Essays in Asset Pricing and Financial Econometrics

by

Dong Meng Ren

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Doctor of Philosophy

in

Economics

Guelph, Ontario, Canada

© Dong Meng Ren, April, 2016
This thesis includes three chapters, the first two in the field of financial econometrics, and the third one in the area of asset pricing. In the first chapter, we compare the finite sample power of short and long-horizon tests in nonlinear predictive regression models of regime switching between bull and bear markets, allowing for time varying transition probabilities. As a point of reference, we also provide a similar comparison in a linear predictive regression model without regime switching. Overall, our results do not support the contention of higher power in longer horizon tests in either the linear or nonlinear regime switching models. Nonetheless, it is possible that other plausible nonlinear models provide stronger justification for long-horizon tests.

Using finite sample simulation methods, we assess the power of long-horizon predictive tests and compare them to their short-run counterparts, when the true underlying model contains financial asset bubbles. Our results indicate that long-run predictive test using valuation predictors -- specifically the dividend price ratio-- do pick up the return predictability inherent in the asset bubbles. However, after size-adjustment, the long-run predictive framework has a small advantage over its short-run counterpart when the predictor is highly persistent and provides a larger, yet still modest power improvement when the predictor is moderately persistent.

The third chapter proposes a simple Bayesian learning framework to assess leverage ratios in the presence of parameter uncertainty about mean log cash flow. In
particular it can explain why firm's leverage ratios have been observed to increase with firm age. Market values are increasing in uncertainty about mean cash flow and leverage ratios are decreasing with market values. Over the life period of firm, the managers and investors rationally learn from realized cash flows. Due to the convex relationship between cash flow and firm value, ceteris paribus, this results in a decrease in market value and an increase in the leverage ratio. Firm level panel data provides empirical evidence consistent with the model predictions after correcting for the endogeneity of the book to market and profitability control variates. The empirical results suggest that the firm leverage ratio increases over firm age due to learning.
Acknowledgements

Foremost I am deeply grateful to my academic advisor Alex Maynard for his constant advising, support and encouragement throughout my doctoral studies. His superb mentorship and detailed instruction have been of great importance to my personal and academic development. Thanks to his intuitive and clear suggestions, I overcome one difficulty after another during my research. Without him, it is impossible for me make the three chapters possible. I also sincerely thank the other members in my advisory committee: Ilias Tsiakas, Paul Anglin, Mei Li, Thanasis Stengos for their invaluable mentoring and constructive suggestions, and for pushing me to the best.

My sincere thanks also go to Laurent Cellarier, Francis Tapon, Kris Inwood, James Amegashie, Mike Hoy, Johanna Goertz, Nikola Gradojevic, Bram Cadsby for their insightful suggestions and creative questions.

Finally I express my thanks to my family and friends for their constant support.
2.4.1 Present value model: parameter settings and restrictions ........................................41
2.4.2 Empirical specifications and tests ........................................................................43
2.5 Results and Interpretation .........................................................................................45
2.6 Conclusion ................................................................................................................49

3 Capital Structure Assessment and Learning about Cash Flow 57

3.1 Introduction .................................................................................................................57
3.2 Assumptions ................................................................................................................60
3.3 No Uncertainty about Cash Flow .................................................................................62
3.4 Learning about Cash Flow ..........................................................................................65
3.2 Empirical Analysis ......................................................................................................69
3.4 Conclusion ..................................................................................................................74
3.5 Appendix ....................................................................................................................74
# List of Figures

1. Fitted Bear State Probabilities ............................................................. 24
2. Size Adjusted Powers of OLS Tests with Linear Model .......................... 25
3. Powers of Oracle Tests with Linear Model ............................................. 26
4. Powers of Bonferroni Tests with Linear Model ....................................... 27
5. Size Adjusted Powers of OLS Tests with Regime Switch Model ............... 28
6. Powers of Oracle Tests with Regime Switch Model .................................. 29
7. Powers of Bonferroni Tests with Regime Switch Model .......................... 30
8. Powers of IVX Tests with Linear Model .................................................. 31
9. Powers of IVX Tests with Regime Switch Model .................................... 32
10. Market Equity Value versus Mean Log Cash Flow ................................... 64
11. Leverage Ratio Versus Mean Log Cash Flow ......................................... 65
12. Posterior Volatility Versus Age ............................................................ 67
13. Leverage Ratio Posterior Volatility ....................................................... 67
14. Leverage Ratio without Learning ........................................................... 68
15. Leverage Ratio with Learning ............................................................... 69
### List of Tables

1. Estimated Coefficients in RSM without Predictors ................................................................. 23
2. Transition Probabilities Coefficients in RSM without Predictors ............................................. 23
3. Estimated Coefficients in RSM with Predictors ............................................................................. 23
4. Transition Probabilities Coefficients in RSM with Predictors ...................................................... 24
5. Fixed Parameter Setting .................................................................................................................. 50
6. Rejection Rates for Standard Linear Alternatives ........................................................................... 51
7. Rejection Rates for Standard Linear Alternatives ........................................................................... 52
8. Rejection Rates for Bubble Alternatives with Fixed Crash Probabilities ....................................... 53
9. Rejection Rates for Bubble Alternatives with Fixed Crash Probabilities ....................................... 54
10. Rejection Rates for Bubble Alternatives with Time Varying Probabilities ..................................... 55
11. Rejection Rates for Bubble Alternatives with Time Varying Probabilities ..................................... 56
12. Summary Statistics ......................................................................................................................... 70
13. Estimates of Static Panel Data Model without IV ......................................................................... 71
14. Estimates of Static Panel Data Model with IV ............................................................................. 73
Chapter 1

Assessing the Power of Long-Horizon Predictive Tests in Models of Bull and Bear Markets

1.1 Introduction

Long-horizon predictive regression tests have been the subject of substantial interest in both empirical finance and financial econometrics, following influential applications to stock return predictability, [Fama and French (1988b), Campbell and Shiller (1988)], exchange rate prediction [Mark (1995), Chinn and Meese (1995)] and the term structure of interest rates [Fama and Bliss (1987), Cutler et al. (1991)]. The strong empirical results, including large R-squares and t-statistics, observed in long-horizon stock return regressions have been particularly influential in the finance literature, as one of several important findings leading to the conclusion by the Royal Swedish Academy of Sciences (2013) Nobel Press release that

There is no way to predict the price of stocks and bonds over the next few days or weeks. But it is quite possible to foresee the broad course of these prices over longer periods, such as the next three to five years.

Similarly, in his survey article, Cochrane (1999, “New Facts in Finance”) cites the predictability of stock returns at long horizons as one of the three most important new facts in finance.

This paper asks whether there is any long-horizon power advantage in empirically plausible nonlinear models using valid econometric tests. In particular, we compare the power of short and long-horizon predictive regressions under the assumption that the data is generated by nonlinear Hamilton (1989) style regime switching models. These models have been successfully used for some time to model long term swings in foreign currency exchange rates [Engel and Hamilton (1990), Bollen et al. (2000), Dewachter (2001), Cheung and Erlandsson (2005)]. A more recent literature develops similar models to capture bull and bear runs (or multiple states) in stock markets [Maheu and McCurdy (2000), Chauvet and Potter (2000), Guidolin and Timmermann (2005), Guidolin and Hyde (2010), Maheu et al. (2012)]. Schaller and Norden (1997) and Guidolin and Timmermann (2005) estimate predictive regressions with regime switches. To our knowledge, no one has considered the power of long-horizon regressions in these models, although Kilian (1999) mentions this avenue.

Overall, our results are not supportive of long-horizon regression. Long-horizon tests can show modest size-adjusted power gains using a standard t-test in the linear model. Using more

\footnote{This chapter is co-authored with Alex Maynard and has been published. See Alex Maynard, Dongmeng Ren (2014), Assessing the Power of Long-Horizon Predictive Tests in Models of Bull and Bear Markets, in Yoosoon Chang , Thomas B. Fomby , Joon Y. Park (ed.) Essays in Honor of Peter C. B. Phillips (Advances in Econometrics, Volume 33) Emerald Group Publishing Limited, pp.673 - 711 for further reference.}
powerful tests, we find no power advantages at long horizons. Short-horizon tests have as good or better power. Long-horizons tests fare no better (and often worse) in bull/bear regime shift models, even when the regressor is predictive of the regime shift itself. In the context of long-horizon IVX tests, we find that the power comparison depends on both the horizon and the tuning parameter determining the degree of filtering used to create the instrument. Nonetheless, we find that with a good choice of this tuning parameter, the short run IVX can provide as good or better power as its long run counterpart in both linear and regime switching predictive regression settings.

The remainder of the paper is organized as follows: Section 1.2 reviews the literature and provides the background and motivation for our study. In Section 1.3 we introduce the regime shifting predictive regression model that is used as a data generating process for our power comparison. Section 1.4 describes the data. Section 1.5 presents the empirical estimates of the regime shift model used to anchor our simulations. Section 1.6 describes the short and long-horizon specifications and tests employed in the power comparisons. Section 1.7 provides an extensive simulated power comparison of short and long-horizon regressions using two different long-horizon specifications and several long-horizon tests.

1.2 Literature, Background, and Motivation

The dependent variables in long-horizon regressions are typically multi-year asset returns observed at a monthly or quarterly frequency, defined by

\[ r_{t+k} = \sum_{j=1}^{k} r_{t+j}, \]  

(1)

where \( k \) is the horizon period and \( r_t \) is the one period return. The long-horizon return is regressed on a pre-determined predictor \( x_t \), such as a dividend or earnings price ratio, resulting in the long-horizon regression specification

\[ r_{t+k}^k = \beta_0(k) + \beta_1(k)x_t + \varepsilon_{1,t+k}^k \]  

(2)

with tests for long-horizon predictability based on the restriction

\[ H_0 : \beta_1(k) = 0. \]  

(3)

The special case in which \( k = 1 \) corresponds to the short-horizon predictive regression

\[ r_{t+1} = \beta_0 + \beta_1 x_t + \varepsilon_{1,t+1}, \]  

(4)

which has also generated substantial interest. Although predetermined, valuation predictors, such as the dividend and earnings price ratios, tend to be both persistent and highly endogenous.
Following Cavanagh et al. (1995), they have been traditionally modelled using the local-to-unity process [Phillips (1987), Chan and Wei (1987)] and this has more recently been generalized to include mildly integrated processes [Phillips and Magdalinos (2007), Phillips and Lee (2012)]. The long horizon, persistence, and endogeneity combine to create a nonstandard inference problem. They have thus generated an increasing interest in financial econometrics, including the development of new long-horizon tests [Valkanov (2003), Liu and Maynard (2007), Hjalmarsson (2012), Phillips and Lee (2013)].

Although his interest in this particular problem appears fairly recent, it seems fair to say that Professor Peter C.B. Phillips has made at least three very important contributions to this literature. The first contribution is an indirect contribution. In fact, the long-horizon regression provides one of many nice examples of the useful insights that can be gained from the nonstationary asymptotics developed in very large part by Professor Phillips. The empirical results from these regressions appear remarkably strong and are of clear importance to economics and finance. At the same time, early simulation studies detect substantial size distortion [Kim and Nelson (1993), Goetzmann and Jorion (1993)]. Allowing the horizon to increase with \( k \) and modelling the predictor as a near unit root, Valkanov (2003) uses the tools of (near) nonstationary asymptotics to provide a clear explanation for the size distortion observed in earlier studies. In his framework, the regressor \( x_t \) on the right hand side of (2) is near integrated, while the dependent variable (1) is partially summed since \( k \) grows with the sample size, \( T \). Consequently, under the null hypothesis in (3), the asymptotics resemble in certain respects the spurious regression asymptotics of Phillips (1986), helping to explain both the size distortion and large \( t \) and R-squared statistics in empirical applications.

Further insights are provided by Hjalmarsson (2012), who shows that when the regression coefficient from the long-horizon regression is scaled by \( k \) it shares the same second order endogeneity bias as the short-horizon regression estimator of \( \beta_1 \) in (4). Often referred to as the Stambaugh (1999) bias, following Cavanagh et al. (1995) this bias is perhaps best understood by viewing the large sample behavior of the predictive regression in (4) as a special case of cointegrating regression asymptotics [Park and Phillips (1988), Park and Phillips (1989), Phillips and Hansen (1990), Phillips (1991), inter alia], generalized to near unit roots [Phillips (1987), Phillips (1988), Chan and Wei (1987)]. Thus, the underlying theoretical frameworks developed earlier by Professor Phillips have been employed to provide a great deal of insight into this important empirical application.

More recently, Professor Phillips has made two important direct contributions to this literature. One of the challenges in predictive regression is that, if the predictor is modelled as a near unit root process, the asymptotic distribution of the test statistic typically depends on the local-to-unity parameter which cannot be estimated. Until recently, many of the econometric methods that have been designed to provide inference in predictive regression, including long-
horizon regressions, rely on a Bonferroni bound employing a first stage confidence interval for the local-to-unity parameter based on the inversion of a unit root test as in Stock (1991). Phillips (2014) has recently uncovered serious concerns with this approach, showing that the Stock (1991) confidence interval for the local-to-unity parameter has zero asymptotic coverage when the data generating process is stationary. This can lead to the invalidity of the resulting Bonferroni based tests when the true process is mildly integrated or stationary.

In a new major breakthrough, Phillips and Magdalinos (2007) generalize the local-to-unity model to allow for mildly integrated processes and Magdalinos and Phillips (2009) use this new theory to propose an alternative solution to the predictive regression known as the IVX method, which uses a mildly filtered version of the predictor as an instrument. Their solution altogether eliminates the endogeneity bias and dependence on an unknown local-to-unity parameter without reliance on Bonferroni methods. It continues to work for stationary and mildly integrated regressors and is easily extended to multivariate settings.\footnote{It also solves the much more general problem of cointegration with near integrated regressors, see Elliott (1998).}


While the econometric literature discussed above is primarily concerned with understanding or eliminating the size distortion in long-horizon predictive regression, in this paper we are interested in the power of long-horizon specifications relative to their short run counterparts in (4). This has been a controversial question, with the presumed power advantages of long-horizon regressions both strongly criticized in Boudoukh et al. (2008) ["The Myth of Long-Horizon Predictability"], and valiantly defended by Cochrane (2008).

There seems to be an unstated assumption in much of the early empirical literature that predictive regressions are more powerful at longer horizons, perhaps due to economic intuition or simply as a result of the stronger empirical results reported at longer horizons, see e.g. Campbell et al. (1997, Chapter 7). This proposition is stated more directly in Campbell (2001). Several works provide support for this conjecture [Campbell (2001), Kilian and Taylor (2003), Wohar and Rapach (2005), Mark and Sul (2006), Cochrane (2008)], while others question the power advantage of longer horizons [Ang and Bekaert (2007), Boudoukh et al. (2008), Hjalmarsson (2008), Hjalmarsson (2012)].

With a few exceptions, the power properties of long-horizon regressions have mainly been examined assuming that the true model for returns is given by a short-horizon linear model similar to (4) [Campbell (2001), Valkanov (2003), Mark and Sul (2006), Cochrane (2008), Hjalmarsson (2008), Hjalmarsson (2012)]. In this case, as both Campbell (2001) and Hjalmarsson (2012) remark, the short-horizon regression in (4) is a correctly specified model. This may lead one to question the advantage of (2), relative to the correctly specified regression in (4). Nonetheless, as
Campbell (2001), Mark and Sul (2006), and Cochrane (2008) all astutely note, the short-horizon regression loses power in the realistic case in which the valuation predictors are both highly persistent and contemporaneously correlated with the residual. They argue that the long-horizon regression offers power improvements in this case.

In fact, this combination of near unit root regressor with contemporaneous endogeneity, gives rise to second order biases in (4), resulting in both size distortion and power loss [Phillips and Hansen (1990), Phillips (1991), Campbell and Yogo (2006)]. However, this arguably reflects only an inefficient use of the information in (4) by OLS based tests that are sub-optimal in the combined presence of endogeneity and (near) nonstationarity. Since more efficient short run estimators and more powerful short run tests are now available [Jansson and Moreira (2006), Campbell and Yogo (2006), Magdalinos and Phillips (2009), inter alia] one can still argue that there is no inherent power advantage to the long-horizon specification when the true model is the linear short-horizon predictive regression. In our simulated power comparisons we do find some small size-adjusted power improvements using OLS based tests when the true model is linear, but these improvements disappear when more powerful tests are used. Hjalmarsson (2012) also provides some results to support this contention.

In our view, any inherent advantage to the long-horizon specification could only arise under nonlinear alternatives to (4). In this case, neither (4) nor (2) is correctly specified, so that the long run specification in (2) may plausibly provide a better approximation to the true, but unknown, nonlinear alternative. Kilian (1999) similarly argues that “the observed pattern of p-values is inconsistent with a linear model”. Yet, we are aware of relatively little literature which investigates the power of long-horizon specifications when the data generating process is nonlinear. Kilian and Taylor (2003) and Wohar and Rapach (2005) simulate power using a nonlinear ESTAR model for the predictor finding power advantages at longer horizons. On the other hand, Ang and Bekaert (2007) simulate power in a nonlinear present value model, but find long-horizon regressions to be less powerful. Marmer (2008) shows that predictability is present only in the short run in the case when the predictive component of returns is a general integral transformation of a nonstationary predictor.3

We compare the power of long and short-horizon predictive tests in the case when the true model is characterized by regime switching between bull and bear markets. We assume that the econometrician does not know the true model and tests predictability using standard linear predictive models without regime shifting in both the short and long-horizon specifications. Since neither linear regression is correctly specified, it is no longer obvious that a powerful

---

3Kasparis et al. (2012) and Cai and Gao (2013) find evidence of nonlinear predictive power in the dividend yield for stock returns, while the results in Juhl (2011) are suggestive, but insignificant. Gonzalo and Pitarakis (2012) find evidence of structural breaks in the predictive regression relationship. However, none of these papers considers the implications for long-horizon predictive tests.
short-horizon test should outperform a long-horizon test. In fact, since the states are themselves persistent, with a probability of switching between states that depends on the predictor, it seems a priori plausible that the long-horizon regressions might outperform the short-horizon regressions without the Markov-switching component.

1.3 Regime Switching Predictive Regressive Models

We consider the following regime switching model

\[ r_{t+1} = \beta_0(s_t) + \beta_1(s_t)x_t + \sigma_1(s_t)\varepsilon_{t+1}, \quad (5) \]

where the state \( s_t \in 1 \ldots N \) is a first order Markov process with transition probabilities

\[ p_{ij} = \text{Prob}(s_t = j|s_{t-1} = i). \quad (6) \]

When \( N = 1 \) and there is just one single state, then the regime switching model specializes to the standard linear predictive model

\[ r_{t+1} = \beta_0 + \beta_1 x_t + \sigma_1 \varepsilon_{t+1}. \quad (7) \]

In order to allow the predictor to predict transitions between states, as well as returns within states, we focus our attention on a model with time varying transition probabilities. Following Ding (2012), we model them as (where \( \Phi \) is the CDF of \( N(0,1) \)):

\begin{align*}
  p_{ij,t} & = \Pi_{l=1}^{j-1} (1 - q_{il,t}) q_{ij,t} \quad (8) \\
  q_{ij,t} & = \begin{cases} 
  \Phi (\theta_{0,ij} + \theta_{1,ij} x_{t-1}) & \text{for } 1 < j < N - 1 \\
  1 & \text{for } j = N.
\end{cases}
\end{align*}

This formulation allows us to model the transition probabilities as a function of the predictor, while still ensuring that these probabilities appropriately sum to one. For example, in the two-state model \( N = 2 \), this simplifies to

\begin{align*}
  p_{11,t} & = q_{11,t} = \Phi (\theta_{0,11} + \theta_{1,11} x_{t-1}) \quad (9) \\
  p_{21,t} & = 1 - p_{11,t} \quad (10)
\end{align*}

in which \( x_t \) may help predict transitions between bull and bear states.

To provide an economic interpretation, we follow the literature in referring to the states as bull and bear states in the two state model and as bull, normal, and bear states in the three state model. To avoid any ambiguity, we define the bull and bear states, respectively, as the states with highest and lowest values of the estimated intercept \( \hat{\beta}_0(s_t) \) in (5).
1.3.1 Persistent, Endogenous Regressors

Although predetermined, common valuation predictors, such as the dividend or earnings price ratios, are both persistent and endogenous. It is typically difficult to reject unit roots in these variables. Yet, taken literally, a unit root in these predictors may not seem natural from a financial modelling perspective. Moreover, time series unit root tests are, by design, not consistent against near unit root alternatives. Accordingly, the local-to-unity or near unit root models [Phillips (1987), Phillips (1988), Chan and Wei (1987)] have been increasingly used to model the predictor. This class of models has been recently generalized by Phillips and Magdalinos (2007) to include mildly integrated series. Consider, for simplicity, a first order autoregressive process for the predictor

$$x_t = \rho_0 + \rho_1 x_{t-1} + v_t.$$  \hspace{1cm} (11)

A very general formulation for $\rho_1$ is provided by Phillips and Lee (2013), who model $\rho_1$ as

$$\rho_1 = 1 + c/T^\eta.$$ \hspace{1cm} (12)

Equation (12) specializes to a unit root process, in the case when $c = 0$, a near integrated process when $\eta = 1$ and $c < 0$ and a mildly integrated series when $0 < \eta < 1$.\(^4\)

In order to realistically capture the endogeneity in the data, we follow the linear predictive regression literature in modelling the innovations as contemporaneously correlated with correlation $\delta = \text{corr}(\varepsilon_t, v_t)$:

$$\left(\varepsilon_t, v_t/\sigma_2\right)' \sim \text{i.i.d.} \left(0, \begin{bmatrix} 1 & \delta \\ \delta & 1 \end{bmatrix}\right).$$ \hspace{1cm} (13)

This error structure is common in both the predictive regression and regime shift literatures. In the absence of regime shift, it nonetheless has the disadvantage of imposing homoskedasticity. In our context, we allow for the conditional return variance $\sigma^2_t(s_t)$ to vary across states. Incorporating time varying conditional heteroskedasticity within states could further enhance the realism of the model, although possibly at the cost of complicating its estimation.

1.4 Data and Variables

Our study is based on 1048 monthly observations of the excess return, including dividends, and the dividend price ratio on the S&P 500 price index starting in July 1926 and ending in January, 2013. The stock return and dividend data are maintained by Robert Shiller and are combined with the monthly risk free rate maintained by Kenneth French. The monthly dividends are linearly interpolated from yearly totals. Defining the level of the S&P 500 price index by $P_t$,

\(^4\)Locally ($c > 0, \eta = 1$) and mildly ($c > 0, 0 < \eta < 1$) explosive processes are also allowed for in (12).
dividends by $D_t$, and the risk free rate by $i_t$, we define the excess return as

$$r_t = \ln \left( \frac{P_t + D_t}{P_{t-1}} \right) - i_t. \quad (14)$$

Following common practice [see for example, Fama and French (1988a) and Chapter 7 of Campbell et al. (1997)], we sum the monthly dividends over the past twelve months to obtain

$$dp_t = \ln \left( \frac{D_t + D_{t-1} + \ldots + D_{t-11}}{P_t} \right). \quad (15)$$

Dividends are often summed as in (15) on account of their strong seasonality. Although not standard practice, this could also be achieved by a yearly average of dividends. It is possible that smoothing the dividends may also mitigate any bias from the linear interpolation of yearly dividends.\(^5\)

Following (Schaller and Norden 1997, footnote 14, page 14), who estimate a regime switching predictive model with fixed coefficients, we then standardize $dp_t$ to have zero-mean and unit variance in order to define our predictor as

$$x_t = \left( dp_t - \bar{dp}_t \right) / \hat{\sigma}_{dp}, \quad (16)$$

where $\bar{dp}_t$ and $\hat{\sigma}_{dp}$ are the sample mean and standard deviation of $dp_t$ respectively. Although this normalization is not typical in the linear predictive regression literature, we also find that it leads to a somewhat better fit and more easily interpretable parameters when estimating a regime switching predictive regression model.

If $dp_t$ is (near) nonstationary, it will have an infinite range, tending to take values far from zero. Without the normalization in (16), this would lead the transition probabilities, defined as an asymptotically homogeneous function of the predictor in (8),\(^6\) to cluster at zero or one depending on the signs of $\theta_1$ and $x_t$. By normalizing the standard error, we avoid this problem. On the other hand, this normalization would present additional challenges for asymptotic inference in the regime shift model, since $x_t$ is a function of the dividend price ratio process in (15), which may itself be characterized by a near unit root, in which case the standard deviation in (16) is increasing in sample size. We leave these developments for future work. In this paper, we simply employ the parameter estimates of the regime switching model to parameterize the data generating process used in our simulated power analysis.

1.5 Empirical Estimates

Before turning to the regime switching regression model, we first estimate a regime switching model for the mean and variance of the excess return in (14). In this way, we explore the distinct

\(^5\)Recent research by Ghysels and Miller (2014) uncovers potentially serious bias in the use of interpolated data in the context of cointegration testing.

properties of the returns in the bear, bull, and normal states of the market. This also allows us to assess the significance of the coefficients and the fit of the regime switching model, without the additional complications induced by time varying transition probabilities and the inclusion of a persistent predictive regressor.

Tables 1 and 2 respectively show the estimated coefficients and transition coefficients in the pure regime switching model without predictors. Table 1 provides the estimates of $\beta_0(s_t)$ and $\sigma_1(s_t)$ for each state $s_t$, when restricting $\beta_1(s_t) = 0$ in (5). Likewise, Table 2 provides estimates of the transition probabilities (6), which are non-time varying in the absence of a predictor.\(^7\) Standard errors are given in parenthesis.\(^8\)

For the purposes of comparison, we include the results for the single state model in column 2 of Table 1.\(^9\) This shows a mean (log) excess return for the S&P500 of 0.0398, or approximately four percent, over the entire sample period and a return variance of 0.0022. Columns 3 and 4 show the estimated bear and bull state values for the two state model. The mean return in the bull state is more than two percentage points higher than in the bear state, whereas the variance of the bear state is nearly seven times as high as that of the bull state, reflecting the greater turbulence that occurs during market downturns. In both cases, the difference between the estimates in the two states is several times bigger than the larger of the two standard errors, lending statistical support to the observation that the market behaves differently during bear and bull states. In the three state model (columns 5-7), the bull and bear states capture the more substantial rallies and downturns, with calmer periods absorbed by the normal state. Consequently, the three state models shows a much larger differential of almost five percent between average returns in the bull and bear markets. The bear market again shows the greatest volatility, while the normal state has the lowest return variance.

Columns 2-3 of Table 2 show the transition probabilities between bear and bull states in the two state model. Both states are quite persistent, but the probability of switching from bear to bull state in any period is substantially higher than the probability of switching from bull to bear. This reflects the greater duration of bull states. Columns 4-6 of the same table provide the transition probabilities for the three state model. All three states are again persistent, but the normal and bull states are more persistent than the bear state. In all cases, the standard errors suggest that the estimates are reasonably precise.

The last row of Table 1 shows the Hannan and Quinn (1979) information criteria (H-Q IC) for each model, with lower, more negative values denoting a better penalized fit.\(^10\) Both regime

\(^7\)Omitting the predictor restricts $\theta_{i,j} = 0$ in (8) for all $i,j$.
\(^8\)We estimate MRSM by using Maximum likelihoods Methods.
\(^9\)We are grateful to Marcelo Perlin for use of his MATLAB code for the estimation of the fixed transition probability Markov Switching Models. His code is available online and documented in Perlin (2012).
\(^10\)Guidolin and Hyde (2010) similarly employ the H-Q IC to aid the selection of the number of states in a regime switching model.
switching models are strongly favored relative to the single state model and the three state model shows a slight improvement over the two state model.\textsuperscript{11}

We now include the dividend price ratio as a state-dependent predictor for the returns themselves in (5).\textsuperscript{12} This arguably allows for a more realistic specification of the predictive regression model, which incorporates the difference in expected return and variance across the bear and bull states. It also allows for the predictive power of the dividend price ratio to vary across states. In addition, we allow the dividend price ratio to help predict the transition probability between bear and bull states via its inclusion in the model for the time-varying transition probabilities in (8). To the extent that valuation variables may predict transitions between bull and bear markets, this may have stronger practical implications than return predictions within a given state.

Drawing inference in the regime shift predictive regression model is presumably complicated by a similar second order bias as in the linear predictive regression model. Deriving techniques that avoid second order bias and size distortion in this highly nonlinear model is a worthy goal for future research, but lies beyond both the scope and needs of our current paper, in which we require only consistent model estimates in order to anchor our simulated power comparisons in Section 1.7. In a related context, Gonzalo and Pitarakis (2012) develop an IVX based predictive regression test with structural breaks. Although their model differs in a number of important respects from that in (5) and (8), their finding offers general support for the existence of breaks or regime changes in predictive regression.

Table 3 shows the parameters estimates and standard errors in parenthesis in the regime switching predictive regression model (5) with time varying transition probabilities given by (8). The results for the two state model are shown in columns 3-4. As in Table 1, the bull state has a higher mean return, whereas the bear state has a higher standard deviation. The predictive power of the dividend yield (if significant) is also about twice as large in the bear state as in the bull state. Gonzalo and Pitarakis (2012) recently come to a similar conclusion in a model of predictive regression with structural breaks. Interestingly, in the three state model shown in columns 5-7, it is the normal state in which the dividend yield is most strongly predictive, with the coefficient on the dividend price ratio negative and estimated with a large standard error in the bear state.

Table 4 shows the estimates and standard errors for the parameters in (8) governing the transition probabilities. Here, we focus our comments on the two state model in columns 2-3,

\textsuperscript{11}In additional results, omitted for brevity, a marginally better H-Q IC value was obtained in the four state, but at the cost of large standard errors, resulting in the poor identification of a number of the model coefficients. 
\textsuperscript{12}We are grateful to Zhuangxin Ding for use of his MATLAB code. As documented in Ding (2012), this code extends Perlin (2012)’s original code in order to allow for time varying transition probabilities. In order to verify that global maximums were obtained, we compared estimation results and likelihood values across several perturbations to our initial starting values for the optimizations.
whose parameters are the most readily interpretable, as seen from (9). The first panel provides the estimates of the intercepts (the $\theta_{o,ij}$ in 8). The diagonal element of the intercepts ($\hat{\theta}_{o,11}$) is large and positive, while the off-diagonal element ($\hat{\theta}_{o,21}$) is negative. This indicates that the states are again persistent, with a relatively small probability of switching in any single period. The second panel in Table 4 shows the slope coefficients (the $\theta_{1,ij}$ in 8). The positive estimate of $\hat{\theta}_{1,11}$ in the second column suggests that the conditional probability of remaining in a bull market is highest when the dividend price ratio is large. Put another way, a switch from bull to bear market, possibly via a market crash, appears more likely when the stocks are highly valued relative to dividends (low dividend price ratio). The negative estimate of $\hat{\theta}_{1,21}$ suggests that the transition from bear to bull market is also more likely when stocks are relatively higher valued. This may reflect the observation that bull markets sometimes end in sudden crashes, whereas the transition from bear to bull often follows a more gradual market rally, which allows the stock prices to partially recover prior to the onset of the next bull market.

The last row of Table 3 shows the Hannan and Quinn (1979) information criteria (H-Q IC) for each model. The inclusion of the dividend price ratio leads to a lower H-Q IC in all models, although it is unclear if the difference is significant. Among the models with the dividend price ratio included in Table 3, the two state model has the lowest H-Q IC. Its value is substantially lower than the single state model or linear predictive regression. The two state model also has a somewhat better fit than the three state model. We therefore employ the two state model with the dividend price ratio as our baseline model in which to anchor our simulations in Section 1.7. Figure 1 shows the fitted probabilities of a bear state in this model, with the calendar date shown on the horizontal axis and the major market downturns listed in the figure caption. Although during a majority of the time the probability of the bear state is low, the model appears to identify nearly all of the major historical downturns and/or market crashes as bear states. Indeed, it would appear to do quite a good job in capturing the major periods that most market observers would classify as bear states.

### 1.6 Long-Horizon Regression Specifications and Tests

In the simulated power exercise that follows, the empirical researcher is not assumed to know the true model. Instead it is assumed that either a short or long-horizon linear predictive regression is employed as an empirical specification. In particular, we consider two long-horizon specifications from the literature, both of which share the same null hypothesis, but which differ in the formulation of their alternative. Both involve the choice of a horizon length ($k$), which in some asymptotic analysis is modelled as a function of the sample size $T$ [Valkanov (2003), Phillips and Lee (2013)] and in others taken to be arbitrarily large, but fixed, as in Hjalmarsson
A fairly general formulation covering both of these possibilities is given by

\[ k = \lambda T^v \quad \text{for} \quad 0 \leq v \leq 1, \quad \lambda > 0, \quad \text{and} \quad \lambda I(v \neq 0) \leq 1 \]  

(17)

where \( I() \) denotes the indicator function. In (17), \( v = 0 \) corresponds to a fixed horizon as in Hjalmarsson (2011), \( (v = 1, \ 0 < \lambda < 1) \) corresponds to a horizon that grows at a fixed fraction of sample size as in Valkanov (2003), and \( (\lambda = 1, \ 0 < v < 1) \) allows a horizon that grows more slowly than \( T \), as in Phillips and Lee (2013).

### 1.6.1 Standard Long-Horizon Specification

One of the most common versions of the long-horizon specification is given by (2), in which \( r^k_{t+k} \) is the long-horizon \((k\) period) return in (1). The null hypothesis of non-predictability is tested via a test of the parameter restriction in (3) under the maintained hypothesis\(^{13}\) that

\[ E_t \varepsilon^k_{1,t+k} = 0. \]  

(18)

### 1.6.2 Rearranged Long-Horizon Regression

We also consider a rearrangement of this regression originally proposed by Cochrane (1991) and Jegadeesh (1991) and given by

\[ r_{t+k} = \beta_0^*(k) + \beta_1^*(k)x_t^k + \varepsilon^*_t \]  

(19)

in which a one period return is regressed on the long-horizon (or partially summed) regressor

\[ x_t^k = \sum_{j=1}^{k} x_{t+k-j}. \]  

(20)

The null hypothesis of non-predictability is then tested using the coefficient restriction

\[ H_0 : \beta_1^*(k) = 0 \]  

(21)

under the maintained assumption that

\[ E_t \varepsilon^*_t = 0. \]  

(22)

Liu and Maynard (2007) use this same rearranged regression to propose an exact nonparametric test of long-horizon predictability based on a sign test. Phillips and Lee (2013) employ this specification to provide a long-horizon version of the IVX solution to the predictive regression problem.

---

\(^{13}\)The maintained hypothesis in (18) ensures the \( r_{t+k} \) is unpredictable under the null in (3). We refer to it as a maintained hypothesis, since it is not usually explicitly tested in this literature.
Since \(x^k_t\) is realized at time \(t + k - 1\), the rearranged regression in (19) can also be viewed as a short-horizon regression that uses a long history of the predictor \(x_t\) as its regressor. Note, however, that the temporal distance between the one period component returns and one period component predictors in the two regressions, (2) and (19), are the same. That is, \((r_{t+k}, x^k_t) = (r_{t+k}, x_{t+k-1} + x_{t+k-2} + \ldots + x_t)\) and \((r^k_{t+k}, x_t) = (r_{t+1} + r_{t+2} + \ldots r_{t+k}, x_t)\) both involve spacings of 1, 2, \ldots and \(k\) periods between \(r_t\) and \(x_t\). Therefore, when \((r_t, x_t)\) are jointly covariance stationary the rearrangement in (19) can be motivated by expressing \(\beta_1(k) = \text{cov}(r^k_{t+k}, x_t) / \text{var}(x_t)\) and \(\beta^*_1(k) = \text{cov}(r_{t+k}, x^k_t) / \text{var}(x^k_t)\) and noting that the covariances in the numerator are equivalent. In other words, the null hypotheses that \(\beta_1(k) = 0\) and that \(\beta^*_1(k) = 0\) impose the same orthogonality restriction, viz. \(\text{cov}(r^k_{t+k}, x_t) = \text{cov}(r_{t+k}, x^k_t) = 0\).

More generally, even when \(x_t\) is not stationary, under the maintained hypothesis in (22), Liu and Maynard (2007) and Phillips and Lee (2013) show that the null hypothesis (21) of the rearranged specification implies the original null hypothesis in (3), together with the maintained assumption (18). Therefore, the two formulations test the same null hypothesis, but may have different power implications under the alternative hypothesis.

### 1.6.3 Long-Horizon Tests for the Standard Specification

We are not currently aware of a valid existing method of testing (3) using the traditional long-horizon formulation in (2) under the general specifications (12) and (17) for the persistence of predictor, \(x_t\) and the choice of the horizon \(k\). Nevertheless, this specification is of particular interest since it is perhaps the mostly widely used in the empirical literature. Within the confines of the local-to-unity model, which imposes \(\eta = 1\) in (12), Valkanov (2003) and Hjalmarsson (2011) provide long-horizon predictive tests based on (2) employing local-to-unity asymptotics and Bonferroni bounds procedures.

Valkanov (2003) provides the asymptotic distribution of the standard t-test for testing the restriction (3) in (2) under a local-to-unity asymptotic specification for \(x_t\) (12, with \(\eta = 1\)) and a horizon that grows at a fixed fraction of the sample size \((v = 1\) in 17). For a given value of the local to unity parameter \(c\), this provides for a correct critical value for the standard test statistic, scaled by the square-root of the sample size. However, since the critical value depends on \(c\), which cannot be consistently estimated in a time series context, a Bonferroni procedure based on a first stage confidence interval is employed. This approach is therefore similar in spirit to the short-horizon predictive test of Cavanagh et al. (1995). Both employ a standard test statistic (possibly after rescaling) with corrected critical values.

---

14As explained in Liu and Maynard (2007), (21,22) imply \(E_{t+j-1}r_{t+j} = E_{t}r_{t+k} = kE_{0}r_{t+k} = k\beta_0(k)\), which is equivalent to (3,18). See also Phillips and Lee (2013) for an insightful explanation of the relationship between the two long-horizon specifications.
Hjalmarsson (2011) instead corrects the test statistic by providing a second order endogeneity correction to the estimator of $\beta_1(k)$ on which the test is based. This is accomplished via an augmented regression of the type first introduced by Phillips (1991): \[ r_{tk}^k = \beta_0(k) + \beta_1(k)x_t + \phi(k)\Delta_c x_t^k + \varepsilon_{1,t+k}^+ \]
\[ \Delta_c x_t^k = \sum_{j=1}^{k} (x_{t-k+j} - (1 + \frac{c}{T})x_{t-k+j-1}) = \sum_{j=1}^{k} v_{t-k+j}. \] The inclusion of the term $\Delta_c x_t^k$ in (23) removes the endogeneity bias in the estimate of $\beta_1(k)$ in (23), denoted by $\hat{\beta}_1^+(k)$. Although the distribution of $\hat{\beta}_1^+(k)$ still depends on the horizon length $k$, it does so in a simple way. As is well explained in Boudoukh et al. (2008), under the null hypothesis in (3), estimates of $\beta_1(k)$ are approximately proportional to $k$. Consequently, Hjalmarsson (2011) shows that when rescaled by the horizon length, $\hat{\beta}_1^+(k)/k$ has a mixed normal null limiting distribution whose variance no longer depends on $k$. Likewise, letting $t_k^+(c)$ denote the standard t-statistic associated with the test of $H_0: \beta_1(k) = 0$ in (23), he shows that \[ t_k^+(c)/\sqrt{k} \rightarrow_d N(0,1) \] under the null hypothesis, justifying the use of standard critical values.

In the simulations below, we will refer to $t_k^+(c)/\sqrt{k}$ as the ‘oracle’ version of Hjalmarsson (2011)’s test. It is infeasible in practice, because the augmentation $\Delta_c x_t^k$ depends on the unknown value of $c$. In practice, a Bonferroni bound based on a first-stage confidence interval for $c$ is therefore required in order to draw inference. Denoting this confidence interval, with confidence level $1 - \alpha_1$ by $\left(\underline{c}_{\alpha_1}, \bar{c}_{\alpha_1}\right)$ a feasible, yet conservative version of the test rejects if \[ t_{k,min}^+ / \sqrt{k} = \inf_{c \in \left(\underline{c}_{\alpha_1}, \bar{c}_{\alpha_1}\right)} t_k^+(c)/\sqrt{k} > z_{\alpha_2}, \] where $z_{\alpha_2}$ is the standard normal critical value associated with the significance level $\alpha_2$, such that $\alpha_1 + \alpha_2 = \alpha$, where $\alpha$ is the desired significance level of the test. In other words, a rejection occurs for the Bonferroni bounds test only if it occurs for every possible value of $c$ in the first stage confidence interval. The requirement that $\alpha_1 + \alpha_2 = \alpha$ can lead to overly conservative tests and in practice adjustments to $\alpha_1$ are typically recommended in order to keep the test from becoming too conservative [Cavanagh et al. (1995), Campbell and Yogo (2006)]. In implementing this test, we select $\alpha_1$ according to the recommendations in Hjalmarsson (2011), to which we refer the reader for further details. In the linear predictive regression context, Hjalmarsson (2012) finds that his test has better power properties than the earlier test by Valkanov (2003). Thus, we employ this version of the test of the restriction (3) in (2).

\[ \text{In (23) the estimator } \hat{\beta}_1^+(k) \text{ is equivalent to a local-to-unity version of the fully modified Phillips and Hansen (1990) estimator specialized to the context of the long-horizon predictive regression.} \]

\[ \text{Due to the inclusion of this term, we include a plus in the superscript of } \varepsilon_{1,t+k}^+ \text{ to distinguish it from } \varepsilon_{1,t+k}^k. \]
In an important recent paper, Phillips (2014) has shown that the Stock (1991) confidence interval for the local-to-unity parameter $c$ has zero asymptotic coverage when the data generating process is stationary. This can lead to either invalidity or excessive conservatism of Bonferroni procedures based on a first stage confidence interval for $c$, when the true process is mildly integrated ($\eta < 1$ in (12)) or stationary. Although Hjalmarsson (2011) employs an alternative first stage bound, based on Chen and Deo (2009), his simulations do indicate that his test becomes increasingly conservative for large negative values of $c$, particularly at long horizons.

We address this in several respects. First, for all of our simulations, we compare the power of the test across the oracle version (24), in addition to the Bonferroni version in (25). Although the oracle version of the test is not feasible, this allows us to assess whether any of the power comparisons across horizons is driven primarily by the implementation of the Bonferroni bound. More importantly, we include power comparisons for the new IVX based long-horizon test of Phillips and Lee (2013). This is the first regression based long-horizon test that we are aware of which does not require Bonferroni bounds. Since it is based not on (2), but rather on the rearranged regression (19), we discuss its implementation in the next section. Currently, we are not aware of any IVX test based on the original long-horizon specification in (2). Given the emphasis in the empirical literature on the specification in (2), we feel that it is important to include feasible tests based on this specification in our power comparisons.

### 1.6.4 Long-Horizon IVX Test for the Rearranged Specification

Phillips and Lee (2013) develop an IVX estimator and predictive test for use in the long-horizon context, employing the rearranged regression in (19). They refer to their test as the long-horizon IVX (LHIVX) test. As mentioned above, unlike other long-horizon regression based tests, the LHIVX test does not rely on Bonferroni bounds or first stage confidence intervals for $c$. Instead it provides a single test statistic with a standard normal limit distribution yielding simple inference. Consequently, its validity is not restricted to the local-to-unity model, and it is easily applied in multivariate contexts. Indeed, Phillips and Lee (2013) allow for a very general model of the persistence in $x_t$ as in (12).

IVX based tests employ an instrument that is a mildly filtered version of the original regressor $x_t$. The instrument is filtered enough so that the endogeneity term is not present in the limit distribution of the instrumental variable estimator, but not so much that it unduly reduces the instrument strength. An important theoretical foundation for its development is Phillips and Magdalinos (2007)’s generalization of the local-to-unity framework to allow for mildly integrated variables. The instrument is designed to be a mildly integrated regressor ($\eta < 1$ in (12)). This intermediate persistence of the instrument is crucial to the success of the IVX approach, since an I(1) or local-to-unity based instrument would still be subject to an endogeneity bias that depends on $c$, whereas an I(0) instrument would only be weakly correlated with an I(1) regressor.
We will focus on the case of a single regressor, since this is the case we consider in Section 1.7. In the short-horizon case, when \( k = 1 \) in (19), their IVX instrument is defined as the mildly filtered series

\[
\tilde{z}_t = \sum_{j=1}^{T} \left(1 + c_z/T^\delta_z\right)^{(t-j)} \Delta x_j
\]

where \( c_z < 0 \) and \( 0 < \delta_z < 1 \) are specified by the researcher. In the context of the long-horizon regression when \( k > 1 \) in (19), Phillips and Lee (2013) propose the long-horizon instrument

\[
\tilde{z}^k_t = \sum_{j=1}^{k} \tilde{z}_{t+j-1},
\]

which is partially summed in the same way as the regressor in (20). Under an additional rate condition that guides the choice of \( c_z \) and \( \delta_z \), Phillips and Lee (2013) show that their estimator is asymptotically mixed normal, resulting in standard normal or Chi-squared inference over the wide range of persistence levels for \( x_t \) in (12), including mildly integrated, nearly integrated, integrated, mildly explosive and locally explosive models.

### 1.7 Power of Long-Horizon Regression in Linear and Regime Switching Predictive Regression Models

We provide a simulation based comparison of the power of short and long-horizon linear regression predictive tests when the true model is given by the regime switching model in (5) with persistent endogenous regressors as in (11,12,13). As a point of comparison, we also provide results when the data is generated from the short-horizon linear predictive regression (7).

The regime switching model has a good number of parameters associated with it. In order to ensure that we simulate from a realistic parameterization of the model, we use the estimates in Tables 3 to set the default values of the parameters in (5). Similarly, we use the estimates in Table 4 to simulate the model for the time varying transition probabilities in (8). We focus on the two regime \((N = 2)\) model, which appears to have the best fit, but also compare it to the linear predictive regression \((N = 1)\).\(^{17}\) In the model for \( x_t \), the parameters in (12) are not estimable and there is a genuine uncertainty regarding the persistence of valuation predictors such as the dividend price ratio. We consider two specifications of (12): \((c = -2.5, \eta = 1)\) and \((c = -20, \eta = 1)\). The first represents a case in which the series is almost as persistent as a unit root, whereas the second represents the case of a much less persistent series. The innovations for return and predictor series are drawn from (13) using a multivariate normal distribution with correlation \( \delta = -0.95 \), which is typical of valuation predictors, such as the dividend price ratio.\(^{18}\)

\(^{17}\)We have also re-run many of our results using a three regime version of our model, based on earlier estimates, obtaining qualitatively similar results.

\(^{18}\)We estimate \( \delta = -0.9424 \) in the linear predictive regression using our data.
We consider $T = 200$ as a fairly small sample and $T = 1048$, the sample size in our data set described in Section 1.4. All simulation results are based on two thousand replications.\footnote{The figure differs slightly from those of the previously published version due to a minor data update. However, the differences remain trivial.}

We restrict certain parameters under the null hypothesis. In the linear predictive regression it is typical to test the null hypothesis that $\beta_1 = 0$ in (7) under the maintained assumption that $E_t \varepsilon_{t+1} = 0$, which is equivalent to the non-predictability condition $E_t r_{t+1} = \beta_0.\footnote{Since, following much of the literature, we do not impose \( \beta_0 = 0 \), this could be more precisely referred to as the null hypothesis of non-time-varying-predictability. The non-zero intercept allows for a non-time varying risk premium and accommodates the positive average returns that have been recorded over long historical periods.}$ We enforce the same non-predictability condition when imposing the null hypothesis in the regime switching model. This implies both zero slope coefficients in all states and intercepts that do not vary across states, i.e.

$$H_0: \begin{pmatrix} \beta_0(s_t) \\ \beta_1(s_t) \end{pmatrix} = \begin{pmatrix} \beta_{0,0} \\ 0 \end{pmatrix} \text{ for all states } s_t.$$ \hspace{1cm} (28)

For $\beta_{0,0}$, the null value of $\beta_0$, we take a weighted average of our empirical estimates of $\beta_0(s_t)$ across the $N$ states, where the weights are given by the unconditional state probabilities implied by the transition probabilities in (8) evaluated at the mean value of the predictor $\bar{x}$.

In order to provide a power curve, rather than simply power at a single point in the alternative, we allow the parameters that are restricted under the null hypothesis to drift back towards their default (estimated) values under the alternative hypothesis, i.e.

$$H_A: \begin{pmatrix} \beta_0(s_t) \\ \beta_1(s_t) \end{pmatrix} = \begin{pmatrix} \beta_{0,0} \\ 0 \end{pmatrix} + \frac{\gamma}{T(1+v_2)^2/2} \begin{pmatrix} \beta_{0}^*(s_t) - \beta_{0,0} \\ \beta_{1}^*(s_t) \end{pmatrix},$$ \hspace{1cm} (29)

where $\beta_{0}^*(s_t)$ and $\beta_{1}^*(s_t)$ are the default values of the parameters taken from the empirical estimates in each state shown in Table 4. The value of $\gamma$ determines the distance from the null hypothesis and is allowed to vary across the horizontal axis. The value of $v_2$ is specified below for each test.

Although the data is generated from the regime switching predictive regression model, we do not assume that the empirical researcher knows that this is the true model. Rather, our interest lies in comparing the power of the most common short and long-horizon linear predictive regressions when the data is generated by a plausible nonlinear model, such as the regime switching predictive regression model. We compare the power of the long and short-horizon regressions using the two well known long-horizon specifications, the standard long-horizon regression (2) and the rearranged long-horizon regression (19), discussed in Section 1.6. We consider only the empirically relevant positive, one-sided alternatives $H_A : \beta_1(k) > 0$ or $H_A : \beta_1^*(k) > 0$ respectively to the null restrictions (3) and (21). To keep our comparisons easily viewable, in each
figure we compare the power curve for the short-horizon specification $k = 1$ to two long-horizon specifications, where $k$ is chosen according to (17). The choice of $\lambda$ and $v$ are discussed separately below for each test.

In the case of the standard long-horizon regression (2), we simulate power comparisons across three different tests:

- Although the OLS based test is invalid in our framework and highly sized distorted, it has nonetheless been influential in the literature. It is therefore of some interest to know whether longer horizon versions of the test lead to any power improvements or whether they merely aggravate size distortion. Thus we compare size-adjusted power of the OLS based test between long and short-horizon tests.

- As an alternative test that has conservative size within the confines of the local-to-model, we consider the feasible, Bonferroni version of the rescaled, endogeneity corrected tests of Hjalmarsson (2011) using the test statistic $t_{k,\min}^+/\sqrt{k}$ in (25). Since this test is not oversized, we compare power rather than size-adjusted power in this case.\(^{21}\)

- Since Phillips (2014) has pointed out some important shortcomings with many of the existing Bonferroni based approaches to predictive testing and Hjalmarsson (2011) reports that when $c$ is large negative, the Bonferroni version of his test becomes more conservative at larger horizons, we also provide results for the oracle version of the same test in (24). Although the oracle version is of course infeasible, it helps us to ensure that none of the comparison are solely driven by the influence of the Bonferroni bounds in the feasible version of the test.

In the case of the rearranged regression, we compare two additional tests.

- We compare size-adjusted power in an OLS based test of the rearranged regression.

- More interestingly, we also compare the power (without size-adjustment) of the IVX and LHIVX tests of Phillips and Lee (2013) discussed in section 1.6.

Since this results in a great many power curves, we include only a few of the curves to illustrate our main findings below. Additional results, that generally confirm the main conclusions arrived at below, are available from the authors upon request. These include results using both the smaller sample size of $T = 200$ and the lower persistence parameter of $c = -20$.

---

\(^{21}\)In principle, it is not clear whether any of the tests we consider here are correctly sized when the residual variance $\sigma_1^2(n_t)$ varies across states under the null hypothesis. However, this does not appear to be an important practical concern in any of our simulations.
1.7.1 Power Comparisons Based on the Standard Long-Horizon Specification

We first compare the size-adjusted power of short and long-horizon tests based on (2) for the OLS based test. We also compare power for both the Bonferroni and oracle versions of Hjalmarsson (2011)’s test discussed in Section 1.6.3. Since these tests have non-trivial power against $O_p\left(T^{-1}\right)$ alternatives, we set $v_2 = 1$ in (29). Similarly, we set $v = 1$ in (17) when selecting $k$ and take $\lambda = 0.05$ and $0.010$, resulting in values of $k = 10$ and $k = 20$ when $T = 200$ and $k = 52$ and $k = 104$ when $T = 1048$. In all cases we compare to $k = 1$.

1.7.1.1 True Model is a Linear Predictive Regression

Figure 2 (using $T = 1048$) provides the size adjusted power curves of standard t test when $k = 1$ and $k = 52$ and $k = 105$ when the true model is the linear predictive model in (7). The three curves are generally quite similar. A small power advantage to the longer horizon tests is detectable towards the middle of the power curve ($50 < \gamma < 150$). The OLS based test does not make efficient use of the endogeneity between the regression error and regressor innovation, which may explain why there is room for some slight power advantages to the long-horizon specification. We next turn to the oracle version of Hjalmarsson (2011)’s test in Figure 3, which makes optimal use of this endogeneity within the confines of the linear predictive model with local-to-unity predictor. With the use of this more powerful test, the slight advantage of the long-horizon test disappears. The short-horizon is more powerful throughout the power curve. The dominance of the short-horizon test remains even after the test is made feasible using its Bonferroni implementation in Figure 4. This confirms that when the true model is linear and the short run regression is therefore correctly specified, the long-horizon tests should not be able to provide power improvements to efficient versions of the short-horizon test.

1.7.1.2 True Model is a Regime Switching Predictive Regression

We now consider the same comparisons in the regime switching predictive regression model, in which the regressor can be predictive for both the transition probabilities of switching between regimes and the return outcomes within regimes. In this case, both the short and long-horizon models are inherently misspecified and act only as linear approximations to the true nonlinear model. In principle, there is no reason to argue that the long-horizon model couldn’t provide an improved approximation and better power than the short-horizon model. Indeed, it would seem surprising if there were not some nonlinear models for which this is the case. The practical question that we now address is whether, using simulations based on empirical estimates of the regime switching model, we obtain power improvements at longer horizons. Put another way, if the regime switching model were the true model, would we be more likely to detect the resulting return predictability at longer horizons?
Our results suggest that this is not the case. Long-horizon tests fare no better (and sometimes worse) when the data is generated from nonlinear regime shift models, capturing bear/bull markets. We find this result even when allowing the regressor to be predictive of the regime shift itself. The OLS test, shown in Figure 5 (using $T = 1048$), again shows modest size-adjusted power advantages at long-horizons close to the null hypothesis, whereas the short-horizon test has the best power for distant alternatives. However, the short-horizon test continues to show the best power over the entire alternative in both the oracle and Bonferroni versions of Hjalmarsson (2011)'s test, shown in Figures 6 and 7. Thus, even in this nonlinear model, the long-horizon tests do not provide a power advantage over their short-horizon counterparts.

### 1.7.2 Power Comparisons Based on Rearranged Long-Horizon Specification

We now compare power across horizons in the rearranged long-horizon empirical specification in (19), with our primary focus on the IVX and LHIVX estimators discussed in Section 1.6. For the short-horizon IVX, $k = 1$. For the LHIVX, Phillips and Lee (2013) consider return horizons that grow more slowly than sample size ($v < 1$ in (17)). We fix $\lambda = 1$ and consider both $v = 0.80$ and $v = 0.90$, resulting in values of $k = 69$ and $k = 117$ when $T = 200$ and $k = 260$ and $k = 522$ when $T = 1048$. Phillips and Lee (2013) show that their test is divergent against alternatives of the rate $T^{-1/2}k^{-1/2} = T^{-(1+v)/2}$. Therefore, we set $v = 0.80$ in the specification of the local alternative in (29), where $\gamma$ again varies across the horizontal axis. Following both the recommendations of Phillips and Lee (2012) and Phillips and Lee (2013) we set $c_z = -5.0$. The selection of $\delta_z$ in (26) requires slightly more attention. For the short-horizon IVX, Phillips and Lee (2012) recommend a value of $\delta_z$ between 0.75 and 0.95. For the long-horizon regression, rate restrictions require $\delta_z < v$ in (17) and Phillips and Lee (2013) suggest $\delta_z = v - 0.05$. Therefore, we employ $\delta_z = 0.75$ when $v = 0.80$ and $\delta_z = 0.85$ when $v = 0.90$. We compare these to two versions of the short-horizon IVX: $k = 1$ with $\delta_z = 0.75$ and $k = 1$ with $\delta_z = 0.85$. With these choices, we use recommended values of $\delta_z$ in all cases and we can compare the short and long-horizon regressions holding $\delta_z$ fixed.

The LHIVX is a newly introduced test and we are not aware of any finite sample power results for the test even in the linear model, so it is interesting to compare its performance across horizons and choices of $\delta_z$. Figure 8 provides this comparison in the linear predictive model for $T = 1048$ and $c = -2.5$. Of the two long-horizon tests, it is interesting to note that using $v = 0.80$ ($k = 261$) with $\delta_z = 0.75$ has both better size and power than the test employing $v = 0.90$ ($k = 525$). This result was not expected, but it may be that $v = 0.90$ comes too close to the restriction that $v < 1$. In fact, the LHIVX test with $v = 0.80$ and $\delta_z = 0.75$ has very good local power, dominating the short-horizon IVX test with the same value of $\delta_z$ for $\gamma < 200$. However, when using a larger value of $\delta_z = 0.85$, the short-horizon test does as well as the long-horizon test near the null hypothesis and provides better power further into the alternative.
These results are interesting in that they highlight the important interplay between the choice of $\delta_z$ and the horizon. Yet, overall they still suggest that with a good choice of $\delta_z$ the short-horizon IVX can provide as much or more power than its long-horizon counterpart.

In Figure 9, we next investigate the power of the LHIVX test when the data is generated by the regime switching model. The general conclusions do not change. The LHIVX test with $v = 0.80$ ($k = 250$) and the short-horizon IVX test with $\delta_z = 0.85$ are equally good and improve on the other two tests near the null hypothesis, while the short-horizon test with $\delta_z = 0.85$ appears to be the best further along the power curve.

Overall, when the short-horizon test is efficient or based on a good selection of $\delta_z$, our simulations show little evidence of power improvements from longer horizon tests in either the linear predictive regression or the regime switching predictive regression model.

1.8 Conclusion

There has been much debate in the literature regarding the presumed power advantages of long-horizon regressions. With a few notable exceptions, most of this debate has taken place within the context of a linear predictive model as the true data generating process. In this case, long-horizon regression tests may have power advantages over short-horizon tests that do not fully exploit the information in the model, especially the contemporaneous endogeneity between the regressor innovation and the regression residual. On the other hand, it seems unlikely that they would have advantages over more efficient short-horizon tests when a linear short-horizon model is the true model.

The class of nonlinear models is so large that it would not be surprising if one could find nonlinear models which are better approximated by long-horizon than short-horizon regression models. However, this would only seem to provide practical support for the long-horizon approach if both the nonlinear model that supports it and its parameter values are realistic. This argues in favor of a comparison based on simulations from an estimated nonlinear model that has already received support in the empirical literature.

In this paper we compare the power of short and long-horizon tests in both the linear model and an empirically plausible nonlinear model, involving regime switching predictive regression with time varying transition probabilities that are also a function of the predictor. Previous empirical literature has shown this model to be successful in capturing the bull and bear states long observed by market participants. We use our empirical estimates to anchor our simulations in a realistic manner.

We then compare the power of short and long-horizon tests when the true model is given by the estimated regime switching predictive regression model. As a point of reference, we provide similar comparisons when the data is generated by a linear regression model. We make these
comparisons using two different versions of the long-horizon regressions: the standard long-horizon regression in (2), which has a long-horizon dependent variable and the rearranged long-horizon regression in (19), which uses a long-horizon predictor. Accordingly, we also consider two different recent long-horizon tests, the Hjalmarsson (2011) test and the Phillips and Lee (2013) long-horizon IVX (LHIVX) test. In addition, we compare size adjusted power for the influential, though size-distorted, OLS based tests used in much of the earlier empirical work. We also make comparisons using an infeasible, oracle version of the Hjalmarsson (2011) test.

Taken as a whole, our results are not overly supportive of the long-horizon regressions. In the case of the standard long-horizon specification, we find that that long-horizon regressions can provide at best a modest improvement in size-adjusted power using OLS based tests. However, in more powerful versions of the tests, such as the oracle version of the Hjalmarsson (2011) test, or even its Bonferroni counterpart, the short-horizon test shows better power in both the linear and regime switching predictive tests.

In the context of the rearranged regression, the power of the LHIVX test depends on both the horizon and the parameter $\delta_z$, which determines the extent of the mild filtering used in the creation of the instrument. The theoretical results provide more flexibility in selecting this parameter at short horizons than in long horizons. For some choices of $\delta_z$ the long-horizon IVX test can provide more power close to the null hypothesis than its short-horizon counterpart. However, with a better choice of $\delta_z$, we find that the short-horizon IVX is generally more powerful than the LHIVX, even close to the null hypothesis. The advantage to the short-horizon test increases as we move further from the null hypothesis.

Overall, our results support the contention that the higher power of long-horizon regressions may simply be a “myth” as argued by Boudoukh et al. (2008). On the other hand, regime switching models are not the only plausibly nonlinear models that could be considered in making these comparisons. Contrary to our results, improved power has been found in long-horizon regressions when using a nonlinear ESTAR model as the data generating process [Kilian and Taylor (2003), Wohar and Rapach (2005)]. Models incorporating financial bubbles [Evans (1991), Phillips et al. (2014)] might also support the conjecture of improved power at longer horizons and this could be an interesting avenue for future research. More general theoretical results on the relative power of long horizon tests in nonlinear models might be possible in the general framework of Park and Phillips (1999).
Table 1: Estimated Model Coefficients in Regime Switching Models without Predictors

<table>
<thead>
<tr>
<th>$s_t$</th>
<th>Single State Model</th>
<th>Two State Model</th>
<th>Three State Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N/A</td>
<td>Bull</td>
<td>Bear</td>
</tr>
<tr>
<td>$\beta_0(s_t)$</td>
<td>0.0398</td>
<td>0.0437</td>
<td>0.0219</td>
</tr>
<tr>
<td></td>
<td>(0.0014)</td>
<td>(0.0011)</td>
<td>(0.0067)</td>
</tr>
<tr>
<td>$\hat{\sigma}^2_1(s_t)$</td>
<td>0.0022</td>
<td>0.0010</td>
<td>0.0071</td>
</tr>
<tr>
<td></td>
<td>(0.0463)</td>
<td>(0.0001)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>H-Q IC</td>
<td>-3.2961</td>
<td>-3.5706</td>
<td>-3.6743</td>
</tr>
</tbody>
</table>

The table shows the estimated expected return $\hat{\beta}_0(s_t)$ and variance $\hat{\sigma}^2_1(s_t)$ for each possible value of the state $s_t$. In other words, it shows the estimated coefficients in a restricted version (5), in which $\beta_1(s_t) = 0$. Standard errors are given in parenthesis. The bottom line provides the value of the Hannan and Quinn (1979) information criteria (H-Q IC), with a lower value indicating a better penalized fit.

Table 2: Transition Probabilities in Regime Switching Models without Predictors

<table>
<thead>
<tr>
<th>$s_t/s_{t-1}$</th>
<th>Two State Model</th>
<th>Three State Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bull</td>
<td>Bull</td>
<td>Bull</td>
</tr>
<tr>
<td></td>
<td>Bear</td>
<td>Bear</td>
</tr>
<tr>
<td>Bull</td>
<td>0.97</td>
<td>0.94</td>
</tr>
<tr>
<td>(0.03)</td>
<td>(0.05)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Normal</td>
<td>0.04</td>
<td>0.96</td>
</tr>
<tr>
<td>(0.02)</td>
<td>(0.05)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Bear</td>
<td>0.03</td>
<td>0.88</td>
</tr>
<tr>
<td>(0.01)</td>
<td>(0.07)</td>
<td>(0.01)</td>
</tr>
</tbody>
</table>

The column headings denote the state in the previous period ($s_{t-1}$), while row headings denote the state in the current period ($s_t$). The entries give the estimated probability of switching to the state denoted by the row headings in time $t$ conditional on being in the state denoted by the column heading in time $t - 1$. Standard errors are given in parenthesis.

Table 3: Coefficient Estimates in Regime Switching Predictive Regression Model with Time Varying Transition Probabilities

<table>
<thead>
<tr>
<th>$s_t =$:</th>
<th>Single State Model</th>
<th>Two State Model</th>
<th>Three State Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
<td>Bull</td>
<td>Bear</td>
<td>Bull</td>
</tr>
<tr>
<td>$\beta_0(s_t)$</td>
<td>0.0398</td>
<td>-0.0052</td>
<td>0.0447</td>
</tr>
<tr>
<td>(0.0013)</td>
<td>(0.0121)</td>
<td>(0.0011)</td>
<td>(0.0125)</td>
</tr>
<tr>
<td>$\beta_1(s_t)$</td>
<td>0.0160</td>
<td>0.0367</td>
<td>0.0158</td>
</tr>
<tr>
<td>(0.0013)</td>
<td>(0.0086)</td>
<td>(0.0011)</td>
<td>(0.0034)</td>
</tr>
<tr>
<td>$\hat{\sigma}^2_1(s_t)$</td>
<td>0.0010</td>
<td>0.0008</td>
<td>0.0009</td>
</tr>
<tr>
<td>(0.0436)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>H-Q IC</td>
<td>-3.4184</td>
<td>-3.7617</td>
<td>-3.6861</td>
</tr>
</tbody>
</table>

The table entries show the estimates of the coefficients in the regime switching predictive regression model (5) for each possible value of the state $s_t$. Standard errors are given in parenthesis. The bottom line provides the value of the Hannan and Quinn (1979) information criteria (H-Q IC), with a lower value indicating a better penalized fit. Note that the inclusion of the dividend price ratio as a predictor may induce a predictive regression problem, a solution for which has yet to be provided in this regime switching model. Therefore, standard errors can only be relied on as a rough indicator of the estimation variance, but should not be employed to draw formal inferences.

23
Table 4: Time Varying Transition Probability Model Coefficients in Regime Switching Predictive Regression Model

<table>
<thead>
<tr>
<th></th>
<th>Two State Model</th>
<th>Three State Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_t/s_{t-1}$</td>
<td>$\theta_{0.11}$</td>
<td>$\theta_{0.21}$</td>
</tr>
<tr>
<td>Bull (i=1) Bear (i=2)</td>
<td>1.9289</td>
<td>-0.5987</td>
</tr>
<tr>
<td></td>
<td>(0.1515)</td>
<td>(0.2912)</td>
</tr>
<tr>
<td>Normal (i=2) Bear (i=3)</td>
<td>-1.9999</td>
<td>-2.0018</td>
</tr>
<tr>
<td></td>
<td>(3.3623)</td>
<td>(0.8731)</td>
</tr>
</tbody>
</table>

Slope Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Two State Model</th>
<th>Three State Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_t/s_{t-1}$</td>
<td>$\hat{\theta}_{1.11}$</td>
<td>$\hat{\theta}_{1.21}$</td>
</tr>
<tr>
<td>Bull (i=1) Bear (i=2)</td>
<td>0.1770</td>
<td>-0.7124</td>
</tr>
<tr>
<td></td>
<td>(0.1346)</td>
<td>(0.2501)</td>
</tr>
<tr>
<td>Normal (i=2)</td>
<td>-0.0001</td>
<td>-0.0019</td>
</tr>
<tr>
<td></td>
<td>(2.3320)</td>
<td>(0.5638)</td>
</tr>
</tbody>
</table>

The table entries show the estimates of the coefficients in (8), the model for the time transition probabilities in the regime switching predictive regression model (5). Standard errors are given in parenthesis. Note that the inclusion of the dividend price ratio as a predictor may induce a predictive regression problem, a solution for which has yet to be provided in this regime switching model. Therefore, standard errors can only be relied on as a rough indicator of the estimation variance, but should not be employed to draw formal inferences.

Figure 1: Fitted Bear State Probabilities in Two State Regime Switching Predictive Regression Model with Time Varying Transition Probabilities.

Figure 2: The size-adjusted power of the OLS based predictive test when the return is generated from the linear predictive regression model in (7), the predictor is generated by the local-to-unity process in (11) and (12) with $c = -2.5$, and the residuals are generated from (13) using a multivariate normal distribution with $\delta = -0.95$. We set $T = 1048$. 
Figure 3: The power of the oracle based predictive test when the return is generated from the linear predictive regression model in (7), the predictor is generated by the local-to-unity process in (11) and (12) with $c = -2.5$, and the residuals are generated from (13) using a multivariate normal distribution with $\delta = -0.95$. We set $T = 1048$. 
Figure 4: The power of the Bonferroni based predictive test when the return is generated from the linear predictive regression model in (7), the predictor is generated by the local-to-unity process in (11) and (12) with $c = -2.5$, and the residuals are generated from (13) using a multivariate normal distribution with $\delta = -0.95$. We set $T = 1048$. 
Figure 5: The size-adjusted power of the OLS based predictive test when the true return model is the two-state regime switching model in (5) with time varying transition probabilities given by (8), the predictor is generated by the local-to-unity process in (11) and (12) with $c = -2.5$, and the residuals are generated from (13) using a multivariate normal distribution with $\delta = -0.95$. We set $T = 1048$. 
Figure 6: The power of the oracle based predictive test when the true return model is the two-state regime switching model in (5) with time varying transition probabilities given by (8), the predictor is generated by the local-to-unity process in (11) and (12) with $c = -2.5$, and the residuals are generated from (13) using a multivariate normal distribution with $\delta = -0.95$. We set $T = 1048$. 

Power across return horizons $H_{\gamma} | \beta_1 > 0$ with $\alpha = 0.05$

\[
\gamma_{k-1/2} \rightarrow t + k, c_{ORAC}, k = 1
\]

\[
\gamma_{k-1/2} \rightarrow t + k, c_{ORAC}, k = 52
\]

\[
\gamma_{k-1/2} \rightarrow t + k, c_{ORAC}, k = 104
\]
Figure 7: The power of the Bonferroni based predictive test when the true return model is the two-state regime switching model in (5) with time varying transition probabilities given by (8), the predictor is generated by the local-to-unity process in (11) and (12) with $c = -2.5$, and the residuals are generated from (13) using a multivariate normal distribution with $\delta = -0.95$. We set $T = 1048$. 
Figure 8: The power of the LHIVX test in a linear predictive regression model. We plot power for both $k = 1$ and for $k$ given by (17) with $\lambda = 1$ and $v = 0.75$ and $v = 0.90$. The horizontal axis shows the value of $\gamma$ in (29) with $v_2 = 0.80$. Here we set $c = -2.5$, $c_z = -5.0$, and $T = 1048$. 
Figure 9: The power of the LHIVX test in the two-state regime switching model with time varying transition probabilities. We plot power for both $k = 1$ and for $k$ given by (17) with $\lambda = 1$ and $v = 0.75$ and $v = 0.90$. The horizontal axis shows the value of $\gamma$ in (29) with $v_2 = 0.80$. Here we set $c = -2.5$, $c_z = -5.0$, and $T = 1048$.  
Chapter 2

The Finite Sample Power of Long-Horizon Predictive Tests in Models with Financial Bubbles

2.1 Introduction

In popular policy discussions, financial bubbles have become closely associated with the term “irrational exuberance” coined in 1996 by former Federal Reserve Chairman Alan Greenspan two days after a meeting in which Professors John Campbell and Robert Shiller famously warned him of the high stock market valuations. Their warning was supported by a regression scatter plot showing a negative relationship between ten year market returns and the long-term price equity ratio. Based on this negative long-horizon predictive relationship, they cautioned that historical periods of high stock valuations had often been followed by below average returns over the next decade - a warning that Carroll (2008) finds to be prescient.

Thus, in the policy context, financial bubbles and valuation based long-horizon predictive regression have been tightly linked. Yet, the academic literature on these two topics remains primarily distinct. As outlined in the following section, one large literature has developed to debate and test for the existence of, model, and date financial bubbles. A second, essentially distinct, literature debates the size and power properties of long-horizon tests and proposes new test procedures.

Our paper is an attempt to bridge these two literatures. We examine the ability of long-horizon predictive tests to detect the predictability inherent in a present value model with asset bubbles. In particular, we assess the statistic power of an appropriately size-corrected test based on a long-horizon predictive regression specification using a valuation based predictor, when return predictability is due to the presence of periodic financial bubbles. We also compare the predictive power of the long-horizon test to that of its short-horizon counterpart, in which returns are predicted just one period ahead.

Several primary findings emerge from our analysis. First, we confirm that valuation based linear predictive regression models have the power to detect return predictability due to more complicated models involving financial bubbles. Provided that properly sized tests are employed, this provides support to the consideration of market valuations in assessing the possibility of a financial bubble. Secondly, even after size correction, we do find some power improvements to predictive tests when conducted at longer horizons in the presence of financial bubbles. This offers some limited support to the notion that financial bubbles can explain or justify higher power or stronger predictability at longer horizons. Finally, however, we note these advantages are relatively modest at best even in this setting in which intuition might suggest that long run predictive tests would prove more powerful than their short-run counterpart. For example,
when the predictor is strongly autocorrelated with a root close to unity, our simulations suggest that after size correction a predictive regression with a ten year horizon is at best two percent more likely to detect bubble driven return predictability than a predictive regression with a one-year horizon. With lower, yet still plausible values of this autoregressive root, the power improvement at the longer horizons increases to as much as six percent, a non-trivial, yet still modest improvement.\footnote{In addition to the ten year return horizon, the analysis of Campbell and Shiller’s referred to above used ten years of smoothed earnings to calculate their valuation ratio. This latter smoothing arguably has important benefits that are not analyzed here, where we focus on the length of the return horizon and employ the lagged dividend price ratio as the valuation based predictor.}

These results also contribute to the long-standing debate over the power of long-horizon predictive regressions. Much of this debate has taken place in a linear predictive model, which is less simple than it appears due to the persistence of the lagged valuation predictors and their correlation with contemporaneous returns. However, an argument can be made that with the use of a powerful short-run test, there can be no inherent advantage to long-horizon regressions when the linear short-horizon model is a correct specification (Hjalmarsson 2011, Maynard and Ren 2014). This paper joins a smaller literature which has considered the debate in the context of non-linear models, in which the short-horizon linear regression is misspecified and there may therefore be stronger a priori reason to suspect improvements at longer horizons (Kilian and Taylor 2003, Wohar and Rapach 2005, Ang and Bekaert 2007, Maynard and Ren 2014).

The long ten-year return horizon was a seemingly critical component of Campbell and Shiller’s analysis. When an asset bubble is present it may continue to expand for many periods before eventually bursting or correcting. Thus they were not arguing that a correction would be immediate, but rather that it was likely to occur over the longer-horizon. The implicit intuitive argument is that valuation based predictors will have greater power to detect bubble driven return predictability at longer horizons. Yet, even in this non-linear setting in which intuition might appear to strongly favour the long-horizon specification, we find only modest advantages to it.

The remainder of the paper is organized as follows. A discussion of the literatures on financial bubbles and long-horizon return predictability is given in the next section. Section 3 presents and derives the present value model with the financial bubble component. Section 4 describes the simulation framework and parameter settings. Section 5 presents and interprets our main results and findings. Section 6 concludes. Tables and figures are included in the appendix.

## 2.2 Literature, Background, and Motivation

What are asset bubbles? Do they exist and is it appropriate to integrate them into asset pricing models? These issues present challenges to both asset pricing and econometric theory. In order
to answer these questions, many scholars provide different models and explanations for financial bubbles. There is no consensus regarding the uniform definition of bubbles in the literature. However, historical experiences, including the dot.com bubble and 2008 sub-prime crisis, appears to many observers to confirm the existence of asset bubbles.

The existence and econometric test of asset bubbles constitute a much debated subject of research in financial economics. Much of the theoretical literature focuses on the question of whether bubbles in asset prices exist. The argument for the absence of bubbles in asset prices builds on the assumption of rational investors as stated by the efficient markets hypothesis of Fama (1965) who posits that the presence of informed traders will correct any overpricing or underpricing errors created by speculators. The justification for the non-existence of bubbles is also provided by neoclassical theorists who claim that bubbles can be precluded if there is a maximum possible price, backward inductions and transversality conditions, as in Santos and Woodford (1997a). The views against bubbles are enforced by the use of nonlinear utility function and also the general equilibrium framework. However, Abreu and Brunnermeier (2003) challenges the traditional view by proposing that the persistence of bubbles is supported by the failure of arbitrageurs to synchronize or coordinate their trading strategies. Speculative bubble models have also been developed in which heterogeneous beliefs of agents generate overvaluation of asset prices and thus create speculative bubbles in both static and dynamic frameworks. Miller (1977), Chen et al. (2002), Scheinkman and Wei (2003) and others discuss speculative bubbles in detail.

A second question addressed in the literature is how to identify and test for bubbles. This question has generated various econometric approaches. An overview of past works on econometric tests for bubbles by Gurkaynak (2008) includes variance bounds tests by Shiller (1981), two-step tests by West (1987), cointegration based tests by Diba and Grossman (1988), and tests of intrinsic bubbles by Froot and Obstfeld (1991) and tests of unobserved bubbles by Wu (1997). However,Gurkaynak (2008) notes that many of these approaches are not robust across models. The most recent advances in the specification and test of bubbles have been made by Phillips et al. (2015a) and Phillips et al. (2014). Phillips et al. (2015b) propose an augmented Dickey Fuller type test for the explosive autoregressive behavior associated with bubbles. Phillips et al. (2015a) provide a recursive algorithm to identify and date bubble episodes.

Despite numerous previous studies devoted to the research on existence and econometric test of asset bubbles, there is not much academic research done to build a link between bubbles and the power of long-run predictability of stock returns. Researchers disagree over how or whether to integrate the asset bubble component into asset pricing models (see the discussion of Cochrane (2005)) and what the implications are for empirical research if asset bubbles are incorporated into the asset pricing models. These different opinions are expressed in Pstor and Veronesi (2006),Cooper (2008). It appears that there is a disconnect between the asset bubble
and asset return literatures as a result of these controversial debates.

Recently, Santos and Woodford (1997b) address this gap in the literature by incorporating a financial bubble into a consumption based asset pricing model. They derive implications for the risk premium and employ a GMM based approach to estimate the model parameters. Working independently, we also incorporate a financial bubble component into a present value model, but use this to address a different question. Specifically, what is the relation between bubbles and the statistical power of long run predictive regression? Can price bubbles help to improve the statistical power of long horizon predictability of stock returns? These questions, which are not answered in previous studies, motivate our paper.

The finance literature generally models a bubble as an increasing over-valuation of an asset relative to its fundamental evaluation that becomes self-fulfilling because investors bid the price of the asset up on the expectation that future investors will bid it up even further. Previous work also indicates that the dividends and earning price ratio should provide indicators of bubbles. They spike during bubbles and then subside. The goal of our paper is to explore whether there is a connection between the presence of bubbles and the power of long-horizon predictive regression.

The statistical power of long-horizon predictability tests remains a topic of debate in the empirical literature on predictive regression of stock returns. As Boudoukh et al. (2008) note, the strong empirical results and large R-squared values from long-horizon predictive results have been hugely influential, cited by Cochrane (1999, “New Facts in Finance”) as one of the three most important new facts in finance. In fact, the 2013 Nobel Press release Royal Swedish Academy of Sciences (2013) made specific reference to the (partial) forecastability of stock and bond returns at long-horizons. This follows highly influential empirical applications to long-horizon predictability for stock returns (Fama and French 1988, Campbell and Shiller 1988), bonds yields (Fama and Bliss 1987, Cutler et al. 1991), and exchange rates (Mark 1995, Chinn and Meese 1995).

These strong empirical results were originally thought to be due to higher statistical power obtained at longer horizons. However, this conclusion became less clear-cut as size distortions in long-horizon regressions became apparent. These were first noted in simulation studies (Kim and Nelson (1993), and Goetzmann and Jorion (1993)) and later formalized asymptotically (Valkanov (2003a) and Hjalmarsson (2012)). Boudoukh et al. (2008) demonstrate that the empirical evidence of predictability appears to strengthen as the horizon increases even under

---

2The main derivation in the next section was derived before coming across a newly posted working paper version of Santos and Woodford (1997b) in early 2015. We also did not share our results with them ahead of their work. Thus both sets of results were arrived at independently.

3Similar, though less severe size distortion, has been found in short-horizon regression when the lagged predictors are persistent and contemporaneously correlated with return innovations (Mankiw and Shapiro 1986, Cavanagh et al. 1995, Stambaugh 1999). This has led to a large financial econometric literature too vast to review here, where our focus is on longer horizon prediction.
the null hypothesis that returns are unpredictable.4

This has led to a debate over the power of long-horizon regressions: is there really higher power at long-horizons or are the strong and influential empirical results simply a result of size distortion – a question first asked explicitly by Campbell (2001). Boudoukh et al. (2008) refer to the claim of higher statistical power at longer horizons as a “myth” and several other papers report evidence against it (Ang and Bekaert 2007, Boudoukh et al. 2008, Hjalmarsson 2008, Hjalmarsson 2012, Maynard and Ren 2014). Others works provide somewhat more favorable evidence (Campbell 2001, Kilian and Taylor 2003, Wohar and Rapach 2005, Mark and Sul 2006, Cochrane 2008), and Cochrane (2008) argues strongly in defense of the proposition of stronger predictability at longer horizons.

Maynard and Ren (2014) note that much of this debate takes place within the context of linear models, in which the short-horizon regression is correctly specified. As they note, due to the persistence of the valuation predictors and their contemporaneously correlated residuals standard short-horizon tests may suffer from power loss even if size corrected (Campbell and Yogo 2006), opening the possibility of power gains at longer horizons (Cochrane 2008). However, once more powerful short-horizon tests are employed there is no obvious reason to expect increased power at longer horizons in linear models for which the short-horizon predictive regression is a correct specification (Hjalmarsson 2011, Maynard and Ren 2014).

More interestingly, in most models with non-linear dynamics or relations, the linear predictive regression is simply a linear approximation to the true underlying model at all horizons. If the longer-horizon regression provides an improved approximation, power improvements at longer horizons may exist even after controlling size. Kilian (1999) first suggests that power at longer horizons may be due to the presence of non-linearities. Kilian and Taylor (2003) and Wohar and Rapach (2005) study the power of long-horizon predictive regressions in the ESTAR model, (Ang and Bekaert 2007) considers the non-linear present value model, while (Maynard and Ren 2014) address the question in a model of regime switching.

By analysing the power of long-horizon regressions in present value models with non-linear price bubble dynamics, we contribute to this small but growing literature. In the empirical asset pricing literature, a common starting point characterizing the present value relation is

\[ dp_t = \lim_{i \to \infty} \rho^i dp_{t+i} + \sum_{i=0}^{\infty} \rho^i (r_{t+1+i} - \Delta d_{t+1+i}) + \frac{k}{1-\rho}, \]

where \( dp_t, r_t, \Delta d_t \) are dividend price ratio, stock returns and dividend growth ratio, \( \rho = \frac{1}{1+exp(d-p)} \) is the discount factor and \( k = -\rho \log \rho - (1-\rho) \log(1-\rho) \). The transversality condition ensures that the first term \( \lim_{i \to \infty} \rho^i dp_{t+i} = 0 \). In this rational expectation present value model, bubbles are assumed away by the transversality condition and the price dividend

---

4Several new methods have been developed in an effort to overcome the size distortion in long horizon regressions, including Valkanov (2003b), Liu and Maynard (2007), Hjalmarsson (2012), and Phillips and Lee (2013).
ratio is interpreted as the discounted value of future cash flows. In our framework, we do not impose the transversality condition and re-derive stock returns as a function of dividend ratio and bubbles in the log linear approximation framework. In our specification, we characterize the true model with this non-linear present value model and provide the power analysis for both long and short horizon predictive tests.

### 2.3 Structural Present-Value Model with Bubble

This section develops the structural relations between long-run predictability of stock returns and bubbles in the log-linearized framework of Campbell and Shiller (1988). However, we propose an extended specification where the bubble is not assumed away due to the finite horizon. In this setting, the bubble is assumed to be a latent variable following Evans (1991) and Phillips, Shi, Yu (2014, 2015) and we allow for it to be incorporated into dividend price dynamics. Thus it has a non-trivial effect on expected stock returns.

Let \( B_t \) denote a periodically explosive bubble component in the aggregate stock market, following the lognormal distribution, \[ B_{t+1} = \begin{cases} \rho_b^{-1} B_t \varepsilon_{B,t+1}, & \text{if } B_t < b \\ \zeta + (\pi \rho_b)^{-1} \theta_{t+1} (B_t - \rho_b \zeta) \varepsilon_{B,t+1}, & \text{if } B_t > b \end{cases} \]

where \( \rho_b^{-1} > 1, \rho_b \) is the discount factor, \( \varepsilon_{B,t} = \exp(y_t - \tau^2/2) \) and \( y_t \) is i.i.d. \( N(0, \tau^2) \). \( \theta_t \) is assumed to follow a Bernoulli process which takes the value 1 with probability \( \pi \) and 0 with probability \( 1 - \pi \), and \( \zeta \) is the remaining size after the bubble collapses. In this case, the bubble has a constant crash probability.

We also consider a logit specification for the bubble crash probability in the spirit of Froot and Obstfeld (1991). In this extension the probability \( \pi_t \) is time varying and dependent on the size of the bubble component (scaled by dividends). More specifically,

\[
\pi_t = \frac{\exp \left( \delta_0 + \delta_1 \frac{B_{t-1}}{D_{t-1}} \right)}{1 + \exp \left( \delta_0 + \delta_1 \frac{B_{t-1}}{D_{t-1}} \right)}
\]

where \( D_t \) denotes real dividend of the stock market and \( \delta_0 \) and \( \delta_1 \) are the logit parameters.\(^5\) In the special case when \( \delta_1 \) equals zero, the time varying crash probability is reduced to a constant crash probability. When \( \delta_1 > 0 \), the crash probability increases with the size of the scaled bubble.

The bubble has the conditional expectation \( E_t(B_{t+1}) = \rho_b^{-1} B_t \) and its collapse is triggered by the Bernoulli process if the bubble is bigger than the cut-off size \( b \).

\(^5\)In our simulations, we employ the same value of \( \delta_0 \) and \( \delta_1 \) as in Froot and Obstfeld (1991). They refer to \( \delta_0 \) and \( \delta_1 \) as \( c \) and \( d \) in their notation use \( D_t \) in place of \( D_{t-1} \).
We model the real dividend $D_t$ as the random walk process

$$D_t = u + D_{t-1} + \varepsilon_D,$$

where $\varepsilon_D \sim$ i.i.d. $\text{LOGN}(0, \sigma_D^2)$. Let $P_t$ denote the stock price which contains a market fundamental component $P^f_t$, where

$$P^f_t = \frac{u \rho}{(1-\rho)^2} + \frac{\rho}{1-\rho} D_t.$$

The actual price is then assumed to depend on both the fundamental price and the asset bubble as:

$$P_t = P^f_t + \kappa B_t,$$

where $\kappa$ controls the relative magnitudes of these two components.

Given the aggregate stock price equation, following Froot and Obstfeld (1991), we can divide both sides by $D_t$, where $D_t$ is real dividend process to derive

$$P_t D_t = P^f_t D_t + \kappa B_t D_t.$$

Taking logs on both sides of the equation and then carrying out a Taylor expansion around $B_t D_t$, we obtain

$$dp_{t+1} = \alpha + \phi dp_t + \varphi \log \left( \frac{B_t}{D_t} \right) + \varepsilon_t dp,$$

where $dp_t$ denotes the log dividend price ratio, $\log \left( \frac{B_t}{D_t} \right)$ denotes the log scaled bubble.

Let $r_{t+1}$ denote the total log return on the aggregate stock market, $\Delta d_{t+1}$ denote dividend growth rate,

$$r_{t+1} = \log \left( \frac{P_{t+1} + D_{t+1}}{P_t} \right),$$

$$\Delta d_{t+1} = \log \left( \frac{D_{t+1}}{D_t} \right).$$

We can relate expected stock returns to the $\log \left( \frac{B_t}{D_t} \right)$, $\Delta d_{t+1}$, through log-linearized return as

$$r_{t+1} = a + \Delta d_{t+1} - \rho dp_{t+1} + dp_t$$

$$\Delta d_{t+1} = u_d + b_d dp_t + \varepsilon_{t,d},$$

where $\rho = \frac{\exp(dp)}{1 + \exp(dp)}$, $dp = E(dp_t)$, as in Campbell and Schiller (1988). If we use this formula and substitute forward, it follows that

$$\sum_{j=1}^{k} \rho^{j-1} r_{t+j} = \sum_{j=1}^{k} \rho^{j-1} \alpha + dp_t + \rho^k (p_{t+k} - d_{t+k}) + \sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}.$$
horizons, thus this term will not disappear. We iterate $dp_t$ and $\Delta d_{t+1}$ forward $k$ periods, and then substitute them into the formula to obtain

$$
\sum_{j=1}^{k} \rho^{j-1} r_{t+j} = \text{const}^k + dp_t + \phi^k(p_t - d_t) + \rho^k \sum_{j=1}^{k} \phi^{j-1} \varphi \log \left( \frac{B_{t+k-j}}{D_{t+k-j}} \right) + \sum_{j=1}^{k} \rho^{j-1} b_d dp_{t+j-1} + \tilde{u}_{t+k}
$$

$$
= \text{const}^k + (1 - \phi^k)(p_t - d_t) - \sum_{j=1}^{k} \phi^{j-1} \varphi \log \left( \frac{B_{t+k-j}}{D_{t+k-j}} \right) + \tilde{u}_{t+k},
$$

where

$$
\text{const}^k = \sum_{j=1}^{k} \rho^{j-1} \alpha + \rho^k \sum_{j=1}^{k} \phi^{j-1} \alpha_d + \rho^k \sum_{j=1}^{k} \phi^{j-1} \alpha_d 
$$

$$
\tilde{u}_{t+k} = \rho^k \sum_{j=1}^{k} \phi^{j-1} \varepsilon_{t+k+1-j} + \sum_{j=1}^{k} \rho^{j-1} \varepsilon_{t+k,d}.
$$

This function provides the long run relation between the expected stock returns and the bubble. When $k = 1$, it reduces to the short run equation

$$
r_{t+1} = \text{const} + (1 + b_d - \rho \phi) dp_t + \rho \varphi \log \left( \frac{B_t}{D_t} \right) + \varepsilon_{t+1,d} - \rho \varepsilon_{t+1,dp} \quad (31)
$$

where $\text{const} = a + a_d - a_{dp} \rho$ and $\tilde{u}_{t+1} = \varepsilon_{t+1,d} - \rho \varepsilon_{t+1,dp}$. Interestingly, (31) can be decomposed into the extended data generating process proposed by Cochrane(2008)(note that we make the same assumption as the paper that constant term is equal to zero) as follows:

$$
r_{t+1} = \Delta d_{t+1} - \rho dp_{t+1} + dp_t
$$

$$
\Delta d_{t+1} = b_d dp_t + \varepsilon_{t+1,d}
$$

$$
dp_{t+1} = \phi dp_t + \varphi \log \left( \frac{B_t}{D_t} \right) + \varepsilon_{t+1,dp}.
$$

Shocks to expected dividend growth rate $\varepsilon_{t+1,d}$ and and dividend price ratio $\varepsilon_{t+1,dp}$ are assumed to be jointly independent and identically distributed over time with mean zero and covariance matrix

$$
\Sigma = \begin{bmatrix}
\sigma_d^2 & \sigma_{d,dp} \\
\sigma_{d,dp} & \sigma_{dp}^2
\end{bmatrix}.
$$

### 2.4 Simulation Design

Based on the model development in the previous section, we now describe the simulation design and parameter settings. Our simulations are designed to provide a (partial) answer to the following hypothetical questions. Suppose that the present value model with bubbles describes the true model from which the data were generated. Suppose also that the econometricians does not know the true model and therefore tests for the presence of return predictability using either
the standard short or long-horizon linear predictive regression specification. Would she uncover predictability when present? If so, which specification, the long or short-run specification, would be more likely to uncover this predictability?

To address this question, we repeatedly simulate pseudo return and dividend data from the structural present value model of the previous section. Then on each simulated pseudo data series, we test for return predictability using the standard short and long-horizon predictive tests described below. The average rejection rate over five thousand simulations is used to calculate the test power.\textsuperscript{6} By repeating this exercise over a range of values for $\varphi$, the parameter that controls the strength of the bubble component, we form a power curve. As a comparison, we also perform a similar exercise for the case in which returns are predictable in the absence of financial bubbles.

In order to undertake this analysis, we need to define the null and alternative hypotheses in both the ‘true’ present value model and the empirical specifications employed by the econometrician. The underlying non-predictability hypothesis that we impose in both cases is given by

$$H_0 : E_t r_{t+1} = \mu. \quad (32)$$

However, this implies different parameter restrictions in each model. In Section 2.4.1 we first discuss the restrictions and parameter settings of the present value model which forms the data generating process for the pseudo return and dividends data. In Section 2.4.2, we then describe the empirical specifications and tests employed by the econometrician.

### 2.4.1 Present value model: parameter settings and restrictions

In the present value model of the previous section, the non-predictability null hypothesis in (32) implies the joint restriction:

$$H_{0}^{PV} : \varphi = 0, \ b_d = \phi \rho - 1 \ and \ E_t \tilde{u}_{t+1} = 0. \quad (33)$$

This both eliminates the financial bubble component by setting $\varphi = 0$ and sets the coefficient $1 - \phi \rho + b_d$ on the dividend price ratio in (31) equal to zero. Thus, the null hypothesis eliminates both the traditional linear source of predictability that arises in the present value model without bubbles and also eliminates the financial bubble.

The traditional alternative in a present value model without bubbles is given by

$$H_{A}^{TRAD} : b_d > \phi \rho - 1 \ and \ \varphi = 0, \quad (34)$$

\textsuperscript{6}Both regression specifications involve size distortions. As discussed below, we address this both by comparing size corrected power in standard regression t-tests and by calculating power in tests designed to control size in both short and long-run predictive regressions with near unit root regressors.
in which the dividend price ratio has predictive power for future returns, despite the fact that no bubble is present ($\varphi = 0$). In order to provide a simulated power curve we simulate the rejection rates for a range of values for $b_d$. In particular, we specify the local alternative

$$b_d = \phi \varphi - 1 + \frac{\gamma}{T} \tag{35}$$

where we vary $\gamma \geq 0$, while holding $\varphi = 0$ in order to vary the distance from the null hypothesis. When $\gamma = 0$, $b_d = \phi \varphi - 1$ and the dividend price ratio has no predictive content for returns. As we increase $\gamma$, we increase the strength of the predictive relation between the returns and lagged dividend price ratio. We consider this alternative first as a basis of comparison.

Our primary focus is on the alternative in which predictability is due to the presence of a bubble process. We specify this alternative as

$$H^B_A : \varphi > 0 \text{ and } b_d = \phi \varphi - 1 \tag{36}$$

Under this alternative, the only source of predictability is the bubble process. The traditional linear predictability in (34) is eliminated by the restriction that $b_d = \phi \varphi - 1$. We simulate over a range of values for

$$\varphi = \frac{\gamma}{T} \tag{37}$$

where we vary $\gamma \geq 0$ in order to vary $\varphi$ and adjust $b_d$ accordingly in order to maintain the restriction $b_d = \phi \varphi - 1$. When $\varphi = \gamma = 0$, there is no bubble component. As we increase $\varphi$, we increase the strength of the bubble component.\(^7\)

We next focus our attention on the remaining parameters of the present value model of the previous section. The coefficient $\phi$ denotes the autoregressive (AR(1)) coefficient for the dividend price ratio under the null hypothesis. Following Cavanagh et al. (1995), it is common in the prediction regression literature to model this parameter using the local to unity framework of (Phillips 1987, Chan and Wei 1987):\(^8\)

$$\phi = 1 + \frac{c}{T}, \tag{38}$$

where the local-to-unity coefficient $c$ determines the strength of the persistance and $T$ denotes the available sample size. This near unit root process is a statistical device designed to model a process that has its largest root close or equal to one. Thus it captures the fact that the dividend price ratio has been observed to be persistent without imposing an exact unit root. Although

\(^7\)Note that the meaning of $\gamma$ in the two alternative specifications is not equivalent since it is used to set different parameters. Mixtures of these two alternative could also be considered. We do not pursue this here since we obtain fairly clear results from these two polar alternatives.

\(^8\)More recently Phillips and Magdalinos (2007) generalize the local-to-unity model to include mildly integrated processes and Magdalinos and Phillips (2009) employ this framework in the predictive regression literature to provide a novel solution for the predictive regression problem.
nor normally employed in asymptotic theory, the parameter $c$ provides a useful way to describe the persistence of the simulated dividend price ratio in a manner that is independent of the sample size. Since the value of the local-to-unity coefficient cannot be consistently estimated in the time series context, we have assessed the robustness of our results across a range of choices for $c$ and thus $\phi$. Specifically, we consider values of $c$ equal to $0$, $-2.5$, $-10$, and $-25$. Using a sample size of $T = 200$ as our base case, these correspond to values of the autoregressive coefficient $\phi$ equal to $1$, $0.9875$, $0.95$, and $0.875$ respectively.\footnote{We also re-ran many of the results using a sample size of $500$ for which the implied values of $\phi$ are $1$, $0.995$, $0.98$, and $0.95$. This yielded qualitatively similar results. However, since our remaining parameters are matched to a yearly frequency, we view $T = 200$ as the more realistic sample size.}

The remaining parameters are held fixed, staying the same under the null and both alternative hypotheses. Their values, provided in Table 5, are set realistically based on those reported by Evans (1991) and Cochrane (2008). As defined in the previous section $\sigma_f, P^f_0, \rho, b, B_0, \zeta, \tau, \kappa, \sigma_d, \sigma_{dp}, \sigma_{d,dp}$ denote the standard deviation of fundamental price innovation ($\sigma_f$), the initial fundamental price ($P^f_0$), the discount factor ($\rho$), the threshold size for the bubble to become large enough to potentially collapse ($b$), the size of the initial bubble ($B_0$), the size of the remaining bubble component after a collapse ($\zeta$), the standard deviation of the bubble component ($\tau$), the relative magnitudes of the fundamental price and bubble ($\kappa$), the standard deviations of the innovation processes driving the dividend growth ($\sigma_d$), rate and dividend price yields ratio ($\sigma_{dp}$), and covariance of the innovation processes for the dividend growth rate and the dividend yield ratio ($\sigma_{d,dp}$). $\pi$ denotes the probability of a bubble collapse in fixed version probability of collapse model. $\delta_0$ and $\delta_1$ are the logit parameters determining this probability in the model with a time varying probability of collapse. All the parameter values are annualized to generate yearly data series for the simulation purpose.

### 2.4.2 Empirical specifications and tests

Specifically, the empirical specification we assume has the form

$$r^k_{t+k} = \beta_0(k) + \beta_1(k)dp_t + \varepsilon^k_{t+k} \quad (39)$$

in which the dependent variable is defined as

$$r^k_{t+k} = \sum_{j=1}^{k} r_{t+j} \quad (40)$$

When $k = 1$ this specializes to the standard short-horizon predictive regression using the dividend price ratio as the predictor. When $k > 1$ it corresponds to what is arguably the most common long-horizon specification employed in practice. We calibrate our simulation to a one-year data
frequency and employ values of $k$ equal to one, five, and ten year horizons. In the empirical specification in (39), the null hypothesis is imposed as

$$H_0^{EMP} : \beta_1(k) = 0 \text{ and } E_{t+k} \epsilon_{t+k} = 0,$$

(41)

although it is the restriction on the regression coefficient $\beta_1(k)$ that is normally tested in practice. The empirically relevant alternative is

$$H_A^{EMP} : \beta_1(k) > 0,$$

(42)

in which case a low dividend price ratio implies higher expected future returns.

Although influential, standard OLS based tests of (41) are size distorted even at short horizons (Mankiw and Shapiro 1986, Cavanagh et al. 1995, Stambaugh 1999), due to both the persistence of the dividend price process and its contemporaneous correlation with the return innovations. As shown by Valkanov (2003b), the size distortion becomes even worse at longer horizons ($k > 1$). Although our primary concern is the test power, a size-distorted test may have power for the wrong reason. In our simulated power exercises, we therefore address this in two ways. First, we size-correct the standard OLS tests by recalculating their critical values via simulation to ensure that all tests have the correct rejection rate under the null hypothesis. Secondly, we provide a power comparison using a modified test by Hjalmarsson (2011) that both improves power and corrects size for tests of (41) under the local-to-unity specification in (38).

One attractive feature of Hjalmarsson (2011)’s test in this context is that it is designed to work without modification at both long ($k > 1$) and short horizons ($k = 1$). At the short-horizons it essentially equivalent to the test of (Campbell and Yogo 2006) and its feasible implementation uses a similar Bonferroni bounds procedure based on the inversion of a unit root test to address the inherent uncertainty regarding the value of $c$ in (38). On the other hand, a short-coming of the test, is that without modification, this Bonferroni procedure is invalid when the predictor is stationary or mildly integrated as shown by Phillips (2014). In order to ensure that none of our results are driven by the use of this Bonferroni procedure, we also provide results using an infeasible or ‘oracle’ version Hjalmarsson (2011)’s test, in which the true value of $c$ is employed in place of the Bonferroni bound.\footnote{Phillips and Lee (2013) provide a long-horizon IVX test of (42) that remains valid for a re-arranged version of the long-horizon test in which a one-period ahead return is regressed on a multi-period predictor. A distinct advantage of their test is that it avoids the need for Bonferroni bounds and remains valid across a much wider range of persistent processes. Further more, for the purpose of comparison, we include this long-horizon IVX test that applies directly to the traditional alternative specification in (42) in the long-horizon case, which is available upon request. Liu and Maynard (2007) propose a long-horizon predictive sign test with exact size, but also in the re-arranged long-horizon specification. Valkanov (2003b) provides long-horizon tests in the traditional long-horizon specification, but these tests require a long-horizon and are less easily compared to the short-horizon case ($k = 1$). There is an extensive financial econometrics literature proposing size corrective procedures in the short-horizon case when $k = 1$, many of which have not been extended to the long-horizon case.}
2.5 Results and Interpretation

As a point of comparison, in Table 7, we first compare the power of short and long-horizon regression predictive tests using the empirical specification in (39), when the true alternative hypothesis is specified by the present value model without a bubble component as in (34). The main table entries in Columns 3–10 provide the simulated rejection rates using five thousand simulations and a simulated sample size of $T = 200$. The top row of each panel provides the value of $\gamma$ in (35). The finite sample powers of the three predictive tests at the one ($k = 1$), five ($k = 5$), and ten ($k = 10$) year horizons are provided in the Columns 4-6 ($\gamma > 0$ in (35)). The third column ($\gamma = 0$ in (35)) shows the rejection rates under the null hypothesis in (33), under which returns are unpredictable.

In each panel three different testing procedures are employed. Rows 2-4 of each panel show the size adjusted power of the standard OLS tests based on the specification in (39). These three separate rows show the one year, five year, and ten year return horizons respectively, allowing us to compare size-adjusted power between the short (1 year) and long (5-10 year) horizon specifications. The size adjustment is implemented by replacing the standard critical values with their simulated counterparts. Without this adjustment the OLS tests would be size distorted, making test power more difficult to interpret. Furthermore, the size distortion would be more severe at longer horizons, so that the power of the tests would also be difficult to compare. Although this simulation based size-adjustment is not feasible in practice, it allows us to compare size-adjusted power across horizons for the standard long horizon tests, the empirical results from which have been influential in the finance literature.

Rows 5-7 of each panel show the ‘oracle’ version of the Hjalmarsson (2011) long horizon predictive test. By design, it has correct size and maximal power, within the context of the linear model in (39) when the dividend price ratio is a local to unity process. However, the oracle version of the test is infeasible in practice because it relies on knowledge of the true value of the local-to-unity parameter in (38). For this reason, in Rows 8-10 of each panel, we also employ the feasible version of Hjalmarsson (2011) long-horizon predictive test based on a first stage confidence interval for the local-to-unity parameter.\textsuperscript{11} Provided that the true process for the dividend price ratio is well described by a local-to-unity process, this provides a practical version of the test that can be applied at all three horizons.

The size adjusted power of the long-horizon OLS test shows a slight power advantage of about 1-2 percent over the short horizon version for alternatives close to the null hypotheses when the dividend price ratio is highly persistent $c = 0$ (unit root) and to a lesser extent $c = −2.5$ (near unit root). This is most apparent in Table 7, Panel A, Rows 2-4 and Columns 4-6 and Panel B, Rows 2-4, Column 5. In this case, because the predictive regressor is strongly persistent and its

\textsuperscript{11}We thank Erik Hjalmarsson for the use of his code.
innovations are correlated with the return innovation it is subject to the Stambaugh (1999) bias. Consequently, it is inefficient even after size-adjustment. Therefore, there is room to improve on the power of the short-horizon OLS based test and the long-horizon version of the same test does appear to provide some modest improvement close to the null hypothesis. This supports Cochrane (2008)’s argument that the long horizon specification may be more powerful than its short-run counterpart due to the strong negative correlation between the return and dividend price ratio innovations. However, the improvements are quite modest.

Campbell and Yogo (2006) show that a powerful short-run predictive test can be constructed by making explicit use of this residual correlation. Hjalmarsson (2011) extends his approach to include long-horizon predictive tests. Under some additional assumptions, he shows that the infeasible oracle version of his test achieves the optimal power available at each horizon. Using this (infeasible) test, there is no longer any power deficiency at the short horizon and therefore no reason to expect improvements at longer horizons. This reasoning is confirmed by the results in Rows 4-6 (Columns 4-10) of Table 7, Panels A and B. Using the oracle version of Hjalmarsson (2011)’s test, which essentially specializes to Campbell and Yogo (2006)’s test when \( k = 1 \), there is no longer any advantage to the longer horizon specifications. Indeed, the power drops by as much as five percent when moving from the one year horizon to the ten year horizon. As expected from theory, the oracle version of the test is also more powerful than the size-adjusted OLS test, with the improvement exceeding ten percent in some cases.

The consideration of the oracle test in the paragraph above is useful in clarifying the point that since the short-run regression is a correct specification within the traditional linear alternative, any power improvements at longer horizons dissipate once a powerful short-horizon test is employed. However, the oracle version of the test is not relevant in practice because it is infeasible and the Bonferroni bounds required to allow practical implementation may normally be expected to reduce its power. Therefore in rows 7-9, we also compare the power of the feasible Bonferroni version at the one, five, and ten year horizons. Generally, the power of this feasible version also declines with the return horizons, but by much less than in the oracle case. There are also some cases in which we find a modest rise in power, particularly Columns 5 and 7 of Panel B, where the power increases by about one percent moving the one year horizon to the five year horizon. As expected, the power of the Bonferroni approach is below that of its infeasible oracle counterpart, but still exceeds that of the size-adjusted OLS test.

Turning to Panels C and D of Table 7, we find that all three tests have their highest power at the one year horizon when the predictive regressor is less strongly persistent. In fact, in Panel D, when \( c = -25 \), corresponding to an autoregressive root of 0.875, the size-adjusted power of the OLS test drops by as much as seven percent moving from a one year horizon to a five year horizon and by as much as twenty percent moving to the ten year horizon (see Column 9). Roughly similar power declines at longer horizons are observed using the oracle version of Hjalmarsson (2011)’s
test, although the declines are bit less pronounced for the feasible Bonferroni version.

To summarize the results for the traditional linear alternatives, there are modest improvements of 1-2 percent in size-adjusted power at longer horizons using a standard OLS test when the predictor is persistent. However, these improvements may be too small to matter much in practice and generally dissipate once powerful short-run predictive tests are employed. On the other hand, there can be somewhat substantial power losses at long horizons when the predictor is less persistent. Overall these results support previous contentions that within the context of the standard linear model there is relatively little power gain at longer horizons and that the stronger empirical results observed at longer horizons may be primarily a consequence of size distortion (Boudoukh et al. 2008, Hjalmarsson 2011). The result is supported by the intuition that when the model is linear, the short horizon empirical specification (39 with $k = 1$) is a correct or approximately correct specification. In this case, there are unlikely to be any advantages to changing the specification or horizon once powerful tests are employed.

A more interesting question, in our view, concerns the difference in power across horizons when the true model is fundamentally non-linear. In this case, neither the short nor the long-horizon predictive regressions are correct specifications and there may therefore be a stronger a priori argument for the possibility of power advantages at longer horizons. Table 9 shows the same power comparisons when predictability is driven by the presence of a financial bubble component with a fixed probability of collapse. Since the empirical specification in (39) does not explicitly allow for the possibility of a financial bubble, both the short and long-horizon tests, may be viewed as misspecifications, which may nonetheless detect a predictive relationship between the valuations (dividend price ratio) and future returns, due to the presence of the financial bubble. Since the bubble is persistent, it is an a priori conceivable conjecture that this relationship could be stronger at longer horizons.

The power results for all three tests confirm that the dividend price ratio has predictive ability for returns in the presence of a financial bubble. For example, we note that the power of all three tests exceeds the size and increases steadily towards one as we move further away from the null hypothesis by moving rightward from Column 4 with a low value of $\gamma$ to Column 10 with a high value of $\gamma$. Given the presence of bubbles, this supports the use of valuation based predictors, such as the dividend price ratio. It thus confirms the intuition of value based investors that when stock prices are high relative to fundamental values this can serve as an indicator of ‘irrational exuberance’ or an asset bubble that may result in lower expected future returns.\(^{12}\)

\(^{12}\)Of course, the predictive regression does not distinguish predictability due to the presence of a bubble component from other sources of predictability, such as the more traditional alternative in (34). Moreover, if one is convinced that a bubble component is present and important, tests specifically designed to detect the bubble may likely be more powerful.
We next ask whether the presence of a bubble can justify the contention that predictive regressions can have higher power at longer horizons. If market valuations are high relative to fundamentals in the presence of an asset bubble, then lower (or negative) returns may be predicted after the bubble bursts. Since the bubble is expected to last or grow for several periods before bursting, intuition might suggest that this prediction could be stronger at longer horizons. For the OLS based tests, we again observe modest power improvements at the five and ten horizons of about one to two percent, whereas using the other two tests the power generally declines at longer horizons, in some cases by ten percent or more. It is also interesting to note that the power advantage of Hjalmarsson (2011)’s Campbell and Yogo (2006)-style-test is actually much larger in the presence of the bubble. For example, in Column 5 of Panel D, at the one year horizon the Bonferroni version of Hjalmarsson (2011)’s test obtains a fifty percent rejection before the size-adjusted OLS obtains even a twenty percent rejection.

The story-line changes when the root of the of the predictor is taken a bit further below one. For the standard linear alternative, Table 7 Panel D, showed a substantial drop in power for all three tests as the return horizon increased. However, in Panel D of Table 9, the size-adjusted power of the OLS test now increases by as much as six percent moving from the one to the ten year horizon. The power of the feasible version of Hjalmarsson (2011)’s test increases by an even greater nineteen percent (Panel D, Column 5, Rows 5-7), although its infeasible oracle counterpart continues to register a decline in power at higher horizons.

Finally, in Table 11, we consider an extension of the model along the lines of Froot and Obstfeld (1991), in which the probability of a collapse depends on the size of the asset bubble as specified in (30) where the values of $\delta_0$ and $\delta_1$ are given in Table 5. Since, all else equal, an increase in bubble size is associated with an increase in stock price level and thus a decreased dividend price ratio, this extension suggests that smaller values of the dividend price ratio will tend to be associated with a greater likelihood of a bubble collapse.

The resulting power comparisons are shown in Table 11. For the most part, they support the robustness of our results from the simpler bubble model in Table 9. First, we note that regardless of either the persistence parameter $c$ or the choice of test method, the dividend price ratio is predictive for returns in the presence of a financial bubble. Secondly, the size-adjusted power of the OLS test generally improves at longer horizons in the presence of the bubble, with modest improvements of no more than two percent when the autoregressive root is equal or very close to one and somewhat larger improvements of up to about five percent in Panel D when the root is a bit further from one. This confirms the robustness of our previous finding that even after size correction long-horizon regressions can improve power modestly when the true model allows for asset bubbles. However, the substantial long-horizon improvements observed for the Bonferroni test in Panel D in the simpler model of Table 9 are no longer present, leading us to conclude that this result, although striking, is somewhat less robust.
Overall, we find strong evidence that the dividend price ratio is predictive for returns in models with financial bubbles using both short and long-horizon regressions. When this predictor is very close to having a unit root, we observe a very modest improvement to the size-adjusted OLS tests at longer horizons, but these improvements don’t extend to more powerful predictive tests. When the predictor is strongly autocorrelated but with a root that is a bit further from one, we find modest improvements of about five percent in size-adjusted power in OLS based tests.

2.6 Conclusion

The paper provides a comprehensive comparative evaluation of the power of long and short horizon predictive regression when the true data generating process is a structural non linear present value model. In particular, we study the relationship between asset bubbles and tests of return predictability using valuation-based predictors, such as the dividend price ratio. An asset bubble may be reflected in high valuations that correct after the bubble bursts. This suggests that valuations may be predictive in models with asset bubbles. Our findings confirm the power of asset valuation ratios to predict asset returns in the presence of bubbles.

Since an asset bubble may last or expand for several periods before eventually bursting, intuition might also suggest stronger predictability at longer horizons. However, we find only modest evidence that this is the case. The ability to detect bubble driven return predictability seems to be relatively similar at both short and long horizons. Our paper therefore adds to existing evidence that large increases in predictive power at longer horizons are difficult to justify after properly controlling for size. Campbell (2001) initially asked whether increased power could be justified even in linear models in which the short-horizon regression would seem to be correctly specified. By contrast, we find only modest advantages to the long-horizon specification even in non-linear models of asset bubbles that seem a priori favourable to prediction at longer horizons.
<table>
<thead>
<tr>
<th>Parameters Employed in all Specifications</th>
<th>( \sigma_f )</th>
<th>( P^f_0 )</th>
<th>( \rho )</th>
<th>( b )</th>
<th>( B_0 )</th>
<th>( \zeta )</th>
<th>( \tau )</th>
<th>( \kappa )</th>
<th>( \sigma_d )</th>
<th>( \sigma_{dp} )</th>
<th>( \sigma_{d,dp} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.869</td>
<td>41.195</td>
<td>0.952</td>
<td>1</td>
<td>0.50</td>
<td>0.50</td>
<td>0.05</td>
<td>20</td>
<td>14</td>
<td>15.3</td>
<td>7.5</td>
</tr>
</tbody>
</table>

Fixed Probability of Bubble Collapse Specification
\[ \pi = 0.85 \]

Time Varying Probability of Bubble Collapse Specification
\[ \delta_0 = 14.63 \quad \delta_1 = 0.26 \]

As defined in Section 2.3, \( \sigma_f, P^f_0, \rho, b, B_0, \zeta, \tau, \kappa, \sigma_d, \sigma_{dp}, \sigma_{d,dp} \) denote the standard deviation of fundamental price innovation (\( \sigma_f \)), the initial fundamental price (\( P^f_0 \)), the discount factor (\( \rho \)), the threshold size for the bubble to become large enough to potentially collapse (\( b \)), the size of the initial bubble (\( B_0 \)), the size of the remaining bubble component after a collapse (\( \zeta \)), the standard deviation of the bubble component (\( \tau \)), the relative magnitudes of the fundamental price and bubble (\( \kappa \)), the standard deviations of the innovation processes driving the dividend growth (\( \sigma_d \)), rate and dividend price yields ratio (\( \sigma_{dp} \)), and covariance of the innovation processes for the dividend growth rate and the dividend yield ratio (\( \sigma_{d,dp} \)). \( \pi \) denotes the probability of a bubble collapse in fixed version probability of collapse model. \( \delta_0 \) and \( \delta_1 \) are the logit parameters determining this probability in the model with a time varying probability of collapse. All the parameter values are annualized to generate yearly data series for the simulation purpose. To check the robustness of our simulation results, we have also generated the power curves using different values of \( \kappa \) controlling the size of asset bubble, finding that simulation results remain unchanged.
Table 6: Rejection Rates for Standard Linear Alternative

<table>
<thead>
<tr>
<th>Test Method</th>
<th>Horizons(k)/γ</th>
<th>0.0000</th>
<th>0.0088</th>
<th>0.0175</th>
<th>0.0263</th>
<th>0.0350</th>
<th>0.0438</th>
<th>0.0525</th>
<th>0.0613</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>1 Year</td>
<td>0.0518</td>
<td>0.1186</td>
<td>0.2704</td>
<td>0.4750</td>
<td>0.6726</td>
<td>0.8278</td>
<td>0.9168</td>
<td>0.9552</td>
</tr>
<tr>
<td></td>
<td>5 Year</td>
<td>0.0600</td>
<td>0.1276</td>
<td>0.2930</td>
<td>0.4872</td>
<td>0.6780</td>
<td>0.8206</td>
<td>0.9076</td>
<td>0.9456</td>
</tr>
<tr>
<td>OLS</td>
<td>10 Year</td>
<td>0.0606</td>
<td>0.1308</td>
<td>0.2818</td>
<td>0.4734</td>
<td>0.6588</td>
<td>0.7882</td>
<td>0.8770</td>
<td>0.9226</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year</td>
<td>0.0538</td>
<td>0.1666</td>
<td>0.3848</td>
<td>0.6060</td>
<td>0.7618</td>
<td>0.8688</td>
<td>0.9280</td>
<td>0.9534</td>
</tr>
<tr>
<td>(2011)</td>
<td>5 Year</td>
<td>0.0544</td>
<td>0.1788</td>
<td>0.3804</td>
<td>0.5828</td>
<td>0.7374</td>
<td>0.8484</td>
<td>0.9090</td>
<td>0.9384</td>
</tr>
<tr>
<td>Oracle</td>
<td>10 Year</td>
<td>0.0560</td>
<td>0.1628</td>
<td>0.3538</td>
<td>0.5510</td>
<td>0.7010</td>
<td>0.8188</td>
<td>0.8792</td>
<td>0.9144</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year</td>
<td>0.0544</td>
<td>0.1688</td>
<td>0.3684</td>
<td>0.5844</td>
<td>0.7534</td>
<td>0.8692</td>
<td>0.9346</td>
<td>0.9630</td>
</tr>
<tr>
<td>(2011)</td>
<td>5 Year</td>
<td>0.0576</td>
<td>0.1666</td>
<td>0.3628</td>
<td>0.5666</td>
<td>0.7300</td>
<td>0.8574</td>
<td>0.9198</td>
<td>0.9492</td>
</tr>
<tr>
<td>Bonferroni</td>
<td>10 Year</td>
<td>0.0572</td>
<td>0.1592</td>
<td>0.3456</td>
<td>0.5460</td>
<td>0.7042</td>
<td>0.8282</td>
<td>0.8920</td>
<td>0.9298</td>
</tr>
</tbody>
</table>

The table shows simulated rejection rates both when the data is generated under the null hypothesis in (33) (Column 3, $\gamma = 0$) and when it is generated under the standard linear alternative hypothesis in (34)-(35) (Columns 4-10, $\gamma > 0$). The three tests listed in Column 1 are all based on the empirical specification in (39) and described in Section 2.4.2. Column 2 indicates the test horizon or value of $k$ in (39). The standard OLS test is (infeasibly) size-adjusted by simulation to allow comparison of size-adjusted power. The local-to-unity parameter describing the persistence of the dividend price ratio is varied across the four panels. The last seven columns of the top row of each panel provides the values of $\gamma$ used in the panel.
The table shows simulated rejection rates both when the data is generated under the null hypothesis in (33) (Column 3, $\gamma = 0$) and when it is generated under the standard linear alternative hypothesis in (34)-(35) (Columns 4-10, $\gamma > 0$). The three tests listed in Column 1 are all based on the empirical specification in (39) and described in Section 2.4.2. Column 2 indicates the test horizon or value of $k$ in (39). The standard OLS test is (infeasibly) size-adjusted by simulation to allow comparison of size-adjusted power. The local-to-unity parameter describing the persistence of the dividend price ratio is varied across the four panels. The last seven columns of the top row of each panel provides the values of $\gamma$ used in the panel.
Table 8: Rejection Rates for Bubble Alternative with Fixed Crash Probability

<table>
<thead>
<tr>
<th>Test Method</th>
<th>Horizons(k)/γ</th>
<th>0.0000</th>
<th>0.2625</th>
<th>0.5625</th>
<th>0.8250</th>
<th>1.1625</th>
<th>1.3875</th>
<th>1.7625</th>
<th>2.2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>1 Year</td>
<td>0.0518</td>
<td>0.0782</td>
<td>0.1652</td>
<td>0.2956</td>
<td>0.5256</td>
<td>0.6936</td>
<td>0.8798</td>
<td>0.9832</td>
</tr>
<tr>
<td></td>
<td>5 Year</td>
<td>0.0600</td>
<td>0.0862</td>
<td>0.1748</td>
<td>0.3152</td>
<td>0.5264</td>
<td>0.6786</td>
<td>0.8624</td>
<td>0.9758</td>
</tr>
<tr>
<td>OLS</td>
<td>10 Year</td>
<td>0.0606</td>
<td>0.0896</td>
<td>0.1762</td>
<td>0.3154</td>
<td>0.5064</td>
<td>0.6492</td>
<td>0.8256</td>
<td>0.9588</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year</td>
<td>0.0538</td>
<td>0.2482</td>
<td>0.8002</td>
<td>0.9758</td>
<td>0.9992</td>
<td>0.9998</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>(2011) 5 Year</td>
<td>0.0544</td>
<td>0.2196</td>
<td>0.6896</td>
<td>0.9048</td>
<td>0.9752</td>
<td>0.9870</td>
<td>0.9972</td>
<td>0.9998</td>
</tr>
<tr>
<td>Oracle</td>
<td>10 Year</td>
<td>0.0560</td>
<td>0.1976</td>
<td>0.5728</td>
<td>0.7758</td>
<td>0.9210</td>
<td>0.9636</td>
<td>0.9910</td>
<td></td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year</td>
<td>0.0544</td>
<td>0.2156</td>
<td>0.7636</td>
<td>0.9702</td>
<td>0.9986</td>
<td>0.9996</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>(2011) 5 Year</td>
<td>0.0576</td>
<td>0.2018</td>
<td>0.6548</td>
<td>0.8836</td>
<td>0.9684</td>
<td>0.9842</td>
<td>0.9950</td>
<td>0.9994</td>
</tr>
<tr>
<td>Bonferroni</td>
<td>10 Year</td>
<td>0.0572</td>
<td>0.1844</td>
<td>0.5458</td>
<td>0.7556</td>
<td>0.8700</td>
<td>0.9096</td>
<td>0.9578</td>
<td>0.9892</td>
</tr>
</tbody>
</table>

Panel A: $c = 0$

<table>
<thead>
<tr>
<th>Test Method</th>
<th>Horizons(k)/γ</th>
<th>0.0000</th>
<th>0.2250</th>
<th>0.5625</th>
<th>0.8250</th>
<th>0.9375</th>
<th>1.2375</th>
<th>1.5750</th>
<th>1.8750</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>1 Year</td>
<td>0.0524</td>
<td>0.0712</td>
<td>0.1968</td>
<td>0.3850</td>
<td>0.4612</td>
<td>0.6986</td>
<td>0.8756</td>
<td>0.9578</td>
</tr>
<tr>
<td></td>
<td>5 Year</td>
<td>0.0556</td>
<td>0.0760</td>
<td>0.2078</td>
<td>0.3922</td>
<td>0.4598</td>
<td>0.6940</td>
<td>0.8608</td>
<td>0.9486</td>
</tr>
<tr>
<td>OLS</td>
<td>10 Year</td>
<td>0.0548</td>
<td>0.0742</td>
<td>0.2036</td>
<td>0.3882</td>
<td>0.4510</td>
<td>0.6732</td>
<td>0.8352</td>
<td>0.9296</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year</td>
<td>0.0520</td>
<td>0.1656</td>
<td>0.7768</td>
<td>0.9810</td>
<td>0.9952</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>(2011) 5 Year</td>
<td>0.0516</td>
<td>0.1502</td>
<td>0.7012</td>
<td>0.9374</td>
<td>0.9700</td>
<td>0.9958</td>
<td>0.9990</td>
<td>1.0000</td>
</tr>
<tr>
<td>Oracle</td>
<td>10 Year</td>
<td>0.0492</td>
<td>0.1362</td>
<td>0.6048</td>
<td>0.8484</td>
<td>0.8930</td>
<td>0.9516</td>
<td>0.9736</td>
<td>0.9860</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year</td>
<td>0.0498</td>
<td>0.0904</td>
<td>0.5152</td>
<td>0.8782</td>
<td>0.9432</td>
<td>0.9990</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>(2011) 5 Year</td>
<td>0.0546</td>
<td>0.0960</td>
<td>0.4936</td>
<td>0.8324</td>
<td>0.9058</td>
<td>0.9846</td>
<td>0.9978</td>
<td>0.9998</td>
</tr>
<tr>
<td>Bonferroni</td>
<td>10 Year</td>
<td>0.0534</td>
<td>0.0992</td>
<td>0.4570</td>
<td>0.7716</td>
<td>0.8430</td>
<td>0.9500</td>
<td>0.9838</td>
<td>0.9946</td>
</tr>
</tbody>
</table>

Panel B: $c = -2.5$

The table shows simulated rejection rates both when the data is generated under the null hypothesis in (33) (Column 3, $\gamma = 0$) and when it is generated under the alternative of a bubble process in (36)-(37) (Columns 4-10, $\gamma > 0$) with fixed probability $\pi$ of collapse. The three tests listed in Column 1 are all based on the empirical specification in (39) and described in Section 2.4.2. Column 2 indicates the test horizon or value of $k$ in (39). The standard OLS test is (infeasibly) size-adjusted by simulation to allow comparison of size-adjusted power. The local-to-unity parameter describing the persistence of the dividend price ratio is varied across the four panels. The last seven columns of the top row of each panel provides the values of $\gamma$ used in the panel.
Table 9: Rejection Rates for Bubble Alternative with Fixed Crash Probability (Continued)

<table>
<thead>
<tr>
<th>Test Method</th>
<th>Horizons(k)/γ</th>
<th>0.0000</th>
<th>0.4500</th>
<th>0.7125</th>
<th>0.9375</th>
<th>1.0875</th>
<th>1.2375</th>
<th>1.5000</th>
<th>1.9500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>1 Year</td>
<td>0.0554</td>
<td>0.1274</td>
<td>0.2806</td>
<td>0.4688</td>
<td>0.6152</td>
<td>0.7430</td>
<td>0.8928</td>
<td>0.9854</td>
</tr>
<tr>
<td></td>
<td>5 Year</td>
<td>0.0530</td>
<td>0.1268</td>
<td>0.2826</td>
<td>0.4682</td>
<td>0.6104</td>
<td>0.7326</td>
<td>0.8794</td>
<td>0.9800</td>
</tr>
<tr>
<td></td>
<td>10 Year</td>
<td>0.0516</td>
<td>0.1330</td>
<td>0.2826</td>
<td>0.4690</td>
<td>0.6038</td>
<td>0.7264</td>
<td>0.8672</td>
<td>0.9732</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year</td>
<td>0.0536</td>
<td>0.3968</td>
<td>0.8170</td>
<td>0.9770</td>
<td>0.9934</td>
<td>0.9990</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>(2011)</td>
<td>5 Year</td>
<td>0.0470</td>
<td>0.3538</td>
<td>0.7526</td>
<td>0.9404</td>
<td>0.9828</td>
<td>0.9952</td>
<td>0.9998</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>10 Year</td>
<td>0.0486</td>
<td>0.3178</td>
<td>0.6630</td>
<td>0.8672</td>
<td>0.9366</td>
<td>0.9676</td>
<td>0.9874</td>
<td>0.9940</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year</td>
<td>0.0528</td>
<td>0.1110</td>
<td>0.2650</td>
<td>0.5666</td>
<td>0.7574</td>
<td>0.8952</td>
<td>0.9802</td>
<td>0.9994</td>
</tr>
<tr>
<td>(2011)</td>
<td>5 Year</td>
<td>0.0476</td>
<td>0.1172</td>
<td>0.2830</td>
<td>0.5622</td>
<td>0.7446</td>
<td>0.8782</td>
<td>0.9706</td>
<td>0.9988</td>
</tr>
<tr>
<td>Bonferroni</td>
<td>10 Year</td>
<td>0.0516</td>
<td>0.1414</td>
<td>0.3266</td>
<td>0.5910</td>
<td>0.7540</td>
<td>0.8756</td>
<td>0.9642</td>
<td>0.9972</td>
</tr>
</tbody>
</table>

The table shows simulated rejection rates both when the data is generated under the null hypothesis in (33) (Column 3, $\gamma = 0$) and when it is generated under the alternative of a bubble process in (36)-(37) (Columns 4-10, $\gamma > 0$) with fixed probability $\pi$ of collapse. The three tests listed in Column 1 are all based on the empirical specification in (39) and described in Section 2.4.2. Column 2 indicates the test horizon or value of $k$ in (39). The standard OLS test is (infeasibly) size-adjusted by simulation to allow comparison of size-adjusted power. The local-to-unity parameter describing the persistence of the dividend price ratio is varied across the four panels. The last seven columns of the top row of each panel provides the values of $\gamma$ used in the panel.
Table 10: Rejection Rates for Bubble Alternative with Time Varying Crash Probability

<table>
<thead>
<tr>
<th>Test Method</th>
<th>Horizons(k)/γ</th>
<th>0.0000</th>
<th>0.3750</th>
<th>0.6750</th>
<th>0.9750</th>
<th>1.2375</th>
<th>1.5375</th>
<th>1.8000</th>
<th>2.3625</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: c = 0</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>1 Year</td>
<td>0.0518</td>
<td>0.1064</td>
<td>0.2186</td>
<td>0.3980</td>
<td>0.5922</td>
<td>0.7696</td>
<td>0.8882</td>
<td>0.9898</td>
</tr>
<tr>
<td></td>
<td>5 Year</td>
<td>0.0600</td>
<td>0.1162</td>
<td>0.2298</td>
<td>0.4054</td>
<td>0.5902</td>
<td>0.7594</td>
<td>0.8702</td>
<td>0.9836</td>
</tr>
<tr>
<td>OLS</td>
<td>10 Year</td>
<td>0.0606</td>
<td>0.1196</td>
<td>0.2350</td>
<td>0.3958</td>
<td>0.5672</td>
<td>0.7270</td>
<td>0.8322</td>
<td>0.9708</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year</td>
<td>0.0538</td>
<td>0.4742</td>
<td>0.9094</td>
<td>0.9944</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>(2011)</td>
<td>5 Year</td>
<td>0.0544</td>
<td>0.4118</td>
<td>0.8120</td>
<td>0.9484</td>
<td>0.9828</td>
<td>0.9922</td>
<td>0.9970</td>
<td>0.9998</td>
</tr>
<tr>
<td>Oracle</td>
<td>10 Year</td>
<td>0.0560</td>
<td>0.3462</td>
<td>0.6824</td>
<td>0.8286</td>
<td>0.9030</td>
<td>0.9426</td>
<td>0.9708</td>
<td>0.9936</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year</td>
<td>0.0544</td>
<td>0.4274</td>
<td>0.8888</td>
<td>0.9930</td>
<td>0.9998</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>(2011)</td>
<td>5 Year</td>
<td>0.0576</td>
<td>0.3762</td>
<td>0.7868</td>
<td>0.9332</td>
<td>0.9786</td>
<td>0.9894</td>
<td>0.9960</td>
<td>0.9998</td>
</tr>
<tr>
<td>Bonferroni</td>
<td>10 Year</td>
<td>0.0572</td>
<td>0.3232</td>
<td>0.6558</td>
<td>0.8090</td>
<td>0.8890</td>
<td>0.9334</td>
<td>0.9624</td>
<td>0.9902</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Panel B: c = −2.5</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>1 Year</td>
<td>0.0524</td>
<td>0.1110</td>
<td>0.2406</td>
<td>0.3850</td>
<td>0.5806</td>
<td>0.7350</td>
<td>0.8756</td>
<td>0.9684</td>
</tr>
<tr>
<td>Adjusted</td>
<td>5 Year</td>
<td>0.0556</td>
<td>0.1168</td>
<td>0.2504</td>
<td>0.3922</td>
<td>0.5710</td>
<td>0.7260</td>
<td>0.8608</td>
<td>0.9588</td>
</tr>
<tr>
<td>OLS</td>
<td>10 Year</td>
<td>0.0548</td>
<td>0.1128</td>
<td>0.2542</td>
<td>0.3882</td>
<td>0.5542</td>
<td>0.7018</td>
<td>0.8352</td>
<td>0.9438</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year</td>
<td>0.0520</td>
<td>0.3474</td>
<td>0.8330</td>
<td>0.9810</td>
<td>0.9992</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>(2011)</td>
<td>5 Year</td>
<td>0.0516</td>
<td>0.3116</td>
<td>0.7532</td>
<td>0.9374</td>
<td>0.9906</td>
<td>0.9976</td>
<td>0.9990</td>
<td>0.9996</td>
</tr>
<tr>
<td>Oracle</td>
<td>10 Year</td>
<td>0.0492</td>
<td>0.2712</td>
<td>0.6510</td>
<td>0.8484</td>
<td>0.9300</td>
<td>0.9580</td>
<td>0.9736</td>
<td>0.9860</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year</td>
<td>0.0498</td>
<td>0.1820</td>
<td>0.5900</td>
<td>0.8782</td>
<td>0.9838</td>
<td>0.9996</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>(2011)</td>
<td>5 Year</td>
<td>0.0546</td>
<td>0.1894</td>
<td>0.5610</td>
<td>0.8324</td>
<td>0.9602</td>
<td>0.9892</td>
<td>0.9978</td>
<td>0.9998</td>
</tr>
<tr>
<td>Bonferroni</td>
<td>10 Year</td>
<td>0.0534</td>
<td>0.1838</td>
<td>0.5204</td>
<td>0.7716</td>
<td>0.9122</td>
<td>0.9606</td>
<td>0.9838</td>
<td>0.9952</td>
</tr>
</tbody>
</table>

The table shows simulated rejection rates both when the data is generated under the null hypothesis in (33) (Column 3, γ = 0) and when it is generated under the alternative of a bubble process in (36)-(37) (Columns 4-10, γ > 0) with a time varying probability of collapse given by (30). The three tests listed in Column 1 are all based on the empirical specification in (39) and described in Section 2.4.2. Column 2 indicates the test horizon or value of k in (39). The standard OLS test is (infeasibly) size-adjusted by simulation to allow comparison of size-adjusted power. The local-to-unity parameter describing the persistence of the dividend price ratio is varied across the four panels. The last seven columns of the top row of each panel provides the values of γ used in the panel.
Table 11: Rejection Rates for Bubble Alternative with Time Varying Crash Probability (Continued)

<table>
<thead>
<tr>
<th>Test Method</th>
<th>Horizons(k)/γ</th>
<th>0.0000</th>
<th>0.3750</th>
<th>0.6375</th>
<th>0.8625</th>
<th>0.9750</th>
<th>1.1250</th>
<th>1.3500</th>
<th>2.4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>1 Year</td>
<td>0.0554</td>
<td>0.1030</td>
<td>0.2252</td>
<td>0.4234</td>
<td>0.5128</td>
<td>0.6552</td>
<td>0.8138</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>5 Year</td>
<td>0.0530</td>
<td>0.1002</td>
<td>0.2228</td>
<td>0.4158</td>
<td>0.5098</td>
<td>0.6464</td>
<td>0.7986</td>
<td>0.9984</td>
</tr>
<tr>
<td>OLS</td>
<td>10 Year</td>
<td>0.0516</td>
<td>0.1064</td>
<td>0.2350</td>
<td>0.4156</td>
<td>0.5132</td>
<td>0.6484</td>
<td>0.7896</td>
<td>0.9968</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year (2011)</td>
<td>0.0536</td>
<td>0.2706</td>
<td>0.7164</td>
<td>0.9426</td>
<td>0.9828</td>
<td>0.9976</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>5 Year (2011)</td>
<td>0.0470</td>
<td>0.2434</td>
<td>0.6492</td>
<td>0.8932</td>
<td>0.9626</td>
<td>0.9914</td>
<td>0.9986</td>
<td>1.0000</td>
</tr>
<tr>
<td>Oracle</td>
<td>10 Year</td>
<td>0.0486</td>
<td>0.2256</td>
<td>0.5692</td>
<td>0.8116</td>
<td>0.8876</td>
<td>0.9538</td>
<td>0.9798</td>
<td>0.9982</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year (2011)</td>
<td>0.0512</td>
<td>0.1712</td>
<td>0.5094</td>
<td>0.8048</td>
<td>0.9148</td>
<td>0.9860</td>
<td>0.9970</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>5 Year (2011)</td>
<td>0.0500</td>
<td>0.1626</td>
<td>0.4602</td>
<td>0.7330</td>
<td>0.8590</td>
<td>0.9480</td>
<td>0.9892</td>
<td>1.0000</td>
</tr>
<tr>
<td>Bonferroni</td>
<td>10 Year</td>
<td>0.0492</td>
<td>0.1568</td>
<td>0.4096</td>
<td>0.6410</td>
<td>0.7676</td>
<td>0.8762</td>
<td>0.9460</td>
<td>0.9974</td>
</tr>
</tbody>
</table>

Panel C: c = −10

<table>
<thead>
<tr>
<th>Test Method</th>
<th>Horizons(k)/γ</th>
<th>0.0000</th>
<th>0.5250</th>
<th>0.8250</th>
<th>1.0125</th>
<th>1.2000</th>
<th>1.4250</th>
<th>1.6125</th>
<th>2.2125</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>1 Year</td>
<td>0.0510</td>
<td>0.1220</td>
<td>0.2682</td>
<td>0.4260</td>
<td>0.5752</td>
<td>0.7808</td>
<td>0.8886</td>
<td>0.9962</td>
</tr>
<tr>
<td></td>
<td>5 Year</td>
<td>0.0532</td>
<td>0.1326</td>
<td>0.3014</td>
<td>0.4628</td>
<td>0.6138</td>
<td>0.8066</td>
<td>0.8992</td>
<td>0.9966</td>
</tr>
<tr>
<td>OLS</td>
<td>10 Year</td>
<td>0.0476</td>
<td>0.1320</td>
<td>0.3112</td>
<td>0.4788</td>
<td>0.6238</td>
<td>0.8056</td>
<td>0.8978</td>
<td>0.9934</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year (2011)</td>
<td>0.0512</td>
<td>0.3332</td>
<td>0.7692</td>
<td>0.9326</td>
<td>0.9890</td>
<td>0.9984</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>5 Year (2011)</td>
<td>0.0500</td>
<td>0.3132</td>
<td>0.7058</td>
<td>0.8756</td>
<td>0.9664</td>
<td>0.9940</td>
<td>0.9988</td>
<td>1.0000</td>
</tr>
<tr>
<td>Oracle</td>
<td>10 Year</td>
<td>0.0492</td>
<td>0.2830</td>
<td>0.6192</td>
<td>0.7874</td>
<td>0.8984</td>
<td>0.9570</td>
<td>0.9798</td>
<td>0.9968</td>
</tr>
<tr>
<td>Hjalmarsson</td>
<td>1 Year (2011)</td>
<td>0.0520</td>
<td>0.7208</td>
<td>0.9810</td>
<td>0.9994</td>
<td>0.9999</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>5 Year (2011)</td>
<td>0.0516</td>
<td>0.6406</td>
<td>0.9374</td>
<td>0.9848</td>
<td>0.9946</td>
<td>0.9984</td>
<td>0.9994</td>
<td>1.0000</td>
</tr>
<tr>
<td>Bonferroni</td>
<td>10 Year</td>
<td>0.0492</td>
<td>0.5498</td>
<td>0.8484</td>
<td>0.9174</td>
<td>0.9440</td>
<td>0.9660</td>
<td>0.9792</td>
<td>0.9932</td>
</tr>
</tbody>
</table>

The table shows simulated rejection rates both when the data is generated under the null hypothesis in (33) (Column 3, γ = 0) and when it is generated under the alternative of a bubble process in (36)-(37) (Columns 4-10, γ > 0) with a time varying probability of collapse given by (30). The three tests listed in Column 1 are all based on the empirical specification in (39) and described in Section 2.4.2. Column 2 indicates the test horizon or value of k in (39). The standard OLS test is (infeasibly) size-adjusted by simulation to allow comparison of size-adjusted power. The local-to-unity parameter describing the persistence of the dividend price ratio is varied across the four panels. The last seven columns of the top row of each panel provides the values of γ used in the panel.
Chapter 3

Capital Structure Assessment and Learning about Cash Flow

3.1 Introduction and Literature Review

What determines firms’ leverage ratios? The puzzle, proposed by the seminal paper of Myers (1984), has been discussed at length in the financial economics literature. Almost all previous theoretical models assume that cash flows follow a Brownian motion with observed known mean cash flow. However, there appears not to be much research devoted to modelling capital structure with parameter uncertainty in a continuous time framework. My paper shows that leverage ratios can be explained by managers’ uncertainty about future mean log cash flows. In my model, firms’ observed leverage ratios will change dynamically with firm age as managers update their subjective beliefs.

Much previous study has been devoted to modelling uncertainty applied in asset pricing. In almost all its sub fields, such as return predictability, equity premium, option markets, credit spreads, the term structure of interest rates and long-run risk, Bayesian learning theory is utilized actively, see, for example Pastor and Veronesi (2002), Brennan and Xia (2001), Xia (2001), Dufrense and Goldstein (2001). However, there remains a disconnect between the research on capital structure and model uncertainty.

The absence of Bayesian learning in continuous time capital structure models may be due to the fact that its introduction complicates the modelling process. In fact, even without learning, previous capital structure models appear very complex, often requiring numerical solutions for an interpretable result. Yang (2013) makes the first attempt to use Bayesian uncertainty to explain corporate structure. In his model, changes are induced by differences of opinion between external investors and managers who hold divergent beliefs. His model requires the use of numerical analysis to simulate the effect of differential belief on the debt ratio.

My paper builds on the structural model of debts (Merton (1974), Black and Cox (1976), Morellec (2005), Streblaev (2007)) based on the assumption that firm managers cannot observe the mean log cash flow, which is only believed to follow some prior distribution. Therefore, managers will form their beliefs about future cash flow to make their assessment of market equity value and leverage ratios. Another feature of the model is that a stochastic discount factor is used in place of the deterministic discount factors commonly used in the capital structure literature. The model incorporates the dynamic properties of the discount factor, which is correlated with future cash flows. Furthermore, I extend the asset pricing model of Pastor and Veronesi (2003) to a corporate finance setting. They provide a stock valuations model with profitability uncertainty.
However, their model assumes that firms’ values come only from equities without having access to external funding or issuing debts. My analysis posits that firms, regardless of their maturity, use debts throughout the operational period of the firms.

Pastor and Veronesi (2003) report, but do not explain, an interesting pronounced increase in median firm leverage ratios with firm age. I use Bayesian learning to explain this increase in the following way. I assume that expected cash flow is unknown but that managers/investors\(^1\) gradually learn about mean cash flow from realized cash flow using rational Bayesian updating. Therefore, the posterior variance of mean log cash flow starts high, but falls gradually due to the resolution of uncertainty.

My parsimonious model abstracts from the long run steady growth level of cash flow by assuming it is zero. An upward growth trend for firm cash flow would imply a higher initial present discount cash flow value and hence a higher firm value that would be equivalent to a higher mean cash flow in a model without cash flow growth. Provided that investors discount future consumption trend would also imply an expected cash flow growth in firm value in the absence of the Bayesian learning captured by my model. Intuitively, firms would be expected to grow at the stochastic interest rate adjusted by a risk premium. This would imply a tendency of firm value to grow with firm age, which is the opposite prediction provided by my Bayesian learning effect. In practice, I expect both countervailing effects to be present.\(^2\) However, while the upward trend in firm value due to the discounting of cash flow growth would be the same across all firm ages,\(^3\) the Bayesian learning effects of my model are most heavily pronounced during the first several years of a firm’s life, during which time substantial cash flow uncertainty is resolved. To be more precisely, I am effectively modelling deviations from expected firm value relative to long term trend. Within this period the drop in expected firm value due to the resolution of uncertainty is likely to dominate the upward trend in firm value that would otherwise be observed. For older firms, the converse may hold true.

The resolution of uncertainty captured by model may also be most relevant to high-tech firm with patents, especially when they are very young or when they just finish initial IPOs. Investors or managers in this sector gamble on the new technology adopted by young firms.

\(^1\)In this article, managers, representative of investors, do the learning. Agency issues are not considered.

\(^2\)Note that Veronesi and Pastor (2003) provide empirical evidence that median firm market to book values decrease in the first several years of a firm’s life. Current book value over its next period ratio over years remains relatively stable as found in Veronesi and Pastor (2006). Despite the fact that the book values also tend to decline during this period, it does so smoothly. Although I model market equities rather than market to book value, following the same approach of Veronesi and Pastor (2006), my model can normalize the stock value by cash flow to get market to cash flow ratio which is comparable to market to book value. Hence, the majority of initial drop in stock value is attributed to declines in market value.

\(^3\)My model assumes risk neutral investors/firm managers. In a model with both risk aversion and firm growth, the discounting and consequent implications for firm growth could vary across firm age to the extent that risk premiums also vary across firm age.
Over the first several years of the median firms’ life, investors learn that its new technology will not lead it to be the next technology grant and trim their assessment of firm value accordingly. My model captures this effect and its implication for firm leverage. Gradually, as these firms age and their future prospects become more predictable the normal dynamics of stable growth can be expected to replace the initial decline in firm value due to uncertainty resolution that is captured succinctly by my model.

It is worth noting that my model does not imply a fall in value for all young firms. Rather, it explains the fall in value for median young firm, which turn out not to be the tremendous success that investors had hoped for. For these typical firms, as well as the low performing firms, the resolution of uncertainty disappoints investors by reducing the upside potential that this firm will might have become a great investment. Of course, some firms do become huge investments and in their case, the resolution of uncertainty can lead to very large gains in firm values. Indeed, it is exactly these out-signed returns in a few very lucky cases that induces investors to place bets on all new firms, thus driving up their initial firm price and driving down their initial firm leverage.

I model the leverage of typical median firms. In order to provide a tractable and interpretable model, I follow the same principle of parsimony as suggested by Veronesi and Pastor (2009). In my framework, homogeneous firms are considered instead of a large number of heterogeneous firms (see Morellec, Nikolov, Schurhoff (2013)). In this case, a single representative agent learns about parameter uncertainty from one single representative firm or its cash flow instead of different means with respect to different firms in different industries. This results in both a parsimonious model and a tractable analytic solution. The advantages of such an approach are explained at length in the survey by Veronesi and Pastor (2009).

Empirically, summary statistics of sample data indicates that mean leverage ratio increases over firm age. Estimates of my dynamic panel data model also suggest that there is a positive relation between firm leverage and firm age, indicating that firm leverage ratio is increasing over firm age.

This motivates an in-depth empirical analysis on the relationship between leverage, firm age, and cash flow uncertainty. I employ a panel data from 1962 to 2007 to estimate the relationship between leverage ratio and mean cash flow uncertainty, after the inclusion of standard control variates, including profitability, market to book ratio, tangibility and cash flow volatility. Two of the controls, profitability and market to book ratio are found to be highly endogenously. I therefore estimate a dynamic panel data model in the first differences using lagged differences as instruments, following the approach of Arellano and Bond (1991) and Blundell and Bond (1998). Consistent with my model prediction, the results show that leverage rates increase as firms age and uncertainty on the mean log cash flow resolves.

My paper is related to Pastor and Veronesi (2003) which propose a parsimonious model to
evaluate the market value of young firms' stocks. However, their model only allows for equity financing and is thus silent on the implications for firm leverage. I provide a non-trivial extension of their model to characterize the effect of uncertainty of mean log cash flow on leverage ratios through its impact on market equity.

This paper is also related to Welch (2004), whose empirical finding suggests that stock returns solely determine cross sectional changes in leverage ratios and that all other firm specific factors, even if they have some explanatory powers on leverage ratios, will only propagate their influences through stock returns. The paper concludes that leverage ratio valuation is made mainly in response to the changes in firms’ equity value. Similarly, in my model, uncertainty affects leverage through its effect on firm equity value.

My paper is also related to Lemmon, Roberts and Zender (2008), Graham, Leary, Roberts (2014) although learning about expected cash flow is not taken into account when explaining swings in leverage ratios in their article. However, the empirical analysis of my paper uses their identified firm specific determinants as control variates and partly validates some of their empirical results.

The paper is divided into five parts. Section two describes the model and assumptions. Section three provides the model solution and comparative statistics without learning. In Section four the leverage ratio is evaluated with learning and the impact of firm age on uncertainty, firm value and leverage is derived. An empirical analysis is provided in Section five to show the learning effect on leverage ratios. A conclusion follows in Section six.

3.2 Assumptions

This section provides a detailed analysis of firms value and leverage ratios without learning. There are no transaction costs involved in the business operation throughout the paper. The representative firm live for $T$ periods, where $T$ is finite and exogenous. In particular, the firm’s cash flow is assumed not to hit default or restructuring boundaries. For each period, the firm issues a constant coupon $c$ for raising money and a flat term structure with interest rate $r$ is assumed. Investors and borrowers finance their projects with this interest rate. Net operation incomes are used to reinvest in the firm operation and pay for taxes. The firm’s log cash flow follows a Brownian motion with an unobserved mean. There are no default, restructuring or bankruptcy costs.\footnote{Following Pastor and Veronesi (2003), a young firm’s cash flow will continue for a relatively long period so that expected cash flows can be used to repay debts without a probability of default. Also, market structure in this paper is not in an equilibrium. Infrequent balancing of financial structure due to transaction costs suggested by Strebulaev (2007) and changes in capital structure mainly attributed to variations in stock returns Welch (2004) motivates this assumption.}

Cash flow, defined here as the instantaneous earnings before income and tax or cash flows
is given by \( f(X_t) = \delta X_t \), where \( X_t \) is the firm’s net cash flow, and \( \delta \) denotes managerial entrenchment.\(^5\)

The firms’ log cash flow follows the Brownian motion process

\[
d\log(X_t) = \varphi (\bar{u} - \log(X_t))dt + \sigma_{x,1}dZ_1 + \sigma_{x,2}dZ_2,
\]

\( \sigma_{x,1} \) is the standard variance of log cash flow driven by systematic shocks \( dZ_1 \), and \( \sigma_{x,2} \) is the standard deviation of log cash flow driven by idiosyncratic shocks \( dZ_2 \). \( \varphi \) is the parameter controlling the mean reversion speed. In this set up, managers, acting in the interests of stockholders, cannot observe the mean of log cash flow \( \bar{u} \) and can only learn through past its realized counterparts. That is, \( \bar{u} \) is assumed to be priorly distributed as \( N(\hat{u}_0, \hat{o}) \) and \( x_t = \log(X_t) \) will be used throughout the article.

The firms’ total value at time \( t \) is given by

\[
V_t = \left( E_t \int_t^T \pi_w(1 - \tau)(\delta X_w - c)/\pi_t dw + \pi_T(1 - \tau)(\delta X_T - c)/\pi_t \right) + \left( E_t \int_t^T \pi_w(1 - \tau_t)/\pi_t cdt + \pi_T(1 - \tau_t)/\pi_t c \right)
\]

conditional on the information set \( F_t : \{H_t = (v_t, \log(X_t)) : 0 \leq \omega \leq t\} \), with \( v_t = \log X_t \pi_t \).

The first term on (43) is the equity value or the value stemming from net operating income, while the second term of (43) is the value at time \( T \). The third term of (43) denotes the firm’s debt level and the fourth term is the terminal value of (43) the firm’s debt at time \( T \). \( \pi_t \) is the stochastic discount factor. \( \tau \) and \( \tau_t \) refers to corporate income tax and personal income tax respectively.

The stochastic discount factor follows the log normal Brownian process:

\[
d\pi_t = -\pi_t r dt - \pi_t \sigma_{\pi,1} dZ_1,
\]

in which the dynamics of \( \pi_t \) are driven by the systematic shock. In this set-up, \( \sigma_{x,1}, \sigma_{x,2}, \sigma_{\pi,1}, \)
\( r, \tau, \tau_t, c \) are all scalars, which do not vary with time.

There are two features distinguishing my model from the previous literature on capital finance. The first different feature is the assumption of parameter uncertainty due to mean log cash flow. This is contrary to known mean cash flow in most continuous time capital structure literature. In my model, log cash flow is correlated with stochastic discount factor which also distinguishes from the asset pricing model of Pastor and Veronesi (2003). The second feature from their model is that the conditional distribution for stochastic discount factor needs to be derived and also debt value is added to my model.

3.3 No Uncertainty about Cash Flow

As a basis of comparison, this section provides an analysis of leverage ratios without learning, assuming mean log cash flow is known. Closed form solutions for equity value, debt value and leverage ratio are derived. Comparative statics are performed to assess the impact of mean log cash flow, current cash flow, systematic and idiosyncratic shocks to log cash flow, stochastic discount factor, and the tax rate on equity value, the debt value and the leverage ratio respectively.

**Proposition 1.** Firm Value $V_t$ takes the following form.

$$V_t = E_t + D_t$$

where $\varepsilon = T - t$ denotes the time to maturity. The leverage ratio at time $t$ is given by

$$\text{Leverage}_t = \frac{D_t}{E_t + D_t}$$

The equity value is given by

$$E_t = \int_0^\varepsilon \delta(1 - \tau)A_1(\bar{u}, x_t, k)dk - \int_0^\varepsilon c(1 - \tau)A_2(k)dk$$

and $A_1$ and $A_2$ are exponential functions given by

$$A_1 = \exp\{x_t - (r + \frac{1}{2}\sigma_{\pi}\sigma'_{\pi})\varepsilon + (1 - e^{-\psi\varepsilon})(\bar{u} - x_t) + (\sigma_{x,1}^2 + \sigma_{x,2}^2 + 2\sigma_{\pi,1}\sigma_{x,2})\varepsilon$$

$$+ \frac{\sigma_{x,1}^2 - \sigma_{x,2}^2 + 2\sigma_{x,2}^2e^{-\psi\varepsilon} + 2e^{-\psi\varepsilon}(\sigma_{x,1}^2 + \sigma_{x,2}^2)}{2\psi}\}$$

$$A_2 = \exp\{-r + \frac{1}{2}\sigma_{\pi}\sigma'_{\pi})\varepsilon + (\sigma_{x,1}^2 + \sigma_{x,2}^2 + 2\sigma_{\pi,1}\sigma_{x,2})\varepsilon$$

$$+ \frac{3\sigma_{x,2}^2(e^{-\psi\varepsilon} - 1) - 2(1 - e^{-\psi\varepsilon})(\sigma_{x,1}^2 + \sigma_{x,1}\sigma_{\pi,1})}{2\psi}\}$$

The closed form solution for firm value and leverage ratio given in Proposition 1 above characterizes a convex relationship between mean log cash flow and firm value. Interestingly, the convex relation also exists between the leverage ratio and mean log cash flow. Mean log cash flow mainly affects the leverage ratio via its effect on market equity value, while there is no effect of it on debt value over age.

**Corollary 1.** Mean log cash flow $\bar{u}$ has an increasing effect on firms’ equity value. Equity value rises in current log cash flow $\log(X_t)$. Interest rate and $\sigma_{\pi}\sigma'_{\pi}$’s effect on equity value is
The more managerial entrenchment, the higher equity value. Equity value is decreasing in the coupon rate. The effect of mean reversion speed on equity value cannot be determined. The cash flow shocks \( \sigma_{x,1}^2, \sigma_{x,2}^2 \) have a positive effect on firm equity value. Firm market value declines over the corporate tax rate.

When the expected cash flow and current level of cash flow increases, the firm’s stock price will rise and thus the valuation of the firm’s market value will increase. This confirms the results of most empirical finding about the determinants of capital structure.\(^6\) The interest rate’s effect is ambiguous. This is because the marginal effect of interest rate on equity value is two fold: on the one hand, its effect on net cash flow is positive, on the other hand, its effect on the debt is negative. Therefore, it is difficult to judge which effect will dominate. If managers hold a greater share of firm’s equity or become more entrenched, they have more incentive to maximize the firm’s value. The larger the positive cash flow shocks, the higher the market value.

**Corollary 2.** Firm equity value is convex in mean log cash flow \( \bar{u} \).

A mathematical proof of convexity is provided in the Appendix. Utilizing the same parameters of Pastor and Veronesi (2003)\(^7\), I illustrate the model implied convex relation between market equity value and mean log cash flow as well as between the leverage ratio and mean log cash flow. In Figure 10 and 11. This confirms their finding that the larger the mean log cash flow or cash flow growth rate, the more pronounced the convexity. One interesting feature is the role of mean reversion speed in affecting convexity.

The debt value mainly includes the tax advantage brought from the net cash flow from business operation before known \( T \).

**Corollary 3.** Debt value rises with the coupon rate. The interest rate and the stochastic discount factor both have a negative effect on debt value. The effect of cash flow shocks on debt is ambiguous. The tax rate has a negative effect on the firm’s debt value.

If the interest rate goes up, this means that firms face a larger external financing constraint, and so debt available to the firm will decrease. If a higher tax rate is imposed on the firms’ earnings then the tax advantage available for them will be diminished. \( \bar{u} \) impacts the leverage ratio mainly through equity value. Because cash flow uncertainty has no effect on debt dynamics, the leverage ratio is adjusted by the changes in firms’ equity value, which is caused by future cash flow uncertainty.

**Corollary 4.**

\(^6\)See Roberts (2008).

\(^7\)Similar to their model implied prediction of convexity, my model also implies convexity between market value and mean log cash flow as well as that among leverage ratio and mean log cash flow. However, due to newly added debt value, my model implied prediction of the effect of interest rate, systematic shocks of log cash flow and stochastic discount factor on market value is different.
Figure 10: Market value is convex in mean cash flow. The vertical axis shows market value. The horizontal axis shows mean log cash flow.

1. **Without learning, the leverage ratio is decreasing in** $\bar{u}$ **and current log cash flow level** $x_t$ **and** $\delta$.

2. **The effect of the interest rate, stochastic discount factor, cash flow shocks and mean reversion speed are ambiguous.**

Note that the mean log cash flow only affects the leverage ratio through its effect on equity value, and according to corollary one, equity value is decreasing in mean cash flow. Thus the leverage ratio is decreasing with $\bar{u}$. The same holds true for $x_t$ and $\delta$. The model implied leverage ratio is plotted in Figure 12 using the same parameters as before. Interestingly, when firms are very young, their leverage ratio appears to be high (about 0.27). This figure is similar to the median values calculated from the data set of CRSP (see Section 5). As cash flow rate goes up, firms become less levered. One explanation is that expected net cash flow increase firms’ retained earning which can be reinvested to positive cash flow projects instead of issuing more debt externally to satisfy the financing needs. This again confirms many of the empirical findings.\(^8\) Even if the leverage ratio diminishes, it does so very slowly towards some steady state values. It is noteworthy that the leverage ratio fluctuates between 0.25 to 0.16 before firms mature at the age of 15. The slow mean reversion speed of the leverage ratio has been observed by Lemmon, Roberts and Zender (2008) and Dufrense and Goldstein (2001). My results are consistent with these empirical results.

\(^8\)Refer to section 5.
3.4 Learning about Cash Flow

In this section, I incorporate Bayesian learning to ask how leverage ratios change with firm age as the uncertainty resolves. Mean log cash flow $\bar{u}$ is no longer assumed to be known. Instead managers and investors have a common prior on mean log cash flow and use Bayesian rules to update their posterior belief based on realized cash flow. They next assess the influences exerted by uncertainty of mean cash flow upon firm equity value. Since firm debt value is not affected by cash flow uncertainty, they then revise the leverage ratios accordingly. The uncertainty, as discussed before, originates from the convexity between positive effect of the cash flow growth rate on firm equity value.

**Lemma 1.** Managers/investors update their posterior mean updates following the rule,

$$d\tilde{u}_t = \frac{\tilde{\sigma}_t^2}{\sigma_{x,2}} d\tilde{Z}_{2,t}$$

(50)

where $\tilde{Z}_{2,t}$ is the expectation error.$^9$

This equation provides the updating formula for the posterior mean log cash flow as a function of posterior variance of mean log cash flow. The smaller the posterior variance of mean log cash flow, the smaller the mean log cash flow.

$^9$Its definition can be found in the section of Proof of Proposition 2 of the Appendix.
After tedious calculations, the posterior variance $\hat{\sigma}_t^2$ of mean log cash flow is equal to

$$\hat{\sigma}_t^2 = 1/(\frac{1}{\hat{\sigma}_0^2} + \frac{\phi^2}{\sigma_x^2} t) \quad (51)$$

The posterior variance of mean log cash flow is a decreasing function of firm age. As firm grows mature, managers and investors learn that future log cash flow will approach its mean closely, thereby reducing the volatility of mean log cash flow.

**Proposition 2.** With unknown $\bar{u}$, firm value is expressed as the following:

$$V_t = E_t\{E_t\{\int_t^T \pi_w(1-\tau)(\delta X_w - c)/\pi_tdw + [\pi_T(1-\tau)(\delta X_T - c)/\pi_t]$$

$$+ \int_t^T \pi_w(1-\tau_i)/\pi_t cdt + \pi_T(1-\tau_i)/\pi_t c | \bar{u}\}\} \quad (52)$$

then, the solution is given by

$$V_t = \int_0^\varepsilon \delta(1-\tau)\tilde{A}_1(\bar{u}_t, x_t, k)dk + \delta(1-\tau)\tilde{A}_1(\bar{u}_t, x_t, \varepsilon)$$

$$+ \int_0^\varepsilon \delta(\tau-\tau_i)cA_2(k)dk + (\tau-\tau_i)cA_2(\varepsilon), \quad (53)$$

where $\varepsilon$ denotes the time to maturity, $A_2$ remains the same as equation (46). $\tilde{A}_1 = A_1 * e^{\psi \varepsilon} \tilde{\sigma}_t^2$, where $A_1$ is defined in equation (46).

Equation (51) shows how posterior belief on cash flow uncertainty affects market values. We can see three parameters in this formula: mean reversion speed $\phi$, prior variance of mean cash flow $\hat{\sigma}_0^2$ and cash flow shock $\sigma_x^2$. The posterior variance falls with mean reversion speed and firm age, but has a positive correlation with both the posterior variance and prior variance of mean cash flow, holding other parameters fixed. Using the same parameters as before, we obtain Figure 12, which indeed verifies the negative relation between cash flow uncertainty and firm age.

Staring from a young age, we see that there is a gradual decline of the variance of cash flow over age. The gradually diminishing variance serves a key factor in explaining the role of learning. To get a deeper understanding of this requires investigating how equity value varies with the posterior variance.

**Corollary 5.** The leverage ratio decreases with the posterior variance.

Recall that firm value is increasing with the posterior variance. Therefore, firm value will go up one-by-one with the rise in the posterior variance, while leverage ratio will decrease since debt value is unaffected by it. Figure 13 validates the argument.
Figure 12: Posterior volatility declines over age. The vertical axis shows posterior volatility. The horizontal axis shows age.

Figure 13: Cash flow volatility and leverage ratios. The vertical axis shows leverage ratio. The horizontal axis posterior variance.
Corollary 6. Firm leverage increases with firm age due to learning.

Figures 14 and 15 provide the contrast of the relationship between the leverage ratio and firm age without learning and with learning. Once learning is present, it is demonstrated clearly that as the firm ages, the volatility of cash flow gradually declines and thus the firms equity value falls. With the resolution of uncertainty, firm value decreases and then firm leverage increases. This is consistent with and potentially explains previous empirical findings. Dufresne and Goldstein (2001) proposes that the leverage ratio is mean-reverting. Lemmon, Roberts, Zender (2008) points out that serial correlation of the leverage ratio is a major feature of leverage ratio and also if a firm initially has a low leverage ratio, then the leverage ratio gradually increases until it reaches a steady level.

In the case of learning, firm equity drops more sharply from a high value of 40. It then gradually flattens towards a number of 5. This results in an initial low debt ratio of 0.055. It gradually rises to 0.12, a level achieved when the firm grows mature. In the case without learning (see figure 15), firm equity value drops steadily with age (the magnitude of fall in equity value appears very small) and debt ratio becomes gradually higher in a very slow pace. Again, the debt ratio tends to converge to a steady level when reaching mature age. The pronounced difference is that due to learning, leverage ratio undergo an jump especially when firms are at a young age. In Figure 15, different values of $\varphi$ and $\delta$ have been employed to see the effects of different parameter values on debt ratio with learning. $\varphi$ is controlling for the mean reversion speed. The larger the value of $\varphi$, the quicker cash flow reverts to its mean value and the faster managers learn. While $\delta$ is
Figure 15: Leverage ratio varies over age with learning. The vertical axis shows leverage ratio. The horizontal axis shows age.

A managerial entrenchment parameter, it controls how shocks to managerial security affect cash flows and also leverage ratio. A higher value of $\delta$ leads to a higher value of cash flows, holding other parameters fixed. Consequently when accounting for the learning effect, the initial value of the leverage ratio will decline more.

### 3.5 Empirical Analysis

The previous sections provided the effect of learning on leverage ratios, finding that managers gradually update their evaluations of leverage ratio upwards as firms age. The model implied simulation results demonstrated the relationship among the posterior variance of mean log cash flow, firm equity value and firm debt ratios. Using panel data methods, I now examine their empirical relation more closely.

The data used in this paper comes from Arthur Korteweg’s website. Unbalanced panel data is used in the empirical section spanning from 1962 to 2007. Firm specific variables include market leverage ratios, market to book value ratios, tangibility, profitability, cash flow volatility and one dummy variable indicating whether the economy is in a boom or recession. All the firm specific variables are winsorized at the 0.99 and 0.01 level to ensure that there is no effect of

---

10 Data in his website (http://www-bcf.usc.edu/~korteweg/) does not include firm age. I thank Bo Han Li at University of Western Ontario for providing me data on firm age.

11 Variables definition are provided in the Appendix part one in great detail. The variables’ definitions include also plain firm age, as well as lagged variables.
Table 12: Summary Statistics

<table>
<thead>
<tr>
<th>Firm Age</th>
<th>Mean Leverage Ratio</th>
<th>Mean M/B Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.029</td>
<td>7.819</td>
</tr>
<tr>
<td>2</td>
<td>0.023</td>
<td>8.85</td>
</tr>
<tr>
<td>3</td>
<td>0.018</td>
<td>6.367</td>
</tr>
<tr>
<td>4</td>
<td>0.221</td>
<td>3.207</td>
</tr>
<tr>
<td>5</td>
<td>0.231</td>
<td>3.847</td>
</tr>
<tr>
<td>6</td>
<td>0.298</td>
<td>3.286</td>
</tr>
</tbody>
</table>

outliers on the estimates. The data is sampled from 232 firms and comes from 22 four digit SIC industries.\(^\text{12}\) All the sampled firms have market value more than 10 million dollars. Financial sector data are not included.

The model of the previous section uses learning to explain the relationship between firm age and the leverage ratio. The leverage ratio increases with the reduction of uncertainty, which is measured by the posterior variance in equation (51). Following Pastor and Veronesi (2003), I employ one over one plus plain firm age as a proxy for the posterior variance in (51). Plain firm age is defined as number of years since the firm’s first appearance in the Compustat data base. Summary statistics\(^\text{13}\) are given in Table 12,\(^\text{14}\) which provides the mean leverage and market book ratios of all firms of given age or 1962 cohort. Interestingly, we can see that the mean leverage ratio starts very low for young firm at about 0.02 and then experiences a big jump in the third year, after which it tends to stabilize at 0.27 as firms age. The results are consistent with those of Pastor and Veronesi (2003). The opposite pattern can be found for the time varying changes in the average market to book ratio, which tends to decrease over time.

I first estimate a static panel data model with serially correlated errors to examine the effect of the posterior variance on the leverage ratio after the inclusion of standard control variates. The model is given by:

\[
\text{LeverageRatio}_{i,s} = \beta_0 + \beta_1 \text{PosteriorVariance}_{i,s} + \beta_2 X_{i,s} + a_i + \eta_t + \varepsilon_{i,s}
\]  

(54)

where I define the proxy for the posterior variance as in (51).

\[
\text{PosteriorVariance}_{i,s} = \frac{1}{1 + \text{FirmAge}_{i,s}}
\]  

(55)

\(^{12}\)This is the same data as Arthur Korteweg (2010).

\(^{13}\)Mean leverage ratio is calculated conditional on age and 1962 cohort.

\(^{14}\)The table one documents summary statistics for M/B and leverage ratio for groups of firms based on the same age. Age definition has been provided in the appendix. The row is firm age. For each age, mean statistics are computed across firms in the 1962 cohort.
The residual $\varepsilon_{i,s}$ is given by

$$
\varepsilon_{i,s} = \rho_1 \varepsilon_{i,s-1} + \omega_{i,s}, \quad \text{for} \ |\rho_1| < 1
$$

(56)

The dependent variable, $\text{LeverageRatio}_{i,s}$, is the leverage ratio at time $s$ for company $i$, $\omega_{i,s}$ is serially and cross-sectionally uncorrelated. The control variate $X_{i,s}$ includes firm specific characteristics including profitability, firm size, tangibility, market to book value, cash flow volatility and macroeconomic recession indicators, $a_i$ is the firm fixed effects and $\eta_s$ is the year effects. The error $\varepsilon_{i,s}$ is serially correlated and assumed to follow an $AR(1)$ process. I compare estimates from several different panel data methods.

Fixed effect effects without instruments have been implemented first in Table 13 to explore the relation between the leverage ratio and the posterior variance. Column 2 of Table 13 reports the regression results of the panel data model in levels without the posterior variance and column 3 presents the same model with the posterior variance. Interestingly, the estimated coefficient on the posterior variance is 1.23, whose sign and magnitude are contradictory to the model implied simulation result, as well as the summary result in Table 12.

A likely explanation for this puzzling result is endogeneity bias due to both simutaneity and omitted variable bias. These biases have been discussed at length in the corporate finance literature, see (Arellano (2003) and Hsiao (2003)). If firms borrow to invest in profitable projects, this may increase both profitability and the market book ratio and both regressors may therefore

<table>
<thead>
<tr>
<th>Variable</th>
<th>Panel Data Model In Level</th>
<th>Panel Data Model with Posterior Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posterior Variance</td>
<td>1.23</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Profitability</td>
<td>-0.19</td>
<td>-0.19</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>M/B</td>
<td>-0.03</td>
<td>-0.32</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Tangibility</td>
<td>0.01</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

This table presents the regression results of a static panel data model without instruments. All the estimates are significant at the 5 percent level. Standard error is provided in the parenthesis and rounded to the second decimal place. All the independent variables are trimmed at the 0.99 and 0.1 level. Variables definitions are in the Appendix.
be correlated with the innovation to the firm leverage. Indeed Pastor and Veronesi (2003) show evidence that the market to book value is negatively determined by the leverage ratio. As a result, simultaneity exists between these two variables. Omitted variables present another source of endogeneity bias since profitability is also determined by the dividend policy and the marginal tax rate.\footnote{Due to data availability, tax rate and dividend payout variable are not included in the model specification and thus profitability is presumably to be endogenous.} Lemmon, Roberts and Zennder (2008) and Roberts and Whited (2013) provide detailed documentation of these two issues concerning the empirical specification of corporate finance regression models. In order to address the endogeneity of profitability and the market value to book value ratio, I next estimate a dynamic panel data model with a one period lagged dependent variable and instrument these two lagged regressors, namely market book ratios and profitability, using the (Blundell and Bond (1998)) and (Arellano and Bond (1991)) approaches in which the lagged variables are differenced to remove the fixed effect and the lag difference are used as instruments. First the dynamic panel data model is given by

$$
LeverageRatio_{i,s} = \beta_0 + \beta_1 LeverageRatio_{i,s-1} + \beta_2 PosteriorVariance_{i,s} + \beta_3 X_{i,s-1} + a_i + \eta_s + \varepsilon_{i,s} \tag{57}
$$

where one period lagged variable of $X_{i,s-1}$ is included in the model specification\footnote{The empirical specification follows Lemmon, Roberts and Zender (2008) closely, the choice of lags and instruments is based on the paper to ensure the reasonable specification.}. Next the first difference of the equation generates

$$
\Delta LeverageRatio_{i,s} = \beta_1 \Delta LeverageRatio_{i,s-1} + \beta_2 \Delta PosteriorVariance_{i,s} + \beta_3 \Delta X_{i,s-1} + \Delta a_i + \Delta \eta_s + \Delta \varepsilon_{i,s} \tag{58}
$$

where all the variables of (57) are expressed in the first differences and $\Delta X_{i,s-2}$, the first difference of the one period lagged variables of $X_{i,s-1}$ of (57) are chosen as instruments for profitability and the market to book ratio.

The Hausman test has been performed first to test endogeneity for market book ratio and profitability. The statistics suggest that the null hypothesis of no endogeneity has been rejected. Thus profitability and market to book ratio are confirmed to be endogenous. I next test for weak instruments using the approach of Stock and Yogo (2005). Test results available upon request indicate that the null hypothesis of weak instrument has been rejected. Thus, the lagged first difference in these variables are strong instruments. The results are shown in Table 14. After controlling for the endogeneity, the posterior variance now enters negatively and significantly. This contrasts with the positive sign in Table 13 before the instrument. Apparently in Table 14
This table presents the regression results of dynamic panel data model with instruments. All the estimates are significant at the 5 percent level. Standard error is provided in the parenthesis and rounded to the second decimal place. GMM One denotes GMM using Arellano-Bond estimator while GMM Two denotes GMM using Arellano-Blundell-Bond estimator. The difference is that the latter is a system estimator that uses moment conditions in which lagged differences of the dependent variable (leverage ratio), the predetermined variables, and the endogenous variables (profitability and M/B ratio) are used as instruments for the level equation in addition to the moment conditions of lagged levels as instruments for the differenced equation. All the independent variables are trimmed at the 0.99 and 0.1 level. Variables definitions are in the Appendix.
the magnitude of previous effect of learning become more pronounced at −0.78 and −0.42. Since the posterior variance falls with firm age, this indicates that firms’ leverage ratios are increasing over age. The large magnitude suggests that the learning effect is especially astute for younger firms. This validates the increasing mean leverage ratio over age in Table 12 and the counterpart of Pastor and Veronesi (2003), suggesting that the learning effect is prominent especially when firms are younger. In summary, due to the presence of endogenous variables and simultaneity, the first empirical specification without consideration of these two issues generates a biased estimate of the effect of the posterior variance on firm leverage. After performing endogeniety and weak instrument test, valid instruments are chosen for profitability and the market to book ratio, the two endogenous variables. The resulting empirical regression result is consistent with the model implied prediction that the leverage ratio is increasing over firm age.

3.6 Conclusion

This paper proposes a simple dynamic model to explore the relationship among the leverage ratio, market value and mean log cash flow. I find that leverage ratios grow over firm age due to learning. Initially when uncertainty about expected cash flow is highly uncertain, firm market value is high, and this in turn implies a low leverage ratio, since there is no effect of mean cash flow uncertainty on debt values. However, as cash flow uncertainty decreases with firm age, market values are found to decline slowly. Consequently the leverage ratio declines gradually to its long run steady state level. This motivates a further empirical analysis of the relationship between leverage and firm age using panel data. After addressing the endogeneity of profitability and the market to book ratios, I find empirical regression results that are consistent with the model implied prediction that the leverage ratio is increasing over firm age.

There are still several directions for future research. In ongoing work, I generalize the model so that managers can choose the optimal leverage ratio in the presence of learning. It may also be interesting to study the behaviour of the debt ratio in an equilibrium environment in which the stochastic discount factor can be endogenized. Learning from peer firms’ leverage ratios can also become a potential topic, as recent papers have found that managers are motivated to learn from peer firms’ leverage ratio when making decisions.

3.7 Appendix

Variables Construction:
This section documents the definition of all the variables incorporated for the empirical specification. The Compustat data sample is used throughout the article. All numbers denote the annual Compustat item number.
Firm Size = \log \text{(book assets)}, where assets are deflated by GDP deflator.

Profitability = \text{operating income before depreciation (13) / book assets.}

Cash Flow Volatility = \text{the standard deviation of historical operating income.}

Market leverage = \text{total debt / (total debt + market equity)}

Market to Book = \text{(market equity + total debt + preferred stock liquidating value (10) - deferred taxes and investment tax credits (35)) / book assets.}

Tangibility = \text{net PPE / book assets}

Firm Age = \text{the number of years since the first time for the firm to appear in the CRSP database or WRDs database.}

Posterior Variance of Cash Flow = \frac{1}{1 + \text{Firm Age}}.

Economy Boom Indicator Dummy is defined according to NBER boom index.

Proofs
The Appendix provides a detailed derivation of Proposition 1 and 2. Because of tedious algebra, proof of all the corollaries will be available upon request.

The cash flow is given by

\[ f(X_t) = \delta X_t \]  

(59)

where \( X_t \) is the firm cash flow, \( \delta \) denotes managerial entrenchment and \( f(X_t) \) is considered to be firms’ net profits.

Firms’ log cash flow is supposed to follow the Brownian motion process below:

\[ d\log(X_t) = \varphi(\bar{u} - \log(X_t))dt + \sigma_{x,1}dZ_1 + \sigma_{x,2}dZ_2 \]  

(60)

Where \( \sigma_{x,1} \) and \( \sigma_{x,2} \) are constants. In this set up, managers, who are in the interests of stockholders, can not observe the mean of log cash flow and can only learn through past realized counterparts. That is, \( \bar{u} \) is assumed to be priorly distributed as \( (\hat{\sigma}_u, \hat{\sigma}_o) \).

The firm value is given by

\[ V_t = E_t \int_t^T \pi_w(1 - \tau)(\delta X_w - c)/\pi_t dw + [\pi_T(1 - \tau)(\delta X_T - c)/\pi_T] + \int_t^T \pi_w(1 - \tau_t)/\pi_t c dt + \int_t^T \pi_T(1 - \tau_t)/\pi_t c \]

conditional on the information set \( F_t : \{H_t = (v_t, \log(X_t)) : 0 \leq \omega \leq t\} \)
The first term and second term of the right hand side is the equity value or values stemming from cash flow while the third and fourth value denote firms’ debt level. \( \pi_t \) is the stochastic discount factor.

The stochastic discount factor following log-normal Brownian process:

\[
d\pi_t = -\pi_t r dt - \pi_t \sigma_{\pi} dZ_t
\]  

(62)

The dynamics of \( \pi_t \) is driven the same systematic shock and correlated with log cash flow. \( V_t \) can be first evaluated by Fubini theorem, by which we can move expectation operator inside the integration. Therefore, we need to evaluate \( E(\pi_t X_t) \).

\textbf{Proof of Proposition 1.} let \( v_t = log(\pi_t X_t) \), then \( v_t = log\pi_t + logX_t \) and also \( \pi_t X_t = e^{v_t} \)

\[
dv_t = dlog\pi_t + dlogX_t
\]

(63)

We can substitute the process of \( dlog\pi_t \), after we use \( \text{ito} \) lemma with respect to \( dlog\pi_t \) and \( dlogX_t \) defined above, and then we have

\[
dv_t = (-r - \frac{1}{2} \sigma_{\pi} \sigma_\nu^2) dt + \sigma_{\pi} dZ_1 + \varphi(\bar{u} - log(X_t)) dt + \sigma_{x,1} dZ_1 + \sigma_{x,2} dZ_2
\]

(64)

\[
dv_t = (-r - \frac{1}{2} \sigma_{\pi} \sigma_\nu^2 + \varphi(\bar{u} - log(X_t)) dt + (\sigma_{\pi,1} + \sigma_{x,1}) dZ_1 + \sigma_{x,2} dZ_2
\]

(65)

where \( \sigma_{\pi} = \begin{bmatrix} \sigma_{\pi,1} & \sigma_{\pi,2}^0 \end{bmatrix} \)

Let \( F_t = (v_t, logX_t)' \), \( (v_t, logX_t)' \) be a vector of state variables, its matrix representation is given by:

\[
d \begin{pmatrix} v_t \\ logX_t \end{pmatrix} = \begin{pmatrix} \varphi \bar{u} - r - \frac{1}{2} \sigma_{\pi} \sigma_\nu^2 \n \varphi u \n 0 - \varphi \end{pmatrix} \begin{pmatrix} v_t \\ logX_t \end{pmatrix} dt + \begin{pmatrix} \sigma_{\pi,1} + \sigma_{x,1} & \sigma_{x,2} \\ \sigma_{x,1} & \sigma_{x,2} \end{pmatrix} \begin{pmatrix} dZ_1 \\ dZ_2 \end{pmatrix}
\]

(66)

In this vector form, \( A = \begin{pmatrix} \varphi \bar{u} - r - \frac{1}{2} \sigma_{\pi} \sigma_\nu^2 \\ \varphi u \n 0 - \varphi \end{pmatrix}, B = \begin{pmatrix} 0 & -\varphi \\ 0 & -\varphi \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{\pi,1} + \sigma_{x,1} & \sigma_{x,2} \\ \sigma_{x,1} & \sigma_{x,2} \end{pmatrix} \)

This is a standard multi-dimensional linear process (Duffie,1996) whose solution is a known closed form solution \( F_T | F_0 \sim N(\mu(F_0,T), \Sigma_F(T)) \), where

\[
\mu(F_0,T) = \psi(T)F_0 + \int_0^T \psi(T - t)Adt
\]

\[
\Sigma_F(T) = \int_0^T \psi(T - t)\Sigma \psi(T - t)'dt
\]

\[
\psi(T) = w exp(\Lambda * T)w^{-1}
\]

The normality of \( F_T \) implies that \( v_T = e_1 F_T \) with \( e_1 = (1,0) \), \( v_T \) is also normally distributed \( V_T | F_0 \sim N(e_1 \mu(F_0,T), e_1 \Sigma_F(T)e_1') \), then we can use the properties for log-normal random variables \( E(\pi X_t) = E(e^{v_T}) = E(exp(e_1 F_T)) = exp(e_1 \mu(F_0,T) + \frac{1}{2} e_1 \Sigma_F(T)e_1') \), First, we can evaluate \( \mu(F_0,T) \),
First, let’s define \( \psi(T-t) = \begin{bmatrix} 1 & -1 + e^{-\varphi(T-t)} \\ 0 & e^{-\varphi(T-t)} \end{bmatrix} \), \( \psi'(T-t) = \begin{bmatrix} 1 & 0 \\ -1 + e^{-\varphi(T-t)} & e^{-\varphi(T-t)} \end{bmatrix} \)

\[
\Sigma = \begin{bmatrix} \sigma_{\pi,1} + \sigma_{x,1} & \sigma_{x,2} \\ \sigma_{x,1} & \sigma_{x,2} \end{bmatrix}, \Sigma' = \begin{bmatrix} \sigma_{\pi,1} + \sigma_{x,1} & \sigma_{x,1} \\ \sigma_{x,2} & \sigma_{x,2} \end{bmatrix}
\]

\[
\mu(F_0, T) = \begin{bmatrix} \frac{v_0}{\log x_0} + \int_0^T \left[ \phi^T \left( \varphi \right) - r - \frac{1}{2} \sigma_{\pi}' \sigma_{\pi} + \varphi (1 - 1 + e^{-\varphi(T-t)}) \right] \right] dt
\]

Then, \( \Sigma F(T) = \int_0^T \psi(T-t) \Sigma' \psi(T-t)' dt \) can be evaluated, set \( K = \psi(T-t) \Sigma \)

\[
K \Sigma' = \begin{bmatrix} (\sigma_{\pi,1} + \sigma_{x,1}) + (-1 + e^{-\varphi(T-t)}) \sigma_{x,1} & (\sigma_{x,2} + (-1 + e^{-\varphi(T-t)}) \sigma_{x,2} \\ e^{-\varphi(T-t)} \sigma_{x,1} & e^{-\varphi(T-t)} \sigma_{x,2} \end{bmatrix}
\]

Solve for \( K \Sigma' \)

\[
K \Sigma' = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \text{ where } a = [(\sigma_{\pi,1} + \sigma_{x,1}) + (-1 + e^{-\varphi(T-t)}) \sigma_{x,1}] (\sigma_{\pi,1} + \sigma_{x,1}) + (\sigma_{x,2} + (-1 + e^{-\varphi(T-t)}) \sigma_{x,2},
\]

\[
b = [(\sigma_{\pi,1} + \sigma_{x,1}) + (-1 + e^{-\varphi(T-t)}) \sigma_{x,1}] * \sigma_{x,1} + (\sigma_{x,2} + (-1 + e^{-\varphi(T-t)}) \sigma_{x,2} \sigma_{x,2}
\]

\[
c = (\sigma_{\pi,1} + \sigma_{x,1}) e^{-\varphi(T-t)} + e^{-\varphi(T-t)} \sigma_{x,2} \sigma_{x,2}
\]

\[
d = \sigma_{x,1}^2 e^{-\varphi(T-t)} + \sigma_{x,2}^2 e^{-\varphi(T-t)}
\]

Multiply \( K \Sigma' \) by \( \psi'(T-t) \),

\[
\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 + e^{-\varphi(T-t)} & e^{-\varphi(T-t)} \end{bmatrix} = \begin{bmatrix} a + b(-1 + e^{-\varphi(T-t)}) & be^{-\varphi(T-t)} \\ c + de^{-\varphi(T-t)} & de^{-\varphi(T-t)} \end{bmatrix}
\]

The algebra involved here is very tedious and also the term \( \frac{1}{2} \sigma_1 \Sigma F(T) \rho \) in the equation \( E(\pi_1 x_1) = E(e^{\nu}) = E(exp(\epsilon_1 F_T)) = exp(\epsilon_1 \mu(F_0, T) + \frac{1}{2} \sigma_1 \Sigma F(T) \rho \) isolate only the first row and column element from the matrix \( \Sigma F(T) \), as a result, we just only need to compute the first row and column element \( \Sigma_{11} \) from the matrix.
\[ \Sigma_F(t) = \begin{bmatrix} a + b(-1 + e^{-\varphi(T-t)}) & be^{-\varphi(T-t)} \\ c + de^{-\varphi(T-t)} & de^{-\varphi(T-t)} \end{bmatrix} \]

\[ \Sigma_{11} = \{(\sigma_{\pi,1} + \sigma_{x,1}) + (-1 + e^{-\varphi(t-T)})\sigma_{x,1}\} (\sigma_{\pi,1} + \sigma_{x,1} + \sigma_{x,2} + (1 - e^{-\varphi(T-t)})\sigma_{x,2}) \]  
\[ = (\sigma_{\pi,1} + \sigma_{x,1})^2 + (-1 + e^{-\varphi(T-t)})\sigma_{x,1}\sigma_{x,1} \]
\[ + \sigma_{x,2}^2 + \sigma_{x,2}^2 - e^{-\varphi(T-t)}\sigma_{x,2}\sigma_{x,2} + (\sigma_{\pi,1} + \sigma_{x,1}) \sigma_{x,1} \sigma_{x,1} + (1 + e^{-\varphi(T-t)}) \sigma_{x,2} \sigma_{x,2} \]
\[ + (1 - e^{-\varphi(T-t)})^2 \sigma_{x,1}^2 + (1 - e^{-\varphi(T-t)})^2 \sigma_{x,2}^2 \]
\[ = (\sigma_{\pi,1} + \sigma_{x,1})^2 + (-1 + e^{-\varphi(T-t)})^2 \sigma_{x,1} \sigma_{x,1} + (1 - e^{-\varphi(T-t)})^2 \sigma_{x,2} \sigma_{x,2} \]
\[ + (1 - e^{-\varphi(T-t)})^2 \sigma_{x,1}^2 + (1 - e^{-\varphi(T-t)})^2 \sigma_{x,2}^2 \]
\[ = (\sigma_{\pi,1} + \sigma_{x,1})^2 + 2(1 - e^{-\varphi(T-t)})^2 \sigma_{x,1}^2 + (1 - e^{-\varphi(T-t)})^2 \sigma_{x,2}^2 \]
\[ + e^{-\varphi(T-t)} \sigma_{x,2}^2 + (1 - e^{-\varphi(T-t)})^2 \sigma_{x,2}^2 \]
\[ = (\sigma_{\pi,1} + \sigma_{x,1})^2 + e^{-\varphi(T-t)} \sigma_{x,1} \sigma_{x,1} + e^{-\varphi(T-t)} \sigma_{x,2} \sigma_{x,2} + (1 - e^{-\varphi(T-t)})^2 \sigma_{x,2}^2 \]
\[ + (1 - e^{-\varphi(T-t)})^2 \sigma_{x,2}^2 \]
\[ = \sigma_{\pi,1}^2 + \sigma_{x,1} \sigma_{x,1} + e^{-\varphi(T-t)} \sigma_{x,2} \sigma_{x,2} + (1 - e^{-\varphi(T-t)})^2 \sigma_{x,2}^2 \]

Then we take the integration from 0 to T, we have the formula

\[ \text{Var}_{11} = \int_0^T \Sigma_{11}(t) dt = \int_0^T (\sigma_{\pi,1}^2 + \sigma_{x,1}^2 + 2 \sigma_{\pi,1} \sigma_{x,1} + e^{-\varphi(T-t)} \sigma_{x,2}^2 - 2 e^{-\varphi(T-t)} \sigma_{x,2}^2 + (\sigma_{\pi,1}^2 + \sigma_{x,2}^2) e^{-2\varphi(T-t)}) \]

\[ = (\sigma_{\pi,1}^2 + \sigma_{x,1}^2 + 2 \sigma_{\pi,1} \sigma_{x,1})(T-t) + \frac{(1-e^{-\varphi(T-t)})(\sigma_{x,2}^2 - 2 e^{-\varphi(T-t)} \sigma_{x,2}^2)}{\varphi} + \frac{1}{(\sigma_{\pi,1}^2 + \sigma_{x,2}^2)(1-e^{-2\varphi T})} \]

\[ = (\sigma_{\pi,1}^2 + \sigma_{x,1}^2 + 2 \sigma_{\pi,1} \sigma_{x,1})(T-t) + \frac{2(\sigma_{x,2}^2 - 2 e^{-\varphi(T-t)} \sigma_{x,2}^2)}{\varphi} + \frac{1}{(\sigma_{\pi,1}^2 + \sigma_{x,2}^2)(1-e^{-2\varphi T})} \]

\[ = (\sigma_{\pi,1}^2 + \sigma_{x,1}^2 + 2 \sigma_{\pi,1} \sigma_{x,1})(T-t) + \frac{2(\sigma_{x,2}^2 - 2 e^{-\varphi(T-t)} \sigma_{x,2}^2)}{\varphi} + \frac{1}{(\sigma_{\pi,1}^2 + \sigma_{x,2}^2)(1-e^{-2\varphi T})} \]

In order to back out conditional variance of \( \pi_t \) conditional on the information set \( F_t \): \( \{H_t = (\nu_t, log(X_t)) : 0 \leq \nu \leq t \} \), we need to use the formula:

\[ \text{Var}_{11} + \text{Var}_{22} - 2 \text{Var}_{12} = (\sigma_{\pi,1}^2 + \sigma_{x,2}^2 + 2 \sigma_{\pi,1} \sigma_{x,1})(T-t) + \frac{(1-e^{-\varphi(T-t)})(\sigma_{x,2}^2 - 2 e^{-\varphi(T-t)} \sigma_{x,2}^2)}{\varphi} + \frac{1}{(\sigma_{\pi,1}^2 + \sigma_{x,2}^2)(1-e^{-2\varphi T})} \]

\[ + \frac{2(\sigma_{x,1} \sigma_{x,1} \sigma_{\pi,1} \varphi) + \sigma_{x,2}^2 (1-e^{-\varphi T})^2 + \sigma_{x,2}^2 (1-e^{-2\varphi T})^2 + \sigma_{x,2}^2 (1-e^{-2\varphi T})^2 + \sigma_{x,2}^2 (1-e^{-2\varphi T})^2}{\varphi} \]

\[ = (\sigma_{\pi,1}^2 + \sigma_{x,1}^2 + 2 \sigma_{\pi,1} \sigma_{x,1})(T-t) + \frac{2(\sigma_{x,2}^2 - 2 e^{-\varphi(T-t)} \sigma_{x,2}^2)}{\varphi} + \frac{1}{(\sigma_{\pi,1}^2 + \sigma_{x,2}^2)(1-e^{-2\varphi T})} \]

\[ \text{Var}(\pi_t) = (\sigma_{\pi,1}^2 + \sigma_{x,2}^2 + 2 \sigma_{\pi,1} \sigma_{x,1})(T-t) + \frac{3(\sigma_{x,2}^2 - 2 \sigma_{x,2}^2)(1-e^{-\varphi T}-1)(1-e^{-2\varphi T})}{\varphi} \]
Then without learning, the closed form solution for firm value is given by
\[ V_t = \int_0^T \delta(1-\tau)A_1(\bar{u}, x_t, k)dk + \delta(1-\tau)A_1(\bar{\bar{u}}, x_t, \varepsilon) + \int_0^T \delta(\tau-\tau_1)cA_2(k)dk + (\tau-\tau_1)cA_2(\varepsilon) \]
where \( \varepsilon \) denotes the time to maturity, \( T \) is known terminal valuation time.

\[ E_t = \int_0^T \delta(1-\tau)A_1(\bar{u}, x_t, k)dk - \int_0^T \delta(1-\tau)A_2(k)dk + \delta(1-\tau)A_1(\bar{u}, x_t, \varepsilon) - c(1-\tau_1)A_2(\varepsilon) \]
\[ D_t = \int_0^T (1-\tau_1)cA_2(k)dk + c(1-\tau_1)A_2(\varepsilon) \]

As a result, the leverage ratio can be evaluated at time \( t \), which is equal to
\[ L_t = \frac{\int_0^T \delta(1-\tau)A_1(\bar{u}, x_t, k)dk + \delta(1-\tau)A_2(k)dk}{\int_0^T (1-\tau_1)cA_2(k)dk + c(1-\tau_1)A_2(\varepsilon)} \]

This formula is actually equivalent to \( L_t = \frac{D_t}{E_t} \), substituting \( E_t \) and \( D_t \) generates the result. The proof of Proposition one is finished.

**Proof of proposition 2:**

If mean log cash flow is unknown, the firm value can be written as
\[ V_t = E_t \left\{ E_T \left\{ \int_0^T \pi(u, \omega) (1-\tau)(\delta X_w - c) / \pi_T dw + [\pi_T (1-\tau) \delta X_T - c] / \pi_T \right\} + \int_0^T \pi(u, 1-\tau) / \pi_T d\tau + \pi_T (1-\tau_1) / \pi_T c (\bar{u}) \right\} = X_t E_t \left\{ A(\bar{u}, x_t, \varepsilon) \right\}, \]

where \( A(\bar{u}, x_t, \varepsilon) \) is already defined as \( A_1 \) and \( A_2 \), based on previous of derived results for them, we have
\[ V_t = \int_0^T \delta(1-\tau)E_t \left\{ A_1(\bar{u}, x_t, k)dk + \delta(1-\tau)E_t \left\{ A_1(\bar{u}, x_t, \varepsilon) + \int_0^T \delta(\tau-\tau_1)cA_2(k)dk + (\tau-\tau_1)cA_2(\varepsilon) \right\} \right\} \]

Now the leverage ratio with learning is defined as
\[ L_t = \int_0^T \delta(1-\tau)(A_1(\bar{u}, x_t, k)dk + \delta(1-\tau)(A_1(\bar{u}, x_t, \varepsilon) + \int_0^T \delta(\tau-\tau_1)cA_2(k)dk + (\tau-\tau_1)cA_2(\varepsilon) \]

To derive the updating process for expected log cash flow, we can let
\[ dF_t = d \left( \frac{v_t}{log X_t} \right) = \left[ \begin{array}{c}
-\frac{1}{2} \sigma_{\pi}^2 \\
0
\end{array} \right] + \left[ \begin{array}{c}
\varphi \\
0
\end{array} \right] \bar{u} + \left[ \begin{array}{c}
-\varphi \\
0
\end{array} \right] \left( \begin{array}{c}
v_t \\
log X_t
\end{array} \right) dt + \left[ \begin{array}{cc}
\sigma_{\pi,1} & \sigma_{\pi,2} \\
\sigma_{x,1} & \sigma_{x,2}
\end{array} \right] \left[ \begin{array}{c}
dZ_1 \\
dZ_2
\end{array} \right] \]

Now denote \( \hat{u}_t = E(\bar{x}) = \bar{u} \), this is the expectation of \( \bar{x} \) conditional on the information set \( F_t = \{ H_t = (\bar{v}_t, log X_t) : 0 \leq \omega \leq t \} \), following Pastor and Veronese (2003).

Define \( d\tilde{Z}_t = \Sigma_1^{-1}(dF_t - E(dF_t)) = \Sigma_1^{-1}(dF_t - (\kappa_0 + \kappa_1 \bar{x} + \kappa_2 F_t)dt, \tilde{Z}_t \) is a standrad weiner process with respect to \( F_t \). Given a prior distribution at time of \( t = 0 \),

We know that \( \hat{u} \sim N(\bar{u}_0, \sigma_0^2) \),while \( \hat{u} \) satisfies the SDE \( d\hat{u}_t = \hat{\sigma}_t^2 \kappa_1 (\Sigma')^{-1} d\tilde{Z}_t = \hat{\sigma}_t^2 \frac{\varphi}{\sigma_{x,2}} d\tilde{Z}_2,t \)

Finally, \( d\hat{u}_t = \hat{\sigma}_t^2 \frac{\varphi}{\sigma_{x,2}} d\tilde{Z}_2,t \)

Then posterior variance of mean log cash flow satisfies the Riccati differentiation \( \frac{d\hat{\sigma}_t^2}{dt} = -\hat{\sigma}_t^2 \kappa_1 (\Sigma')^{-1} \kappa_1 \]
\[ = -\hat{\sigma}_t^2 \left( \begin{array}{c}
\varphi \\
\varphi
\end{array} \right) \left[ \begin{array}{cc}
\sigma_{\pi,1} + \sigma_{x,1} & \sigma_{x,2} \\
\sigma_{x,1} & \sigma_{x,2}
\end{array} \right] \left[ \begin{array}{cc}
\sigma_{\pi,1} + \sigma_{x,1} & \sigma_{x,2} \\
\sigma_{x,1} & \sigma_{x,2}
\end{array} \right]^{-1} \left( \begin{array}{c}
\varphi \\
\varphi
\end{array} \right) \]
\[ \hat{\sigma}_t^2 = 1/(\sigma_0^2 + \frac{\varphi^2}{\sigma_{x,2}^2} t). \]
References


