{Q^*}})}$ . Hence, the difference between the Kalman estimator and the Bayesian estimator is given by:

$$\hat{x}_{1kf} - \hat{x}_{1bf} = m(1 - HK_{kf}) + (s_1 - H(Ax_0 + B^*)) (K_{kf} - K_{bf}) - \frac{n(\alpha_1)}{Z_1 \sqrt{a_1}} \quad (34)$$

One can immediately verify the following limiting cases:

- 1) If  $R = 0$  then  $K_{1kf} = K_{1bf} = 1/H$ , and  $P_{1bf} = P_{1kf} = 0$  and  $\hat{x}_{1bf} = \hat{x}_{1kf} = s_1$ .
- 2) If  $Q^* = 0$  then  $K_{1kf} = K_{1bf} = 0$ , and  $P_{1bf} = P_{1kf} = 0$  and  $\hat{x}_{1bf} = \hat{x}_{1kf} = Ax_0 + B^*$ .
- 3) As  $B^* \rightarrow +\infty$  then  $P_{1bf} \rightarrow P_{1kf}$  and  $\hat{x}_{1bf} \rightarrow \hat{x}_{1kf}$ .

We can now derive the Bayesian filter for  $x_2$ . From Bayes' rule it follows that:

$$p(x_2|s_2, s_1) = \frac{p(s_2|x_2)p(x_2|s_1)}{W_2} \quad (35)$$

where the normalization constant  $W_2$  is defined as  $= \int p(s_2|x_2)p(x_2|s_1)dx_1$ . The predictive distribution of the state  $x_2$  at time 2 can be computed by the Chapman-Kolmogorov equation:

$$p(x_2|s_1) = \int p(x_2|x_1)p(x_1|s_1)dx_1 \quad (36)$$

We know that  $p(x_1|s_1)$  and  $p(x_2|x_1)$  are both truncated normal distributions. The sum of two truncated normal distributions is, however, no longer truncated normal. Since  $x_2 = Ax_1 + B^* + w_1^*$  with  $w_1^* \sim N(0, Q^*; -B^*)$  and since  $x_1 \in (Ax_0, +\infty)$ , it follows that  $x_2 \in (A^2x_0, +\infty)$ . More generally,  $x_t \in (A^t x_0, +\infty)$ .

Hence, while  $p(x_2|s_1)$  is still truncated, the distribution is neither normal nor truncated normal, but something in between. As a result analytical results are no longer available.

In general, the posterior distribution at time  $T$  is given by:

$$p(x_0, x_1, \dots, x_T|s_1, \dots, s_T) \propto [\prod_{j=1}^T p(s_j|x_j)] [p(x_0)\prod_{j=1}^T p(x_j|x_{j-1})] \quad (37)$$

Assuming  $A > 0$  (for  $A = 0$  the process has no memory, and the solution is as described by the above one-period model), it follows from the central limit theorem that the posterior distribution at  $T$  resembles more closely a normal distribution, and the Kalman filter will move closer towards the optimal Bayesian filter.

Panel A of Figure 1 shows that the distribution of the (standardized) latent variable converges to its long run stationary distribution within 5 time steps. The parameters used to generate panel A are  $B^* = 0$ ,  $Q^* = 1$ ,  $A = 0.7$  and  $R = 0.5$ . Since  $-B^*/\sqrt{Q^*} = 0$ , 50% of the distribution for the disturbances is truncated, which is the maximum possible amount of truncation. Yet, even for this extreme set of parameters, the stationary distribution (fine dotted line) is close to being standard normal (solid line).

Panel B of Figure 1 considers the long run stationary distribution of the latent variable for different parameter sets. The plot shows that the stationary distribution converges rapidly towards the normal distribution as we reduce the amount of truncation. E.g. for  $B^* = 6$  and  $Q^* = 4$ , the stationary distribution for the latent variable coincides with the normal distribution. Since the stationary distribution for the latent variable is close to being normal this means that the Kalman filter is (near) optimal. We verify in the last section of this appendix that the Kalman filter is indeed optimal.

## 4.2 Bayesian filter versus Kalman filter: a comparative analysis

In what follows we first numerically compare the one-period Kalman filter with the one-period Bayesian filter. In final section we show that the multi-period, stationary Kalman filter indeed approaches the optimal filter.

Figure 2 examines the accuracy of the linear one-period Kalman filter developed in the paper, with the accuracy of the non-linear one-period Bayesian filter developed in the previous section. Panels A and C plot the histogram of the relative error in percent, i.e.  $100 * (\hat{x}_1 - x_0)/x_0$ , for 10000 simulated values of  $x_1$  using the Kalman filter and Bayesian filter, respectively. The starting value,  $x_0$ , is the same for all simulations, and given by the long run stationary value  $x_0 = B/(1 - A)$ . Therefore, by multiplying the percent errors by  $x_0/100$ , one can also retrieve the errors in levels.

As for the truncation point, we consider the most extreme case by setting  $B^* = 0$ . Hence, since  $Q^* = 1$ ,  $B^*/\sqrt{Q^*} = 0$ , and therefore  $N(-B^*/\sqrt{Q^*}) = 0.5$ . This means that the distribution of the disturbances is truncated by 50%. Since  $B^*$  must be positive this is the highest level of truncation we can adopt.

Panels A and C show that both error distributions are centered around zero. In fact, the average relative error for the Kalman and Bayesian filter are, respectively, 2.73% and 2.67% with standard errors of respectively, 17.25% and 16.56%. Both filters have an average error that is statistically not different from zero.

The square root of the mean square error (RMSE) (i.e.  $\sqrt{\sum_i^n (\hat{x}_i - x_i)^2}$ ) is 0.461 and 0.449 for the Kalman and Bayesian filters, respectively. This means that the  $RMSE_{kf}/RMSE_{bf} = 1.027$ , indicating that, as to be expected, the KF does not minimize the RMSE when truncation

is severe. However, the performance of the KF, compared to the optimal Bayesian filter, is still good despite the high degree of truncation. The average difference in percent between the Kalman filter estimate and the Bayesian estimate (i.e.  $100 * (\hat{x}_{1kf} - \hat{x}_{1bf}) / \hat{x}_{1bf}$ ) is 0.03% and the standard deviation is 3.94%.

Panels B, D and F plot a histogram of the Kalman estimates ( $\hat{x}_{1kf}$ ), Bayesian estimates ( $\hat{x}_{1bf}$ ) and actual values ( $x_1$ ), respectively. The distribution of the actual values is truncated at 1.86 ( $Ax_0$ ). The posterior distribution of the Bayesian estimates (panel D) successfully enforces this truncation, as no estimates are below 1.86. The distribution of the Kalman estimates (panel B), however, is normally distributed, and has a small tail of values below 1.86, causing the Kalman filter to be inaccurate when  $x_t$  approaches the truncation point. This point is illustrated in panel E which gives a scatter plot of the Kalman filter estimates against the measurements. The thin straight line represents the 45 degree line through the origin. As to be expected, the slope of the scatter plot is less than 45 degrees, reflecting the fact that the Kalman filter estimate only partially responds to measurements due to the noise inherent in the measurements. The Kalman filter is a linear function of the measurements causing the estimates to drop below the truncation point for sufficiently low measurements. In contrast, the Bayesian filter (panel G) is highly non-linear in the region of the truncation point and manages to keep all estimates above the truncation point, now matter how small the measurements. Finally panel H plots the Kalman estimates (Y-axis) against the Bayesian estimates (X-axis). Both estimates are roughly identical (and coincide with the 45 degree straight line through the origin) for average values of the latent variable. When the latent variable approaches the truncation point for values just above 1.86, the Kalman filter increasingly underestimates the Bayesian filter. The KF estimates are also below the Bayesian estimates for values in the right tail of the distribution, but the relative errors are much lower than for the left tail.

We conclude that when truncation is severe, the one-period KF estimates are inaccurate when the latent variable approaches the truncation point. We should, however, stress that we are considering a worst case scenario by focussing both on the one-period model and on high degrees of truncation. Errors are much smaller for the stationary (i.e. multi-period) distribution of the latent variable, or when the disturbances to the latent variable are subject to less truncation.

While figure 2 focuses on a specific set of parameters, table 1 tries to put things in a broader context by comparing the two filters for a wider range of parameter values. The key determinants of the accuracy of the Kalman filter are the degree of truncation ( $-B^*/\sqrt{Q^*}$ ) and the autoregression coefficient ( $A$ ). We fix  $Q^*$  at 1. We vary  $B^*$  from 10 to zero, corresponding to, respectively, no and 50% truncation of the disturbance distribution. We set  $x_0 = B/(1-A)$  and keep the standard deviation of the measurement error at a constant 5% of the expected a priori value for  $x_1$  (i.e.  $R = (0.05 * (A * x_0 + B))^2$ ).

Table 1 shows that the Bayesian filter and Kalman filter are pretty much identical in terms of minimizing the mean square error. The last column gives the ratio of the root mean square errors ( $RMSE_{kf}/RMSE_{bf}$ ), and shows that the ratio is everywhere between 1.00 and 1.03. The Bayesian Filter, as to be expected, always has the lower RMSE. Finally, note that the standard deviation of the relative error become very large for both filters as the autoregression coefficient,  $A$ , and the truncation point  $B^*$  both go to zero because in those cases the values for  $x_t$  tend to be clustered at tiny values just above zero. Furthermore, when both  $B^* = 0$  and  $A = 0$ , the latent variable only contains pure noise and no longer has a predictable component.

We conclude from table 1 that for practical purposes (i.e.  $A > 0.5$  and  $B^*/\sqrt{Q^*} > 1$ ) the one-period Kalman filter is identical to the optimal non-linear Bayesian filter.<sup>5</sup> Only for very low levels of income persistence (i.e. low values for  $A$ ), combined with high truncation levels are the filters different. In those cases the Kalman filter has a RMSE that is between 1% and 3% higher than the RMSE of the Bayesian filter.

### 4.3 On the optimality of the steady-state Kalman filter

So far we have focussed on the accuracy of the one-period model. What about the multi-period, infinite horizon model? While  $x_1$  has a truncated normal distribution,  $x_2, x_3, \dots$  are in general no longer truncated normal because the sum of truncated normal disturbances is not truncated normal but something in between a truncated normal and a normal distribution

---

<sup>5</sup>In our model  $A$  determines the degree of persistence in both the latent variable  $X_t$  and earnings  $\pi_t$  (since earnings are proportional to  $x_t$ ). The empirical workhorse model for earnings and cash flows is the first-order autoregressive process (see Frankel and Litov (2009), among others). Frankel and Litov (2009) report values for the autoregressive coefficient ranging from 0.5 to 0.81 (and 0.51 to 0.85) for earnings (cash-flows), suggesting that earnings are highly persistent.

(see panel A, figure 1). As shown before the truncation point  $A^t x_0$  converges to zero as  $t \rightarrow \infty$ , and the stationary distribution for  $x_t$  resembles fairly closely a normal distribution (see panel B in figure 1). As a result the steady state Kalman Filter is (near) optimal. This can be shown by applying standard diagnostic tests to the Kalman filter innovations. A key result from the literature (going back to Kailath (1968) and Mehra (1970), see also Durbin and Koopman (2012) and Simon (2006)) states that the Kalman filter is optimal if and only if the innovations  $innov_t^- \equiv s_t - H(A\hat{x}_{t-1} + B)$  are white noise.

To check the optimality of the Kalman Filter in our paper, we simulate  $x_t$  and  $s_t$  over 200 periods, and then calculate the corresponding Kalman filter estimates. We set  $A = 0.7$ , which represents a typical level of persistence in corporate cash flows and earnings (see Frankel and Litov (2009)). With  $B^* = 0.5$  and  $Q^* = 1$  we enforce a high, but still reasonable level of truncation (marginal costs become infinitely high as  $x_t \rightarrow 0$ , so we want to set  $B^*$  high enough above 0 to ensure marginal costs remain bounded, or in other words, we do not want to truncate the distribution of the disturbances too much). Finally,  $R = 0.5$ .

Panel A of figure 3 plots the actual values ( $x_t$ ) and the prior estimates ( $\hat{x}_t^-$ ) for the first 32 observations, while panel C plots the actual values ( $x_t$ ), the posterior Kalman estimates ( $\hat{x}_t$ ) and the measurements  $s_t$ . Panel B plots the prior and posterior standard errors of the Kalman filter estimates. Note that the standard errors converge to a constant level within 4 time steps, which confirms the usefulness of the stationary Kalman filter as the workhorse for our theoretical analysis.

Panel D plots the actual values ( $x_t$ ) against the Kalman estimates ( $\hat{x}_t$ ). The fit is very good and reveals no bias. Indeed, a simple regression  $x_t = \alpha + \beta\hat{x}_t + \zeta_t$  reveals that  $R^2 = 0.764$ , with parameter estimates  $\hat{\alpha} = -0.26$  (standard error = 0.157)  $\hat{\beta} = 1.09$  (standard error = 0.04).

To check whether the innovations  $innov_t \equiv s_t - H\hat{x}_t^-$  (where  $\hat{x}_t^- \equiv A\hat{x}_{t-1} + B$ ) are white noise, we first execute the following preliminary regression:  $s_t = \alpha + \beta\hat{x}_t^- + \zeta_t$ . We find that  $\hat{\alpha} = 0.3$  (SE= 0.484) and  $\hat{\beta} = 0.945$  (SE = 0.136), with a Durbin Watson statistic equal to 2.01. Hence,  $\hat{\alpha}$  is insignificant,  $\hat{\beta}$  is not significantly different from 1 and there is no first-order serial correlation in the innovations.

Figure 4 plots the cumulative distribution of the standardized Kalman Filter innovations.

Panel A presents the benchmark case of no truncation. As expected, the cumulative distribution of the standardized innovations (solid line) coincides fairly closely with the standard normal distribution, indicating that the innovations are Gaussian white noise.

Panel B presents the case with truncation. Even though no longer perfect, the fit with the cumulative standard normal distribution is still very good, suggesting that the innovations of the Kalman filter are still close to being Gaussian.

As a final, and more definite test, we perform Bartlett's cumulative periodogram white-noise test (see figure 5 for a visual representation). The test shows that the probability of exceeding the test statistic is 0.968, and therefore we decisively accept the null hypothesis that the innovations are white noise. Table 2 gives the autocorrelations for up to 5 lags. In all cases we comfortably accept the null hypothesis that there is no serial correlation in the Kalman filter innovations. The small autocorrelations confirm that the innovations are white noise. As a result, we cannot reject the null hypothesis that the Kalman filter in our paper is optimal.

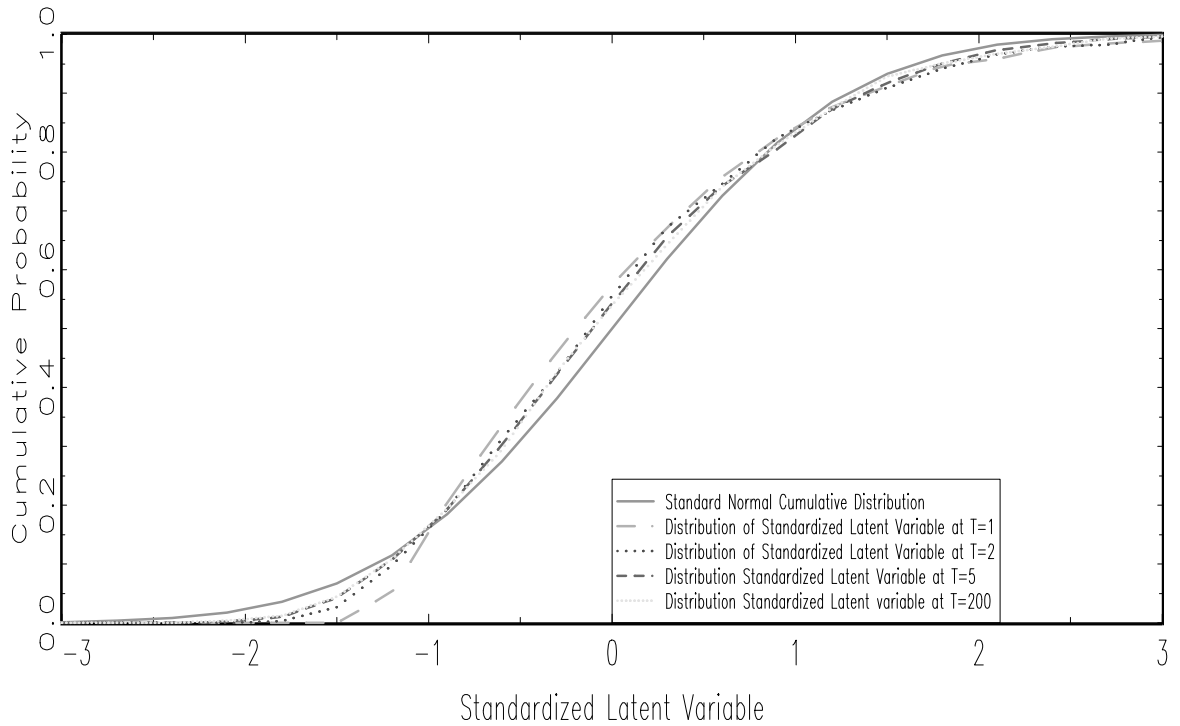
## References

- BENMELECH, E., E. KANDEL, AND P. VERONESI (2010): “Stock-Based Compensation and CEO (Dis)Incentives,” *The Quarterly Journal of Economics*, 125(4), 1769–1820.
- DURBIN, J., AND S. J. KOOPMAN (2012): *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, United Kingdom, second edition edn.
- FRANKEL, R., AND L. LITOV (2009): “Earnings persistence,” *Journal of Accounting and Economics*, 47(1-2), 182–190.
- KAILATH, T. (1968): “An Innovations Approach to least Squares Estimation Part I: Linear Filtering in Additive Noise,” *IEEE Transactions on Automatic Control*, 13, 646–654.
- MEHRA, R. (1970): “On the Identification of Variances and Adaptive Kalman Filtering,” *IEEE Transactions on Automatic Control*, 15, 175–184.
- SIMON, D. (2006): *Optimal State Estimation: Kalman,  $H_\infty$ , and Nonlinear Approaches*. John Wiley & Sons, Hoboken, New Jersey.

Figure 1: Distribution of the Latent Variable

The figure examines the distribution of the latent variable. Panel A plots the distribution of the standardized latent variable at various points in time ( $T=1, 2, 5$  and  $200$ ), and compares this distribution with the standard normal distribution. The parameter values used to generate panel A in this figure are:  $A = 0.7$ ,  $B^* = 0$ ,  $Q^* = 1$ ,  $R = 0.5$  and  $H = 1$ . Setting the truncation point at zero for the disturbances of the latent variable implies that 50% of the distribution is truncated. Panel B plots the stationary distribution for the standardized latent variable for various parameter combinations, reflecting various degrees of truncation. The base parameters used are  $A = 0.7$ ,  $R = 0.5$  and  $H = 1$ .

Cumulative Distribution of standardized Latent Variable (with truncation)



Cumulative Distribution of standardized Latent Variable (with truncation)

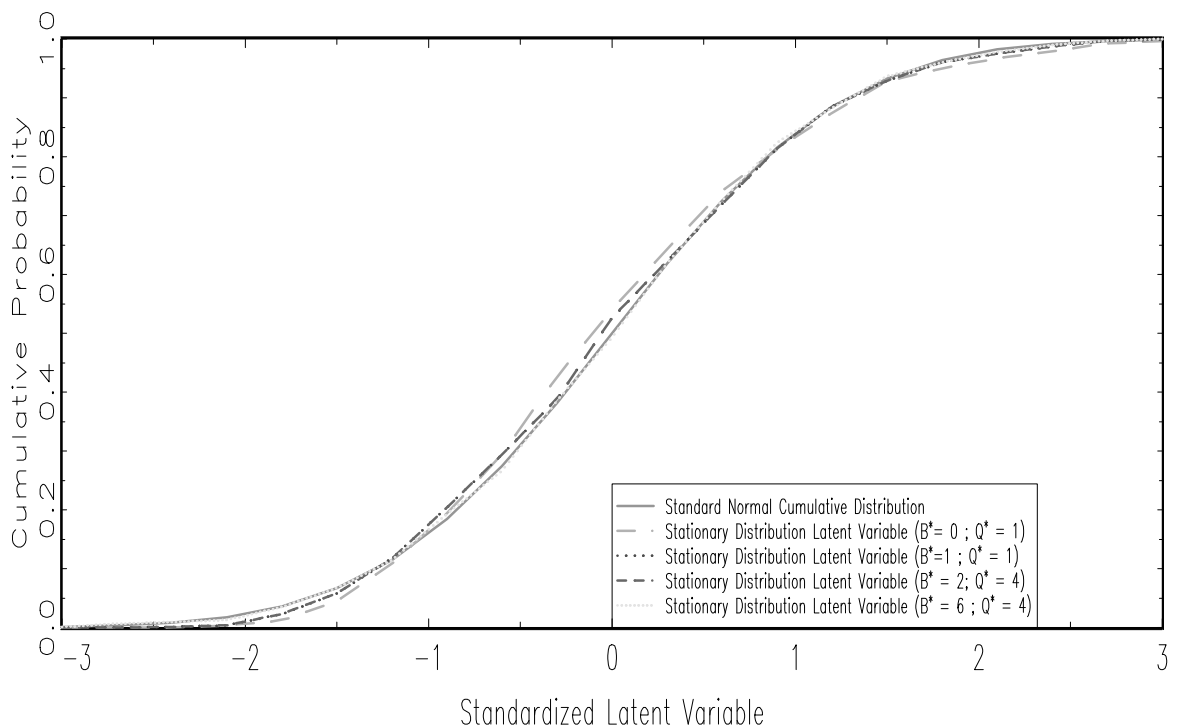


Figure 2: Kalman filter versus optimal Bayesian filter

The figure gives a histogram of the relative error in percent ( $100 * (\hat{x}_1 - x_1)/x_1$ ) for the one-period linear Kalman Filter (panel A) and the one period non-linear Bayesian filter (Panel C) for a sample of 10000 observations. Panel B, D and F plot histograms of, respectively, the Kalman filter estimates, the Bayesian Filter estimates, and the actual values for the latent variable. Panel E and G plot, respectively, the Kalman and Bayesian filter estimates against the measurements; the thin straight line is the 45 degree line through the origin. Panel H plots the Kalman filter estimate against the Bayesian filter estimate; the thin straight line is the 45 degree line through the origin. The parameter values used to generate the panels in this figure are:  $A = 0.7$ ,  $B^* = 0$ ,  $Q^* = 1$ ,  $R = 0.5$  and  $H = 1$ .

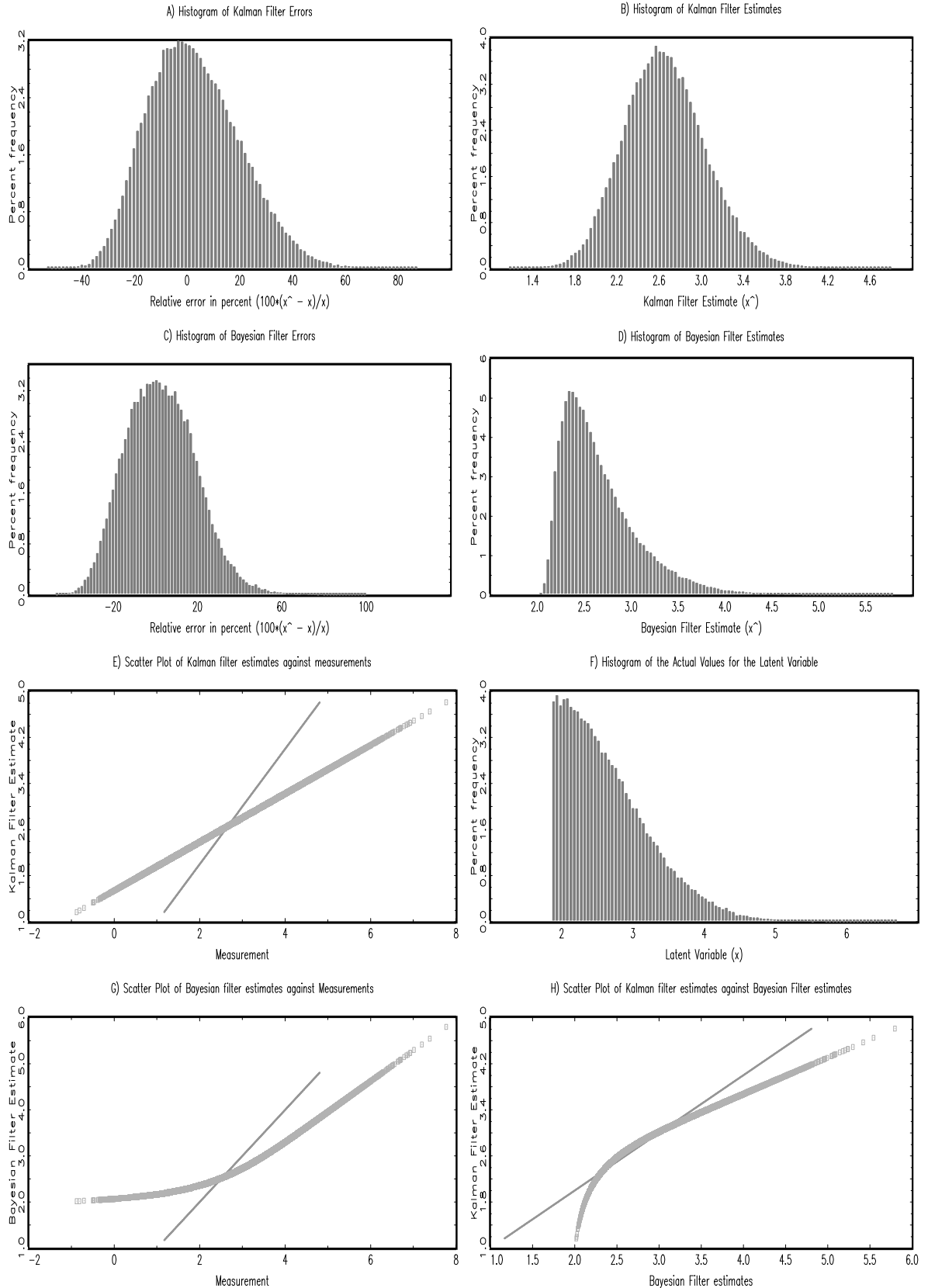
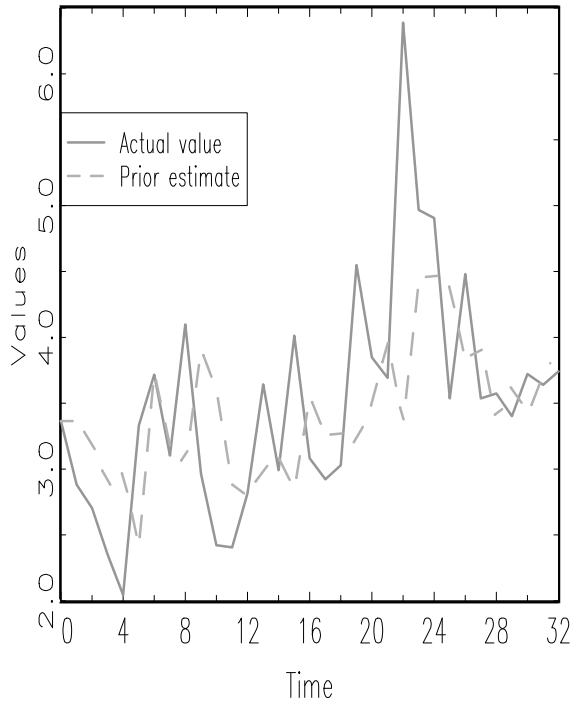


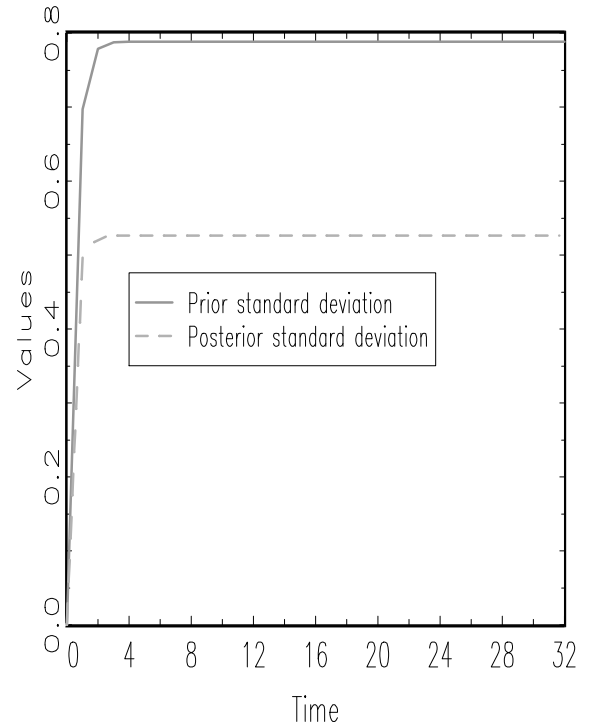
Figure 3: Kalman filter estimates and actual values

Panel A simulates the actual value of the latent variable  $x_t$  (solid line) and plots the prior Kalman filter estimate  $\hat{x}^-$  (dashed line). Panel C plots the actual value,  $x_t$  (solid line), the posterior Kalman filter estimate,  $\hat{x}_t$  (dashed line), and the measurement  $s_t$  (dotted line). Panel B plots prior and posterior standard deviation of the Kalman filter estimate. Panel D gives a scatter plot of the actual values  $x_t$  against the Kalman filter estimates,  $\hat{x}_t$ . The parameter values used to generate the panels in this figure are:  $A = 0.7$ ,  $B^* = 0.5$ ,  $Q^* = 1$ ,  $R = 0.5$ ,  $H = 1$ ,  $x_0 = B/(1 - A)$  and  $P_0 = 0$ . Setting the truncation point at  $B^* = 0.5$  for the disturbances of the latent variable implies that  $N(-0.5/1) = 0.31$  or 31% of the distribution is truncated.

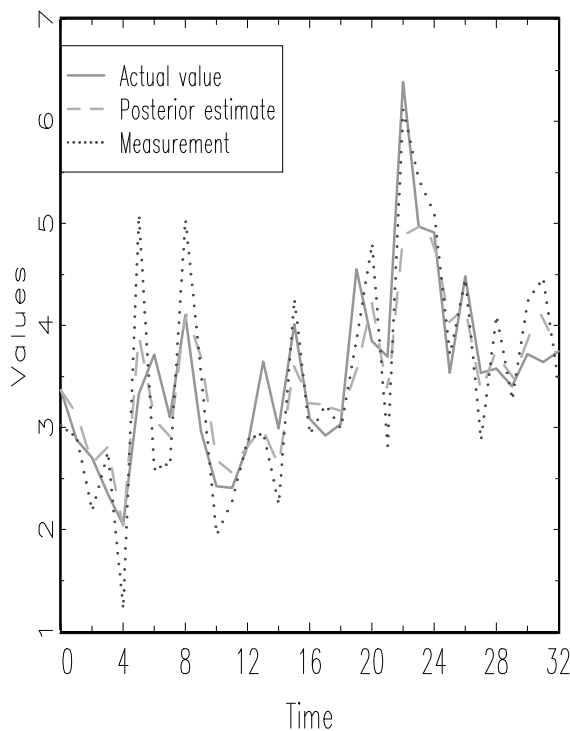
A) Actual value and Prior KF Estimates



B) Prior and Posterior standard deviation



C) Actual value and Posterior KF Estimates



D) Actual values against Posterior KF estimates

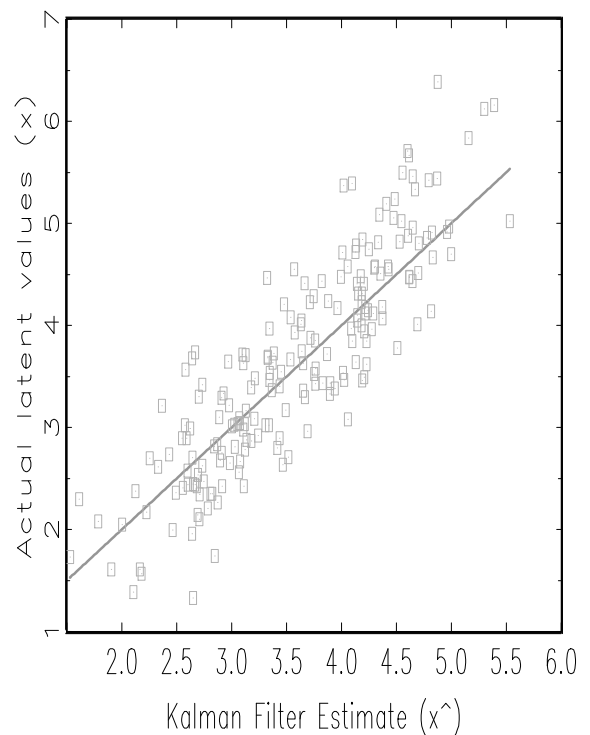
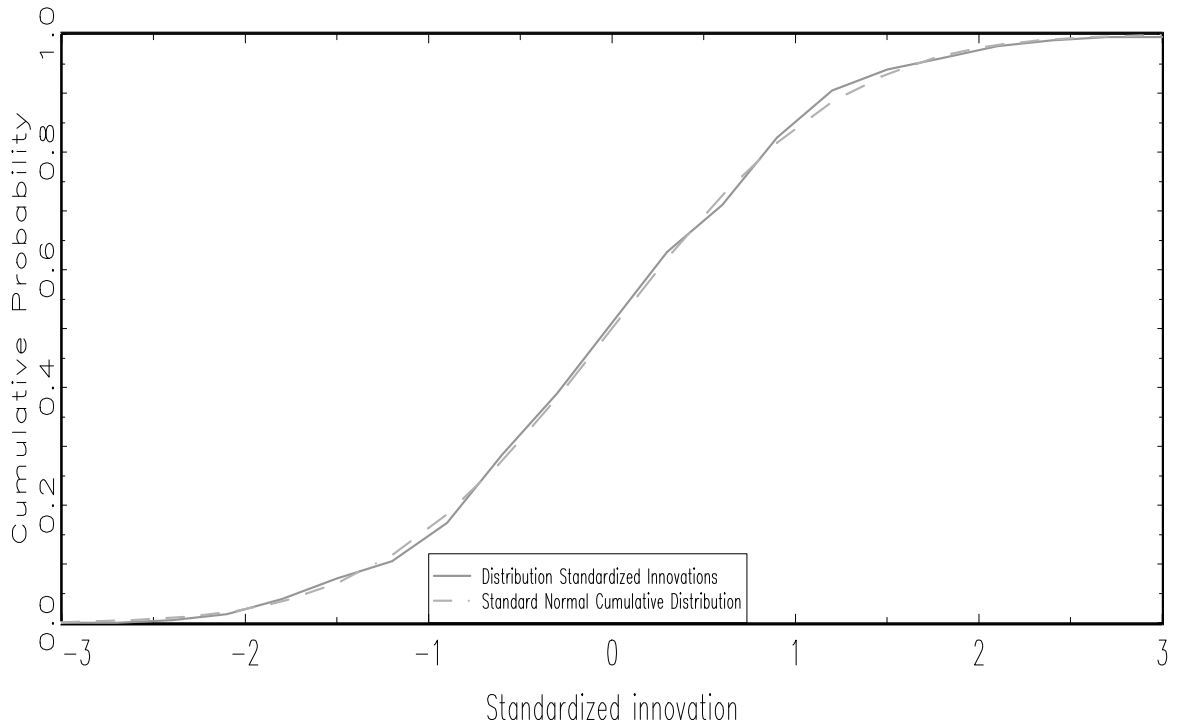


Figure 4: Cumulative distribution of standardized Kalman filter innovations.

The solid line in Panel A plots the distribution of the standardized Kalman filter innovations (for a time series of 200 observations for  $x_t$ ) assuming there is no truncation ( $B^* = +\infty$ ). The dashed line plots the standard normal cumulative distribution. The solid line in Panel B plots the distribution of the standardized Kalman filter innovations assuming the distribution of the disturbances to the latent variable is truncated (i.e.  $B^* = 0.5$ ). The dashed line plots the standard normal cumulative distribution. The other parameter values used to generate the panels in this figure are:  $A = 0.7$ ,  $Q^* = 1$ ,  $R = 0.5$ ,  $H = 1$ ,  $x_0 = B/(1 - A)$  and  $P_0 = 0$ . Setting the truncation point at  $B^* = 0.5$  for the disturbances of the latent variable implies that  $N(-0.5/1) = 0.31$  or 31% of the distribution is truncated.

Cumulative Distribution of standardized KF innovations (no truncation)



Cumulative Distribution of standardized KF innovations (with truncation)

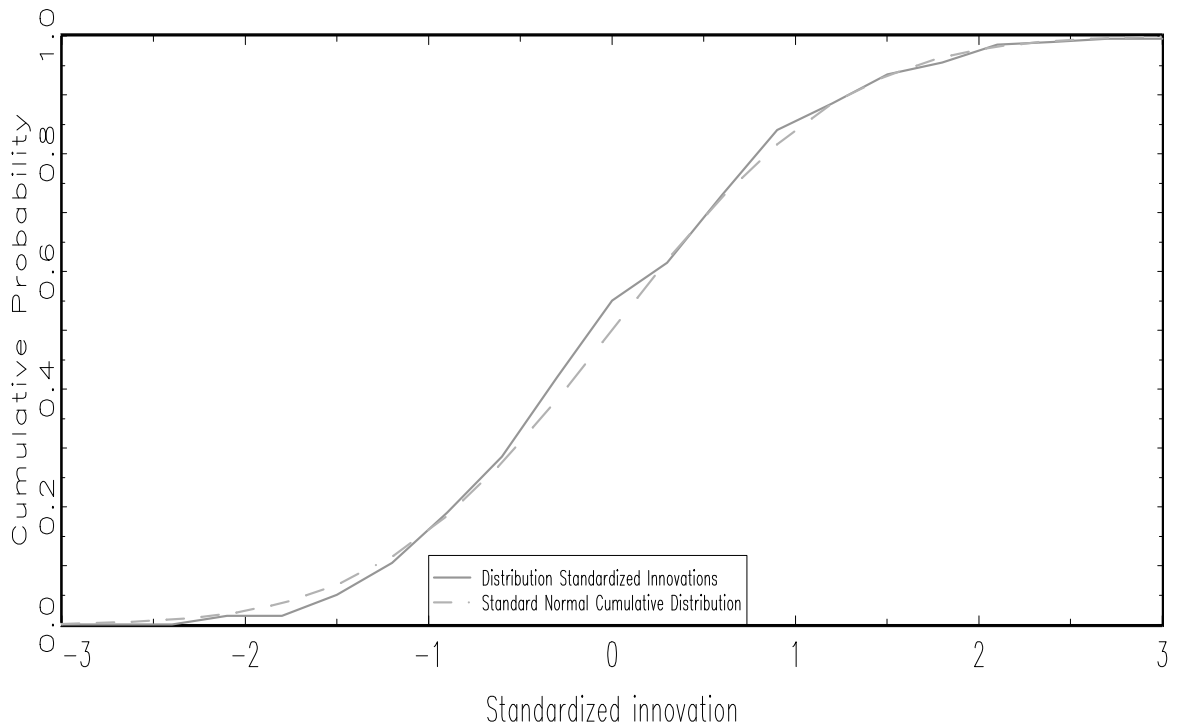


Figure 5: Cumulative Periodogram White-Noise test

The figure plots the cumulative periodogram of the Kalman filter innovations (for a time series of 200 observations) assuming the distribution of the disturbances to the latent variable is truncated (i.e.  $B^* = 0.5$ ). The two parallel lines on either side of the observations constitute a 95% confidence interval. The other parameter values used to generate the time series of Kalman filter innovations in this figure are:  $A = 0.7$ ,  $Q^* = 1$ ,  $R = 0.5$ ,  $H = 1$ ,  $x_0 = B/(1 - A)$  and  $P_0 = 0$ . Setting the truncation point at  $B^* = 0.5$  for the disturbances of the latent variable implies that  $N(-0.5/1) = 0.31$  or 31% of the distribution is truncated.

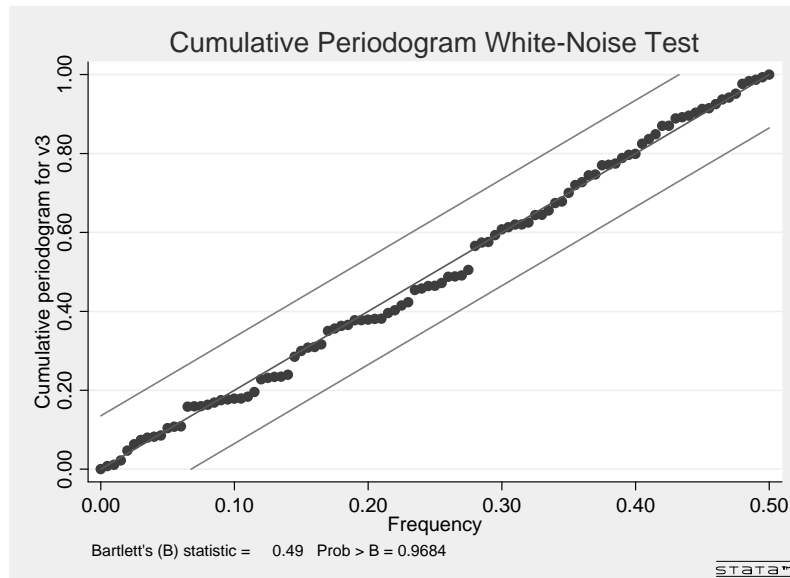


Table 1: Kalman filter versus Bayesian Filter: a comparative analysis

The table examines the errors of the Kalman Filter (kf) and Bayesian Filter (bf), i.e.  $e_j = 100 * (\hat{x}_{1j} - x_t)/x_t$  for  $j = kf$  or  $j = bf$ .  $\bar{e}_j$  denotes the average relative error.  $RMSE_j \equiv \sqrt{\sum_i^n (x_{1i} - x_t)^2}$  for  $j = kf, bf$ . The parameter values used to generate the table are  $Q^* = 1$ ,  $H = 1$  and  $R = [0.05 * (A * B / (1 - A) + B)]^2$ .

$A$	$B^*$	$\bar{e}_{kf}$	$\bar{e}_{bf}$	$stdev(e_{kf})$	$stdev(e_{bf})$	$m_1 \equiv RMSE_{kf}$	$m_2 \equiv RMSE_{bf}$	$m_1/m_2$
0.99	10.00	0.00	0.00	0.10	0.10	1.00	1.00	1.00
0.99	1.00	0.01	0.01	0.61	0.61	0.79	0.79	1.00
0.99	0.50	0.01	0.01	0.68	0.68	0.69	0.69	1.00
0.99	0.00	0.01	0.01	0.74	0.74	0.60	0.60	1.00
0.50	10.00	0.15	0.15	3.59	3.59	0.71	0.71	1.00
0.50	1.00	0.27	0.31	5.73	5.52	0.13	0.13	1.01
0.50	0.50	0.26	0.32	5.88	5.59	0.10	0.10	1.02
0.50	0.00	0.24	0.31	5.99	5.65	0.08	0.08	1.02
0.25	10.00	0.20	0.20	4.25	4.25	0.56	0.56	1.00
0.25	1.00	0.34	0.49	7.41	7.04	0.09	0.08	1.01
0.25	0.50	0.34	0.53	7.84	7.34	0.07	0.07	1.01
0.25	0.00	0.32	0.55	8.16	7.58	0.05	0.05	1.02
0.00	10.00	0.23	0.23	4.61	4.61	0.45	0.45	1.00
0.00	1.00	5.13	14.35	532.69	636.31	0.06	0.06	1.01
0.00	0.50	6.74	37.23	4094.83	6742.08	0.05	0.05	1.01
0.00	0.00	6.57	24.80	732.37	833.43	0.04	0.04	1.01

Table 2: Autocorrelations and Partial Autocorrelations of the Kalman filter innovations

The table presents the autocorrelations (AC) and partial autocorrelations (PAC) for the Kalman filter innovations. Test statistics (Q) for the null hypothesis of zero autocorrelation and the probability of exceeding the test statistic (Prob > Q) are provided. The parameter values used to generate the Kalman filter innovations are  $Q^* = 1$ ,  $H = 1$ ,  $B^* = 0.5$ ,  $A = 0.7$  and  $R = 0.5$ . This implies that about 31% of the distribution for the disturbances to the latent variable is truncated. The number of observations generated in the time series is 200.

Lag	AC	PAC	Q	Prob > Q
1	-0.026	-0.026	0.139	0.71
2	-0.037	-0.0383	0.421	0.81
3	0.0459	0.0441	0.853	0.84
4	-0.030	-0.031	1.04	0.90
5	-0.0127	-0.012	1.075	0.96