ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Master Specialisation Financial Economics

MEAN REVERSION IN STOCKS

A univariate linear time series analysis of book equity to market equity sorted U.S. portfolios

Author:	S. Kourdoupalos
EUR study number:	305213
Thesis supervisor:	Dr. P.J.P.M.Versijp
Finish date:	May 2009

ABSTRACT

This master thesis treats the question whether prices of book equity to market equity sorted stock portfolios follow a random walk. In the theoretical part of the thesis, the random walk model is derived as a testable expression of market informational efficiency assuming constant expected returns and its statistical tests, namely the regression beta and the variance ratio, are discussed. As alternatives to the random walk, the structural models of fads and time varying expected returns are presented. A new model which incorporates the characteristics of the previous two models is also proposed. In the empirical part, the random walk test statistics are estimated. The random walk is rejected for portfolios with low BE/ME ratio by the variance ratio. The observationally equivalent ARMA forms of the alternative models are estimated and an ARMA(2,2) process is found to fit better the data. Moreover, to measure the ability of the statistical tests to reject the random walk when the alternative models considered are true, the power of the variance ratio and regression beta is calculated. The power of the variance ratio is higher than the regression beta and it deteriorates exponentially with the return interval.

Keywords: Mean reversion, random walk, market efficiency, permanent/transitory components, structural model.

JEL classification: C22, G14

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 Introduction

1.1 Introduction to univariate time series models

In univariate time series analysis, there are two fundamental approaches of modeling [Kirchgässner and Wolters (2007, pp.3-4), Brockwell and Davis (2002, pp.23-24)]. The first approach assumes that time series are composed of different unobserved components; these are a trend, a seasonal, a cyclical[1] component and a noise residual. Seasonal components have a certain period, where the term period is the length of time to complete a full cycle, whereas cyclical components have a non-specific period. In this paper only cyclical components are considered. Early time series models used components which were deterministic functions of time. In particular, the models of economic time series simply used a deterministic time trend, plus noise. It has only been since the work of Nelson and Plosser (1982) that economic time series are modeled with a stochastic trend. Such models with stochastic components are also named structural, which obviously is because they impose structure to the examined time series.

The types of components mentioned so far and considered in this paper are trend/cycle and stochastic/deterministic. Depending on the effect that a change of a component has on a time series, components can also be classified into permanent/transitory (or temporary). Naturally, deterministic components are permanent. Concerning stochastic components, trends are permanent (non-stationary) and cycles are transitory (stationary). In modern literature, since components are considered to be stochastic, the term permanent component refers to the stochastic trend and the transitory component refers to the stochastic cycle.

The second modeling approach assumes that series are generated from a single stochastic process instead of a sum of processes. In short, the method is to difference the data until they become stationary and then fit an ARMA (autoregressive moving-average) process. Note that the two modeling approaches are related. The previously mentioned simple unobserved components models with either a deterministic or a stochastic trend, plus a residual noise, can be converted into stationary and have an ARMA representation. To do this the deterministic trend needs to be subtracted or first differences to be taken respectively. As a result, deterministic trend and stochastic trend processes are named trend stationary and difference stationary.

Beveridge and Nelson (1981, pp.154-158) shown how the two modelling approaches under certain restrictions are interchangeable. An ARMA process has an unobserved components representation with a stochastic trend, given by a random walk, a cyclical component and perfect correlation between the components' errors. Conversely, Nelson and Plosser (1982, p.155) proved that an unobserved components model of the previous form with arbitrary correlation between the two components' residuals may have one or more ARMA representations.

In financial economics, modeling stock prices as random walk used to be the convention. A good reasoning is that it can be derived under market informational efficiency assuming constant expected

---

[1] The difference between the terms seasonal and cyclical is subtle and in some texts not pointed out properly. In this text the terms are used as in Chatfield (1995, p.9).

returns [Fama (1976, pp.133-151)]. Within the framework discussed previously a random walk is a stochastic trend component. A trend given by a random walk is a permanent (non-stationary) component due to a property of random walk, which is, shocks are incorporated in the price level permanently. The mean reversion literature tries to identify transitory (stationary) components in stock prices and model them accordingly. A transitory component has the property to revert to its mean. Transitory components are of interest since it is more or less expected that stocks have an upward (stochastic) trend to compensate their holders with a return. The question of transitory components is whether there are predictable fluctuations above and below the trend.

1.2 The literature

The estimated components of structural models show the magnitude of each component in the series. Early papers that perform maximum likelihood estimation of structural models using a state space representation are Harvey (1985), Watson (1986), Clark (1987) and Conrad and Kaul (1988). After the permanent and transitory components are estimated, Stock and Watson (1988, p.157) and Conrad and Kaul (1988, p.417) regress the original series on each component to calculate the proportion of variance that each explains. This is a measure of the magnitude of each component which is simply equivalent to the ratio of each component's variance to the variance of the series.

ARMA models have no obvious method to give an indication about any possible components' magnitude. Campbell and Mankiw (1987a, 1987b) perform exact maximum likelihood estimation of ARMA models using a state space representation and measure the permanent effect of a shock using its impulse response function. This method can say if the permanent effect deviates from one, which would be its value under random walk, but it does not shed light to the issue of whether there is a transitory component and what its magnitude is. Still, using the Beveridge-Nelson decomposition, an ARMA can be converted into an unobserved components model. The impulse response function as measure of persistence is a parametric one and can be affected by a possible model misspecification. Poterba and Summers (1988, p.31) do not estimate ARMA processes because the estimated models can not be recovered when artificial data are generated by them.

A non-parametric measure of persistence is of Cochrane (1988, p.898) who developed the variance of long differences of a series to reveal the variance of its permanent component. Then the ratio of the permanent component's variance to the variance of the series indicates the magnitude of the permanent component. This ratio, called the variance ratio, was shown to be a function of sample autocorrelations. This latter interpretation of the variance ratio is used by Poterba and Summers (1988, p.30) as a statistic to test the random walk model, that is, whether sample autocorrelations are jointly different from zero. Lo and MacKinlay (1988, pp.45-50) derive a slightly different version of the variance ratio. Another test method of the random walk is of Fama and French (1988, p.249) who regress multiperiod returns on their first lag. Under the null hypothesis the regression coefficient should not be statistically different form zero. Jegadeesh (1991, p.1429) generalises the regression test by assessing different number of periods for the multiperiod return regressand and regressor.

Statistical tests might not be powerful enough to reject the random walk null hypothesis. Summers (1986, p.594) proposes the fads model as an alternative for which statistical tests lack power to accept when it is the true model. Among the tests some can be more powerful depending on the alternative hypothesis. Jegadeesh (1991, p.1430) and Richardson and Smith (1994, pp.384-392) perform asymptotic power comparisons, whereas Poterba and Summers (1988, pp.31-34), Lo and MacKinlay (1989, pp.425-435) finite power comparisons.

1.3 Motivation

The research question of this master thesis is whether there are mean reverting components in the prices of BE/ME sorted portfolios. The issues related with mean reversion are addressed using linear univariate methods. Of course, real world relations are multivariate and possibly nonlinear. Yet, univariate models can serve as a benchmark to multivariate models, that is, they have to perform at least as well as the univariate ones. Evidence against the random walk model of both univariate and multivariate analysis is disputable, see for example Kaul (1996, pp.284-286) for a review. In addition, structural models are capable of capturing nonlinear behaviour of series by including time varying parameters, for example a time varying drift.

To illustrate the foundation of the random walk model and why it can be considered as the null hypothesis when prices are modeled, it is derived as a testable expression of market efficiency. Other alternative models studied in the literature that imply a mean reverting behaviour for prices are the fads [Summers (1986, p.594)] and the time-varying expected returns models [Conrad and Kaul (1988, p.411)]. The thesis proposes a new composite model that incorporates the characteristics and tackles the problems of the aforementioned models. The problem related with the fads model is that it only implies negative first-order return autocorrelations for every return interval [Summers (1986, p.595), Jagadeesh (1991, p.1429)], whereas the observed first-order autocorrelations of short horizon returns are positive [Poterba and Summers (1988, p.37), Lo and MacKinlay (1988, p.52)]. It is shown that the theoretical return autocorrelation function of this composite model can account for the observed positive first order autocorrelations. Moreover, the fads and the time varying expected returns models are indistinguishable, with respect to which model is the true data generating process, since both models are observationally equivalent to an ARMA(1,1) process. The composite model has an ARMA(2,2) observationally equivalent process, which is shown to fit better the data than an ARMA(1,1).

The variance ratio as a test statistic and a function of sample autocorrelations has been widely used to detect deviation from random walk. The original variance ratio derivation of Cochrane (1988) interpreted it as a measure of persistence, which apparently attracted less interest. In this study, the interpretation as a measure of persistence is shown not to be clear since under observationally equivalent processes the ratio has a different mathematical expression. The variance ratio version of Lo and MacKinlay (1988) and the regression beta statistic of Fama and French (1988) are estimated for BE/ME sorted portfolios in comparison to the papers of Fama and French (1988), Lo and MacKinlay (1988), Conrad and Kaul (1988), Jagadeesh (1991) who apply their analyses to size sorted and industry portfolios. To consider the small sample properties of the test statistics, their sampling distribution under the null hypothesis of random walk is constructed using the resampling technique of randomisation applied in Kim, Nelson and Startz (1991, p.519).

The statistical tests might not be able to detect mean reversion in the time series. Finite sample power calculations are used to examine whether the variance ratio and the regression test have the power to reject the null hypothesis when the considered alternative models are correct. Poterba and Summers (1988) and Lo and MacKinlay (1988) perform power calculations using returns simulated by ARMA(1,1) models with arbitrarily chosen parameters. In this study the simulated returns are generated from the estimated ARMA models by resampling the estimated residuals. To take into account the uncertainty about the estimated parameters, a sensitivity analysis of the power calculations for plus/minus half the standard errors of the estimates is conducted.

Chapter 2 includes the theoretical analysis of the thesis. Section 2.1 derives the random walk model and illustrates its testable implications. In section 2.2, the concept of mean reverting component is presented and the properties of the alternative structural models are discussed. The random walk test statistics are presented in section 2.3. Chapter 3 includes the empirical analysis of the thesis. The dataset selection is discussed in section 3.1. Section 3.2 explains the methodology used, which includes bias corrections of the statistics, structural models estimation issues, diagnostics for model selection and the method of power calculation. In section 3.3, after the descriptive statistics of the series are analysed, the empirical results of the random walk tests, the alternative model estimation and the power calculations are presented. Chapter 4 summarises and concludes with respect to the findings of the empirical analysis.

CHAPTER 2 Theoretical models

2.1 Foundation of the random walk model

In this section the random walk model for prices is derived as a testable expression of informational market efficiency assuming constant expected returns. Efficiency is a sufficient condition for a testable expression that leads to the random walk model. The assumption of constant expected returns is necessary because expected returns are included in the testable expression and they are unobservable. Unavoidably the testable expression depends on a joint null hypothesis, that is, the market is efficient and expected returns are constant. After the random walk is derived, its implications are considered. Namely, returns are unpredictable and shocks have a permanent effect on prices.

2.1.1 Definition of market efficiency

Fama (1976, pp.133-151) provides a formulated and detailed definition of efficiency, which has become the standard definition in the literature, along with a testing methodology. To simplify the notation, the market efficiency is derived here for a single security, although the derived expressions would implicitly be valid for each security or portfolios of them. The complete information set available at time *t*-1 is I_{t-1} and the information set used by the market is I_{t-1}^m, with the latter being a subset of the former $I_{t-1}^m \subseteq I_{t-1}$. For every point in time, each information set determines a conditional probability density function of future prices $f_m(P_{t+\tau-1}|I_{t-1}^m)$ and $f(P_{t+\tau-1}|I_{t-1})$ respectively, where $P_{t+\tau-1}$ is the price of the security τ periods (τ=1,2,…) ahead of *t*-1. Prices are considered to include reinvested dividends. Informational efficiency, that is, prices fully reflect available information, implies that the market fully uses the complete information set to assess the probability density function of future prices, and thus

$$I_{t-1}^m = I_{t-1}. \tag{1}$$

This in turn implies that the density functions under the two information sets are equal,

$$f_m(P_{t+\tau-1}|I_{t-1}^m) = f(P_{t+\tau-1}|I_{t-1}). \tag{2}$$

It is now understood that the informational efficient market definition should be: *future* (expected) prices always fully reflect the complete current information. Efficiency is a necessary and sufficient condition for (2). Equality (2) implies that conditional on information at time *t*-1, expected values of the two densities are also equal,

$$E_m(P_{t+\tau-1}|I_{t-1}^m) = E(P_{t+\tau-1}|I_{t-1}) \tag{3}$$

or if from both sides current prices are subtracted and also divided by them, then the equivalent expression for the special case of τ=1 is

$$E(R_t|I_{t-1}) = E_m(R_t|I_{t-1}^m). \tag{4}$$

Where R_t denotes the return realized between *t*-1 and *t*. Efficiency is a sufficient condition for (3) but not a necessary one. If equality (3) is proved to be valid, (2) is not necessarily valid and the market is not necessarily efficient. If (3) is not valid, (2) is not valid and the market is inefficient. Within this setting, the efficiency hypothesis cannot be proved to be valid. Efficiency is testable using expression (3) as long as efficiency is the null hypothesis. This means that efficiency holds unless proved otherwise.

2.1.2 Testable expressions of market efficiency

The formulated definition of efficiency is not directly testable. Expression (1) is not testable because the complete information set I_{t-1} is unobservable. Expression (2) is not testable as well, because both the probability density functions are unknown. Known are the prices observed in the market, which are generated by the true density $f(P_{t+\tau-1}|I_{t-1})$. That is, the informational efficiency framework considers *observed* prices as indeed reflecting all available information. Still, the market-assessed density $f_m(P_{t+\tau-1}|I_{t-1}^m)$ and the prices generated by it are unknown. It is necessary to assume a specific equilibrium model to be valid, which will generate the market assessed future security prices or equivalently the market-assessed future security returns. The model used in the expected return predictability literature is the so-called martingale model with a constant drift[2]. Taking conditional expectations on both sides of the single period return definition (including dividends) yields the equilibrium model

$$E_m(R_t|I_{t-1}^m) = \frac{E_m(P_t|I_{t-1}^m) - P_{t-1}}{P_{t-1}}. \tag{5}$$

Its interpretation is that the market sets the expected value of the market assessed density, given current information, such as it satisfies the equilibrium condition. Assuming constant expected returns, the equilibrium becomes

$$E_m(R_t|I_{t-1}^m) = R. \tag{6}$$

From equations (4) and (6), if the assumption that the equilibrium model is valid is added to the null hypothesis of efficiency, the null hypothesis model in terms of returns is derived as $E(R_t|I_{t-1}) = R$ or as stochastic

$$R_t = R + \varepsilon_t. \tag{7}$$

The error term ε_t needs only to be serially uncorrelated and mean stationary. Even though the martingale model with constant drift implies efficiency[3], a joint null hypothesis is made, that is, the market is efficient and the chosen equilibrium model is correct. If the model with constant drift was the only one which can imply efficiency then the joint hypothesis would have reduced to a single hypothesis.

---

[2] For an introduction to martingales see e.g. Cuthbertson (1996, pp.102-104) and Campbell, Lo and MacKinlay (1997, pp.30-31).

[3] To see this first examine the martingale without a drift $E_m(P_t|I_{t-1}^m) = P_{t-1}$, which says that current information, as reflected in current prices, is already incorporated in expectations of future prices. Differently written as $P_t = P_{t-1} + \varepsilon_t$, it means that there are changes in future prices only if new information arises and these changes are unpredictable. Thus, future prices fully reflect available current information. The problem is that the simple martingale does not account for risk. This is solved including a constant drift as in (1) and the model becomes $P_t/(1+R) = P_{t-1} + \varepsilon_t$. The argument that future prices fully reflect available information is still valid.

However, there are also other models, such as martingale with time-varying drift or discounted dividend models which can imply efficiency.

It was shown that the null hypothesis is that the true density $f(R_t | I_{t-1})$ is mean stationary, i.e. its conditional expectation is constant for different I_{t-1} and *t*. The alternative hypothesis is that conditional expected returns vary with I_{t-1},

$$E(R_t | I_{t-1}) = a + \mathbf{b}'\mathbf{I}_{t-1} , \qquad (8)$$

where $\mathbf{b}'$ is the coefficient vector of the independent variables $\mathbf{I}_{t-1}$ constituting the information set, which in univariate analysis includes historical returns. A special case is when $\mathbf{I}_{t-1}$ includes only the return of the previous period,

$$E(R_t | I_{t-1}) = a + bR_{t-1} . \qquad (9)$$

Writing the expression as stochastic leads to what Fama (1991, p.1576) defines as test of return predictability or test of weak form efficiency, according to Fama’s (1970, p.388) terminology,

$$R_t = a + bR_{t-1} + \varepsilon_t . \qquad (10)$$

The error term ε_t only needs to be serially uncorrelated and mean stationary. Under efficiency, the estimated coefficient *b* is expected to be zero. A zero coefficient indicates that past information, as perceived by historical returns, does not explain future returns. If coefficient α is moved to the right hand side of equation (10), the interpretation of informational market efficiency is that past returns cannot predict abnormal returns. The testable expression (10) is an AR(1) model. If higher order lagged returns and error terms are added into (10) then the expression becomes an ARMA model. Again the coefficient estimates of the ARMA model, under the efficiency hypothesis, are not expected to be statistically different from zero.

2.1.3 The random walk model

The martingale model with constant expected returns can also lead to a testable expression of market efficiency formulated in terms of log prices, namely the random walk model. Equation (7) has an equivalent in terms of prices

$$P_t = P_{t-1}(1 + R) + \varepsilon_t . \qquad (11)$$

Taking the natural logarithm of (11), setting $\varepsilon_t^* = \ln(1 + \varepsilon_t)$ and $r = \ln(1 + R)$, which is the continuously compounded return, and using lower case letters for prices to denote their natural logarithms leads to

$$p_t = p_{t-1} + r + \varepsilon_t^* . \qquad (12)$$

The notation used for the error does not really matter. It can simply be written without the star superscript. The derived expression is the null hypothesis model in terms of log prices, which is a random walk with a drift *r*

$$p_t = p_{t-1} + r + \varepsilon_t . \tag{13}$$

Depending on the properties of the error term two forms of random walk are mainly distinguished (Campbell, Lo and MacKinlay, 1997 pp.31-33). In the first form the error term is IID (independently and identically distributed) noise with zero mean and constant variance: $\varepsilon_t \sim \mathrm{IID}(0, \sigma_\varepsilon^2)$. The second form has less restrictions, which are the same as those in (7), namely, errors are uncorrelated and mean stationary, i.e. white noise (WN): $\varepsilon_t \sim \mathrm{WN}(0, \sigma_{\varepsilon,t}^2)$. The second form is more involved since it allows non-linear dependencies between errors and heteroskedasticity. Depending on the advances of the literature either the first or the second form will be considered. Expression (13) can be written as

$$r_t = r + \varepsilon_t . \tag{14}$$

Returns are stationary since the error term is stationary and the random walk model is named (first-order) difference stationary process. It is apparent that (14) and (7) are equivalent. The only difference is that the former is formulated in terms of continuously compounded returns. The expression simply implies that returns are uncorrelated at all leads and lags

Equation (13) with recursive substitution becomes

$$p_t = p_0 + t\,r + \sum\nolimits_{i=1}^{t} \varepsilon_i . \tag{15}$$

The current price level includes all past shocks, which makes the term $\sum\nolimits_{i=1}^{t} \varepsilon_i$ a stochastic permanent (non-stationary) component. In addition, the long-term effect of a shock to the price is the shock's value itself, which again means shocks are incorporated in the price level permanently. Hamilton (1994, p.439) proves this as follows. Write a future price at time *t*+*τ* as a sum of price first differences

$$p_{t+\tau} = (p_{t+\tau} - p_{t+\tau-1}) + (p_{t+\tau-1} - p_{t+\tau-2}) + \ldots + (p_{t+1} - p_t) + p_t .$$

By taking the derivative of the conditional expectation of prices, it is shown that the long term sensitivity of future expected prices to a shock equals unity,

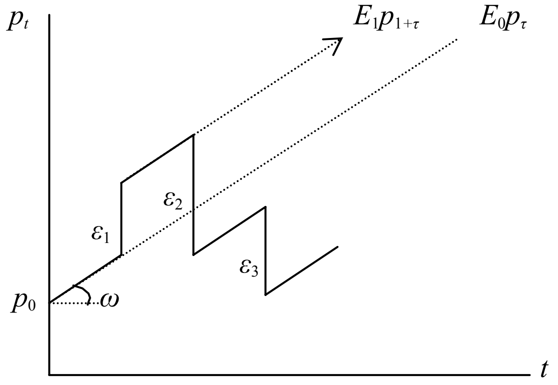
$$\lim_{\tau\to\infty} \frac{\partial E_t p_{t+\tau}}{\partial \varepsilon_t} = \lim_{\tau\to\infty} \frac{\partial}{\partial \varepsilon_t} (p_t + \tau r) = 1 . \tag{16}$$

The equation $E_t p_{t+\tau} = p_t + \tau r$ shows that any future expected price $E_t p_{t+\tau}$ depends on the realised current price, plus a deterministic component which is function of time, and thus $E_t p_{t+\tau}$ can be characterised as a trend.

To make the aforementioned concepts of random walk and permanent effect clearer, the course of prices as a function of time is examined in figure 1. For *t*=0 the trend line is $E_0 p_\tau = p_0 + r\tau$. The trend is a linear function of time with slope equal to $\tan\omega = r$ and vertical axis intercept p_0. From t_0 to t_1 the price increases by *r* to become $p_1 = p_0 + r$. Suppose there is at time t_1 an exogenous random innovation ε_1, the price then becomes $p_1 = (p_0 + \varepsilon_1) + r$. From this new price value, it is seen that the axis intercept is changed to $p_0 + \varepsilon_1$ although the slope remains constant. The same result applies for every subsequent innovation $\varepsilon_2, \varepsilon_3, \ldots$. In general, the price level is determined by $p_t = \left(p_0 + \sum_{i=1}^{t} \varepsilon_i\right) + rt$. Thus, innovations add up each period to the price level and change the vertical-axis intercept. The price level does not revert back to its original trend but it rather has a random course since the sum of the innovations is random. After the ε_1 shock is realized, its long term effect on prices for any τ period ahead is $E_1 p_{1+\tau} = (p_0 + \varepsilon_1) + r(1+\tau) = p_1 + r\tau$, which is a new trend line parallel to the original one. Thus, the effect of innovation ε_1 is considered permanent in the sense that prices are expected to develop along the new trend line determined by the innovation.

Figure 1 The course of prices under random walk

The figure shows the course of prices following a random walk as a function of time. Prices develop on a trend line given by their conditional expectation. The slope of the trend equals the drift rate $\tan\omega = r$. The response of prices to exogenous shocks ε_i and the subsequent change of long term trend τ periods ahead are illustrated.



Note that expression (14) is derived for single period returns. However, the uncorrelated returns implication can be generalised for multiperiod returns. Multiperiod returns will be examined throughout this paper since some transitory effects caused by factors like the business cycle may only be revealed in the long-run. Define a non-overlapping *k* period return of a prices sample $(p_0, p_1, ..., p_{T-1})$ and its τ period lead/lag using (13) as

$$r_{tk}(k) = \sum_{i=0}^{k-1} r_{tk-i} = p_{tk} - p_{(t-1)k}, \quad t = 1,2,\ldots,T-1, \quad k = 1,2,\ldots \tag{17}$$

$$r_{(t+\tau)k}(k) = \sum_{i=0}^{k-1} r_{(t+\tau)k-i} = p_{(t+\tau)k} - p_{(t+\tau-1)k}.$$

The complex subscript notation is necessary to distinguish between non-overlapping and overlapping multiperiod returns.[4] For the random walk, the non-overlapping *k* period return and its τ period lead/lag are

---

[4] The reader should not be confused by the lack of an operator between (functions of) subscript variables since it implicitly is the multiplication operator. Within this notational framework single period returns for *k*=1 are written as usually, that is, $r_{t+\tau}(1) = p_{t+\tau} - p_{t+\tau-1} = r_{t+\tau}$.

$$r_{tk}(k) = kr + \sum\nolimits_{i=(t-1)k+1}^{tk} \varepsilon_i, \qquad t = 1,2,\ldots,T-1, \qquad k = 1,2,\ldots \tag{18}$$

$$r_{(t+\tau)k}(k) = kr + \sum\nolimits_{i=(t+\tau-1)k+1}^{(t+\tau)k} \varepsilon_i .$$

The two expressions do not have common increments and thus multiperiod returns are uncorrelated for all leads and lags. Moreover, the expected value and variance of multiperiod returns are *k* times the single period returns. The problem with multiperiod returns is that the data observations become less as the number of periods increase. A solution to this is the use of overlapping data. The overlapping *k* period return equivalents are defined as

$$r_t(k) = \sum\nolimits_{i=0}^{k-1} r_{t+i} = p_t - p_{t-k} \tag{19}$$

$$r_{t+\tau}(k) = \sum\nolimits_{i=0}^{k-1} r_{t+\tau+i} = p_{t+\tau} - p_{t+\tau-k} .$$

For the random walk, the non-overlapping *k* period return and its *τ* period lead/lag are

$$r_t(k) = kr + \sum\nolimits_{i=t-k+1}^{t} \varepsilon_i \tag{20}$$

$$r_{t+\tau}(k) = kr + \sum\nolimits_{i=t+\tau-k+1}^{t+\tau} \varepsilon_i .$$

This time the two expressions do have common increments for $|\tau| \le k-1$ and thus overlapping multiperiod returns calculated from prices which follow a random walk are correlated. Using proper econometric techniques to take into account this correlation, overlapping observations have the advantage that the number of exploitable observations increases and thus the estimated sample statistics are more probable to converge to their asymptotic values.

2.2 Alternative models

In this section alternative models to the random walk are considered. A natural alternative to the random walk model is a model that includes shocks that have only a transitory stochastic effect. Such a model is a deterministic trend process. A more general model should include both a permanent and a transitory stochastic effect. Using the Beveridge-Nelson decomposition method it is shown that a general difference stationary process can be decomposed into a permanent and a transitory component.

Structural models with an economic rationale that decompose prices into a permanent and a transitory component need to be considered. Such are the fads and time-varying expected returns models. A new composite model which incorporates the characteristics of the two aforementioned models is proposed as another alternative.

2.2.1 Permanent and transitory components

Regarding the random walk model, it was discussed that shocks have a permanent effect on the series. Alternative models would include transitory components. A transitory component is in fact a stationary process. All mean stationary with finite variance time series processes are mean reverting, in the sense that the process fluctuates within limited amplitude and will tend to revert back to its expected value.

Considering this reversion, stationary processes have only a transitory stochastic effect on series. The speed of mean reversion is determined by the covariance function of the process. IID noise processes, because of their zero autocovariance, are instantaneously mean reverting. The random walk model contains an IID noise process and thus generates single and multiperiod returns which are instantaneously mean reverting[5]. In the case of covariance non-stationarity, the speed of mean reversion is time varying. An AR(1) zero mean process is covariance stationary and a measure of its mean reversion speed is given by the half-life index $\mathrm{H} = \ln(0.5)/\ln(\phi_1)$, where ϕ_1 is the autoregressive coefficient. The higher ϕ_1 is, the slower the mean reversion. This index measures how many periods will take for an effect of a shock to become half of its original value when there are no other shocks in the meanwhile. The half-life index can be generalised for higher order autoregressive processes [see e.g. Mark (2001, p.42)].

An alternative to the random walk is a model with deterministic trend and a stochastic error part which has only a temporary effect on the process. Such a stochastic process is named trend stationary and is represented as

$$p_t = tr + \psi(L)\upsilon_t, \qquad \upsilon_t \sim \mathrm{IID}(0,\sigma_\upsilon^2). \tag{21}$$

Where $\psi(L) = 1 - \psi_1 L^1 - \psi_2 L^2 - \psi_3 L^3 - \ldots$ is a lag polynomial of infinite order and L is is the lag operator with $L^i \upsilon_t = \upsilon_{t-i}$ and i integer. In order for the variance of the process to take a finite value, the convergence condition $\sum_{i=1}^{\infty} \psi_i^2 < \infty$ must be valid. This implies $\lim_{i\to\infty} \psi_i = 0$. This particular deterministic time trend is linear, but more complex ones can be in a polynomial form. The convergence condition implies

$$\lim_{\tau\to\infty} \frac{\partial E_t p_{t+\tau}}{\partial \upsilon_t} = 0. \tag{22}$$

Thus the long-term effect sensitivity of a shock to the price is zero. This model with a deterministic trend and errors having a transitory effect is the other extreme compare to random walk, where errors have a permanent effect. More involved models which can incorporate both permanent and transitory components are considered in the next sub-section.

A more general model is now considered which includes random walk as a special case for $\psi(L) = 1$,

$$p_t = p_{t-1} + r + \psi(L)\upsilon_t, \qquad \upsilon_t \sim \mathrm{IID}(0,\sigma_\upsilon^2), \tag{23}$$

with $\sum_{i=1}^{\infty} \psi_i^2 < \infty$ and $\lim_{i\to\infty} \psi_i = 0$. In (23) the difference of prices is generated by a stationary MA(∞) process. The effect of a shock to the long-run expected level of a price is weighted by [Hamilton (1994, p.442)][6]

---

[5] There is a common misconception about whether the alternative hypothesis tested against the null of random walk is prices being mean reverting or returns being mean reverting. Having in mind that returns under random walk are (instantaneously) mean reverting the latter does not make sense. Mean reversion as an alternative to random walk should refer only to prices.

[6] It is proved like (16), $\lim_{\tau\to\infty} \frac{\partial E_t p_{t+\tau}}{\partial \upsilon_t} = \lim_{\tau\to\infty} \frac{\partial}{\partial \upsilon_t}\left[p_t + \tau r + (\psi_\tau + \psi_{\tau-1} + \ldots + \psi_1)\upsilon_t + (\psi_{\tau+1} + \psi_\tau + \ldots + \psi_2)\upsilon_{t-1} + \ldots\right] = \psi(1)$.

$$\lim_{\tau\to\infty}\frac{\partial E_t p_{t+\tau}}{\partial \upsilon_t} = \psi(1). \tag{24}$$

In other words this is the cumulative response of prices to a unitary shock. The papers of Campbell and Mankiw (1987a p.112, 1987b p.861) use this measure to examine the permanence of a shock's effect.

2.2.2 Structural models

Structural models constitute of stochastic components. A popular model of this kind is of Beveridge and Nelson (1981 pp.154-158) who show that an ARMA(p,q) model of returns

$$r_t = p_t - p_{t-1} = \theta_0 + \theta_1 r_{t-1} + \ldots + \theta_{\mathrm{p}} r_{t-\mathrm{p}} + \upsilon_t - \zeta_1 \upsilon_{t-1} - \ldots - \zeta_{\mathrm{q}} \upsilon_{t-\mathrm{q}},$$

which has a MA(∞) representation (21) with $r = \dfrac{\theta_0}{1-\theta_1-\ldots-\theta_{\mathrm{p}}}$ and $\psi(L) = \dfrac{\zeta_{\mathrm{q}}(L)}{\theta_{\mathrm{p}}(L)}$, can be decomposed into a permanent and a transitory component (see appendix A):

$$\begin{aligned} p_t &= p_t^* + u_t \\ p_t^* &= p_{t-1}^* + r + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{IID}(0, [\psi(1)]^2 \sigma_\upsilon^2) \\ u_t &= \phi_0 \upsilon_t + \phi_1 \upsilon_{t-1} + \ldots, \qquad |\phi| < 1, \qquad \upsilon_t \sim \mathrm{IID}(0, \sigma_\upsilon^2). \end{aligned} \tag{25}$$

Where p_t^* is the permanent component following a random walk and u_t is the transitory component following an MA(∞) process. Stock and Watson (1988, p.171), assuming $p_0 = 0$, $\upsilon_t = 0$ for $t \le 0$, derive the transitory component as a finite-order MA process.

A structural model of the form (25) decomposes an observable time series into two unobserved components. Thus it can be named unobserved components model. A major characteristic of the unobserved component model derived by the Beveridge-Nelson decomposition is that the shocks affecting the permanent and transitory compoents are perfectly correlated by derivation. The Beveridge-Nelson decomposition leads to two interchangeable representations, i.e. ARMA and unobserved components. This one-to-one relationship between the two representations always exists.

The Beveridge-Nelson decomposition is restrictive as far as the dependence between the shocks is concerned. To avoid this restriction, alternatively a time series can be explicitly modelled as a sum of a permanent component (random walk with drift) and a transitory component with zero correlation between their errors. Such a model is named UC-ARMA[7]. UC stands for unobserved components and the ARMA term refers to the process of the transitory component. Nelson and Plosser (1982 pp.153-155) show that a UC-ARMA model of the form (25) with uncorrelated errors and a finite-order MA process as transitory component has an econometrically identifiable ARMA representation. If the correlation of the shocks is a priori unspecified, then the ARMA representation in unidentifiable in the sense that its parameter values are not unique.

---

[7] In this text, either this term or "unobserved components model with uncorrelated errors" shall be used to distinguish from unobserved components models with perfectly correlated errors.

The general case to transform an ARMA model to an unobserved components representation is for correlation values between zero and one, $\rho_{\varepsilon\upsilon} \in [0,1)$, excluding the value of 1 which is the case of the Beveridge-Nelson decomposition. Such transformation with arbitrary correlation between the errors is not guaranteed to exist[8] [Nelson and Plosser (1982, p.155), Watson (1986, p.53)] but if it exists then it might have more than one representations because the parameters are not unique. Even so, the innovation variance of the permanent component is always the same [Cochrane (1988, p.904)] and the permanent and temporary components equal the ones of the Beveridge-Nelson decomposition [Morley, Nelson and Zivot (2003, p.237)].

2.2.3 The fads model

Alternative models to the random walk need to be supported by a plausible economic interpretation. Summers (1986, p.594) proposed an alternative model for which correlation tests lack power to reject the random walk. Under this alternative model, actual prices deviate from the efficient market prices p_t^* by a pricing error term u_t which follows an AR(1) process. This process models the fads, i.e. a persistent pricing error that will decay slowly back to zero after some periods.

$$\begin{aligned} & p_t = p_t^* + u_t \\ & p_t^* = p_{t-1}^* + r + \varepsilon_t, \qquad \varepsilon_t \sim \text{IID}(0, \sigma_\varepsilon^2) \\ & u_t = \phi_1 u_{t-1} + \upsilon_t, \quad |\phi_1| < 1, \quad \upsilon_t \sim \text{IID}(0, \sigma_\upsilon^2) \end{aligned} \tag{26}$$

Model (26) assumes constant expected returns. Under the fads model, the autoregressive coefficient of the transitory component would be expected to be positive. This means positive (negative) pricing errors at a point of time are followed by subsequent positive (negative) pricing errors of a smaller magnitude. The state space representation of the model is

$$p_t = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} p_t^* \\ u_t \end{bmatrix}$$

$$\begin{bmatrix} p_t^* \\ u_t \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & \phi_1 \end{bmatrix} \begin{bmatrix} p_{t-1}^* \\ u_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ v_t \end{bmatrix}, \quad \begin{bmatrix} \varepsilon_t \\ v_t \end{bmatrix} \sim \text{IID}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_v^2 \end{bmatrix} \right),$$

which can be estimated using exact maximum likelihood.

Summers (1986, p.595) and Poterba and Summers (1988, p.32) show that the model (26) has the following reduced form

$$(1 - \phi_1 L)(1 - L) p_t = (1 - \phi_1) r + \varepsilon_t - \phi_1 \varepsilon_{t-1} + \upsilon_t - \upsilon_{t-1} \,. \tag{27}$$

The reduced form is not a regular ARMA process since the right-hand side is a sum of two different error processes. The sum of the two MA processes is however equivalent to a single MA process with order

---

[8] This argument is used in Cochrane (1988, p.904) although it has only been proved for $\rho_{\varepsilon\upsilon} = 0$ in Watson (1986, pp.52-53). In particular, only processes whose spectral density function has a global minimum at zero frequency have an unobserved components representation. For example, this restriction rules out processes with positive first-order autocorrelation at lag one.

equal to the bigger order of the two processes [see Hamilton (1994, p.106)]. Thus (27) has an observationally equivalent ARMA(1,1) representation

$$(1-\phi_1 L)(1-L)p_t = \phi_0 + \omega_t - \zeta_1 \omega_{t-1}, \qquad \omega_t \sim \text{IID}(0,\sigma_\omega).$$

The ARMA model can be estimated and then by equating the moments of the two equivalent processes the reduced form is identified. Notice that the autoregressive polynomials are the same and identifiable directly from the ARMA model. The unknowns of the reduced form are $r,\sigma_\varepsilon^2,\sigma_\upsilon^2,\sigma_{\varepsilon\upsilon}$, but there are only three equations to solve, namely the mean, the variance and the first-order autocovariance expressions. To overcome this identifiability problem Poterba and Summers (1988, p.32) assume $\sigma_{\varepsilon\upsilon}=0$. However, such an assumption restricts the autocovariance structure of the process (27) and as Lippi and Reichlin (1992, p.91) show it also restricts the cumulative response to $\psi(1)<1$. Morley, Nelson and Zivot (2003, p.237) show that there is still a way not to assume $\sigma_{\varepsilon\upsilon}=0$ and calculate the permanent and transitory components of model (26). It is shown that the Beveridge-Nelson decomposition can be performed to the estimated ARMA model and the permanent and transitory estimated components are actually the same with the ones form the state space representation estimates of (26).

Let us now illustrate the concepts of mean reverting and permanent components of the fads model. The reduced form of the fads model can be written as $p_t = p_{t-1} + r + \varepsilon_t + (1-\phi_1 L)^{-1}(1-L)\upsilon_t$. The long term effect sensitivities of the errors to prices are[9]

$$\lim_{\tau\to\infty}\frac{\partial E_t p_{t+\tau}}{\partial \varepsilon_t} = 1, \qquad \lim_{\tau\to\infty}\frac{\partial E_t p_{t+\tau}}{\partial \upsilon_t} = 0. \tag{28}$$

The course of prices as a function of time based on an example is examined in figure 2. Prices are a sum of a random walk, for which what was said in figure 1 is valid, and a mean reverting component, $p_t = \left(p_0^* + \sum_{i=1}^t \varepsilon_i + u_t\right) + rt$. Assuming the transitory component equals its unconditional mean at *t*=0, i.e. $u_0 = 0$, the trend line is $E_0 p_\tau = p_0 + r\tau$. From t_0 to t_1 the price increases by r to become $p_1 = p_0 + r$. Suppose there are at time t_1 two positively correlated[10] random innovations $\varepsilon_1, \upsilon_1$. The new price is $p_1 = \left(p_0^* + \varepsilon_1 + \upsilon_1\right) + r$. The long term effect on prices for several periods *τ* ahead is

$$E_1 p_{1+\tau} = \left(p_0 + \varepsilon_1 + u_{1+\tau}\right) + r(1+\tau) = \left(p_0 + \varepsilon_1 + \phi_1^\tau \upsilon_1\right) + r(1+\tau) \approx \left(p_0 + \varepsilon_1\right) + r(1+\tau)$$

which is the trend line of the permanent component seen in the random walk course of figure 1. Notice that the term $\phi_1^\tau \upsilon_1$ converges to zero for large *τ*. In other words, the effect of the transitory component

---

[9] Similarly to (16), $\lim_{\tau\to\infty}\frac{\partial E_t p_{t+\tau}}{\partial \varepsilon_t} = \lim_{\tau\to\infty}\frac{\partial}{\partial \varepsilon_t}\begin{bmatrix} p_t + \tau r + (\phi_1^{\tau-1}\upsilon_{t+1} + \phi_1^{\tau}\upsilon_t + ...)(1-L) \\ + (\phi_1^{\tau-2}\upsilon_{t+1} + \phi_1^{\tau-1}\upsilon_t + ...)(1-L) + ... + (\phi_1^{0}\upsilon_t + \phi_1^{1}\upsilon_{t-1} + ...)(1-L) \end{bmatrix} = 1$, and

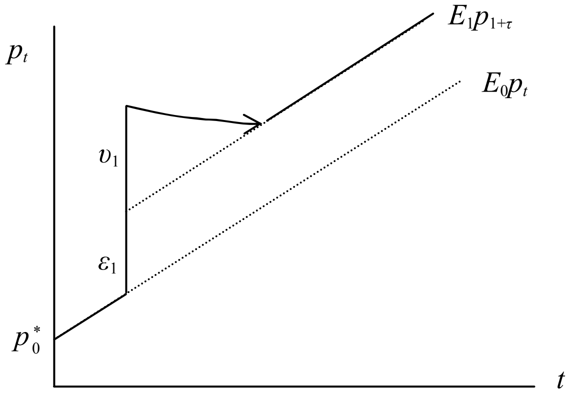
$$\lim_{\tau\to\infty}\frac{\partial E_t p_{t+\tau}}{\partial \upsilon_t} = \lim_{\tau\to\infty}\frac{\partial}{\partial \upsilon_t}\left[p_t + (\phi_1^\tau\upsilon_t - \phi_1^{\tau-1}\upsilon_t) + (\phi_1^{\tau-1}\upsilon_t - \phi_1^{\tau-2}\upsilon_t) + ... + (\phi_1^{1}\upsilon_{t-1} - \phi_1^{0}\upsilon_t)\right] = \lim_{\tau\to\infty}\frac{\partial}{\partial \upsilon_t}\phi_1^\tau\upsilon_t = 0$$

[10] The choice of the correlation sign does not really matter.

will gradually revert to its zero mean as time passes. Overall, the random walk error variance σ_ε^2 determines the amplitude of changes of the trend, the transitory component error variance σ_υ^2 determines the amplitude of the transitory deviations from the trend and the autoregressive coefficient ϕ_1 determines the rate of mean reversion to the trend.

Figure 2 The course of prices under the fads model

The figure shows the course of prices following the fads model as a function of time. Prices develop on a trend line given by their conditional expectation. The slope of the trend equals the drift rate *r*. The response of prices to the net effect of exogenous shocks of the trend ε_t and of the cyclical component υ_t at time *t*=1 is illustrated. The shock υ_1 is shown to have a transitory effect since prices converge to their long term trend several periods τ ahead.



2.2.4 The univariate time-varying expected return model

Conrad and Kaul (1988, p.411) suggest a model in which prices follow a random walk but with time-varying expected returns. The time-varying expected returns are modelled as a non-zero mean AR(1). The non-zero mean in the autoregressive process is added by Kaul (1996, p.274) to take into account the return compensation required by investors to hold stocks.

$$p_t = p_t^* \tag{29}$$

$$p_t^* = p_{t-1}^* + E_{t-1}(r_t) + \varepsilon_t, \qquad \varepsilon_t \sim \text{IID}(0,\sigma_\varepsilon^2)$$

$$E_{t-1}(r_t) = a_0 + a_1 E_{t-2}(r_{t-1}) + v_{t-1}, \qquad |a_1| < 1, \qquad v_t \sim \text{IID}(0,\sigma_v^2)$$

The model already is in a state space representation and can be estimated using exact maximum likelihood. Note that the two errors may covary, i.e. $\sigma_{\varepsilon v} \neq 0$.

Model (29) implies that returns has the reduced form

$$(1-a_1L)(1-L)p_t = a_0 + \varepsilon_t - a_1\varepsilon_{t-1} + v_{t-1}. \tag{30}$$

which again has an observationally equivalent ARMA(1,1). A problem related with the observationally equivalent representation is that both the fads and the time-varying returns models follow an ARMA(1,1), which in turn makes the two of them observationally equivalent. Even if the data do indeed follow an

ARMA(1,1), it can not be distinguished which of the two models, (26) and (29), is true. The reduced form is non-identifiable from the ARMA(1,1) model without assuming $\sigma_{\varepsilon v} = 0$, which imposes restriction to the autocovariance structure of the process (30). The model (29) is not in the form trend plus cycle and thus the result of Morley, Nelson and Zivot (2003, p.237) to estimate the components of the original model using the Beveridge-Nelson decomposition of the ARMA model is not applicable.

Let us now illustrate the concepts of mean reverting and permanent components of the time varying expected returns model. The reduced form of the fads model can be written as $p_t = p_{t-1} + \frac{a_0}{1-a_1} + \varepsilon_t + (1-a_1 L)^{-1}(1-L)v_t$. The long term effect sensitivities of the error terms of the two components are derived similarly to (16)[11]

$$\lim_{\tau\to\infty}\frac{\partial E_t p_{t+\tau}}{\partial \varepsilon_t} = 1, \quad \lim_{\tau\to\infty}\frac{\partial E_t p_{t+\tau}}{\partial v_t} = \frac{1}{1-a_1}, \quad \lim_{\tau\to\infty}\frac{\partial E_t r_{t+\tau}}{\partial v_t} = 0 \tag{31}$$

Interestingly, the long term effect sensitivity of error v_t for expected returns is zero, whereas for prices is either higher or lower than unity depending on the sign of a_1. The course of prices as a function of time based on an example is examined in figure 3. Assuming the mean reverting drift equals its unconditional mean at time t_0 the trend line is $E_0 p_t = p_0 + \frac{a_0}{1-a_1} t$. From t_0 to t_1 the price increases by the drift to become $p_1 = p_0 + \frac{a_0}{1-a_1}$. Suppose there are at time t_1 two positively correlated innovations ε_1, v_1. The new price at t_1 is $p_1 = p_0 + \frac{a_0}{1-a_1} + \varepsilon_1$ and the price at t_2 will be

$$p_2 = \left(p_0 + \frac{a_0}{1-a_1} + a_1 \frac{a_0}{1-a_1} + \varepsilon_1 + v_1\right) + a_0 .$$

After innovation v_1 materialises, the trend line of prices becomes mean reverting. The slope from t_1 to t_2 is simply $E_2(r_3) = p_2 - p_1 = a_0 + a_1 \frac{a_0}{1-a_1} + v_1$ and so on for the following time intervals. From τ to $\tau + 1$, the long term slope is

$$\lim_{\tau\to\infty} E_\tau(r_{1+\tau}) = \lim_{\tau\to\infty}\left[a_0\left(1 + a_1 + a_1^2 + \ldots + a_1^{\tau-1} + \frac{a_1^\tau}{1-a_1}\right) + a_1^{\tau-1} v_1\right] = \frac{a_0}{1-a_1},$$

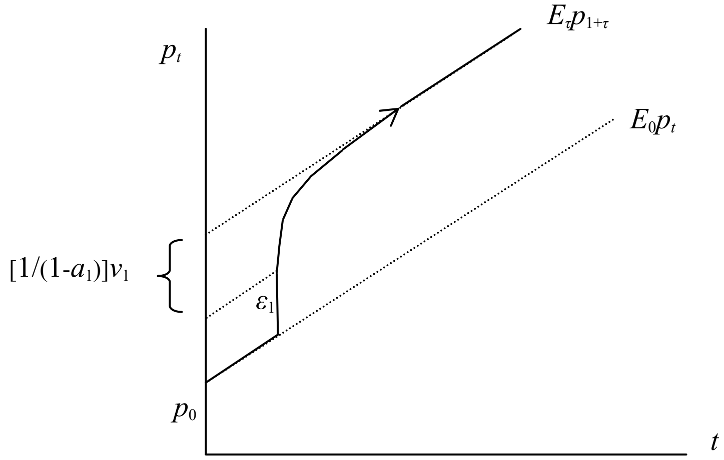
---

[11] $\lim_{\tau\to\infty}\frac{\partial E_t p_{t+\tau}}{\partial \varepsilon_t} = \lim_{\tau\to\infty}\frac{\partial}{\partial \varepsilon_t}\left[p_t + \tau\frac{a_0}{1-a_1} + (a_1^{\tau-1} v_t + a_1^{\tau} v_{t-1} + \ldots) + (a_1^{\tau-2} v_t + a_1^{\tau-1} v_{t-1} + \ldots) + \ldots + (a_1^0 v_t + a_1^1 v_{t-1} + \ldots)\right] = 1$,
$\lim_{\tau\to\infty}\frac{\partial E_t p_{t+\tau}}{\partial v_t} = \lim_{\tau\to\infty}\frac{\partial}{\partial v_t}(a_1^0 + a_1^1 + \ldots + a_1^{\tau-1})v_t = \frac{1}{1-a_1}$ and $\lim_{\tau\to\infty}\frac{\partial E_t r_{t+\tau}}{\partial v_t} = \lim_{\tau\to\infty}\frac{\partial E_t(p_{t+\tau} - p_{t+\tau-1})}{\partial v_t} = \lim_{\tau\to\infty}\frac{\partial}{\partial v_t} a_1^{\tau-1} v_t = 0$.

that is, it reverts back to its unconditional mean. Overall, the random walk error variance σ_{ε}^{2}, the transitory component error variance σ_{v}^{2} and the autoregressive coefficient a_1 determine the amplitude of changes of the trend. The autoregressive coefficient a_1 determines the rate of mean reversion of the drift.

Figure 3 The course of prices under the time varying expected returns model

The figure shows the course of prices following the time varying expected returns model as a function of time. Prices develop on a trend line given by their conditional expectation. The response of prices to the net effect of exogenous shocks of the trend ε_t and of the time varying drift v_t at time *t*=1 are illustrated. Shock v_1 is shown to have a transitory effect on expected returns and shocks v_1 and ε_1 have a permanent effect on prices.



2.2.5 A composite model

Since the fads and the time-varying returns interpretations are both plausible, consider a composite model that combines their characteristics,

$$p_t = p_t^* + u_t \tag{32}$$

$$p_t^* = p_{t-1}^* + E_{t-1}(r_t) + \varepsilon_t, \qquad \varepsilon_t \sim \text{IID}(0, \sigma_{\varepsilon}^{2})$$

$$E_t(r_{t-1}) = a_0 + a_1 E_{t-1}(r_{t-2}) + v_t, \qquad |a_1| < 1, \qquad v_t \sim \text{IID}(0, \sigma_{v}^{2})$$

$$u_t = \phi_1 u_{t-1} + \upsilon_t, \qquad |\phi_1| < 1, \qquad \upsilon_t \sim \text{IID}(0, \sigma_{\upsilon}^{2}).$$

The state space representation of the model is

$$p_t = \phi_1 p_{t-1} + \begin{bmatrix} 1 & -\phi_1 & 0 \end{bmatrix} \begin{bmatrix} p_t^* \\ p_{t-1}^* \\ E_t(r_{t-1}) \end{bmatrix} + \upsilon_t$$

$$\begin{bmatrix} p_t^* \\ p_{t-1}^* \\ E_t(r_{t-1}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ a_0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & a_1 \end{bmatrix} \begin{bmatrix} p_{t-1}^* \\ p_{t-2}^* \\ E_{t-1}(r_{t-2}) \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ v_t \end{bmatrix}, \qquad \begin{bmatrix} \upsilon_t \\ \varepsilon_t \\ v_t \end{bmatrix} \sim \text{IID}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\upsilon}^{2} & \sigma_{\varepsilon\upsilon} & \sigma_{v\upsilon} \\ \sigma_{\varepsilon\upsilon} & \sigma_{\varepsilon}^{2} & \sigma_{\varepsilon v} \\ \sigma_{v\upsilon} & \sigma_{\varepsilon v} & \sigma_{v}^{2} \end{bmatrix} \right),$$

which can be estimated using exact maximum likelihood.

The two middle processes of model (32) can be combined to give

$$p_t^* = p_{t-1}^* + \frac{a_0}{1-a_1} + \varepsilon_t + (1-a_1 L)^{-1} v_{t-1},$$

which is observationally equivalent to a first-order difference MA(∞). In other words, the trend of model (32) has an observationally equivalent difference stationary process of the form (21). This means that model (32) allows the permanent component to be a more general difference stationary process rather than a random walk. It can be shown that the composite model has the reduced form

$$\begin{aligned}[1-(a_1+\phi_1)L + a_1\phi_1 L^2](1-L)p_t &= (1-\phi_1)a_0 + [1-(a_1+\phi_1)L + a_1\phi_1 L^2]\varepsilon_t \\ &\quad + (1-\phi_1 L)v_{t-1} + [1-(1+a_1)L + a_1 L^2]\upsilon_t,\end{aligned} \tag{33}$$

which is observationally equivalent to an ARMA(2,2) process

$$(1-\theta_1 L - \theta_2 L^2)(1-L)p_t = \theta_0 + \omega_t - \zeta_1 \omega_{t-1} - \zeta_2 \omega_{t-2}, \qquad \omega_t \sim \text{IID}(0,\sigma_\omega).$$

When the ARMA(2,2) is estimated the coefficients of the original AR polynomial can be found from $\theta_1 = a_1 + \phi_1, \theta_2 = a_1\phi_1$. The reduced form is not identifiable since the autocovariances are non-zero up to the second order, which including the variance and mean implies four equations, but there are seven unknowns $a_0, \sigma_\varepsilon^2, \sigma_\upsilon^2, \sigma_\upsilon^2, \sigma_{\varepsilon v}, \sigma_{e\upsilon}, \sigma_{v\upsilon}$. Oh and Zivot (2006, p.10) propose to either impose three extra conditions, obviously $\sigma_{\varepsilon v}, \sigma_{e\upsilon}, \sigma_{v\upsilon} = 0$, or modify the original model by increasing the lag order of the fads component so that the number of non-zero autocovariances suffices to solve for the unknowns. In this case the model has an observationally equivalent ARMA(5,5) process.

As suggested in Nelson and Plosser (1982, p.154), before adopting an unobserved components model, it should be examined whether the theoretical autocorrelation function coincides with the observed one. In what follows, Zhou and Qing (2000, pp.526-529) apply a similar analysis in a multivariate framework. The autocorrelation function of the composite model can be shown to be

$$\rho(\tau) = \left[\frac{a_1^\tau}{1-a_1^2}\sigma_v^2 + \frac{\left(2\phi_1^\tau - \phi_1^{\tau+1} - \phi_1^{\tau-1}\right)}{1-\phi_1^2}\sigma_\upsilon^2\right] \Big/ \left[\frac{1}{1-a_1^2}\sigma_v^2 + \sigma_\varepsilon^2 + \frac{2}{1-\phi_1^2}\sigma_\upsilon^2\right], \tag{34}$$

and the first order autocorrelation is

$$\rho(1) = \left(\frac{a_1}{1-a_1^2}\sigma_v^2 - \frac{1-\phi_1}{1+\phi_1}\sigma_\upsilon^2\right) \Big/ \left(\frac{1}{1-a_1^2}\sigma_v^2 + \sigma_\varepsilon^2 + \frac{2}{1-\phi_1^2}\sigma_\upsilon^2\right). \tag{35}$$

An important issue concerning the observed first-order autocorrelations of short horizon returns (or single period returns) is that they are positive and the fads model is unable to explain them [Poterba and Summers (1988, p.37), Lo and MacKinlay (1988, p.56)]. For the composite model, the sign of the autocorrelation coefficient is assessed for $0 < \phi_1 < 1$ and $0 < a_1 < 1$. The motivation for this selection is as follows. Whether ϕ_1 is positive or negative is indifferent to the sign of the autocorrelation coefficient. So ϕ_1 is assumed positive. Besides, a positive ϕ_1 characterizes the fads. If a_1 is negative, then the autocorrelation coefficient is restricted to be always negative, which is of no interest since first order autocorrelation coefficients are observed to be positive. So a_1 is also assumed positive. Now, notice that the denominator of (35) is always positive. To assess the sign of the numerator of (35) the conditions $0<\phi_1<1$ and $0<a_1<1$ are used to prove that the bounds of the numerator are[12]

$$-\sigma_\upsilon^2 < \frac{a_1}{1-a_1^2}\sigma_v^2 - \frac{1-\phi_1}{1+\phi_1}\sigma_\upsilon^2 < \frac{1}{1-a_1}\sigma_v^2 .$$

The upper bound is positive and the lower bound is negative. The issue is when the numerator of (35) is positive and when negative. The factor $a_1/(1-a_1^2)$ is positive and the factor $-(1-\phi_1)/(1+\phi_1)$ is negative under the specifications of the model. Since the positive factor increases with a_1, i.e. $\frac{d}{da_1}\left[a_1/(1-a_1^2)\right]>0$, and the negative factor decreases with ϕ_1, i.e. $\frac{d}{d\phi}\left[(1-\phi_1)/(1+\phi_1)\right]<0$, $\rho(1)$ will be positive for sufficiently large a_1 and ϕ_1. The conclusion is that if the fads' pricing errors and the time-varying returns are generated by slowly mean reverting processes, it would be expected to observe a positive fist-order single period return autocorrelation.

2.3 Random walk tests

In this section the random walk is tested exploiting the implication of the null hypothesis that returns are uncorrelated. The most obvious way to test this implication is to estimate the sample autocorrelations and see whether they are statistically different from zero. More elaborate test statistics like the regression beta coefficient and the variance ratio are joint tests of the sample autocorrelations.

2.3.1 The regression beta coefficient

The regression test of Fama and French (1988, p.249), which was initially developed to support the alternative of the fads model being true, exploits the fact that when prices follow a random walk multiperiod non-overlapping returns should be uncorrelated. A non-overlapping *k* period return is regressed on its first lag

$$r_{tk}(k) = a + b(k) r_{(t-1)k}(k) + \varepsilon_t, \qquad k = 2,3,\dots.$$

The ordinary least squares beta coefficient is given by

$$b(k) = \frac{\operatorname{cov}\left[r_{tk}(k), r_{(t-1)k}(k)\right]}{\operatorname{var}\left[r_{tk}(k)\right]}.$$

For a sample of prices $(p_0, p_1, \dots, p_{T-1})$ the estimated coefficient is

---

[12] Using $0<a_1<1$ and $0<1-a_1^2<1$ leads to the first inequality $0<\frac{a_1}{1-a_1^2}<\frac{1}{1-a_1^2}$. Then $0<\phi_1<1$ can be used to show the second inequality $\frac{-1}{1+\phi_1}<-\frac{1-\phi_1}{1+\phi_1}<0$, where $-1<\frac{-1}{1+\phi_1}$. Multiply the first inequality with σ_v^2 and the second with σ_υ^2 and sum them.

$$b(k) = \frac{\sum_{t=1}^{(T-1)/k}\left[\left(\sum_{i=0}^{k-1} r_{tk-i}(k) - k\mu\right)\left(\sum_{i=0}^{k-1} r_{(t-1)k-i}(k) - k\mu\right)\right]}{\sum_{t=1}^{T}\left[\left(\sum_{i=0}^{k-1} r_{tk-i}(k) - k\mu\right)^{2}\right]} \tag{36}$$

$$\mu = \frac{1}{T-1}\sum_{t=1}^{T-1} r_t \, . \tag{37}$$

If expression (36) is developed and the numerator and denominator divided by the sample variance, then the regression coefficient can be written as a function of sample autocorrelations of single period returns.

$$b(k) = \frac{\rho(1) + 2\rho(2) + \ldots + k\rho(k) + (k-1)\rho(k+1) + \ldots + \rho(2k-1)}{k + 2\left[(k-1)\rho(1) + (k-2)\rho(2) + \ldots + \rho(k-1)\right]} \tag{38}$$

$$= \frac{\sum_{i=1}^{2k-1} \min(i, 2k-i)\rho(i)}{k + 2\sum_{i=1}^{k-1}(k-i)\rho(i)} .$$

In small samples the statistic is biased and a finite sampling distribution under the null hypothesis is simulated to test its significance.

Jegadeesh (1991, p.1429) generalises the regression test of Fama and French (1988, p.249) by assessing different number of periods for the multiperiod return regressand and regressor. In his paper it is suggested, using the approximate slope criterion of Geweke (1981), that single period return as a regressand produces a more powerful test asymptotically compare to tests with multiperiod return regressands.

2.3.2 The variance ratio

The variance ratio was initially developed by Cochrane (1988, p.898). His measure is the variance of long differences and it identifies the magnitude of the permanent component in a time series. The long differences of prices refer to their *k*-th order difference as *k* tends to infinity. The measure exploits the behaviour of the variance of multiperiod returns under models with stochastic trend and deterministic trend. Its algebraic form is

$$\lim_{k\to\infty} V(k) = \lim_{k\to\infty} \frac{\operatorname{var}[r_{tk}(k)]}{k} = \lim_{k\to\infty} \frac{\operatorname{var}(p_{tk} - p_{(t-1)k})}{k} . \tag{39}$$

The ratio of the variance of long differences to the variance of the series indicates the magnitude of the permanent component.

$$VR(k) = \frac{\lim_{k\to\infty} V(k)}{\operatorname{var}(r_t)} = \lim_{k\to\infty} \frac{\operatorname{var}[r_{tk}(k)]}{k \operatorname{var}(r_t)} . \tag{40}$$

Under the random walk and using (18), the variance of long differences is $\operatorname{var}[r_{tk}(k)] = k\sigma_{\varepsilon}^{2}$ and thus

$$VR(k) = \lim_{k\to\infty} \frac{k\sigma_{\varepsilon}^{2}}{k \operatorname{var}(r_t)} = \frac{\sigma_{\varepsilon}^{2}}{\operatorname{var}(r_t)} .$$

Under the trend stationary process (23), the variance of long differences is shown in appendix B (first derivation) to be given by $\text{var}[r_{tk}(k)] = 2\left[\psi(L)\right]^2 \sigma_\upsilon^2$ and thus in this case

$$VR(k) = \lim_{k\to\infty} \frac{2\left[\psi(L)\right]^2 \sigma_\upsilon^2}{k\,\text{var}(r_t)} = 0 .$$

In the case of a first-order difference stationary process (21), the variance ratio is (see appendix B, second derivation)

$$VR(k) = \lim_{k\to\infty} \frac{k[\psi(1)]^2 \sigma_\upsilon^2}{k[\psi(1)]^2 \sigma_\upsilon^2} = 1 .$$

It would be expected to find the same result for the observationally equivalent models of the difference stationary process. However, it is shown in appendix B (third derivation) that under the Beveridge-Nelson decomposition the variance ratio is

$$VR(k) = \lim_{k\to\infty} \frac{[\psi(1)]^2 k\sigma_\upsilon^2 + 2[\phi(L)]^2 \sigma_\upsilon^2 + 2\psi(1)\phi(1)\sigma_\upsilon^2}{k\left[[\psi(1)+\phi(1)]^2 + [\phi(1)]^2\right]\sigma_\upsilon^2} = \frac{[\psi(1)]^2}{[\psi(1)+\phi(1)]^2 + [\phi(1)]^2} .$$

Similarly, it is shown in appendix B (fourth derivation) that under a structural model of the form (25) with arbitrary correlation between the errors, the variance ratio has not the same algebraic expression. Hence, the interpretation of the variance ratio as a measure of the magnitude of the permanent component is not clear. Cochrane proves that the variance ratio is a function of sample autocorrelations of single period returns.

$$VR(k) = \frac{\text{var}[r_{tk}(k)]}{k\,\text{var}(r_t)} = 1 + 2\sum_{i=1}^{k-1} \frac{k-i}{k}\rho(i) \tag{41}$$

This interpretation of variance ratio is which Poterba and Summers (1988, p.30) use as a statistic to test whether sample autocorrelations are jointly different from zero.

Lo and MacKinlay (1988 pp.45-50) provide another derivation of the variance ratio statistic and derive its asymptotic properties. Their version of the variance ratio is defined as

$$VR(k) = \frac{\text{var}[r_{tk}(k)]}{\text{var}(r_t)}, \quad k = 2,3,\dots . \tag{42}$$

The sample variance, when the sample size[13] of prices is T $(p_0, p_1, \dots, p_{T-1})$, is

$$\text{var}[r_{tk}(k)] = \frac{1}{T-1}\sum_{t=1}^{(T-1)/k} (p_{tk} - p_{(t-1)k} - k\mu)^2 , \tag{43}$$

where μ is the average single period return as in (37). The variance ratio statistic will follow asymptotically

$$\sqrt{T-1}VR(k) \sim N(1, 2(k-1)) .$$

If overlapping returns are to be used and correct for the small sample bias, then the sample variance becomes

---

[13] Because $(T-1)/k$ in the k period return variance must be integer, in practice the endpoints of the sample size T are adjusted accordingly during estimation.

$$\operatorname{var}[r_{tk}(k)] = \frac{1}{k(T-k)\left(1-\dfrac{k}{T-1}\right)} \sum\nolimits_{t=k}^{T-1} (p_t - p_{t-k} - k\mu)^2, \qquad k = 2,3,\ldots. \tag{44}$$

In this case the variance ratio statistic will follow asymptotically

$$\sqrt{T-1}VR(k) \sim N\left(1, \frac{2(2k-1)(k-1)}{3k}\right).$$

A heteroskedastic consistent variance of the sample statistic is derived using the fact that the variance ratio can be written as a function of sample autocorrelations.

$$\sigma^2_{VR(k)} = \sum_{i=1}^{k-1}\left[\left[\frac{2(k-i)}{k}\right]^2 \frac{\sum\nolimits_{t=i+1}^{T-1}(p_t - p_{t-1} - \mu)^2 (p_{t-i} - p_{t-i-1} - \mu)^2}{\left[\sum\nolimits_{t=1}^{T-1}(p_t - p_{t-1} - \mu)^2\right]^2}\right], \qquad k = 2,3,\ldots. \tag{45}$$

For its detailed derivation see Lo and MacKinlay (1988, pp.48-50).

CHAPTER 3 Empirical analysis

3.1 Data

The dataset used is 5 BE/ME sorted portfolios' monthly value weighted returns (including reinvested dividends) of NYSE-AMEX-NASDAQ stocks from the CRSP database, which are available online by K. French. The CRSP U.S. stock returns database contains high quality data that span the period 1926:07-2008:07. A simple argument in favour of the choice of value weighted returns is stated in Fama and French (1993, p.10); value weighted portfolio returns correspond to realistic investment opportunities.

Portfolios of stocks are examined because portfolio returns have lower variance than individual stocks, which means statistical tests have more power to detect deviations from random walk. This argument is brought forth in Hawawini and Keim (1994, p.31). Besides, individual stocks contain idiosyncratic noise [Lo and MacKinlay (1988, p.56)]. In particular, the choice of the BE/ME sorted portfolios is because so far the literature has tested size sorted and industry portfolios. The intention is to identify transitory components related to BE/ME. The choice of 5 sorted portfolios is made for no other reason than trying to keep the results as easily presentable as possible.

A problem that comes with sorted portfolios is that their composition changes since stocks move across quantiles over time. An analysis which intends to identify mean reversion in such portfolios can be flawed [Fama and French (1988, p.252)]. However, the average cross-correlation of yearly BE/ME of the considered portfolios is 0.90 which means that the relative BE/ME of portfolios does not change much. It can be concluded that the composition of portfolios is relatively stable. It would have been ideal to examine mean reversion in size and BE/ME sorted portfolios so that the two possible factors, size and value, contributing to mean reversion are examined separately[14]. However, the compositions of double-sorted portfolios show lower stability. The average cross-correlation of yearly size of the considered portfolios is 0.37.

The minimum return period considered is a month. This return interval is chosen to avoid microstructure effects and noise of daily or weekly data. From these monthly portfolio returns, longer period returns of overlapping observations are constructed. The reason for choosing longer periods rather than single is because the testable implications of the random walk model are also applicable to multiperiod returns. In addition, it might be the case that some transitory effects caused by factors like the business cycle may only be revealed in the long-run. A problem with multiperiod returns is that the data observations become less as the number of periods increase and thus estimates do not have the desirable asymptotic properties. To increase the asymptotic consistency of the estimates, overlapping observations are constructed.

Returns that include reinvested dividends are used. If this is not the case, log prices do not follow a martingale [see Cuthbertson (1996, p.103)] and to derive the random walk model under informational efficiency prices are needed to follow a martingale. The original return data are converted into

---

[14] The paper that examines size and BE/ME sorted portfolios is of Ho and Sears (2006) but no provision for the portfolio composition changes is taken.

continuously compounded returns. Equivalently, prices are converted into log prices[15]. Taking the log of prices is a standard method that makes exponential growth of prices linear and eliminates the increasing variability of prices [Tsay (2005, p.73)]. To explain the latter note that, although the variance of the data generating process may be stable, the magnitude of oscillations increases as the level of prices increase. Additionally, taking logs allows multiperiod returns to be sums of single year returns which simplifies the time series analysis [Campbell, Lo and MacKinlay (1997, p.11)].

3.2 Methodology

The methodology to treat the question whether stock prices follow a random walk includes three parts. First, the regression beta and variance ratio statistics are estimated to test the null hypothesis of prices following a random walk against the alternative hypothesis of prices not following a random walk. Second, If the random walk model is rejected the next step would be to estimate alternative (not in the strict statistical sense of hypothesis testing) models and see whether they fit adequately the data. Even if the random walk is not rejected by the statistical tests, it is interesting to examine whether they actually are powerful enough to reject the null hypothesis. To do so, artificial data are generated by the so called alternative models and the percentage of random walk rejections by the statistical test is calculated.

The first part of the methodology involves the random walk tests or, differently put, tests of whether autocorrelations are separately or jointly different from zero. The sample return autocorrelations are estimated with the usual simplified formula [see Chatfield (1995, pp.19-20)]. Under the random walk, the τ-th order standardised sample autocorrelation coefficient $\rho(\tau)$ should follow

$$\sqrt{T}\rho(\tau)+(T-\tau)^{-1} \sim N(0,1) .$$

Note that the there is a negative bias since $E[\rho(\tau)]=-(T-\tau)^{-1}$ is different from zero. The regression beta and variance ratio are functions of return autocorrelations and thus this bias should be taken into account. To adjust for this bias regarding the regression beta, the expected value of the statistic calculated from (38) is added to the sample estimate. The bias correction of the variance ratio is already incorporated in (44).

To avoid a finite sample bias, the significance of the sample estimates is tested using sampling distributions constructed by calculating 2000 times the statistics from randomised (shuffled) return series. A way to simulate the statistic distribution under random walk followed by Fama and French (1988) is by drawing random errors from a normal distribution and calculating returns with (14). This is problematic for two reasons. The parameters of the normal distribution are chosen arbitrarily and it is implied that returns follow a normal distribution. If returns do not actually follow a normal distribution, the calculated sampling distribution is wrong. The method used in Kim, Nelson and Startz (1991) to construct the sample distribution is by the resampling technique of randomisation. With randomisation, the original data are shuffled to destroy any dependence between the observations. As a result, the data can be

---

[15] The prices are constructed from the return series by arbitrarily setting the initial price equal to a monetary value of 10.

considered as if they were generated by a random walk. In our case, the construction of the sampling distribution is based on 2000 shuffles.

The second part of the methodology involves the estimation of the alternative models. The alternative models were not possible to be estimated using the exact maximum likelihood method. Identification problems of state space models are usual [see e.g. Hamilton (1994, pp.387-388)]. The models are not identifiable even when the cross-covariances of errors are restricted to zero. The identification problem practically means that while the parameter values change, the log likelihood function cannot improve. In other words, different sets of parameter values produce the same log likelihood value and thus the optimisation algorithm cannot find the direction to converge to the maximum. The optimisation algorithm of Berndt, Hall, Hall, and Hausman, and that of Marquardt, available in EViews were both used. The initial parameters values were the ones specified by EViews since there is no prior information about their values. As a result, the ARMA models are estimated with the conditional maximum likelihood method and the structural models parameters are calculated by equating their covariances with those of the ARMA. The closed form solutions of the parameters for the fads model are

$$r = \frac{\phi_0}{1-\phi_1}, \quad \sigma_\varepsilon^2 = \left(\frac{1-\zeta_1}{1-\phi_1}\right)^2 \sigma_\omega^2, \quad \sigma_\upsilon^2 = \left[\zeta_1 - \phi_1\left(\frac{1-\zeta_1}{1-\phi_1}\right)^2\right]\sigma_\omega^2$$

and for the time varying expected returns are

$$a_0 = \phi_0, \quad a_1 = \phi_1, \quad \sigma_\varepsilon^2 = \frac{\zeta_1}{\phi_1}\sigma_\omega^2, \quad \sigma_v^2 = \left[1+\zeta_1^2 - \frac{\zeta_1(1+\phi_1^2)}{\phi_1}\right]\sigma_\omega^2 .$$

The composite model has two sets of closed form parameter solutions due to a quadratic equation. They are found using MATLAB but they are not reported due to their length. In some cases the estimated ARMA parameters imply a negative error variance for a structural model. When this happens, the ARMA model does not have an observationally equivalent structural model [Morley, Nelson and Zivot (2003, p.237)].

The diagnostics used to see which of the two ARMA processes fits the data best is the Schwarz criterion, the *Q* statistic and the autocorrelation function. The Schwarz criterion is chosen on the basis of its consistency and its formula is

$$SC = \frac{2(c\log T - \log L)}{T}$$

where *c* is number of estimated parameters and *L* is the likelihood value. The smaller the criterion value, the better the fit. The *Q* statistic is a joint test of the significance up to the τ-th order of the residual autocorrelations and its finite sample adjusted version is given by

$$Q(\tau) = T(T+2)\sum_{i=1}^{\tau}\left[\rho^2(i)/(T-i)\right]$$

following asymptotically a χ_τ^2 distribution. The *Q* statistic is estimated for arbitrarily chosen value of τ, which in practice is 12 or 15. The Jarque-Bera statistic is used to test the normality of the estimated residuals of the models. The *JB* statistic follows asymptotically a χ_2^2 distribution and its formula is

$$JB = \frac{T}{6}\left(S^2 + \frac{(K-3)^2}{4} \right),$$

where S is the skewness and K the kurtosis of the series. A limitation with the estimation of the alternative models is that diagnostics like the Schwarz criterion, the Q statistic and the autocorrelation function pattern can say which model fits better than another. However, unless the time series data have a clear pattern, it is difficult to accurately specify which the data generating process is.

The third part of the methodology involves the calculation of the test statistics power when the true model is one of the alternatives. The alternative ARMA models simulate 2000 return series by randomising the estimated residuals. The randomisation method is chosen to avoid assuming the estimated residuals are normally distributed when they are actually not. After the returns are simulated and the statistics calculated, it is calculated in what percentage of the 2000 repetitions the statistical tests can actually reject the random walk. Since there is some uncertainty about the parameter estimates of the ARMA models, a sensitivity analysis of the power calculations with the original estimates plus/minus 0.5 their standard error is conducted. Using the resampled residuals, the estimated standard error of the residuals is implicitly kept fixed.

3.3 Empirical results

3.3.1 Descriptive statistics

The descriptive statistics of the data can be a preliminary guide towards the identification of the data generating process. The data can be analysed as returns or prices. When the data are in the form of returns their sample moments can describe the distribution which the realised returns have been drawn from. Table 1 exhibits the sample moments up to fourth order and the Jarque-Bera normality test statistic of the 5 sorted portfolios.

Table 1 Descriptive statistics of returns

The table reports the sample moments up to fourth order and the Jarque-Bera statistic (observed significance level in parenthesis) of the 5 BE/ME sorted portfolio return series.

Portfolio	1	2	3	4	5
Mean	0.009	0.009	0.010	0.012	0.013
Variance	0.003	0.003	0.003	0.005	0.006
Skewness	-0.051	-0.034	0.938	2.023	1.530
Kurtosis	8.264	8.918	18.171	26.175	18.783
JB	1137 (0.000)	1437 (0.000)	9590 (0.000)	22714 (0.000)	10607 (0.000)

The 5 sorted portfolios are in ascending order of their BE/ME value. In other words, portfolio 1 includes stocks with the lowest BE/ME ratio which are named growth stocks and portfolio 5 includes the ones with the highest BE/ME ratio, named value stocks. The stocks in portfolios 2 and 4 can also be characterised as growth and value respectively, of course, at a lesser degree than the two extremes. The most obvious characteristic to notice in table 1 is that all portfolios do not follow a normal distribution. The two growth portfolios have quite similar moments. They both are negatively skewed and have kurtosis. In comparison, the rest of the portfolios are positively skewed and have much higher kurtosis. The moments of the two value portfolios, although not similar, are close and certainly distinctive from the rest. The middle portfolio seems to resemble more to the growth stocks with respect to the mean and variance, and to the value stocks with respect to the skewness and kurtosis.

The descriptive statistic used in univariate time series analysis to capture the linear dynamics of the data is the autocorrelation function. Panel A of figure 4 shows the autocorrelation functions of the 5 sorted portfolios[16]. In general, when sample statistics are estimated, it is expected to find a percentage of them outside the imposed significance interval of the sampling distribution under the null hypothesis. For example, if 36 autocorrelations are estimated for a single portfolio, it would be expected to find on average 3.6 (rounded is 4) significant autocorrelations for a two-sided significance level of 10%. Moreover, not all significant autocorrelations are important. Usually those at low lead/lag are more important since they may be more meaningful by means of having an economic interpretation. For example, if returns have a significant autocorrelation at lead 15, an economic rationale would be difficult to find. If returns have a significant positive first order autocorrelation, it may be attributed to fads. In panel A, there is no clear seasonality in returns since series with seasonal fluctuations in regular fixed intervals are expected to show oscillations in the autocorrelation function at the corresponding regular lags. The autocorrelation function has resemblances across all portfolios but the similarities are more for portfolios with similar BE/ME value. All portfolios have a positive significant first order autocorrelation and for value portfolios it is higher than growth portfolios. At lead/lag three, the value portfolios have again more profoundly significant autocorrelations. For subsequent lead/lags the similarities within the two groups of portfolios, value and growth, continue. For all portfolios, more significant autocorrelations are found than it is expected for 10% and 5% double sided significance level.

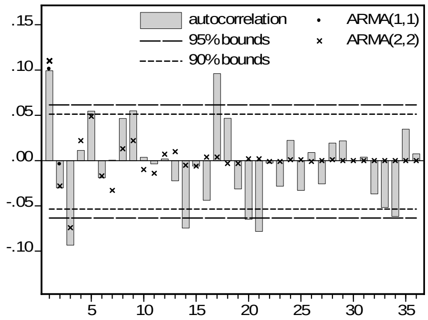
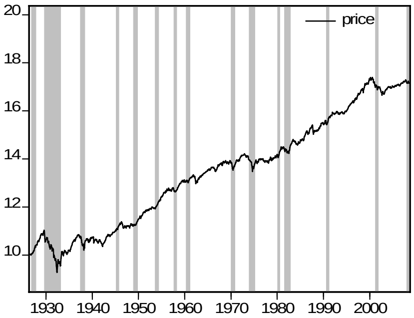
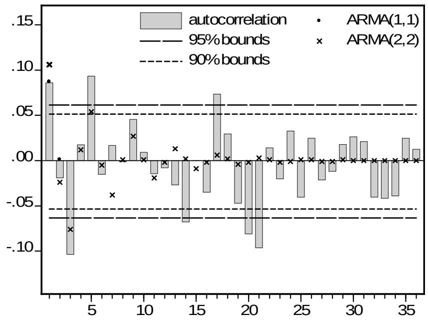
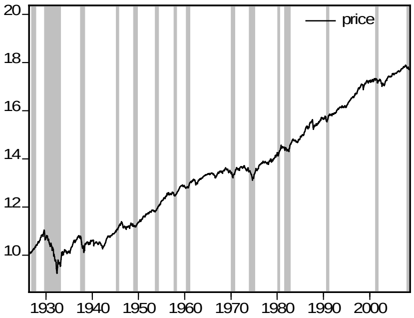
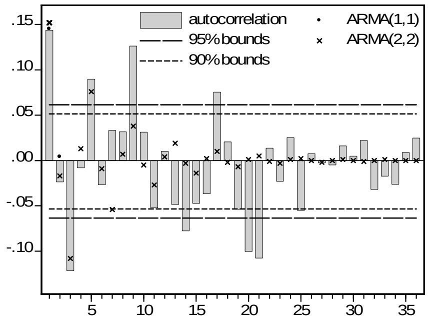
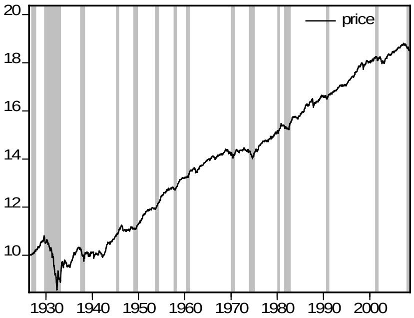
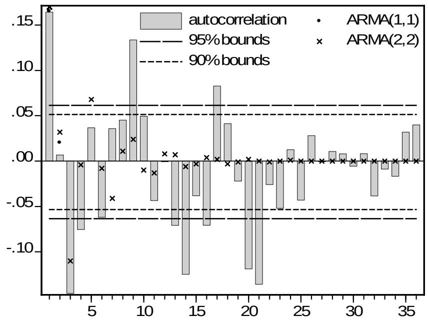
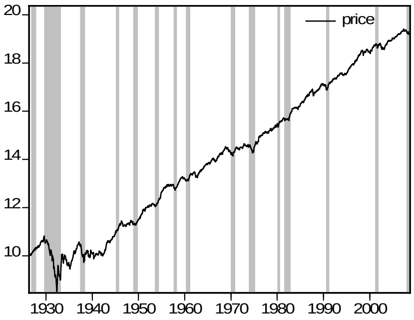
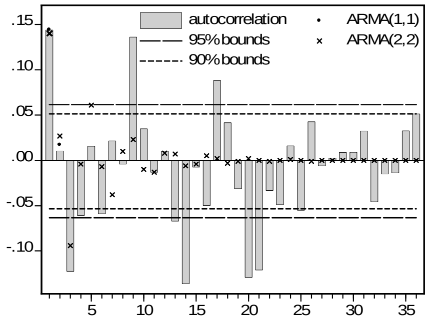
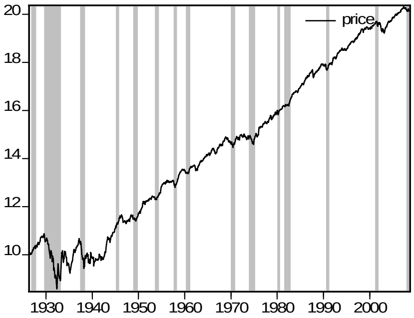
To describe the data in the form of prices, which are not stationary like their first difference, their course can be plotted. Panel B of figure 4 exhibits the course of prices along with the U.S. NBER (National Bureau of Economic Research) recession periods (gray shaded areas) measured from peak to trough of the business cycle. The prices of diversified portfolios, like the ones examined, should respond mainly to aggregate shocks related to new information concerning the point of the cycle. Inside the shaded areas representing recessions, informational shocks inflict a negative effect to prices. The patterns of prices explained in figures 1 to 3 may be present in the actual courses, but the actual courses are much more complex and a graphical examination is simply not feasible. What can be said is that the courses of all portfolio prices show a strong upward trend with small deviations around it. If prices follow a random walk then the innovations (shocks) should have small variability compared to the magnitude of the upward trend.

---

[16] Throughout the empirical analysis figures within a single panel have identical axis scales to avoid visual misinterpretations.

Figure 4 Return autocorrelations and course of prices

Panel A shows the sample autocorrelations and the autocorrelations implied by the estimated ARMA(1,1) and ARMA(2,2) as functions of lead/lag τ=1,…,36 (horizontal axis). For the ARMA(1,1), only the non-zero autocorrelations are depicted. The dashed lines are the 90% and 95% confidence intervals of the sampling distribution under random walk. Panel B shows the course of log prices as a function of time (horizontal axis). The shaded areas are the NBER recession periods.

Portfolio	Panel A: Return autocorrelations	Panel B: Course of prices
1		
2		
3		
4		
5		

3.3.2 Random walk tests

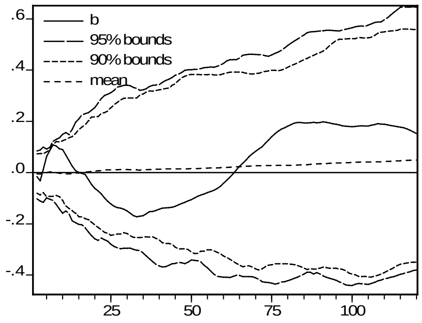
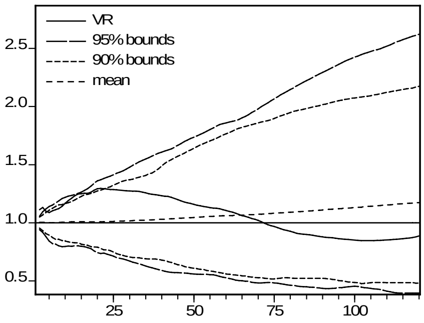
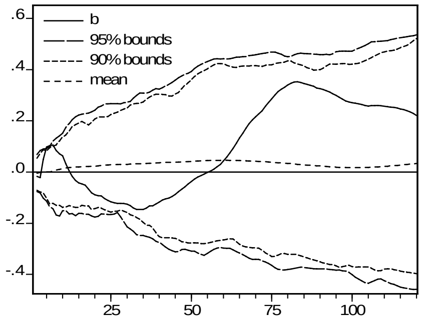
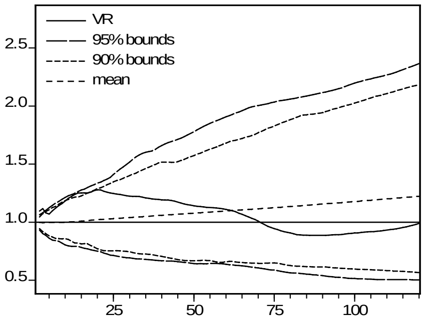
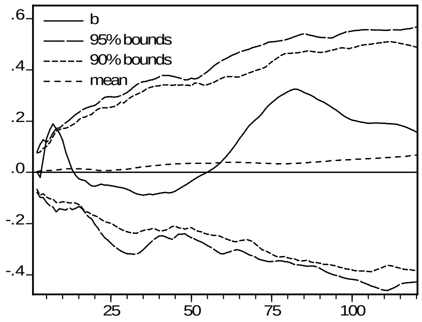
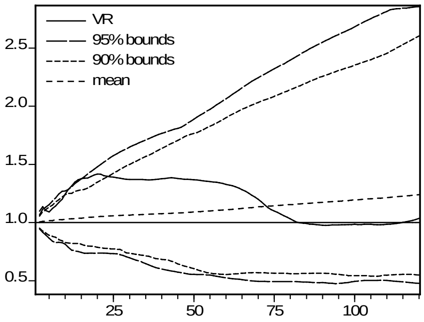
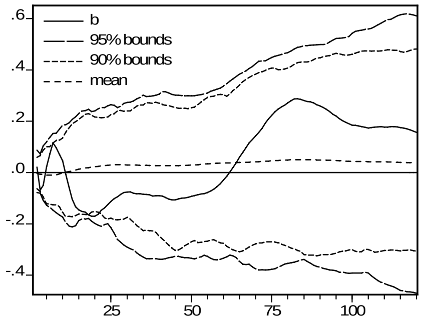
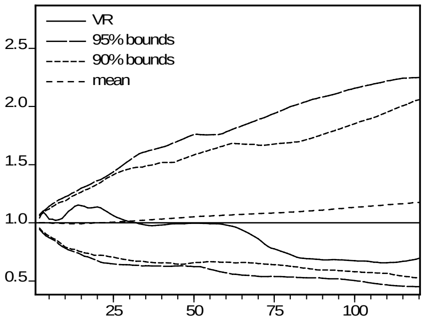
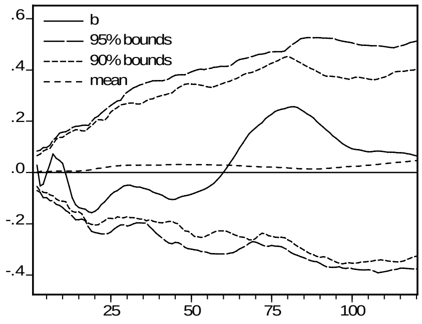
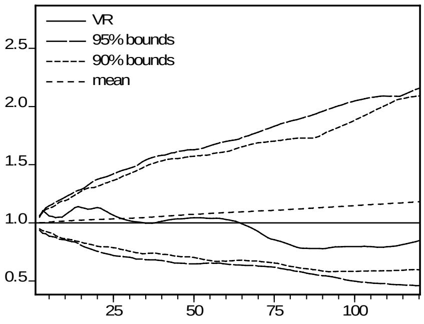
Figure 5 shows the regression beta and the variance ratio sample statistics, the 90% and 95% confidence bounds of the sampling distribution and its mean. As it was shown in expressions (38) and (41), both statistics are a function of sample autocorrelations, albeit with different weights applied to them. Under random walk the expected value of the regression beta and the variance ratio would asymptotically (i.e. for an infinite sample) be zero and one respectively for all return periods k. It can be seen in both panels of figure 5 that the finite-sample expected value of the sampling distribution is positively biased. Based on (38) and (41), the factors which determine whether the statistics are above (below) their asymptotic expected value are the number of positive (negative) autocorrelations, their magnitude and the weights applied to them. This being said, the pattern of variance ratio can be explained. All portfolios have variance ratio higher than one for small k which can be mainly attributed to the magnitude of the positive first order autocorrelation observed in panel A of figure 4 and the weight applied to it. For subsequent k, the variance ratio of portfolios 4 and 5 becomes lower than one due to high negative autocorrelations at orders 20 and 21. The variance ratio of the rest portfolios becomes lower than one for larger k. Regarding the regression beta of the portfolios, its value is again positive for low k due to the significant first-order positive autocorrelation. For subsequent but still low k, the regression beta becomes negative, which is the pattern described to be captured by the variance ratio for larger k. The regression beta $b(k)$ captures at smaller k the observed negative sample autocorrelations since it is a function of autocorrelations up to order $2k$-1 whereas the variance ratio $VR(k)$ is up to order k-1.

All significance sample statistics are found for k lower than 25. The two value portfolios show no significance for both statistics. The middle portfolio has the most significant regression beta sample statistic values for values of k from 5 to 8, i.e. four significant. However, for 119 sample estimates of each portfolio and a two sided significance level of 10%, it is expected to find 12 significant. The variance ratio for the first three portfolios is significant for k from 2 to 4. It is also significant from 10 to 22, from 10 to 18 and from 10 to 24 respectively for these three portfolios. As a conclusion, the random walk null hypothesis is rejected convincingly only for the first and third portfolios.

The paper of Fama and French (1988) finds for size sorted and industry portfolios a U-shaped pattern of the regression beta which is negative up to k=96 and then it become positive. It also finds that the sample statistic is significant for values of k from 36 to 60. The significance can be attributed to the way the sampling distribution was constructed, which is by choosing an arbitrary generating process to simulate returns. The BE/ME sorted portfolios examined herein have positive regression beta for low values of k and the negative sort of U-shaped pattern turns positive at the region of k=60. Lo and MacKinlay (1988) estimate the variance ratio for weekly and monthly returns of size sorted portfolios and significant variance ratios, which are higher than one, are found for the weekly data when k<16. In their paper, the asymptotic sampling distribution of the statistic is used to conduct the tests.

Figure 5 Regression beta and variance ratio sample statistics

The figure shows the sample regression beta (b), the variance ratio (VR), the 90% and 95% confidence bounds and the mean of their sampling distribution as functions of return interval *k*=2,…,120 (horizontal axis).

Portfolio	Panel A: Regression beta	Panel B: Variance ratio
1		
2		
3		
4		
5		

3.3.3 Alternative models estimation

The estimated parameters of the ARMA(1,1) process are exhibited in table 2. There is not much confidence about the estimated parameter values since their standard errors are many times larger than the parameters themselves. The Schwarz criterion and the *Q* statistic show that the ARMA(1,1) process fits better the growth portfolios since they have lower *SC* values and higher p values for the *Q* statistic. In fact, the null hypothesis of uncorrelated residuals is not rejected for portfolio 1. The persistence of the effect of a unit shock is measured by the cumulative response $\psi(1)$ for a sufficient amount of periods ahead. Recall that the estimated ARMA(1,1) model can be written as a first order difference stationary (23) to which the cumulative response $\psi(1)$ applies. All portfolios have cumulative response higher than one. The growth portfolios show the lower shock effect persistence and the value stocks the higher. This practically means that the price of value portfolios will respond in the long run to a higher degree to the impact of new information.

Table 2 ARMA(1,1) model estimated parameters

The table reports the estimated parameters of the ARMA(1,1) and its diagnostics, the Schwarz criterion, the *Q* statistic, the *JB* statistic of residuals and the cumulative response $\psi(1)$ to a shock. In parenthesis are the standard errors of parameters and $\psi(1)$, and the observed significance levels of *Q*(12) and *JB* statistic.

Portfolio	1	2	3	4	5
$(1-\phi_1 L)r_t = \phi_0 + (1-\zeta_1 L)\omega_t, \quad \omega_t \sim \text{IID}(0,\sigma_\omega^2)$					
ϕ_0	0.009 (0.002)	0.009 (0.002)	0.010 (0.002)	0.011 (0.002)	0.013 (0.003)
ϕ_1	-0.036 (0.374)	0.012 (0.486)	0.030 (0.410)	0.121 (0.409)	0.116 (0.396)
ζ_1	0.139 (0.401)	0.076 (0.526)	0.118 (0.462)	0.044 (0.493)	0.029 (0.461)
σ_ω^2	0.003	0.003	0.003	0.004	0.006
SC	-2.931	-3.015	-2.879	-2.571	-2.229
Q(12)	16.833 (0.078)	23.027 (0.011)	46.503 (0.000)	55.180 (0.000)	43.749 (0.000)
JB	1014 (0.000)	1272 (0.000)	6726 (0.000)	14338 (0.000)	8252 (0.000)
$\psi(1)$	1.099 (0.047)	1.089 (0.048)	1.153 (0.068)	1.189 (0.083)	1.164 (0.080)

Table 3 exhibits the parameter values of the fads and time-varying expected returns models corresponding to the ARMA(1,1) estimates. In some cases the structural model representation is not applicable since solving for the parameters leads to at least one negative error variance. Panel A of table 3 shows that the trend variability σ_ε^2 is multiple times larger than the variability of the transitory component σ_υ^2. Panel B shows that the value portfolios has high autoregressive parameters a_1, relatively low random walk error variance σ_ε^2 and the variability of the time varying drift σ_v^2 is multiple times larger than the variability of the random walk error. Thus, informational shocks are transmitted to the price level of value portfolio through changes in the drift, assuming the time varying expected returns model is valid. At the other extreme, portfolios 2 and 3 have small autoregressive coefficient and high random walk error variance.

Table 3 Parameter values of the fads and the time varying expected returns models

Panel A and B report the parameter values of the fads and the time varying expected return models corresponding to the estimated parameters of the ARMA(1,1). The cross covariance of the errors is set to zero. An NA is used to denote portfolios for which a valid structural representation is not applicable. Panel B also shows the cumulative response of the trend.

Portfolio	1	2	3	4	5
Panel A: Fads model $(1-\phi_1 L)r_t = (1-\phi_1)r + \varepsilon_t - \phi_1\varepsilon_{t-1} + \upsilon_t - \upsilon_{t-1}, \varepsilon_t \sim \text{IID}(0,\sigma^2_\varepsilon), \upsilon_t \sim \text{IID}(0,\sigma^2_\upsilon)$					
r	0.0085	0.0093	0.0106	NA	NA
ϕ_1	-0.0363	0.0118	0.0302	NA	NA
σ^2_ε	0.0021	0.0025	0.0027	NA	NA
σ^2_υ	0.0005	0.0002	0.0003	NA	NA
$\sigma_{\varepsilon\upsilon}$	0	0	0	0	0
Panel B: Time varying expected return model $(1-a_1 L)r_t = a_0 + \varepsilon_t - a_1\varepsilon_{t-1} + v_{t-1}, \varepsilon_t \sim \text{IID}(0,\sigma^2_\varepsilon), v_t \sim \text{IID}(0,\sigma^2_v)$					
a_0	NA	0.0092	0.0103	0.0115	0.0134
a_1	NA	0.0118	0.0302	0.1215	0.1158
σ^2_ε	NA	0.0180	0.0126	0.0016	0.0015
σ^2_v	NA	0.0026	0.0029	0.0042	0.0060
$\sigma_{\varepsilon v}$	0	0	0	0	0
$1/(1-a_1)$	NA	1.0119	1.0311	1.1383	1.1310

The estimated parameters of the ARMA(2,2) process are exhibited in table 4. Similarly to the ARMA(1,1), the growth portfolios have better diagnostics. In particular, they have a lower Schwarz criterion value and the null hypothesis of uncorrelated residuals is not rejected. The cumulative response $\psi(1)$ for the difference stationary representation of the ARMA(2,2) is higher than one for all portfolios, with the growth stocks showing the lower shock effect persistence and the value stocks the higher. The composite model parameter identification by using the ARMA(2,2) estimates is applicable for none of the portfolios since solving for the parameters leads to at least one negative error variance.

A comparison of the ARMA(1,1) and ARMA(2,2) estimated processes can be done by comparing the Schwarz criterion a portfolio achieves under each process. Under the ARMA(2,2), for portfolio 1 and 2 the criterion is marginally higher by 0.001 and for the rest portfolios is clearly lower. Thus, the ARMA(2,2) fits better the data. The other way to compare the two processes is by inspecting what autocorrelation function each of them implies and whether it can capture the dynamics of the data. In panel A of figure 4 the autocorrelations implied by the estimated ARMA processes are plotted against the sample autocorrelations of the data. The ARMA(1,1) model has a positive first order autocorrelation for all portfolios which coincides with the value of the actual one, and a second order autocorrelation which does not come close to the actual one. The autocorrelations of the process for lead/lag three or higher are all zero. The ARMA(2,2) process has richer autocorrelation function and can capture better the dynamics of the actual data. Especially for portfolios 1 and 2, the autocorrelation function of the ARMA(2,2) process has values close to the sample autocorrelations for up to 6 leads/lags. As a conclusion, the ARMA(2,2) model fits better the data than the ARMA(1,1).

Table 4 ARMA(2,2) model estimated parameters

The table reports the estimated parameters of the ARMA(2,2) and its diagnostics Schwarz criterion, *Q* statistic, the *JB* statistic of residuals and cumulative response $\psi(1)$ to a shock. In parenthesis are the standard errors of parameters and $\psi(1)$, and the observed significance level of *Q*(12) and *JB*.

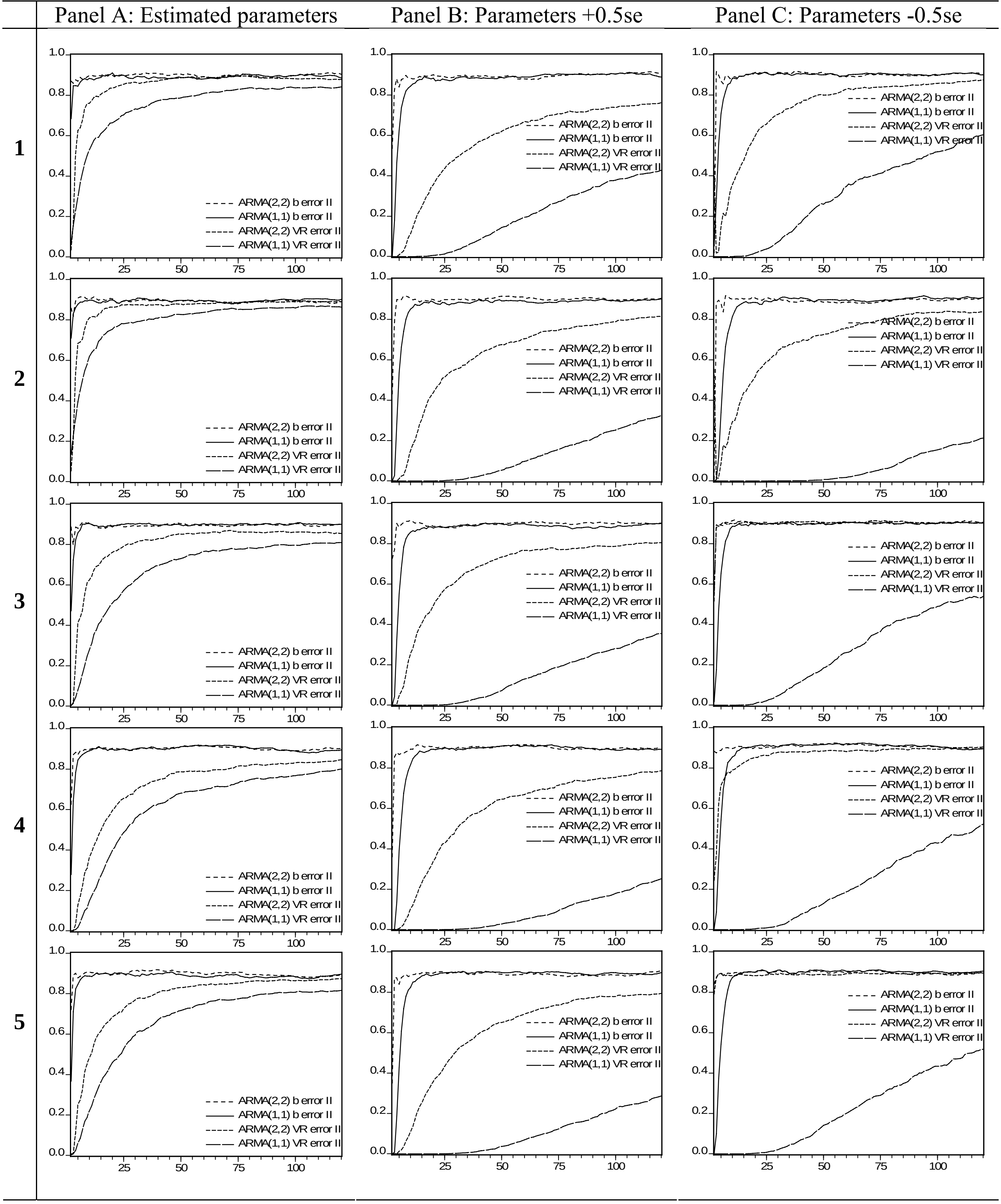
Portfolio	1	2	3	4	5
$(1-\theta_1 L-\theta_2 L^2)r_t = \theta_0 + (1-\zeta_1 L-\zeta_2 L^2)\omega_t, \quad \omega_t \sim \text{IID}(0,\sigma_\omega)$					
θ_0	0.009 (0.002)	0.009 (0.002)	0.010 (0.002)	0.011 (0.002)	0.013 (0.003)
θ_1	-0.041 (0.163)	0.067 (0.154)	-0.006 (0.098)	-0.146 (0.097)	-0.147 (0.111)
θ_2	-0.680 (0.111)	-0.699 (0.108)	-0.708 (0.073)	-0.620 (0.080)	-0.641 (0.089)
ζ_1	0.155 (0.164)	0.042 (0.157)	0.164 (0.096)	0.320 (0.089)	0.290 (0.104)
ζ_2	0.673 (0.114)	0.684 (0.112)	0.728 (0.071)	0.717 (0.067)	0.717 (0.078)
σ_ω	0.003	0.003	0.003	0.004	0.006
SC	-2.930	-3.016	-2.892	-2.587	-2.234
Q(12)	3.789 (0.876)	7.659 (0.467)	19.991 (0.010)	28.649 (0.000)	26.679 (0.001)
JB	777 (0.000)	1034 (0.000)	5121 (0.000)	12714 (0.000)	6746 (0.000)
$\psi(1)$	1.062 (0.029)	1.057 (0.029)	1.103 (0.028)	1.154 (0.031)	1.122 (0.029)

3.3.4 Power calculations

The regression beta and the variance ratio statistic did not reject the random walk null hypothesis for most of the portfolios. This might have been the case because the test statistics are not powerful enough, that is, their type II error was high against certain alternative hypotheses. Panel A of figure 6 shows the type II error of both statistics when the true models are the estimated ARMA(1,1) and ARMA(2,2). To calculate the type II error return series are simulated by randomizing the estimated residuals 2000 times for each ARMA model and the significance level (type I error) is set to 10%. Apart from the result that the variance ratio is less powerful for growth portfolios, the results have common characteristics across all portfolios.

Figure 6 Type II error of the regression beta and variance ratio test statistics

Panel A shows the type II error of the regression beta (b) and variance ratio (VR) test statistics as functions of return interval *k*=2,…,120 (horizontal axis) when the true models are the estimated ARMA(1,1) and ARMA(2,2). Panels B and C report the type error calculations for estimated parameter values ±0.5 standard error. Calculations are based on 2000 randomisations of estimated errors and type I error 10%.



The power of the regression beta is rather low for all *k* and roughly equals the 10% significance level, whereas the power of the variance ratio is high for low *k* but deteriorates exponentially as *k* increases. The variance ratio is strictly more powerful than the regression beta and the statistics are more powerful when the true model is an ARMA(1,1). The latter simple means that the statistics are more able to detect a deviation from the random walk when the true model is an ARMA(1,1). But, as it was concluded from the comparison between the *SC* values and the autocorrelation function of ARMA(1,1) and ARMA(2,2), an ARMA(2,2) is actually more possible to be the true model. Moreover, the variance ratio has the least power for those portfolios which the alternatives are most meaningful, i.e. those which have better diagnostics.[17] The two aforementioned facts are not encouraging for the effectiveness of the random walk test. Since there is uncertainty about the estimated ARMA models a sensitivity analysis of the power calculation is conducted with parameter values plus/minus half their standard error and their results are exhibited in panel B and C of figure 6. The power of the regression beta improves slightly only against the ARMA(1,1) and for very small values of *k*, whereas the power of the variance ratio improves significantly except for the case of minus the standard error when the true model is the ARMA(2,2) for portfolios 3, 4, 5.

The papers of Poterba and Summers (1988) and Lo and MacKinlay (1988) find for returns simulated by an ARMA(1,1) with arbitrarily chosen parameters that the variance ratio is more powerful than the regression beta and the *Q* statistic respectively. As in Lo and MacKinlay (1989, p.228, 235) it is found that for specific return intervals, statistics are more powerful to detect deviations from random walk. This is the reason why joint tests of the significance of sample statistics across return intervals are not examined as proposed in Richardson (1992, p.5, 8). Joint tests would have been appropriate if rejections of the null hypothesis in individual return intervals are random and due to sampling error as suggested in Kim, Nelson and Startz (1991, p.520).

A way to search for the reason why the test statistics are not powerful enough would include the inspection of the sample distribution of the statistic under the null and the alternative hypotheses. It is not however practically feasible to report all the sampling distributions. What is feasible to report is the first moment of the sampling distributions. Figure 7 exhibits the mean of the sampling distribution of the two test statistics under the different models and the sample estimates of the statistics. Under the random walk and the ARMA models, the first moment is almost the same for the regression beta but clearly different for the variance ratio. Then why the variance ratio has so low power? A hint can come from exhibiting the sampling distribution of the two statistics for one portfolio, take for example portfolio 3, and for selected values of *k* in panels A and B of figure 8. Regarding the variance ratio, although the sampling distributions have different moments, their range is fairly close. Even though the null hypothesis and an alternative may imply very different values for the statistic, if they have similar range, there is a very small chance the null will ever be rejected. Regarding the regression beta, it seems that not only is the first moment of the sampling distributions fairly close but also the higher moments.

---

[17] Recall that portfolios 1 and 2 were concluded to have better diagnostics regarding the ARMA models based on comparisons between their *SC* and *Q* statistic values and the ones of the other portfolios.

Figure 7 First moment of the regression beta and variance ratio sampling distributions under the null and alternative hypotheses

Figure 7 shows the first moment of the regression beta (b) and variance ratio (VR) sampling distributions under the null of random walk and the alternative hypotheses ARMA(1,1)-ARMA(2,2), and the sample statistic as a function of return interval *k*=2,…,120 (horizontal axis). The first moments under the ARMA models are calculated from simulated return series by randomizing 2000 times the estimated residuals and under the null from 2000 randomisations of the original return series.

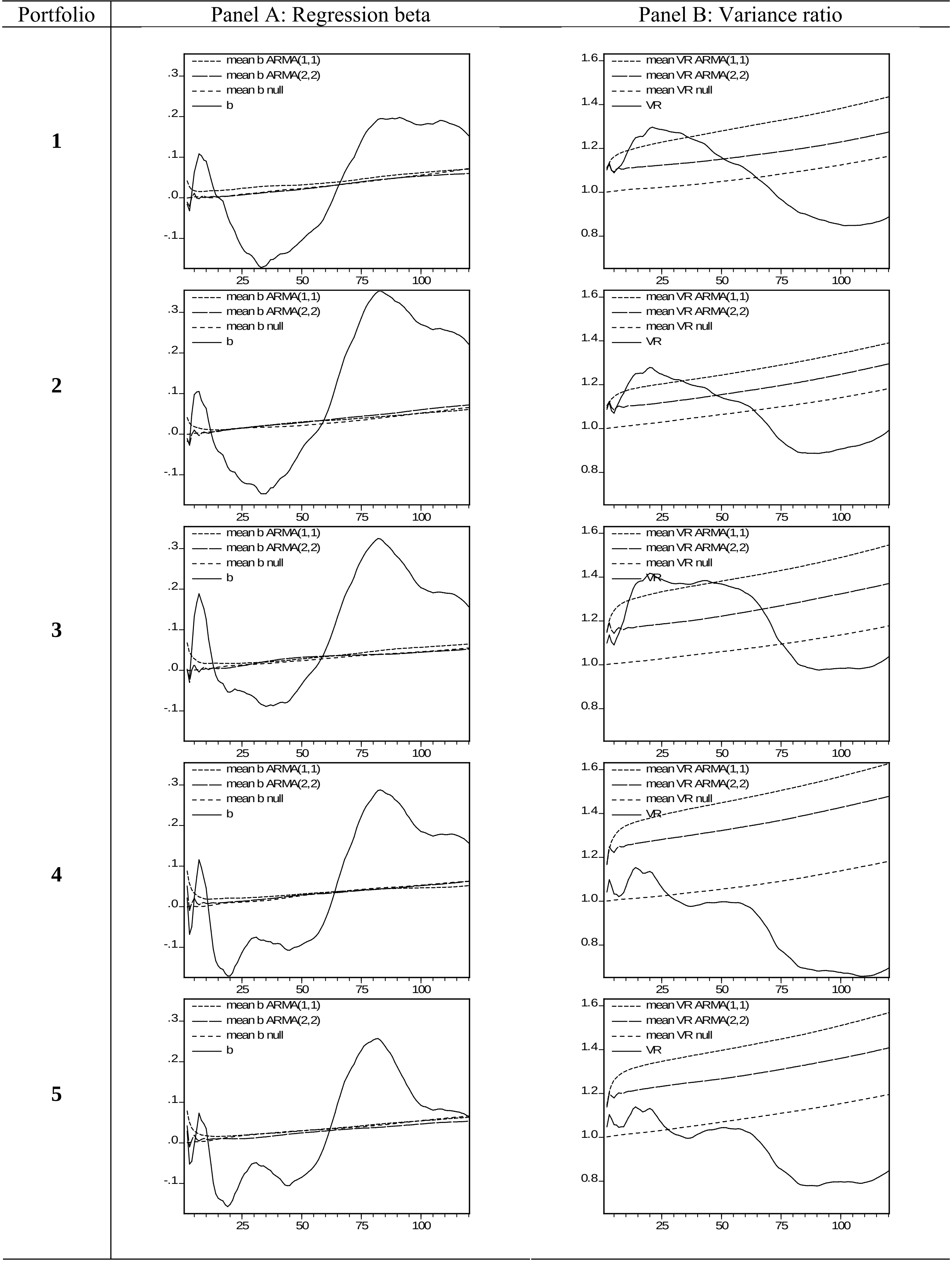
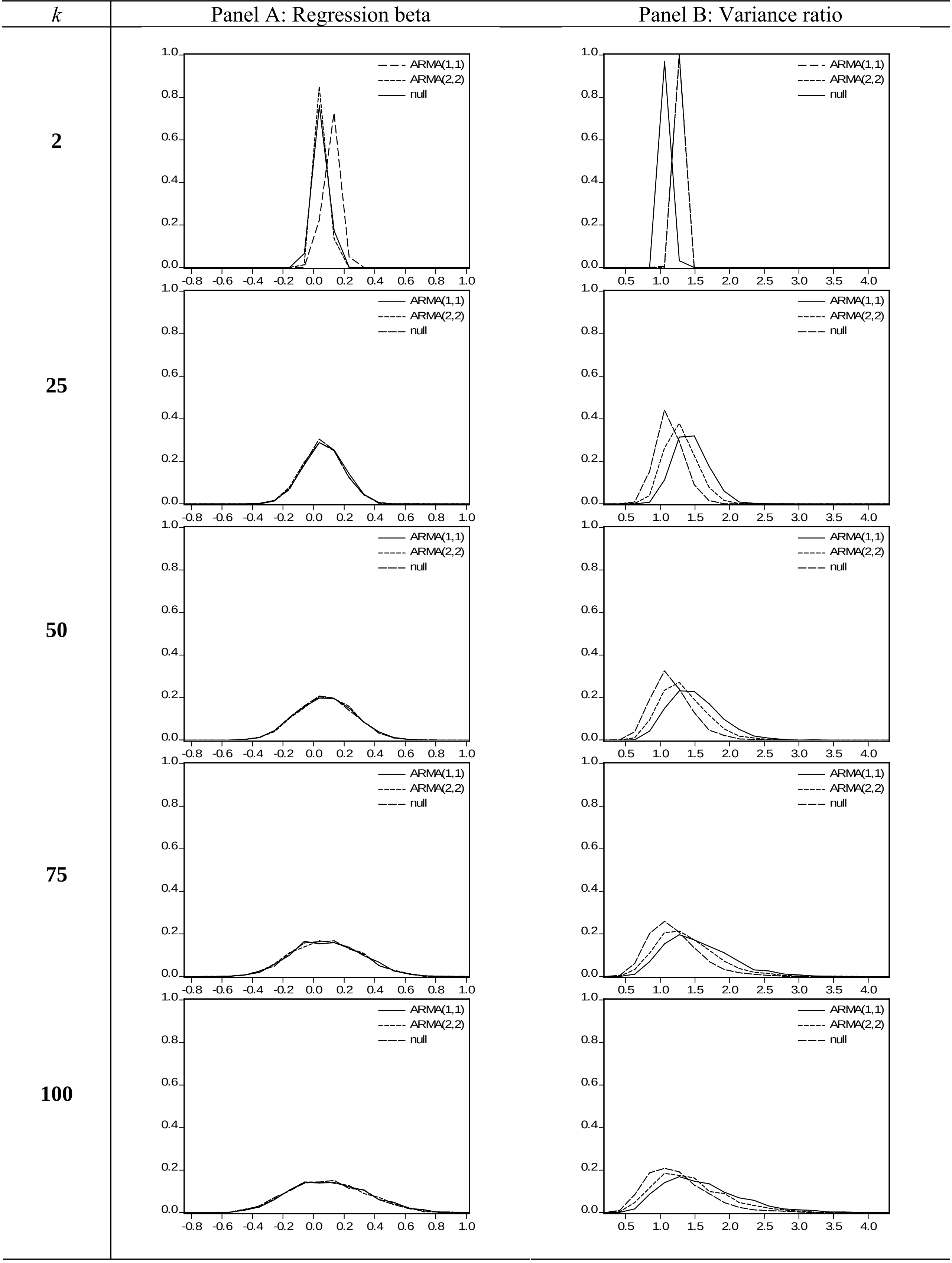


Figure 8 Sampling distribution of the regression beta and variance ratio under the null and alternative hypotheses for portfolio 3

Figure 8 shows the frequency distribution of the regression beta and variance ratio under random walk, ARMA(1,1) and ARMA(2,2) for return interval values *k*=2, 25, 50, 75, 100 for portfolio 3. The distributions under the ARMA models are calculated from simulated return series by randomizing 2000 times the estimated residuals and under the null from 2000 randomisations of the original return series.



CHAPTER 4 Summary and conclusions

In section 2.1 the random walk model is derived as a testable expression of market informational efficiency assuming constant expected returns. If markets are efficient, prices follow a random walk. In section 2.3 the random walk test statistics, namely the regression beta and the variance ratio, are discussed. Both statistics are a function of sample autocorrelations, which simply makes them a test of whether sample autocorrelations are jointly different from zero. Subsection 3.3.2 estimates the sample statistics for overlapping monthly return interval k=2,...,120 and tests their significance. To construct their sampling distributions under the null hypothesis of random walk, the statistics are calculated 2000 times from shuffled return series. When the return interval k is small, the random walk model is rejected by the variance ratio for portfolios with low BE/ME ratio.

Section 2.2 discusses the alternative structural models of fads and time varying expected returns. These structural models constitute of permanent/transitory components and are observationally equivalent to an ARMA(1,1) process. A new composite model, which has an ARMA(2,2) observationally equivalent, is also proposed to incorporate the characteristics of the previous two models. The structural models have a state space representation and their parameters can be estimated with exact maximum likelihood. This estimation however is problematic since the algorithms used to maximise the log likelihood function were not able to converge to a solution. A solution was not found even when the cross covariances of errors were restricted to zero. The method finally used in subsection 3.3.3 is to estimate the observationally equivalent ARMA processes of the structural models and solve for the unknown parameters. For many cases the structural models do not have a valid representation. As a result, the estimation problems does not allow for conclusions to be made with respect to the magnitude of the permanent and transitory components. Comparing the structural models in terms of their observationally equivalent processes, the ARMA(2,2) fits better the data, i.e. has a lower Schwarz criterion, and has a richer autocorrelation function that can capture better the dynamics of the data.

The ability of the statistics to detect deviations from random walk when one of the alternative models is true is examined in subsection 3.3.4 and is found to be low. The regression beta has large type II error which equals the significance level (type I error) for almost every return interval k. This would have been expected if the null and the alternative hypotheses implied the same sampling distribution for the statistic. Indeed, panel A of figure 7 and 8 actually support this expectation. The power of the test statistic does not improve (almost at all) even when returns are simulated from ARMA models with the estimated parameters plus/minus half their standard error. The variance ratio is strictly more powerful than the regression beta. However, there are two results that are not encouraging for the effectiveness of the variance ratio. When the true model is an ARMA(1,1) the statistic has a lower type II error than the ARMA(2,2) case. But, as it was shown, an ARMA(2,2) is more possible to be indeed the true model. Moreover, the variance ratio has the least power for those portfolios (growth) which the alternatives can be indeed the true. Overall, given the low power of the statistics, if the random walk is actually rejected, as with portfolios 1, 2 (marginally) and 3, this should be considered as a strong result.

The original intention for examining BE/ME sorted portfolios has been to identify transitory components related to the sorting variable. Indeed, some of the results show a distinction between growth and value portfolios. The random walk model is marginally rejected for the growth portfolios by the variance ratio test, but it is not rejected for value portfolios. The ARMA(1,1) and ARMA(2,2) processes fit better the growth portfolios and the variance ratio is less powerful for growth portfolios.

REFERENCES

Beveridge, S. and C. Nelson, 1981, A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle, *Journal of Monetary Economics* 7, 151-174.

Brockwell, P. and R. Davis, 2002, Introduction to time series and forecasting, Second edition, Springer.

Campbell, J., A. Lo, and C. MacKinlay, 1997, The econometrics of financial markets. Princeton University Press.

Campbell, J. and G. Mankiw, 1987a, Permanent and transitory components in macroeconomic fluctuations, *The American Economic Review* 77, 111-117.

Campbell, J. and G. Mankiw, 1987b, Are output fluctuations transitory?, *The Quarterly Journal of Economics* 102, 857-880.

Chatfield, C., 1995, The analysis of time series: An introduction, Fifth edition, Chapman and Harp/CRC.

Clark, P., 1987, The cyclical component of U.S. economic activity, *The Quarterly Journal of Economics* 102, 797-814.

Cochrane, J., 1988, How big is the random walk in GNP?, *The Journal of Political Economy* 96, 893-920.

Conrad, J. and G. Kaul, 1988, Time-variation in expected returns, *Journal of Business* 61, 409-425.

Cuthbertson, K., 1996, Quantitative financial economics: Stocks, bonds and foreign exchange, Wiley.

Fama, E., 1970, Efficient capital markets: A review of theory and empirical work, *The Journal of Finance* 25, 383-417.

Fama, E., 1976, Foundations of finance, Basic Books, New York.

Fama, E., 1991, Efficient capital markets: II, *The Journal of Finance* 46, 1575-1617.

Fama, E. and K. French, 1988, Permanent and temporary components of stock prices, *Journal of Political Economy* 96, 246-273.

Fama, E. and K. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Geweke, J., 1981, The approximate slope of econometric tests, *Econometrica* 49, 1427-1442.

Hamilton, J., 1994, Time Series Analysis, Princeton University Press.

Harvey, A., 1985, Trends and cycles in macroeconomic time series. *Journal of Business and Economic Statistics* 3, 216-227.

Hawawini, G. and D. Keim, 1994, On the predictability of common stock returns: World-wide evidence, working paper.

Ho, C. and S. Sears, 2006, Is there conditional mean reversion in stock returns? *Quarterly Journal of Business and Economics* 45, pages 91-112.

Jagadeesh, N., 1991, Seasonality in stock mean reversion: Evidence from the U.S. and the U.K., *The Journal of Finance* 46, 1427-1444.

Kaul, G., 1996, Predictable components in stock returns, *Handbook of Statistics* 14, 269-296.

Kim, M., C. Nelson and R. Startz ,1991, Mean Reversion in stock prices? A Reappraisal of the Empirical Evidence. *Review of Economic Studies* 58, 515-528.

Kirchgässner, G. and U. Wolters, 2007, Introduction to modern time series analysis, Springer.

Lippi, M. and L. Reichlin, 1992, On persistence of shocks to economic variables. A common misconception, *Journal of Monetary Economics* 29, 87–93.

Lo, A. and C. MacKinlay, 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of Financial Studies* 1, 41-66.

Lo, A. and C. MacKinlay, 1989, The size and power of the variance ratio test in finite samples: A Monte Carlo investigation, *Journal of Econometrics* 40, 203-238.

Mark, N., 2001, International macroeconomics and finance: Theory and empirical methods. Blackwell.

Morley, J., C. Nelson and E. Zivot, 2003, Why are the Beveridge-Nelson and the unobserved components decomposition of GDP so different?, *The Review of Economics and Statistics* 85, 235-243.

Nelson, C. and C. Plosser, 1982, Trends and random walks in macroeconomic time series: Some evidence and implications, *Journal of Monetary Economics* 10, 139-62.

Oh, K. and E. Zivot, 2006, The Clark model with correlated components, Unpublished manuscript, Department of economics, University of Washington.

Poterba, J. and L. Summers, 1988, Mean reversion in stock prices: Evidence and implications, *Journal of Financial Economics* 22, 27-59.

Richardson, M., 1992, Temporary components of stock prices: A skeptic's view, Working paper, Wharton School, University of Pennsylvania.

Richardson, M. and T. Smith, 1994, A unified approach to testing for serial correlation in stock returns, *Journal of Business* 67, 371-399.

Stock, J. and M. Watson, 1988, Variable trends in economic time series, *The Journal of Economic Perspectives* 2, 147-174.

Summers, L., 1986, Does the stock market rationally reflect fundamental values?, *The Journal of Finance* 41, 591-601.

Tsay, R., 2005, Analysis of financial time series, Second edition, Wiley.

Watson, M., 1986, Univariate detrending methods with stochastic trends, *Journal of Monetary Economics* 18, 49-75.

Zhou, C. and C. Qing, 2000, A state-space model of short- and long-horizon stock returns, *Journal of Financial Research* 23, 523-554.

APPENDIX A The Beveridge-Nelson (1981) decomposition

Consider an ARMA(p,q) model $\theta_{\mathrm{p}}(L)r_t = \theta_0 + \zeta_{\mathrm{q}}(L)\upsilon_t$ which can also be written as a MA(∞) process $r_t = p_t - p_{t-1} = r + \psi(L)\upsilon_t$. The future expected price $p_{t+\tau}$ will only be affected by innovations incorporated in its level permanently. Innovations which cause a transitory effect have zero expected value. From the definition of returns at time *t*+*τ* we have

$$\begin{aligned} p_{t+\tau} &= r_{t+\tau} + p_{t+\tau-1} \\ &= r_{t+\tau} + \ldots + r_{t+2} + r_{t+1} + p_t \end{aligned}$$

$$E_t p_{t+\tau} = E_t\left(r_{t+\tau} + \ldots + r_{t+2} + r_{t+1}\right) + p_t .$$

The expected values of returns can be found from the MA(∞) representation

$$\begin{aligned} E_t r_{t+1} &= r + \psi_1 \upsilon_t + \psi_2 \upsilon_{t-1} + \ldots \\ E_t r_{t+2} &= r + \psi_2 \upsilon_t + \psi_3 \upsilon_{t-1} + \ldots \\ &\vdots \\ E_t r_{t+\tau} &= r + \psi_\tau \upsilon_t + \psi_{\tau+1} \upsilon_{t-1} + \ldots \end{aligned}$$

which can be substituted in the previous expression to get

$$E_t p_{t+\tau} = \tau r + p_t + \left(\sum\nolimits_{i=1}^{\tau} \psi_i\right)\upsilon_t + \left(\sum\nolimits_{i=2}^{\tau+1} \psi_i\right)\upsilon_{t-1} + \ldots .$$

For $\tau \to \infty$ the expected price is

$$\lim_{\tau\to\infty} E_t p_{t+\tau} = \lim_{\tau\to\infty}\left[\tau r + p_t + \left(\sum\nolimits_{i=1}^{\tau} \psi_i\right)\upsilon_t + \left(\sum\nolimits_{i=2}^{\tau+1} \psi_i\right)\upsilon_{t-1} + \ldots\right].$$

The first product at the right-hand side is the deterministic trend and the rest is a stochastic permanent trend. Set the stochastic permanent trend equal to

$$p_t^* = \lim_{\tau\to\infty}\left(E_t p_{t+\tau} - \tau r\right) = p_t + \left(\sum\nolimits_{i=1}^{\infty} \psi_i\right)\upsilon_t + \left(\sum\nolimits_{i=2}^{\infty} \psi_i\right)\upsilon_{t-1} + \ldots$$

which also implies an expression for prices

$$p_t = p_t^* - \left(\sum\nolimits_{i=1}^{\infty} \psi_i\right)\upsilon_t - \left(\sum\nolimits_{i=2}^{\infty} \psi_i\right)\upsilon_{t-1} - \ldots .$$

The stochastic permanent trend p_t^* follows a random walk. To see this take its first difference

$$p_t^* - p_{t-1}^* = p_t - p_{t-1} + \left(\sum\nolimits_{1}^{\infty} \psi_i\right)\upsilon_t - (\psi_1 \upsilon_{t-1} + \psi_2 \upsilon_{t-2} + \ldots)$$

$$\begin{aligned} p_t^* &= p_{t-1}^* + r + \left(\sum\nolimits_{i=0}^{\infty} \psi_i\right)\upsilon_t \\ &= p_{t-1}^* + r + \psi(1)\upsilon_t . \end{aligned}$$

Set $\varepsilon_t = \psi(1)\upsilon_t$ and it is derived that the stochastic permanent component follows a random walk

$$p_t^* = p_{t-1}^* + r + \varepsilon_t .$$

Now take the previous expression for prices

$$p_t = p_t^* - \left(\sum\nolimits_{i=1}^{\infty} \psi_i\right) \upsilon_t - \left(\sum\nolimits_{i=2}^{\infty} \psi_i\right) \upsilon_{t-1} - ...$$

which decomposes prices into the permanent component p_t^* and a transitory component

$$u_t = -\left(\sum\nolimits_{i=1}^{\infty} \psi_i\right) \upsilon_t - \left(\sum\nolimits_{i=2}^{\infty} \psi_i\right) \upsilon_{t-1} - \dots .$$

If $\phi_i = -\sum\nolimits_{j=i+1}^{\infty} \psi_j$ is set, the transitory component is written as

$$u_t = \phi_0 \upsilon_t + \phi_1 \upsilon_{t-1} + \dots .$$

APPENDIX B The variance ratio as a measure of persistence

The mathematical expression of the variance ratio (40) is derived under different processes. In each case, using the subscript notation of multiperiod returns, the variance of long differences (the numerator) and the variance of the process (part of the denominator) are initially derived. First, the variance ratio is derived for a trend stationary process like (21). As a first step create the long differences of the process.

$$p_{tk} = tr + \psi(L)\upsilon_{tk}$$

$$p_{(t-1)k} = (t-1)r + \psi(L)\upsilon_{(t-1)k}$$

$$p_{tk} - p_{(t-1)k} = r + \psi(L)\upsilon_{tk} - \psi(L)\upsilon_{(t-1)k}$$

Then the variance of the long differences is

$$\mathrm{var}(p_{tk} - p_{(t-1)k}) = \mathrm{var}\left[r + \psi(L)\upsilon_{tk} - \psi(L)\upsilon_{(t-1)k}\right]$$

$$\mathrm{var}[r_{tk}(k)] = 2[\psi(1)]^2\sigma_\upsilon^2.$$

The variance of the trend stationary process is

$$\mathrm{var}[r_t] = [\psi(1)]^2\sigma_\upsilon^2.$$

The variance ratio is

$$VR(k) = \lim_{k\to\infty}\frac{\mathrm{var}[r_{tk}(k)]}{k\,\mathrm{var}(r_t)} = \lim_{k\to\infty}\frac{2[\psi(1)]^2\sigma_\upsilon^2}{k[\psi(1)]^2\sigma_\upsilon^2} = 0\,.$$

Second, the variance ratio is derived for a first-order difference process like (23). Take the first-order difference stationary process and recursively substitute prices *k*-1 times.

$$p_{tk} = p_{tk-1} + r + \psi(L)\upsilon_{tk}$$

$$p_{tk} = p_{tk-k} + kr + \psi(L)(\upsilon_{tk} + \upsilon_{tk-1} + ... + \upsilon_{tk-k+1})$$

$$p_{tk} - p_{(t-1)k} = kr + \psi(L)(\upsilon_{tk} + \upsilon_{tk-1} + ... + \upsilon_{tk-k+1})$$

Then the variance of the long differences is

$$\mathrm{var}(p_{tk} - p_{(t-1)k}) = \mathrm{var}\left[kr + \psi(L)(\upsilon_{tk} + \upsilon_{tk-1} + ... + \upsilon_{tk-k+1})\right]$$

$$\mathrm{var}[r_{tk}(k)] = [\psi(1)]^2 k\sigma_\upsilon^2.$$

The variance of the first-order difference process itself is

$$\mathrm{var}[r_t] = [\psi(1)]^2\sigma_\upsilon^2.$$

Now the variance ratio is

$$VR(k) = \lim_{k\to\infty}\frac{\mathrm{var}[r_{tk}(k)]}{k\,\mathrm{var}(r_t)} = \lim_{k\to\infty}\frac{[\psi(1)]^2 k\sigma_\upsilon^2}{k[\psi(1)]^2\sigma_\upsilon^2} = 1\,.$$

Third, the variance ratio is derived under the Beveridge-Nelson decomposition of a first-order difference stationary process $p_{tk} = p_{tk-1} + r + \psi(L)\upsilon_{tk}$. The decomposed form (see appendix A) is

$$p_{tk} = p^*_{tk} + u_{tk}$$

$$p^*_{tk} = p^*_{tk-1} + r + \psi(1)\upsilon_{tk}$$

$$u_{tk} = \phi(L)\upsilon_{tk}$$

Take the *k*-th order difference of prices to get

$$\begin{aligned} p_{tk} - p_{tk-k} &= p^*_{tk} - p^*_{tk-k} + \phi(L)(\upsilon_{tk} - \upsilon_{tk-k}) \\ &= rk + \psi(1)(\upsilon_{tk} + \upsilon_{tk-1} + ... + \upsilon_{tk-k+1}) + \phi(L)(\upsilon_{tk} - \upsilon_{tk-k}) \end{aligned}$$

$$\begin{aligned} \mathrm{var}[r_{tk}(k)] &= \mathrm{var}[\psi(1)(\upsilon_{tk} + \upsilon_{tk-1} + ... + \upsilon_{tk-k+1}) + \phi(L)(\upsilon_{tk} - \upsilon_{tk-k})] \\ &= [\psi(1)]^2 k\sigma^2_\upsilon + 2[\phi(L)]^2 \sigma^2_\upsilon + 2\psi(1)\phi(L)\,\mathrm{cov}(\upsilon_{tk} + \upsilon_{tk-1} + ... + \upsilon_{tk-k+1}, \upsilon_{tk} - \upsilon_{tk-k}) \\ &= [\psi(1)]^2 k\sigma^2_\upsilon + 2[\phi(L)]^2 \sigma^2_\upsilon + 2\psi(1)\phi(1)\sigma^2_\upsilon. \end{aligned}$$

The variance of the process is

$$\begin{aligned} \mathrm{var}(r_t) &= \mathrm{var}[p^*_t - p^*_{t-1} + \phi(L)(u_t - u_{t-1})] \\ &= \mathrm{var}[r + \psi(1)\upsilon_t + \phi(L)(\upsilon_t - \upsilon_{t-1})] \\ &= [\psi(1)]^2 \sigma^2_\upsilon + 2[\phi(1)]^2 \sigma^2_\upsilon + 2\psi(1)\phi(1)\sigma^2_\upsilon \\ &= \left[[\psi(1) + \phi(1)]^2 + [\phi(1)]^2 \right] \sigma^2_\upsilon. \end{aligned}$$

And the variance ratio is

$$VR(k) = \lim_{k\to\infty} \frac{\mathrm{var}[r_{tk}(k)]}{k\,\mathrm{var}(r_t)} = \lim_{k\to\infty} \frac{[\psi(1)]^2 k\sigma^2_\upsilon + 2[\phi(L)]^2 \sigma^2_\upsilon + 2\psi(1)\phi(1)\sigma^2_\upsilon}{k\left[[\psi(1) + \phi(1)]^2 + [\phi(1)]^2 \right]\sigma^2_\upsilon}$$

$$= \frac{[\psi(1)]^2}{[\psi(1) + \phi(1)]^2 + [\phi(1)]^2}.$$

Fourth, the variance ratio is derived under a structural model of the form (25) with arbitrary correlation between the errors of the two components.

$$p_{tk} = p^*_{tk} + u_{tk}$$

$$p^*_{tk} = p^*_{tk-1} + r + \varepsilon_{tk}$$

$$u_{tk} = \phi(L)\upsilon_{tk}$$

Take the *k*-th order difference of prices to get

$$\begin{aligned} p_{tk} - p_{tk-k} &= p^*_{tk} - p^*_{tk-k} + u_{tk} - u_{tk-k} \\ &= rk + \varepsilon_{tk} + \varepsilon_{tk-1} + ... + \varepsilon_{tk-k+1} + \phi(L)(\upsilon_{tk} - \upsilon_{tk-k}) \end{aligned}$$

$$\begin{aligned} \mathrm{var}[r_{tk}(k)] &= \mathrm{var}[rk + \varepsilon_{tk} + \varepsilon_{tk-1} + ... + \varepsilon_{tk-k+1} + \phi(L)(\upsilon_{tk} - \upsilon_{tk-k})] \\ &= k\sigma^2_\varepsilon + 2[\phi(L)]^2 \sigma^2_\upsilon + 2\phi(1)\sigma_{\varepsilon\upsilon}. \end{aligned}$$

The variance of the process is

$$\begin{aligned}\mathrm{var}(r_t) &= \mathrm{var}[p_t^* - p_{t-1}^* + \phi(L)(\upsilon_t - \upsilon_{t-1})] \\ &= \sigma_\varepsilon^2 + \mathrm{var}[\phi(L)(\upsilon_t - \upsilon_{t-1})] + \mathrm{cov}[\varepsilon_t, \phi(L)\upsilon_t] \\ &= \sigma_\varepsilon^2 + 2[\phi(1)]^2 \sigma_\upsilon^2 + 2\phi(1)\sigma_{\varepsilon\upsilon}.\end{aligned}$$

And the variance ratio is

$$VR(k) = \lim_{k\to\infty} \frac{\mathrm{var}[r_{tk}(k)]}{k\,\mathrm{var}(r_t)} = \lim_{k\to\infty} \frac{k\sigma_\varepsilon^2 + 2[\phi(L)]^2\sigma_\upsilon^2 + 2\phi(1)\sigma_{\varepsilon\upsilon}}{k\left[\sigma_\varepsilon^2 + 2[\phi(1)]^2\sigma_\upsilon^2 + 2\phi(1)\sigma_{\varepsilon\upsilon}\right]} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + 2[\phi(1)]^2\sigma_\upsilon^2}.$$