Internet Appendix for:

Sequential Learning, Predictability, and Optimal Portfolio Returns

MICHAEL JOHANNES, ARTHUR KORTEWEG,

and NICHOLAS POLSON *

Section I of this Internet Appendix describes the full set of parameter estimates for the four Bayesian models (CV, CV-DC, SV, and SV-DC). Section II shows the simulation results used to determine statistical significance. Section III describes the portfolio weights for both cash dividends and net payout yields. Section IV shows the conditional and unconditional excess market return distributions. Section V describes the particle filter algorithms for the four Bayesian models in detail, and Section VI explains the calculation of Savage Density ratios for the hypothesis tests in Figure IA.10.

*Citation format: Johannes, Michael, Arthur Korteweg, and Nicholas Polson, Internet Appendix for “Sequential Learning, Predictability, and Optimal Portfolio Returns,” *Journal of Finance*, DOI:10.1111/jofi.12121. Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the authors of the article.

I. Full Parameter Estimates

A. CV Model

The CV model jointly specifies the process for returns and payout ratios, as shown in equations (2) and (7) in the published article, where the volatilities are assumed to be constant. Figure IA.1 shows the sequential parameter estimates of the CV model, using the cash dividend yield as the predictor variable, x_t. For each parameter, we summarize the posterior distribution at each point in time via its mean (the solid line) and a $(1, 99)\,\%$ posterior probability interval (the shaded area). The figure reveals wide posterior bands at the beginning of the sample, consistent with the relatively uninformative priors from the generally short training sample. As new data arrive, the investor's view of the location and uncertainty of the parameters changes drastically. Most notably, the volatility in the returns equation declines substantially over time. This is merely an implication of the large fluctuations in market volatility, which are easiest to detect when the sample starts in a period of high volatility, such as the 1927 to 1930 period. Since nearly all studies begin in 1926, discarding the data and starting after World War II merely generates additional sample selection issues with regard to volatility. The perceived equity premium is $E\left[\alpha + \beta x_t | y^t\right]$. Interestingly, there is little significant variation in the location of α and β although the posterior confidence bands are naturally much tighter towards the end of the sample. The estimates of β follow the general pattern of the OLS estimates in Figure 1 of the main paper.

Regarding the dividend yield process, estimates of β_x trend upwards, although the

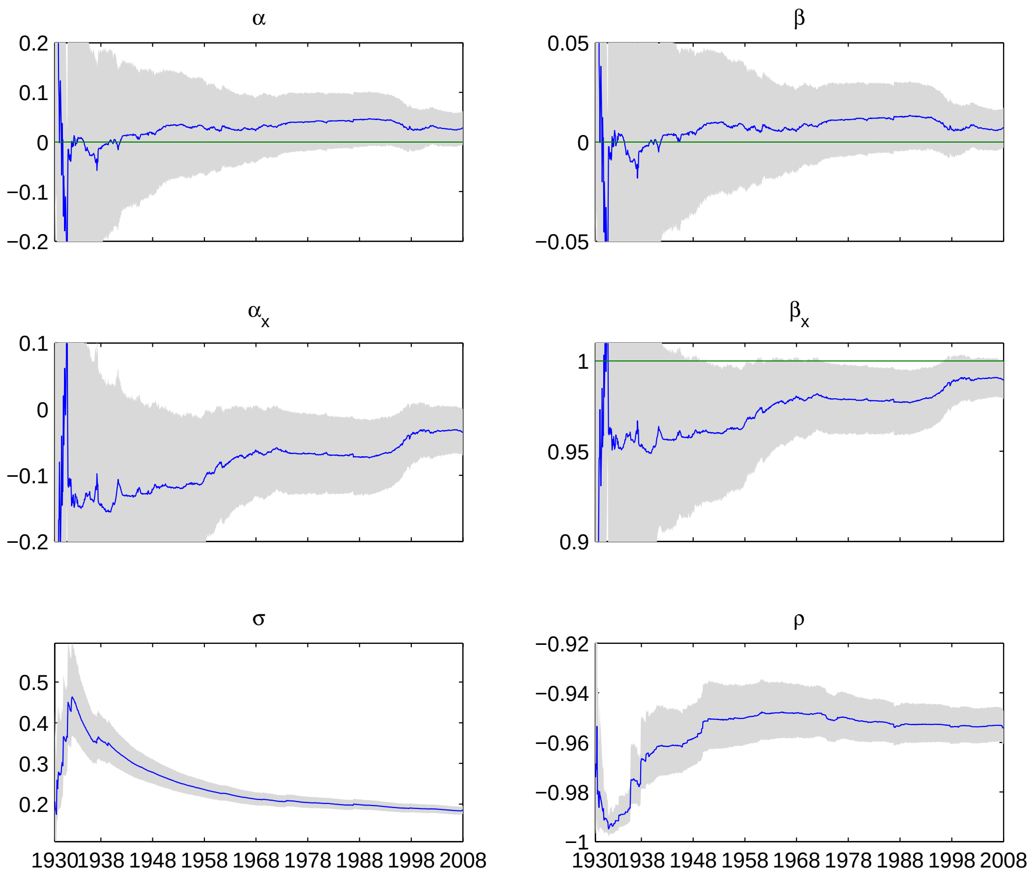

Figure IA.1. Sequential parameter estimates: CV model with dividend yield. This figure plots sequential parameter estimates for the CV model,

$$r_{t+1} = \alpha + \beta x_t + \sigma \varepsilon^r_{t+1}$$
$$x_{t+1} = \alpha_x + \beta_x x_t + \sigma_x \varepsilon^x_{t+1},$$

where r_{t+1} is the return on the market portfolio in excess of the risk-free rate from month t to month $t+1$. The predictor variable, x_t, is the traditional cash dividend yield. The shocks ε^r_{t+1} and ε^x_{t+1} are distributed standard normal with correlation coefficient ρ. Each panel displays the posterior means and (1,99)% posterior probability intervals (the grey shaded area) for each time period. Excess market return volatility, σ, is annualized.

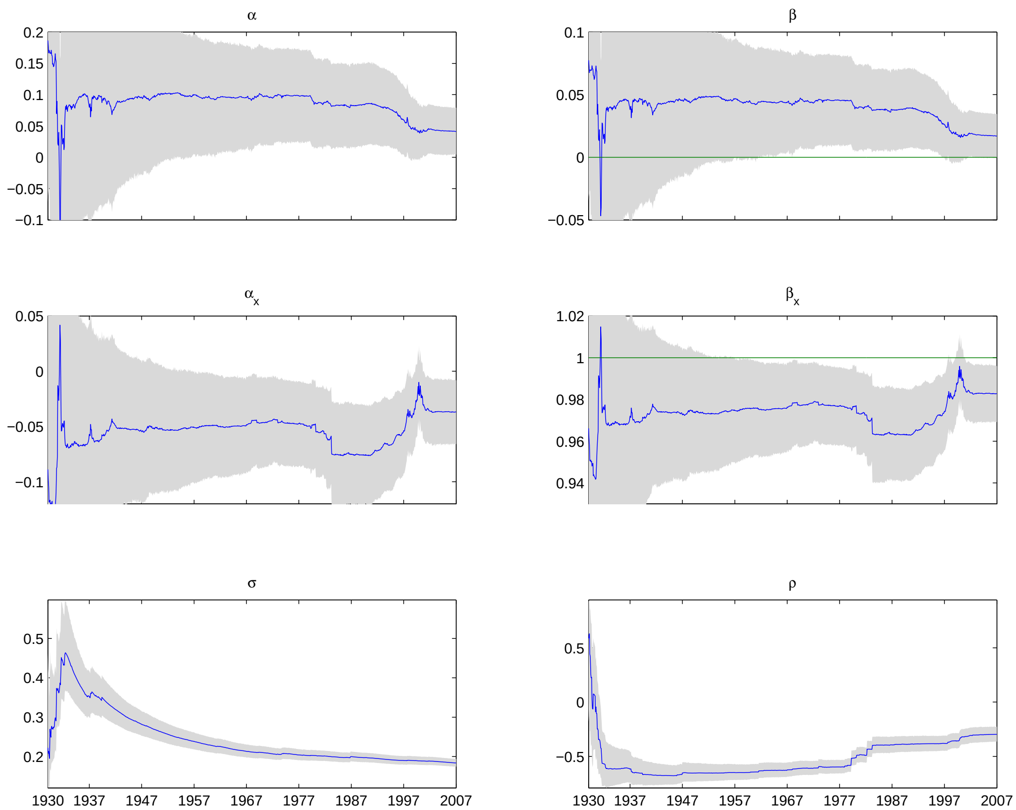

Figure IA.2. Sequential parameter estimates: CV model with net payout yield. This figure plots sequential parameter estimates for the CV model,

$$r_{t+1} = \alpha + \beta x_t + \sigma \varepsilon^r_{t+1}$$
$$x_{t+1} = \alpha_x + \beta_x x_t + \sigma_x \varepsilon^x_{t+1},$$

where r_{t+1} is the return on the market portfolio in excess of the risk-free rate from month t to month $t+1$. The predictor variable, x_t, is the net payout yield of Boudoukh et al. (2007). The shocks ε^r_{t+1} and ε^x_{t+1} are distributed standard normal with correlation coefficient ρ. Each panel displays the posterior means and (1,99)% posterior probability intervals (the grey shaded area) for each time period. Excess market return volatility, σ, is annualized.

movement is not large. Brav et al. (2005) present evidence that the speed of mean reversion for dividends has slowed in the second half of the 20th century, making dividend yields more persistent. If so, our results indicate the rate of structural change is rather slow, inconsistent with either a regime-switching or an abrupt structural break model.

Figure IA.2 uses the net payout yield as the predictor variable. There are a number of important differences. Although the estimates of α, β, and σ are quite similar, the estimates for the payout ratio series are not. There is an abrupt change in the parameters of the net payout process in the early 1980s, in particular, in α_x, β_x, and ρ. This result is suggestive of a structural break in the dynamics of net payouts, something more substantial than just parameter uncertainty. Interestingly, Boudoukh et al. (2007) formally test for a structural break and find no evidence; however, we use monthly data, whereas they focus on annual data. The source of the variation can be found in Figure 1 of the main paper, where we can see that in the early 1980s the net payout variable had a series of high frequency shocks. The net effect of making the process less persistent is that it reduces the autocorrelation (β_x). These high frequency fluctuations have an even greater impact on ρ, as the relatively stable link between payout ratio shocks and market returns is broken. The source of these shocks is a sudden increase in net repurchases, which no doubt corresponds to a structural economic change following the adoption in 1982 of SEC rule 10b-18, providing safe harbor from liability for firms repurchasing shares in accordance with the rule's conditions.

To formally assess the strength of predictability, Figure IA.3 summarizes the posterior probabilities in the benchmark model for tests of $\beta = 0$ and $\beta_{dp} = 1$ – the unit root case. Using our particle filter, we calculate the Bayes factor for $\mathcal{H}_0 : \beta = 0$ versus $\mathcal{H}_1 : \beta \neq 0$ as

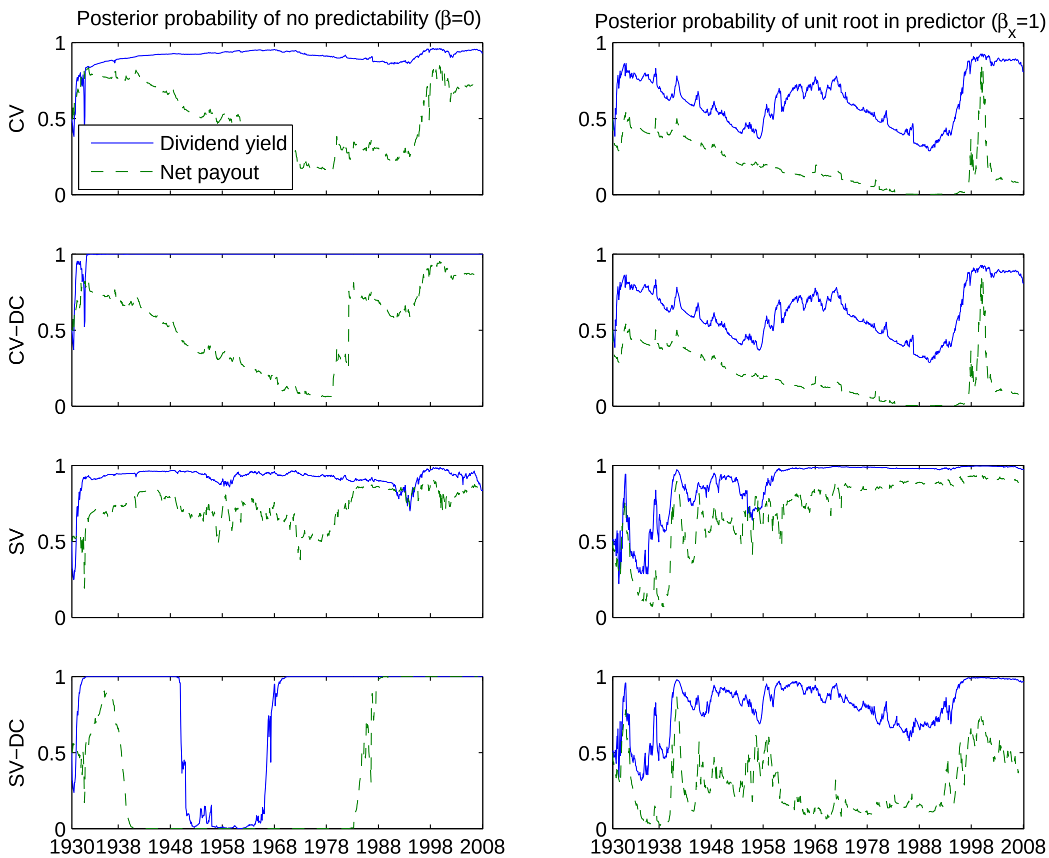

Figure IA.3. Hypothesis tests. This figure plots posterior probabilities for hypothesis tests of no predictability (left-hand-side plots) and a unit root in the predictor process (right-hand-side plots). The predictor variable is either dividend yield (solid line) or net payout yield (striped line). CV and SV represent models with expected return predictability and constant volatility (CV) and stochastic volatility (SV), respectively. DC stands for drifting coefficients and represents models where the predictability coefficient is allowed to vary over time. The null hypothesis for no predictability is $\mathcal{H}_0 : \beta = 0$ in the CV and SV models or $\mathcal{H}_0 : \beta_0 + \beta_t = 0$ in the DC and SV-DC models. The null hypothesis for a unit root in the predictor process is $\mathcal{H}_0 : \beta_x = 1$.

follows:

$$\mathcal{BF}_{0,1}^{t} = \frac{p\left(\beta = 0 | y^t, \mathcal{H}_1\right)}{p\left(\beta = 0 | \mathcal{H}_1\right)} \approx \frac{1}{N} \sum_{i=1}^{N} \frac{p\left(\beta = 0 | \theta^{(i)}, s_t^{(i)}, y^t, \mathcal{H}_1\right)}{p\left(\beta = 0 | \mathcal{H}_1\right)}.$$

Here $p(\beta|\theta^{(i)}, s_t^{(i)}, y^t)$ is the normal distribution (see Section D) and $\theta^{(i)}$ is the filtered parameter vector for particle i at time t. Since we use a Bayesian regression to train the prior, the denominator is easy to calculate from $p\left(\beta | \mathcal{H}_1\right)$, which is a Student t distribution. The calculation for $\mathcal{H}_0 : \beta_{dp} = 1$ is analogous. For more details on calculating Bayes factors, see Section E of this appendix.

For the traditional dividend yield measure in the top-left panel of Figure IA.3, we find that there is little statistical evidence in favor of predictability. In the net payout data, the posterior probability of $\mathcal{H}_0 : \beta = 0$ slowly decreases to around 20% as parameter uncertainty decreases, but the decline in the predictability coefficient starting in the early 1990s reverses this trend, and ends up around 70%. This confirms the findings Boudoukh et al. (2007), as the weight of evidence against the hypothesis that $\beta = 0$ is much stronger using net payouts instead of the traditional cash dividend yield measure. The posterior probability of a unit root in dividend yields fluctuates significantly for the benchmark model, but the results generally favor a unit root and for net payout yield there is little evidence of any nonstationarity.

B. CV-DC Model

Figure IA.4 shows the parameter estimates for the drifting coefficients (CV-DC) model in equations (5) through (7) in the main article, using the cash dividend yield as predictor and assuming constant variances. In this model predictability consists of two components, a long-run average, β_0, and a time-varying component, β_t, with expected excess returns given by $E[\alpha + (\beta_0 + \beta_t)x_t|y^t]$. The long-run average predictability, β_0, is statistically indistinguishable from zero for virtually the entire sample period, as shown by the $(1, 99)\,\%$ posterior probability interval. The time-varying component, β_t, reveals substantial variation around the long-run average. Figure IA.6 shows that this variation is related to real GDP growth, with β_t higher in recessions and lower in expansions (Henkel, Martin, and Nardari (2011); Dangl and Halling (2012) also document this countercyclicality). However, the variation is not economically large and rarely statistically significant. The β_t process is highly persistent, with an autoregressive coefficient β_β of about 0.97, and the volatility of the shocks, σ_β, is around 0.001.

The posterior probability for the hypothesis test $\beta_0 + \beta_t = 0$ is analyzed in Figure IA.3. The probability is close to one for the entire sample, strongly supporting no predictability. This pattern in predictability from the CV-DC model is distinctly different from that of the cumulative OLS regressions in Figure 1 of the main text and the benchmark model in Figure IA.1.

The results for the net payout measure in Figure IA.5 are closer to the CV model. Figure IA.3 shows that the posterior probability of the null tracks closely with the benchmark model. At one point in the late 1960s, there was strong evidence for predictability, but that predictability quickly vanished in the 1970s. The posterior probability of no predictability has remained between 50% and 90% since then. The long-run average predictability, β_0, is high compared to the dividend yield variable, and is close to the estimate of β_x in the CV model. The time-varying component is stable, showing little variation around the long-run

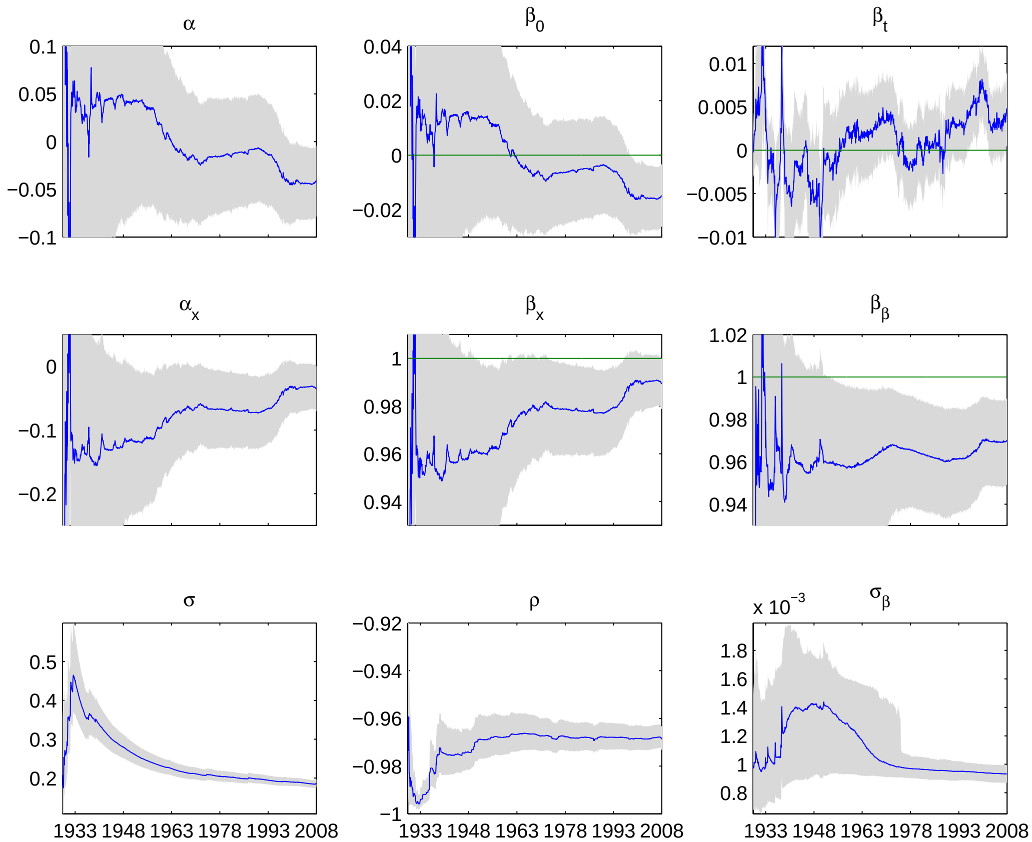

Figure IA.4. Sequential parameter estimates: CV-DC model with dividend yield. This figure plots sequential parameter estimates for the drifting coefficients model using the traditional dividend yield as predictor. The time-varying predictability coefficient follows an AR(1) process, $\beta_{t+1} = \beta_\beta \beta_t + \sigma_\beta \varepsilon_{t+1}^\beta$. The other coefficients are as defined in figure IA.1. Each panel displays the posterior means and (1,99)% posterior credible intervals for each time period. Return volatility is annualized.

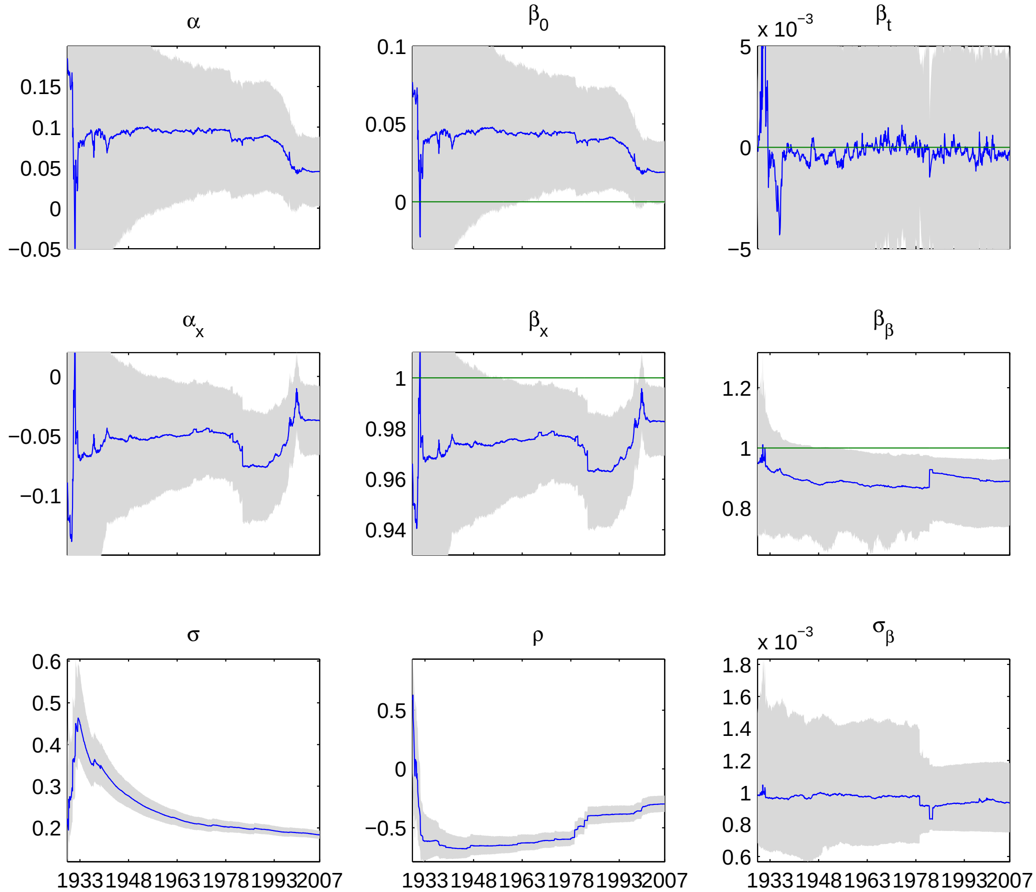

Figure IA.5. Sequential parameter estimates: CV-DC model with net payout yield. This figure plots sequential parameter estimates for the drifting coefficients model using the net payout yield as predictor. The time-varying predictability coefficient follows an AR(1) process, $\beta_{t+1} = \beta_\beta \beta_t + \sigma_\beta \varepsilon^{\beta}_{t+1}$. The other coefficients are as defined in figure IA.2. Each panel displays the posterior means and (1,99)% posterior credible intervals for each time period. Return volatility is annualized.

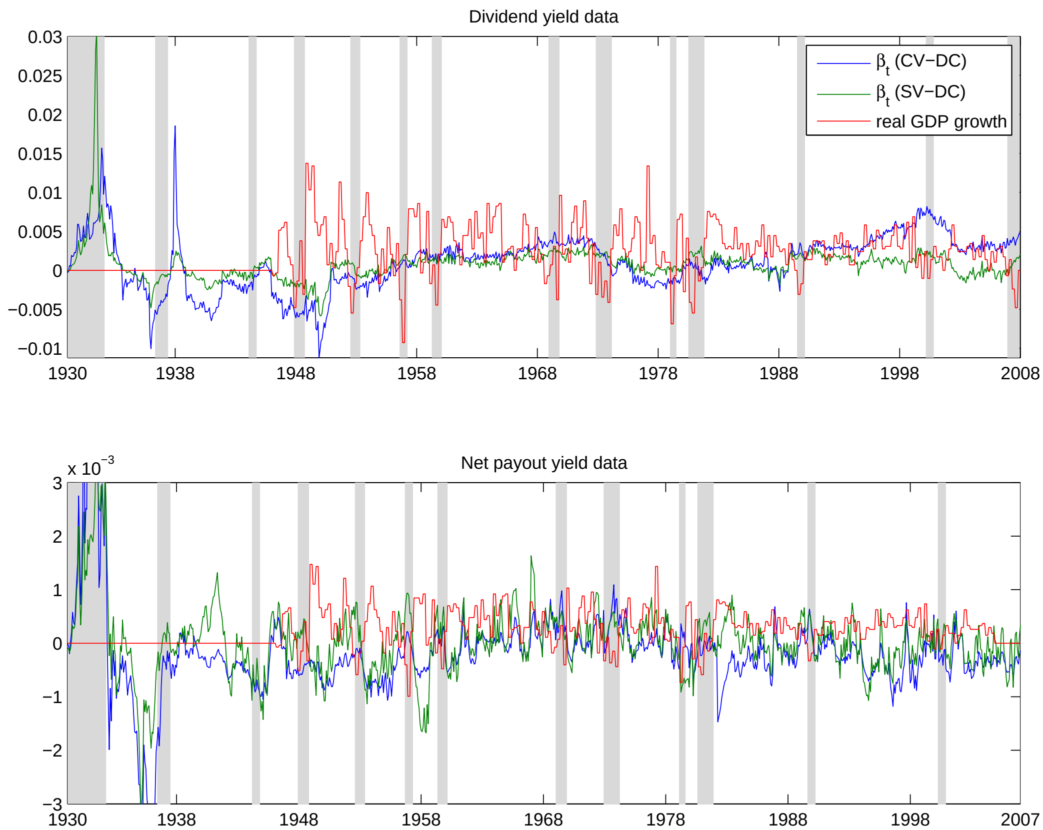

Figure IA.6. Drifting predictability coefficient. This figure depicts time series plots of the posterior mean of β_t, the drifting component of the predictability coefficient, plotted against real GDP growth (normalized to have the same standard deviation as β_t). The grey shaded areas are NBER peak-to-trough recessions. The top plot uses the traditional cash dividend yield as the predictor variable, whereas the bottom plot uses the net payout yield. The CV-DC model has constant volatility and the SV-DC model has stochastic volatility.

mean for most of the sample. The total predictability coefficient, $\beta_0 + \beta_t$, therefore hovers very closely around the CV estimate.

Investors learn about the level of long-term predictability, β_0, as evidenced by the tightening $(1, 99)\,\%$ bounds in Figures IA.4 and IA.5. The bounds for the time-varying component, β_t, tighten only slightly over the sample period, and only to the extent that investors learn about β_β and σ_β, the parameters that govern the process of the time-varying coefficient. In other words, investors learn the long-run mean predictability and the process of the time-varying component, but never learn the exact predictability coefficient at a given point in time. The economic implication is that learning remains important in drifting coefficient models, even in very large data sets, whereas the effects of learning on portfolio formation wanes over time in constant coefficient models (such as our CV model). This is true in general for models with unobserved and time-varying state variables.

C. SV Model

Figure IA.7 displays the stochastic volatility (SV) model estimates using cash dividends. This model is expressed in equations (2) and (7) in the main article. It has constant regression coefficients, like the benchmark model, but allows for stochastic volatility in both the excess return and payout yield equations.

The posterior mean estimates of the regression coefficients, α, β, α_x, and β_x, are different from those obtained in the constant volatility (CV) model due to the "GLS versus OLS" distinction, where periods of high volatility are down-weighted in the SV model but not in the CV model.

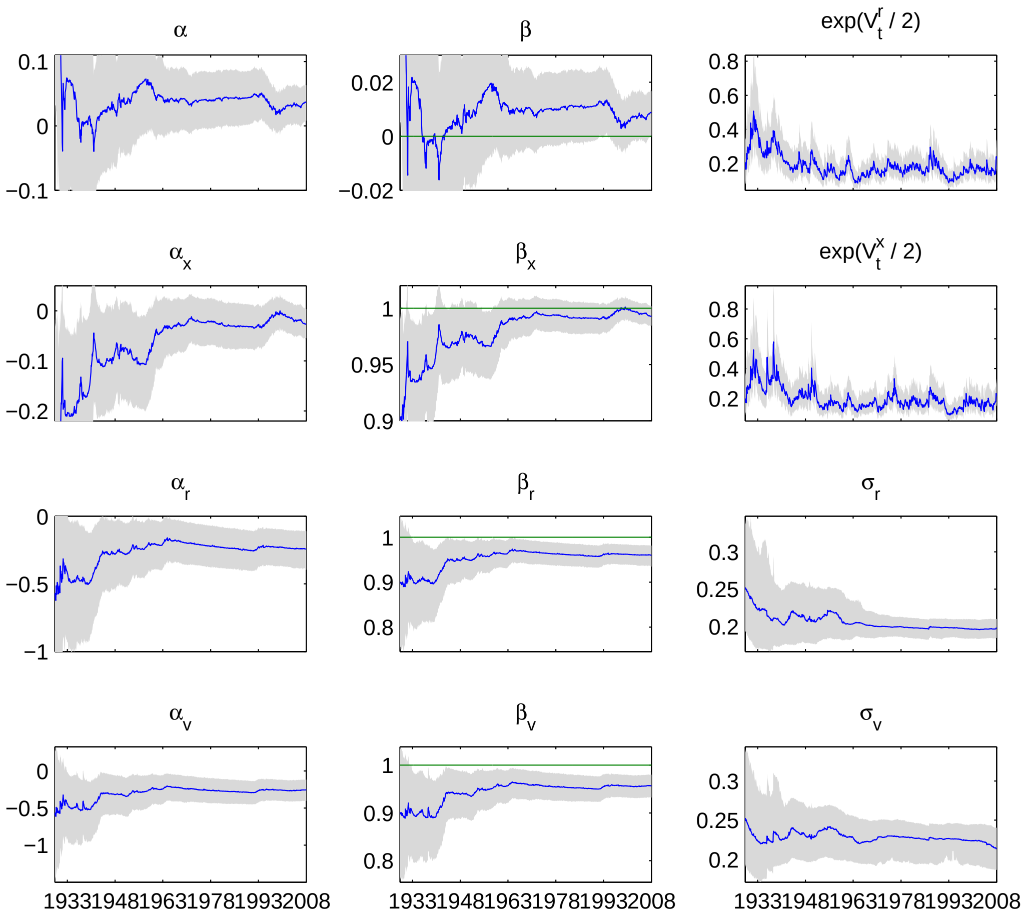

Figure IA.7. Sequential parameter estimates: SV model with dividend yield. This figure plots sequential parameter estimates for the SV model using the traditional dividend yield as predictor. The log-volatilities follow AR(1) processes

$$V^r_{t+1} = \alpha_r + \beta_r V^r_t + \sigma_r \eta^r_{t+1}$$
$$V^x_{t+1} = \alpha_v + \beta_v V^x_t + \sigma_v \eta^v_{t+1}.$$

All other parameters are as in figure IA.1. Each panel displays the posterior means and (1,99)% posterior credible intervals for each time period. Return and dividend yield volatilities are annualized.

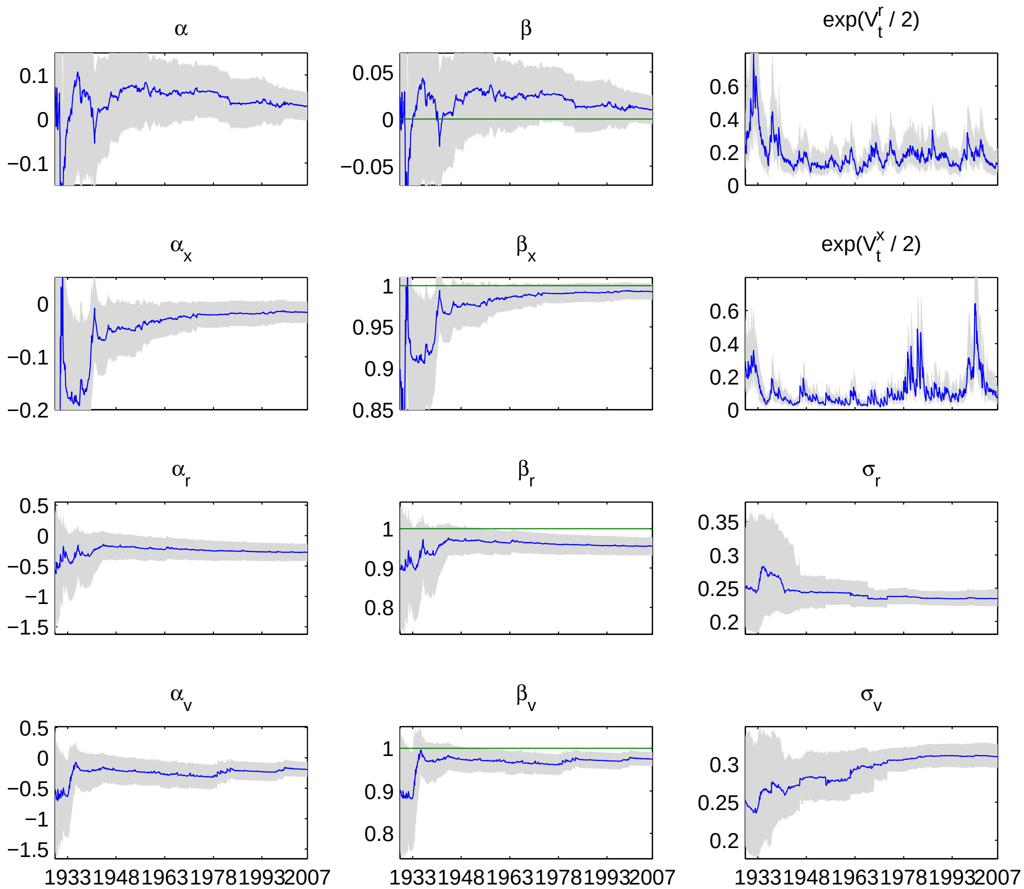

Figure IA.8. Sequential parameter estimates: SV model with net payout yield. This figure plots sequential parameter estimates for the SV model using the net payout yield as predictor. The log-volatilities follow AR(1) processes

$$V^r_{t+1} = \alpha_r + \beta_r V^r_t + \sigma_r \eta^r_{t+1}$$
$$V^x_{t+1} = \alpha_v + \beta_v V^x_t + \sigma_v \eta^v_{t+1}.$$

All other parameters are as in figure IA.2. Each panel displays the posterior means and (1,99)% posterior credible intervals for each time period. Return and net payout yield volatilities are annualized.

Excess return volatility, V^r, is high in the 1930s, during the oil crisis of the 1970s, during the crash of 1987, in the Internet period 1997 to 2001, and during the credit crisis of 2008. Dividend yield volatility, V^x, is high in the 1930s and during the 1970s and early 1980s. The volatility processes of excess returns and dividend yields are very persistent, with autoregressive coefficients around 0.95. In comparison, Markov Chain Monte Carlo (MCMC) estimates of the autoregressive coefficient of excess return volatility are around 0.98 (Johannes, Polson, and Stroud (2002)).

For the net payout data, excess return volatility shows the same pattern as cash dividends, but the volatility for the predictor variable exhibits a different pattern. In contrast to dividend yield, net payout volatility is low in the 1930s but high during the early 2000s, when dividends remained stable but issuances and repurchases spiked (Boudoukh et al. (2007)). There are two extreme volatility spikes, in the early 1980s and in 2000. The volatility shocks absorb the large shocks to net payouts during that period, and α_x, β_x, and ρ do not show the breaks that we find in the CV and CV-DC models.

D. SV-DC Model

Both features of the SV and DC models are present in the full-fledged SV-DC model in equations (5) through (7) in the published article. Figures IA.9 and IA.10 show sequential parameter estimates of the SV-DC model using the cash dividend yield and the net payout yield, respectively, as the predictor variable. For the dividend yield estimates in Figure IA.9, the most notable difference is that the process for β_t is slightly less persistent (i.e., the autoregressive coefficient β_β is lower), which causes shocks to β_t to dissipate faster. The

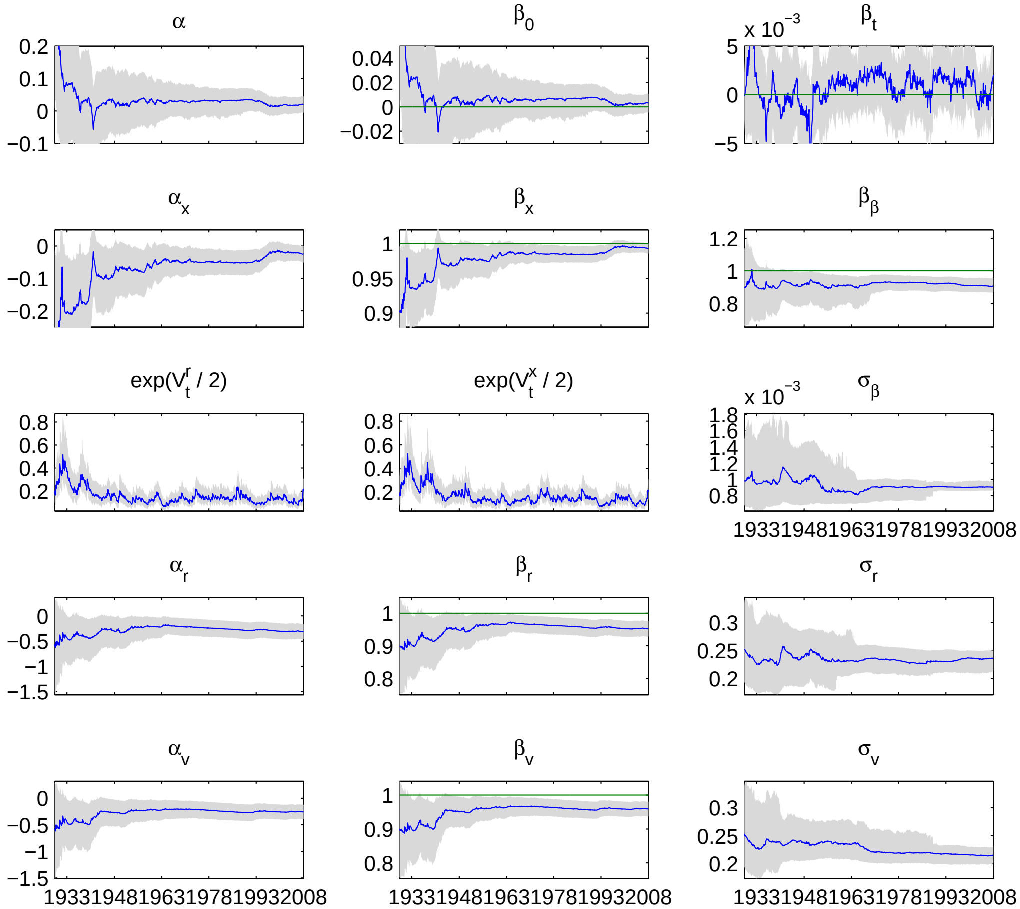

Figure IA.9. Sequential parameter estimates: SV-DC model with dividend yield. This figure plots sequential parameter estimates for the stochastic volatility and drifting coefficients model using the traditional dividend yield as predictor. The parameters are as defined in figures IA.4 and IA.7. Each panel displays the posterior means and (1,99)% posterior credible intervals for each time period. Return and dividend yield volatilities are annualized.

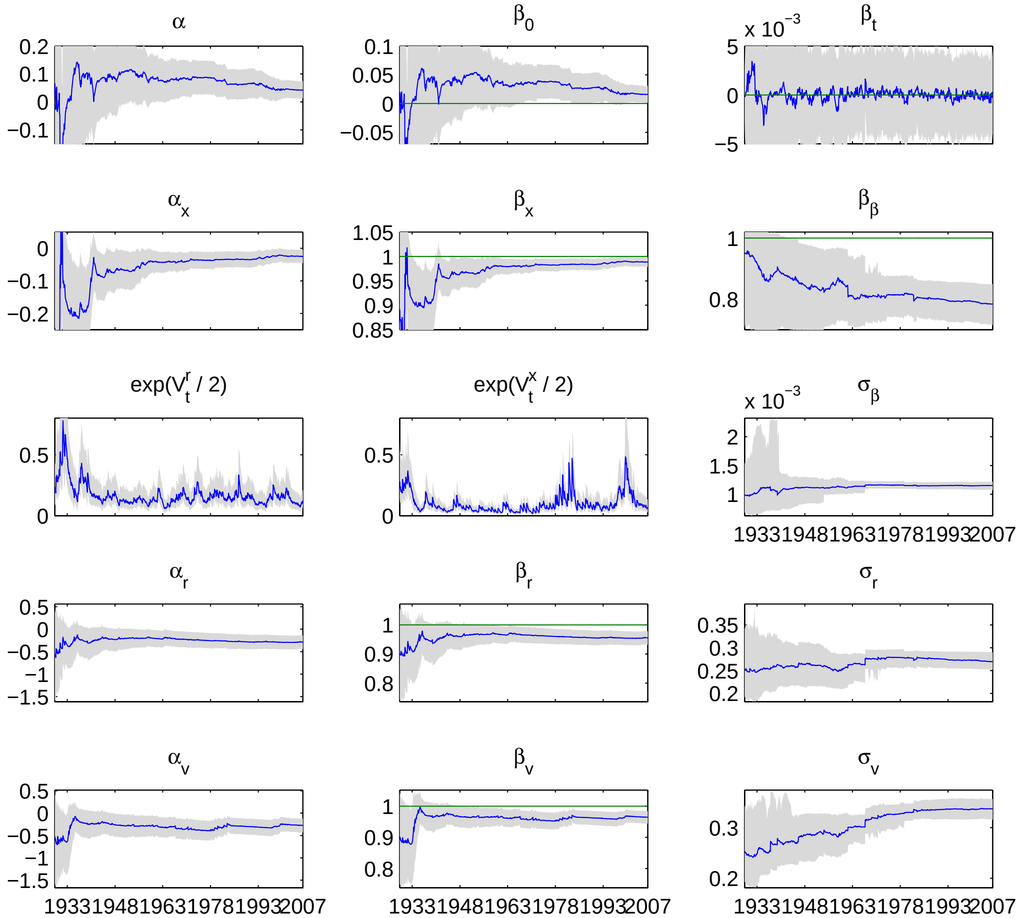

Figure IA.10. Sequential parameter estimates: SV-DC model with net payout yield. This figure plots sequential parameter estimates for the stochastic volatility and drifting coefficients model using the net payout yield as predictor. The parameters are as defined in figures IA.5 and IA.8. Each panel displays the posterior means and (1,99)% posterior credible intervals for each time period. Return and net payout yield volatilities are annualized.

evidence for or against predictability is quite volatile, as both volatility and the regression coefficients move over time. At the end of the sample, there is very strong evidence against predictability, using either data set.

II. Simulation Results for Statistical Significance

To judge statistical significance, we simulate 500 data sets of a model that has no predictability by construction:

$$r_{t+1} = \alpha + \sigma \varepsilon^{r}_{t+1} \tag{IA.1}$$

$$x_{t+1} = \alpha_x + \beta_x x_t + \sigma_x \varepsilon^{x}_{t+1}. \tag{IA.2}$$

We calibrate the model to the observed data, generating returns and predictors with the same expected returns and variances as our empirical data set. The predictors are also calibrated to have the same autocorrelation as we observe in the data. We then estimate our various models on the simulated data sets and report the mean certainty equivalent return and Sharpe ratio, as well as the 90th and 95th percentiles across data sets. Tables IA.I and IA.II show the results for the simulated dividend yield and net payout yield, respectively.

In Tables IA.III and IA.IV we perform a similar exercise for dividend yield and net payout yield, respectively, but we simulate data with stochastic volatility in both the returns and the predictor variables while maintaining no predictability by construction. We chose

Table IA.I

Statistical Significance: Dividend Yield Data

This table reports summary statistics of annualized certainty equivalent returns (Panel A) and monthly Sharpe ratios (Panel B) across 500 simulated data sets for a power utility investor with risk aversion γ. Simulated data sets are of the same size, and the same means and covariances, as the empirical data set, but with no predictability. The top line for each model shows the mean statistic across data sets, followed by the 90th and 95th percentiles.

Panel A: Certainty equivalent returns (in % per annum)

		$\gamma = 4$			$\gamma = 6$		
		1m	1y	2y	1m	1y	2y
CV-CM	mean	2.60	4.10	4.10	3.26	3.93	3.93
	90th percentile	5.85	5.92	5.92	5.12	5.16	5.16
	95th percentile	6.20	6.17	6.17	5.35	5.31	5.31
CV-OLS	mean	-4.84	-4.73	-4.85	-3.45	-4.01	-6.45
	90th percentile	5.19	5.30	5.31	4.68	4.74	4.75
	95th percentile	5.99	5.90	5.92	5.22	5.15	5.15
CV-rolling OLS	mean	-23.84	-23.88	-25.03	-42.46	-45.95	-47.55
	90th percentile	-8.95	-8.69	-9.54	-12.69	-13.28	-14.58
	95th percentile	-6.15	-6.33	-7.03	-9.96	-10.17	-11.53
CV	mean	-3.89	3.66	3.51	-2.32	3.63	3.48
	90th percentile	5.47	5.84	5.84	4.87	5.08	5.11
	95th percentile	6.16	6.30	6.30	5.32	5.42	5.42
CV-DC	mean	-5.09	4.49	4.53	-1.71	4.14	4.17
	90th percentile	5.09	5.91	5.92	4.59	5.18	5.16
	95th percentile	5.74	6.18	6.19	5.05	5.34	5.37
SV-CM	mean	4.37	5.20	5.21	4.00	4.64	4.65
	90th percentile	5.89	6.10	6.10	5.14	5.26	5.26
	95th percentile	6.13	6.25	6.25	5.31	5.36	5.37
SV	mean	3.32	5.10	5.10	3.18	4.57	4.57
	90th percentile	5.55	6.30	6.34	4.88	5.42	5.43
	95th percentile	6.00	6.50	6.53	5.17	5.55	5.56
SV-DC	mean	-0.20	5.01	5.00	-0.84	4.53	4.52
	90th percentile	5.18	6.11	6.07	4.64	5.30	5.26
	95th percentile	5.78	6.42	6.37	5.04	5.49	5.45

Panel B: Sharpe ratios (monthly)							
		$\gamma = 4$			$\gamma = 6$		
		1m	1y	2y	1m	1y	2y
CV-CM	mean	0.086	0.093	0.093	0.086	0.093	0.093
	90th percentile	0.125	0.128	0.128	0.124	0.129	0.129
	95th percentile	0.133	0.136	0.136	0.133	0.136	0.136
CV-OLS	mean	0.059	0.072	0.072	0.059	0.071	0.071
	90th percentile	0.110	0.117	0.116	0.108	0.116	0.116
	95th percentile	0.126	0.129	0.130	0.125	0.129	0.130
CV-rolling OLS	mean	0.063	0.071	0.072	0.059	0.067	0.068
	90th percentile	0.108	0.114	0.114	0.104	0.109	0.110
	95th percentile	0.121	0.123	0.123	0.117	0.120	0.121
CV	mean	0.072	0.085	0.086	0.071	0.085	0.085
	90th percentile	0.118	0.124	0.126	0.118	0.125	0.126
	95th percentile	0.131	0.136	0.135	0.131	0.135	0.135
CV-DC	mean	0.068	0.089	0.090	0.067	0.088	0.089
	90th percentile	0.113	0.124	0.126	0.112	0.125	0.127
	95th percentile	0.123	0.134	0.135	0.123	0.134	0.135
SV-CM	mean	0.088	0.103	0.104	0.086	0.102	0.102
	90th percentile	0.129	0.132	0.132	0.129	0.132	0.132
	95th percentile	0.137	0.140	0.140	0.137	0.139	0.140
SV	mean	0.073	0.099	0.100	0.071	0.098	0.099
	90th percentile	0.117	0.138	0.138	0.116	0.137	0.138
	95th percentile	0.126	0.143	0.143	0.125	0.143	0.143
SV-DC	mean	0.064	0.096	0.096	0.063	0.095	0.095
	90th percentile	0.113	0.133	0.133	0.113	0.134	0.133
	95th percentile	0.122	0.142	0.142	0.122	0.142	0.142

Table IA.II

Statistical Significance: Net Payout Yield Data

This table reports summary statistics of annualized certainty equivalent returns (Panel A) and monthly Sharpe ratios (Panel B) across 500 simulated data sets for a power utility investor with risk aversion γ. Simulated data sets are of the same size, and the same means and covariances, as the empirical data set, but with no predictability. The top line for each model shows the mean statistic across data sets, followed by the 90th and 95th percentiles.

Panel A: Certainty equivalent returns (in % per annum)							
		$\gamma = 4$			$\gamma = 6$		
		1m	1y	2y	1m	1y	2y
CV-OLS	mean	-5.11	-1.24	-1.20	-7.05	-6.44	-6.39
	90th percentile	5.68	6.32	6.31	5.02	5.42	5.43
	95th percentile	6.33	6.74	6.75	5.47	5.71	5.73
CV-rolling OLS	mean	-20.88	-17.06	-16.56	-40.85	-40.60	-40.04
	90th percentile	-5.14	-4.02	-3.94	-8.75	-6.88	-6.71
	95th percentile	-3.10	-1.70	-1.64	-5.68	-4.80	-4.64
CV	mean	-3.80	4.43	4.53	-3.97	4.21	4.26
	90th percentile	6.10	6.42	6.43	5.27	5.49	5.50
	95th percentile	6.56	6.79	6.78	5.59	5.74	5.76
CV-DC	mean	-4.08	4.94	5.01	-3.59	4.47	4.54
	90th percentile	5.68	6.35	6.34	4.99	5.45	5.45
	95th percentile	6.19	6.75	6.75	5.34	5.73	5.71
SV	mean	4.32	5.72	5.72	3.92	5.01	5.02
	90th percentile	6.09	6.74	6.74	5.29	5.71	5.70
	95th percentile	6.49	6.89	6.90	5.52	5.82	5.83
SV-DC	mean	1.54	5.58	5.55	1.60	4.92	4.90
	90th percentile	5.63	6.59	6.55	4.97	5.62	5.59
	95th percentile	6.18	6.83	6.76	5.33	5.79	5.73

Panel B: Sharpe ratios (monthly)							
		$\gamma = 4$			$\gamma = 6$		
		1m	1y	2y	1m	1y	2y
CV-OLS	mean	0.069	0.088	0.089	0.068	0.087	0.087
	90th percentile	0.120	0.135	0.135	0.119	0.134	0.135
	95th percentile	0.135	0.145	0.145	0.135	0.145	0.145
CV-rolling OLS	mean	0.079	0.088	0.088	0.076	0.084	0.084
	90th percentile	0.123	0.127	0.127	0.119	0.121	0.121
	95th percentile	0.132	0.137	0.137	0.128	0.132	0.132
CV	mean	0.083	0.102	0.103	0.082	0.102	0.102
	90th percentile	0.129	0.142	0.142	0.129	0.142	0.142
	95th percentile	0.142	0.151	0.151	0.142	0.150	0.151
CV-DC	mean	0.078	0.104	0.105	0.077	0.104	0.104
	90th percentile	0.124	0.142	0.143	0.123	0.142	0.143
	95th percentile	0.134	0.152	0.153	0.134	0.151	0.152
SV	mean	0.091	0.121	0.121	0.088	0.120	0.120
	90th percentile	0.128	0.150	0.150	0.128	0.150	0.150
	95th percentile	0.140	0.157	0.156	0.140	0.158	0.157
SV-DC	mean	0.077	0.117	0.116	0.075	0.116	0.116
	90th percentile	0.120	0.149	0.149	0.120	0.150	0.150
	95th percentile	0.129	0.155	0.154	0.129	0.155	0.156

Table IA.III

Statistical Significance: Dividend Yield Data with Stochastic Volatility

This table reports summary statistics of annualized certainty equivalent returns (Panel A) and monthly Sharpe ratios (Panel B) across 500 simulated data sets for a power utility investor with risk aversion coefficient γ. The simulated data sets are of the same size and with the same means as the empirical data set, but with no predictability. Stochastic volatility parameters are calibrated to match the observed data. The top line for each model shows the mean statistic across data sets, followed by the 90th and 95th percentiles.

Panel A: Certainty equivalent returns (in % per annum)							
		$\gamma = 4$			$\gamma = 6$		
		1m	1y	2y	1m	1y	2y
CV-CM	mean	0.51	2.99	2.99	0.84	3.42	3.42
	90th percentile	5.83	5.83	5.83	5.09	5.11	5.11
	95th percentile	6.13	6.07	6.07	5.29	5.27	5.27
CV-OLS	mean	-6.57	-6.62	-7.78	-6.48	-14.88	-15.98
	90th percentile	5.20	5.31	5.34	4.67	4.74	4.76
	95th percentile	5.70	5.70	5.67	5.01	4.98	4.99
CV-rolling OLS	mean	-24.42	-24.95	-25.94	-45.01	-57.23	-59.25
	90th percentile	-7.07	-6.99	-7.47	-10.08	-10.48	-10.85
	95th percentile	-4.51	-3.87	-4.20	-8.15	-6.93	-7.50
CV	mean	-5.49	2.78	2.54	-4.25	3.15	2.85
	90th percentile	5.48	5.66	5.67	4.85	4.99	4.98
	95th percentile	5.90	6.00	6.02	5.14	5.23	5.23
CV-DC	mean	-7.20	4.37	4.42	-5.65	4.06	4.10
	90th percentile	5.05	5.66	5.66	4.59	4.98	4.98
	95th percentile	5.54	5.93	5.92	4.89	5.16	5.15
SV-CM	mean	4.24	5.16	5.16	3.92	4.62	4.62
	90th percentile	5.88	6.03	6.03	5.12	5.22	5.22
	95th percentile	6.16	6.26	6.26	5.31	5.37	5.37
SV	mean	2.66	5.08	5.09	2.13	4.57	4.57
	90th percentile	5.50	6.04	6.06	4.81	5.23	5.25
	95th percentile	5.87	6.38	6.38	5.09	5.46	5.49
SV-DC	mean	-1.95	5.03	5.01	-4.33	4.51	4.50
	90th percentile	5.32	6.15	6.11	4.75	5.30	5.27
	95th percentile	5.85	6.43	6.39	5.09	5.49	5.47

Panel B: Sharpe ratios (monthly)							
		$\gamma = 4$			$\gamma = 6$		
		1m	1y	2y	1m	1y	2y
CV-CM	mean	0.084	0.091	0.091	0.084	0.091	0.091
	90th percentile	0.127	0.131	0.131	0.128	0.131	0.131
	95th percentile	0.135	0.137	0.137	0.135	0.137	0.137
CV-OLS	mean	0.052	0.067	0.067	0.051	0.066	0.066
	90th percentile	0.108	0.113	0.114	0.108	0.113	0.114
	95th percentile	0.120	0.125	0.125	0.120	0.124	0.124
CV-rolling OLS	mean	0.061	0.069	0.070	0.057	0.065	0.066
	90th percentile	0.107	0.112	0.113	0.105	0.108	0.109
	95th percentile	0.121	0.121	0.122	0.116	0.118	0.118
CV	mean	0.066	0.082	0.082	0.065	0.081	0.081
	90th percentile	0.118	0.122	0.123	0.117	0.122	0.124
	95th percentile	0.128	0.134	0.135	0.127	0.136	0.136
CV-DC	mean	0.061	0.085	0.086	0.060	0.084	0.085
	90th percentile	0.111	0.123	0.123	0.110	0.123	0.124
	95th percentile	0.123	0.132	0.132	0.121	0.132	0.132
SV-CM	mean	0.085	0.101	0.102	0.084	0.100	0.101
	90th percentile	0.127	0.131	0.131	0.127	0.132	0.132
	95th percentile	0.134	0.138	0.138	0.135	0.137	0.137
SV	mean	0.069	0.100	0.100	0.067	0.099	0.099
	90th percentile	0.114	0.135	0.136	0.113	0.135	0.135
	95th percentile	0.126	0.142	0.143	0.125	0.141	0.142
SV-DC	mean	0.065	0.099	0.098	0.063	0.098	0.097
	90th percentile	0.116	0.138	0.137	0.116	0.137	0.138
	95th percentile	0.127	0.145	0.146	0.126	0.145	0.146

Table IA.IV

Statistical Significance: Net Payout Yield Data with Stochastic Volatility

This table reports summary statistics of annualized certainty equivalent returns (Panel A) and monthly Sharpe ratios (Panel B) across 500 simulated data sets for a power utility investor with risk aversion coefficient γ. The simulated data sets are of the same size and with the same means as the empirical data set, but with no predictability. Stochastic volatility parameters are calibrated to match the observed data. The top line for each model shows the mean statistic across data sets, followed by the 90th and 95th percentiles.

Panel A: Certainty equivalent returns (in % per annum)

		$\gamma = 4$			$\gamma = 6$		
		1m	1y	2y	1m	1y	2y
CV-OLS	mean	-6.54	-3.47	-2.89	-8.54	-6.78	-4.42
	90th percentile	5.69	6.16	6.17	5.01	5.34	5.33
	95th percentile	6.11	6.60	6.64	5.28	5.61	5.63
CV-rolling OLS	mean	-21.64	-18.38	-17.74	-40.95	-39.35	-36.99
	90th percentile	-4.10	-3.27	-3.09	-6.95	-5.04	-4.97
	95th percentile	-2.22	-0.59	-0.53	-5.10	-3.83	-3.78
CV	mean	-5.69	4.31	4.36	-5.33	4.08	4.10
	90th percentile	5.96	6.35	6.36	5.19	5.46	5.47
	95th percentile	6.40	6.72	6.74	5.49	5.71	5.72
CV-DC	mean	-6.60	4.90	4.96	-6.64	4.43	4.49
	90th percentile	5.77	6.31	6.31	5.08	5.45	5.43
	95th percentile	6.26	6.65	6.65	5.42	5.65	5.66
SV	mean	3.90	5.65	5.66	3.60	4.97	4.97
	90th percentile	6.16	6.76	6.74	5.30	5.71	5.73
	95th percentile	6.56	7.05	7.02	5.59	5.93	5.93
SV-DC	mean	-0.36	5.56	5.53	-3.47	4.91	4.89
	90th percentile	5.97	6.61	6.54	5.19	5.62	5.59
	95th percentile	6.50	6.88	6.79	5.53	5.79	5.75

Panel B: Sharpe ratios (monthly)							
		$\gamma = 4$			$\gamma = 6$		
		1m	1y	2y	1m	1y	2y
CV-OLS	mean	0.067	0.086	0.087	0.067	0.085	0.086
	90th percentile	0.122	0.136	0.137	0.121	0.135	0.137
	95th percentile	0.132	0.146	0.146	0.132	0.146	0.146
CV-rolling OLS	mean	0.080	0.089	0.089	0.077	0.085	0.085
	90th percentile	0.121	0.130	0.131	0.118	0.125	0.126
	95th percentile	0.132	0.140	0.140	0.128	0.134	0.134
CV	mean	0.082	0.101	0.101	0.081	0.101	0.101
	90th percentile	0.130	0.141	0.141	0.131	0.141	0.140
	95th percentile	0.141	0.150	0.151	0.140	0.151	0.151
CV-DC	mean	0.078	0.104	0.104	0.077	0.103	0.104
	90th percentile	0.125	0.139	0.140	0.125	0.139	0.140
	95th percentile	0.134	0.151	0.149	0.134	0.150	0.150
SV	mean	0.089	0.118	0.119	0.086	0.118	0.118
	90th percentile	0.133	0.153	0.153	0.132	0.153	0.152
	95th percentile	0.143	0.158	0.158	0.142	0.158	0.159
SV-DC	mean	0.078	0.116	0.116	0.076	0.115	0.115
	90th percentile	0.128	0.149	0.148	0.128	0.148	0.149
	95th percentile	0.139	0.155	0.155	0.138	0.156	0.155

the parameters of the volatility processes to match the empirical data, with autoregressive coefficients equal to 0.98, and matching the long-run mean of the volatility process to the unconditional volatility in the data.

III. Portfolio Weights

Figures IA.11 and IA.12 provide a term structure perspective on the portfolio weights. The figures display the portfolio weights on different dates for different models and investment horizons. The various models can generate very different long-horizon moments and return distributions, due to the time-varying state variables, estimation risk, and predictability. The differences arise because parameter uncertainty and mean-reversion (in expected returns and volatilities) impacts predictive moments differently as a function of investment horizon.

Table IA.V reports means and standard deviations of the portfolio weights, as well as correlations between the weights and the latent volatility states from the *SV* model, for $\gamma = 4$. The broad patterns are quite clear. First, the correlation between the portfolio weights in the SV models and actual volatility is negative and much higher (in an absolute sense) than in the constant volatility models.[1] This clearly demonstrates the volatility timing result from using models with stochastic volatility. Second, the stochastic volatility

[1] To calculate the correlation between portfolio weights and the volatility state, we omitted the first 10 years when there is still a lot of updating about the mean and variance in the constant model that can introduce a spurious correlation.

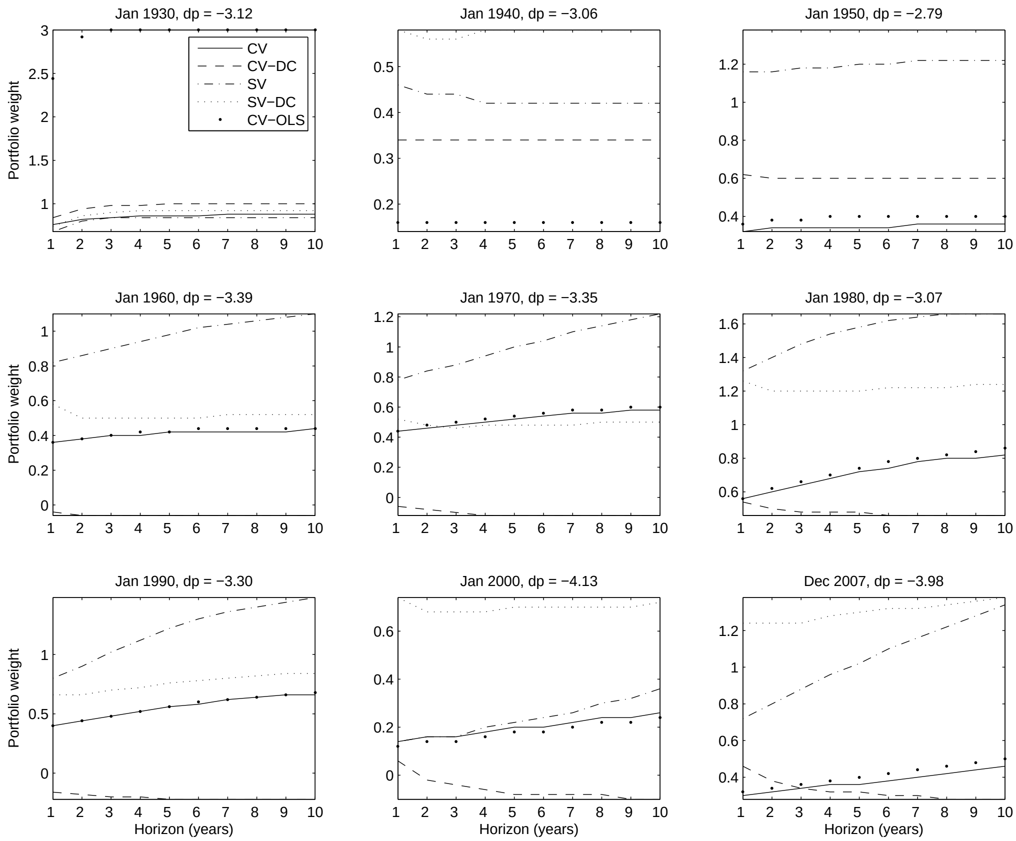

Figure IA.11. Optimal portfolio weights by investor horizon: dividend yield data. This figure plots optimal portfolio weights for an investor who allocates wealth between the market portfolio of stocks and a risk-free one-period bond, with an investment horizon spanning from one to 10 years. The plots show the optimal weights on the stock portfolio at the beginning of each decade in our sample period, as well as at the final datapoint in our sample (December 2008, bottom-right plot). The investor has power utility with risk aversion $\gamma = 4$, and rebalances annually while accounting for all parameter and state uncertainty. CV and SV represent models with expected return predictability and constant volatility (CV) and stochastic volatility (SV), respectively. DC stands for drifting coefficients and represents models where the predictability coefficient is allowed to vary over time. CV-OLS uses the OLS point estimates of equation (1) in the published paper, with data up to time t.

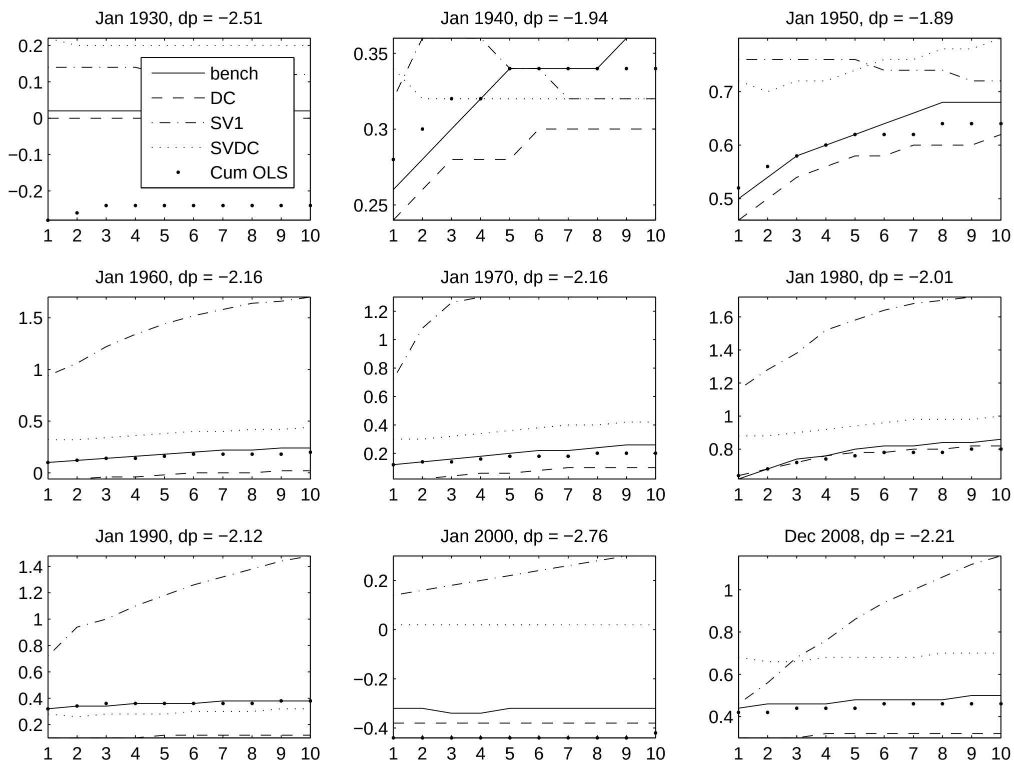

Figure IA.12. Optimal portfolio weights by investor horizon: net payout yield data. This figure plots optimal portfolio weights for an investor who allocates wealth between the market portfolio of stocks and a risk-free one-period bond, with an investment horizon spanning from one to 10 years. The plots show the optimal weights on the stock portfolio at the beginning of each decade in our sample period, as well as at the final datapoint in our sample (December 2008, bottom-right plot). The investor has power utility with risk aversion $\gamma = 4$, and rebalances annually while accounting for all parameter and state uncertainty. CV and SV represent models with expected return predictability and constant volatility (CV) and stochastic volatility (SV), respectively. DC stands for drifting coefficients and represents models where the predictability coefficient is allowed to vary over time. CV-OLS uses the OLS point estimates of equation (1) in the published paper, with data up to time t.

Table IA.V

Portfolio Weights Statistics, $\gamma = 4$

This table reports basic statistics of portfolio weights across models and investment horizons, for a power utility investor with risk aversion coefficient γ. The columns "mean" and "std" show the mean and standard deviation of portfolio weights across all months in our sample. The column "corr" shows the correlation coefficient between portfolio weights and the conditional volatility state from the SV model, after a burn-in period of 10 years.

Panel A: Dividend yield data

	1m			1y			2y		
	mean	std	corr	mean	std	corr	mean	std	corr
Constant volatility models									
CV-CM	0.41	0.18	-0.13	0.41	0.16	-0.09	0.41	0.16	-0.09
CV-OLS	0.14	0.27	-0.11	0.30	0.30	0.07	0.33	0.36	0.08
CV-rolling OLS	1.23	1.85	-0.01	1.46	1.80	-0.02	1.54	1.81	-0.01
CV	0.26	0.26	0.11	0.28	0.19	0.06	0.30	0.20	0.06
CV-DC	0.19	0.46	0.08	0.22	0.27	0.13	0.19	0.26	0.13
Stochastic volatility models									
SV-CM	1.09	0.56	-0.42	0.75	0.28	-0.18	0.75	0.28	-0.18
SV	0.88	0.50	-0.59	0.57	0.24	-0.25	0.59	0.25	-0.26
SV-DC	0.75	0.70	-0.43	0.40	0.26	-0.12	0.31	0.22	-0.12

Panel B: Net payout yield data

	1m			1y			2y		
	mean	std	corr	mean	std	corr	mean	std	corr
Constant volatility models									
CV-OLS	0.10	0.33	-0.22	0.25	0.29	-0.12	0.27	0.30	-0.13
CV-rolling OLS	1.39	1.72	-0.16	1.55	1.62	-0.09	1.57	1.62	-0.09
CV	0.23	0.33	-0.22	0.26	0.27	-0.12	0.27	0.28	-0.12
CV-DC	0.21	0.37	-0.25	0.23	0.26	-0.16	0.24	0.27	-0.16
Stochastic volatility models									
SV	0.64	0.64	-0.46	0.43	0.34	-0.29	0.47	0.36	-0.29
SV-DC	0.78	0.62	-0.53	0.51	0.27	-0.34	0.54	0.29	-0.34

models generally have higher average weights than the constant volatility models (ignoring the CV-rolling OLS results). This result occurs for two reasons: (1) the stochastic volatility models can take significantly larger positions when volatility is low than the constant volatility models (portfolio weights increase convexly as volatility falls), and (2) volatility estimates in constant volatility models trend down for the first portion of the sample, as realized equity volatility fell in the 1940s, 1950s, and 1960s. The constant volatility models just cannot handle persistent changes in volatility. In the dividend yield model, the SV model also has higher forecasts of expected returns for much of the sample. Third, constant mean models (for both stochastic and constant volatility specifications) have slightly higher average portfolio weights than the models with predictability. This is likely due to a combination of slightly less parameter uncertainty and drifting in predictability coefficients (see, for example, Figure 1 of the published paper).

IV. Conditional and Unconditional Return Distributions

We analyze the conditional and unconditional moments of the excess market return distribution from the various models through the use of two sets of simulations. We simulate 100,000 one-month returns from the *conditional* distribution of the excess market returns using a two-step approach. First, we randomly draw a set of parameters from the joint posterior distribution of parameters at the midpoint of the sample time series (December

1967). Second, we draw a return conditional on the set of parameters, where we set the state variables (payout yields and volatility states) equal to their sample means.[2] Thus, the simulated distribution is conditional in the sense that it conditions on specific values of the state variables, while fully reflecting the effects of parameter uncertainty. Given that the distribution absent uncertainty should be normal by assumption, this distribution is useful to gauge the impact of uncertainty about parameters and latent volatility states.

The second set of simulations is set up to analyze the *unconditional* return distribution from the models. We simulate a time series of 100,000 returns, starting the payout yield and volatility states at their sample means, and simulating them forward (rather than re-setting them to their sample means for each draw, as we did when sampling the conditional return distribution). Hence, this procedure involves an additional step compared to the simulation of the conditional return distribution above. The unconditional distribution not only fully reflects the effects of parameter uncertainty, but also, for the models with stochastic volatility, uncertainty about latent volatility states.

Tables IA.VI and IA.VII show the results. The conditional and unconditional return distributions for the models with constant volatility (CV-CM, CV-OLS, CV-rolling, CV, and CV-DC) are similar, with kurtosis slightly above three, and skewness that tends to be slightly positive. Across all models, the rolling OLS model has the worst fit. In contrast, the unconditional distributions for the stochastic volatility models (SV-CM, SV, SV-DC, and SV-corr) have considerably larger variance and kurtosis than the conditional distributions.

[2]Note that the specification of the model in equation (2) of the paper requires simulating the volatility state one month ahead.

Table IA.VI

Conditional and Unconditional Return Distributions: Dividend Yield Data

This table reports moments of the unconditional and conditional annualized excess market return distribution (where 0.01 = 1%), based on the posterior distribution of parameters at the midpoint of the sample time series (December 1967). The top row shows the moments of the data. For each of the models, the moments of the unconditional excess return distribution are calculated from a time series of 100,000 simulated returns. The conditional excess return distribution is calculated from 100,000 draws of one-month returns, with state variables (dividend yield and volatilities) set equal to their long-run means. Both distributions fully reflect the effects of parameter uncertainty and the uncertainty about latent volatility states, where applicable. "kurt" is kurtosis (where the Normal distribution has kurtosis of three).

	Unconditional excess returns				Conditional excess returns			
	mean	st.dev.	skew	kurt	mean	st.dev.	skew	kurt
data	0.056	0.190	-0.535	9.528	-	-	-	-
Constant Volatility models								
CV-CM	0.071	0.212	0.011	3.035	0.071	0.212	0.011	3.035
CV-OLS	0.072	0.215	0.010	3.015	0.072	0.215	0.010	3.015
CV-rolling OLS	0.321	0.120	-0.054	2.925	0.324	0.064	0.010	3.015
CV	0.064	0.251	-0.006	3.000	0.064	0.251	-0.002	3.002
CV-DC	0.048	0.287	0.013	3.025	0.049	0.286	0.006	3.014
Stochastic Volatility models								
SV-CM	0.173	0.251	0.396	20.422	0.173	0.179	0.036	6.435
SV	0.062	0.288	-0.068	12.997	0.059	0.210	-0.026	3.889
SV-DC	0.173	0.258	-0.069	12.937	0.189	0.181	-0.001	4.005
SV-corr	0.073	0.235	0.019	11.267	0.072	0.201	-0.013	3.173

This is due to the latent volatility state moving around over time, which is incorporated into the unconditional, but not the conditional, distribution. The unconditional distribution of these models gets much closer to the distribution of the data compared to the models with constant volatility. These results underscore once again that stochastic volatility is important for fitting the kurtosis observed in the data. The SV models also generate more negative skewness compared to the CV models, although the high negative skewness observed in the data is more difficult to match.

Table IA.VII

Conditional and Unconditional Return Distributions: Net Payout Yield Data

This table reports moments of the unconditional and conditional annualized excess market return distribution (where 0.01 = 1%), based on the posterior distribution of parameters at the midpoint of the sample time series (December 1967). The top row shows the moments of the data. For each of the models, the moments of the unconditional excess return distribution are calculated from a time series of 100,000 simulated returns. The conditional excess return distribution is calculated from 100,000 draws of one-month returns, with state variables (net payout yield and volatilities) set equal to their long-run means. Both distributions fully reflect the effects of parameter uncertainty and the uncertainty about latent volatility states, where applicable. "kurt" is kurtosis (where the Normal distribution has kurtosis of three).

	Unconditional excess returns				Conditional excess returns			
	mean	st.dev.	skew	kurt	mean	st.dev.	skew	kurt
data	0.056	0.190	-0.535	9.528	-	-	-	-
Constant Volatility models								
CV-OLS	0.068	0.216	0.009	3.011	0.065	0.213	0.010	3.015
CV-rolling OLS	0.132	0.095	-0.022	2.989	0.127	0.064	0.010	3.015
CV	0.065	0.265	-0.014	3.000	0.063	0.263	-0.002	3.004
CV-DC	0.063	0.271	-0.007	3.019	0.065	0.269	0.006	3.013
Stochastic Volatility models								
SV	0.062	0.278	-0.064	12.278	0.064	0.199	-0.028	3.770
SV-DC	0.084	0.261	0.022	15.618	0.086	0.195	-0.011	4.162
SV-corr	0.072	0.216	-0.004	11.065	0.073	0.179	-0.011	3.069

V. Particle Filter Algorithms

Our particle filtering and learning algorithm is as follows. First, express $p\left(L_{t+1}|y^{t+1}\right)$ relative to $p\left(L_t, s_t, \theta|y^t\right)$. The s_t are sufficient statistics for the distribution of the parameters, θ. The continuous distributions are approximated by a set of N particles, $p^N\left(L_t, s_t, \theta|y^t\right)$. The particles essentially form a histogram, where each particle represents a point in the support (L_t, s_t, θ) with a "weight", w, that corresponds to the height of the histogram:

$$\begin{aligned} p^N\left(L_{t+1}|y^{t+1}\right) &= \int p\left(y_{t+1}|L_t, s_t, \theta\right) p\left(L_{t+1}|L_t, s_t, \theta, y_{t+1}\right) dp^N\left(L_t, s_t, \theta|y^t\right) \\ &= \sum_{i=1}^{N} w\left(\left(L_t, s_t, \theta\right)^{(i)}\right) p\left(L_{t+1}|\left(L_t, s_t, \theta\right)^{(i)}, y_{t+1}\right), \end{aligned}$$

with weights given by

$$w\left(\left(L_t, s_t, \theta\right)^{(i)}\right) = \frac{p\left(y_{t+1}|\left(L_t, s_t, \theta\right)^{(i)}\right)}{\sum_{i=1}^{N} p\left(y_{t+1}|\left(L_t, s_t, \theta\right)^{(i)}\right)}.$$

The distribution $p^N\left(L_{t+1}|y^{t+1}\right)$ is then a discrete mixture distribution. To sample from this distribution, first draw

$$\textit{Step 1:} \quad k^{(i)} \sim \textit{Multi}\left(w\left(L_t, s_t, \theta\right)^{(i)}\right).$$

Now propagate the states and sufficient statistics to $\left(L_{t+1}\right)^{(i)}$

$$\textit{Step 2:} \quad L_{t+1}^{(i)} \sim p\left(L_{t+1}|\left(L_t, s_t, \theta\right)^{k^{(i)}}, y_{t+1}\right)$$

$$\textit{Step 3:} \quad s_{t+1}^{(i)} = \mathcal{S}\left(s_t^{k^{(i)}}, L_{t+1}^{(i)}, y_{t+1}\right).$$

Given sufficient statistics, the parameters are propagated with

$$\textit{Step 4:} \quad \theta^{(i)} \sim p\left(\theta|s_{t+1}^{(i)}\right) .$$

Given these particles, it is easy to estimate parameters, state variables, and marginal likelihoods. The likelihood component $p(y^t|\mathcal{M}_j) = \prod_{j=1}^{t} p(y_j|y^{j-1}, \mathcal{M}_j)$ is estimated recursively using

$$p(y_t|y^{t-1}, \mathcal{M}_j) \approx \frac{1}{N}\sum_{i=1}^{N} p(y_t|\theta^{(i)}, s_{t-1}^{(i)}, L_{t-1}^{(i)}, y^{t-1}, \mathcal{M}_j).$$

In the remainder of this section we provide the particle filter for each model in detail.

A. CV Model

The benchmark CV model has no latent state variables since both r_t and x_t are observed:

$$r_{t+1} = \alpha + \beta x_t + \sigma \varepsilon_{t+1}^r \tag{IA.3}$$

$$x_{t+1} = \alpha_x + \beta_x x_t + \sigma_x \varepsilon_{t+1}^x. \tag{IA.4}$$

The shocks are standard normal random variables with correlation ρ, and the parameter vector $\theta = (\alpha, \beta, \alpha_x, \beta_x, \sigma, \sigma_x, \rho)$. The initial resampling step uses weights proportional to the predictive likelihood of the new data, $y_{t+1} = [\ r_{t+1}\quad x_{t+1}\ ]'$:

$$w(s_t, \theta) \propto p\left(y_{t+1}|y_t, s_t, \theta\right) = \mathcal{N}\left(\begin{bmatrix} \alpha + \beta x_t \\ \alpha_x + \beta_x x_t \end{bmatrix}, \Sigma\right),$$

where

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma\sigma_x \\ \rho\sigma\sigma_x & \sigma_x^2 \end{bmatrix}.$$

The parameter posteriors follow from the theory of multivariate normal linear and conjugate

Normal-Inverse Wishart priors:

$$p\left(\Sigma|s_{t+1}\right) \sim \mathcal{IW}\left(c_{t+1}, C_{t+1}\right)$$

$$p\left(\alpha, \beta, \alpha_x, \beta_x|\Sigma, s_{t+1}\right) \sim \mathcal{N}\left(vec(\mu_{t+1}), \Sigma \otimes A_{t+1}^{-1}\right),$$

where $\mu_{t+1} = A_{t+1}^{-1} a_{t+1}$. The sufficient statistics, s_{t+1}, are updated using the recursions:

$$A_{t+1} = A_t + Z_t Z_t'$$

$$a_{t+1} = a_t + Z_t \cdot y_{t+1}'$$

$$W_{t+1} = W_t + y_{t+1} \cdot y_{t+1}'$$

$$c_{t+1} = c_t + 1$$

$$C_{t+1} = C_0 + W_{t+1} + \mu_{t+1} A_{t+1} \mu_{t+1} - \mu_{t+1}' \cdot a_{t+1} - a_{t+1}' \cdot \mu_{t+1} + \left(\mu_{t+1} - \mu_0\right)' A_0 \left(\mu_{t+1} - \mu_0\right),$$

where $W_0 = 0$ and $Z_t = [\ 1 \quad x_t\ ]'$.

B. CV-DC Model

The drifting coefficients model has one latent state variable, $L_t = \beta_t$:

$$r_{t+1} = \alpha + \beta x_t + \beta_{t+1} x_t + \sigma \varepsilon_{t+1}^r \tag{IA.5}$$

$$x_{t+1} = \alpha_x + \beta_x x_t + \sigma_x \varepsilon_{t+1}^x, \tag{IA.6}$$

$$\beta_{t+1} = \beta_\beta \beta_t + \sigma_\beta \varepsilon_{t+1}^\beta. \tag{IA.7}$$

The innovations in the predictability coefficient, ε_{t+1}^β, are independent of the shocks ε_{t+1} and ε_{t+1}^x. First, resample the particles with weights proportional to the predictive likelihood:

$$w(L_t, s_t, \theta) \propto p\left(y_{t+1}|y_t, L_t, s_t, \theta\right) = \mathcal{N}\left(\begin{bmatrix} \alpha + \beta x_t + \beta_\beta \beta_t x_t \\ \alpha_x + \beta_x x_t \end{bmatrix}, \Sigma + \begin{bmatrix} \sigma_\beta^2 x_t^2 & 0 \\ 0 & 0 \end{bmatrix}\right),$$

Next, update the latent state using the Kalman filter recursion

$$\beta_{t+1}|L_t, \theta, y_{t+1} \sim \mathcal{N}\left(V \cdot \left[\beta_\beta \beta_t/\sigma_\beta^2 + [x_t \ 0]\Sigma^{-1} \left[\begin{array}{c} r_{t+1} - \alpha - \beta x_t \\ x_{t+1} - \alpha_x - \beta_x x_t \end{array}\right]\right], V\right),$$

where $V^{-1} = 1/\sigma_\beta^2 + [x_t \ 0]\Sigma^{-1}[x_t \ 0]'$.

The parameters of the observation equations are drawn as in the benchmark model, with Normal-Inverse Wishart conjugate priors. The parameters of the drifting coefficient evolution are drawn from a linear regression with Normal-Inverse Gamma conjugate prior:

$$p(\Sigma|s_{t+1}) \sim \mathcal{IW}(c_{t+1}, C_{t+1})$$

$$p(\alpha, \beta, \alpha_x, \beta_x|\Sigma, s_{t+1}) \sim \mathcal{N}\left(vec(\mu_{t+1}), \Sigma \otimes A_{t+1}^{-1}\right)$$

$$p\left(\sigma_\beta^2|s_{t+1}\right) \sim \mathcal{IG}(g_{t+1}, G_{t+1})$$

$$p\left(\beta_\beta|\sigma_\beta^2, s_{t+1}\right) \sim \mathcal{N}\left(\mu_{t+1}^\beta, \sigma_\beta^2\left(S_{t+1}^\beta\right)^{-1}\right),$$

where $\mu_{t+1} = A_{t+1}^{-1}a_{t+1}$, and $\mu_{t+1}^\beta = \left(S_{t+1}^\beta\right)^{-1} m_{t+1}^\beta$. The sufficient statistics, s_{t+1}, are updated using the recursions

$$A_{t+1} = A_t + Z_t Z_t'$$

$$a_{t+1} = a_t + Z_t \cdot Y_{t+1}'$$

$$W_{t+1} = W_t + y_{t+1} \cdot y_{t+1}'$$

$$c_{t+1} = c_t + 1$$

$$C_{t+1} = C_0 + W_{t+1} + \mu_{t+1} A_{t+1} \mu_{t+1} - \mu_{t+1}' \cdot a_{t+1} - a_{t+1}' \cdot \mu_{t+1} + (\mu_{t+1} - \mu_0)' A_0 (\mu_{t+1} - \mu_0)$$

$$S_{t+1}^\beta = S_t^\beta + \beta_t^2$$

$$m_{t+1}^\beta = m_t^\beta + \beta_t \cdot \beta_{t+1}$$

$$B_{t+1} = B_t + \beta_{t+1}^2$$

$$g_{t+1} = g_t + 1$$

$$G_{t+1} = G_0 + B_{t+1} + \left(\mu_{t+1}^{\beta}\right)^2 S_{t+1}^{\beta} - \left(\mu_{t+1}^{\beta} \cdot m_{t+1}^{\beta}\right)^2 + \left(\mu_{t+1}^{\beta} - \mu_0^{\beta}\right)^2 S_0^{\beta},$$

where $W_0 = 0$, $B_0 = 0$, $Z_t = [\begin{array}{cc} 1 & x_t \end{array}]'$, and $Y_t = [\begin{array}{cc} r_t - \beta_t x_{t-1} & x_t \end{array}]'$.

C. SV Model

Rewrite the SV model with log-stochastic volatility, with an innocuous change of variables for the volatility process for convenience:

$$r_{t+1} = \alpha + \beta x_t + \exp\left(V_{t+1}^r/2\right)\varepsilon_{t+1}^r \tag{IA.8}$$

$$x_{t+1} = \alpha_x + \beta_x x_t + \exp\left(V_{t+1}^x/2\right)\varepsilon_{t+1}^x, \tag{IA.9}$$

$$V_{t+1}^r = \alpha_r + \beta_r V_t^r + \sigma_r \eta_{t+1}^r \tag{IA.10}$$

$$V_{t+1}^x = \alpha_v + \beta_v V_t^x + \sigma_v \eta_{t+1}^v. \tag{IA.11}$$

This model contains two latent state variables: the volatilities of returns and payout yields, $L_t = (V_t^r, V_t^x)$. The innovations in volatilities are independent of each other and to the shocks to returns and payout yields.

First, we propagate the volatility states,

$$V_{t+1}^r \sim \mathcal{N}\left(\alpha_r + \beta_r V_t^r, \sigma_r^2\right)$$

$$V_{t+1}^x \sim \mathcal{N}\left(\alpha_v + \beta_v V_t^x, \sigma_v^2\right).$$

Next, we resample particles using weights

$$w(L_{t+1}, s_t, \theta) \propto p\left(y_{t+1}|y_t, L_{t+1}, s_t, \theta\right) = \mathcal{N}\left(\begin{bmatrix} \alpha + \beta x_t \\ \alpha_x + \beta_x x_t \end{bmatrix}, \Sigma_{t+1}\right),$$

with

$$\Sigma_{t+1} = \begin{bmatrix} \exp(V^r_{t+1}) & \rho \exp(V^r_{t+1}/2 + V^x_{t+1}/2) \\ \rho \exp(V^r_{t+1}/2 + V^x_{t+1}/2) & \exp(V^x_{t+1}) \end{bmatrix}.$$

The parameters, $\theta = (\alpha, \beta, \alpha_x, \beta_x, \alpha_r, \beta_r, \sigma_r, \alpha_v, \beta_v, \sigma_v, \rho)$, are drawn from standard linear regression posteriors with the exception of the correlation coefficient, ρ,

$$p\left(\alpha, \beta, \alpha_x, \beta_x|s_{t+1}\right) \sim \mathcal{N}\left(vec(\mu_{t+1}), A^{-1}_{t+1}\right)$$

$$p\left(\sigma^2_r|s_{t+1}\right) \sim \mathcal{IG}\left(c_{t+1}, C_{t+1}\right)$$

$$p\left(\alpha_r, \beta_r|\sigma^2_r, s_{t+1}\right) \sim \mathcal{N}\left(\mu^r_{t+1}, \sigma^2_r\left(S^r_{t+1}\right)^{-1}\right)$$

$$p\left(\sigma^2_v|s_{t+1}\right) \sim \mathcal{IG}\left(d_{t+1}, D_{t+1}\right)$$

$$p\left(\alpha_v, \beta_v|\sigma^2_v, s_{t+1}\right) \sim \mathcal{N}\left(\mu^v_{t+1}, \sigma^2_v\left(S^v_{t+1}\right)^{-1}\right),$$

where $\mu_{t+1} = A^{-1}_{t+1}a_{t+1}$, $\mu^r_{t+1} = \left(S^r_{t+1}\right)^{-1} m^r_{t+1}$, and $\mu^v_{t+1} = \left(S^v_{t+1}\right)^{-1} m^v_{t+1}$.

The vector of sufficient statistics, s_{t+1}, is updated using the recursions:

$$A_{t+1} = A_t + Z'_t\Sigma^{-1}_{t+1}Z_t$$

$$a_{t+1} = a_t + Z'_t\Sigma^{-1}_{t+1}y_{t+1}$$

$$S^r_{t+1} = S^r_t + R_tR'_t$$

$$m^r_{t+1} = m^r_t + R_t \cdot V^r_{t+1}$$

$$W^r_{t+1} = W^r_t + \left(V^r_{t+1}\right)^2$$

$$c_{t+1} = c_t + 1$$

$$C_{t+1} = C_0 + W^r_{t+1} + {\mu^r_{t+1}}' S^r_{t+1}\mu^r_{t+1} - 2\, {\mu^r_{t+1}}' \cdot m^r_{t+1} + \left(\mu^r_{t+1} - \mu^r_0\right)' S^r_0 \left(\mu^r_{t+1} - \mu^r_0\right)$$

$$S^v_{t+1} = S^v_t + X_t X_t'$$

$$m^v_{t+1} = m^v_t + X_t \cdot V^x_{t+1}$$

$$W^v_{t+1} = W^v_t + \left(V^x_{t+1}\right)^2$$

$$d_{t+1} = d_t + 1$$

$$D_{t+1} = D_0 + W^v_{t+1} + {\mu^v_{t+1}}' S^v_{t+1}\mu^v_{t+1} - 2\, {\mu^v_{t+1}}' \cdot m^v_{t+1} + \left(\mu^v_{t+1} - \mu^v_0\right)' S^v_0 \left(\mu^v_{t+1} - \mu^v_0\right),$$

where $W^r_0 = 0$, $W^v_0 = 0$, and $Z_t = J \otimes [\ 1 \quad x_t\ ]$, J is the two-dimensional identity matrix, $R_t = [\ 1 \quad V^r_t\ ]'$, and $X_t = [\ 1 \quad V^x_t\ ]'$.

The correlation between the residuals of the return and payout regressions, ρ, is estimated from a grid. The probability of drawing a particular ρ is proportional to

$$\begin{vmatrix} 1 & \rho \\ \rho & 1 \end{vmatrix}^{-t/2} \cdot \exp\left(-1/2(S^{(11)}_{t+1} + S^{(22)}_{t+1} - 2\rho S^{(12)}_{t+1})/(1-\rho^2)\right),$$

where

$$S_{t+1} = S_t + \begin{bmatrix} \varepsilon^r_{t+1} \\ \varepsilon^x_{t+1} \end{bmatrix} [\ \varepsilon^r_{t+1} \quad \varepsilon^x_{t+1}\ ].$$

D. SV-DC Model

The SV-DC model has both stochastic volatility and a drifting predictability coefficient,

$$r_{t+1} = \alpha + \beta x_t + \beta_{t+1} x_t + \exp\left(V^r_{t+1}/2\right)\varepsilon^r_{t+1} \tag{IA.12}$$

$$x_{t+1} = \alpha_x + \beta_x x_t + \exp\left(V^x_{t+1}/2\right)\varepsilon^x_{t+1}, \tag{IA.13}$$

$$V^r_{t+1} = \alpha_r + \beta_r V^r_t + \sigma_r \eta^r_{t+1} \tag{IA.14}$$

$$V^x_{t+1} = \alpha_v + \beta_v V^x_t + \sigma_v \eta^v_{t+1} \tag{IA.15}$$

$$\beta_{t+1} = \beta_\beta \beta_t + \sigma_\beta \varepsilon^\beta_{t+1}. \tag{IA.16}$$

The model contains three latent state variables: the volatilities of returns and payout yields and the drifting coefficient, $L_t = (V^r_t, V^x_t, \beta_t)$.

The particle filter for this model combines the filter for the SV and CV-DC models. We propagate the volatilities as in the SV model. The resampling weights and drifting coefficient are calculated as in the CV-DC model, replacing Σ by Σ_{t+1} as defined above.

The sufficient statistics and posterior parameter distributions are the same as in the SV model, using $Y_t = [\ r_t - \beta_t x_{t-1} \quad x_t\ ]'$. The sufficient statistics for the drifting coefficient are as shown in the CV-DC model.

To summarize the particle filter algorithm, first we propagate the volatility states

$$\left(\begin{bmatrix} V^r_{t+1} \\ V^x_{t+1} \end{bmatrix}\right)^{(i)} \sim p\left(\begin{bmatrix} V^r_{t+1} \\ V^x_{t+1} \end{bmatrix} | L^{(i)}_t, \theta^{(i)}, y_{t+1}\right).$$

Second, we resample the particles $(L_{t+1}, s_t, \theta)^{(i)}$ with weights $w(L_{t+1}, s_t, \theta)$. Third, we update $\beta^{(i)}_{t+1} \sim p(\beta|(y_{t+1}, L^{(i)}_t, \theta^{(i)})$. Fourth, we update sufficient statistics $s^{(i)}_{t+1} = \mathcal{S}\left(s^{(i)}_t, L^{(i)}_{t+1}, y_{t+1}\right)$, and in the last step we draw $\theta^{(i)} \sim p\left(\theta|s^{(i)}_{t+1}\right)$.

VI. Savage Density Ratios

If we partition the parameter vector as $\theta = (\theta_{\mathcal{M}}, \theta_{-\mathcal{M}})$, then we are interested in the models given by $\mathcal{M}_0 : \theta_{-\mathcal{M}} = 0$ and $\mathcal{M}_1 : \theta_{-\mathcal{M}} \neq 0$. Here nesting means that the priors over the unrestricted parameters are the same across the two models,

$$p\left(\theta_{\mathcal{M}}|\mathcal{M}_0\right) = p\left(\theta_{\mathcal{M}}|\theta_{-\mathcal{M}} = 0, \mathcal{M}_1\right),$$

and that the likelihoods of the observed data are equal,

$$p\left(y_t|\theta_{\mathcal{M}}, \mathcal{M}_0\right) = p\left(y_t|\theta_{\mathcal{M}}, \theta_{-\mathcal{M}} = 0, \mathcal{M}_1\right).$$

These are formal definitions of "nesting." The main result is that the Bayes factor, $\mathcal{BF}_{0,1}$, equals

$$\mathcal{BF}_{0,1}^t = \frac{p\left(\theta_{-\mathcal{M}} = 0|y^t, \mathcal{M}_1\right)}{p\left(\theta_{-\mathcal{M}} = 0|\mathcal{M}_1\right)}.$$

To see this, note that by Bayes rule

$$\begin{aligned}
\frac{p\left(\theta_{-\mathcal{M}} = 0|y^t, \mathcal{M}_1\right)}{p\left(\theta_{-\mathcal{M}} = 0|\mathcal{M}_1\right)} &= \frac{p\left(y^t|\theta_{-\mathcal{M}} = 0, \mathcal{M}_1\right)}{p\left(y^t|\mathcal{M}_1\right)} \\
&= \frac{\int p\left(y^t|\theta_{\mathcal{M}}, \theta_{-\mathcal{M}} = 0, \mathcal{M}_1\right) p\left(\theta_{\mathcal{M}}|\theta_{-\mathcal{M}} = 0, \mathcal{M}_1\right) d\theta_{\mathcal{M}}}{p\left(y^t|\mathcal{M}_1\right)} \\
&= \frac{\int p\left(y^t|\theta_{\mathcal{M}}, \mathcal{M}_0\right) p\left(\theta_{\mathcal{M}}|\mathcal{M}_0\right) d\theta_{\mathcal{M}}}{p\left(y^t|\mathcal{M}_1\right)} \\
&= \frac{p\left(y^t|\mathcal{M}_0\right)}{p\left(y^t|\mathcal{M}_1\right)} = \mathcal{BF}_{0,1}^t.
\end{aligned}$$

This takes the convenient form of a ratio of ordinates, both computed under the more general model. The denominator is just an ordinate of the prior distribution, $p\left(\theta_{-\mathcal{M}} = 0|\mathcal{M}_1\right)$. The numerator is

$$p\left(\theta_{-\mathcal{M}} = 0|y^t, \mathcal{M}_1\right) = \int p\left(\theta_{-\mathcal{M}} = 0|\theta_{\mathcal{M}}, s_t\right) p\left(\theta_{\mathcal{M}}, s_t^{\mathcal{M}}|y^t\right) d\left(\theta_{\mathcal{M}}, s_t^{\mathcal{M}}\right),$$

which is in the familiar "Rao-Blackwellization" form common for efficient Monte Carlo in MCMC settings. The estimate is given by

$$p^N\left(\theta_{-\mathcal{M}}=0|y^t,\mathcal{M}_1\right)=\frac{1}{N}\sum_{k=1}^{N}p\left(\theta_{-\mathcal{M}}=0|\left(\theta_{\mathcal{M}},s_t^{\mathcal{M}}\right)^k\right),$$

where $(\theta_{\mathcal{M}},s_t)^k$ are samples from $p^N\left(x_t,s_t^{\mathcal{M}_i},\theta_{\mathcal{M}_i}|\mathcal{M}_i,y^t\right)$.

REFERENCES

Boudoukh, Jacob, Roni Michaely, Michael Roberts, and Matthew Richardson, 2007, On the importance of measuring payout yield: Implications for empirical asset pricing, *Journal of Finance* 62, 877-915.

Brav, Alon, John R. Graham, Campbell R. Harvey, and Roni Michaely, 2005, Payout policy in the 21st century, *Journal of Financial Economics* 77, 483-527.

Dangl, Thomas, and Michael Halling, 2012, Predictive regressions with time-varying coefficients, *Journal of Financial Economics* 106, 157-181.

Henkel, Samuel J., J. Spencer Martin, and Federico Nardari, 2011, Time-varying short-horizon return predictability, *Journal of Financial Economics* 99, 560-580.

Johannes, Michael, Nicholas G. Polson, and Jonathan R. Stroud, 2002, Sequential optimal portfolio performance: Market and volatility timing, Working paper, University of Chicago.