Department of Economics

Working Paper No. 0401

http://nt2.fas.nus.edu.sg/ecs/pub/wp/wp0401.pdf

On the Estimation of Cost of Capital and its Reliability

Wing-Keung Wong
Department of Economics
National University of Singapore

Raymond H. Chan
Department of Mathematics
Chinese University of Hong Kong

March 15, 2004

Abstract: Gordon and Shapiro (1956) first equated the price of a share with the present value of future dividends and derived the well-known relationship. Since then, there have been many improvements on the theory. For example, Thompson (1985, 1987) combined the "dividend yield plus growth" method with Box-Jenkins time series analysis of past dividend experience to estimate the cost of capital and its "reliability" for individual firms. Thompson and Wong (1991, 1996) proved the existence and uniqueness of the cost of capital and provided formula to estimate both the cost of capital and its reliability. However, their approaches cannot be used if the "reliability" does not exist or if there are multiple solutions for the "reliability". In this paper, we extend their theory by proving the existence and uniqueness of this reliability. In addition, we propose the estimators for the reliability and prove that the estimators converge to a true parameter. The estimation approach is further simplified, hence rendering computation easier. In addition, the properties of the cost of capital and its reliability will be analyzed with illustrations of several commonly used Box-Jenkins models.

 The first author would also like to thank Robert B. Miller and Howard E. Thompson for their continuous guidance and encouragement. The research is partially supported by the grants from the National University of Singapore and from the Hong Kong Research Grant Council.

1 Introduction

Assuming constant rate for the future dividends and income, Gordon and Shapiro (1956) first equated the price of a share with the present value of future dividends and derived the venerable and durable "dividend yield plus growth" method for estimating the cost of capital. Since the cost of capital plays a prominent role in setting rates that customers pay, estimating the "dividend yield plus growth" method is therefore an important element in rate cases for regulated firms.

Miller and Modigliani (1966), Litzenberger and Rao (1971), McDonald (1971), Higgens (1974) and Thompson (1979) have all used a variant of the "dividend yield plus growth" method to estimate the cost of equity capital for a cross section of electric utilities. Makhija and Thompson (1984) have compared the various cross-sectional models using this method regards to their efficiency as a tool for rate cases. Thompson (1984) used the same technique along with cross-sectional data to estimate the cost of capital for individual utilities, but measures of reliability of the estimates were not obtained. To cope with the "reliability" question, Thompson (1985, 1987) combined the "dividend yield plus growth" method with Box-Jenkins time series analysis of past dividend experience to estimate the cost of capital and its "reliability" for individual firms. His approach has the desirable feature of relaxing the constant growth rate assumption which had served as the basis for all the preceding models.

The credibility of cost of capital estimates from statistical forecasts using time series methodology have also been examined by Thompson and Wong (1991). Their analysis raises the question of whether the estimation procedure developed by Thompson (1987) would always produce finite estimates of dividends and ultimately the cost of capital. Moreover there is the question of simplification of Thompson's estimation approach. In his approach, the cost of capital is solved from a non-linear equation which is in terms of past dividend realizations, the parameters of the Box-Jenkins model as well as the covariance matrix of the parameters. Thus a change in the Box-Jenkins model will result in a change of the form of the non-linear equation for solving the cost of capital estimates. This makes the estimation procedure complicated. In fact, since the reliability relies on the parameters of the Box-Jenkins model, it makes the estimation "model-dependent" and the computation difficult.

To resolve this issue, Thompson and Wong (1996) proved the existence and uniqueness of the cost of capital and provided formulae to estimate both the cost of capital and its reliability. In their approach, the equation to solve for the cost of capital is only in terms of forecasted future dividends while the reliability is only in terms of forecasted future dividends and their covariance matrix. The parameters of the Box-Jenkins model and the covariance matrix of those parameters are no longer needed in the development of a measure of "reliability". Thus their approach to estimating the cost of capital and its "reliability" is "model free" – The same program can be used for any Box-Jenkins model or any time series model so long as the covariance of future dividends forecasts can be estimated. However, their approaches cannot be used if the "reliability" does not exist or if there are multiple solutions for the "reliability". This paper extends their theory by proving the existence and uniqueness of the reliability. This enables their approach to be carried out in practice.

Conceptually the formulae for estimating both the cost of capital and its reliability are in terms of infinite sums and infinite-dimensional matrices for the estimate and its reliability. Computation in this case is impossible. Thompson and Wong (1996) developed the formula for the estimators in terms of finite sums only such that computation can be carried out. However the proposed estimators for the "reliability" did not provide evidence that the estimators converge to the true parameter. Thus this paper will propose another set of estimators for the reliability and will also prove that the estimators converge to the true parameter. The estimation approach is further simplified, hence rendering computation easier. In addition, the properties of the cost of capital and its reliability will be analyzed with illustrations of several commonly used Box-Jenkins models.

The next section will state the theory of the cost of capital and the condition for the existence and uniqueness of the reliability of the cost of capital. Section 3 investigates the validity of the conditions made in Section 2 by examining three typical ARIMA models and the situation for general models. Section 4 includes a study on the estimation procedure of finding the cost of equity capital and its reliability. The paper concludes with a discussion on the applicability of the procedure.

2 The Theory

Assume that the dividends are issued m times a year and the expected dividend, discount rate, cost of capital and stock price at time t are denoted by d_t, r_t, ρ_t and P_t respectively. The discount rate, r_t, at time t is defined such that the price of a share is equal to the present value of the expected future dividends (Gordon and Shapiro 1956, Thompson 1985, and Thompson and Wong 1991, 1996):

$$P_t = \sum_{i=1}^{\infty} \frac{d_{t+i}}{(1+r_t)^i} \tag{1}$$

and the cost of capital, ρ_t, at time t is defined as:

$$\rho_t = F(r_t) = (1+r_t)^m - 1 \,. \tag{2}$$

Let

$$\mathbf{d}_t = (d_{t+1}, d_{t+2}, \ldots, d_{t+i}, \ldots) \tag{3}$$

and consider the set $\mathbf{S}_t$ of the collection of $\mathbf{d}_t$ satisfying the following assumptions:

Assumption 1: *There exists a positive number K such that $\sum_{i=1}^{\infty} d_{t+i} > K$.*

Assumption 2: *There exists a number $r > -1$ such that*

$$\sum_{i=1}^{\infty} \frac{d_{t+i}}{(1+r)^i} < \infty \,.$$

Assumption 3: *The series of dividends per share follows a time series model such that the expected future dividends can be forecasted.*

As the dividends are non-negative, Assumptions 1 and 2 together imply that for any positive value of stock price P_t, there exists a number $r_0 > -1$ such that

$$P_t < \sum_{i=1}^{\infty} \frac{d_{t+i}}{(1+r_0)^i} < \infty. \tag{4}$$

Let $f_t : \mathbf{S}_t \times (r_0, \infty) \longrightarrow \mathsf{R}$ be defined by

$$f_t(\mathbf{d}_t, r) = \sum_{i=1}^{\infty} \frac{d_{t+i}}{(1+r)^i} - P_t \tag{5}$$

where R is the set of real numbers.

For each non-negative $\mathbf{d}_t$ and for each n such that $\sum_{i=1}^{n} d_{t+i}$ is positive, one can easily show that there exists a unique function $g_{t,n}$ and a variable $r_{t,n}$ such that

$$f_{t,n}(\mathbf{d}_t, r_{t,n}) = 0 \qquad \text{and} \qquad r_{t,n} = g_{t,n}(\mathbf{d}_t) \qquad \text{for any} \quad \mathbf{d}_t \in \mathbf{S}_t\,. \tag{6}$$

and show that there exists a function $g_{t,n}$ which is continuously differentiable with respect to $d_{t+1}, d_{t+2}, \cdots, d_{t+n}$ and its partial derivative of $g_{t,n}$ is

$$\frac{\partial g_{t,n}(\mathbf{d}_t)}{\partial d_{t+i}} = \left[(1 + g_{t,n}(\mathbf{d}_t))^i \sum_{j=1}^{n} \frac{j d_{t+j}}{(1 + g_{t,n}(\mathbf{d}_t))^{j+1}} \right]^{-1}$$

for each $i \leq n$ and the derivative is equal to zero for $i > n$.

From the theory of equations we know that the number of positive roots of a polynomial is related to the number of sign changes of its coefficients. One can apply this idea to determine that if there is only one change in sign you could be assured of only a single root. This method can also be used to determine the uniqueness of the solution for $f_{t,n}$ because all d_{t+i} are non-negative.

However, the uniqueness of the solution for f_t cannot be proved directly from the results of the uniqueness of the solution for $f_{t,n}$ because f_t is defined in an infinite-dimensional space but $f_{t,n}$ is not. It is well-known that a function is continuous or differentiable in a finite-dimensional space may not be continuous or differentiable in the infinite-dimensional space. A function which has solution in any finite-dimensional space may not even obtain a solution in the infinite-dimensional space. The estimation of the cost of capital and its reliability requires the existence and uniqueness of the solution for f_t. It also requires the condition of continuity and differentiablity of f_t. The following lemma shows that there exists a variable $r_t > r_0$ such that $f_t(\mathbf{d}_t, r_t) = 0$.

Lemma 1 *For the function f_t defined in (5), if $\mathbf{d}_t$ satisfies Assumptions 1 and 2, then*

1. *f_t is continuously differentiable,*
2. *there exists a unique continuously differentiable function g_t and a variable $r_t > r_0$ such that*
$$f_t(\mathbf{d}_t, r_t) = 0 \qquad and \qquad r_t = g_t(\mathbf{d}_t) \tag{7}$$
where r_0 is defined in (4), and
3. $$\frac{\partial g_{t,n}(\mathbf{d}_t)}{\partial(\mathbf{d}_t)} \partial d_{t+i} \longrightarrow \frac{\partial g_t(\mathbf{d}_t)}{\partial d_{t+i}} \qquad as \quad n \to \infty \tag{8}$$
where
$$\frac{\partial g_t(\mathbf{d}_t)}{\partial d_{t+i}} = [(1 + g_t(\mathbf{d}_t))^i \sum_{j=1}^{\infty} \frac{j d_{t+j}}{(1 + g_t(\mathbf{d}_t))^{j+1}}]^{-1}.$$

The proof is in Appendix A1.

The estimation of the cost of capital and its reliability requires the existence and uniqueness of the solution for $f_t = 0$. It also requires the condition of continuity and differentibility of f_t. Lemma 1 proves that the solution r_t exists and is unique. The conditions of continuity and differentibility of f_t were also stipulated. Once the estimate of r_t is obtained, Equation (2) can be applied to obtain the estimate of ρ_t.

Nevertheless, the estimation cannot be obtained if the reliability does not exist or if there are multiple solutions for the reliability. This paper seeks to substantiate the existence and uniqueness of the reliability; which guarantees the estimation is possible. To do this, the following assumption is introduced:

Assumption 4: *The covariance matrix of the forecast errors*[1]

$$\Sigma_t = E[\ (\hat{\mathbf{d}}_t - \mathbf{d}_t)(\hat{\mathbf{d}}_t - \mathbf{d}_t)'\ ] = (\sigma_{ij})$$

can be estimated and there exist constants M and k such that for all $\{(i, j)\}$

$$|\sigma_{ij}| < M\, k^{i+j} \qquad and \qquad k < 1 + \hat{r}_t \tag{9}$$

[1] For simplicity, we omit the subscript t in σ_{ij}.

except for a finite set of $\{(i,j)\}$ *where* $\mathbf{d}_t$ *is defined in (3),* $\hat{\mathbf{d}}_t$ *is the estimate of* $\mathbf{d}_t$ *and* $\hat{r}_t$ *satisfies*

$$f_t(\hat{\mathbf{d}}_t, \hat{r}_t) = 0 \tag{10}$$

with f_t *defined in (5).*

We then extend the theory of the cost of capital by proving the existence and uniqueness of the reliability as stated in the following theorem:

Theorem 1 *Suppose that a sequence of dividends* $\{d_\iota\}$ *issued* m *times a year satisfying Assumptions 1 to 4 is observed from* $\iota = 1$ *to* $\iota = t$. *The discount rate* r_t *and the cost of capital* ρ_t *are defined in Equations (1) and (2) respectively. Let the function* f_t *be defined as in Equation (5). For any* $\mathbf{d}_t = (d_{t+1}, d_{t+2}, \ldots, d_{t+i}, \ldots)$ *and any positive price* P_t, *we have*

1. *for the estimator* $\hat{r}_t$ *of* r_t *satisfying (10), there exists a unique solution for its variance* σ^2_{rt} *satisfying*

$$\sigma^2_{rt} = \left[\sum_{i=1}^{\infty} \frac{i\tilde{d}_{t+i}}{(1+\tilde{r}_t)^{i+1}}\right]^{-2} \tilde{\mathbf{a}}'_t \Sigma_t \tilde{\mathbf{a}}_t, \tag{11}$$

   where $\tilde{\mathbf{d}}_t = (\tilde{d}_{t+1}, \tilde{d}_{t+2}, \cdots, \tilde{d}_{t+i}, \cdots)$ *lies between* $\mathbf{d}_t$ *and* $\hat{\mathbf{d}}_t$, $\tilde{\mathbf{a}}_t = (\tilde{a}_t, \tilde{a}_t^2, \cdots, \tilde{a}_t^n, \cdots)'$ *with* $\tilde{a}_t = 1/(1+\tilde{r}_t)$, $\tilde{r}_t$ *lies between* r_t *and* $\hat{r}_t$, *and*

2. *for the estimator* $\hat{\rho}_t$ *of* ρ_t, *there exists a unique solution for its variance* $\sigma^2_{\rho t}$ *satisfying*

$$\sigma^2_{\rho t} = m^2(1+\check{r}_t)^{2m-2}\sigma^2_{rt}, \tag{12}$$

   where $\check{r}_t$ *lies between* r_t *and* $\tilde{r}_t$, *and the estimate* $\hat{\rho}_t$ *is obtained by* $F(\hat{r}_t)$ *using Equation (2).*

The proof is in Appendix A2. Next we study the validity of Assumption 4.

3 Covariance Matrix of the Forecast Errors

Theorem 1 provides all the necessary and sufficient conditions for the estimation for the reliability of the cost of capital. In spite of the conditions, the theory is still not considered complete if Assumption 4 does not hold. In this connection, the validity of Assumption 4 is tested by examining the covariance structure on three ARIMA models and discussing the situations for the general models. For simplicity, in this section $\tilde{\mathbf{a}}_t$ is replaced by $\mathbf{a}$.

Firstly consider the covariance matrix of the forecast errors when the dividends $\{d_t\}_{t=1}^T$ follow an ARIMA(0,1,1) model:

Model A: $$(1 - B)d_t = \delta + (1 - \theta B)\varepsilon_t \, .$$

For this model, referring to the proof in Appendix A3, we have

$$\mathbf{a}'\Sigma_t\mathbf{a} = \frac{a^2(1-\theta a)^2\sigma^2}{(1-a^2)(1-a)^2}. \tag{13}$$

For this example, we can set $k = 1 + r/2$. Then, Assumption 4 holds automatically.

Next will be a study of the covariance matrix of the forecast errors when the dividends $\{d_t\}_{t=1}^T$ follow an ARIMA$(0, 1, q)$ model:

Model B: $$(1 - B)d_t = \delta + (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)\varepsilon_t \, .$$

For this example, Assumption 4 holds automatically (see Appendix A4).

Finally, the covariance matrix of the forecast errors when the dividends $\{d_t\}_{t=1}^T$ follows an ARIMA(0,1,1) is analyzed as shown in the model:

Model C: $$(1 - \phi B)d_t = \delta + \varepsilon_t \, .$$

As proved in Appendix A5, we have

$$\mathbf{a}'\Sigma_t\mathbf{a} = \frac{a^2\sigma^2}{(1-a^2)(1-a\phi)^2}. \tag{14}$$

For this example, Assumption 4 holds if $|\phi/(1+r)| < 1$. As $|\phi/(1+r)| < 1$ can be obtained easily by applying Assumption 2, Assumption 4 holds automatically.

Model C is important in the theory of the estimation of the cost of capital because: (i) it is common for academics or financial practitioners to use an AR model with $|\phi| > 1$[2] to include situations in which the growth rate is considered in the dividends; (ii) in practice, most of the dividend series will be stationary after differencing once. In this case, ARIMA$(0, 1, q)$ will be the right model. However, the covariance matrix of the forecast errors for the ARIMA$(0, 1, q)$ will be dominated by the AR(1) with $|\phi| > 1$ for large $\{(i, j)\}$. Hence, the study of Model C guarantees that Theorem 1 holds for any ARIMA model. Refer to the discussion after (21) in the next section, we can further drop the requirement of ARIMA model to be any general time series model used for the forecasting of the dividends series.

4 The Estimation Procedure

This section describes the estimation procedure and explores the properties for the estimation of the reliability for the cost of capital. Note that the iterative procedure for estimating the cost of capital itself has been fully explored by Thompson (1985, 1987) and Thompson and Wong (1991, 1996).

Begin with an estimation of a time series model of past dividends. From the time series model, all parameters can be estimated and future dividends, $\mathbf{d}_t$, can be estimated by $\hat{\mathbf{d}}_t$ using the statistical procedures germane to the time series model. Thereafter r_t can be estimated by $\hat{r}_t$ which satisfies (10).

It is difficult to obtain $\hat{r}_t$ by solving Equation (10) directly in most of the situations because it involves an infinite sum. Summation has been completed by Thompson (1985) algebraically together with the procedure of applying Newton's method to estimate the cost of capital. Thompson and Wong (1996) introduce an alternative iterative approach to get $\hat{r}_t$. The estimate $\hat{\rho}_t$ can then be obtained by $F(\hat{r}_t)$ using Equation (2).

[2] see Thompson (1987).

After obtaining the estimates $\hat{r}_t$ and $\hat{\rho}_t$, the variance σ^2_{rt} can be estimated by

$$\hat{\sigma}^2_{rt} = \left[\sum_{i=1}^{\infty} \frac{i\hat{d}_{t+i}}{(1+\hat{r}_t)^{i+1}}\right]^{-2} \hat{\mathbf{a}}_t' \hat{\Sigma}_t \hat{\mathbf{a}}_t \tag{15}$$

and the variance $\sigma^2_{\rho t}$ can be estimated by

$$\hat{\sigma}^2_{\rho t} = m^2(1+\hat{r}_t)^{2m-2}\hat{\sigma}^2_{rt}. \tag{16}$$

The reliabilities of the discount rate r_t and of the cost of capital ρ_t can be measured by their standard deviations $\hat{\sigma}_{rt}$ and $\hat{\sigma}_{\rho t}$ respectively. We note that in Theorem 1, the estimator for σ^2_{rt} is in term of $\tilde{d}_{t+i}$, $\tilde{\mathbf{a}}_t$ and $\tilde{r}_t$ while the estimator for $\sigma^2_{\rho t}$ is in term of $\check{r}_t$. In practice, we use $\hat{d}_{t+i}$ to estimate $\tilde{d}_{t+i}$, use $\hat{\mathbf{a}}_t$ to estimate $\tilde{\mathbf{a}}_t$ and use $\hat{r}_t$ to estimate both $\tilde{r}_t$ and $\check{r}_t$.

In order to estimate σ^2_{rt}, two sequences $\{\hat{\sigma}^2_{1,t,n}\}$ and $\{\hat{\sigma}^2_{2,t,n}\}$ have been proposed by Thompson and Wong (1996) such that:

$$\hat{\sigma}^2_{1,t,n} = \left[\sum_{i=1}^{2n} \frac{i\hat{d}_{t+i}}{(1+\hat{r}_t)^{i+1}}\right]^{-2} \hat{\mathbf{a}}_{t,n}' \hat{\Sigma}_{t,n} \hat{\mathbf{a}}_{t,n} \tag{17}$$

$$\hat{\sigma}^2_{2,t,n} = \left[\sum_{i=1}^{n} \frac{i\hat{d}_{t+i}}{(1+\hat{r}_t)^{i+1}}\right]^{-2} \hat{\mathbf{a}}_{t,n}' \hat{\Sigma}_{t,n} \hat{\mathbf{a}}_{t,n} \tag{18}$$

where $\hat{\mathbf{a}}_{t,n} = (\hat{a}_t, \hat{a}_t^2, \cdots, \hat{a}_t^n)'$ with $\hat{a}_t = 1/(1+\hat{r}_t)$ and $\Sigma_{t,n} = E[\ (\hat{\mathbf{d}}_{t,n} - \mathbf{d}_{t,n})(\hat{\mathbf{d}}_{t,n} - \mathbf{d}_{t,n})'\ ]$ with $\mathbf{d}_{t,n} = (d_{t+1}, d_{t+2}, \ldots, d_{t+n})$ and $\hat{\mathbf{d}}_{t,n} = (\hat{d}_{t+1}, \hat{d}_{t+2}, \cdots, \hat{d}_{t+n})$. For the Wisconsin Power Pte Ltd data, it has been observed that the sequence $\{\hat{\sigma}^2_{1,t,n}\}$ (and respectively $\{\hat{\sigma}^2_{2,t,n}\}$) is an increasing (respectively decreasing) sequence converging to $\hat{\sigma}^2_{rt}$. Thus they can be used in the estimation of σ^2_{rt}. For a tolerance level α, we can then find n such that $\hat{\sigma}^2_{1,t,n}$ and $\hat{\sigma}^2_{2,t,n}$ satisfy

$$|\hat{\sigma}_{1,t,n} - \hat{\sigma}_{2,t,n}| \leq \alpha\ . \tag{19}$$

In this situation, both $\hat{\sigma}^2_{1,t,n}$ and $\hat{\sigma}^2_{2,t,n}$ or any of their linear combinations can be used as an estimate of σ^2_{rt}. Thereafter, $\sigma^2_{\rho t}$ can be estimated by applying Equation (16).

However, it is well-known that in general $\{\hat{\sigma}^2_{1,t,n}\}$ and $\{\hat{\sigma}^2_{2,t,n}\}$ may not necessarily be an increasing function and a decreasing function respectively. If they are not, then $|\hat{\sigma}_{1,t,n} - \hat{\sigma}_{rt}|$ and/or $|\hat{\sigma}_{2,t,n} - \hat{\sigma}_{rt}|$ can be greater than α even if (19) holds. In this situation, neither $\hat{\sigma}^2_{1,t,n}$ nor $\hat{\sigma}^2_{2,t,n}$ can be used as an estimate for σ^2_{rt}. To overcome the difficulty, we define

$$\hat{\sigma}^2_{t,m,n} = G(t,m,n) = \left[\sum_{i=1}^{m} \frac{i\hat{d}_{t+i}}{(1+\hat{r}_t)^{i+1}}\right]^{-2} \hat{\mathbf{a}}'_{t,n}\hat{\Sigma}_{t,n}\hat{\mathbf{a}}_{t,n} \tag{20}$$

and introduce the following theorem to make the theory of the estimation for the cost of capital complete:

Theorem 2 *There exist subsequences $\{n_1\}, \{n_2\}, \{n_3\}$ and $\{n_4\}$ such that*

1. *$\hat{\sigma}^2_{t,n_1,n_2}$ defined in (20) is an increasing series converging to $\hat{\sigma}^2_{rt}$; and*
2. *$\hat{\sigma}^2_{t,n_3,n_4}$ defined in (20) is a decreasing series converging to $\hat{\sigma}^2_{rt}$.*

The proof is shown in Appendix A6.

To estimate σ^2_{rt}, the most difficult way will be computing $\hat{\mathbf{a}}'_{t,n}\hat{\Sigma}_{t,n}\hat{\mathbf{a}}_{t,n}$, especially since each entry in the matrix $\Sigma_t = (\sigma_{ij})$ depends on the time series model for $\{d_t\}$. To make the entries of the matrix Σ_t independent of the model for the class of ARIMA models, we assume $\{d_t\}$ follows the model:

$$\Phi(B)(1-B)^d d_t = \Theta(B)\varepsilon_t\ .$$

Alternatively the model can easily be re-written as:

$$d_t = \Psi(B)\varepsilon_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \tag{21}$$

with $\psi_0 = 1$.

Actually the assumption that the dividends follow an ARIMA model can be omitted as Equation (21) can be obtained by Wold's Representation Theorem (see Box et al 1994) for nearly any time series model. As long as ψ_i can be estimated for any i, the estimation of the cost of capital and its reliability in the paper can also be obtained and hence it becomes "model-free". Nevertheless, estimating σ^2_{rt} is still the most difficult part. To make the computation easier, the following theorem is introduced:

Theorem 3 *The product $\hat{\mathbf{a}}'_{t,N}\hat{\Sigma}_{t,N}\hat{\mathbf{a}}_{t,N}$ defined in (20) can be written as*

$$\hat{\mathbf{a}}'_{t,N}\hat{\Sigma}_{t,N}\hat{\mathbf{a}}_{t,N} = \sigma^2 \sum_{i=1}^{N} \psi_{i-1} \sum_{j=1}^{N} a^{i-j} \o_j, \tag{22}$$

where

$$\o_j = \psi_{j-1} \left(\frac{a^{2j} - a^{2N+2}}{1 - a^2} \right).$$

Thus, the computation of $\hat{\mathbf{a}}'_{t,N}\hat{\Sigma}_{t,N}\hat{\mathbf{a}}_{t,N}$ can be done in $O(N \log N)$ operations.

The proof is in Appendix A7.

Here, $\sum_{j=1}^{N} a^{i-j}\o_j$ is the product of a Toeplitz matrix times a vector. This can be done in $O(N \log N)$ operations by embedding the Toeplitz matrix in a circulant matrix and then using Fast Fourier Transform, see Chan and Ng (1996). Hence $\hat{\mathbf{a}}'_{t,N}\hat{\Sigma}_{t,N}\hat{\mathbf{a}}_{t,N}$ can be done in $O(N \log N)$ operations too, and Equation (22) speeds up the estimation procedure.

The reliabilities of both the discount rate and the cost of capital ρ_t can be measured by Equations (15) and (16) respectively, which unfortunately involve infinite sums. Thompson and Wong (1996) use the estimates in Equations (17) and (18) for the reliability of the discount rate, and they involve only finite sums. This makes the estimation possible. Application of Equation (22) further reduces the computation complexity, resulting in higher estimation accuracy.

5 Discussion

In this paper we have been concerned with the applicability of the old, but venerable, "dividend yield plus growth" model. Our analysis rests squarely on four assumptions to guarantee that there will be a solution, in terms of r_t to the equation

$$P_t = \sum_{i=1}^{\infty} \frac{d_{t+i}}{(1+r_t)^i}.$$

Thompson and Wong (1996) have discussed in detail the validity of Assumptions 1 to 3 in reality. Hence, the only assumption that concerns the application of our method is the fourth one, that the $\{(i,j)\}$ entry of the covariance matrix of the forecast errors is bounded by $M\,k^{i+j}$ as in Equation (9). As discussed in Section 3, Assumption 4 is valid as ARIMA$(0,1,q)$ is a good approximation for most of the models used for the dividends and the estimate of the forecast errors of nearly all models should be bounded by the forecast errors of the AR(1) model with $|\phi| \geq 1$. Applying Wold's representation Theorem, one can conclude that the methods presented here can be applied to most, if not all, practical situations and can therefore be used without fear of troubling anomalies.

Above all, the approaches shown in this paper to determine the cost of capital is adaptable to PCs. It consists of calculating a sequence on cost of capital estimates which are guaranteed to converge to the cost of capital. The calculation of a sequence on the reliability of the cost of capital are also certain to converge to the reliability.

The formula of the reliabilities for the discount rate had previously involve infinitive dimensional vectors and matrices, hence the estimation is not feasible. Thompson and Wong (1996) have therefore introduced the estimates of the reliability for the discount rate in which all vectors and matrices in the formula are finite. This enables estimation to be carried out. Nevertheless, when the dimension of the vectors and matrices are large, the estimation will take up considerable computation time and incur more rounding error in the estimation process. In this paper, a formula is introduced to reduce the computation complexity, thus it speeds up the estimation procedure leading to higher accuracy rate.

The method presented here rests solidly on the basis where past historical observations are relevant to the future dividends, notwithstanding situations where the estimates cannot be precise. However it is common knowledge that estimates are inherently inaccurate. Thompson and Wong (1996) concluded that the methods developed for the estimation of the cost of capital and its reliability in most situations are still relevant especially to regulated industries. This is because the statistical time series models have the ability to track gradual changes and adapt to them.

After the 1980s, one cannot help but be struck by the massive changes taking place in the business world. Changes would include precipitous declines in the business fortunes of many highly regarded firms; deregulation in the trucking, airline, and banking industries; restructuring in the oil industry; the rise and fall of internet stocks; a move toward globalization and enhanced competition; and countless other changes which were unexpected prior to the 1980s. One may wonder how well the cost of capital can be applied in this changing environment. Even though the estimation may not be so accurate, our approach is still useful due to the following reasons: (i) the estimated cost of capital and its reliability provides the best information we can get based on the present price and past and present dividends, which gives investors the figure for estimated returns if the time series model for the past and present dividends is correct, (ii) the model may change as time varies, and our approach provides the formula for investors to update the cost of capital and its reliability from time to time, and (iii) for those companies with significant dividend fluctuations, the forecast errors of the future dividends will consequently be large. In return the reliability of the cost of capital will become immense and hence the confidence interval for the cost of capital will be wider. Thus, the approach demonstrated in this paper still provides investors useful information on the returns and reliability of the stocks purchased.

Nevertheless, investors may incorporate other approaches to improve the estimation of the cost of capital and its reliability. One such technique is the Bayesian approach (Matsumura et al 1990 and Wong and Bian 2000), while another is the repeated time series approach, (Wong and Miller 1990 and Wong et al 2001b). Once the cost of capital is computed, it may be applied in stock selection. It will definitely be better if some other methodologies are included, e.g. stochastic dominance approach (Wong and Li 1999 and Li and Wong 1999), technical analysis approach (Wong et al 2001a, 2003) and to incorporate the economic and financial situations of the market (Manzur et al 1999, Wan and Wong 2001 and Wong et al 2004) in the decision-making process.

This paper has developed the estimators for the cost of capital and its reliability. However, we are still not able to construct the confidence interval for the cost of capital as the distribution of the estimator for the cost of capital is unknown. To study its distribution, one may have to use Monte Carlo methods. The distribution may still be normal but it is more likely that it is non-normal or even be skewed. One may refer to Tiku et al (2000) for the flat-tailed symmetric distribution or to Tiku et al (1999) for the asymmetric distribution. After acquiring information on the distribution, one can then construct simulation to obtain the critical values and thereafter the confidence intervals can be achieved.

Finally, although academics and finance practitioners usually believe that the dividends series will be stationary, even after differencing once, the series may remain stationary. In this situation, a unit root test and cointegration test (Tiku and Wong 1998 and Wong et al 2004) should be incorporated in the estimation.

References

Box, G.E.P., G.M. Jenkins, & Gregory C. Reinsel (1994). *Time Series Analysis Forecasting and Control,* 3rd. Edition, Prentice-Hall, Inc.

Chan, R.H. & M.K. Ng (1996). Conjugate gradient method for Toeplitz systems. *SIAM Review*, 38, 427–482.

Dieudonne, J. (1960). *Foundations of Modern Analysis.* Academic Press, New York: Academic Press.

Gordon, M.J. & E. Shapiro (1956). Capital equipment analysis: the required rate of profit. *Management Science,* X, 102–110.

Higgens, R.C. (1974). Growth dividend policy, and capital costs in the electric utility industry. *Journal of Finance,* X, 1189–1202.

Li, C.K. & W.K. Wong (1999). A note on stochastic dominance for risk averters and risk takers. *RAIRO Recherche Operationnelle.* 33, 509-524.

Litzenberger, R.H. & C.V. Rao (1971). Estimates of the marginal time preference and average risk aversion of investors in electric utility shares, 1960–1966. *Bell Journal of Economics and management Science,* Spring, 265–277.

Makhija, A. & H.E. Thompson (1984). Comparison of alternative models for estimating the cost of capital for electric utilities. *The Journal of Economics and Business,* 36, 107–132.

Manzur, M., W.K. Wong, & I.C. Chau (1999), Measuring international competitiveness: experience from East Asia. *Applied Economics,* 31, 1383–1391.

Matsumura, E.M., K.W. Tsui, & W.K. Wong (1990). An extended multinomial-Dirichlet model for error bounds for dollar-unit sampling. *Contemporary Accounting Research.* 6(2-I), 485–500.

McDonald, J.G. (1971). Required return on public utility equities: a national and regional analysis, 1958-1969. *Bell Journal of Economics and management Science,* Autumn, 503–514.

Miller, M. & F. Modigliani (1966). Some estimates of the cost of capital to the electric utility industry, 1954–1957. *American Economic Review,* 56, 333–391.

Thompson, H.E. (1979). The cost of capital for electric utilities: 1958–1976. *The Bell Journal of Economics,* 10, 619–635.

Thompson, H.E. (1984). Estimating individual company costs of capital from cross sectional data: a random coefficients approach. *Managerial and Decision Economics,* 5, 130–140.

Thompson, H.E. (1985). The magnitude and reliability of equity capital cost estimates: a statistical approach. *Managerial and Decision Economics,* 6, 132–140.

Thompson, H.E. (1987). Determination of benchmark rates of return. *Managerial and Decision Economics,* 8, 321–332.

Thompson, H.E. & W.K. Wong (1991). On the unavoidability of "unscientific" judgement in estimating the cost of capital. *Managerial and Decision Economics,* 12, 27–42.

Thompson, H.E. & W.K. Wong (1996). Revisiting 'dividend yield plus growth' and its applicability. *Engineering Economist,* 41, 123–147.

Tiku, M.L. & W.K. Wong (1998). Testing for unit root in AR(1) model using three and four moment approximations. *Communications in Statistics: Simulation and Computation,* 27(1), 185–198.

Tiku, M.L., W.K. Wong, & G. Bian (1999). Time series models with asymmetric innovations. *Communications in Statistics: Theory and Methods,* 28(6), 1331–1360.

Tiku, M.L., W.K. Wong, D.C. Vaughan, & G. Bian (2000). Time series models with nonnormal innovations: symmetric location-scale distributions. *Journal of Time Series Analysis,* 21(5), 571–596.

Wan, Henry Jr. & W.K. Wong (2001). Contagion or inductance: crisis 1997 reconsidered. *Japanese Economic Review,* 52(4), 372–380.

Wong, W.K. & G. Bian (2000). Robust estimation in capital asset pricing estimation.

Journal of Applied Mathematics & Decision Sciences, 4(1), 65–82.

Wong, W.K., B.K. Chew, & D. Sikorski (2001a). Can P/E ratio and bond yield be used to beat stock markets? *Multinational Finance Journal,* 5, 59–86.

Wong, W.K. & C.K. Li (1999). A note on convex stochastic dominance theory. *Economics Letters,* 62, 293–300.

Wong W.K., M. Manzur, & B.K. Chew (2003). How rewarding is technical analysis? Evidence from Singapore stock market. *Applied Financial Economics,* 13(7), 543–551.

Wong, W.K. & R.B. Miller (1990). Analysis of ARIMA-noise models with repeated time series. *Journal of Business and Economic Statistics,* 8(2), 243–250.

Wong, W.K., R.B. Miller & K. Shrestha (2001b). Maximum Likelihood Estimation of ARMA Model with Error Processes for Replicated Observation. *Journal of Applied Statistical Science,* 10(4), 287-297.

Wong, W.K., J.H.W. Penm, R.D. Terrell, & K.Y.C. Lim (2004). The Relationship between stock markets of major developed countries and Asian emerging markets. *Journal of Applied Mathematics and Decision Sciences,* (forthcoming).

Appendix

For simplicity, we use $\mathbf{a}$, $\mathbf{a}_N$, a, r and $\Sigma_{t,N}$ for $\tilde{\mathbf{a}}_t$, $\tilde{\mathbf{a}}_{t,N}$, $\tilde{a}_t$, $\tilde{r}_t$ and $\tilde{\Sigma}_{t,N}$ respectively for all the proofs below.

A1. *Proof of Lemma 1:*

We consider the norm in the vector space R of the set of real numbers to be $\|r\| = |r|$ for $r \in \mathsf{R}$. It is well-known that R is a Banach space. We define the norms:

$$\|\mathbf{d}_t\|_{r_0} = \sum_{i=1}^{\infty} \frac{|d_{t+i}|}{(1+r_0)^i} \qquad \text{and} \qquad \|(\mathbf{d}_t, r)\|_{r_0} = \|\mathbf{d}_t\|_{r_0} + |r|$$

where $\mathbf{d}_t = (d_{t+1}, d_{t+2}, \ldots, d_{t+i}, \ldots) \in \mathbf{R}^\infty$; $r, r_0 \in \mathsf{R}$; and $r_0 > -1$. Let the normed space $\mathbf{E}_{t,r_0}$ be defined as

$$\mathbf{E}_{t,r_0} = \{\mathbf{d}_t \in \mathbf{R}^\infty : \|\mathbf{d}_t\|_{r_0} < \infty\}$$

and the norm of the normed space $\mathbf{E}_{t,r_0} \times \mathsf{R}$ be $\|(\mathbf{d}_t, r)\|_{r_0}$. We show that $\mathbf{E}_{t,r_0}$ and $\mathbf{E}_{t,r_0} \times \mathsf{R}$ are complete as follows:

Let the function $F : \mathbf{E}_{t,0} \to \mathbf{E}_{t,r_0}$ to be

$$F(\mathbf{d}_t) = \left(\frac{d_{t+1}}{1+r_0}, \frac{d_{t+2}}{(1+r_0)^2}, \cdots, \frac{d_{t+n}}{(1+r_0)^n}, \cdots\right).$$

It is easy to check that F is a linear isometric isomorphism. $\mathbf{E}_{t,0}$, which is equal to the Banach space l^1, is well-known to be complete. Hence, $\mathbf{E}_{t,r_0}$ is complete. Consequently, $\mathbf{E}_{t,r_0} \times \mathsf{R}$ is also complete. And therefore both are Banach spaces.

In the following we prove that if r_0 satisfies (4), then the function $f_t : \mathbf{E}_{t,r_0} \times \mathsf{R} \to \mathsf{R}$ defined in (5) is continuously differentiable with its derivative $Df_t(\mathbf{d}_t, r)$ satisfying:

$$Df_t(\mathbf{d}_t, r) \cdot (\mathbf{u}, v) = \sum_{i=1}^{\infty} \frac{u_i}{(1+r)^i} + \sum_{i=1}^{\infty} \frac{-i d_{t+i}}{(1+r)^{i+1}} \cdot v \tag{23}$$

where $\mathbf{u} = (u_1, u_2, \cdots, u_i, \cdots) \in \mathbf{E}_{t,r_0}$ and $v \in \mathsf{R}$. First notice that

$$\Delta \equiv \|f_t(\mathbf{d}_t + \mathbf{u}, r+v) - f_t(\mathbf{d}_t, r) - [\sum_{i=1}^{\infty} \frac{u_i}{(1+r)^i} + \sum_{i=1}^{\infty} \frac{-i d_{t+i}}{(1+r)^{i+1}} \cdot v]\|$$

$$
\begin{aligned}
&= \quad \| \sum_{i=1}^{\infty} \frac{d_{t+i}}{(1+r_0)^i} [\frac{(1+r_0)^i}{(1+r+v)^i} - \frac{(1+r_0)^i}{(1+r)^i} + \frac{iv(1+r_0)^i}{(1+r)^{i+1}}] \\
&\qquad + \sum_{i=1}^{\infty} \frac{u_i}{(1+r_0)^i} [\frac{(1+r_0)^i}{(1+r+v)^i} - \frac{(1+r_0)^i}{(1+r)^i}] \| \\
&\leq \quad \|\mathbf{d}_t\|_{r_0} \cdot | \sum_{i=1}^{\infty} [\frac{(1+r_0)^i}{(1+r+v)^i} - \frac{(1+r_0)^i}{(1+r)^i} + \frac{iv(1+r_0)^i}{(1+r)^{i+1}}] \\
&\qquad + \|\mathbf{u}\|_{r_0} \cdot | \sum_{i=1}^{\infty} [\frac{(1+r_0)^i}{(1+r+v)^i} - \frac{(1+r_0)^i}{(1+r)^i}] |
\end{aligned}
$$

where $\|(\mathbf{u}, v)\|_{r_0}$ goes to zero implies both $\|\mathbf{u}\|_{r_0}$ and $|v|$ tend to zero. As $v \to 0$, we can set v such that $r_0 < r + v$ and therefore

$$
\begin{aligned}
\Delta \quad &\leq \quad \|\mathbf{d}_t\|_{r_0} \cdot | \frac{1+r_0}{r+v-r_0} - \frac{1+r_0}{r-r_0} + \frac{v(1+r_0)}{(r-r_0)^2} | \\
&\qquad + \|\mathbf{u}\|_{r_0} \cdot | \frac{1+r_0}{r+v-r_0} - \frac{1+r_0}{r-r_0} | \\
&\leq \quad M_1 v^2 + M_2 \|\mathbf{u}\|_{r_0} \cdot |v|
\end{aligned}
$$

where M_1 and M_2 are finite. Hence

$$
\frac{\Delta}{\|(\mathbf{u}, v)\|_{r_0}} \longrightarrow 0 \qquad \text{as} \quad \|(\mathbf{u}, v)\|_{r_0} \to 0 \, ,
$$

and therefore f_t is differentiable with its derviative $Df_t(\mathbf{d}_t, r)$ satisfying (23). Similarly, one can show that for any $\varepsilon > 0$, there exists a δ such that for $\|(\mathbf{s}, t)\|_{r_0} \leq \delta$ and $\|(\mathbf{u}, v)\|_{r_0} \leq 1$,

$$
\begin{aligned}
&\sup \|[Df_t(\mathbf{d}_t + \mathbf{s}, r+t) - Df_t(\mathbf{d}_t, r)] \cdot (\mathbf{u}, v)\| \\
\leq \quad &\left| \frac{1+r_0}{(r+t-r_0)(r-r_0)} \right| \cdot |t| \\
&\qquad + \|\mathbf{d}\|_{r_0} \cdot \left| \frac{2r+t-2r_0}{(r+t-r_0)^2 (r-r_0)^2} \right| \cdot |t| \|\mathbf{s}\|_{r_0} \cdot \left| \frac{1+r_0}{(r+t-r_0)^2} \right| \\
\leq \quad &\varepsilon
\end{aligned}
$$

for any $(\mathbf{d}_t, r) \in \mathbf{E}_{t,r_0} \times (r_0, \infty)$. Hence f_t is continuously differentiable.

Let $D_1 f_t(\mathbf{d}_t, r)$ and $D_2 f_t(\mathbf{d}_t, r)$ be the partial derivatives of f_t with respect to $\mathbf{d}_t$ and r respectively. By Theorem 8.9.1 in Dieudonne (1960), the mappings $(\mathbf{d}_t, r) \to$

$D_1 f_t(\mathbf{d}_t, r)$ and $(\mathbf{d}_t, r) \rightarrow D_2 f_t(\mathbf{d}_t, r)$ are continuous in $\mathbf{E}_{t,r_0} \times (r_0, \infty)$, and

$$D f_t(\mathbf{d}_t, r) \cdot (\mathbf{u}, v) = D_1 f_t(\mathbf{d}_t, r) \cdot \mathbf{u} + D_2 f_t(\mathbf{d}_t, r) \cdot v.$$

For any point $(\mathbf{d}_t, r_t) \in \mathbf{S}_t \times (r_0, \infty)$ satisfying the equation of f_t in (5), the second partial derivatives $D_2 f_t(\mathbf{d}_t, r_t)$ is linear homeomorphism since $D_2 f_t(\mathbf{d}_t, r_t) \neq 0$. We remark that $\mathbf{S}_t$ is the set of $\mathbf{d}_t$ which satisfies Assumptions 1 to 2.

Finally, by applying Theorem 10.2.1 in Dieudonne (1960), we have the following results: There exists an open neighborhood U_0 of $\mathbf{d}_t$ in $\mathbf{E}_{t,r_0}$ such that for every open connected neighborhood U of $\mathbf{d}_t$, contained in U_0, there is a unique continuous mapping g_t of U into R such that $g_t(\mathbf{d}_t) = r_t$, $(\mathbf{d}_t, g_t(\mathbf{d}_t)) \in \mathbf{S}_t \times (r_0, \infty)$ and $f_t(\mathbf{d}_t, g_t(\mathbf{d}_t)) = 0$ for any $\mathbf{d}_t \in U$. Furthermore, g_t is continuously differentiable in U and its derivative is given by

$$D g_t(\mathbf{d}_t) = -D_2 f(\mathbf{d}_t, g_t(\mathbf{d}_t))^{-1} D_1 f(\mathbf{d}_t, g_t(\mathbf{d}_t)).$$

Besides, it is easy to show that

$$\sum_{i=1}^{n} \frac{i d_{t+i}}{(1+r)^i} \longrightarrow \sum_{i=1}^{\infty} \frac{i d_{t+i}}{(1+r)^i} \qquad \text{and} \quad g_{t,n} \longrightarrow g_t \quad \text{as} \quad n \rightarrow \infty$$

where $g_{t,n}$ is defined in (6). Hence, the equation in (8) holds.

A2. *Proof of Theorem 1:*

We only prove the finiteness of $|\mathbf{a}'\Sigma_t\mathbf{a}|$ here. The rest of the proof is either straightforward or can be modified from the proof in Thompson and Wong (1996). From Assumption 4, as $|\sigma_{ij}| < M\, k^{i+j}$ and $k < 1 + r$ except for a finite set of $\{(i, j)\}$, there exists a constant A such that

$$|\mathbf{a}'\Sigma_t\mathbf{a}| < A + M \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \left(\frac{k}{1+r} \right)^{i+j} = A + M \left(\frac{1}{1 - \frac{k}{1+r}} \right)^2 < \infty. \qquad \square$$

A3. *Proof of Equation (13):*

One can easily show that the covariance for the future dividend (σ_{nm}) at time $T+n$ is

$$\sigma_{nm} = \mathrm{cov}(e_{T+n}, e_{T+m}) = \begin{cases} \sigma^2[1+(n-1)(1-\theta)^2] & n = m \geq 1, \\ \sigma^2[(1-\theta)+(m-1)(1-\theta)^2] & n > m \geq 1. \end{cases}$$

For simplicity, we let $\Theta = 1-\theta$. Then, we have

$$\begin{aligned}
\tilde{\mathbf{a}}_t' \Sigma_t \tilde{\mathbf{a}}_t &= \sigma^2 \mathbf{a}' \begin{pmatrix} 1 & \Theta & \Theta & \cdots & \cdots & \Theta & \cdots \\ \Theta & 1+\Theta^2 & \Theta+\Theta^2 & \cdots & \cdots & \Theta+\Theta^2 & \cdots \\ \Theta & \Theta+\Theta^2 & 1+2\Theta^2 & \cdots & \cdots & \Theta+2\Theta^2 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & 1+(n-1)\Theta^2 & \Theta+n\Theta^2 & \cdots \\ \Theta & \Theta+\Theta^2 & \Theta+2\Theta^2 & \cdots & \Theta+n\Theta^2 & 1+n\Theta^2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \mathbf{a} \\
&= \sigma^2 \mathbf{a}' \begin{pmatrix} 1 & \Theta & \Theta & \cdots & \cdots & \Theta & \cdots \\ \Theta & 1 & \Theta & \cdots & \cdots & \Theta & \cdots \\ \Theta & \Theta & 1 & \cdots & \cdots & \Theta & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & 1 & \Theta & \cdots \\ \Theta & \Theta & \Theta & \cdots & \Theta & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \mathbf{a} \\
&\quad + a^2\sigma^2 \mathbf{a}' \begin{pmatrix} \Theta^2 & \Theta^2 & \cdots & \cdots & \Theta^2 & \cdots \\ \Theta^2 & 2\Theta^2 & \cdots & \cdots & 2\Theta^2 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & (n-1)\Theta^2 & n\Theta^2 & \cdots \\ \Theta^2 & 2\Theta^2 & \cdots & n\Theta^2 & n\Theta^2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \mathbf{a} \\
&= \theta\sigma^2 \mathbf{a}' \mathbf{I} \mathbf{a} + \Theta\sigma^2 \mathbf{a}' \mathbf{1} \mathbf{a} + \ a^2\Theta^2\sigma^2 \mathbf{a}' \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & \cdots \\ 1 & 2 & 2 & \cdots & 2 & \cdots \\ 1 & 2 & 3 & \cdots & 3 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 2 & 3 & \cdots & n & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \mathbf{a}
\end{aligned}$$

$$= \quad \theta\sigma^2 \mathbf{a}'\mathbf{I}\mathbf{a} + \Theta\sigma^2\mathbf{a}'\mathbf{1}\mathbf{a} + \ a^2\Theta^2\sigma^2\mathbf{a}'E\mathbf{a},$$

where $\mathbf{I}$ and $\mathbf{1}$ are the identity matrix and matrix of all ones respectively. As

$$E = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & \cdots \\ 1 & 1 & 1 & \cdots & 1 & \cdots \\ 1 & 1 & 1 & \cdots & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & \cdots \\ 0 & 1 & 1 & \cdots & 1 & \cdots \\ 0 & 1 & 2 & \cdots & 2 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 2 & \cdots & n-1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

we have

$$\mathbf{a}'E\mathbf{a} = \mathbf{a}'\mathbf{1}\mathbf{a} + \ a^2\mathbf{a}'E\mathbf{a}.$$

Hence,

$$\mathbf{a}'E\mathbf{a} = \frac{1}{1-a^2}\mathbf{a}'\mathbf{1}\mathbf{a}.$$

If we let $s = a + a^2 + a^3 + \cdots$, we have

$$\mathbf{a}'\mathbf{1}\mathbf{a} = \mathbf{a}'\begin{pmatrix} s \\ s \\ s \\ \cdots \end{pmatrix} = a\,s + a^2 s + a^3 s + \cdots = s^2 = \left(\frac{a}{1-a}\right)^2.$$

Since

$$\mathbf{a}'\mathbf{I}\mathbf{a} = a^2 + a^4 + a^6 + \cdots = \frac{a^2}{1-a^2},$$

we then have

$$\begin{aligned} \mathbf{a}'\Sigma_t\mathbf{a} &= \theta\sigma^2\mathbf{a}'\mathbf{I}\mathbf{a} + \Theta\sigma^2\mathbf{a}'\mathbf{1}\mathbf{a} + \ a^2\Theta^2\sigma^2\frac{1}{1-a^2}\mathbf{a}'\mathbf{1}\mathbf{a} \\ &= \theta\sigma^2\frac{a^2}{1-a^2} + \Theta\sigma^2\left(\frac{a}{1-a}\right)^2 + \ a^2\Theta^2\sigma^2\frac{1}{1-a^2}\left(\frac{a}{1-a}\right)^2 \\ &= \frac{\theta a^2\sigma^2}{1-a^2} + \frac{\Theta a^2\sigma^2}{(1-a)^2} + \frac{a^4\Theta^2\sigma^2}{(1-a^2)(1-a)^2} = \frac{a^2(1-\theta a)^2\sigma^2}{(1-a^2)(1-a)^2}. \qquad \square \end{aligned}$$

A4. *Proof of the finiteness of* $\mathbf{a}'\Sigma_t\mathbf{a}$ *in Model B:*

Let $k_i = 1-\theta_1-\cdots-\theta_i$ with $k_0 = 1$ and let $s_i = 1+(1-\theta_1)^2+\cdots+(1-\theta_1-\cdots-\theta_i)^2$ for $1 \le i \le q$ with $s_0 = 1$. One can easily show that the covariance for the future dividend (σ_{nm}) at time $T+n$ is:

$$\sigma_{nn} = \begin{cases} \sigma^2 s_{n-1} & 1 \le n \le q+1, \\ \sigma^2[s_{q-1}+mk_q^2] & n = m+q, m \ge 1, \end{cases} \tag{24}$$

$$\sigma_{n1} = \begin{cases} k_{n-1}\sigma^2 & 2 \le n \le q+1, \\ k_q\sigma^2 & n = m+q, m \ge 1, \end{cases} \tag{25}$$

$$\sigma_{np} = \begin{cases} \sigma^2 \sum_{i=0}^{p-1} k_{n-p+i}k_i & 1 \le p \le q,\ p < n \le q+p-1, \\ \sigma^2 k_q K_0^{p-1} & n = q+m, m \ge p,\ q \ge p \ge 1, \end{cases} \tag{26}$$

$$\sigma_{q+p+n,q+p} = \begin{cases} \sigma^2(\sum_{i=0}^{q-1} k_{n+i}k_i + pk_q^2) & 1 \le n \le q-1,\ p \ge 1, \\ \sigma^2(k_q K_0^{q-1} + pk_q^2) & n \ge q,\ p \ge 1 \end{cases} \tag{27}$$

where $K_0^p = \sum_{i=0}^p k_i$ for $p > 0$. From (24)–(27), we can tell that $\sigma_{m,n}$ is bounded by an arithmetic process while a^n is a geometric process. Hence, $\mathbf{a}'\Sigma_t\mathbf{a}$ is finite and Assumption 4 holds. $\square$

A5. *Proof of Equation (14):*

For the AR(1) model, one can easily show that the covariance for the future dividend (σ_{nm}) at time $T+n$ is:

$$\sigma_{nm} = \begin{cases} \sigma^2 \sum_{i=0}^{n-1} \phi^{2i} & n = m \ge 1, \\ \sigma^2(\phi^{n-m}+\phi^{n-m+2}+\cdots+\phi^{n+m-2}) & n > m \ge 1. \end{cases}$$

For simplicity, we let $\Phi_m^n = \phi^m + \phi^{m+2} + \phi^{m+4} + \cdots + \phi^n$ for $n \ge m$,

$$D = \begin{pmatrix} 1 & \phi & \phi^2 & \phi^3 & \cdots & \cdots \\ 0 & 1 & \phi & \phi^2 & \phi^3 & \cdots \\ 0 & 0 & 1 & \phi & \phi^2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

and let $\Sigma = D' + D - \mathbf{I}$. Then we have

$$\Sigma_t = \sigma^2 \begin{pmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{n-1} & \cdots \\ \phi & \Phi_0^2 & \Phi_1^3 & \cdots & \Phi_{n-2}^n & \cdots \\ \phi^2 & \Phi_1^3 & \Phi_0^4 & \cdots & \Phi_{n-3}^{n+1} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi^{n-1} & \Phi_{n-2}^n & \Phi_{n-3}^{n+1} & \cdots & \Phi_0^{2n-2} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

and

$$\mathbf{a}'\Sigma_t\mathbf{a} = \sigma^2\mathbf{a}'\Sigma\mathbf{a} + a^2\phi^2\mathbf{a}'\Sigma_t\mathbf{a} = \frac{\sigma^2}{1-a^2\phi^2}\mathbf{a}'\Sigma\mathbf{a}. \tag{28}$$

As

$$\begin{aligned} \mathbf{a}'D\mathbf{a} &= \mathbf{a}'D'\mathbf{a} = \mathbf{a}' \begin{pmatrix} a + a^2\phi + a^3\phi^2 + a^4\phi^3 + \cdots \\ a^2 + a^3\phi + a^4\phi^2 + a^5\phi^3 + \cdots \\ a^3 + a^4\phi + a^5\phi^2 + a^6\phi^3 + \cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{pmatrix} = \mathbf{a}' \begin{pmatrix} a\,s \\ a^2 s \\ a^3 s \\ a^4 s \\ \cdots \end{pmatrix} \\ &= a^2 s + a^4 s + a^6 s + \cdots = s(a^2 + a^4 + a^6 + \cdots) \\ &= a^2 s\, t = a^2 \left(\frac{1}{1-a\phi}\right)\left(\frac{1}{1-a^2}\right), \end{aligned}$$

where $s = 1 + a\phi + a^2\phi^2 + a^3\phi^3 + \cdots$ and $t = 1 + a^2 + a^4 + a^6 + \cdots$. Hence,

$$\begin{aligned} \mathbf{a}'\Sigma\mathbf{a} &= 2a^2 s\, t - \mathbf{a}'\mathbf{I}\mathbf{a} = 2a^2 s\, t - ||\mathbf{a}||^2 \\ &= 2a^2\left(\frac{1}{1-a\phi}\right)\left(\frac{1}{1-a^2}\right) - \frac{a^2}{1-a^2} = \left(\frac{a^2}{1-a^2}\right)\left(\frac{1+a\phi}{1-a\phi}\right). \end{aligned} \tag{29}$$

From (28) and (29), we have

$$\mathbf{a}'\Sigma_t\mathbf{a} = \left(\frac{\sigma^2}{1-a^2\phi^2}\right)\left(\frac{a^2}{1-a^2}\right)\left(\frac{1+a\phi}{1-a\phi}\right) = \frac{a^2\sigma^2}{(1-a^2)(1-a\phi)^2}. \qquad \square$$

A6. *Proof of Theorem 2:*

As the sequence

$$\left\{ \left[\sum_{i=1}^{n} \frac{i\hat{d}_{t+i}}{(1+\hat{r}_t)^{i+1}} \right]^{-2} \right\}$$

is a decreasing sequence while the sequence $\{\hat{\mathbf{a}}'_{t,n}\hat{\Sigma}_{t,n}\hat{\mathbf{a}}_{t,n}\}$ is an increasing sequence, there exist subsequences

$$\left\{ \left[\sum_{i=1}^{n_1} \frac{i\hat{d}_{t+i}}{(1+\hat{r}_t)^{i+1}} \right]^{-2} \right\}, \quad \left\{ \left[\sum_{i=1}^{n_2} \frac{i\hat{d}_{t+i}}{(1+\hat{r}_t)^{i+1}} \right]^{-2} \right\},$$

$\{\hat{\mathbf{a}}'_{t,n_3}\hat{\Sigma}_{t,n_3}\hat{\mathbf{a}}_{t,n_3}\}$ and $\{\hat{\mathbf{a}}'_{t,n_4}\hat{\Sigma}_{t,n_4}\hat{\mathbf{a}}_{t,n_4}\}$ such that

$$\left\{ \left[\sum_{i=1}^{n_1} \frac{i\hat{d}_{t+i}}{(1+\hat{r}_t)^{i+1}} \right]^{-2} \hat{\mathbf{a}}'_{t,n_3}\hat{\Sigma}_{t,n_3}\hat{\mathbf{a}}_{t,n_3} \right\}$$

is increasing while the sequence

$$\left\{ \left[\sum_{i=1}^{n_2} \frac{i\hat{d}_{t+i}}{(1+\hat{r}_t)^{i+1}} \right]^{-2} \hat{\mathbf{a}}'_{t,n_4}\hat{\Sigma}_{t,n_4}\hat{\mathbf{a}}_{t,n_4} \right\}$$

is decreasing. $\square$

A7. *Proof of Theorem 3:*

One can easily show that the forecast error for the future dividend (e_{T+n}), and the covariance for the future dividend (σ_{nm}) at time $T+n$ are respectively

$$\begin{aligned}
e_{T+n} &= \sum_{i=0}^{n-1} \psi_i \varepsilon_{T+n-i}, \\
\sigma_{nm} &= \operatorname{cov}(e_{T+n}, e_{T+m}) \quad n > m \geq 1 \\
&= \operatorname{cov}\,(\sum_{i=0}^{n-1} \psi_i \varepsilon_{T+n-i}\,, \sum_{j=0}^{m-1} \psi_j \varepsilon_{T+m-j}\,) \\
&= \sigma^2 \sum_{i=0}^{m-1} \psi_i \psi_{n-m+i}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbf{a}_N' \Sigma_{t,N} \mathbf{a}_N &= \sum_{m=1}^{N} \sum_{n=1}^{N} a^{n+m} (\sigma_{nm}) \\
&= \sigma^2 \sum_{m=1}^{N} \sum_{n=1}^{N} a^{n+m} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \psi_i \psi_j \delta_{n-i,m-j} \\
&= \sigma^2 \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{i=1}^{m} \sum_{j=1}^{n} a^{n+m} \psi_{i-1} \psi_{j-1} \delta_{n-i,m-j} \\
&= \sigma^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{m=j}^{N} \sum_{n=i}^{N} \psi_{i-1} \psi_{j-1} a^{n+m} \delta_{n-i,m-j} \\
&= \sigma^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{m=j}^{N} \psi_{i-1} \psi_{j-1} a^{i-j} \sum_{m=j}^{N} a^{2m} \\
&= \sigma^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{m=j}^{N} \psi_{i-1} \psi_{j-1} a^{i-j} \left(\frac{a^{2j} - a^{2N+2}}{1-a^2} \right) \\
&= \sigma^2 \sum_{i=1}^{N} \psi_{i-1} \sum_{j=1}^{N} a^{i-j} \o_j,
\end{aligned}$$

where

$$\o_j = \psi_{j-1} \left(\frac{a^{2j} - a^{2N+2}}{1-a^2} \right). \qquad \square$$