

MOVEMENTS IN THE EQUITY PREMIUM: EVIDENCE FROM A TIME-VARYING VAR*

MASSIMILIANO DE SANTIS[†]

ABSTRACT. Previous literature has recognized the importance of regime changes in the calculation of ex-ante equity premia. However, the methodologies used to estimate equity premia only allow for very restrictive forms of regime transitions. This paper addresses the issue by postulating an evolving model for the law of motion of dividend growth, consumption growth and dividend-price ratio. Model parameters are then used to compute conditional and unconditional U.S. equity premia. We substantially extend and confirm previous work on the declining equity premium and perform exploratory data analysis in search of clues about macroeconomic factors driving the equity premium. We find that the equity premium has declined, particularly from 1950 to 1971 and from 1988 to 2000, and this decline suggests that high post-war equity returns represent the end of a high equity premium. Our results point to changing consumption volatility as an important priced factor. We find that volatility of consumption growth is a good indicator of economic uncertainty, and as such, its changes are reflected in expected returns, and are priced by the market.

*I am grateful to Timothy Cogley, Oscar Jorda, Louis Makowski, Klaus Nehring, Giorgio Primiceri, Martine Quinzii, Kevin Salyer, Aaron Smith, and participants at the European Financial Management Meetings (Basel, 2004) for many helpful comments. Special thanks go to Timothy Cogley, whose suggestions helped improve substantially a previous version of the paper.

[†]Department of economics, Dartmouth College, 6106 Rockefeller Hall, Hanover, NH 03755. E-mail: massimiliano.desantis@dartmouth.edu.

1. INTRODUCTION

Previous literature has recognized the importance of regime changes in the calculation of ex-ante equity premia. However, the methodologies used to estimate equity premia only allow for very restrictive forms of regime changes. For example, Blanchard (1993) uses rolling samples to estimate conditional equity premia. Jagannathan, McGrattan and Scherbina (2000), and Fama and French (2002) use non-overlapping subsamples to estimate unconditional equity premia. In this paper we use an optimal filter that allows for a wide class of regime transitions to efficiently estimate ex-ante equity premia.¹

Calculation of expected equity returns and equity premia is crucial to correctly price assets and guide portfolio allocation decisions. The work cited above emphasizes that use of historical averages of excess returns may result in a poor estimate of the ex-ante equity premium. This is because the calculation misses changes in prices that would accompany an unexpected change in the premium. A more precise measure of the expected equity premium is then calculated from the yield derived from present value relations.

There are two related problems with the use of present value formulas to estimate expected returns. First, they require a log-linearization of returns around a steady-state value of the dividend-price ratio. Second, they require a specification of the law of motion of dividend growth. The usual assumption is that dividend growth follows a stationary ARMA process. While this is a convenient simplifying assumption, there is no reason to assume that the law of motion of dividend growth should follow a stationary process. For example, the Modigliani-Miller theorem states that firm value maximization does not constrain the form of dividend policy. Firm managers then have no incentive to follow a constant law of motion for dividends. This may lead to inconsistent estimation of expected returns. If prices are invariant to dividend policy while dividends are subject to regime changes, the law of motion of the dividend-price ratio will also be time-varying. This implies that log-linearization around an invariant steady state value may be inappropriate. Since the approximating constants enter as geometric weights in the infinite sum of future dividend growth required to estimate expected returns, even small changes in the steady-state value of the dividend-price ratio imply a larger bias in estimated expected returns.

¹The distinction between conditional and unconditional will be defined below. In this paper we provide estimates of both conditional and unconditional equity premia.

If the laws of motion of dividend growth and the dividend-price ratio evolve, so should their joint behavior with other macro-economic variables like consumption growth. In this paper we address these issues by modeling dividend growth, the dividend-price ratio, and consumption growth as a reduced form VAR with two ingredients: time-varying coefficients and time-varying variance-covariance matrix. In contrast to previous literature that examines the behavior of the ex-ante equity premium over time, we use an optimal filter to provide Bayesian estimates of the annual equity premium from 1928 to 2002 that use the entire sample. Also in contrast to previous literature, we include consumption in the system to relate movements of expected returns and equity premia to sources of macroeconomic risk, as measured by fluctuations in per capita consumption growth. Moreover, Parker and Julliard (2005) find that consumption contains information about expected returns at multi-year horizons for a cross section of stock portfolios.

Results from our empirical model substantially extend and confirm previous work of Blanchard (1993), Siegel (1999), Jagannathan, McGrattan and Scherbina (2000), and Fama and French (2002), on the declining equity premium. We find that the equity premium in recent years is closer to levels implied by standard consumption models and that it has been declining in the post-war period from the unusually high levels of the 1930's and 1940's. This decline suggests that the high equity returns in the post war period may represent the end of a high equity premium, as opposed to a puzzle. Furthermore, we find a common low frequency component between volatility of consumption growth and the level of the equity premium.

We also perform exploratory data analysis in search of clues about factors that drive risk premia at business cycle frequencies. Results point to changing consumption volatility as an important priced factor. We find that volatility of consumption growth is a good indicator of economic uncertainty, and as such, changes are reflected in expected returns, and are priced by the market.

For time variation to be relevant, it should also be detected in the data. Timmermann (2001) presents empirical evidence on the existence of multiple structural breaks in the U.S. dividend process using monthly data on the S&P 500 from 1871-1999. Evans (1998) shows evidence of breaks using annual series. Here we review some of that evidence using a quarterly data set and performing stability tests on the VAR equations for dividend

growth, consumption growth, and dividend-price ratio. The results of these tests are presented in the next section.

That expected excess returns evolve over time has been well documented and is at the heart of the predictability literature (Fama and French, 1998; DeBont and Thaler, 1985). Evans (1998) gives an example in which ex-post returns are in-sample forecastable even though agents have rational expectations as a result of regime switching in VAR parameters.

Evans uses Campbell and Shiller's (1988) log-linear approximation of returns and allows the dividend process to switch between two regimes. He then uses the estimated time series process to derive implications for asset prices. Evans is not directly interested in measures of the equity premium, so he does not use the implied VAR parameters to estimate expected excess returns. Further, we do not restrict dividend growth to switch between two regimes. Discrete-switches models either impose a finite number of recurring states, or a finite number of non-recurring states and the switch between regimes is a discrete jump. These models may well describe rapid changes in the joint behavior of the variables of interest, but seem less suitable to capture changes in aggregate stock market behavior, where aggregation among agents smooths out most of the changes. Finally, VAR parameters may vary as a result of economy-wide changes other than dividends, such as changes in preferences or risk attitudes, which can affect the time series properties of the dividend-price ratio.

Other work has looked at the movements in the equity premium over time using present value relations. Of particular importance for our work is Blanchard (1993). As we will see below, our conditional measure of the premium is intimately related to Blanchard's. Blanchard recognizes that the relationship between fundamentals and the premium varies over time. He is more concerned about an unstable inflation process over the sample he considered. This is important in Blanchard's framework, since he needs an estimate of a long-run real return on bonds. Because of this, he chooses to use 40 years rolling samples in his estimations of both expected stock returns and expected bond yields. While rolling regressions allow for smooth regime changes of the type modeled here, they throw away some of the sample at each point in time and the sample size used at time t is chosen in a deterministic way. The procedure used in this paper extends the work of Blanchard in this direction, by providing estimates of expected equity premium at each point in time that use the entire sample. It is

left to the data to decide how much weight to give to observations far from date t . Finally, Blanchard's procedure requires to make assumptions on dividend growth after the terminal date in the sample (1993 in his case). Use of VAR parameters as in our case does not require this.

Like Blanchard, Fama and French (2002), and Jagannathan McGrattan and Scherbina (2000) provide evidence that the unconditional equity premium has declined in the last 50 years and suggest that high realized returns over the period are a consequence of a declining equity premium. Both papers base their analysis on unconditional measures based on ten years sub-samples. Therefore they implicitly assume that the stochastic process underlying stock prices is stable within each decade.

Finally, by introducing consumption growth in the VAR, we link our work to recent developments in the relationship between asset prices and macroeconomic risk, as measured by volatility of consumption. Consumption volatility is found to be time varying and predictable by valuation ratios. Recent work by Bansal, Khatcharian, and Yaron (2003) and Lettau, Ludvigson, and Wachter (2003) show that this relationship is consistent with existing general equilibrium models. Here, using our conditional measures, we provide direct empirical evidence on the relationship between consumption volatility and expected returns, and consumption volatility and expected excess returns.

The remaining of the paper is organized as follows. Section 2 reports stability tests on the VAR equations. Section 3 outlines the econometric model used and discusses the relevant assumptions. Section 4 motivates the Bayesian inference, specifies the priors used in the analysis and gives an overview of the Gibbs sampler. Section 5 details the used measures of expected returns and discusses the results. Section 7 concludes the paper. Two appendices at the end of the paper provide robustness checks of our Bayesian inferences, and details of the Gibbs sampler used in estimation

2. STABILITY TESTS

Stability tests are conducted using quarterly data on dividend growth (Δd_t), per-capita consumption growth (Δc_t) and the log of dividend-price ratio (δ_t). Data on dividends and prices refer to the S&P 500 and are downloaded from Robert Shiller's website as

well as the CPI used to convert to real figures. Data on consumption and population are from the FREDII website.²

The top panel of Table 1 presents summary statistics and Phillips-Perron unit root tests for δ_t , Δc_t , Δd_t . The second part of the table presents residual analysis for AR(2) models of each variable. Ljung-Box statistics (lag length = 12) indicate that two lags are enough to model the dynamics of dividend yields, consumption growth, and dividend growth. Absence of autocorrelation is not rejected in the residuals of the estimated equations. The Phillips-Perron test statistic does not reject the unit root hypothesis for dividend yields, although it does reject the hypothesis for dividend growth and consumption growth. One possible reason for failing to reject is that the dividend yield is a very persistent process, and so the data is not informative enough to distinguish between the two types of processes. Alternatively, and this is the view we take here, the time series model for δ_t may not be stable over the sample, while being stationary within sub-intervals in the sample.

To explore this hypothesis, we present results from the Bai and Perron (1998) tests for structural breaks in Table 2.

The battery of tests by Bai and Perron provides a way to test for deterministic breaks, i.e. at this point we do not seek to model the probability of a break in the processes governing the variables. For instance, a process for dividend-yields δ_t with m deterministic breaks can be written as

$$\begin{aligned}
 (1) \quad & \delta_t = x_t' \beta_1 + u_t && t = 1, 2, \dots, T_1 \\
 & \delta_t = x_t' \beta_2 + u_t && t = T_1 + 1, \dots, T_2 \\
 & \vdots && \vdots \\
 & \delta_t = x_t' \beta_{m+1} + u_t && t = T_m + 1, \dots, T.
 \end{aligned}$$

where T is the sample size, $T_1 < T_2 < \dots < T_m < T$ are the break points, u_t is a disturbance term, and β_j are the time-varying parameters. The deterministic procedure of Bai and Perron provides tests and consistent estimation of the number and location of breakpoints. The tests were conducted using a Gauss procedure provided by Bai and Perron (2001)(henceforth BP). We allowed up to 8 breaks and used a trimming

²Evans, 1998 provides evidence of an unstable dividend growth process using Shiller's annual data. Timmermann (2001) presents evidence at monthly frequency.

$\epsilon = 10\%$,³ hence each segment has a minimum of 19 observations. Consider the dividend yield first. The first issue is the determination of the number of breaks. The first column of Table 2 shows results for dividend yield modelled as an AR(2) process to limit the number of estimated parameters. SupF tests of zero breaks versus 1 up to 5 breaks are all significant at the 1% level. The two double maximum tests, unweighted double maximum test (UDmax) and weighted double maximum (WDmax), test for zero breaks against an unspecified number of breaks and show significance at the 1% level. The following two SupF statistics test the presence of $l + 1$ breaks given that l breaks are present. I only report here statistics up to three breaks given two, which are both significant at the 1% level. The sequential procedure is a procedure that BP suggest for estimating the number of breaks, which corresponds to a sequential application of the $\text{SupF}_T(l + 1|l)$ test to estimate the breaks. The procedure finds evidence of four breaks at the 5% level.

The third column of Table 2 presents results from a model for dividend yields that includes AR(2) terms augmented by Δc_t and Δd_t to guard against the possibility that the breaks are due lack of appropriate conditioning information. If anything, the results strengthen the evidence of breaks in the dividend yield process.

Next, we analyze the behavior of consumption growth and dividend growth. Consumption growth shows pretty strong evidence of breaks: most of the SupF tests are highly significant as well as the double maximum tests. Dividend growth shows less evidence of breaks, although the WDmax is significant at the 1% level. Stronger evidence for breaks is found in the absolute value of dividend growth, $|\Delta d_t|$, or in the square of dividend growth, which may be thought as proxies for volatility. This indicates that a better model for dividend growth should be non-linear. The last column shows results for $|\Delta d_t|$. The SupF statistics, as well as the double maximum statistics are all highly significant, suggesting rejection of constant parameters in the variance of dividend growth. The $\text{SupF}_T(l + 1|l)$ statistics fail to reject. Bai and Perron (2001) points out that the sequential procedure may fail to reject in the presence of breaks if there are recursive states. This could be the case here given the evidence of breaks indicated by the double max statistics. It would also be consistent with the observation that volatility varies over the business cycle.

³See Bai and Perron 2001 for details. Other values for the trimming parameters and number of breaks were tried for robustness check, but results are very similar.

To summarize, we find significant evidence of parameter instability in the behavior of $\delta_t, \Delta c_t, \Delta d_t$ over the sample considered. This and the existing evidence reported in Timmermann (2001) and Evans (1998) gives support to a time varying specification of the joint behavior of the series.⁴

3. THE ECONOMETRIC MODEL

We model the joint behavior of the dividend-price ratio, dividend growth and consumption growth as a VAR with time-varying parameters:

$$(2) \quad \begin{aligned} y_t &= X_t' \theta_t + u_t \\ X_t' &= I_n \otimes [1, y_{t-1}', \dots, y_{t-k}'] \end{aligned}$$

where y_t includes dividend growth, consumption growth, and the log of dividend-price ratio in this order. In general this is an $n \times 1$ vector. \otimes denotes the Kronecker product, so in general X_t' is an $n \times k$ matrix. θ_t is the $k \times 1$ vector of coefficients. The u_t are disturbance terms with variance covariance matrix Ω_t . Without loss of generality consider the following decomposition

$$(3) \quad \Omega_t = A_t^{-1} \Sigma_t A_t'^{-1}$$

where A_t is lower triangular

$$A_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{21,t} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ a_{n1,t} & \cdots & a_{nn-1,t} & 1 \end{bmatrix}$$

and $\Sigma_t^{1/2}$ is the diagonal matrix

$$\Sigma_t^{1/2} = \begin{bmatrix} \sigma_{1,t} & 0 & \cdots & 0 \\ 0 & \sigma_{2,t} & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{n,t} \end{bmatrix}.$$

It follows that (2) is equivalent to

$$(4) \quad y_t = X_t' \theta_t + A_t^{-1} \Sigma_t^{1/2} \varepsilon_t$$

⁴We also conduct Hansen's (1992) test of parameter instability and find similar evidence. Results (not reported) are available in an appendix upon request.

The drifting coefficients are meant to capture possible nonlinearities or time variation in the lag structure of the model. The multivariate time varying variance covariance matrix captures possible heteroskedasticity of the shocks and time variation in the simultaneous relations among the variables in the system. In the context of time varying VAR models, a similar specification has been proposed by Primiceri (2003) and Cogley and Sargent (2002), though the latter has a time invariant A_t matrix. As emphasized in Primiceri (2003), a time variant A_t is highly desirable if the objective is to model time variation in a simultaneous equation model.

Let α_t be the vector of non-zero and non-one elements of the matrix A_t (stacked by rows) and σ_t be the vector of diagonal elements of $\Sigma_t^{1/2}$. The model's time varying parameters evolve as follows:

$$(5) \quad \theta_t = \theta_{t-1} + \nu_t,$$

$$(6) \quad \alpha_t = \alpha_{t-1} + \zeta_t,$$

$$(7) \quad \log \sigma_t = \log \sigma_{t-1} + \eta_t,$$

with the distributional assumptions regarding $(\varepsilon_t, \nu_t, \zeta_t, \eta_t)$ stated below. Time varying parameters θ_t and A_t are modeled as driftless random walks and the standard deviations are assumed to evolve as geometric random walks. Thus, the model belongs to the class of stochastic volatility models, which constitutes an alternative to ARCH models. The crucial difference with ARCH is that the variances generated by (7) are unobservable components.

Equations (4)-(7) form a state space representation for the model. (4) is termed the observation equation, and (5)-(7) are the state equations.

An undesirable feature of the random walk assumption is that the process hits any upper or lower bound with probability one. Our objective though is to uncover the values of the parameters θ_t , A_t and σ_t as they evolve in our finite sample. As long as (5)-(7) are thought to be in place for a finite period of time, the random walk assumption should be quite innocuous and provides flexibility while reducing the number of parameters in the estimation procedure. This is particularly true if, quite plausibly, the variances of parameter innovations are small.

All the innovations in the model are assumed to be jointly normally distributed with a block diagonal covariance matrix:

$$(8) \quad V = \text{Var} \begin{pmatrix} \varepsilon_t \\ \nu_t \\ \zeta_t \\ \eta_t \end{pmatrix} = \begin{bmatrix} I_n & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W \end{bmatrix},$$

where I_n is the identity of dimension n , Q , S , and W are positive definite matrices. We will further assume that S is block diagonal with blocks corresponding to parameters belonging to separate equations in the structural model. This assumption simplifies inference and increases the efficiency of the estimation algorithm.

4. BAYESIAN METHODS

The model in (4)-(7) is basically a regression model with random coefficients and covariances. The Bayesian framework, which views parameters as random variables, is the most natural approach in this setting. The Kalman filter, which is the algorithm used to make inferences about the history of θ_t , also fits naturally in a Bayesian framework (see Meinhold and Singpurwalla, 1983). This section gives an overview of the estimation strategy and the algorithm used in estimation. Two other important reasons make Bayesian methods particularly suitable for this class of models. First, if the variance of the time varying coefficients is small, as one would expect here, then the maximum likelihood estimator is biased towards a constant coefficients VAR. As a consequence, Numerical optimization methods are very likely to get stuck in uninteresting regions of the likelihood (see for instance Stock and Watson, 1998 for a discussion on the subject). The second and related drawback is that numerical optimization methods have to be employed in a highly dimensional problem. Multiple peaks are highly probable in such a non-linear model. This makes MLE quite unreliable if in fact a peak is reached at all. In a Bayesian setting with uninformative or weakly informative priors on reasonable regions of the parameter space, these types of misbehavior are limited.

The problem of estimating a highly dimensional parameter vector is dealt with by means of the Gibbs sampler, which allows to divide the task in smaller and simpler ones. The Gibbs sampler is a stochastic algorithm, and as such it is more likely to escape local maxima.

Finally, MCMC methods, of which Gibbs sampling is a variant, deliver smoothed series i.e. conditional on observing the sample. This is what we want here, as the objective is to uncover how economic quantities of interest have evolved over time in our observed sample.

4.1. Priors.

We choose prior distributions following Cogley and Sargent (2002) and Primiceri (2003). The choice is based on intuitiveness and statistical convenience of the distributions for the application at hand. Following the Bayesian literature, θ_t , A_t , Σ_t will be called “parameters” and the elements of Q , S , W “hyperparameters”.

The hyperparameters are assumed to be distributed as independent inverse-Wishart. The Wishart distribution can be thought of as the multivariate analog of the χ -square, and it is used to impose positive definiteness of the blocks of V as defined in (8). The prior is

$$p(V) = IW(\bar{V}^{-1}, T_0),$$

where $IW(Sc, df)$ represents the inverse-Wishart with scale matrix Sc and degrees of freedom df .

The priors for the initial states of the regression coefficients, the covariances, and log volatilities, $p(\theta_0)$, $p(\alpha_0)$, $p(\log \sigma_0)$, are conveniently assumed to be normally distributed, independent of each other and of the hyperparameters. The VAR is further assumed to be stationary at each point in time. This can be written as

$$p(\theta_0) \propto I(\theta_0)N(\bar{\theta}, \bar{P}),$$

where $I(\theta_0) = 0$ if the roots of the associated VAR polynomial are on or inside the unit circle.

The assumption of a normal prior may be thought as the asymptotic distribution of the parameters θ_0 , α_0 , and $\log \sigma_0$ in the frequentist approach. As the sample size on which the prior is calibrated (T_0) gets large, the frequentist estimate of the parameters would approach a normal distribution under mild assumptions by a central limit theorem. Here, the assumption is made mostly for simplicity. These assumptions, together with (5)-(7) imply normal priors for the evolving parameters. For instance, the vector of covariance states evolves according to $p(\alpha_{t+1}|\alpha_t, S) \sim N(\alpha_t, S)$, and similarly for volatility states. The vector of coefficient states on the other hand, evolves according

to

$$(9) \quad p(\theta_{t+1}|\theta_t, Q) \propto I(\theta_{t+1})f(\theta_{t+1}|\theta_t, Q)\pi(\theta_{t+1}, Q)$$

where $I(\theta_t) = 0$ if the roots of the associated VAR polynomial are inside the unit circle and

$$(10) \quad f(\theta_{t+1}|\theta_t, Q) \sim N(\theta_t, Q).$$

The multiplication by $I(\theta_t)$ reflects the assumption that the log dividend-price ratio, dividend growth and consumption growth evolve as a stationary VAR, given the state. This is important if we want to estimate long-run expected returns, as we will see below. The constant $\pi(\theta_{t+1}, Q)$ is derived in Cogley and Sargent (2002) . It represents the probability that random walk paths emanating from θ_{t+1} will remain in the non-explosive region going forward in time. Thus $I(\cdot)$ truncates the unconstrained normal distribution $f(\theta_{t+1}|\theta_t, Q)$ and $\pi(\cdot)$ downweights values of θ_{t+1} that are likely to become explosive.

The normal prior on θ is standard. The non-unit roots prior is proposed by Cogley and Sargent (2001, 2002) . Primiceri (2003), Smith and Kohn (2002) use the same decomposition of Ω_t and place a similar prior on the elements of A , as well as Cogley and Sargent (2002). The log normal prior on the volatility parameters is common in the stochastic volatility literature modelling η_t as Gaussian (see Kim, Shephard and Chib, 1998). Such prior is not conjugate, but has the advantage of tractability, as detailed in the Appendix.

4.2. Overview of the Simulation method.

The complete Gibbs sampling procedure is detailed in the Appendix, as well as a description of how priors are calibrated. Here I sketch the MCMC algorithm used to sample from the joint posterior of $(\theta^T, A^T, \Sigma^T, V)$. Here and throughout the paper, a superscript T denotes complete histories of data (e.g. $\theta^T = \theta'_1, \dots, \theta'_T$).

Sampling from the joint posterior is complicated, so sampling is carried out in four steps by sequentially drawing from the conditional posterior of the four blocks of parameters: coefficients θ^T , simultaneous relations A^T , variances Σ^T , and hyperparameters V . Posteriors for each block of the Gibbs sampler are conditional on the observed data and the rest of the parameters.

Conditional on A^T and Σ^T , the state space form given by (4) and (5) is linear and Gaussian. Therefore, the conditional posterior of θ^T is a product of Gaussian densities and θ^T can be drawn using a standard simulation smoother (see for instance Fruhwirth-Schnatter (1994) or Cogley and Sargent 2002). This consists in drawing an initial state θ_0 , then use of the Kalman filter produces a trajectory of parameters. From the terminal state, a backward recursion produces the required “smoothed” draws (i.e. draws of θ ’s given Y^T). Similarly, the posterior of A^T conditional on θ^T and Σ^T is a product of normal densities, so A^T is drawn in the same way.

Drawing from the conditional posterior of Σ^T is a little more involved because the conditional state-space representation for $\log \sigma_t$ is not Gaussian. This stage of the Gibbs sampler uses a method proposed by Kim, Shephard and Chib (1998) . This consists of transforming the non-Gaussian state space form in an approximately Gaussian one (by using a mixture of normal distributions), which allows us again to use the standard simulation smoother conditional on a member of the mixture.

Finally, drawing from the conditional posterior of the hyperparameters (V) is standard, since it is a product of independent inverse-Wishart distributions.

The same Gibbs sampler is used by Primiceri (2003) in the context of a VAR for the US economy. Still in the contest of evolving monetary policy, Cogley and Sargent’s (2002) algorithm is similar, though differs in the assumption of A^T being constant, and in the use of a different method to draw from the posterior for volatility states.

After a transitional period (“burn-in” period), the sequence of draws of the four blocks from their respective conditional posteriors converges to a draw from the joint distribution $p(\theta^T, A^T, \Sigma^T, V|Y)$.

5. RESULTS FROM THE TIME VARYING VAR

In this section, we present results from two types of VAR(2) estimated from two different samples. A VAR(2) for dividend growth (Δd_t), consumption growth (Δc_t), and the log of dividend-price ratio ($d_t - p_t \equiv \delta_t$) is used to estimate expected stock returns in both samples.⁵ A second VAR(2), in which dividend growth is replaced by dividend growth in excess of the risk free rate ($\Delta d_t - r_t^f$) is estimated to measure the equity premium, defined as the expected excess return on stocks over the risk free rate.

⁵Here and throughout the rest of the paper, lower case represent natural logs of the variables, e.g. p_t is log of price at time t.

The equity premium is inferred using a time-varying version of what Campbell and Shiller call the *dynamic Gordon growth model*, which will be detailed below.

Data on dividends and prices refer to the S&P 500. The data is downloaded from Robert Shiller’s web-site. The analysis uses two separate data sets. The first data set is annual and runs from 1890 through 2002. The second is quarterly and runs from the third quarter of 1949 through the second quarter of 2002. The annual data set includes (apart from S&P 500 data) the consumption series, CPI and the rate on 4-6 month commercial paper (the “risk-free” rate) available on Shiller’s page. This dataset is the one used in most the work reported in Shiller (1989) and much other subsequent work in asset pricing. Quarterly data on the S&P and consumption are the ones described in section 2. The risk free rate considered in quarterly data is the 3 month T-bill.

We focus on the annual sample to uncover movements in the last 75 years and relate this to the discussion on the declining equity premium. We then use the quarterly sample to further explore movements in the equity premium during the last 40 years and to relate this to some recent literature on the premium and macroeconomic risk that uses the same data set.

Our results on the equity premium are conditional on the index used, the S&P 500. While it may be argued that the S&P 500 index is too narrow a measure for overall market performance, Campbell and Shiller (1988a) document striking similarities between the S&P 500 and the CRSP index over the period 1926-1986. The indices have a correlation coefficient of 0.985 in annual data. Their mean is 0.044 and 0.042 respectively, and the standard deviations are 0.200 and 0.208. Similar results are found for dividends and dividend-price ratio series. Also, the S&P 500 includes something like 75% of US securities in value.

For each estimated VAR, we repeat the algorithm detailed in Appendix 24,000 times, dropping the first 4,000 draws and keeping one every two draws of the remaining 20,000. This yields a sample of 10,000 draws. We use posterior draws to compute expected returns and risk premia as detailed below.

5.1. Prior Calibration.

The priors are calibrated on a constant parameter VAR(2) estimated using an initial sample of 36 observations. This corresponds to the years 1892-1927 in annual data and

1952:Q1-1961:Q3 in the quarterly sample. Priors for parameters and hyperparameters are modeled as follows:

$$\begin{aligned}
\theta_0 &\sim N\left(\hat{\theta}_{OLS}, V(\hat{\theta}_{OLS})\right) \\
A_0 &\sim N\left(\hat{A}_{OLS}, V(\hat{A}_{OLS})\right) \\
\log \sigma_0 &\sim N(\log \hat{\sigma}_{OLS}, I_n) \\
Q &\sim IW\left(k_Q^2 T_0 V(\hat{\theta}_{OLS}), T_0\right) \\
W &\sim IW(k_W^2 I_n, 4) \\
S_1 &\sim IW\left(k_S^2 V(\hat{A}_{1,OLS}), 2\right) \\
S_2 &\sim IW\left(k_S^2 V(\hat{A}_{2,OLS}), 3\right)
\end{aligned}
\tag{11}$$

The prior on θ_0 is standard. For σ_0 we simply use the log of the OLS estimate. The prior on A is calibrated using the residual from the OLS regressions $\hat{u}_t = A_0^{-1} \Sigma_0 \varepsilon_t$. Since A_0 is lower triangular, we can get estimates of the coefficients in A by regressing $\hat{u}_{t,2}$ on $\hat{u}_{t,1}$, and $\hat{u}_{t,3}$ on $\hat{u}_{t,2}$ and $\hat{u}_{t,1}$. The regressions also provide estimates of $V(\hat{A}_{OLS})$. The prior for the hyperparameters are inverse-Wishart with scale matrices set to a fraction of the OLS covariance matrix of the respective parameters. So for Q , the scale matrix k_Q^2 times the covariance of the OLS estimates for θ_0 , times T_0 . We set $k_W = 0.025$. With $k = 0.025$ our prior attributes 2.5% of the estimated total variation in parameters to time variation. This should be a quite conservative value, letting the likelihood add variability if needed. T_0 , the prior degrees of freedom, is set to 22, the minimum required for the prior to be proper ($22 = \dim(Q) + 1$). We multiply the variance by T_0 so that we have a scale matrix, as opposed to a covariance. Cogley and Sargent (2001) and Primiceri (2003) use similar values. For k_W and k_S I choose the same values as Primiceri (2003), i.e. 0.01 and 0.1 respectively. Some robustness checks are discussed in Appendix A. These values seem to be plausible for both data sets and the conclusions drawn below are not altered for alternative sensible specifications of the parameters.

5.2. Measures of Expected Returns and Equity Premiums.

To estimate expected returns and excess returns, we use the log linear approximation

of returns of Campbell and Shiller (1988a,b) . This is

$$(12) \quad h_{t+1} \simeq k + \delta_t + \Delta d_{t+1} - \rho \delta_{t+1}$$

where h denotes log returns, k and ρ approximating constants, and δ_t the log of dividend-prices ($\delta_t = d_t - p_t$). The constants of log-linearization are evaluated at the mean of δ in Campbell and Shiller (19881,b), so that they are defined by $\rho = 1/(1 + \exp(\overline{d - p}))$ and $k = -\log(\rho) - (1 - \rho) \log(1/\rho - 1)$.

Campbell and Shiller derived the linear approximation on the assumption that dividend yield is a stationary process and so choose the sample mean as point of approximation. In a time varying context, it is more appropriate that the approximating point varies over time. We do this by approximating around μ_t^δ , the time-varying unconditional mean of dividend-prices, i.e. we calculate k and ρ at each t (therefore we have k_t and ρ_t) using μ_t^δ as opposed to $\overline{d - p}$.

The first order difference equation defined by (12) can be solved for δ_t imposing the terminal condition $\lim_{j \rightarrow \infty} \rho_t^j \delta_{t+j} = 0$. Taking expected values conditional on an information set containing δ_t we obtain

$$(13) \quad \delta_t = -\frac{k_t}{1 - \rho_t} + E_t \sum_{j=0}^{\infty} \rho_t^j h_{t+j+1} - E_t \sum_{j=0}^{\infty} \rho_t^j \Delta d_{t+j+1}.$$

So the dividend-price ratio (log of it) is equal to a constant plus a weighted sum of future expected returns minus a weighted sum of future expected dividend growth. Campbell and Shiller call this the *dynamic Gordon growth model* for it generalizes Gordon's valuation formula $D_t/P_t = r - g$.

The first sum on the right hand side of (13) is a weighted sum of future expected returns whose weights sum to $(1 - \rho_t)$. We can therefore get a measure of expected returns as follows, which we term ER_c

$$(14) \quad ER_c \equiv (1 - \rho_t) E_t \sum_{j=0}^{\infty} \rho_t^j h_{t+j+1} = k_t + (1 - \rho_t)(\delta_t + E_t \sum_{j=0}^{\infty} \rho_t^j \Delta d_{t+j+1}).$$

This measure of expected returns is simply the average *yield* on the asset. Notice that even small changes in ρ_t may have an impact in the measurement of expected returns, as the error propagates in the infinite sum. VAR parameters are used to calculate the

expectation on the right hand side as

$$E_t \sum_{j=0}^{\infty} \rho_t^j \Delta d_{t+j+1} = s^{dg} \left(\frac{\mu_t}{1 - \rho_t} + F_t (I - \rho_t F_t)^{-1} \xi_t \right) .$$

F_t contains the time- t VAR parameters re-written in “companion form”, ξ_t contains the state vector in deviation from the (time-varying) unconditional means (as in Hamilton p. 259). The unconditional means are computed as $\mu_t = (I - F_t)^{-1} c_t$, where c_t is the vector of intercepts in the VAR. s^{dg} is a row vector that selects dividend growth from the VAR. This is analogous to what one would do with a constant parameter VAR. Here we use a different set of parameters at each date. The conditional expectation E_t is therefore conditional on the variables at time t , y_t in the VAR, and conditional on the VAR parameters θ_t . If we condition only on θ_t we can get a measure of time- t unconditional expected returns. In other words, if we had a time invariant VAR, the unconditional mean of (12) gives $\mu^h = k + \mu^\delta(1 - \rho) + \mu^{dg}$. Analogously, in a time varying VAR we have:

$$(15) \quad ER_u \equiv \mu_t^h = k_t + \mu_t^\delta(1 - \rho_t) + \mu_t^{dg} .$$

Values on the right hand side are calculated from the VAR parameters at each point in time. This is our unconditional measure. ER_u can also be derived by averaging over y_t in (14).

Our conditional measure of expected returns ER_c is an average yield on the risky asset, and can be thought as the average expected return over a period say of 15-20 years, in annual data. At each date t , expected 15-20 years annualized returns will depend on the price level relative to dividends at time t . If stocks are expensive relative to dividends compared to some mean reverting value, yields will be lower. The second measure ER_u represents expected returns as if one bought stocks at their mean price relative to dividends. The fact that our unconditional measure varies over time is meant to capture non-stationarity of $d_t - p_t$ due to structural shifts in dividend policy, productivity, or preferences that change expected returns and/or expected growth rates.

Our analysis focuses mostly on measures of equity premia, or expected *excess* returns. To calculate expected excess returns (the equity premium), notice that excess log returns can be approximated using (12) as

$$h_{t+1} - r_{t+1}^f = k_t + \delta_t + (\Delta d_{t+1} - r_{t+1}^f) - \rho_t \delta_{t+1}$$

where r_{t+1}^f is the return on the risk free rate between t and $t + 1$. This implies that to get measures of the equity premium using the VAR we can just run a VAR with $\Delta d_t - r_t^f$ instead of Δd_t , and use the same formulas above for this VAR with excess dividend growth. This procedure will yield a conditional equity premium (denoted by EP_c) and an unconditional equity premium (denoted by EP_u). Notice that this way of calculating the premium will automatically yield a real equity premium, as inflation corrections cancel out. This also mean that we do not have to worry about calculating expected inflation in our measure of the premium.

It is worth emphasizing the importance of ex-ante measures of the equity premium. Suppose the expected returns on stocks decline slowly and unexpectedly over some period of time. Then, simply looking at Gordon's valuation formula, the price of the security will be rising and realized returns will be higher. So there is a negative correlation between expected returns and realized returns. This is found to be the case in the data, and it lies at the basis of the return predictability literature. If expected returns are time varying with some degree of persistence, then variables that change with expected returns should have some correlation with realized returns and therefore should predict returns in the data.⁶ This observation has also prompted some theoretical research that has led to numerous models with this feature, namely time variation in expected returns negatively correlated with ex-post returns, as in Campbell and Cochrane (1999) and Barberis, Huang, and Santos (2001).

This misalignment between ex-ante and realized returns suggests that using returns directly in the VAR may result in a less precise estimate of the ex-ante equity premium. This is indeed what Fama and French (2002) suggest. Our measure ER_c can be regarded as an extension to the work of Blanchard (1993). Our unconditional measure ER_u is an extension to Jagannathan, McGrattan, and Scherbina (2000). Our approach is therefore part of a growing literature that uses valuation models to estimate expected returns. The approach is more general because it provides time-varying measures using the entire history of data. Blanchard's or Jagannathan et al.'s measures use rolling samples to account for time variation in the distribution of state variables.

5.2.1. *The Declining Equity Premium.*

Our measures of expected returns and equity premia for annual data are reported in

⁶See Cochrane 2001 and Campbell, Lo, MacKinlay 1997 for complete discussions about return predictability.

Figure 1. The equity premium is defined as excess returns on the S&P 500 relative to 6-month commercial paper. Figure 2 does the same for the quarterly sample, which uses the three-month T-bill as risk-free rate. The first noticeable fact is the decline of the equity premium over the past 75 years. This is reflected in both our conditional and unconditional measures. The sample mean of realized excess returns for the period 1928-2002 is 6.5%. Our measure of unconditional equity premium is close to this value only for the period 1928-46, when it is constant at about 6.4%, correcting for Jensen's inequality.⁷ Ex-post excess returns for the period 1928-46 average at about 6%. From 1946 to 1971 we observe a continuous decline, the decline being sharper from 1963. In this sub-sample, the realized excess return is 8.6%, but the unconditional mean return moves from 6% to 3.5 % (from 5% to about 2.5% using log-returns). This confirms both Fama and French (2002) and Jagannathan et McGrattan and Scherbina (2000) conjecture that the ex-post returns are a distorted view of expected excess returns on equities and are a result of a declining equity premium. Similarly, notice that our measure of log expected excess returns stays constant at about 2.5% between 1971 and 1988, or 3.2% in terms of expected excess returns. The ex-post excess returns during the period average at 3.2%. Succinctly, during periods of constant expected returns, the ex-post returns are a better measure of expected returns than in periods of changing expected returns. The evidence is summarized in Table 3. This should warn us about the use ex-post returns in equity valuation, a point also made by Jagannathan, McGrattan, Scherbina (2000) and Siegel (1999).

How can we explain the long run decline of the equity premium? Part of the high equity premium of the period 1928-46 can be explained by the turbulent years of the Great Depression. The feeling of aversion to the stock market generated by the volatile years during and after the Great Depression lasted until well after the war. The early thirties were indeed a period of higher volatility for both dividend growth and dividend yield, as can be seen from estimated volatilities (see Figure 4), particularly for dividend growth. Also, this is a period in which participation to the stock market was quite limited and mutual funds were not available to investors. This made it harder to share risks across people. Fear of a catastrophic event, limited participation, and costly diversification can explain the high ER_u and EP_u of the thirties and why the premium

⁷The VAR produces log excess returns. We correct our measure using the estimated variance of returns. The correction is on average 1.25%.

stayed high for so long. Investors in the 1930's could not know for certain that the U.S. would be the most successful capitalist country in history. Even a small probability of a catastrophic event like the Great Depression can generate a substantial premium, see Rietz (1988) . Even with the economy getting out of the depression, investors' revision of beliefs about the economy could be very slow, given the size of the depression. This could explain the persistence and slow decline of the premium afterwards. As memories of the Great depression started to fade, the premium gradually declined until 1971.

The increased desirability of stocks over this period (and therefore the declining premium) can be further reinforced by the perception that the business cycle has become less severe over time. A measure of the severity of the business cycle is the conditional standard deviation of consumption growth. Macroeconomic risk measured this way increased between 1928 and 1946 (see Figure 4). After 1946, it declines until 1970, strongly supporting the idea of declining macro-risk.

The unconditional equity premium is more or less constant in the period 1976-1988, though volatility of consumption growth keeps increasing until 1981. It then declines again from 1988 to 2002. Increased diversification from the availability of index mutual funds and other new financial instruments in the seventies offsets the increased volatility of consumption growth in the period 1976-1988 and, as a result, the premium remained more or less constant. The premium then declines again with lower uncertainty and greater opportunities for portfolio diversification.

A similar interpretation applies to our conditional measure EP_c . Recall from its definition that EP_c can be considered as an approximation to an average expected excess return on the stock market over a period say of 15-20 years, given the state of the economy at time t . EP_c peaks in 1951 and declines more or less steadily from 1952 to 1973. It then peaks in 1975 and 1985 to start declining again. The seventies are a period of greater uncertainty (see Figures 4 and 5) and higher inflation, and as a consequence, EP_c stops declining and it is about constant until 1988. EP_c peaks in 1975 and stays high until 1985, at a level of about 4%. It then declines to historical lows.

The brief shock of October 1987 is reflected in EP_c , our conditional measure, which increases in 1988 and declines little for about 3 years afterwards.

Adjusting for Jensen's inequality, the unconditional equity premium in 2002 is about 3% (2.95%), rather than the 6% recorded by Mehra and Prescott (1985).

The quarterly measures are basically a blow-up of the annual counterparts for the period 1961-2002, and tell a similar story. There is a sharp drop in the equity premium starting in 1994 which is not as pronounced in annual data. One possible explanation is a drastic regime shift in the payout policy of corporations at the end of the sample that is not well captured by our model. Grullon and Michaely (2002) report evidence for the period 1972-2000 that repurchases have become an important source of payout for U.S. corporations and that firms finance their share repurchases with funds that would otherwise be used to increase dividends. In the sample examined by Grullon and Michaely, repurchases amounted to an average of 10% of dividend payments up to 1983. Between 1984 and 2000, repurchases were 57.7% of dividends, reaching a maximum of 113.11% in 2000. Because this shift is so drastic and it occurs at the end of the sample, it is possible that our model interprets part of this regime shift as a decrease in expected future dividend growth, rather than a change in the law of motion for dividends. With a decrease in expected dividend growth, price can only rise relative to dividends if yields decrease. If this is the case, we can correct our measure of expected excess returns using a payout ratio, rather than the dividend price ratio, for the later part of the sample. Figure 3 graphs the series EP_c and EP_u calculated from a VAR that uses the payout ratio as opposed to the dividend-price ratio.⁸ As the figure shows, repurchases account for most the sharp drop in the unconditional premium of the late nineties.

5.2.2. *Equity Premium and Macroeconomic Risk.*

In this section we explore the relationship between expected returns and selected macroeconomic variables. We summarize the co-movements of our measure of the equity premium and variables that should contain, or have been found to contain information about the premium. We run exploratory regressions here, and while we recognize that the regressions may be subject to some measurement error and do not have a full structural interpretation, we claim that this is a useful exercise that provides some new empirical evidence. Further, the variables we include in the regressions are justified by existing literature that tries to build a bridge between the behavior of the stock market and macroeconomic risk. As some of the variables depend on our estimation procedure, they may have some complicated time series properties, so the standard errors

⁸To calculate the payout ratio, we adjust the S&P 500 dividend-price ratio using the data from Table I in Grullon and Michaely (2002). The assumption is therefore that the same dividend-payout ratio in the sample used by Grullon and Michaely is representative for the S&P 500

are Newey-West autocorrelation-corrected for lags of ten periods in all the regressions discussed below.

Also, we use posterior medians of estimated quantities such as equity premia and volatility of consumption growth, i.e. we do not conduct a fully Bayesian analysis. In other words, we view the prior on the parameters, and the posterior, as theoretical tools to obtain useful economic objects, without attaching to them subjective significance.⁹

We first look at the relationship between the equity premium and the conditional standard deviation of consumption growth in greater detail. Results are presented in Table 4. Bansal and Yaron (2004) show that for a class of exchange economies (as in Lucas, 1978) with Epstein-Zin-Weil preferences and conditionally heteroskedastic consumption growth, the equity premium is an affine function of the volatility of consumption growth:

$$(16) \quad E_t h_{t+1} - r_{t+1}^f = \gamma_0 + \gamma_1 \sigma_t^2(\Delta c_{t+1}).$$

If the equity premium EP_c is regressed on consumption volatility as in (16), the coefficient is 1.29 and it is highly significant, with an R^2 of 41% in annual data. The same regression using quarterly data produces similar results, with an R^2 of 26%. These two regressions confirm the discussion of section 5.2.1 on the long-run relationship between macroeconomic risk and the equity premium.

In the remaining regressions of Table 4, we use the quarterly data set to provide evidence that there is also a relationship between the conditional standard deviation of consumption growth and asset prices at higher frequencies. Notice first that the quarterly volatility measure shows much more variation than its annual counterpart (see Figures 4 and 5). This is consistent with the ARCH literature on excess returns. Changes in the conditional variance of stock returns are most dramatic in monthly data, but weaker at lower frequencies.

Consider the following three regressions:¹⁰

$$(17) \quad d_t - p_t = b_0 + b_1 \sigma_{c,t-1} + u_{1,t}$$

$$(18) \quad E_t[h_{t+1}] = c_0 + c_1 \sigma_{c,t-1}$$

$$(19) \quad \sigma_{c,t+j} = d_0 + d_1(d_t - p_t) + u_{2,t},$$

⁹This is consistent with the discussion in Bickel and Doksum (2000).

¹⁰Apart from being motivated by the findings above, the same projections are implied by the model studied by Bansal and Yaron (2004).

where $E_t h_{t+1}$ denotes a conditional expected real return on stocks. In (17)-(18), positive b_1 and c_1 indicates financial markets dislike uncertainty, and so dividend-prices should increase as prices decrease, and returns should be expected to go up. The last equation provides information on the extent to which volatility is long lasting and time-varying. If volatility is time varying and has some persistence, this should be captured by valuation ratios and the coefficient should be positive and significant. Bansal, Khatcharian, and Yaron (2003) (henceforth BKY) also run the regressions (17) and (19) but, because they do not have a measure of expected returns, they cannot estimate the projection in (18).

We run the first regression on the dividend-price ratio in deviation from the time varying unconditional mean, given our assumption of time-varying parameters. This will also remove the low frequency component. The coefficient on consumption-growth volatility is positive and significant, even after correcting for autocorrelation, and even if we include past values of the dividend price ratio, to limit the possibility of spurious results. This corroborates the results in BKY. Notice that although consumption growth volatility is estimated, in this particular model the t -statistic is still valid asymptotically for the test of $b = 0$.¹¹

We run the second regression for both ER_c (row 4 in Table 4) and $ER_c - ER_u$ (on row 5), to filter out a low frequency component in expected returns. We find c_1 to be positive and significant. It is worth emphasizing that a significant coefficient in the regression of $ER_c - ER_u$ implies that the conditional standard deviation is important in explaining *short-run* changes in expected returns. The coefficient is also economically significant. A 1% increase in volatility of consumption growth implies a 1.156% overall increase in expected real return on stocks, and .692% in terms of deviation from the time t unconditional return.

For the third regression, (17), which summarizes the ability of the valuation ratio to predict future volatility, we find a positive and significant coefficient for 1, 4, and 8 lags.¹² Notice that this last regression, which suggests that volatility of consumption growth is priced, does not suffer from measurement error bias, as the estimated variable is on the left hand side. If anything, this error should inflate standard errors.

¹¹This is because the asymptotic bias in the standard errors with estimated regressors depends multiplicatively on b , as shown in Pagan (1984), so it disappears under the null of $b = 0$.

¹²Lags are chosen to match the work by BKY. Notice that coefficients in BKY are negative, because they use $p_t - d_t$ the negative of what we use here.

In Table 5 we add explanatory variables to the regressions of excess returns. These provide robustness tests on the importance of consumption growth-volatility, and allow us to explore the role of other variables for movements in the premium. We use both quarterly and annual data, though we focus our discussion on the quarterly data set, as it is the data set on which the literature on macroeconomic risk has mostly focused on. For completeness, the tables present regressions using both EP_c and $EP_c - EP_u$, but we focus on the latter variable in which the low frequency component is filtered out (this also reduces the risk of spurious regressions).

The first regression of $EP_c - EP_u$ (the left panel of the two tables) includes expected inflation, dividend growth (in deviation from the time-varying unconditional mean), a measure of consumption-wealth ratio in logs (cay_t), the payout ratio (dividend over earnings), and the volatility of consumption growth estimate from our model. The reason for including inflation or expected inflation in a regression of the equity premium comes from the large body of research which links inflation to the premium through the interaction of inflation and taxation, inflation and risk, or inflation and money illusion.¹³

The consumption wealth ratio is the cay_t variable computed by Lettau and Ludvigson (2001). Data on this variable, both annual and quarterly, can be downloaded from their webpages. The idea behind the variable is that deviations of consumption from a long-run trend with wealth should contain information about expected returns. Under some assumptions, Lettau and Ludvigson compute the variable from a co-integrating relationship between consumption and wealth derived from the intertemporal budget constraint.

It is plausible to think that the payout ratio should predict expected returns. Payout ratios move in response to cyclical variation in earnings and to permanent changes in future expected growth. If payout ratios are higher because companies anticipate higher future growth, then this information should be reflected in prices and expected returns.

The R^2 is high and above 60% in quarterly data. Evidence on expected inflation is mixed, it is positive and significant in quarterly data, but not significant in annual data.

The payout ratio is not significant in the quarterly regression, but significant in annual data and significant at the 10% level in the quarterly regression of EP_c . This confirms

¹³That investors may be comparing nominal rates on bonds to dividend yields for stocks was argued by Modigliani and Cohn (1979), who predicted that if inflation came down, the equity premium would also decrease. The prediction may have held in the eighties, during which we assist to both declining inflation and premium.

the idea that low frequency movements in the payout ratio forecast future returns, whereas cyclical fluctuations do not, as they are mostly driven by cyclical variation in earnings.

The cay_t variable is positive and significant in both data sets, corroborating the results in Lettau and Ludvigson (2001). The regression provides an important robustness test of the effect of consumption-growth volatility on movements of the equity premium. The standard deviation of consumption growth is positive and significant. Also, notice that the size of the coefficient (.56) is not significantly altered by the inclusion of other control variables.

The cay_t variable, being derived from an intertemporal budget constraint, does not impose any particular asset pricing theory. The fact that volatility of consumption growth is still significant means that the variable captures additional information about the premium. This is important because it means that direct testing of asset pricing models as in Hansen and Singleton (1983) or Epstein and Zin (1991) may understate the importance of consumption risk.

6. ECONOMIC IMPORTANCE OF THE TIME VARYING VAR

We presented statistical evidence of parameter variation in Section 2, where we conducted stability tests on VAR equations. A natural question is whether parameter time variation is important economically. In this section we confront the question by presenting summary statistics from the posterior distribution of the difference in the price-dividend ratio implied by the time-varying VAR and the constant parameter VAR.

The time-varying parameters VAR (TVPVAR) nests the constant parameter VAR as a special case. If modeling time variation turns out to be economically unimportant, modeling expected returns by means of a constant parameter VAR would be a simpler way to calculate the cost of capital. So, how much do we lose by using a simple constant parameter VAR if the true model is statistically better represented by a VAR with evolving parameters?

In Section 5.2 we derived our measure of expected returns by inferring it from the dynamic growth model: Given the dividend price ratio and expected future dividend growth, we used the model to calculate expected returns. So, if we use our measure of expected returns in the dynamic growth model and use expected future dividend growth from the TVPVAR, the implied divided price ratio (log of it) will be equal to

the dividend price ratio in the data. But if we use the expected return from a constant parameter VAR in the dynamic growth model with time varying parameters as in (13), the implied price would be different. Under the null hypothesis that the two models are equivalent, the pricing errors should be small and should not have a systematic component. If the two models are not equivalent, then the pricing errors would be large and would have a systematic component.

For each draw from the parameter posterior, we calculate the price implied by the constant parameter model. I.e., we use the time-varying parameters to calculate k_t , ρ_t , and expected dividend growth in (13), but a constant parameter VAR for expected returns (for each draw of the parameters). This yields a $\hat{\delta}_t$. Statistics about the difference, $\hat{e}_t \equiv \hat{\delta}_t - \delta_t$, scaled by δ_t , is what we report in Table 6.¹⁴ For each series of \hat{e}_t we calculate the median error, the the median absolute error, and the first autocorrelation. Then we calculate the 5%, 25%, 50%, 75% and 95% quantiles.

The median absolute error is large. The 5% quantile is 26%. I.e., by using a constant parameter VAR one would make a median pricing error of at least 26% with 95% probability. The median pricing error is also large, and at least 50% of the time it is negative, meaning the constant parameter expected returns would underprice stocks by at least 14% at least 50% of the time. The posterior distribution of the autocorrelation coefficient tells us that the pricing errors are strongly positively correlated. The median autocorrelation coefficient is 15%, the 75% quantile is 64%, and the correlation coefficient is positive 93% of the times.

Figure 6 shows the price dividend ratio and the median price dividend ratio implied by the constant parameter VAR. In terms of prices, the difference is large, and it is quite persistent. The median absolute error is 33% for this series, and the autocorrelation coefficient is 98%.

We take the results in Table 6 and Figure 6 seriously as indicating that the constant parameter VAR is missing an important component of prices.

¹⁴We are assuming that the true model is the TVPVAR here. Given the statistical evidence in Section 2 this seems the appropriate thing to do. Since the TVPVAR includes the constant parameter as a special case, even if the true model was a constant parameter VAR the errors should be small and should not contain a predictable (systematic) component. If that was the case, one should reject the constant parameter as the true model.

7. SUMMARY AND CONCLUSIONS

This paper estimates expected returns and expected equity premia in U.S. stock markets using a valuation formula and allowing both the simultaneous relation between dividend growth, consumption growth, and dividend-prices, and their dynamic lag structure to vary over time. We motivate the importance of time-varying parameters both theoretically and statistically.

In particular, the paper focuses on the size and movements of the equity premium over the last 75 years, and on the relationship between the equity premium and sources of macroeconomic risk. We extend and confirm previous work on the declining equity premium and run exploratory data analysis in search of clues about factors determining movements in the equity premium. We find that the equity premium has declined and it is now much closer to levels predicted by standard consumption models. This implies that asset pricing models that aim at understanding the relationship between risk and returns should try to replicate aggregate stock market features matching a value closer to 3% rather than 6%.

A low estimate for the equity premium has important consequences for portfolio allocation decisions, the cost of capital, and how much of Social Security funds should be put into stocks. Jagannathan McGrattan and Scherbina (2000) indicate institutional changes occurred in the U.S. in the last 30 years, technological improvements in particular. Other reasons for a lower premium may include greater opportunities for portfolio diversification. Since the 1970's, there have been enormous changes in the financial instruments available to the public: think of money market funds, floating-rate notes, index mutual funds, emerging market funds, equity REITs, zero-coupon bonds, S&P index futures and options, and many more. These, as well as technological advances and a decrease in macroeconomic risk, have increased participation in equity markets, and equity premia have decreased.

A low estimate relative to ex-post excess returns makes one wonder whether the calculations of this paper are somehow biased downward. Evidence from the book to market value ratio in post war data support our low estimate. Fama and French (2002) report that the book to market ratio values are on average 0.66 for the period 1950-2000, and in only 6 years the ratio is above 1. The average real income return on book equity is 7.60%, less than the average real stock return for the period, which is about 9%. If the

expected return was indeed greater than the average real income return on equity, then the present value of the investment would be negative, because the expected return is a measure of the cost of capital. Then why would the book to market ratio be less than one? A high cost of capital would also make investment unprofitable. A lower expected return on the other hand makes investment profitable and is more consistent with book to market ratios less than one.

Our second finding is a statistically and economically significant relationship between the premium and volatility of consumption growth, which underscore the role of consumption growth for risk. Consumption growth has been given a relatively low weight in asset pricing because it is thought to be too smooth and too close to i.i.d. in quarterly data. As Bansal and Yaron (2002) show, a small persistent component in the volatility of consumption growth can be hardly detectable in the data and yet have important implications for asset prices. Our results indicate the presence of persistent, and economically meaningful, time-varying conditional volatility of consumption growth.

APPENDIX A. CONVERGENCE AND ROBUSTNESS

A.1. Convergence Checks.

I perform the usual informal checks (changing the starting point in the Markov chain, different number of draws, burning periods and “thinning” ratios) and some more formal ones, such as Geweke (1992) χ^2 convergence diagnostic (CD) and relative numerical efficiency (RNE), and a test on the standardized CUMSUM statistic proposed by Yu and Mykland (1994).¹⁵ The χ^2 statistic of Geweke compares the estimate of a posterior mean from the first N_A draws with the estimate of the last N_B draws of the chain. If the two subsamples are well separated, they should be independent. If the number of draws is sufficiently large, the following statistics is a χ^2

$$CD(\theta) = \frac{(\bar{\theta}_A - \bar{\theta}_B)^2}{nse_A^2 + nse_B^2},$$

where $\bar{\theta}_i$ is the estimate of the posterior mean of the parameter, and nse_i^2 is its numerical standard error, formed from subsample i . In other words, $\bar{\theta}_i$ is the sample mean of Monte Carlo draws, and nse_i^2 is an estimate of the variance of the sample mean. Given that the Monte Carlo sample is a Markov chain, it is not an independent sample, and the variance of the sample mean is function of the variance and autocovariance of the process generating the draws. The variance of the sample mean is therefore estimated with a spectral estimator evaluated at zero, $S_\theta(0)$. In this application, the number of draws is 10,000, N_A is the first 2,000 draws, N_B is the last 5,000 draws, and the variance of the two sample means is estimated using Newey-West weights on a number of covariances equal to 10% N_i . The Newey-West estimator corresponds to the Bartlett Spectral estimator (see Hamilton p.167).

Geweke’s RNE is a measure of efficiency of the algorithm, relative to the i.i.d. case. If the Monte Carlo sample is i.i.d., the variance of the sample mean is the variance of the population divided by the sample size, say γ_0/N . Given dependence of the draws, the variance is instead $2\pi S_\theta(0)/N = (\gamma_0 + 2 \sum_j \gamma_j)/N$. The relative numerical efficiency is

$$RNE(\theta) = \frac{\gamma_0}{2\pi S_\theta(0)} = (1 + 2 \sum_j \rho_j)^{-1}$$

Notice that this statistic is not bounded between zero and one, and values greater than one indicate that the variance of the sample mean in the chain is smaller than in the i.i.d. case. This is desirable and means that convergence can be achieved with a relatively smaller number of draws.

Given the N draws, a standardized version of Yu and Mykland (1994) convergence check is the statistic

$$CS_t(\theta) = \frac{\frac{1}{t} \sum_{n=1}^t \theta^{(n)} - \mu_\theta}{\sigma_\theta}$$

¹⁵Gauss routines that perform the checks described in this section are available upon request.

where μ_θ and σ_θ are the empirical mean and standard deviation of the N draws. If the Markov chain converges, the graph of $CS(\theta)$ against t should converge smoothly towards zero. This statistic is performed on randomly selected parameters.

Results from $CD(\theta)$ and $RNE(\theta)$ are reported in Table 7. They are quite satisfactory, the algorithm is efficient according to RNE for most sets of parameters, although not as efficient in the case of the hyperparameters. A value above 0.5 for the RNE is very satisfactory. For the matrix V , the value of 0.06 is low, but acceptable, given the high number of iterations in the Markov chain. The value is similar to Primiceri's (2003). He reports $1/RNE$ equal to 18 for hyperparameters, which correspond to 0.056. A value of $RNE = 0.06$ means that the algorithm requires 16.7 times more iterations than the i.i.d. case, for a given level of precision. Given the fact that our results are unchanged by changes in starting points of the chain, and greater number of draws, we believe that 10,000 draws are enough and conclude that the algorithm is converging to the ergodic distribution. This is confirmed by the CD statistics and graphs of randomly selected CS_t (not reported). The CD statistics are all well below the χ_1^2 critical value of 3.84. Medians are between 0.36 and 0.77 and convergence is not rejected for even a single parameter.

A.2. Robustness to Prior Calibration.

Robustness checks are conducted for the parameters k_Q, k_S, k_W . Given the great number of parameters, the robustness checks are based on the long run values of the VAR $\mu_{t|T}$ and by looking at the impact on the diagnostic statistics of the Markov chain.

Varying k_Q does not affect the behavior of the long run values of the VAR variables, unless extreme values are used. With values of the order of magnitude of the benchmark level (0.025), from 0.01 to 0.05, both long-run posteriors and diagnostic statistics are not affected. The long run trajectory of dividend growth is affected by high values say greater than 0.1, and large values of k_Q produce implausible values for the long run trajectory of dividend yield. The long run of dividend growth becomes negative at some dates, meaning that time variation in the parameters captures noise. The condition on unit-root parameters does not impact the amount of time variation in θ 's in a sensible way, if anything it reduces it, and it does not affect time variation in Ω_t . Changing k_W or k_S does not have any particular impact on the behavior of long run values and on convergence properties of the algorithm. The most important parameter is clearly k_Q . The reason is that Q affects the amount of time variation and it is a matrix of big dimension, possibly singular. The fact that the model behaves sensibly with a value of k_Q close to what is used in other research with quarterly data (Cogley and Sargent, 2001, 2002, and Primiceri, 2003) is taken as a good sign.

APPENDIX B. THE GIBBS SAMPLER

B.1. Step1: Coefficient States.

Conditional on Y^T, A^T, Σ^T and hyperparameters V , the state equation describing the transition of θ_t is linear with normal innovations. We can therefore use the Kalman

filter to infer the history of θ_t . As shown in Fruhwirth-Schnatter (1994) and Carter and Kohn (1994), the density $f(\theta^T|Y^T, A^T, \Sigma^T, V)$ can be factored as:

$$(20) \quad f(\theta^T|Y^T, A^T, \Sigma^T, V) = f(\theta_T|Y^T, A^T, \Sigma^T, V) \prod_{t=1}^{T-1} f(\theta_t|\theta_{t+1}, Y^t, A^T, \Sigma^T, V),$$

where Y^t is the history of y_t up to date t . Given normality of ν_t in the state equation $\theta_t = \theta_{t-1} + \nu_t$, the conditional distributions $f(\theta_t|\theta_{t+1}, Y^t, A^T, \Sigma^T, V)$ will be Gaussian. Hence, it is enough to update their conditional means and variances. Starting from period 0 with our prior $p(\theta_0) \sim N(\bar{\theta}, \bar{P})$ let

$$\begin{aligned} \theta_{t|t} &= E(\theta_t|Y^t, A^T, \Sigma^T, V), \\ P_{t|t-1} &= \text{Var}(\theta_t|Y^{t-1}, A^T, \Sigma^T, V), \\ P_{t|t} &= \text{Var}(\theta_t|Y^t, A^T, \Sigma^T, V) \end{aligned}$$

represent conditional means and variances. These are computed recursively from the prior by iterating on

$$\begin{aligned} K_t &= P_{t|t-1}X_t'(X_t'P_{t|t-1}X_t + R)^{-1}, \\ \theta_{t|t} &= \theta_{t|t-1} + K_t(y_t - X_t'\theta_{t-1|t-1}), \\ P_{t|t-1} &= P_{t-1|t-1} + Q, \\ P_{t|t} &= P_{t|t-1} - K_tX_t'P_{t|t-1}. \end{aligned}$$

The matrix K_t is the Kalman gain (see Hamilton, 1994). At the end of the sample, these iterations yield the conditional mean and variance for the terminal state $f(\theta_T|Y^T, A^T, \Sigma^T, Q) = N(\theta_{T|T}, P_{T,T})$. To find the remaining terms in (20), notice that we need mean and variances of θ_t conditional on θ_{t+1} and the history up to time t .¹⁶ Let

$$\begin{aligned} \theta_{t|t+1} &= E(\theta_t|\theta_{t+1}, Y^t, A^T, \Sigma^T, V), \\ P_{t|t+1} &= \text{Var}(\theta_t|\theta_{t+1}, Y^t, A^T, \Sigma^T, V) \end{aligned}$$

represent estimates of mean and variances based on the full set of data. Because of the Gaussian assumption, these can be calculated by going backwards from T as

$$(21) \quad \begin{aligned} \theta_{t|t+1} &= \theta_{t|t} + P_{t|t}P_{t+1|t}^{-1}(\theta_{t+1} - \theta_{t|t}), \\ P_{t|t+1} &= P_{t|t} - P_{t|t}P_{t+1|t}^{-1}P_{t|t}. \end{aligned}$$

So the remaining elements in the factorized posterior are

$$(22) \quad f(\theta_t|\theta_{t+1}, Y^t, A^T, \Sigma^T, V) = N(\theta_{t|t+1}, P_{t|t+1}).$$

¹⁶As Cogley and Sargent pointed out, this is not the Kalman smoother, which would be $E(\theta_t|Y_{T+1})$, see Hamilton, 1994.

Accordingly, a random sequence of parameters may be drawn from the prior at time 0 by using the Kalman filter to find $f(\theta_T|Y, V) = N(\theta_{T|T}, P_{T,T})$, draw from this distribution, then draw θ_{T-1} using (21) to calculate the mean and the variance. Then, iterate backward to the beginning of the sample.

B.2. Step 2: Covariance States.

The system of equations (4) can be written as

$$(23) \quad A_t(y_t - X_t'\theta_t) = A_t u_t = \Sigma_t^{1/2} \varepsilon_t,$$

The first row in (23) is just $u_{1,t} = \sigma_{11,t} \varepsilon_{1,t}$, whereas the two other equations in our model are given by

$$(24) \quad \begin{aligned} u_{2,t} &= -a_{21,t} u_{1,t} + \sigma_{2,t} \varepsilon_{2,t} \\ u_{3,t} &= -a_{31,t} u_{1,t} - a_{32,t} u_{2,t} + \sigma_{3,t} \varepsilon_{3,t} \end{aligned}$$

Conditional on Y^T, θ^T, Σ^T the model given by (24) and (6) has a Gaussian but non linear state space representation, the non-linearity given by the presence of contemporaneous u 's in the observation equations. However, under the additional assumption of S being block diagonal, the problem is solved by applying the Kalman filter and the backward recursion equation by equation. The procedure is as above, though now we recover

$$\begin{aligned} \alpha_{i,t|t+1} &= E(\alpha_{i,t} | \alpha_{i,t+1}, Y^t, \theta^T, \Sigma^T, V) \\ \Lambda_{i,t|t+1} &= \text{Var}(\alpha_{i,t} | \alpha_{i,t+1}, Y^t, \theta^T, \Sigma^T, V), \end{aligned}$$

where $\alpha_{i,t}$ is the i -th block of α_t , corresponding to the coefficients of the i -th equation in (24). As in the first step of the Gibbs sampler, $\alpha_{i,t}$ can be drawn recursively from a $N(\alpha_{i,t|t+1}, \Lambda_{i,t|t+1})$. Here is where we make use of the block diagonal S . If the coefficients of the contemporaneous relations among variables were not assumed to evolve independently in each equation, it would not be possible to apply the recursion equation by equation.

B.3. Step 3: Volatility States.

Consider now the system of equations

$$(25) \quad A_t(y_t - X_t'\theta_t) = y_t^* = \Sigma^{1/2} \varepsilon_t$$

Conditional on θ^T, A^T , and Y^T , y_t^* is observable. This is a system of non linear measurement equations, but can be easily converted in a linear one by squaring and taking logs of every element of (25). This leads to the state space form

$$(26) \quad \begin{aligned} y_t^{**} &= 2h_t + e_t \\ h_t &= h_{t-1} + \eta_t \end{aligned}$$

where $y_{it}^{**} = \log[(y_{it}^*)^2 + c]$, $h_t = \ln \sigma_{it}$ and $e_t = \log \varepsilon_{it}$. The ‘‘offset’’ c is introduced to robustify the stochastic volatility estimator in the case of $(y_{it}^*)^2$ being very small (see Fuller (1996) p. 494-497). As in Kim, Shephard and Chib (1998), is set to 0.001. Notice that the system written in this form has a linear, but a non-Gaussian state space form. In fact, the e_{it} are independently distributed as $\log \chi^2(1)$. The procedure of Kim, Shephard

and Chib approximates the $\chi^2(1)$ with a mixture of normal distributions. Because the $\text{Var}(\varepsilon_t)$ is the identity, the same (independent) mixture of normal approximation for any element of e_t can be used. The approximation is written as

$$f(e_{j,t}) \simeq \sum_{i=1}^K q_i f_N(e_{j,t} | m_i - 1.2704, v_i^2),$$

where j refers to the j^{th} element of e ($j = 1, 2, 3$ in our case), and so the density of e_t is a mixture of K normal densities f_N with component probabilities q_i , means $m_i - 1.2704$, and variances v_i^2 . The constants q_i, m_i, v_i^2 are selected to closely approximate the exact density of e_t . Kim Shephard and Chib find that $K = 7$ is enough to match the first four moments of the distributions with their selected constants q_i, m_i, v_i^2 (see table 4 of Kim Shephard and Chib, 1998).

Define $s^T = [s_1, \dots, s_T]^T$, the matrix of indicator variables selecting at every point in time which member of the mixture approximation is to be used for each element of e . Conditional on θ^T, A^T, V , and s^T , the system in (26) is linear and Gaussian. Again, like in the previous steps of the sampler, we can recursively recover

$$\begin{aligned} h_{t|t+1} &= E(h_t | h_{t+1}, Y^t, A^T, \theta^T, V, s^T), \\ H_{t|t+1} &= \text{Var}(h_t | h_{t+1}, Y^t, A^T, \theta^T, V, s^T) \end{aligned}$$

and draw every h_t from $N(h_{t|t+1}, H_{t|t+1})$. Conditional on h^T and y^{**T} , the new s^T can be drawn by independently sampling each $s_{t,j}$ from the probability mass function

$$P(s_{t,j} = i | y_t^*, h_t) = q_i f_N(y_{t,j}^{**} | 2h_t + m_i - 1, 2704, v_i^2), \quad i = 1, \dots, 7, \quad j = 1, 2, 3.$$

B.4. Step 4: Hyperparameters given states.

The hyperparameters of the model are the diagonal blocks of V : Q, W , and the diagonal blocks of S . Conditional on θ^T, A^T, Σ^T , and Y^T , the likelihood of the disturbance vectors in each block is Gaussian by assumption and the prior for the hyperparameters is inverse-Wishart. When a Gaussian is combined with an inverse-Wishart, the posterior is inverse-Wishart too, so each square has an inverse-Wishart posterior distribution, independent of the other blocks. Conditional on θ^T, A^T, Σ^T , and Y^T , it is simple to draw from these inverse-Wishart posteriors because the innovations are observable. As an example, consider drawing from the posterior for Q using the innovations ν_t

$$p(Q | Y^T, \theta^T, A^T, \Sigma^T) = IW(Q_1^{-1}, T_1),$$

where $T_1 = T_0 + T$, $Q_1 = \bar{Q} + \bar{V}_T$, and \bar{Q}_T is proportional to the usual covariance estimator,

$$(1/T)\bar{Q}_T = (1/T) \sum_{t=1}^T \nu_t \nu_t'.$$

To simulate a draw from the inverse-Wishart $IW(S, df)$ we draw df independent vectors η_i from a $N(0, S)$ density, then $V = \left(\sum_{i=1}^{df} \eta_i \eta_i^T \right)^{-1}$ is a draw from an inverse-Wishart. This concludes the Gibbs sampler.

B.5. Summary.

To summarize, the Gibbs sampler iterates on 4 simulations, drawing states conditional on hyperparameters, and then hyperparameters conditional on states. After a "burn-in" period, the sequence of draws approximates a sample from the posterior $p(\theta^T, A^T, \Sigma^T, V|Y^T)$. The sampler can be summarized as follows:

- (1) Initialize A^T, Σ^T, s^T , and V .
- (2) Sample θ^T from $p(\theta^T|Y^T, A^T, \Sigma^T, V)$.
- (3) Sample A^T from $p(A^T|Y^T, \theta^T, \Sigma^T, V)$.
- (4) Sample Σ^T from $p(\Sigma^T|Y^T, \theta^T, A^T, s^T, V)$.
- (5) Sample s^T from $p(s^T|Y^T, \theta^T, A^T, \Sigma^T, V)$.
- (6) Sample V , by sampling Q, W, S from $p(Q, W, S, |Y^T, \theta^T, A^T, \Sigma^T)$
- (7) Go to 2.

REFERENCES

- [1] J. Bai and P. Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66:47–78, 1998.
- [2] Jushan Bai and Pierre Perron. Computation and analysis of multiple structural change models. manuscript, 23 pages, 2001.
- [3] Ravi Bansal, Varoujan Khatcharian, and Amir Yaron. Interpretable asset markets? manuscript, 39 pages, May 2003.
- [4] Ravi Bansal and Amir Yaron. Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance*, 59:1491–1509, 2004.
- [5] Peter J. Bickel and Kjell A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*, volume I. Prentice Hall, 2nd edition, September 2000.
- [6] Olivier Blanchard. Movements in the equity premium. *Brooking Papers on Economic Activity 2*, pages 75–138, 1993.
- [7] J. H. Campbell and R.J. Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1:195–227, 1988.
- [8] J. H. Campbell and R.J. Shiller. Stock prices, earnings, and expected dividends. *Journal of Finance*, 43:661–676, 1988.
- [9] J. Y. Campbell, A.W. Lo, and C. A. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, 1997.
- [10] C.K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81:541–553, 1994.
- [11] John Cochrane. *Asset Pricing*. Princeton University Press, Princeton, New Jersey, 2001.
- [12] Timothy Cogley and Thomas Sargent. Evolving post-World War II U.S. inflation dynamics. *NBER Macroeconomics Annual.*, 2001.
- [13] Timothy Cogley and Thomas Sargent. Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S. manuscript, 2002.
- [14] Werner DeBondt and Richard Thaler. Does the stock market overreact? *Journal of Finance*, 40:793–805, 1985.
- [15] L. Epstein and S. Zin. Substitution, risk aversion and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy*, 99:263–286, 1991.
- [16] Martin D. Evans. Dividend variability and stock market swings. *The Review of Economic Studies*, 65:711–740, 1998.
- [17] Eugene F. Fama and Kenneth R. French. Dividend yields and expected stock returns. *Journal of Financial Economics*, 25:3–43, 1988.
- [18] Eugene F. Fama and Kenneth R. French. The equity premium. *The Journal of Finance*, 57:637–659, 2002.
- [19] Sylvia Fruhwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–202, 1994.
- [20] G. Grullon and R. Michaely. Dividends, share repurchases, and the substitution hypothesis. *Journal of Finance*, 57:1649–1684, 2002.
- [21] Bruce E. Hansen. Testing for parameter instability in linear models. *Journal of Policy Modeling*, 14:517–533, 1992.
- [22] Lars Peter Hansen and Kenneth J. Singleton. Stochastic consumption, risk aversion and the temporal behavior of asset returns. *Journal of Political Economy*, 91(2):249–265, 1983.
- [23] R. Jagannathan, E.R. McGrattan, and A. Scherbina. The declining U.S. equity premium. *Federal Reserve Bank of Minneapolis Quarterly Review*, 24:3–19, 2000.
- [24] S. Kim, N. Shepard, and S. Chib. Stochastic volatility: Likelihood inference and comparison with arch models. *Review of Economic Studies*, 65:361–393, 1998.

- [25] M. Lettau and S. Ludvigson. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 56:815–849, 2001a.
- [26] Martin Lettau, Sidney C. Ludvigson, and Jessica A. Wachter. The declining equity premium: What role does macroeconomic risk play? manuscript, Dec 2003.
- [27] Robert Jr. Lucas. Asset prices in an exchange economy. *Econometrica*, 46:1429–1446, 1978.
- [28] R. Mehra and E. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15:145–162, 1985.
- [29] Richard Meinhold and Nozer Singpurwalla. Understanding the Kalman filter. *The American Statistician*, 37:123–127, 1983.
- [30] Andrew Pagan. Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, 25(1):221–247, Feb 1984.
- [31] Jonathan A. Parker and Christian Julliard. Consumption risk and the cross section of stock returns. *Journal of Political Economy*, 113(1):185–222, February 2005.
- [32] Giorgio E. Primiceri. Time varying structural vector autoregression and monetary policy. Princeton University, manuscript, 42 pages, 2003.
- [33] Tom Rietz. The equity risk premium: A solution? *Journal of Monetary Economics*, 21:117–132, 1988.
- [34] Jeremy J. Siegel. The shrinking equity premium. *Journal of Portfolio Management*, 26:10–17, 1999.
- [35] M Smith and R. Kohn. Parsimoiuous covariance matrix estimation for longitudinal data. *Journal of the American Statistical Association*, 97:1141–1153, 2002.
- [36] J.H. Stock and M.W. Watson. Asymptotically median unbiased estimation of coefficient variance in a time varying parameter model. *Journal of the American Statistical Association*, 93:349–358, 1998.
- [37] A. Timmermann. Structural breaks, incomplete information and stock prices. *Journal of Business and Economic Statistics*, 19:299–314, 2001.

TABLE 1. Time Series Properties

Variable	Mean	Standard Dev.	Max	Min
δ_t	-3.43	0.379	-2.78	-4.52
Δc_t	0.0198	0.0269	0.1	-0.0773
Δd_t	0.0105	0.0601	0.233	-0.253
Variable	PP Test Statistic	10% Critical Value	1% Critical Value	
δ_t	-1.2823	-2.5744	-3.4643	
Δc_t	-4.9577	-2.5744	-3.4643	
Δd_t	-9.8786	-2.5744	-3.4643	
Variable		OLS Estimates	Autocorrelations	Ljung-Box
δ_t	const	-0.0667	lag 1: 0.0055	8.594
	ϕ_1	1.0950	lag 2: -0.034	
	ϕ_2	-0.1130	lag 3: 0.045	
			lag 4: -0.039	
			lag 5: -0.014	
			lag 12: 0.09	
Δc_t	const	0.0157	lag 1: -0.012	16.29
	ϕ_1	0.1677	lag 2: -0.033	
	ϕ_2	0.0383	lag 3: -0.19	
			lag 12: 0.099	
Δd_t	const	0.0058	lag 1: -0.00028	12.76
	ϕ_1	0.2897	lag 2: -0.031	
	ϕ_2	0.1963	lag 3: -0.019	
			lag 12: 0.0012	
Ljung-Box	12 lags Critical Value	21.03		

Notes: δ_t is the log of dividend-price ratio. PP Test indicates the Phillips-Perron test for unit roots. Four lags of differenced variables are included in the Phillips-Perron test. The coefficients const, ϕ_1 , ϕ_2 are coefficients of an AR(2) model for the respective variables. The number of Lags included in the calculation of the Ljung-Box statistic is 12, and the 5% critical value of the χ^2_{12} is 21.03.

TABLE 2. Tests for deterministic breaks

Test	δ_t	δ_t	Δc_t	Δd_t	$ \Delta d_t $
	AR(2)	AR(2), $\Delta c_t, \Delta d_t$	AR(2)	AR(2)	AR(2)
SupF _T (1)	70.8168**	147.4168**	11.1902	9.9162	23.1189**
SupF _T (2)	74.2344**	113.2692**	19.4990**	13.0373*	16.4380**
SupF _T (3)	82.8105**	129.6741**	15.4024**	13.0187*	12.7310*
SupF _T (4)	72.7899**	101.4919**	13.1659*	10.7374*	17.1773**
SupF _T (5)	101.5448**	67.9402**	11.5209*	9.9750*	14.2001**
UDmax	166.1071**	270.9723**	19.4990**	13.0373	23.1189**
WDmax	346.5638**	509.3399**	26.3582**	22.9696**	26.8638**
SupF _T (2 1)	41.2933**	61.7071**	23.3230**	10.74	4.0465
SupF _T (3 2)	45.7777**	55.3380**	8.9841		4.8124
sequential proc. 5%	finds 4 breaks	finds 3 breaks	–	–	
sequential proc. 1%	finds 3 breaks	finds 3 breaks	–	–	

* means significance at 5%, ** at 1%

SupF_T(*j*) tests for zero breaks versus *j* breaks. UDmax and WDmax statistics test for zero breaks versus an unknown number of breaks. The SupF_T(*l* + 1|*l*) tests for *l* + 1 breaks given that *l* breaks are present.

TABLE 3. Ex-ante and Ex-post Returns

Subperiod	EP_c	EP_u	EP_u^*	Excess Returns
1928-1940	5.9%	5.1%	6.4%	3.0%
1928-1946	5.9%	5.0%	6.0%	6.1%
1946-1971	4.6%	3.9%	4.8%	8.6%
1971-1988	3.5%	2.4%	3.2%	3.2%
1976-1988	3.8%	2.3%	3.1%	4.5%
1988-2002	2.0%	1.8%	2.7%	6.7%
1994-2002	1.4%	1.5%	2.5%	5.7%
1995-1999	1.7%	1.5%	2.5%	21.0%

The table reports averages of the variables during the subperiods on the left. EP_c is our measure of the conditional equity premium. EP_u is the unconditional equity premium. * means that EP_u is adjusted for Jensen inequality. The last column has sub-samples averages of historical excess returns. Periods of high Excess Returns correspond to periods of low equity premia.

TABLE 4. Volatility of Consumption Growth and Asset Returns

	Dependent Variable	Regressor	Coeff.	Std.Err.	t-ratio	R²
1 eq (16)	EP_c -Annual	const.	-0.017	0.014	-1.192	0.41
		$\sigma_t(\Delta c)$	1.298	0.302	4.294	
		Var	Coeff.	Std.Err.	t-ratio	R²
2 eq (16)	EP_c -Quarterly	const.	-0.031	0.013	-2.465	0.26
		$\sigma_t(\Delta c)$	1.686	0.290	5.810	
		Var	Coeff.	Std.Err.	t-ratio	R²
3 eq (17)	$(d_t - p_t)$	const.	0.779	0.276	2.822	0.18
		$\sigma_t(\Delta c)(-1)$	20.228	6.837	2.958	
		Var	Coeff.	Std.Err.	t-ratio	R²
4 eq (18)	ER_c	const.	0.000	0.021	0.010	0.17
		$\sigma_t(\Delta c)(-1)$	1.156	0.505	2.291	
		Var	Coeff.	Std.Err.	t-ratio	R²
5 eq(18)	$ER_c - ER_u$	const.	-0.024	0.006	-3.744	0.34
		$\sigma_t(\Delta c)(-1)$	0.692	0.161	4.292	
		Var	Coeff.	Std.Err.	t-ratio	R²
6 eq (19)	$\sigma_t(\Delta c)$	const.	0.031	0.004	7.389	0.16
		$d_{t-1} - p_{t-1}$	0.006	0.003	2.099	
		Var	Coeff.	Std.Err.	t-ratio	R²
7 eq (19)	$\sigma_t(\Delta c)$	const.	0.034	0.003	12.085	0.11
		$d_{t-4} - p_{t-4}$	0.004	0.002	1.902	
		Var	Coeff.	Std.Err.	t-ratio	R²
8 eq (19)	$\sigma_t(\Delta c)$	const.	0.036	0.002	20.489	0.070
		$d_{t-8} - p_{t-8}$	0.003	0.001	1.765	
		Var	Coeff.	Std.Err.	t-ratio	R²

Notes: EP_c is the conditional equity premium, $\sigma_t(\Delta c)$ is the volatility of consumption growth, conditional on time t information. $d_t - p_t$ is the log of the dividend price ratio at time t . Standard error and t-ratios are 10-lag autocorrelation consistent. eq(j) refers to equation (j) in the text.

TABLE 5. Equity Premium and Fundamentals

Quarterly Regressions									
Variable	$EP_c - EP_u$				Var	EP_c			
	Coeff.	Std.Err.	t-ratio	R2		Coeff.	Std.Err.	t-ratio	R2
const.	-0.0222	0.0052	-4.31	0.66	const.	-0.02510	0.0135	-1.86	0.61
$E_t\pi_{t+1}$	0.1157	0.0423	2.73		$E_t\pi_{t+1}$	0.26638	0.1278	2.08	
Δd_t	-0.0062	0.0133	-0.46		Δd_t	0.04453	0.0241	1.85	
cay_t	0.1310	0.0375	3.50		cay_t	0.41302	0.1075	3.84	
payout	-0.0072	0.0085	-0.84		payout	0.03929	0.0244	1.61	
$\sigma_t(\Delta c)$	0.5565	0.1945	2.86		$\sigma_t(\Delta c)$	0.73421	0.5165	1.42	

Annual Regressions									
Variable	$EP_c - EP_u$				Variable	EP_c			
	Coeff.	Std.Err.	t-ratio	R2		Coeff.	Std.Err.	t-ratio	R2
const.	-0.0293	0.0079	-3.71	0.33	const.	-0.0290	0.0208	-1.40	0.64
$E_t\pi_{t+1}$	0.0162	0.0430	0.38		$E_t\pi_{t+1}$	-0.1582	0.0811	-1.95	
Δd_t	0.0097	0.0090	1.07		Δd_t	0.0253	0.0117	2.16	
cay_t	8.5E-06	3.6E-06	2.36		cay_t	6.5E-06	8.2E-06	0.79	
payout	0.0167	0.0065	2.55		payout	0.0415	0.0160	2.60	
$\sigma_t(\Delta c)$	0.6405	0.1615	3.97		$\sigma_t(\Delta c)$	1.1953	0.2856	4.18	

Notes: EP_c is conditional equity premium. EP_u is the unconditional equity premium. This is subtracted from EP_c in the left panel to filter out a low frequency component from the conditional equity premium. $E_t\pi_{t+1}$ is expected inflation. Δd_t is dividend growth. cay_t is the consumption-wealth ratio variable of Lettau and Ludvigson (2001). The variable payout is the ratio of dividend to earnings. $\sigma_t(\Delta c)$ is the standard deviation of consumption conditional on time t . MAC10, MAR10, MAC5, MAR5 are 10 and 5 periods moving averages of consumption (C) and returns (R) respectively. Standard errors and t -ratios are 10-lag autocorrelation consistent.

TABLE 6. Pricing Errors

	$\text{med}(\widehat{e}_t/\delta_t)$	$\text{med}(\widehat{e}_t/\delta_t)$	$\rho(\widehat{e}_t, \widehat{e}_{t-1})$
$q_{.05}$	0.26	-0.62	-0.02
$q_{.25}$	0.44	-0.38	0.10
$q_{.50}$	0.61	-0.14	0.15
$q_{.75}$	0.83	0.21	0.64
$q_{.95}$	1.43	1.43	0.98

Notes: med denotes the median over each time series, $\rho(\cdot)$ is the correlation coefficient, $\widehat{e}_t = \widehat{\delta}_t - \delta_t$, and $q_{.05}, \dots, q_{.95}$ denote quantiles of the posterior distributions of the quantities in the columns.

TABLE 7. Convergence Diagnostics

Parameter block	20 th Autocorrelation					
	Median	Mean	Min	Max	10%	90%
θ^T	0.01189	0.01213	-0.01817	0.05877	-0.0052	0.02846
A	0.01306	0.01827	-0.02812	0.08148	-0.00185	0.04305
Σ	-0.00126	-0.0011	-0.02842	0.02842	-0.01443	0.01156
V	0.11797	0.12056	-0.01664	0.42024	0.05679	0.18159
Parameter block	RNE					
	Median	Mean	Min	Max	10%	90%
θ^T	0.61489	0.65593	0.14499	2.45819	0.29221	1.08640
A	0.67781	0.63823	0.11269	1.82291	0.16013	1.24765
Σ	0.99855	1.09247	0.38967	3.17533	0.58277	1.69928
V	0.05998	0.08196	0.01341	1.0552	0.03679	0.11555
Parameter block	CD					
	Median	Mean	Min	Max	10%	90%
θ^T	0.51137	1.04140	3E-08	2.45819	0.02275	2.92498
A	0.36281	0.94815	1.79E-06	1.82291	0.01137	2.74856
Σ	0.76759	1.57906	6.63E-06	3.17533	0.03186	4.02212
V	0.50669	1.16395	0.00017	1.0552	0.01355	2.80397

Notes: RNE is Geweke (1992) relative numerical efficiency, a measure of the efficiency of the algorithm. CD is Geweke (1992) χ_1^2 convergence diagnostic.

FIGURE 1. Expected Returns and Expected Excess Returns, Annual Data

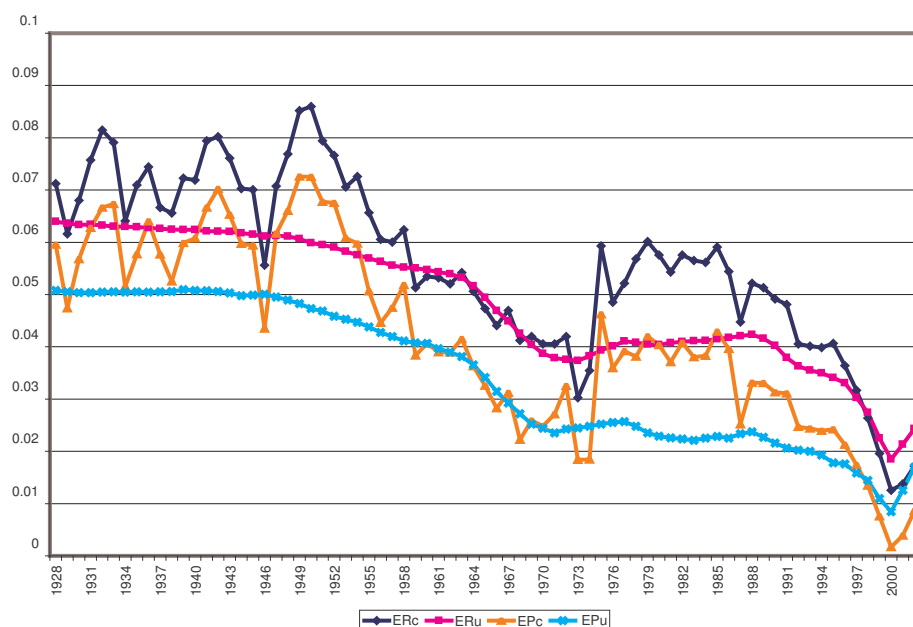


FIGURE 2. Expected Returns and Expected Excess Returns, Quarterly Data

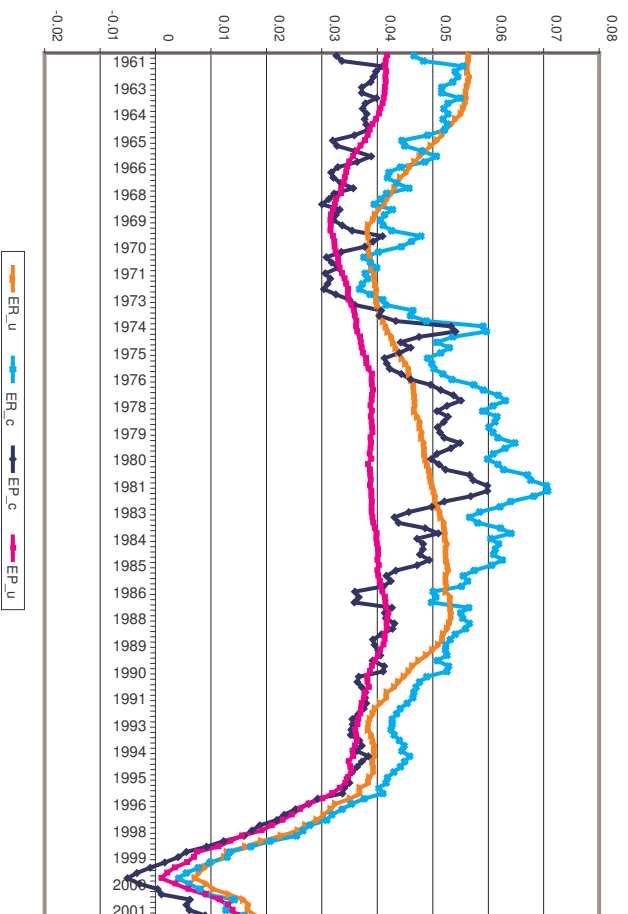


FIGURE 3. Adjusting For Repurchases in Quarterly Data

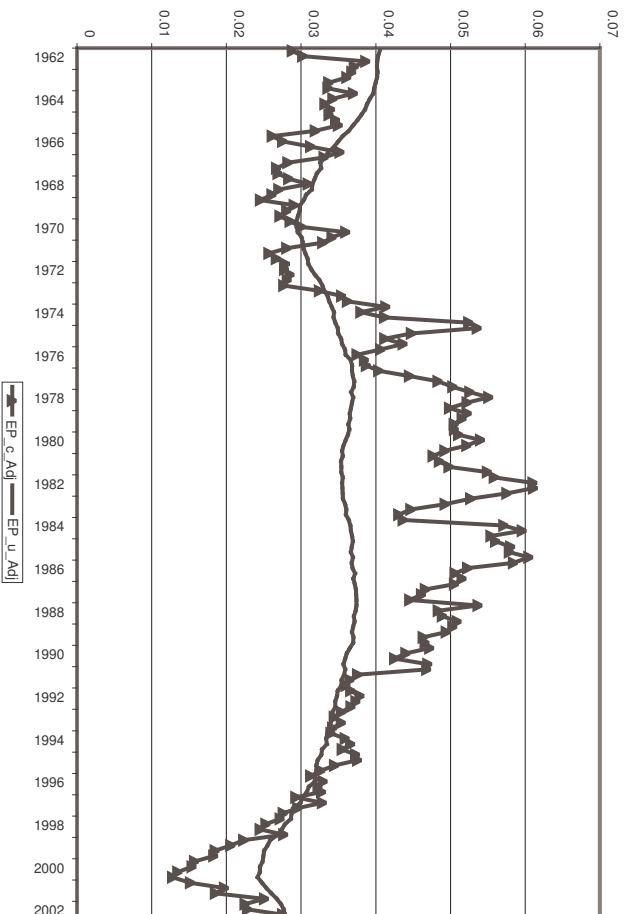


FIGURE 4. Volatility of Dividend Growth, Consumption Growth, and log Dividend-Prices (Annual)

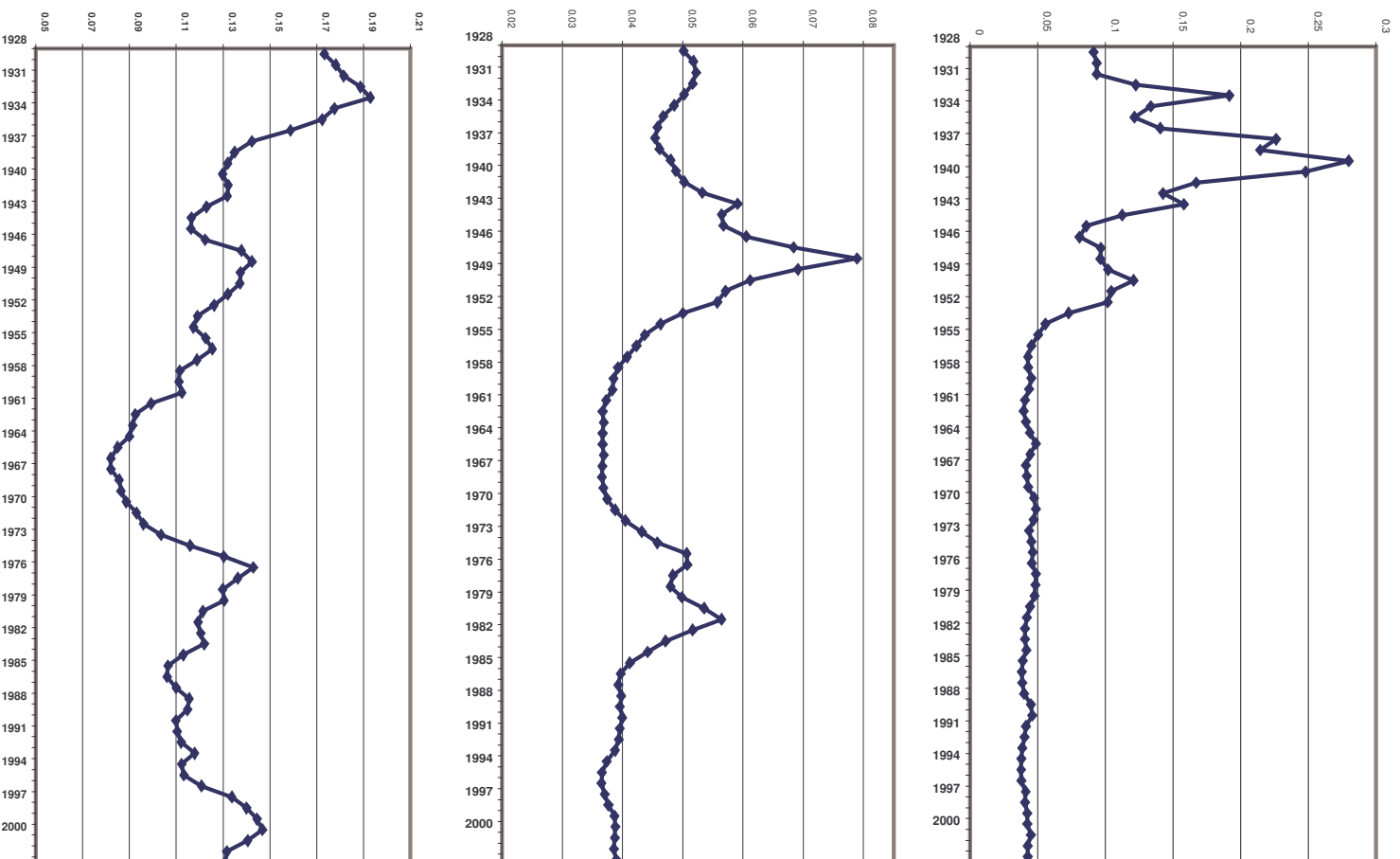


FIGURE 5. Volatility of Dividend Growth, Consumption Growth, and log Dividend-Prices (Quarterly)

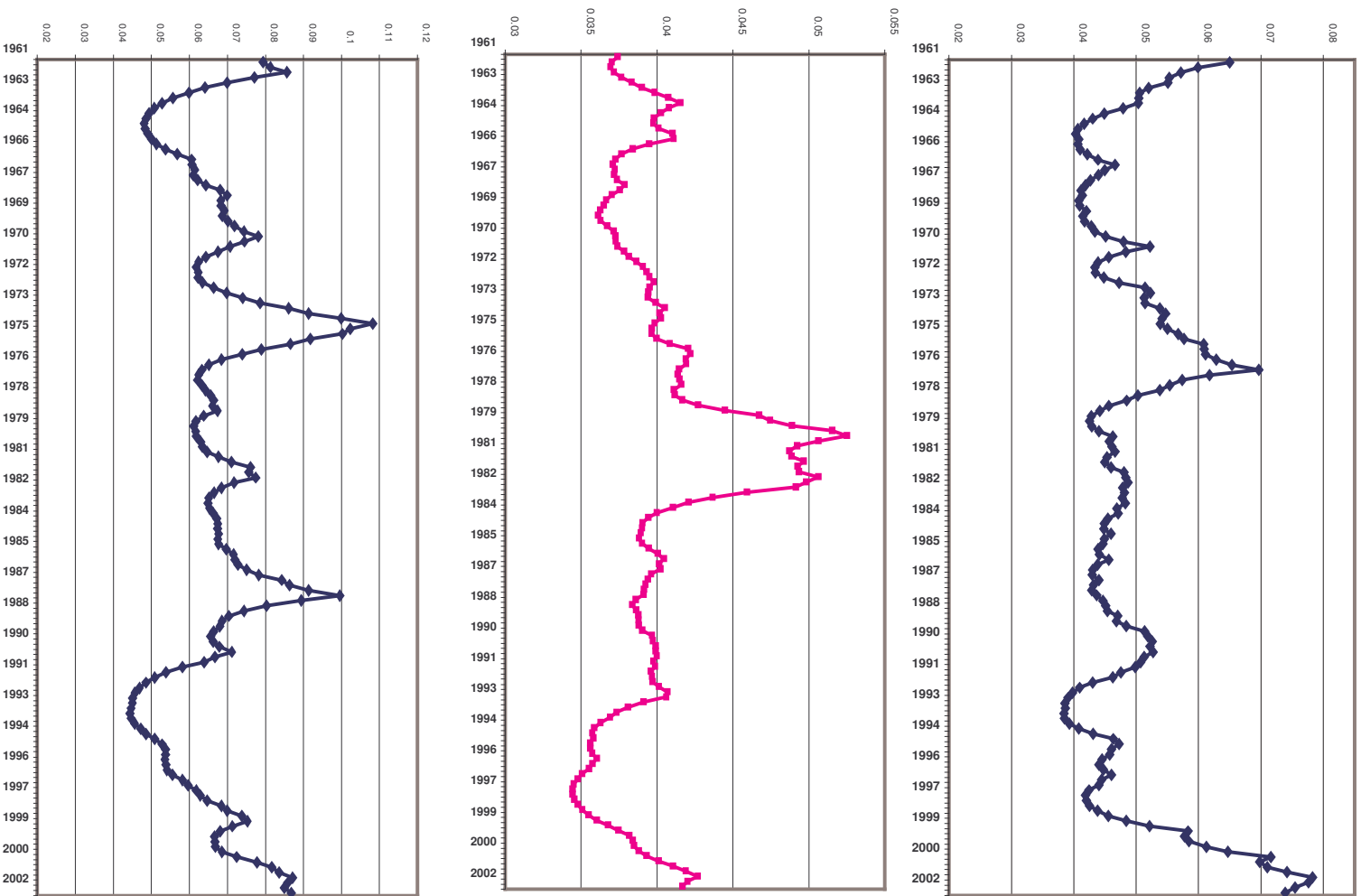


FIGURE 6. Price Dividend Ratios

