2004

Forecasting the equity premium in the Australian market

David E. Allen

Lurion DeMello

Forecasting the Equity Premium in the Australian Market

By

D. E. Allen* and L. DeMello**

*School of Accounting, Finance and Economics
Edith Cowan University

**School of Economics and Finance
Curtin University of Technology

School of Accounting, Finance and Economics & FIMARC Working Paper Series
Edith Cowan University
October 2004
Working Paper 0403

Correspondence author:

David E. Allen
School of Accounting, Finance and Economics
Faculty of Business and Public Management
Edith Cowan University
100 Joondalup Drive
Joondalup WA 6027
Australia
Phone: 61+ (8) 6304 5471
Fax: 61+ (8) 6304 5271
Email: d.allen@ecu.edu.au

Abstract

This paper provides a forecasting methodology for estimating the market risk premium in Australia. We employ an in-sample and out-of-sample forecast estimate using various dividend yield measures. The lagged dividend yield model is used to predict future equity premia on a data series that includes the top 85 percent of the Australian stock market. An important concern in this paper is the accuracy of dividend yield in forecasting the equity premium in the Australian market. We find that the level of predictability in the later part of the series is very weak compared to in-sample prediction during the 70s and 80s. This finding is similar to many claims in most U.S. studies that find that other macroeconomic factors such as the business cycle, inflation and the level of economic growth can play a part in the prediction process.

JEL classification: M4

Key words: Equity premium; dividend yield; conditional/unconditional models

1. Introduction

The equity risk premium is measured as the extra return that equity holders expect to achieve over risk-free assets on average. In the traditional Capital Asset Pricing Model, the risk premium is the additional return required in compensating investors for one unit of (beta) risk. The spread is also known as the spread between equities and bonds. Despite numerous attempts to estimate the value of this premium, there is some debate as to, which of the many empirical estimates represents the true premium required by equity investors.

While there are many topics in the area of finance upon which academics agree, a topic as basic as the equity risk premium still can produce some vigorous debate. Equity risk premiums are a central component of every risk and return model in finance. The concept of equity risk premium is important to an investor as he or she makes an investment decision. The equity risk premium is the reward that investors require, accepting the uncertain outcomes associated with owning equity securities.

The most common approach to estimating equity risk premiums remains the use of historical returns, with the difference in annual returns on stocks and bonds over a long time period comprising the expected risk premium, in the future. There are limitations to this approach given that the attitude to holding assets has changed over time. Mehra and Prescott (1985) henceforth as MP (1985) introduced this approach. It arises from the observation that the average real return on equity over the last the last century in the USA has been about 7% while the average rate of return on riskless, short term securities has been about 1%. According to Ibbotson Associates, stocks have returned 11 percent a year since 1926, compounded annually, and bonds have returned 5 percent. Accordingly, the historical equity risk premium is around six percent.

MP (1985) argues that the historical level of the ex post US equity premium (over the post 1926 period) is puzzlingly high. In their model, individuals were assumed to have additively separably utility functions and constant relative risk aversion. The relevant parameter in their model is the coefficient of relative risk aversion, A, whose interpretation is that if consumption falls by 1 per cent, then the marginal value of a dollar of income increases by A per cent. In their model MP found that to explain historic equity risk premium, A needed to be between thirty and forty, which was deemed to be much too high[1].

The notion that risk matters, and that riskier investments should have a higher expected return than safer investments, to be considered good investments, is intuitive. Thus, the expected return on any investment can be written as the sum of the riskfree rate and an extra return to compensate for the risk. The disagreement, in both theoretical and practical terms, remains on how to measure this risk, and how to convert the risk measure into an expected return that compensates for risk. This paper looks at the estimation of an appropriate risk premium to use in risk and return models, in general, and in the capital asset pricing model, particularly in an Australian context.

The full study is split into two periods for the pre and post 1987 crash periods. Section 2 describes the model used in this study while section three suggests a model for changes in variance of the risk premium while the data in described in Section 3. Results for the analysis are reported in Section 4 and conclusions are drawn in Section 5.

1.1 Risk and Return Models

While there are several competing risk and return models in finance, they all share some common views about risk. First, they all define risk in terms of variance in actual returns around an expected return; thus, an investment is riskless when actual returns are always equal to the expected return. Second, they all argue that risk has to be measured from the perspective of the marginal investor in an asset, and that this marginal investor is well diversified. Therefore, the argument goes, it is only the risk that an investment adds on to a diversified portfolio that should be measured and compensated.

In fact, it is this view of risk that leads risk models to break the risk in any investment into two components. There is a firm-specific component that measures risk that relates only to that investment or to a few investments like it, and a market component that contains risk that affects a large subset or all investments. It is the latter risk that is not diversifiable and should be rewarded. All risk and return models agree on this fairly crucial distinction but they are different when it comes to the issue of how to measure this market risk.

Modern asset pricing theory suggests that equity risk premia are predictable. Fama (1991) confirms the predictability of U.S. stock market in his survey of empirical studies while Bonomo, Ferris and Lamy (1991), Bekaert and Hodrick (1992), Campbell and Hamao (1992), Clare and Thomas (1992), Cochran, Defina and Mills (1993, 1994) appear to corroborate the same pattern of predictability among international stock markets. The levels of predictability of stock returns are only predictable if we expect the dividend yield over the entire sample to be mean reverting. Goetzmann and Jorion (1995) show the dividends yields show only marginal ability to predict stock returns in the United States and the United Kingdom. They also argue that tests over long periods may be affected by survivorship. Simulations show that regression statistics based on a sample drawn solely from surviving markets can seriously be biased toward finding predictability.

Siegel (1999) says that most studies on US markets dating as far back as 1889 and 1926 are unlikely to predict the equity premium for the future. The real rate of return on fixed income assets is likely to be significantly higher than that estimated on earlier data. This Siegel says is confirmed by the yields available on treasury inflation-linked securities, which currently approach 4%. Furthermore, the return on equities is likely to fall from its former level due to the reduction in transactions costs and other factors which have driven equity prices higher relative to fundamentals.

All of the above factors says Siegel make it very surprising that Ivo and Welch (1998) found that most economists still estimate the equity premium at 5% to 6%. This would require a 9% to 10% return on stocks given the current real yield on treasury inflation-indexed securities. To prevent the P-E ratio from expanding further, real per share earnings would have to grow by nearly 8% to 9% per year given the current 1.2% dividend yield.”

Siegel’s study emphasises reversion to the mean. It seems to imply the bull market could rage on only if history was made or it we were entering a new paradigm. While not making predictions, the author is offering a warning based on available data.

We must note that Siegel’s study is related to the US market and the reasoning behind Ivo and Welch (1998) findings doesn’t necessarily apply to non-US markets. A very interesting paper by Goetzmann and Jorion (1999) says that the US is a very unique market in comparison to other large world equity markets. They state that in the beginning of this century, stock markets in countries like Russia, France, Germany, Japan and Argentina have suffered political turmoil, war and hyperinflation. Assuming there was some probability of disruption for the U.S. market, this probability is not effected in the observed U.S. data. In turn, this will bias the estimates of the equity premium.

Lamont (1998) says that dividend and earnings are important, but only for forecasting short-term movements in expected returns. The relative rate is uniformly unimportant. For long-horizon returns, price is all that matters. Recent low forecasts of returns are due to the fact that stock prices are high. Forecasting models suggest investors look for dividends and earnings short-term, but for long-term buy at low stock prices. The deduction from this behaviour suggests that today's market is appealing from a short-term and not a long-term view.

Goyal and Welch (1999) present a conditional and unconditional model in predicting the equity premium. The dividend yield is commonly thought to predict stock returns as does the historical equity premium average (unconditional) model. Goyal and Welch (1999) find that dividend yield regressions fail to predict out of sample but are good predictors for in sample estimates. Their main argument is the time-varying correlation between the dividend yield and expected returns. They then introduce a learning/changing market model, which suggests time-decay in the dividend yield coefficient. The challenge is to find a model other than the unconditional mean to predict the equity premium. We cannot assume that the dividend yield model can predict the equity premium in the simple linear fashion usually presumed.

We use the lagged dividend yield to predict equity premia or stock returns. The argument put forward my Goyal and Welch is the ability of the model to predict in-sample and out-of-sample. The literature on the dividend yield model is covered in Ball (1978) and more recently in Rozeff (1984) Shiller (1984), Campbell and Shiller (1988), Fama French (1988) and Fama and French (1989). Cochrane (1997) provides an excellent survey of the equity premium literature.

Although Goyal and Welch show good in-sample predictive ability for annual equity premia, the dividend yield has poor out-of-sample predictive ability. There is some doubt about their procedure as the authors have used 20 years of data from 1926 to 1946 to predict from 1947 to 1997. This could result in biased estimates and hence make the results inaccurate.

An Australian study by Bellamy and Heaney (1997) looks at the effect of the dividend yield, yield curve slope and level of interest rates. There is some evidence of statistically significant stock return volatility effect in the risk premium though this only appears in the post crash period. The use of dummy variables is ignored in this study but further advancements in this area should lead to these types of analysis.

Rozeff (1984) showed that dividend yields forecast equity risk premia, as would be predicted by a deterministic dividend discount model. For example, if the stock price represents a claim to the future stream of dividends, the price can be exactly determined assuming constantly growing dividends and a known discount rate. Under the Gordon growth model,

$$P(t\text{-}1) = \frac{D(t)}{r\text{-}g} \tag{1}$$

$$r = g + \frac{D(t)}{P(t\text{-}1)} = g + \frac{(1+g)\times D(t-1)}{P(t-1)} \tag{2}$$

where P is the stock price, D is the dividend, r is the discount rate and g is the constant growth rate of dividends. In our study the stock price is the All Ordinaries price index (an index excluding dividends).[2] In the certainty model, the discount rate is the expected return on the stock. Although the model is not directly applicable to the case in which growth rates and discount rates vary through time, the model suggests that dividend yield should capture variations in expected stock returns.

2. Data

Table 1 lists the data used in this paper. The study is based on monthly stock market data gathered from January 1973 to Oct 1999. The data is sampled from Datastream International™. The total market indices (TOTMKAU) are calculated by Datastream International™ and are a market capitalisation weighted index incorporating approximately

80% of the market value at any given point in time. The Australian index is comprised of 154 stocks with new stock to be added in the near future. The reason we use the total market series (TOTMKAU) from Datastream is due to the lack of available indices calculated on a national basis.

The variables include the return index (RI) which includes reinvested dividends, the price index (PI) which excludes dividends reinvested. We used a 5-year bond rate as an approximate to the risk-free rate. The return index is equivalent to the value-weighted index (VWR) used in most United States studies and the price index is equivalent to the value-weighted index excluding dividends (VWRX). The equity premium is simply the difference between the return on the market with dividends and the risk free rate. Due to inconsistency of data availability from other sources, for example the unavailability of data on the return index (RI) from 1973 we had to use the Datastream series.

We split the sample into two sub-samples, one before the October 87 crash and one after the crash.[3] We need to exclude the crash period so see if the sample after the crash gives us different results, in particular values for the average equity premium and dividend yield.

The derived series of interest will be the equity premium, EQP, and the dividend yield, D(t-1)/P(t-2) and D(t-1)/P(t-1). The dividend yield is calculated as the difference between the value-weighted index with dividends and the value-weighted index excluding dividends:

$$\mathrm{VWR(t-1,t)} - \mathrm{VWRX(t-1,t)} \equiv \frac{\mathrm{RI(t)} - \mathrm{RI(t-1)}}{\mathrm{RI(t-1)}} - \frac{\mathrm{PI(t)} - \mathrm{PI(t-1)}}{PI(t-1)} \equiv \frac{D(t-1,t)}{P(t-1)} \tag{3}$$

We should make a note that $D(t-1,t)$ is the same as $D(t)$, given that we assume $D(t-1,t)$ are flows from last period to this period. The last term can be written as $D(t)/P(t-1)$. To compute $D(t)/P(t)$, we multiply by the market capitalisation ratio $P(t-1)/P(t)$.

[Insert Table 1: Data Sources]

[Insert Table 2: Descriptive Statistics]

Table 2a and 2b provides the descriptive statistics for the series. The mean, standard deviation and median are calculated as annual percentage returns, while the other statistics are based on findings from the monthly data. In table 2a the average log equity premium for the entire sample is 2.95% and the average log dividend yield is 4.13% per annum. The average log equity premium for the period before the 1987 crash is 4.70% and the average log dividend yield is 4.40%, which is higher than the entire sample period. The period after the crash gives an average log equity premium of 5.32% and a dividend yield figure of 4.13%. The average equity premium seems to be higher in the period after the crash when compared to the period before the crash. This finding is not surprising, as the Australian market has performed exceptionally well since the crash. The general movement in the all ordinaries index has been on the up and this further shows a good performance in the Australian stock market since the dramatic downturn in October 1987. There has also been a low level of speculation with the government playing a crucial part by not making any drastic announcements.

The reported skewness for the three periods is negative and statistically significant, indicating that large negative returns are more frequent than large positive returns. The skewness for the entire period is very high compared to the two smaller sub-samples. This is not surprising, as there were significantly high negative returns in the 70s and 80s. The skewness after the crash is relatively low, indicating once again a steady flow of positive returns and very small negative returns on the market. Finally, the reported measure of excess kurtosis indicates that large returns occur more frequently than would be the case if returns were normally distributed. As is pointed in Fama (1965), one explanation for the excess kurtosis in stock returns is that the variance of returns is not constant over time.

Figure 1 plots the time series of the equity premium and the dividend yield. The EQP graph shows significant volatility from the mid-70s to the late 80s. The bond market has performed significantly better than the returns on equities, thus giving us an overall low equity premium. The equity premium seems fairly stationary unlike the dividend yield, which, like most studies is found to be non-stationary. The figure also shows structural breaks over the sample period making it difficult in setting up a model for future predictions, but the graphs are good indicators of the movements in the dividend yield and EQP.

Stambaugh and Pastor (1998) express concern over the estimation of the equity premium when structural break are present. Data before a break are relevant if one believes that large shifts in the premium are unlikely or that the premium is associated, to some degree, with volatility. Stambaugh and Pastor (1998) develop and apply a Bayesian framework for estimating the equity premium in the presence of structural breaks. This study is beyond the scope of this paper and we will try to apply different methods to forecasting the equity premium.

[Insert Figure 1: Time Series Graphs]

3. In-Sample Fit and Out-of-Sample Forecasting

3.1 Forecast Evaluation statistics

Although the creation of good parameter estimates is often viewed as the primary goal of econometrics, to many a goal of equal importance is the production of good economic forecasts. We define the best forecast as the one, which yields the forecast error with the minimum variance. In the single equation regression model, ordinary least-squares estimation yields the best forecast among all linear unbiased estimators. One important statistic is the forecast error variance that there are several ways in which we can measure the forecasting accuracy of a model. In this study we look at the mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE)[4]. The mean square error of a predictor can be broken down into three parts. The first, called the *bias proportion*, corresponds to that part of the MSE resulting from a tendency to forecast too high or too low, reflected by the extent to which the intercept term in the regression of actual changes on predicted changes is nonzero. The third, called the *variance proportion,* measures that part of the MSE resulting from an unpredictable error (measured by the variance of the residuals from this regression). This decomposition (see Theil, 1966) provides useful information to someone attempting to evaluate a forecasting method.

A common statistic found in the forecasting context is Theil's inequality (or "U") statistic (see Theil, 1966), given as the square root of the ratio of the mean square error of the predicted change to the average squared actual change. For a perfect forecaster, the statistic is zero; a value of unity or close to it corresponds to a forecast of "no change." (Note that an earlier version of this statistic has been shown to be defective; see Bliemel, 1973).

3.2 In-sample forecasts

Table 3 correlates the equity premium with the lagged dividend yield and lagged dividend yield changes. We have again estimated bivariate regressions on the three samples. The first equation is based on the lag of the dividend yield based on last year's price; the second uses the lag of the current price. The last two equations were estimated on an experimental basis to see whether or not the changes in dividend yield or the differences, have any explanatory power for movements in the equity premium. The bivariate regressions are based on the following equation.

$$EQP = \alpha + \beta DVYIELD \tag{4}$$

where dividend yield is either $D(t-1)/P(t-2)$ or $D(t-1)/P(t-1)$ and the last two equations are the difference between last years dividend yield and the year before, again using the two different prices.

The results in all three samples are very weak. In the first sample (Feb 73 to Oct 99), our specifications differ slightly from earlier work (as earlier work does from one another), but our conclusions are different to Goyal and Welch (1999) and the Fama and French (1988) specifications. The sample dividend yield regressions for the three sample sizes show different results when compared to the findings in Goyal and Welch (1999).

The more common $D(t-1)/P(t-2)$ performs better for the entire sample and the period before the crash than the (perhaps more uncommon) $D(t-1)/P(t-1)$. Table 3 also shows that, although the dividend yield is a non-stationary variable, changes in the dividend yield do not offer improved fit for the first two sample estimates. The sample after the crash shows a better fit with the dividend yield changes as the independent variable. Our interpretation of a good fit is based on the adjusted R^2 estimates. Using the dividend yield changes based on the difference report some mixed results in predicting the equity premium. The model represented by the lagged differences, using the current price, ie., [D(t-1)/P(t-1) – D(t-2)/P(t-2)] shows the best fit between all four regressions when we the entire sample period is used.

The conclusions drawn from Table 3 are as follows. The dividend yield seem to loose it explanatory power as we progress through the sample period. The more common lagged dividend yield model based on last year's price seems to be a better predictor. This finding has been supported in most of the literature, except where Goyal and Welch (1999) find the model based on the current price is the best predictor. They do however note that their finding may not necessarily be true for other markets in different countries, but they do question previous studies based on US data.

3.3 Out-of-Sample Forecasts

Unfortunately, even a sophisticated trader could not have used the regression in Table 3 to predict the equity premium. Most rational decision-makers do not work with complex model to make their decisions. A trader could only have used a simple model based on past values of the equity premium in attempting to forecast next years value. This is why we display statistics on the prediction errors when the dividend yields model and the unconditional equity premium means forecast are estimated only with historical data. We forecast using two different data sets, one utilising the full sample period and the other from December 1987 to Oct 1999. We exclude the period between February 1973 and the October 1987 due to the high inaccuracy in prediction with out-of-sample estimates. These forecast evaluation statistics are reported in Table 4.

In each box, the first two data columns contain the in-sample prediction errors from the single full period regression model as in Table 3. To do an out-of-sample comparison, we need an initial period to estimate coefficients. Thus we chose (ex ante) the post-crash period (Dec 1987 - Oct 1999) as our out-of sample window. The second two data columns display the in sample Dec 1987 - Oct 1999 residuals standard error from out single full-period regression. The final two data columns display the statistics of most interest: the performance of the out-of-sample rolling prediction errors for the Dec 1987 to Oct 1999 period. Again, each year we use only available historical information to estimate the dividend yield regression. The regression coefficients are used to forecast the equity premium, and the statistics are over the sum-total of out-of-sample single year forecasts errors. The out-of-sample benchmark and null hypothesis is that the next year's equity return is simply the same as the historical average, up to this date. This is denoted as **UNC** (unconditional [ie., without dividend yield conditioning]).

Table 3 had indicated that annual equity premia are well predicted in sample. The top panel considers $D(t-1)/P(t-2)$. The first two data columns of Table 4 show that, when compared to the dividend yield model the unconditional model results in a lower RMSE (root mean square error) than the dividend yield model. The lowest RMSE is observed when we forecast in and out-of-sample for the period after the crash. The Theil's coefficient is very high in most of the sample windows, but during the period prior to the crash is found to be 0.67 for the dividend yield model and 0.68 for the unconditional. This finding indicated that a model based on this period is good one for predicting the equity premium.

One might object to our findings based on issues of statistical power. However, it is unclear what modifications one should make to increase power. Both the null hypothesis (unconditional mean) and the alternative hypothesis (the regression model) are clearly defined in the literature, as are the metrics on which they are compared (RMSE, MAE). Alternative, more powerful models are not at debate here. (We will investigate other models in further studies to come). It is the simple linear regression model, which has been proposed in the academic finance literature as an improvement over the null hypothesis, with a new data set being used.

3.4 A Simple Test for parameter Instability

Given the poor out-of-sample performance, our first question is how an investor should view the out-of-sample misprediction evidence in evaluating the linear dividend yield models. We thus develop a simple test for model stability in the dividend-yield prediction context.

We must adjust for the fact that when the dividend yield is almost a random walk, it can bias the estimated dividend yield coefficient, as pointed out by Stambaugh (1999) and Yan (1999).

$$EQP(t) = x_1 + x_2 \times DVY(t-1) + \varepsilon_E(t)$$
$$DVY(t) = x_3 + x_4 \times DVY(t-1) + \varepsilon_D(t)$$

$$\begin{pmatrix} \varepsilon_E(t) \\ \varepsilon_D(t) \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} x_5 & x_6 \\ x_6 & x_7 \end{pmatrix} \times 10^{-3}\right]$$

where EQP is the equity premium and DVY is the dividend yield, either *D*(*t* - 1)/ *P*(*t* - 1) or *D*(*t* - 1) / *P*(*t* - 2). As before, EQP and DVY are quoted in logs.) We want to match (using some functional specification) the empirically observed 1973:03 to 1999:07 data sample moments:

Dividend Yield	X_1	X_2	X_3	X_4	X_5	X_6	X_7
$D(t-1)/P(t-2)$	-0.035	10.816	0.000	0.897	28.35	0.882	0.092
$D(t-1)/P(t-1)$	-0.029	8.925	0.000	0.827	26.56	-0.731	0.007

One can of course not use the data moments as best models estimates. It is well-known that if the true x_4 is close to 1, the sample x_4 is biased downward; and Stambaugh (1999) and Yan (1999) suggest that inference on x_2 is similarly biased if the true (not sample!) x_4 is close to 1. Our own goal is not to obtain inferences (i.e., significance levels) about x_2, but to test if the best stable model that fits the 1973:3 to 1999:7 data can generate poor out-of-sample performance in line with that observed in the real empirical data.

3.5 Impulse Response Functions

In response to the rigid identifying assumptions used in theoretical macroeconomics during the seventies, Sims (1980) provided what has become the standard in empirical macroeconomic research; vector autoregressions (VAR). Since then, researchers in macroeconomics often compute dynamic multipliers of interest (such as impulse response and forecast-error variance decompositions) by specifying a VAR, even though the VAR per se is, often times, of no particular interest. However, VAR-based impulse response functions are restrictive in a manner seldom recognised. In particular impulse responses are constrained to have the following properties[5] : (1) symmetry, responses to positive and negative shocks are mirror images of each other; (2) share invariance, responses to shocks of different magnitudes are scaled version of one another; (3) history independence, the shape of the responses is independent of the local conditional history; (4) multidimensionality, responses are nonlinear functions of high-dimensional parameter estimates which complicate the calculation of standard errors and have the potential of compounding misspecification errors; and (5) linearity, a VAR is a representation of linear, stochastic difference equations that may not appropriately represent more general economic processes.

Impulse responses are important statistics in their own right and thus avoiding these constraints is a natural empirical objective. Based on this we test the response function of our simple dividend yield regression model. Even though we have not approached our analysis on the basis of a Vector Auto Regressive model this does not necessarily mean we that cannot use an impulse response function approach in testing the dynamics of the estimating equation. Jorda (2004) introduces methods for computing impulse response functions that do not require specification and estimation of the unknown dynamic multivariate system itself. The central idea behind this suggested method is to estimate flexible local projections at each period of interest rather than extrapolating into increasingly distant horizons from a given model, as it is usually done in vector autoregressions (VAR). The advantages of local projections are numerous: (1) they can be estimated by simple regression techniques with standard regression packages; (2) they are more robust to misspecification; (3) standard error calculation is direct; and (4) they easily accommodate experimentation with highly non-linear and flexible specifications that may be impractical in a multivariate context. Therefore, these methods are a natural alternative to estimating impulse response functions from VARs.

The generalised impulse responses from a one standard deviation shock to each of the variables are traced out in Figure 3a and 3b. We use generalised impulse response functions because they are not sensitive to the ordering of the variables in the equation and do not assume that when one variable is shocked, all other variables are switched off. We are simply trying to attempt to guage to what extent shocks to certain variables are explained by other variables particularly the impact on Equity Premium when other variables are shocked.

In Figure 3a we show the reponses prior to 1987 in a bid to see if dividend yields shocks had a different impact on equity premia as opposed to post 1987. This was merely done due to the fact that most of the literature in the 1980s supported the role of dividend yield in predicting stock returns and or market risk premiums. The response of EQP to LDVYIELD shows the lasting effect on the equity premium when lagged dividend yields are shocked. This can be interpreted as, shocks in dividend yield impact the risk premium and hence could play a significant role is explaining movements in the risk premium. The response of equity premium in figure 3b shows that risk premiums settle back to their pre-shock level rather quickly. This could be the fact that dividend yields have become poor estimators of risk premiums since the 1987 crash and are no longer suitable. The application of this form of analysis is original in this context and other researcher's have used different models and approaches to show that dividend yields are no longer good predictors of the equity premium.

4. Conclusion

Although the objective of this research was to provide some insight into changes in the stock market risk premium over time, the CAPM and other asset pricing models show that the risk premium of interest to investors is an ex-ante measure. As a result the direct observation of this premium is not feasible. Fan (2003) makes some interesting observations about misinterpretations of ex-post realisations of the CAPM risk premium in tests of the model.

This paper has shown that the predictive capability of the dividend yield model has declined for out-of-sample estimates, but generally results in good in-sample estimates. Good in-sample performance is no guarantee of out-of-sample performance in the equity premium prediction context. The simple dividend-yield predictions over the 1987:12 to 1999:07 period cannot beat the unconditional historical average equity premium on average, much less do so in a statistically significant manner. A naive market-timing trader who just assumed that the equity premium was "like it has been" would typically have outperformed a trader who had employed the dividend yield model.

Appendix A

Mean absolute Error (MAE): This is the average of the absolute values of the forecast errors. It is appropriate when the cost of forecast errors is proportional to the absolute size of the forecast error. This criterion is also called MAD (mean absolute deviation).

Root mean square error (RMSE) This is the square root of the average of the squared values of the forecast errors. This measure implicitly weights large forecasts errors more heavily than small and is appropriate to situations in which the cost of an error increases as the square of that error. This “quadratic loss function” is the most popular in use.

Mean absolute percentage error (MAPE) This is the average of the absolute values of the percentages errors; it has the advantage of being dimensionless. It is more appropriate when the cost of the forecast error is more closely related to the percentage error than to the numerical size of the error.

Notes

1. To see why this is so, consider a gamble where there is a 50 per cent to double your wealth, and a 50 per cent chance to have your wealth fall by half. If A = 30, then you have the absurd implication of being willing to pay 49 per cent of your wealth to avoid the 50 per cent chance of losing half your wealth. The MP paper has been enormously influential, and has spawned a whole new literature.

2. We must note that the return index (RI) used in this study should not be substituted for the price index (PI) as the RI includes dividends, not the type of price index used in the Gordon growth model.

3. We exclude the observation in November rather than October as the data shows significant changes in the month after the crash.

4. For a brief description of these forecasting measurements see Appendix A.

5. The following list of properties is mostly in Koop et al., 1996.

References

Ball, R., 1978, Anomalies in Relationship Between Securities, yields and yield surrogates, Journal of Financial Economics, 6, 103-126.

Bliemel, F. 1973, Theil's Forecast Accuracy Coefficient: A Clarification. Journal of Marketing Research 10, 444-446.

Bollerslev, T., 1986, Generalised Autoregressive Conditional Heteroscedasticity, Journal of Econometrics 31, 307-327.

Bekaert, G., and R. Hodrick, Characterising Predictable Component in Excess Returns on Equity and Foreign Exchange Markets, Journal of Finance, 47, 467-509.

Bonomo, V., et al., 1991, The impact of discount rate changes on equity markets: the evidence from G-7 countries, Journal of International Financial Markets, Institutions and Money, 1, 85-101.

Brailsford, T., 1995, Market closures and time-varying volatility in the Australian equity market, Journal of Empirical Finance 2, 165-172.

Campbell, J., Y. Hamao, 1992, Predictable stock returns in the U.S. and Japan: a study of Long-term Capital Market Integration, Journal of Finance, 47, 43-69.

Campbell, J., and R. Shiller, 1988, The dividend price ratio and expectations of future dividends and discount factors, Review of Financial Studies 1, 195-228.

Clare, A.D. and S.H. Thomas, 1992, International Evidence for the Predictability of Bond and Stock Returns, Economics Letters, 40, 105-112.

Cochran, Defina and Mills, 1993, International Evidence on the Predictability of Stock returns, The Financial Review, 28, 159-180.

Dhillon, A. and H. Johnson, 1994, The effect of dividend changes on stock and bond prices, Journal of Finance 49, 281-289.

Fama, E., 1965, The Behaviour of Stock-Market Prices, The Journal of Business, 34-105.

Fama, E., 1991, Efficient capital markets, Journal of Finance, 46, 1575-1617.

Fama, E.F., and K. R. French, 1988, Dividend yields and expected stock returns, Journal of Financial Economics 22(1), 3-25.

Fan, S., 2003, GCAPM (1): A Microeconomic Theory of Investments, Working Paper, Social Science Research Network, www.ssrn.com.

Froot, K. and M. Obsfeld, 1991, Intrinsic bubbles: the case of stock prices, American Economic Review 81, 1189-1214.

Goetzmann, W.N., and P. Jorion, 1995, A Longer Look at Dividend Yields, Journal of Business, 68(4), 663-679.

Gordon, M. and E. Shapiro, 1956, Capital Equilibrium Analysis: The Required Rate Of Profit, Management Science 3, 102-110.

Goyal, A. I. Welch, I, 1999, Predicting The Equity Premium, working paper, UCLA

Jorda, O., 2003, Model Free Impulse Responses, Working Paper, Department of Economics, University of California Davis.

Koop, G., Pesaran, M. Hashem and Potter, S. M., 1996. Impulse response analysis in nonlinear multivariate models, Journal of Econometrics, Vol. 74(1), pages 119-147.

Lamont, O., 1998, Earnings and Expected Returns, The Journal of Finance 3, 1563-1587.

Mehra, R. and E. Prescott, 1985, The Equity Premium: a Puzzle, Journal of Monetary Economics 15, 145-161.

Rozeff, M 1984, Dividends Yields Are Equity Risk Premiums, Journal of Portfolio Management 11 (Fall): 68-75.

Shiller, R.J., 1984, Stock Price and Social Dynamics, Brookings Papers on Economic Activity, 457-498.

Siegel, J., 1999, The Shrinking Equity Premium: Historical Facts and Future Forecasts, Working Paper, The Wharton School of the University of Pennsylvania.

Sims, C. A, 1980. Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered, American Economic Review, American Economic Association, Vol. 70(2), pages 250-57.

Stambaugh, R., 1999, Predictive Regressions, Working paper, The Wharton School.

Theil, H., 1966, Applied Economic Forecasting. Amsterdam: North Holland.

Timmerman, A., 1993, How Learning In Financial Markets Generates Excess Volatility And Predictability In Stock Prices, Quarterly Journal of Economics 108, 1135-1145.

Williams, J.B., 1938, The Theory Of Investment Value, Harvard University Press, Cambridge, Mass.

Yan, S., 1999, Can Dividend Yield Predict Future Returns: A Re-Examination of Some Early Evidence, working paper UCLA.

Table 1: Data Sources

Name	Description	Source	Availability
RI	Return Index (including dividends)	Datastream	Feb-1973-Oct 1999
PI	Price index (excluding dividends)	Datastream	Feb-1973-Oct 1999
BD	1 month bond rate	Australian Bureau of Statistics	Feb-1973-Oct 1999
EQP	The equity premium	VWR-BD	Feb 1973-Oct 1999
D(t-1)/P(t-2)	The dividend yield	VWR-VWRX	Feb 1973-Oct 1999
D(t-1)/P(t-1)	The dividend yield	D(t-1)/P(t-2)×[P(t-1)/P(t)]	Feb1973-Oct 1999

Explanation: Parenthesised expressions denote timing. When omitted, assume a time subscript of zero. In all regressions that follow, EQP will lead its predictors by one period. For example, the January 1988 dividend yield (e.g., D (December 1987 to January 1988)/P(Dec1987) would be used to forecast February 1988 equity premia EQP(January 1988 to February 1988)

Table2a: Descriptive Statistics

Entire Sample Period, Feb-1973 to Oct-1999

	In Levels				In Logs								
	N	**Mean**	**Sdev.**	**Median**	**Mean**	**Sdev.**	**Median**	**Min**	**Max**	**Skew**	**Kurt**	**JqBr**	**ADF**
VWR	318	16.55	6.31	17.70	13.55	6.43	17.28	-44.52	20.40	-1.248	11.017	911.97	-8.027
VWRX	318	11.78	6.29	12.86	9.08	6.43	12.79	-44.79	20.11	-1.254	10.957	899.39	-8.013
BD	318	10.34	0.22	10.00	10.33	0.23	9.99	0.38	1.29	-0.044	1.944	14.04	-2.021
EQP	318	5.67	0.06	8.43	2.95	6.44	7.66	-45.52	19.57	-1.266	11.084	925.70	-7.937
D(t)/P(t-1)	318	4.31	0.08	4.12	4.13	0.44	4.09	-6.09	3.13	-8.875	152.056	85.98	-3.778
D(t)/P(t)	318	4.27	0.08	4.08	4.13	0.43	4.08	-5.75	3.22	-8.151	142.066	97.48	-3.798

Sample Period Prior to the 1987 crash, Feb-1973 to October-1987

	In Levels				In Logs								
	N	**Mean**	**Sdev.**	**Median**	**Mean**	**Sdev.**	**Median**	**Min**	**Max**	**Skew**	**Kurt**	**JqBr**	**ADF**
VWR	177	20.12	7.01	18.15	16.61	6.95	18.02	-22.87	20.4	-0.29	3.97	9.34	-5.922
VWRX	177	15.05	6.98	13.94	11.74	6.96	13.85	-23.26	20.11	-0.30	3.96	9.38	-5.908
BD	177	11.42	0.19	11.78	11.42	0.19	11.78	0.43	1.29	-0.12	2.18	5.44	-1.855
EQP	177	7.88	7.01	6.84	4.70	6.95	6.73	-23.66	19.57	-0.28	3.96	9.13	-5.718
D(t)/P(t-1)	177	4.46	0.09	4.31	4.40	0.001	4.31	0.21	0.69	0.92	4.42	39.96	-2.041
D(t)/P(t)	177	4.41	0.09	4.31	4.38	0.09	4.23	0.2	0.75	0.93	4.74	47.56	-2.254

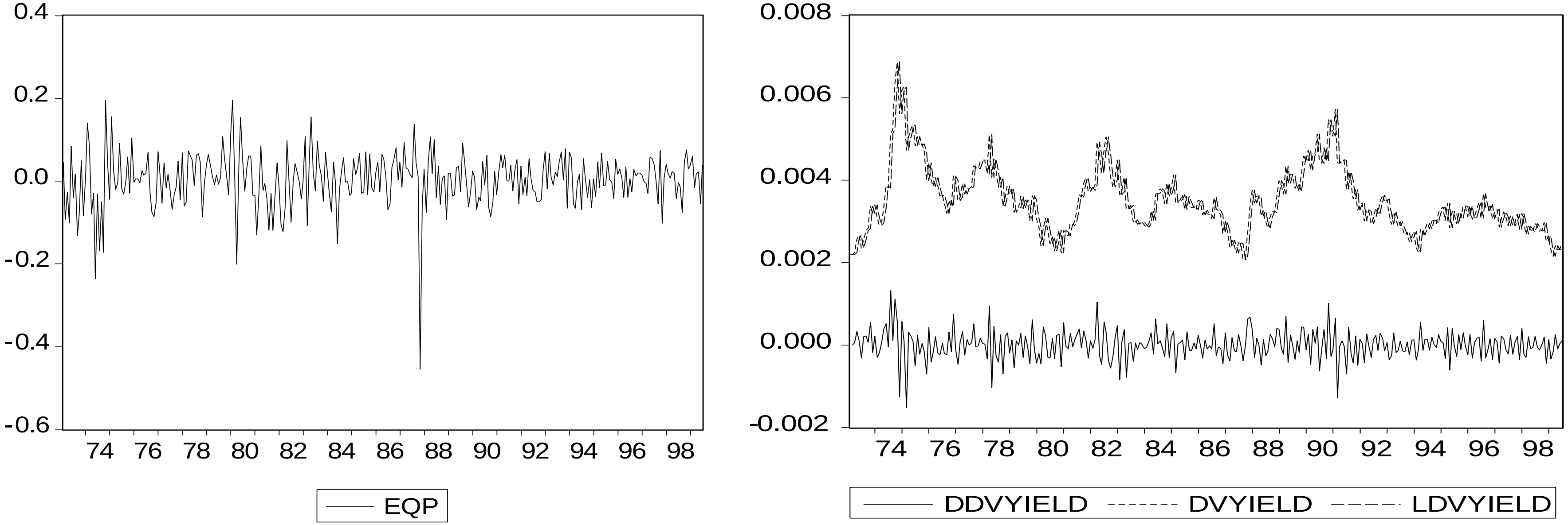
Table2b: Descriptive Statistics

	Sample Period after 1987 crash, Dec-1987 to Oct-1999												
		In Levels			In Logs								
	N	**Mean**	**Sdev.**	**Median**	**Mean**	**Sdev.**	**Median**	**Min**	**Max**	**Skew**	**Kurt**	**JqBr**	**ADF**
VWR	140	16.26	4.44	19.40	14.84	4.39	19.24	-9.75	11.6	-0.13	2.41	2.41	-5.922
VWRX	140	11.64	4.43	15.15	10.32	4.4	15.05	-10.07	11.24	-0.13	2.4	2.45	-5.908
BD	140	9.08	0.23	8.42	9.08	0.23	8.43	0.38	1.15	0.35	1.9	10.04	-1.855
EQP	140	6.63	4.46	8.63	5.32	4.42	8.52	-10.23	10.71	-0.14	2.37	2.78	-5.718
D(t)/P(t-1)	140	4.17	0.07	3.93	4.13	0.07	3.93	0.22	0.57	1.02	3.8	27.79	-2.041
D(t)/P(t)	140	4.13	0.07	3.94	4.10	0.07	3.90	0.22	0.57	1.01	3.82	27.6	-2.254

Explanation: All series are described in Table 1. Throughout the paper, they are measured on a continuously compounding basis. Except where otherwise indicated, the paper reports only results using log variables. Log always means the natural log of 1 plus the value. Every mean and median is significantly different from zero at the 1% level. JqBr is the Jarque-Bera (Jarque and Bera (1987)) test of normality. The critical level of reject normality is 5.99 at the 95% level, 9.21 at the 99% level. ADF is *Augmented Dickey-Fuller* including constant and time trend (Dickey and Fuller (1979)) test for the absence of a unit root. For sample period up to the 1987 crash the ADF values of -4.01 reject the presence of a unit root at the 1% level (-3.437 at the 5% level; -3.142 at the 10% level). For the sample period after the 1987 crash the ADF the critical values are -4.03 at 1% level; -3.44 at 5% level and -3.15 at the 10% level. For the entire sample the critical values are -3.99 at 1% level; -3.43 at 5% and -3.14 at the 10% level.

The results from the three tables are discussed in the main text. All the variables are reported on monthly data, except, the mean and median are reported on an annual basis. We split the sample into 3 different periods to see if the equity premium and dividend yield vary differently prior to the October 87 crash are the period after the crash. We compare the results from these two periods with the entire sample. We exclude data from November 1987 due to the dramatic decline resulting from the October 1987 crash.

Figure 1: Time series Graphs



Explanation: The left graph plots the time series of the log equity premium (EQP). The right graph plots the dividend yield and changes in the dividend yield.

Table 3: Bivariate Regressions Predicting the Equity Premium (EQP) In-Sample

	Sample Period: Feb73 to Oct 99					
Dividend Yield is	**CONST**	**Dividend yield**	R^2	$\overline{R}$	**s.e.**	**N**
$\left[\frac{D(t-1)}{P(t-2)}\right]$	-3.538 -2.159 -2.038	10.816 2.369 2.338	1.75	1.44	6.44	317
$\left[\frac{D(t-1)}{P(t-1)}\right]$	-2.862 -1.822 -1.636	8.925 2.034 1.862	1.30	0.98	6.46	317
$\left[\frac{D(t-1)}{P(t-2)}\right]-\left[\frac{D(t-2)}{P(t-3)}\right]$	-2.750 -1.668 -1.630	8.639 1.883 1.955	1.12	0.80	6.45	316
$\left[\frac{D(t-1)}{P(t-1)}\right]-\left[\frac{D(t-2)}{P(t-2)}\right]$	-4.760 -2.011 -1.231	13.658 2.127 1.488	2.52	1.96	7.66	316

Explanation: Variables are described in Table 1, their descriptive statistics are in Table 2. The dependent variable, the (log) equity premium at year t (in percent), leads the independent variables by one year in all cases expect in the first case where we use the current dividend yield. The first row of each regression model is the coefficient, the second line its OLS t-statistic, the third line its Newey-West heteroskedasticity and autocorrelation adjusted t-statistic. The standard error (s.e.), R^2 and $\overline{R}^2$ (adjusted R^2) are quoted in percent.

Dividend Yield is	Sample Period: Feb73 to Oct 87						Sample Period: Dec 87 to Oct 99					
	CONST	**Dividend yield**	R^2	$\overline{R}$	**s.e.**	**N**	**CONST**	**Dividend yield**	R^2	$\overline{R}$	**s.e.**	**N**
$\left[\frac{D(t-1)}{P(t-2)}\right]$	-3.841 -1.688 -1.780	11.655 1.896 2.147	2.02	1.46	6.92	176	0.235 0.125 0.121	0.586 0.108 0.101	0.01	-0.72	4.43	140
$\left[\frac{D(t-1)}{P(t-1)}\right]$	-2.91 -1.345 -1.257	9.135 1.557 1.482	1.37	0.81	6.94	176	0.404 0.223 0.213	0.586 0.016 0.015	0.00	-0.72	4.44	140
$\left[\frac{D(t-1)}{P(t-2)}\right]-\left[\frac{D(t-2)}{P(t-3)}\right]$	-2.184 -0.945 -1.057	7.193 1.155 1.424	0.77	0.19	6.94	176	-0.369 -0.198 -0.191	2.377 0.439 0.409	0.14	-0.58	4.43	140
$\left[\frac{D(t-1)}{P(t-1)}\right]-\left[\frac{D(t-2)}{P(t-2)}\right]$	-2.509 -1.376 -1.231	8.138 1.376 1.488	1.08	0.51	6.93	176	-0.369 -0.191 -0.087	2.377 0.409 0.336	0.14	-0.58	4.43	140

Explanation: The above two sample periods are used to investigate any differences in parameter values caused by the October 87 crash. The table shows good results up to the crash period, but since December 87 the dividend yield model does not fit the given data. Our conclusion here is simply that the dividend yield has lost its explanatory power in predicting future equity premia movements.

Table 4 continued: Properties of Forecast Errors Predicting the Equity Premium (EQP) Out-of-Sample

$D(t-1)/P(t-1)$ as Forecaster

	in-sample (Feb-1973 to Oct 1999)		in-sample (Feb-1973 to Oct 1987)		in-sample(Dec-87 to Oct 99)		out-of-sample (Dec-87 to Oct 99)	
	DV	**UNC**	**DV**	**UNC**	**DV**	**UNC**	**DV**	**UNC**
RMSE	6.42	6.38	6.90	6.39	4.57	4.63	4.46	4.49
MAE	4.71	4.62	5.33	5.64	3.67	3.69	3.66	3.75
MAPE	199.98	188.23	195.57	193.56	181.58	183.56	184.49	188.56
TIC	0.87	0.81	0.88	0.85	0.88	0.83	0.87	0.91
Bias Proportion	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Variance	0.77	0.71	0.79	0.69	0.71	0.78	0.71	0.73
Covariance	0.23	0.18	0.21	0.26	0.28	0.31	0.29	0.29

Explanation: From the above table and the one previous, we can see that the statistics are very similar to the one in the previous table. A further explanation to these findings is reported in the main body of this paper.

Figure 2: Dividend Yield forecast Evaluation Graphs using $D(t-1)/P(t-2)$ as a forecaster

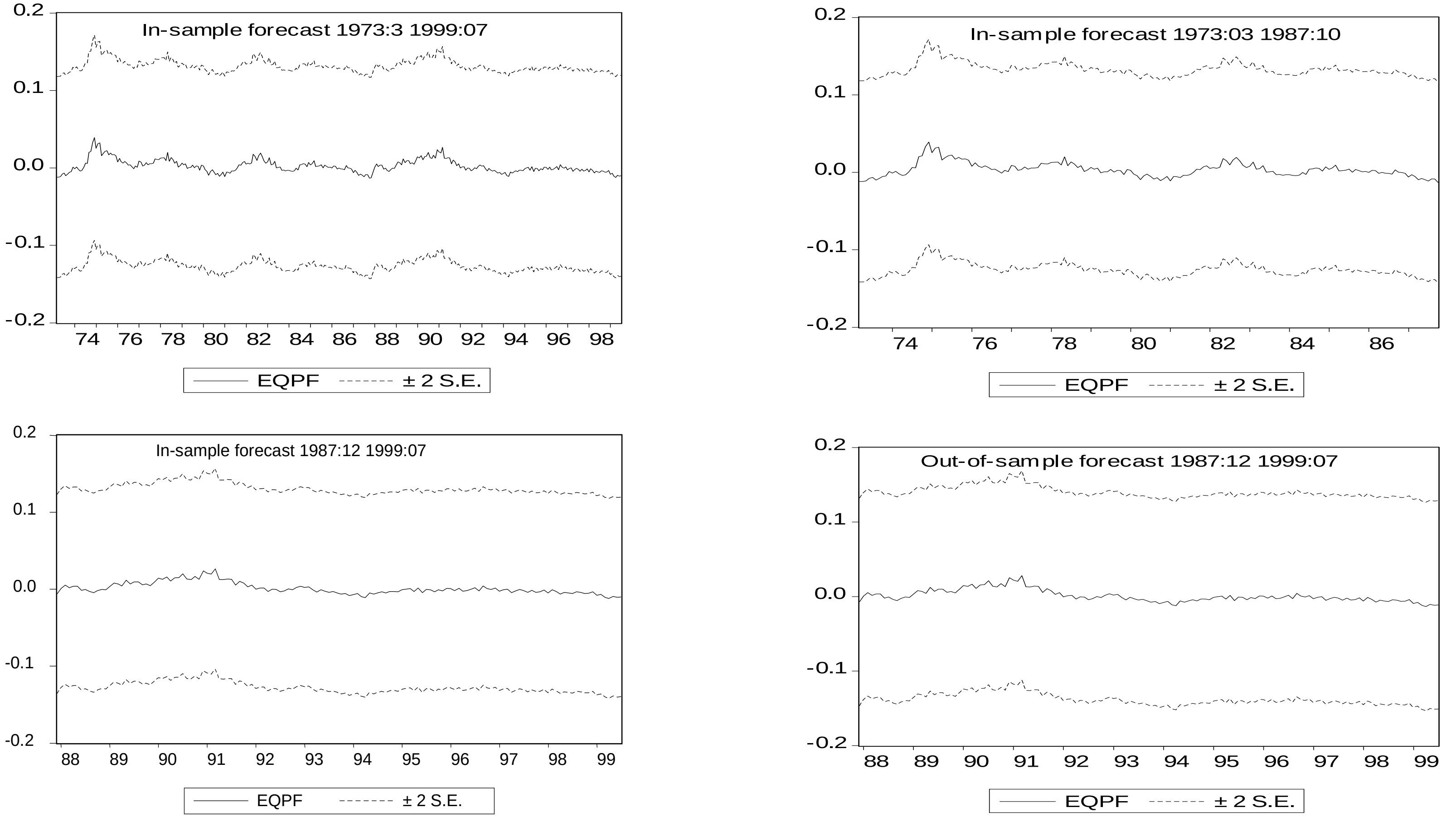
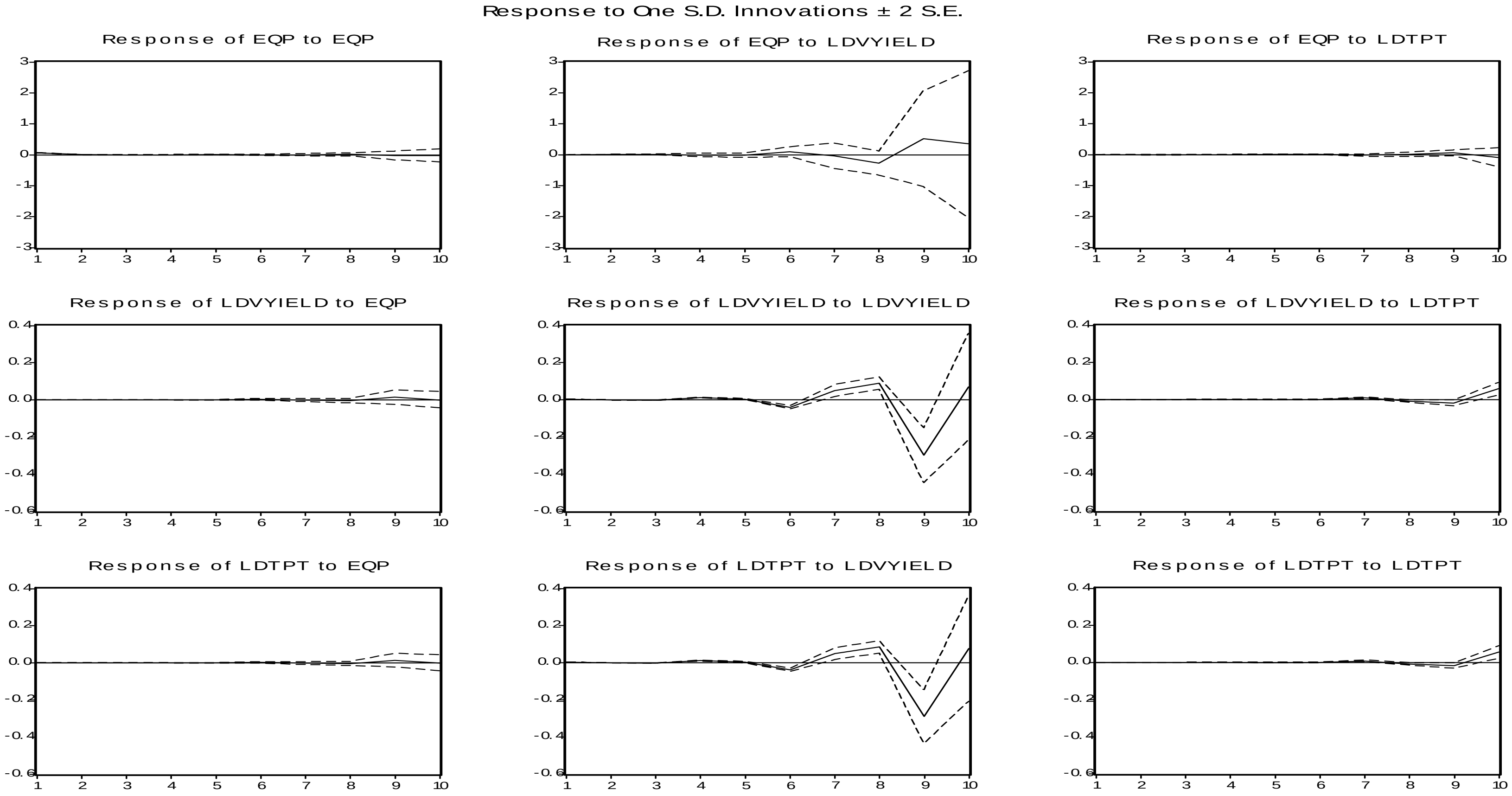


Figure 3: Impulse response functions on the different variables for the entire sample



The above impulse response functions show response of different variables when a one standard deviation shock is imposed of a particular variable. The most important observation is the response of EQP to LDVYIELD.

Figure 3 continued: Impulse response functions on the different variables for the sample period after the 87 crash.