Particle Filtering and Learning in Finance

Nick Polson[1]

[1]University of Chicago GSB
Joint work with Michael Johannes (Columbia) and Arthur Korteweg (Stanford)

October, 2008

Overview

- Particle Filtering, Learning and Smoothing
- Conditional Sufficient Statistics
- Filtering and Learning works for empirical asset pricing problems.
- Sequential Predictive Regressions: Dividend yields and Payout ratios
- Carvalho, Johannes, Lopes and Polson (2008). Particle Learning and Smoothing

Filtering

- Data y_t depends on a latent state variable, x_t.

$$\text{Observation equation: } y_t = f\left(x_t, \varepsilon_t^y\right)$$
$$\text{State evolution: } x_{t+1} = g\left(x_t, \varepsilon_{t+1}^x\right),$$

- Posterior distribution of $p\left(x_t | y^t\right)$ where $y^t = (y_1, ..., y_t)$
- Kalman filter, FFBS (Filter Forward Backwards Sample) Discrete HMM (Baum-Welch, Viterbi, Scott)
- Prediction and Bayesian updating.

$$p\left(x_{t+1} | y^t\right) = \int p\left(x_{t+1} | x_t\right) p\left(x_t | y^t\right) dx_t, \tag{1}$$

updated by Bayes rule

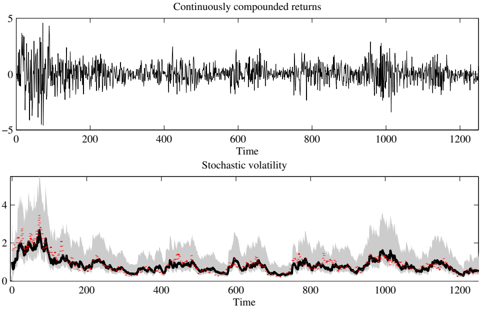
$$\underbrace{p\left(x_{t+1} | y^{t+1}\right)}_{\text{Posterior}} \propto \underbrace{p\left(y_{t+1} | x_{t+1}\right)}_{\text{Likelihood}} \underbrace{p\left(x_{t+1} | y^t\right)}_{\text{Prior}}. \tag{2}$$

SV: Motivating Example

- SV models (JPR, EJP)

$$y_{t+1} = \sqrt{V_{t+1}}\varepsilon^{y}_{t+1}$$
$$\log\left(V_{t+1}\right) = \alpha_v + \beta_v \log\left(V_t\right) + \sigma_v \varepsilon^{v}_{t+1},$$

Typical parameters $\alpha_v = 0$, $\beta_v = 0.95$, and $\sigma_v = 0.10$.


Continuously compounded returns
5
0
−5
0
200
400
600
800
1000
1200
Time
Stochastic volatility
4
2
0
0
200
400
600
800
1000
1200
Time

Algorithm

- Track $(\theta, x_t, s_t)^{(i)}$ including sufficient statistics.
- Algorithm: Exact state filtering and parameter learning

Step 1: Draw $(\theta, x_t, s_t)^{(i)} \sim Multi_N\left(\left\{w\left((x_t, \theta)^{(i)}\right)\right\}_{i=1}^{N}\right)$

Step 2: Draw $x_{t+1}^{(i)} \sim p\left(x_{t+1} | (x_t, \theta)^{(i)}, y_{t+1}\right)$

Step 3: Update sufficient statistics:
$s_{t+1}^{(i)} = \mathcal{S}\left(s_t^{(i)}, x_{t+1}^{(i)}, y_{t+1}\right)$ for $i = 1, ..., N$

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

Conditional Suffi cient Statistics

- Parameter posteriors and sufficient statistics,

$$x_{t+1} = Z_t'\beta + \sigma_x \sqrt{\omega_{t+1}}\epsilon_{t+1}$$

  where $Z_t = (1, x_t)'$ and $\beta = (\alpha_x, \beta_x)'$.

- Posteriors $p\left(\sigma^2 | s_{t+1}\right) \sim \mathcal{IG}\left(a_{t+1}, A_{t+1}\right)$, $p\left(\sigma_x^2 | s_{_{t+1}}\right) \sim \mathcal{IG}\left(b_{t+1}, B_{t+1}\right)$, and $p\left(\beta | \sigma_x^2, s_{t+1}\right) \sim \mathcal{N}\left(c_{t+1}, \sigma_x^2 C_{t+1}^{-1}\right)$.

- Hyperparameters

$$A_{t+1} = \frac{(y_{t+1} - x_{t+1})^2}{\lambda_{t+1}} + A_t$$

$$B_{t+1} = B_t + c_t' C_t c_t + \frac{x_{t+1}^2}{\omega_{t+1}} - c_{t+1}' C_{t+1} c_{t+1}$$

$$c_{t+1} = C_{t+1}^{-1} \left(C_t c_t + \frac{Z_t' x_{t+1}}{\omega_{t+1}} \right)$$

$$C_{t+1} = C_t + \frac{Z_t Z_t'}{\omega_{t+1}},$$

which defines the vector of sufficient statistics, $s_{t+1} = (A_{t+1}, B_{t+1}, c_{t+1}, C_{t+1})$.

Predictive Regressions

- Stock market return predictability often regarded as stylized fact.
- Standard predictive regression model:

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

- How does the evidence for predictability arise?
    - Slow accumulation over time?
    - Or concentrated in particular sub-periods?

Big Picture

- How do investors make (portfolio) decisions?
  - They use models, considering multiple ones: Standard model? Stochastic volatility? Is β constant?
  - They learn about which model(s) work as new information arrives.
  - They care about uncertainty: estimation risk, state variables and models themselves.
- Current academic research:
  - Rational expectations: exact nature of predictability is known.
  - Model diagnostics: chose one "best" model.
  - Bayesian learning: one-shot problem.

Our Approach

- Sequential learning
- Portfolio formation
  - Based on currently available evidence
  - Taking all uncertainty into account: parameters, states and models.

Overview of Results

- Inference on predictability:
  - Predictability is stronger for net payout than for dividend yield.
  - The classical constant coeffi cients, constant volatility models is rejected for both payout measures.
  - Learn about predictability faster in stochastic volatility models.
- Portfolio formation:
  - Form portfolios that take into account all sources of uncertainty (parameter, state and model).
    - Learning.
    - Conditional skewness and kurtosis.
  - Stochastic volatility important for portfolio formation.

Models

- Standard predictive regression model:

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

  - r_{t+1} is the value weighted excess market return; x_t is a maesure of yield (cash dividend yield or net payout yield of Boudoukh et al., 2007).
  - The errors have correlation ρ (Stambaugh, 1999)
- Statistical issues: asymptotics due to unit roots, correlated errors, x_0, long-horizon regressions.
- Specification issues:
  - Time-varying volatility and the relationship between x_t and expected returns

Drifting Coeffi cients

- Theory: Common models (e.g. Santos and Veronesi, 2005) imply that the loading on x_t varies over time (surplus consumption ratio).
- Empirically, ample evidence that relationship between predictors and equity premium changes over time.
  - Lettau and Van Nieuwerburgh (2007).
  - Abrupt shifts vs. slow drifts: how does the economy change over time?

Model Extensions

- Stochastic volatility (SV) and "drifting" coefficients (DC):

$$
\begin{aligned}
r_{t+1} &= \alpha + \beta_0 x_t + \beta_t x_t + \exp\left(V_t^r/2\right)\varepsilon_{t+1}^r \\
x_{t+1} &= \alpha_x + \beta_x x_t + \exp\left(V_t^x/2\right)\varepsilon_{t+1}^x
\end{aligned}
$$

- Latent state variables:
  - V_t are stochastic volatilities (AR(1) processes)
  - β_t is an AR(1) with long-run mean zero:

$$
\beta_{t+1} = \beta_\beta \beta_t + \sigma_\beta \varepsilon_{t+1}^\beta
$$

- Models: benchmark, SV, DC, SVDC.

Bayesian Inference

- We want the distribution of the unknown, conditional on the known: the posterior distribution.
  - Unknowns: parameters (θ) and latent states (L_t), and models.
  - Known: observed returns and dividend yields.
- With $j = 1 \ldots J$ models ($\mathcal{M}_j$), need to compute

$$p(\theta, L^t, \mathcal{M}_j | y^t)$$

where $y^t = \{r_\tau, x_\tau\}_{\tau=1}^t$ is the observed data.

Model Inference

- It is useful to decompose the posterior into two components:
  - Parameter and state variable inference within a model.
  - Inference across models.

$$p\left(L^t, \theta, \mathcal{M}_j | y^t\right) = p\left(L^t, \theta | \mathcal{M}_j, y^t\right) p\left(\mathcal{M}_j | y^t\right)$$

- The fist term is what we usually focus on, decomposed into a likelihood and prior:

$$p\left(L^t, \theta | \mathcal{M}_j, y^t\right) \propto p(y^t | L^t, \theta, \mathcal{M}_j) p(L^t, \theta | \mathcal{M}_j)$$

- The second term is a posterior model probability.
  - Classical methods pick the "best" model. Bayesians compare models and average across models, following the rules of probability.

Sequential Model Choice

- The criterion for comparing models is the posterior odds ratio:

$$odds\left(\mathcal{M}_j \text{ vs. } \mathcal{M}_k | y^t\right) = \frac{p\left(\mathcal{M}_j | y^t\right)}{p\left(\mathcal{M}_k | y^t\right)} = \frac{p\left(y^t | \mathcal{M}_j\right)}{p\left(y^t | \mathcal{M}_k\right)} \frac{p\left(\mathcal{M}_j\right)}{p\left(\mathcal{M}_k\right)}$$

- The term $p\left(\mathcal{M}_j\right)/p\left(\mathcal{M}_k\right)$ is the prior odds ratio.
- The Bayes Factor is:

$$\mathcal{BF}_{j,k}^t = \mathcal{LR}_{j,k}^t = \frac{p(y^t|\mathcal{M}_j)}{p(y^t|\mathcal{M}_k)} = \frac{p(y^t|y^{t-1},\mathcal{M}_j)}{p(y^t|y^{t-1},\mathcal{M}_k)}\mathcal{LR}_{j,k}^{t-1}$$

  - Essentially a likelihood ratio, but where all other aspect of uncertainty (parameters, states) are accounted for.
  - "Automatic" Occam's razor: Bayes Factors automatically punish needlessly complicated models by integrating out all sources of uncertainty.
  - Sequential model monitoring: how and when does the weight of evidence shift?

Optimal Portfolio Allocation

- Investors maximize expected utility:

$$\max_{\omega_t} E_t \left[U \left(r^p_{t+1} \right) \right]$$

where $r^p_{t+1} = 1 + r^f_t + \omega_t \left(r_{t+1} - r^f_t \right)$ and $U \left(r^p_{t+1} \right) = \frac{\left(r^p_{t+1} \right)^{1-\gamma}}{1-\gamma}$.

- Static and not fully rational:
  - Does not take into account wealth in future states.
  - No intertemporal hedging.
  - Dynamic problem is intractable: how to deal with the fact that I know today that I will be revising my beliefs in the future?
- Model averaging:

$$E_t \left[U \left(r^p_{t+1} \right) \right] = \sum_{j=1}^{J} E_t \left[U \left(r^p_{t+1} \right) | \mathcal{M}_j \right] p \left(\mathcal{M}_j | y^t \right)$$

$$= \sum_{j=1}^{J} \int U \left(r^p_{t+1} \right) p \left(r^p_{t+1} | \mathcal{M}_j, y^t \right) p \left(\mathcal{M}_j | y^t \right) dr^p_{t+1}$$

Performance Evaluation

- After computing portfolio weights, a Bayesian investor waits for data.
  - After realizing portfolio returns, revise beliefs: recompute parameter and state variables estimates, update model probabilities and recompute expected utility.
- What about out-of-sample performance evaluation? Sharpe Ratios, CE, etc.
  - Play no formal role in the Bayesian approach: model probabilities take care of everything.
  - Example: $\mu =$ 25%/year, which generates very high expected utility and certainty equivalents compared to other models.
  - At best, some sort of crude model specifi cation tool.
- Interesting contrasts between "performance" of optimal strategies compared to others.

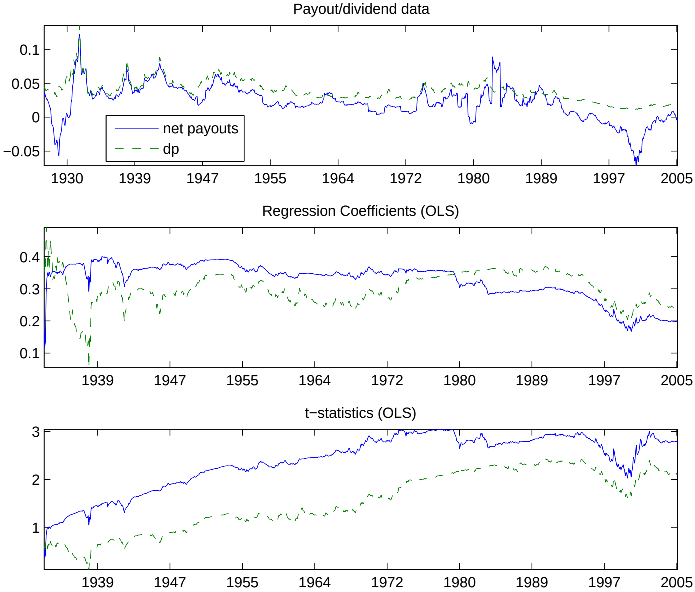
Computation: Particle Learning (PL)

- Compute posteriors for each model and model probabilities $\forall t$.
- Markov Chain Monte Carlo (MCMC) – too expensive
  Difficult to compute marginal likelihoods for model probabilities.
- Particle Learning: fully recursive
  1. Approximate filters for nonlinear or nonnormal settings (Kalman doesn't apply).
  2. Dominant paradigm in all signal tracking problems: robots, radar (military or weather), missile guidance, traffic, GPS, etc.
  3. Accurate and computationally attractive.
- Problem: typically used where models and parameters are known. Here add learning.

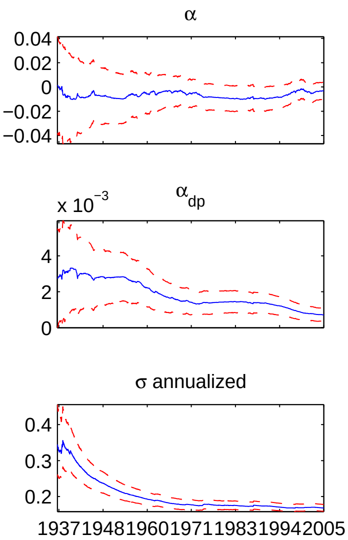
Data

- Value-weighted monthly equity returns, 1927-2005.
- Dividend yield:
  - Traditional cash dividends.
  - Net payout yield: Boudoukh et al. (2007)
- Measuring cash payouts is not easy:
  - Prior to 1983: CRSP changes in shares outstanding.
  - From 1983 onwards: Use issuances and repurchases from statement of cash flows
- Structural breaks?
  - Boudoukh et al. (2007) test for and reject a structural change.

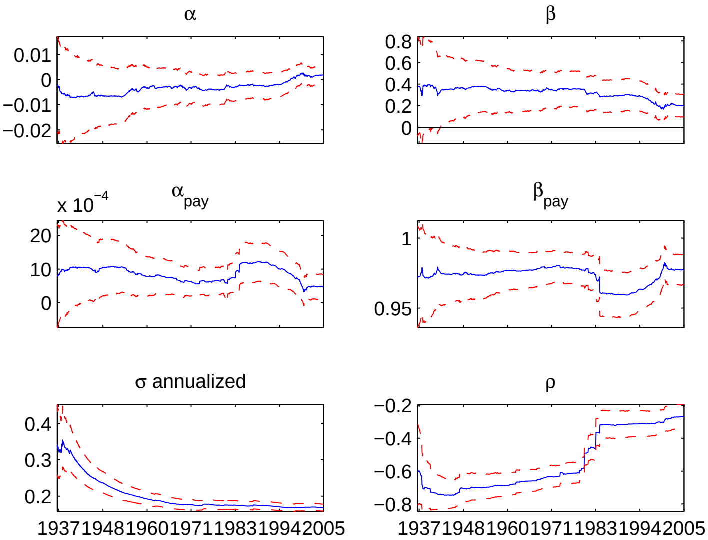
Cumulative Regressions (OLS)



Particle Estimates: Benchmark Model, Dividend Yield



Benchmark Model, Net Payout Yield



Hypothesis Testing

- Test $\mathcal{H}_0 : \beta = 0$ by computing Bayes factors:

$$\mathcal{BF}_{0,1}^{t} = \frac{p\left(\mathcal{H}_0 | y^t\right)}{p\left(\mathcal{H}_1 | y^t\right)} = \frac{p\left(\beta = 0 | y^t, \mathcal{H}_1\right)}{p\left(\beta = 0 | \mathcal{H}_1\right)}$$

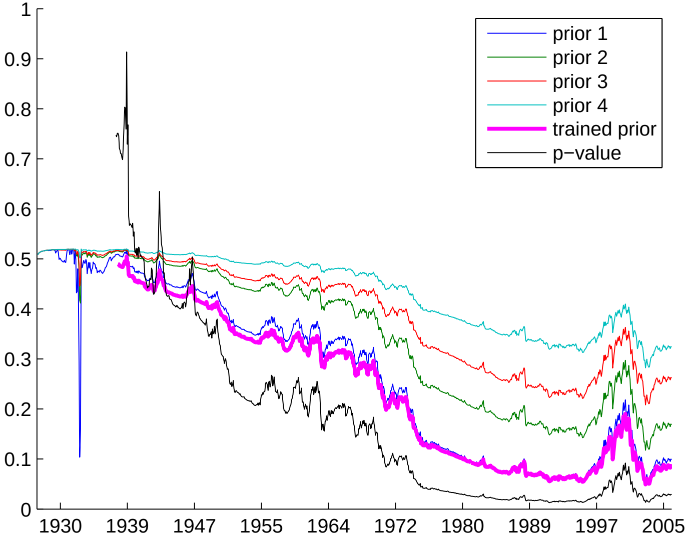
- Convert to probabilities:

$$\text{Prob}\left[\mathcal{H}_0 | y^t\right] = \frac{\mathcal{BF}_{0,1}^{t}}{1 + \mathcal{BF}_{0,1}^{t}}$$

Priors

- We use conjugate priors, which admit sufficient statistics for posteriors.
- We use the 1927-1936 period to “train” the priors.
  - Estimate regressions, take point estimates and standard errors to formulate priors.
  - Avoids the need for subjective priors.
  - Exception: Drifting coeffi cients model: calibrate prior to imply on average high persistence (0.95 monthly autocorrelation, with standard deviation 0.1).

Lindley's Paradox



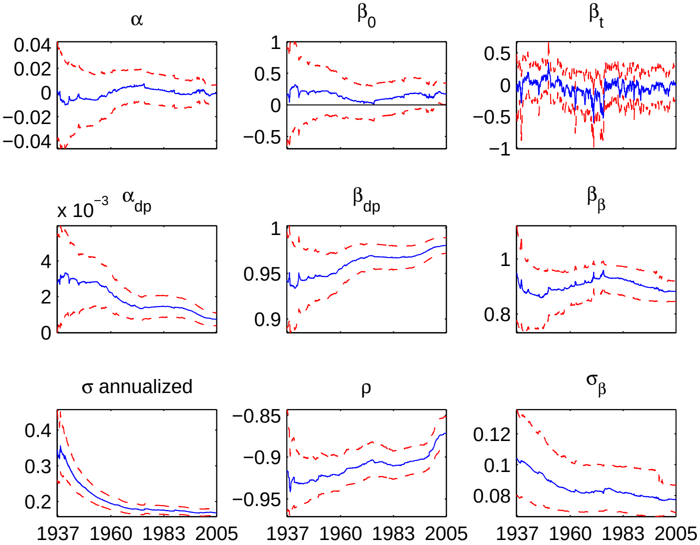
Lindley's Paradox

- Conflict between p-values and Bayes factors in large samples.
  - T-statistic is about 2 to 2.5: signifi cant at the 5% level.
  - Posterior probability for d/p in benchmark model is about 10-15%.
- Why? Large-sample approximation:

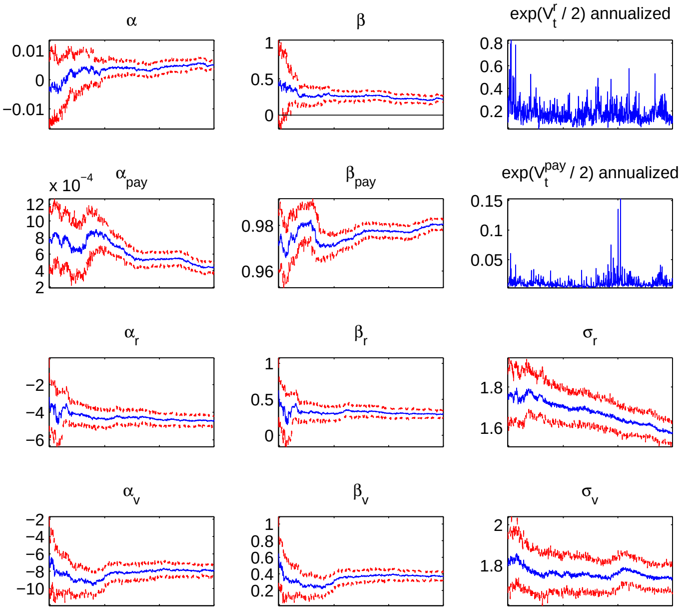
$$\mathcal{BF}_{0,1}^{T} \approx \sqrt{T} \exp\left(-\frac{t_T^2}{2}\right)$$

  - $T = 900$, $t_T = 2$ implies $\mathcal{BF} = 1.3$ and posterior probability is approximately 80%.
  - $T = 900$, $t_T = 3$ implies $\mathcal{BF} = 0.32$ and posterior probability is approximately 25%.
- Need to reduce p-values in large samples.

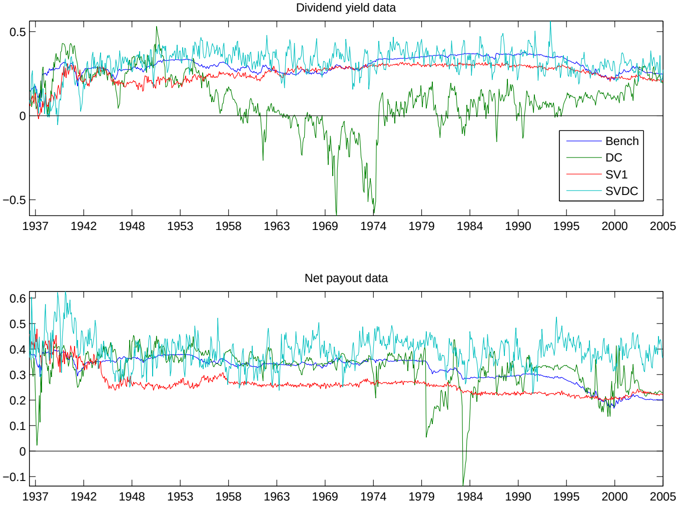
DC, Dividend Yield



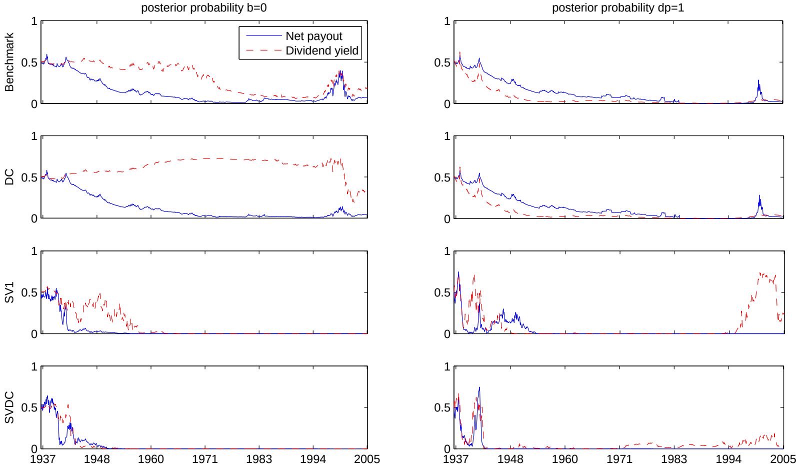
SV, Net Payout Yield



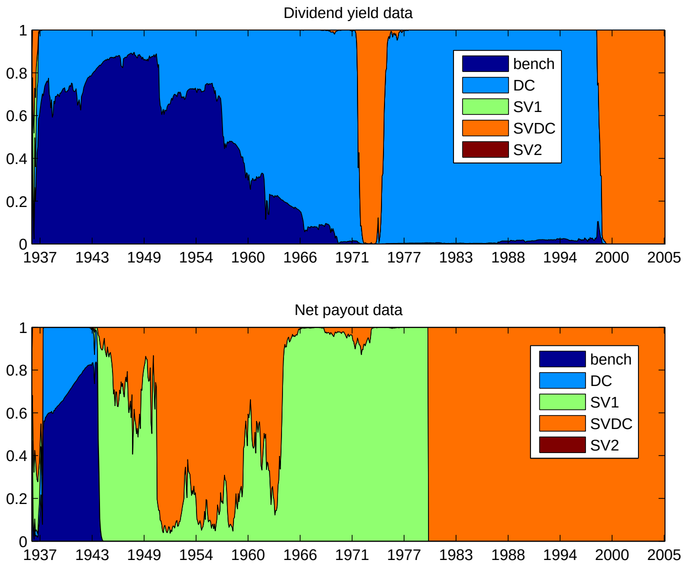
Predictability Estimates across Models



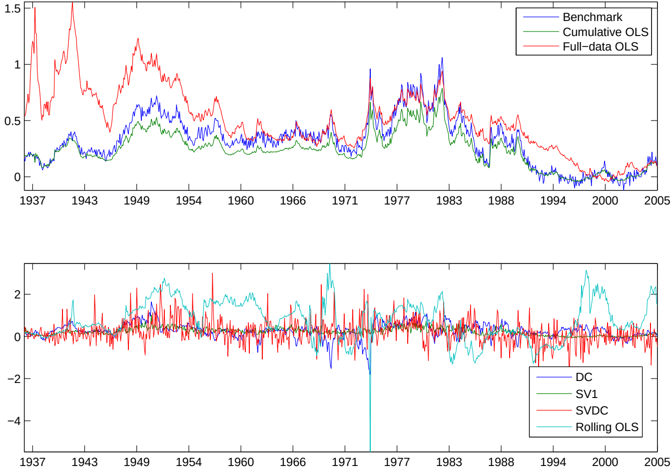
Hypothesis Tests



Sequential Model Probabilities



Optimal Portfolio Weights, Dividend Yield ($\gamma = 4$)



Portfolio Stats: Dividend Yield

$\gamma = 4$	mean	s.d.	skew	kurt	S.R.	C.E.	APY
Full-data OLS	9.16	10.57	0.60	24.35	0.14	0.78	8.58
Cumulative OLS	6.20	4.72	1.30	13.57	0.14	-0.34	6.08
Rolling OLS	10.11	20.43	-4.41	66.20	0.09	-175.45	6.78
Benchmark	6.89	6.32	1.33	15.20	0.13	0.00	6.67
DC	7.34	8.03	-1.54	42.58	0.12	-0.23	6.99
SV1	11.03	12.42	1.13	13.53	0.16	1.96	10.24
SVDC	8.87	13.22	-1.04	22.61	0.11	-1.39	7.94
Model-avge	7.28	7.71	-1.88	50.35	0.12	-0.20	6.85
SV2	7.52	8.09	-0.16	6.31	0.13	0.09	7.17
$\gamma = 6$	mean	s.d.	skew	kurt	S.R.	C.E.	APY
Full-data OLS	7.43	7.05	0.55	24.24	0.14	0.53	7.16
Cumulative OLS	5.45	3.20	1.43	13.36	0.14	-0.23	5.39
Rolling OLS	8.06	13.61	-4.39	66.09	0.09	-22.28	6.93
Benchmark	5.91	4.25	1.47	15.33	0.13	0.00	5.81
DC	6.23	5.39	-1.41	41.29	0.12	-0.12	6.06
SV1	8.82	8.94	1.79	19.86	0.16	1.20	8.40
SVDC	7.46	9.48	-1.14	30.14	0.11	-1.26	6.98
Model-avge	6.22	5.18	-1.69	48.67	0.12	-0.07	6.01
SV2	6.40	5.60	-0.13	6.37	0.13	0.07	6.23
$\gamma = 8$	mean	s.d.	skew	kurt	S.R.	C.E.	APY
Full-data OLS	6.56	5.30	0.49	23.98	0.14	0.41	6.40
Cumulative OLS	5.08	2.46	1.51	12.92	0.14	-0.17	5.04
Rolling OLS	7.03	10.21	-4.36	65.85	0.09	-11.52	6.43
Benchmark	5.42	3.24	1.57	14.89	0.13	0.00	5.35
DC	5.64	4.09	-1.23	39.64	0.12	-0.10	5.54
SV1	7.67	7.00	2.24	23.75	0.15	0.88	7.41
SVDC	6.65	7.22	-0.80	29.51	0.06	-0.83	6.37
Model-avge	5.67	3.93	-1.48	46.95	0.12	-0.02	5.54
SV2	5.79	4.29	-0.10	6.45	0.12	0.04	5.69

Portfolio Stats: Net Payout Yield

$\gamma = 4$	mean	s.d.	skew	kurt	S.R.	C.E.	APY
Full-data OLS	9.52	11.54	0.38	18.66	0.14	0.99	8.83
Cumulative OLS	6.55	6.60	1.14	14.54	0.11	-0.12	6.32
Rolling OLS	11.43	18.31	-2.70	29.45	0.12	-12.83	9.39
Benchmark	7.21	8.49	1.21	15.93	0.11	0.00	6.84
DC	7.64	8.63	0.51	17.63	0.12	0.34	7.25
SV1	6.69	6.01	0.80	10.52	0.13	0.16	6.49
SVDC	6.91	10.10	1.28	16.75	0.08	-0.87	6.40
Model-avge	6.79	6.04	0.79	10.35	0.13	0.25	6.59
SV2	7.46	8.16	-0.33	6.50	0.12	0.29	7.10
$\gamma = 6$	mean	s.d.	skew	kurt	S.R.	C.E.	APY
Full-data OLS	7.71	7.70	0.33	18.65	0.14	0.68	7.39
Cumulative OLS	5.72	4.43	1.20	14.56	0.11	-0.08	5.61
Rolling OLS	8.98	12.19	-2.70	40.89	0.12	-4.80	5.61
Benchmark	6.17	5.71	1.29	16.16	0.11	0.00	5.99
DC	6.49	5.92	0.97	22.05	0.12	0.23	6.30
SV1	5.89	4.23	1.03	12.05	0.13	0.15	5.79
SVDC	6.08	7.22	2.19	28.03	0.08	-0.61	5.81
Model-avge	5.97	4.27	0.96	12.03	0.13	0.21	5.86
SV2	6.40	5.63	-0.29	6.68	0.12	0.21	6.22
$\gamma = 8$	mean	s.d.	skew	kurt	S.R.	C.E.	APY
Full-data OLS	6.80	5.78	0.28	18.55	0.14	0.51	6.61
Cumulative OLS	5.31	3.36	1.25	14.38	0.11	-0.06	5.24
Rolling OLS	7.75	9.14	-2.69	40.64	0.12	-2.94	7.29
Benchmark	5.64	4.30	1.34	16.07	0.11	0.00	5.54
DC	5.87	4.46	1.00	22.16	0.12	0.16	5.75
SV1	5.44	3.29	1.12	12.91	0.12	0.09	5.38
SVDC	5.62	5.79	2.95	43.27	0.08	-0.54	5.44
Model-avge	5.51	3.31	1.09	12.90	0.13	0.15	5.44
SV2	5.83	4.30	-0.23	6.58	0.12	0.15	5.72

Extensions

- Economic restrictions on priors
- Intertemporal hedging
- Long-horizon predictability
    - Larger role for parameter uncertainty in portfolio formation.

Conclusions

- Stock return predictability examined sequentially.
  - Predictability is stronger for net payout than for dividend yield.
  - The classical constant coeffi cients, constant volatility models is rejected for both payout measures.
  - Learn about predictability faster in stochastic volatility models.
- Portfolio formation:
  - Form portfolios that take into account all sources of uncertainty (parameter, state and model).
    - Learning.
    - Conditional skewness and kurtosis.
  - Stochastic volatility important for portfolio formation.