On the Long Run Volatility of Stocks

Robert Mcculloch
University of Chicago

Joint work With:
Carlos Carvalho, University of Texas
Hedibert Lopes, George Washington University

University of Minnesota, October 24, 2013

The Question

A Conventional Wisdom:

If I'm investing over the long haul (eg pension for a young person),

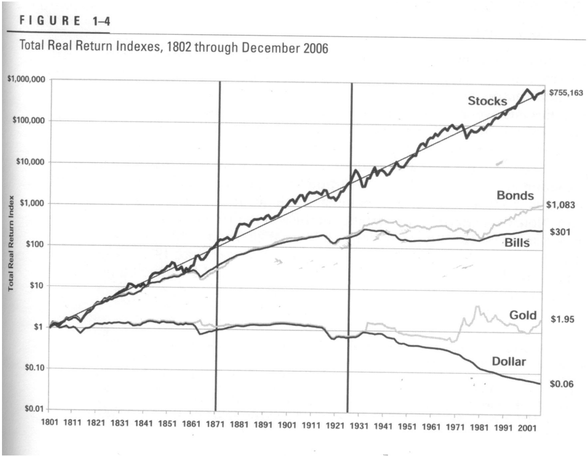
I'll get a good return out of the stock market

and I'm fairly safe !!!

..over the last century, accumulations in stocks have always outperformed other financial assets for the patient investor.

Page 5, "Stocks for the Long Run", 4^{th} ed, by Jeremy Siegel. (Russell E. Palmer professor of Finance at Wharton, U Penn).

Stocks for the Long Run: Conventional Wisdom



3

Page 11, Siegel.

Simple Benchmark Model:

r_t is return in t^{th} period,
the annual real log return on the U.S. equity market:

$$r_t \sim N(\mu, \sigma^2),\ iid, \quad \approx \quad \textit{random walk on prices}$$

Returns k periods in the future:

$$r_{t,t+k} = r_{t+1} + r_{t+2} + \ldots + r_{t+k}$$

$$Var(r_{t,t+k}) = k\,\sigma^2$$

The variance, per period, is constant over any investment horizon.

We use

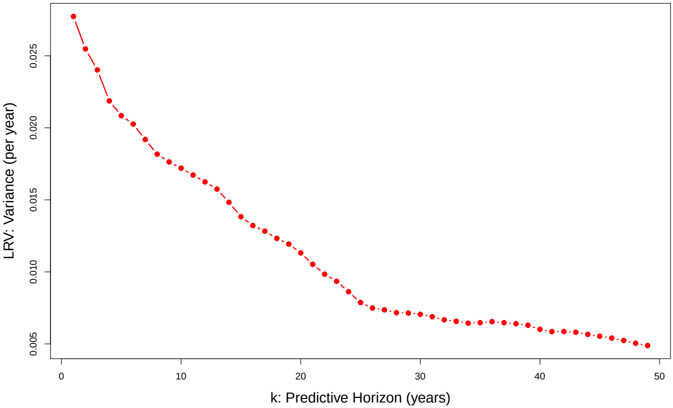
$$LRV = \frac{Var(r_{t,t+k})}{k} = \frac{Var(r_{t+1} + r_{t+2} + \ldots + r_{t+k})}{k}$$

the variance per period over horizon k as our measure of "long run volatility".

- Under certain utility functions, this is the thing that matters.
- Easy to compare other models to simple benchmark model.

Under simple iid benchmark model, LRV is flat over the investment horizon.

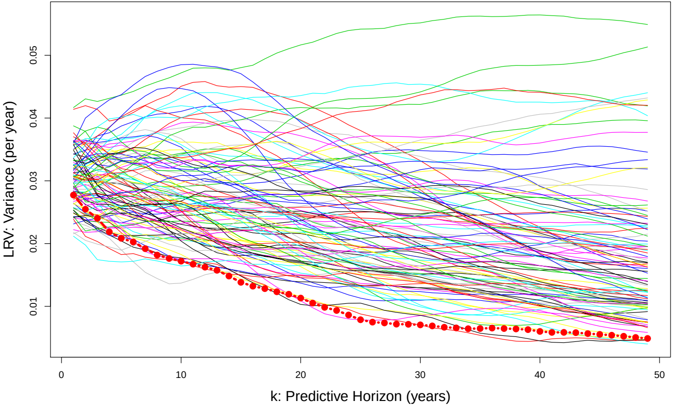
x axis: k, number of years in holding period = investment horizon.
y axis: Long run volatility.



Based on annual data on “the market return” from 1802 to 2009. For each horizon, roll a window adding up blocks of returns having length equal to the horizon.

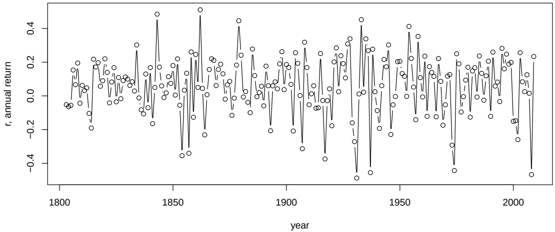
The LRV decreases with the length of the horizon !!

However, you can get a wide range of paths from this rolling window if you simulate iid data.

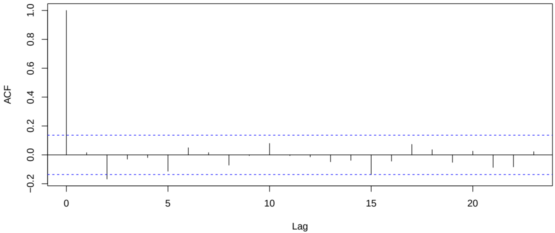


Red dots is from the data (same as previous slide), rest are from simulated data: $r_t \sim N(\bar{r}, s_r^2)$ iid.

Here is the data.



Looks close to iid !!



But, if we can find the smallest amount of "structure", it really matters !!!

We will also have *predictor variables*.

So, there is a tremendous amount of uncertainty.

Can we model the returns process to help us investigate LRV?

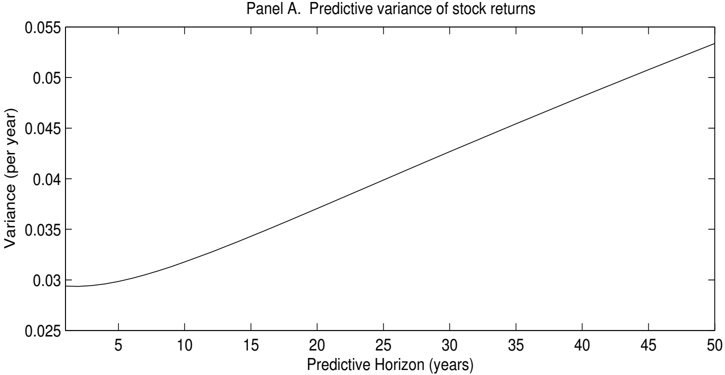
The two key papers are Barberis (2000) and Pastor and Stambaugh (2011).

Both papers take a Bayesian approach.

Barberis (2000) used a simple dynamic model and obtained the result that the LRV declines in the horizon.

Paster and Stambaugh (2011) used a richer model *and* strong prior information to obtain the result that ....

Stocks for the Long Run:
Pastor and Stambaugh 2011 "main result"



LRV increases with the horizon !!!!

5

Flatly contradicts stocks for the long run !!

Summary:

- A simple view of the world suggests stocks are less volatile over long horizons (Barberis (2000)).
- ... while a more complex view states that stocks could be more volatile over long horizons (Pastor and Stambaugh (2011)).
- Which direction is right? Our work hopes to address:
    1. Better understand the sensitivity of the results to prior specification.
    2. Enrich PS2011 framework to potentially improve the assessment of long run volatility.
- Have (1), still working on (2).
- Our prelminary results indicate that Carlos is not crazy for having 100% equities in his retirement portfolio !

Questions:

- How do Pastor and Stambaugh get a model that allows for the possibility of increasing *or* decreasing LRV?
- What is the Barberis model (he got decreasing LRV)?
- What prior information do the financial economists bring to the model?
- How do we introduce "predictor variables" to the model?

This is crazy.

I have low signal data, and I need to investigate the consequences of a long run prediction.

The Models

Pastor and Stambaugh Predictive System (simple no predictor version)

r_t: return on market, period t.

$$\begin{aligned} r_{t+1} &= \mu_t + u_{t+1} \\ \mu_{t+1} &= \alpha + \beta\,\mu_t + w_{t+1} \end{aligned}$$

We would want $|\beta| < 1$ to ensure stationarity.

This looks like our simplest state-space model:

State equation: AR(1)
Observation equation: state plus error.

Would usually have (in "statistical applications") $u_{t+1} \perp w_{t+1}$.

$$
\begin{aligned}
r_{t+1} &= \mu_t + u_{t+1} \\
\mu_{t+1} &= \alpha + \beta\, \mu_t + w_{t+1}
\end{aligned}
$$

However, PS interpret μ_t as

$$\mu_t = E(r_{t+1} \mid I_t)$$

where I_t is "information available at time t".

Then, it is natural to imagine:

$$\begin{pmatrix} u_{t+1} \\ v_{t+1} \end{pmatrix} \sim N(0, \Sigma)$$

where Σ need not be diagonal since both u and w are impacted by the same set of new information.

$$
\begin{aligned}
r_{t+1} &= \mu_t + u_{t+1} \\
\mu_{t+1} &= \alpha + \beta\,\mu_t + w_{t+1}
\end{aligned}
$$

Key prior information (from PS):

> *The predictive system allows us to explore roles for a variety of prior beliefs about the behavior of expected returns, chief among which is the belief that unexpected returns (u_{t+1}) are negatively correlated with innovations in expected return (w_{t+1}).*

Pastor and Stambaugh use economic arguments to motivate the strong belief that

$$\rho_{u,w} < 0.$$

$$
\begin{aligned}
r_{t+1} &= \mu_t + u_{t+1} \\
\mu_{t+1} &= \alpha + \beta\, \mu_t + w_{t+1}
\end{aligned}
$$

With this model we have

$$
\frac{Var(r_{t,t+k})}{k} = \sigma_{11}\,[1 + 2A(k)\,\rho_{uw} + B(k)].
$$

where $A(k)$ and $B(k)$ are positive and σ_{11} is the variance of u_t..

With negative ρ_{uw} we can get LRV to decline with the horizon!!

But, it doesn't have to !!

In the data, the short run dependence is very low (looks iid).

But, the model allows us to look for a long run, low strength, dependent structure giving variance reduction over long horizons.

In particular, with $0 < \beta < 1$ we get mean reversion in the underlying state.

We are not trying to fit the data ($\mu_t \approx r_t$), we are trying to find a long run, low signal dependence structure which gives declining LRV for long horizons.

What is the Barberis model?

$$\begin{aligned} r_{t+1} &= a + b\,x_t + u_{t+1} \\ x_{t+1} &= \alpha + \beta\,x_t + w_{t+1} \end{aligned}$$

$$\begin{pmatrix} u_{t+1} \\ w_{t+1} \end{pmatrix} \sim N(0, \Sigma)$$

Here, x is a variable that we observe and think can help us predict the returns (one x above for simplicity, full model will have more).

This is a very natural model:

- The first equation tells us how the next r relates to our current x.
- To predict several periods ahead, we will have to predict x.

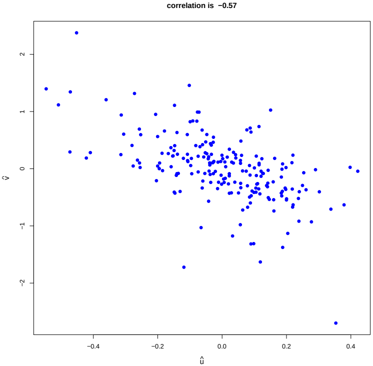
Barberis also allows for correlated errors.

$$
\begin{aligned}
r_{t+1} &= a + b\,x_t + u_{t+1} \\
x_{t+1} &= \alpha + \beta\,x_t + w_{t+1}
\end{aligned}
$$

x: dividend yield.
$\hat{u}$: resids from regression of r_{t+1} on x_t.
$\hat{w}$: resids from regression of x_{t+1} on x_t.

Indirect, empirical evidence in support of negative correlation.

corr = -.57.



Barberis, PS, and Imperfect Predictors:

PS:

$$\begin{aligned} r_{t+1} &= \mu_t + u_{t+1} \\ \mu_{t+1} &= \alpha + \beta\, \mu_t + w_{t+1} \end{aligned}$$

Barberis:

$$\begin{aligned} r_{t+1} &= a + b\, x_t + u_{t+1} \\ x_{t+1} &= \alpha + \beta\, x_t + w_{t+1} \end{aligned}$$

PS interpret the Barberis model as making the strong assumption of *perfect predictors*.

That is, they think of the Barberis model as assumming

$$\mu_t = a + b\, x_t.$$

But with, $\mu_t = E(r_{t+1} \mid I_t)$, it does not seem reasonable to assume x_t can capture all of I_t.

PS describe their model as "living with imperfect predictors".

It seem more likely that the predictors are imperfect, in that they are correlated with μ_t*, but cannot deliver it perfectly.*

PS also want to include x as part of the system.

They add x_t to the system, again using a dependent error structure to capture relationships.

Predictive Systems (Pastor and Stambaugh 2011):

$$
\begin{aligned}
r_{t+1} &= \mu_t + u_{t+1} \\
\mu_{t+1} &= \alpha + \beta\, \mu_t + w_{t+1} \\
x_{t+1} &= A + B\, x_t + v_{t+1}
\end{aligned}
$$

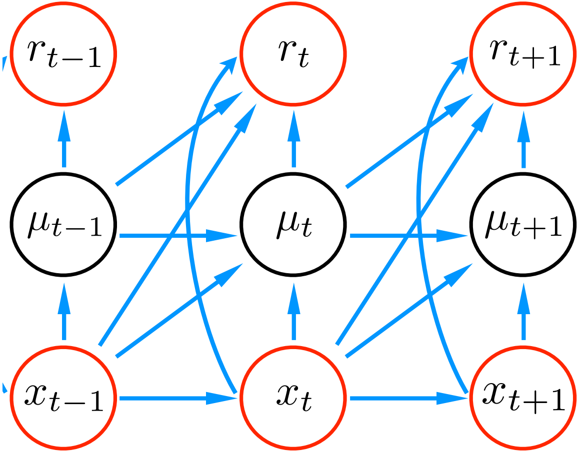
$$
\begin{pmatrix} u_{t+1} \\ w_{t+1} \\ v_{t+1} \end{pmatrix} \sim N(0, \Sigma)
$$

In PS, the dimension of x is 3.

x1: The first predictor is the market-wide dividend yield, which is equal to total dividends paid over the previous 12 months divided by the current total market capitalization.

x2: bond yield, x3: term spread.

Predictive Systems (Pastor and Stambaugh, 2011)



19

This is a complex model!!

Parameter Uncertainty and LRV:

Both Barberis and PS emphasis the importance of accounting for "parameter uncertainty".

That is, when we predict into the future, we should use the full Bayesian predictive, rather than plugging in parameter estimates.

That is,

$$LRV = \frac{Var(r_{t,t+k} \mid D_t)}{k} = \frac{Var(r_{t+1} + r_{t+2} + \ldots + r_{t+k} \mid D_t)}{k}$$

(we added D_t, data up to time t).

Which depends on the *predictive distribution*:

$$p(r_{t+1}, r_{t+2}, \ldots, r_{t+k} \mid D_t).$$

The full Bayesian joint distribution of future returns over the horizon with the parameters integrated out conditional on the observed data.

In Predictive Systems, what parameters are being integrated out?

$$
\begin{aligned}
r_{t+1} &= \mu_t + u_{t+1} \\
\mu_{t+1} &= \alpha + \beta\,\mu_t + w_{t+1} \\
x_{t+1} &= A + B\,x_t + v_{t+1}
\end{aligned}
\qquad
\begin{pmatrix} u_{t+1} \\ w_{t+1} \\ v_{t+1} \end{pmatrix} \sim N(0, \Sigma)
$$

Let T be sample size (the time you are looking forward from).

We have to integrate out:

$$(\{\mu_t\}_{t=1}^{T+k}, \alpha, \beta, A, B, \Sigma)$$

Parameter uncertainty in the simple iid model:

Suppose $r_t \sim N(\mu, \sigma^2)$ iid with only μ unknown.

$$\begin{aligned} Var(r_{t,t+k} \mid D_t) &= Var((\mu + \epsilon_1) + (\mu + \epsilon_2) + \ldots + (\mu + \epsilon_k) \\ &= Var(k\,\mu + (\epsilon_1 + \epsilon_2 + \ldots + \epsilon_k)) \\ &= k^2\, Var(\mu \mid D_t) + k\,\sigma^2 \end{aligned}$$

Your uncertainty about μ hits each return!

$\Rightarrow$ LRV increases linearly with k !!

Parameter uncertainty could really matter !!

PS argue that ignoring parameter uncertainty and imperfect predictors, fails to account for important sources of risk.

After fully accounting for these, they find increasing LRV !!!!

How do we integrate out?

Do MCMC on

$$(\{\mu_t\}_{t=1}^{T+k}, \alpha, \beta, A, B, \Sigma)$$

with a somewhat non-standard FFBS on the latent $\{\mu_t\}$.

In some models we can analytically compute part of LRV, but in all cases we can do it by simulation.

Our Goals

Stay within basic PS Predictive System framework, but:

- Elaborate model to account for multivariate stochastic volatility (Σ_t instead of just Σ).

  Need this for analysis of quarterly data.

- Using annual data, explore the sensitivity of results to prior information.

  In particular, we employ a prior on Σ that allows us to simply specify and vary the prior.

We actually started with the first goal, and the fun part there is we have to put a prior of $\{\rho_t\}$ rather than just ρ.

But, in learning for the first goal, we stumbled into the second and got results that suprised us!

Our Prior on Σ:

$\Sigma \Leftrightarrow (\sigma, \theta, \phi)$.

$$
\begin{aligned}
u_t &= \sigma_1 Z_{t1} \\
w_t &= \theta u_t + \sigma_2 Z_{t2} \\
v_{t1} &= \phi_1 u_t + \phi_2 w_t + \sigma_3 Z_{t3} \\
v_{t2} &= \phi_3 u_t + \phi_4 w_t + \phi_5 v_{t1} + \sigma_3 Z_{t4} \\
v_{t3} &= \phi_6 u_t + \phi_7 w_t + \phi_8 v_{t1} + \phi_9 v_{t2} + \sigma_3 Z_{t5}
\end{aligned}
$$

Z's are standard normal.

Inverted chi-squared on each σ_i^2 and normal on θ and each ϕ_i.

Priors on σ_1, σ_2, and θ control prior on ρ.

PS used a very "non-informative" version of the inverted Wishart.

The Results

Start with IID Model:

$$r_t \sim N(\mu, \sigma^2) \quad \text{iid}$$

With μ and σ unknown.

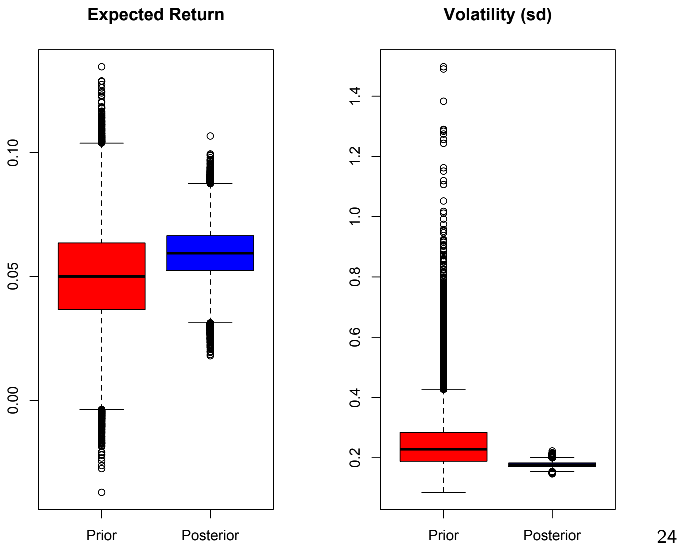
"Reasonably flat" priors on μ and σ.

Very important:

When we say "reasonably flat", we don't mean $p(\mu) \propto 1$!!

Can't get declining LRV, but let's see what happens.

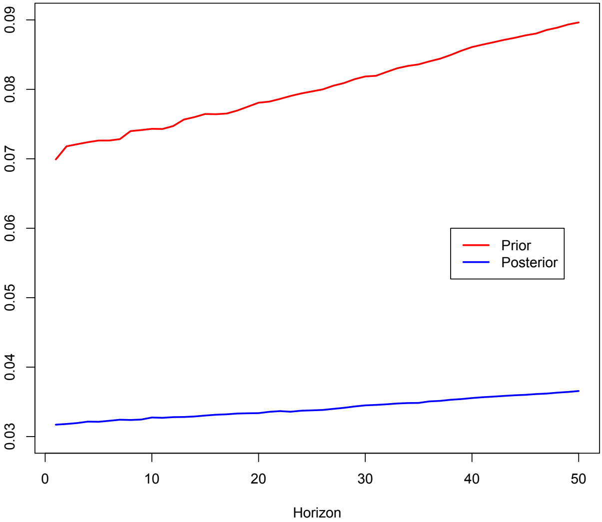
IID model



Left: μ, Right: σ.

Learn a lot about σ, not so much about μ.

IID model - Predictive Variance (per year)



Red: is prior, or, more precisely, *prior predictive*, (no $\pm$!!)[25].
Data very informative, but LRV still goes up (of course).
Recall, PS starts at about .03 and increases to .055. !!!

Simple Predictive System, "Flat Priors"

Now we use the simple predictive system with no x's.

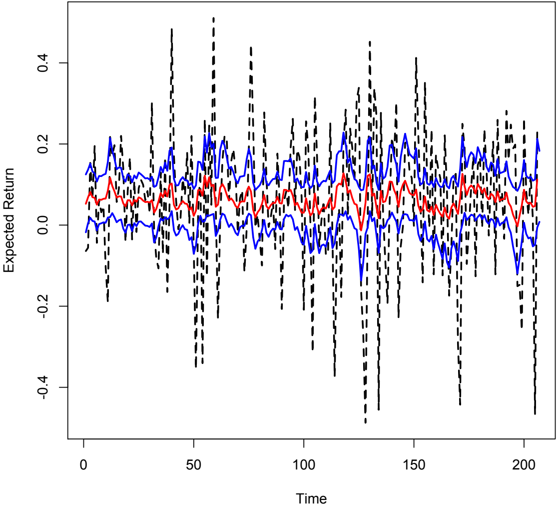
$$
\begin{aligned}
r_{t+1} &= \mu_t + u_{t+1} \\
\mu_{t+1} &= \alpha + \beta\,\mu_t + w_{t+1}
\end{aligned}
$$

$$
\begin{pmatrix} u_{t+1} \\ v_{t+1} \end{pmatrix} \sim N(0, \Sigma)
$$

Prior: "Reasonably Flat".

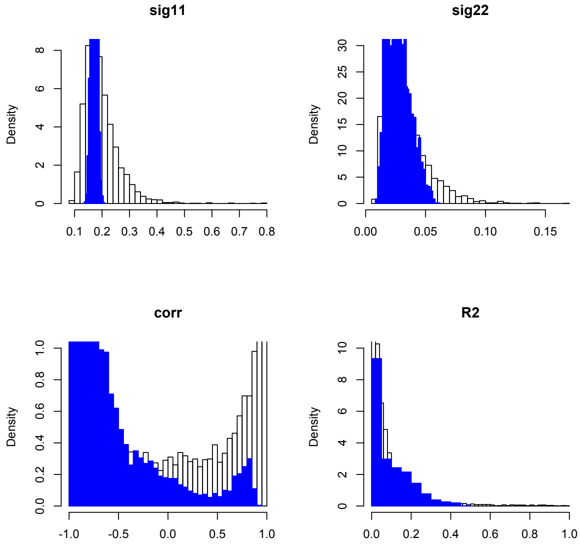
LRV could go up or down!!
Could be a lot of parameter uncertainty!!

Dynamic Model – "Flat Priors"



30

Red: posterior mean of μ_t. Blue: 95% pointwise intervals for μ_t.
Black: data



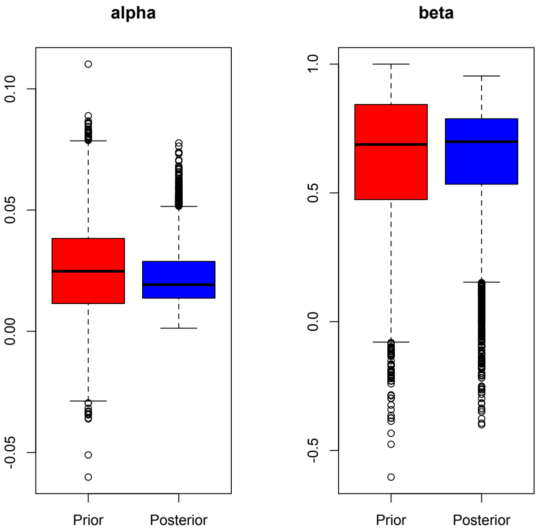
31

Blue: posterior, Open: prior

Data is quite informative for both σ_{11} *and* ρ.

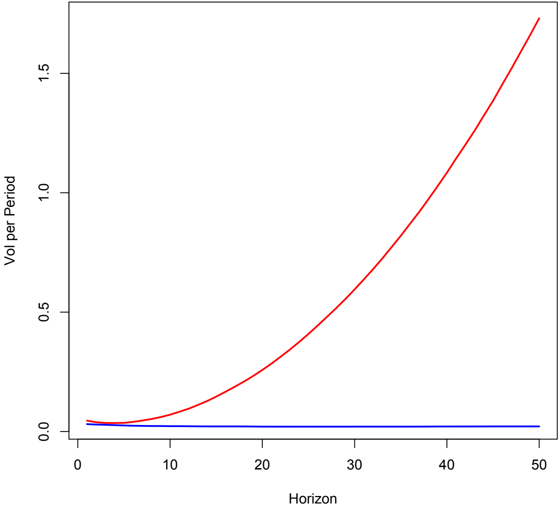
R^2: stationary $Var(\mu_t)/Var(r_{t+1})$, measure of fit.

Dynamic Model – "Flat Priors"



32

Dynamic Model – "Flat Priors"

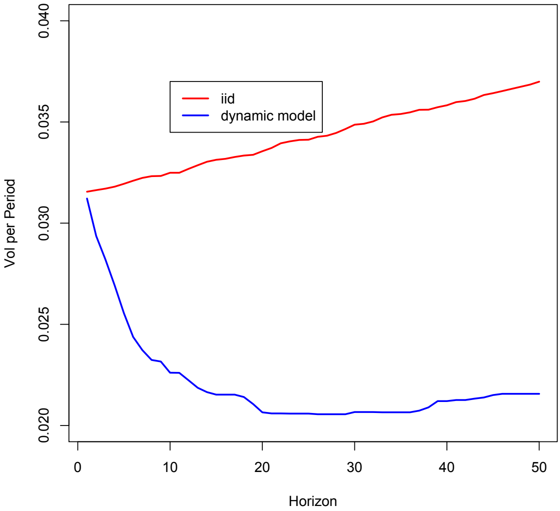


33

Red: prior, Blue:posterior

?Suggests prior is too flat?

Dynamic Model – "Flat Priors"



34

Red: iid posterior, Blue: flat prior dynamic posterior.
It goes down for 25 years !!!!!

Simple Predictive System DLM, Flat Prior:

- Enough info in data to overcome parameter uncertainty and find dynamic structure giving declining LRV.
  In particular we find $\rho < 0$.

- This is a "worst case", we had a low signal to noise ratio and did not use a lot of prior or predictive x's.

- 30 years out we are at .022 (14%) while PS have .042 (20%).

This is our main result:
in the PS framework, we get the Barberis result !!

Simple Predictive System DLM, More Informative Priors

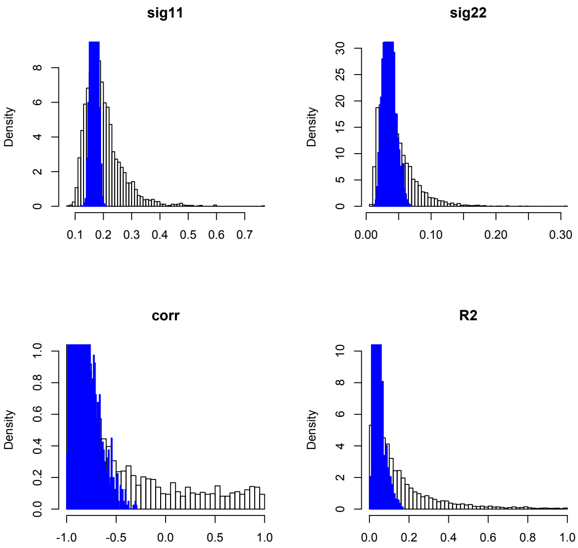
Now we inject more prior information.

In particular, we tell the model $\rho < 0$.

We'll look at results from a

"Prior 1: Somewhat Informative Prior" and a
"Prior 2: More Informative Prior".

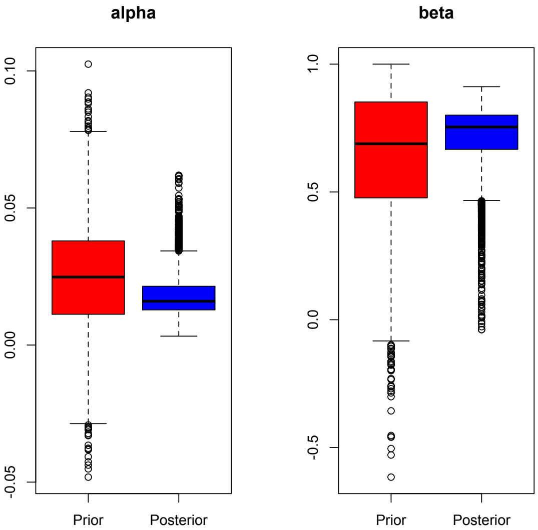
DLM – "Somewhat Informative Prior'



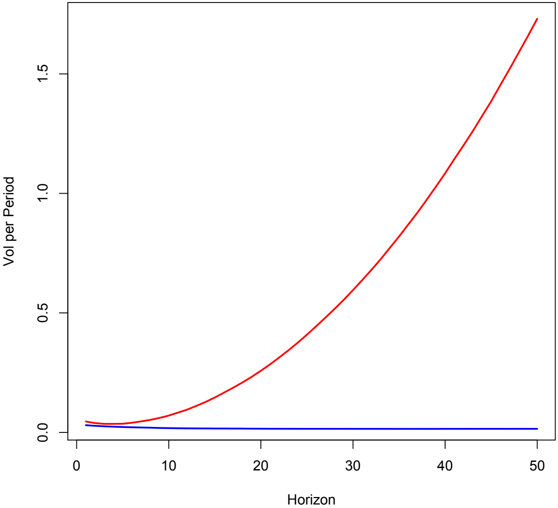
38

This time we suggest the ρ is not positive.

DLM – "Somewhat Informative Prior"



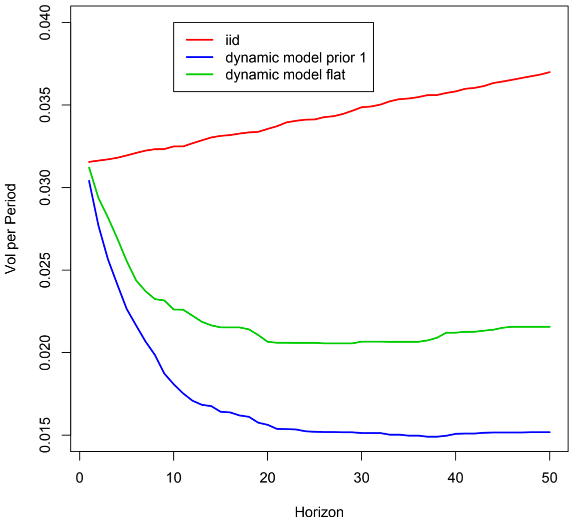
DLM – “Somewhat Informative Prior”



40

Red: prior, Blue: posterior.

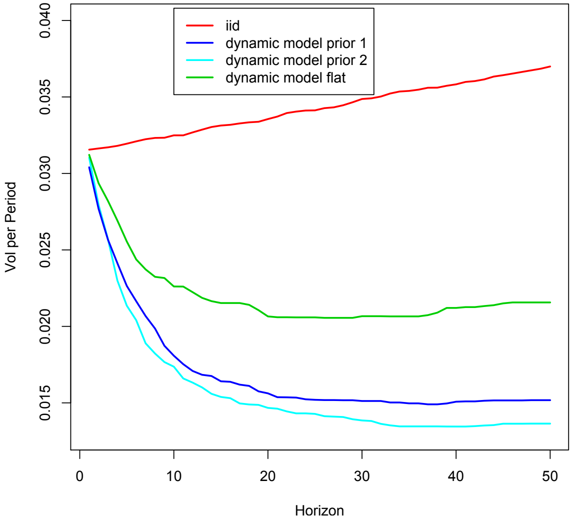
DLM – "Somewhat Informative Prior'



41

Results sensitive to prior choice.
If we put in $\rho < 0$, get stronger decrease in LRV.

DLM – "More Informative Prior'



42

Even stronger prior on $\rho < 0$ does not make much difference.

Comments:

With “prior1 - Somewhat Informative” and “prior2 - More Informative” the priors are more informative in sequence... both in ρ and for the parameters.

The main goal was to achieve something reasonable for both the unconditional mean and variance of the return.

We actually think the implied priors in “flat” are just not reasonable yet, we still get a result that contradicts PS.

The main message here is that with reasonable priors and $\rho < 0$ *the PS result is not achievable in this worst case scenario for the “predictive system” framework.*

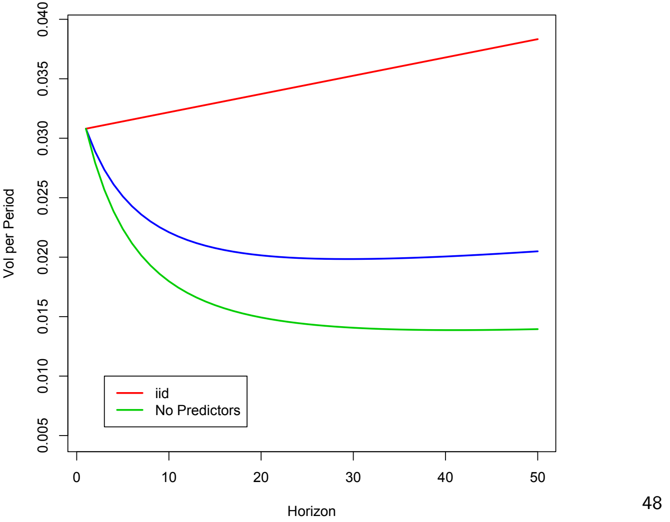
Predictive System with x

Ok, let's add the x's into the system and see what that does.

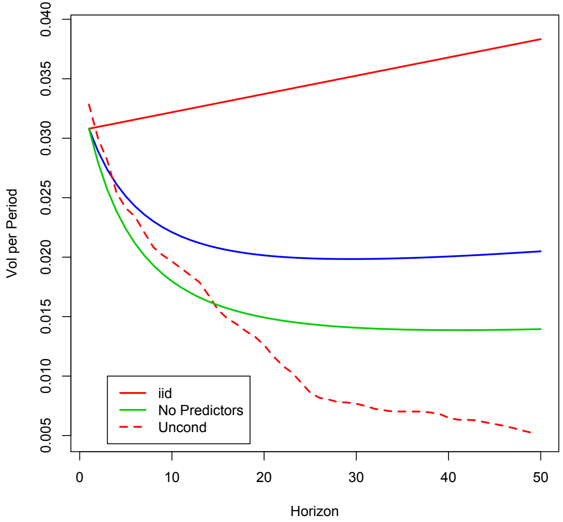
This gives us results which are most comparable to PS.

$$
\begin{aligned}
r_{t+1} &= \mu_t + u_{t+1} \\
\mu_{t+1} &= \alpha + \beta\,\mu_t + w_{t+1} \\
x_{t+1} &= A + B\,x_t + v_{t+1}
\end{aligned}
\qquad
\begin{pmatrix} u_{t+1} \\ w_{t+1} \\ v_{t+1} \end{pmatrix} \sim N(0, \Sigma)
$$

Predictability via "Predictive Systems"



Blue: with predictors. Different, but still declines out to 20 years.



49

Ucond: rolling window we did at the beginning).
Dynamic model and rolling window similar out to 20 years!!

How can the x's hurt?

Regardless, these results are still are in agreement with conventional wisdom!

No matter how hard we try, we can't get increasing LRV !!!

Time Variation

Instead of just Σ, we want Σ_t , and we want to easily incorporate the prior belief that $\rho_t = cor(u_t, w_t) < 0$ for all t and possibly other prior beliefs as well.

For the annual data we have been looking at, this does not seem crucial.

But we have quarterly data and maybe we can get more info out of that, but, we need to have time-varying variance.

We start with the Choleski Stochastic Volatility approach of Lopes, McCulloch, and Tsay.

With one x we have:

$$
\begin{array}{rcll}
u_t & = & \exp(\theta_{t1}/2)\, Z_{t1} & p(w_t) \\
w_t & = & \theta_{t3}\, u_t + \exp(\theta_{t2}/2)\, Z_{t2} & p(u_t \mid w_t) \\
v_t & = & \phi_{t2}\, u_t + \phi_{t3}\, w_t + \exp(\phi_{t1}/2)\, Z_{t3} & p(v_t \mid w_t, u_t)
\end{array}
$$

At each t, the three θ's and three ϕ's are one to one with Σ_t.

With more x's the ϕ's go on (more rows).

Let's just focus on the θ's because they determine ρ_t.

We have,

$$\begin{aligned} u_t &= \exp(\theta_{t1}/2)\, Z_{t1} \\ w_t &= \theta_{t3}\, u_t + \exp(\theta_{t2}/2)\, Z_{t2} \end{aligned}$$

$$\rho_t = \rho(\theta_{t1}, \theta_{t2}, \theta_{t3}) = \frac{\theta_{t3}}{(\theta_{t3}^2 + \exp(\theta_{t2} - \theta_{t1}))^{1/2}}.$$

We want to put a prior on the θ_t so that we get ρ_t we like.

The usual prior for the θ_{ti} series is

$$\theta_{ti} = a_i + b_i\,\theta_{t-1,i} + s_i\,z_{ti}$$

Let's call this $q(\theta_{ti}\,|\,\theta_{t-1,i})$.

Letting $\theta_t = (\theta_{t1}, \theta_{t2}, \theta_{t3})$, let,

$$q(\theta_t\,|\,\theta_{t-1}) = \Pi_{i=1}^{3} q(\theta_{ti}\,|\,\theta_{t-1,i}).$$

We usually choose the s_i so that
successive θ are not "too different".

Our prior formulation is

$$p(\theta_t \,|\theta_{t-1}) \propto q(\theta_t \,|\, \theta_{t-1}) \; f(\theta_t).$$

To get our ρ_t prior, we use,

$$f(\theta_t) = \exp(\frac{-(\rho(\theta_t) - \bar{\rho})^2}{\kappa})$$

q:
usual smoothness, don't let θ's jump around to much

f:
have preference for each θ_t, small κ means each θ_t should be such that $\rho_t \approx \bar{\rho}$

Bivariate Stochastic Volatility with Flexible Prior

$$(w_t, u_t)' \sim N(0, \Sigma(\theta_t)), \;\; \theta_t = (\theta_{t1}, \theta_{t2}, \theta_{t3})$$

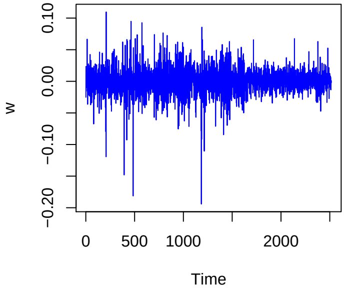
$$\begin{aligned} u_t &= \exp(\theta_{t1}/2)\, Z_{t1} \\ w_t &= \theta_{t3}\, u_t + \exp(\theta_{t2}/2)\, Z_{t2} \end{aligned}$$

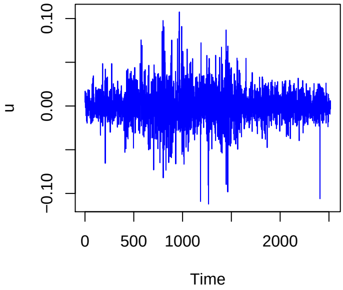
$$\begin{aligned} p(\theta_t \,|\theta_{t-1}) &\propto q(\theta_t \,|\, \theta_{t-1})\; f(\theta_t) \\ &= q(\theta_t \,|\, \theta_{t-1})\; f(\theta_t)\; K(\theta_{t-1}) \end{aligned}$$

$$p(\theta_0) \propto f(\theta_0)\; \Pi_{i=1}^{3} p(\theta_{i0})$$

Simple Example:

Let w and u be the observed bivariate series consisting of daily returns from two stocks in the S&P100.





Prior:

$$f(\theta_t) = \exp(\frac{-(\rho(\theta_t) - \bar{\rho})^2}{\kappa})$$

For this data, it is more reasonable to believe that $\rho_t > 0$!

I'll hide the details about q and show results for

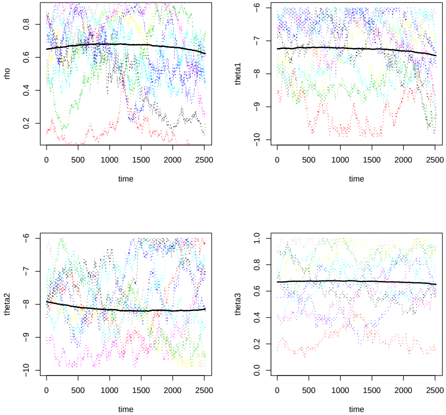
$$\bar{\rho} = .8, \;\; \kappa = .01, \; .25$$

$\kappa = .01$: tight prior.

$\kappa = .25$: loose prior.

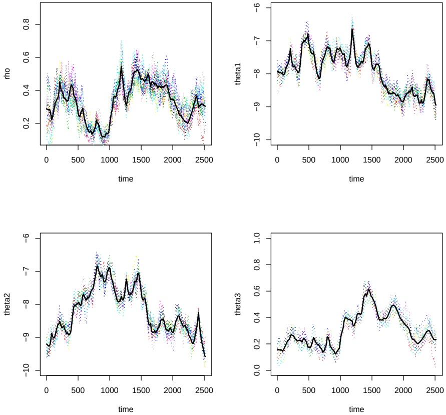
loose prior: draws from prior

black is average draw, others are individual draws
(1,1): ρ_t, (1,2): θ_{t1}
(2,1): θ_{t2}, (2,2): θ_{t3}



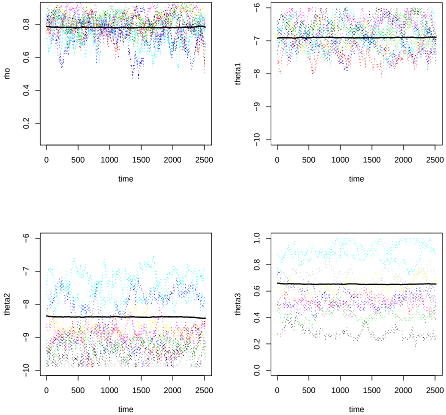
loose prior: draws from posterior

black is average draw, others are individual draws
(1,1): ρ_t, (1,2): θ_{t1}
(2,1): θ_{t2}, (2,2): θ_{t3}



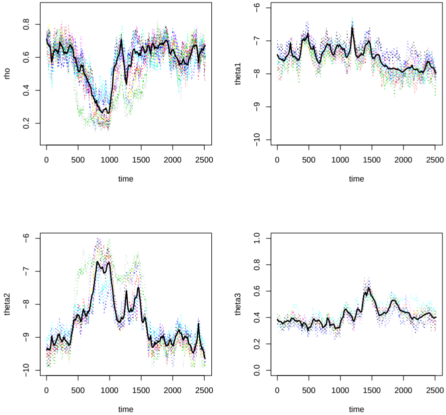
tight prior: draws from prior

black is average draw, others are individual draws
(1,1): ρ_t, (1,2): θ_{t1}
(2,1): θ_{t2}, (2,2): θ_{t3}



tight prior: draws from posterior

black is average draw, others are individual draws
(1,1): ρ_t, (1,2): θ_{t1}
(2,1): θ_{t2}, (2,2): θ_{t3}



Now we take the time varying Σ to the LRV problem:

- quarterly data.
- time varying Σ, with prior on ρ_t.
- simplify predictive system model.

Proposal

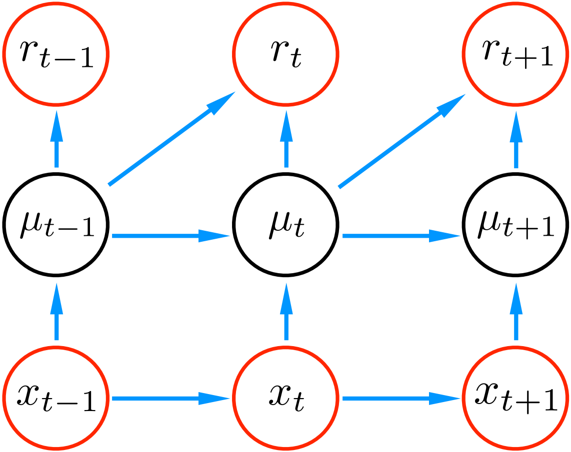
$$
\begin{aligned}
r_{t+1} &= \mu_t + u_{t+1} \\
\mu_{t+1} &= \alpha + \beta\mu_t + x'_{t+1}\gamma + w_{t+1} \\
x_{t+1} &= A + Bx_t + v_{t+1}
\end{aligned}
$$

where

$$
\begin{pmatrix} u_{t+1} \\ w_{t+1} \\ v_{t+1} \end{pmatrix} \sim N\left(0, \begin{bmatrix} \Sigma_{t+1} & 0 \\ 0 & \Sigma^X_{t+1} \end{bmatrix}\right)
$$

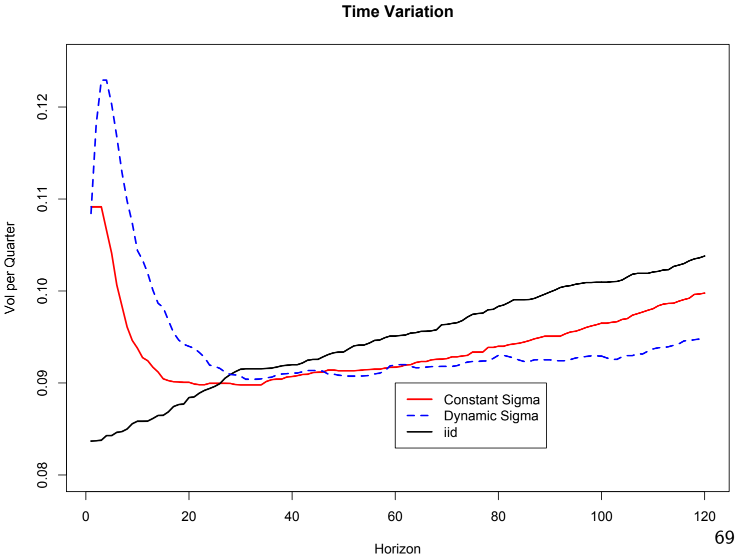
66

Proposal



67

Preliminary Results – Time Varying Covariance Matrix



69

Closing Comments:

- With reasonable priors (or maybe unreasonable) and 200 years of data stocks look very attractive for long horizon portfolios.
- This result appears robust to the added complexity of time varying relationships between predictors and expected returns and stochastic volatility.
- Lots to do still... In particular, we are still working on understanding how to better use the predictors in helping access the long-run volatility.