Problem Set 4 Answers

The point of this probem is to think through agents that have more information than we do, and the point and nature of state-space models.

1a)

$$\begin{aligned}(1-\phi L)r_{t+1} &= (1-\phi L)\,x_t + (1-\phi L)\,\varepsilon^r_{t+1} = \varepsilon^x_t + \varepsilon^r_{t+1} - \phi\varepsilon^r_t \\ (1-\phi L)r_{t+1} &= v^r_{t+1} - \theta v^r_t\end{aligned}$$

$$\begin{aligned}\sigma^2\left(\varepsilon^x_t + \varepsilon^r_{t+1} - \phi\varepsilon^r_t\right) &\Rightarrow \sigma^2_x + \left(1+\phi^2\right)\sigma^2_r - 2\phi\sigma_{xr} = (1+\theta^2)\sigma^2_v \\ E\left(\varepsilon^x_t + \varepsilon^r_{t+1} - \phi\varepsilon^r_t\right)\left(\varepsilon^x_{t-1} + \varepsilon^r_t - \phi\varepsilon^r_{t-1}\right) &\Rightarrow \sigma_{xr} - \phi\sigma^2_r = -\theta\sigma^2_v \Leftarrow E\left(v^r_{t+1} - \theta v^r_t\right)\left(v^r_t - \theta v^r_{t-1}\right)\end{aligned}$$

$$\frac{\phi\sigma^2_r - \sigma_{xr}}{\sigma^2_x + \left(1+\phi^2\right)\sigma^2_r - 2\phi\sigma_{xr}} = \frac{\theta}{1+\theta^2}$$

$$\frac{\phi - \frac{\sigma_{xr}}{\sigma^2_r}}{\left(1+\phi^2\right) + \frac{\sigma^2_x - 2\phi\sigma_{xr}}{\sigma^2_r}} = \frac{\theta}{1+\theta^2}$$

b)

$$(1-\phi L)r_{t+1} = v^r_{t+1} - \theta v^r_t$$

$$\begin{aligned}(1-\phi L)r_{t+1} &= (1-\theta L)v^r_{t+1} \\ \frac{(1-\phi L)}{(1-\theta L)}r_{t+1} &= v^r_{t+1} \\ \left\{1 + \frac{(1-\phi L)-(1-\theta L)}{(1-\theta L)}\right\}r_{t+1} &= v^r_{t+1} \\ \left\{1 + \frac{(\theta-\phi)L}{(1-\theta L)}\right\}r_{t+1} &= v^r_{t+1} \\ r_{t+1} &= \frac{\phi-\theta}{(1-\theta L)}r_t + v^r_{t+1} \\ r_{t+1} &= (\phi-\theta)\sum_{j=0}^{\infty}\theta^j r_{t-j} + v^r_{t+1}\end{aligned}$$

c)

$$\begin{aligned}\hat{x}_t &= (\phi-\theta)\sum_{j=0}^{\infty}\theta^j r_{t-j} \\ \hat{x}_t &= \frac{(\phi-\theta)}{1-\theta L}r_t \\ \hat{x}_t &= \frac{(\phi-\theta)}{1-\theta L}\frac{1-\theta L}{1-\phi L}v^r_t \\ \hat{x}_t &= (\phi-\theta)\frac{1}{1-\phi L}v^r_t \\ \hat{x}_{t+1} &= \phi\hat{x}_t + (\phi-\theta)\,v^r_{t+1}\end{aligned}$$

So the state-space version of the Wold representation is

$$\begin{aligned}\hat{x}_{t+1} &= \phi\hat{x}_t + (\phi-\theta)\, v^r_{t+1} \\ r_{t+1} &= \hat{x}_t + v^r_{t+1}\end{aligned}$$

d)

$$\begin{aligned}\hat{x}_{t+1} &= \phi\hat{x}_t + (\phi-\theta)\, v^r_{t+1} \\ r_{t+1} &= \hat{x}_t + v^r_{t+1}\end{aligned} \tag{6}$$

$$\begin{aligned}(1-\phi L)r_{t+1} &= (\phi-\theta)\, v^r_t + v^r_{t+1} - \phi v^r_t \\ (1-\phi L)r_{t+1} &= v^r_{t+1} - \theta v^r_t\end{aligned}$$

Using the formula for matching autocorrelations, and denoting $\tilde{\theta}$ the MA coefficient that comes out of the process,

$$\begin{aligned}\sigma^2_{v_{\hat{x}}} &= (\phi-\theta)^2\,\sigma^2_{v^r} \\ \sigma_{v_{\hat{x}}v_r} &= (\phi-\theta)\,\sigma^2_{v_r}\end{aligned}$$

$$\frac{\phi - \frac{\sigma_{xr}}{\sigma^2_r}}{\left(1+\phi^2\right) + \frac{\sigma^2_{\hat{x}} - 2\phi\sigma_{xr}}{\sigma^2_r}} = \frac{\tilde{\theta}}{1+\tilde{\theta}^2}$$

$$\frac{\phi - (\phi-\theta)}{\left(1+\phi^2\right) + (\phi-\theta)^2 - 2\phi\,(\phi-\theta)} = \frac{\tilde{\theta}}{1+\tilde{\theta}^2}$$

$$\frac{\theta}{1+\phi^2+\phi^2-2\theta\phi+\theta^2-2\phi^2+2\theta\phi} = \frac{\tilde{\theta}}{1+\tilde{\theta}^2}$$

$$\frac{\theta}{1+\theta^2} = \frac{\tilde{\theta}}{1+\tilde{\theta}^2}$$

d)

$$\begin{aligned}E\left(x_{t+1}|I_{t+1}\right) &= E\left(x_{t+1}|I_t\right) + \left\{E\left(x_{t+1}|I_{t+1}\right) - E\left(x_{t+1}|I_t\right)\right\} \\ E\left(x_{t+1}|I_{t+1}\right) &= E\left(\phi x_t + \varepsilon^x_{t+1}|I_t\right) + \left\{E\left(x_{t+1}|I_{t+1}\right) - E\left(x_{t+1}|I_t\right)\right\} \\ E\left(x_{t+1}|I_{t+1}\right) &= \phi E\left(x_t|I_t\right) + \left\{E\left(x_{t+1}|I_{t+1}\right) - E\left(x_{t+1}|I_t\right)\right\} \\ \hat{x}_{t+1} &= \phi\hat{x}_t + k\times v^r_{t+1}\end{aligned}$$

The last equality follows because any innovation from I_t to I_{t+1} can only be a function of v^r_{t+1}. Note that the error term is not just $E\left(\varepsilon^x_{t+1}|I_{t+1}\right)$. Since future r can tell us about past expectations,

$$E\left(x_{t+1}|I_{t+1}\right) - E\left(x_{t+1}|I_t\right) = \phi\left[E\left(x_t|I_{t+1}\right) - E\left(x_t|I_t\right)\right] + E\left(\varepsilon^x_{t+1}|I_{t+1}\right)$$

By definition,

$$r_{t+1} = E(r_{t+1}|I_t) + v^r_{t+1} = \hat{x}_t + v^r_{t+1}.$$

OK, so now we have

$$\begin{aligned}\hat{x}_{t+1} &= \phi\hat{x}_t + k\times v^r_{t+1} \\ r_{t+1} &= \hat{x}_t + v^r_{t+1}.\end{aligned}$$

We just need to pick k so that the Wold representation of r has an MA root θ, which you've already done.

2.

$$\begin{aligned}\mu_t &= \phi\mu_{t-1} + \sigma_\mu v_t\\ g_t &= \theta g_{t-1} + \sigma_g v_t.\end{aligned}$$

a)

$$\begin{aligned}dp_t &= \sum \rho^{j-1} E_t r_{t+j} - E_t \sum \rho^{j-1} E_t \Delta d_{t+j}\\ &= \sum \rho^{j-1} E_t \mu_{t+j-1} - E_t g_{t+j-1}\\ dp_t &= \frac{\mu_t}{1-\rho\phi} - \frac{g_t}{1-\rho\theta}\end{aligned}$$

b)

$$dp_t = \left(\frac{\sigma_\mu}{1-\rho\phi}\frac{1}{1-\phi L} - \frac{\sigma_g}{1-\rho\theta}\frac{1}{1-\theta L}\right) v_t$$

$$(1-\theta L)(1-\phi L)\, dp_t = \left(\frac{\sigma_\mu}{1-\rho\phi}(1-\theta L) - \frac{\sigma_g}{1-\rho\theta}(1-\phi L)\right) v_t$$

We have to express this as an ARMA(2,1) and find the root to make sure it's less than one

$$\begin{aligned}(1-\theta L)(1-\phi L)\, dp_t &= \left(\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right) - \left(\frac{\sigma_\mu\theta}{1-\rho\phi} - \frac{\sigma_g\phi}{1-\rho\theta}\right) L\right) v_t\\ (1-\theta L)(1-\phi L)\, dp_t &= \left(1 - \frac{\frac{\sigma_\mu}{1-\rho\phi}\theta - \frac{\sigma_g}{1-\rho\theta}\phi}{\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}} L\right)\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right) v_t\end{aligned}$$

An example of what I think is "reasonable,"

$$\frac{\frac{\sigma_\mu}{1-\rho\phi}\theta - \frac{\sigma_g}{1-\rho\theta}\phi}{\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}} = \frac{\frac{.2}{1-0.96\times 0.94}0.4 - \frac{.1}{1-0.96\times 0.4}0.94}{\frac{.2}{1-0.96\times 0.94} - \frac{.1}{1-0.96\times 0.4}} = 0.35$$

Note though that because of the - sign, ξ is not between θ and ϕ. A bit more generally:

$$\begin{aligned}\frac{\frac{\sigma_\mu}{1-\rho\phi}\theta - \frac{\sigma_g}{1-\rho\theta}\phi}{\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}} &< 1?\\ \frac{\sigma_\mu}{1-\rho\phi}\theta - \frac{\sigma_g}{1-\rho\theta}\phi &< \frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\\ \frac{\sigma_\mu}{1-\rho\phi}(\theta-1) - \frac{\sigma_g}{1-\rho\theta}(\phi-1) &< 0\\ \frac{\sigma_\mu}{1-\rho\phi}(\theta-1) &< \frac{\sigma_g}{1-\rho\theta}(\phi-1)\\ \frac{\frac{\sigma_\mu}{1-\rho\phi}}{\frac{\sigma_g}{1-\rho\theta}} &> \frac{(\phi-1)}{(\theta-1)}\\ \frac{\frac{\sigma_\mu}{1-\rho\phi}}{\frac{\sigma_g}{1-\rho\theta}} &> \frac{1-\phi}{1-\theta}\end{aligned}$$

The left hand side is the contribution of expected returns to dp volatility, relative to the contribution of expected dividend growth to that volatility. If $\theta < \phi$ the right side is even less than one. So this is a quite mild restriction, it says we need “enough” expected return contribution to dividend yield volatility. Perturbing our $\sigma_g = 0$ view, most numbers will satisfy it.

c)

$$
\begin{aligned}
(1-\theta L)(1-\phi L)\, dp_t &= (1-\xi L)\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right) v_t \\
(1-\theta L)\, dp_t &= (1-\xi L)\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right) \frac{1}{\sigma_\mu}\frac{\sigma_\mu}{1-\phi L} v_t \\
(1-\theta L)\, dp_t &= (1-\xi L)\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right) \frac{1}{\sigma_\mu}\mu_t \\
\frac{\sigma_\mu}{\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right)}\frac{(1-\theta L)}{(1-\xi L)} dp_t &= \mu_t
\end{aligned}
$$

$$
\begin{aligned}
r_{t+1} &= \frac{\sigma_\mu}{\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right)}\frac{(1-\theta L)}{(1-\xi L)} dp_t + \varepsilon^r_{t+1} \\
r_{t+1} &= (1-\rho\phi)\frac{\frac{\sigma_\mu}{1-\rho\phi}}{\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right)}\frac{(1-\theta L)}{(1-\xi L)} dp_t + \varepsilon^r_{t+1}
\end{aligned}
$$

Here you see the old coefficient, $(1-\rho\phi)$. If $\sigma_g = 0$, that's it (and $\xi = \theta$ in that case too). If $\sigma_\mu = 0$ on the other hand, it's all zero. if $\sigma_\mu = \sigma_g$ and $\phi = \theta$, then dp never moves (expected returns and dividend growth offset), so naturally the coefficient explodes

$$
\begin{aligned}
(1-\theta L)(1-\phi L)\, dp_t &= (1-\xi L)\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right) v_t \\
(1-\phi L)\, dp_t &= (1-\xi L)\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right) \frac{1}{\sigma_g}\frac{\sigma_g}{1-\theta L} v_t \\
(1-\phi L)\, dp_t &= (1-\xi L)\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right) \frac{1}{\sigma_g} g_t \\
\frac{\sigma_g}{\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right)}\frac{(1-\phi L)}{(1-\xi L)} dp_t &= g_t
\end{aligned}
$$

$$
\begin{aligned}
\Delta d_{t+1} &= \frac{\sigma_g}{\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right)}\frac{(1-\phi L)}{(1-\xi L)} dp_t + \varepsilon^d_{t+1} \\
\Delta d_{t+1} &= (1-\rho\theta)\frac{\frac{\sigma_g}{1-\rho\theta}}{\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right)}\frac{(1-\phi L)}{(1-\xi L)} dp_t + \varepsilon^d_{t+1}
\end{aligned}
$$

Conversely, here you see the coefficient in the “dividend growth” world $(1-\rho\theta)$, times a term which is -one in that world ($\sigma_\mu = 0$). if $\sigma_g = 0$, we're back to 0 as promised in the return world.

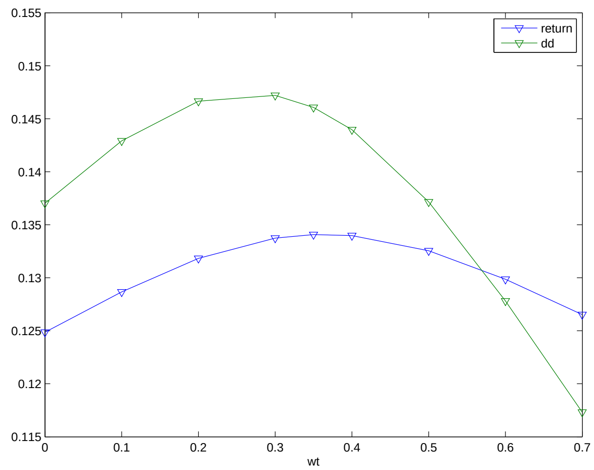
In sum, our VAR should look like this.

$$
\begin{aligned}
r_{t+1} &= (1-\rho\phi)\frac{\frac{\sigma_\mu}{1-\rho\phi}}{\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right)}\frac{(1-\theta L)}{(1-\xi L)} dp_t + \varepsilon^r_{t+1} \\
\Delta d_{t+1} &= (1-\rho\theta)\frac{\frac{\sigma_g}{1-\rho\theta}}{\left(\frac{\sigma_\mu}{1-\rho\phi} - \frac{\sigma_g}{1-\rho\theta}\right)}\frac{(1-\phi L)}{(1-\xi L)} dp_t + \varepsilon^d_{t+1} \\
(1-\theta L)(1-\phi L)\, dp_t &= (1-\xi L)\,\varepsilon^{dp}_{t+1}
\end{aligned}
$$

d) The return and dividend growth coefficients have the pattern

$$\begin{aligned} r &: \frac{(1-\theta L)}{(1-\xi L)} = 1 + \frac{(\xi - \theta) L}{1 - \xi L} \\ d &: \frac{(1-\phi L)}{(1-\xi L)} = 1 + \frac{(\xi - \phi) L}{1 - \xi L} \end{aligned}$$

I searched a bit and settled on $\xi = 0.3$ as maximizing R^2. (To to this for real, you minimize sum of mse across equations.) Here's a graph of the R^2 as a function of the weight ξ :



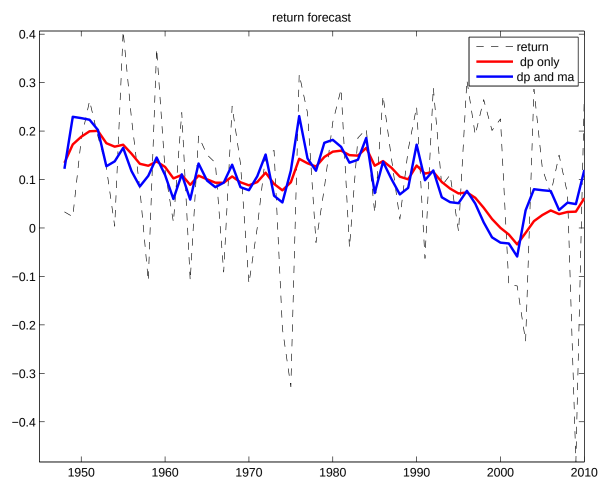
Next, my results are

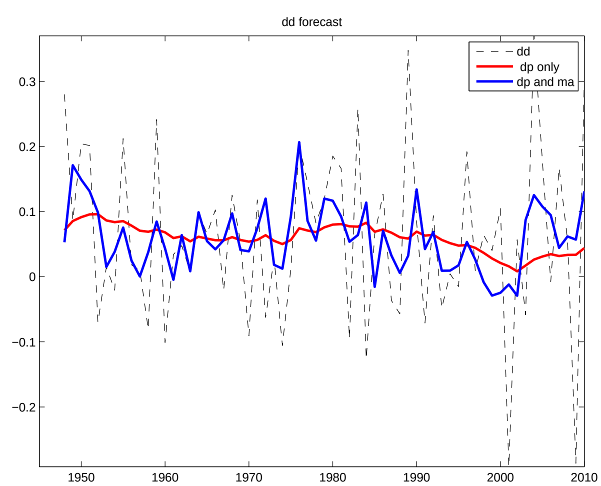
```
rho = mean D/P / (1+mean ( D/P))
    0.9682

start date (first lhv) 19471231 end date 20091231
forecasts using one lag of dp
         b         t        R2
r     0.126     2.563    0.101
dd    0.047     1.024    0.024
dp    0.950    23.697    0.902
forecasts using one lag of dp and   5 year MA with weight 0.30
         b         t       bma       tma        R2
r     0.109     2.073    0.220     1.706     0.134
dd    0.021     0.467    0.329     3.253     0.147
dp    0.941    25.154    0.116     1.145     0.904
```

The MA term has a small increase in the return R^2, but a big impact on dividend growth. "Recent changes" in dp ratio do seem to help to forecast dividend growth. They forecast returns in the same direction, though they also forecast dp a little bit In terms of the identity $c_r - c_d \approx -\rho c_{dp}$ we have $0.184 - 0.289 = -0.105 \approx -0.96 \times 0.112 = -0.10$. So, the reason it doesn't help r as much as we thought is that it also helps *dp* a bit. Still, it has to show up *somewhere*, we can't just take the 1.3 and 1.0 t statistics and conclude *both* return and dp coefficients are zero. So, if you swallow the dividend coefficient, you have to swallow one of the others.

This is essentially the Koijen-Van Binsbergen result, though their paper uses a ML estimation and includes a moving average of past dividend growth and returns. (That's the same thing given the identity of course.)

Here are the plots.. The new variable is clearly matching a lot of the "wiggles" in dividend growth, and corresponding wiggles in returns. Note how the wiggles line up really well; by forecasting dividend growth you almost mechanically forecast returns (actually if you forecast dividend growth without forecasting dp, yes, you mechanically forecast returns! Darn identities again!) However, ex post dividend growth is less volatile than returns, which is why the improvement in R2 is more dramatic for dividends, and why the t stat is better for dividends even though the increase in variance of expected component is about the same. Once again, beware making decisions based on t stats!





To do: What does the impulse response function look like? Since the dividend forecast is very short term, I don't think it's much affected, but I haven't worked it out.

Now, why am I not in the end ecstatic about all this? It seems we have learned so much by the state-space model! We were able to infer what seemed like hidden state variables. Alas, not really. Suppose you had just started with the final VAR, as an empirical finding and you asked "what time series process for expected returns and dividend growth follow here? Start with

$$
\begin{aligned}
r_{t+1} &= (1-\rho\phi)\frac{\frac{\sigma_\mu}{1-\rho\phi}}{\left(\frac{\sigma_\mu}{1-\rho\phi}-\frac{\sigma_g}{1-\rho\theta}\right)}\frac{(1-\theta L)}{(1-\xi L)}dp_t+\varepsilon^r_{t+1} \\
\Delta d_{t+1} &= (1-\rho\theta)\frac{\frac{\sigma_g}{1-\rho\phi}}{\left(\frac{\sigma_\mu}{1-\rho\phi}-\frac{\sigma_g}{1-\rho\theta}\right)}\frac{(1-\phi L)}{(1-\xi L)}dp_t+\varepsilon^d_{t+1} \\
(1-\theta L)(1-\phi L)\,dp_t &= (1-\xi L)\,\varepsilon^{dp}_{t+1}
\end{aligned}
$$

Then

$$
\begin{aligned}
E_t\left(r_{t+1}\right) &= (1-\rho\phi)\frac{\frac{\sigma_\mu}{1-\rho\phi}}{\left(\frac{\sigma_\mu}{1-\rho\phi}-\frac{\sigma_g}{1-\rho\theta}\right)}\frac{(1-\theta L)}{(1-\xi L)}dp_t \\
&= (1-\rho\phi)\frac{\frac{\sigma_\mu}{1-\rho\phi}}{\left(\frac{\sigma_\mu}{1-\rho\phi}-\frac{\sigma_g}{1-\rho\theta}\right)}\frac{(1-\theta L)}{(1-\xi L)}\frac{(1-\xi L)}{(1-\theta L)\,1-\phi L)}\varepsilon^{dp}_{t+1t} \\
&= (1-\rho\phi)\frac{\frac{\sigma_\mu}{1-\rho\phi}}{\left(\frac{\sigma_\mu}{1-\rho\phi}-\frac{\sigma_g}{1-\rho\theta}\right)}\frac{1}{(1-\phi L)}\varepsilon^{dp}_{t+1t}
\end{aligned}
$$

"Oh," you exclaim "The VAR implies that expected returns follow an AR(1) process with parameter ϕ, driven by the dividend yield shock!" The derivation works both ways, *the "state space model" is nothing more than an implication of the Wold representation.* If we had an *economic* reason to impose an AR(1) on expected returns, then there would be some reason to impose the resultant "smoothness" on the VAR.

The best I can say is, if we want to summarize the lessons of a VAR it might be interesting to summarize them in terms of simple models for expected returns and dividend growth rather than simple ARMA structures on the VAR coefficients. But there is nothing more than the Wold representation here.

To put the argument another way, it appears as if we are learning a lot about agent's information sets g and μ. Alas, we are not, and the correlation of shocks assumption was not at all innocuous. In reality $E_t(r_{t+1})$ incorporates a lot of information from other variables (e.g. cay) that we do not see at all, and this information *is not revealed* by the $\{dp, \Delta d, r\}$ dataset, no matter what we do. For getting closer to that true picture of the world, for understanding a bit more what $E_t(r_{t+1}|\text{agent information})$ really looks like, I still think that adding *other variables* is more important than mining the lag structure of the VAR, even as cleverly as I've done here.

e)The MA forecast for comparison. Then the forecast with just the difference. As you see the R2 is almost the same. Now, how do you compare two obviously very correlated variables? It's easy to make one or the other absorb the common movement. I chose to display "what if you assign most of the common signal to the first difference" by running the regression

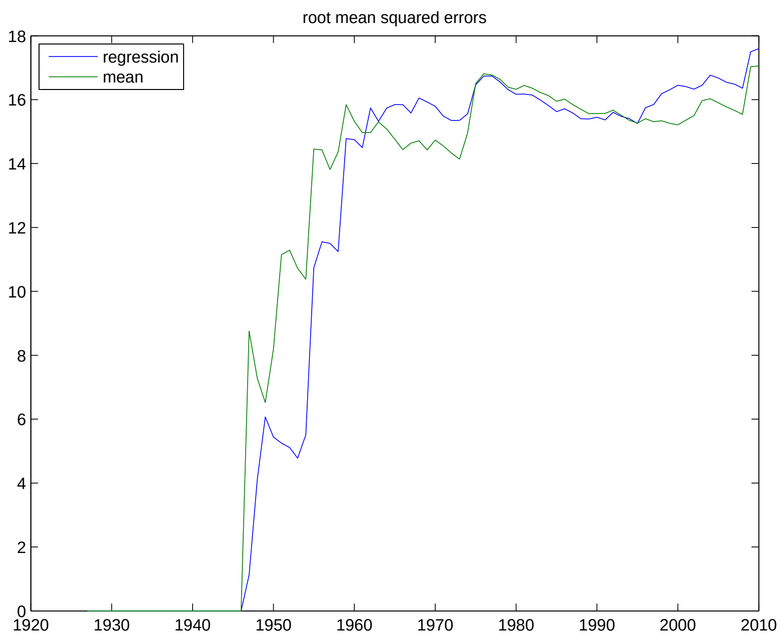
$$
r_{t+1} = a + b dp_t + c(dp_t - dp_{t-1}) + d(dp_{t-1} - \sum \xi^j dp_{t-1}) + \varepsilon^r_{t+1}
$$

As you see, it says the MA really doesn't add that much.

```
forecasts using one lag of dp and   5 year MA with weight 0.30
          b        t      bma      tma       R2
```

```
r      0.109     2.073     0.220     1.706     0.134
dd     0.021     0.467     0.329     3.253     0.147
dp     0.941    25.154     0.116     1.145     0.904
forecasts using one lag of dp and dp difference
           b         t        bD        tD        R2
r      0.116     2.231     0.198     1.378     0.125
dd     0.030     0.646     0.332     3.231     0.137
dp     0.943    25.187     0.141     1.240     0.904
forecasts using one lag of dp, dp difference, and   5 year MA with weight 0.30
           b         t       b,D       t,D       bma       tma        R2
r      0.104     1.965     0.195     1.405     0.463     1.110     0.138
dd     0.022     0.458     0.330     3.400     0.321     0.883     0.147
dp     0.946    24.829     0.142     1.243     0.321     0.883     0.905
```

3. Here is my plot



As you can see, the regression does a bit better for a while, then the mean does better through the 70s; the regression catches up, but the mean comes back in the 90s boom. At the last point, I find a RMSE of

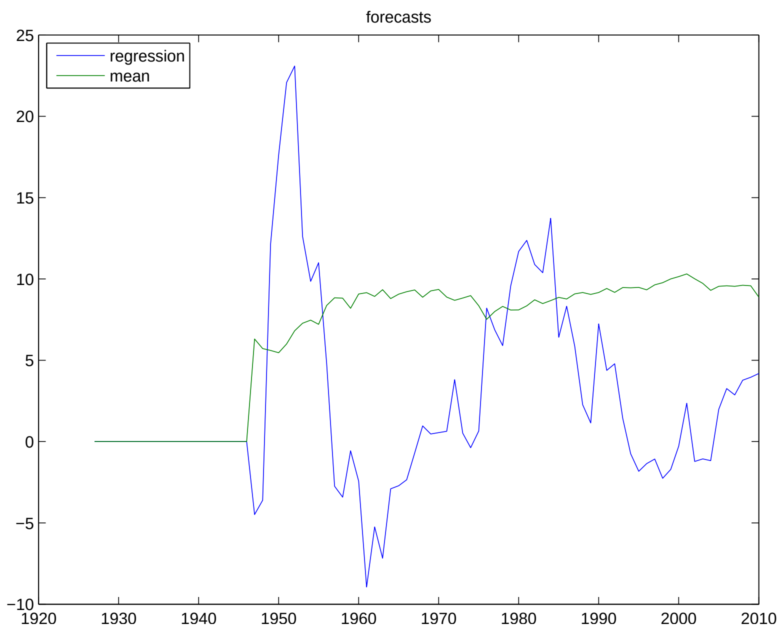
```
17.5986   17.0556
```

for the regression and sample mean respectively. Just using the sample mean would have done better "out of sample!"'

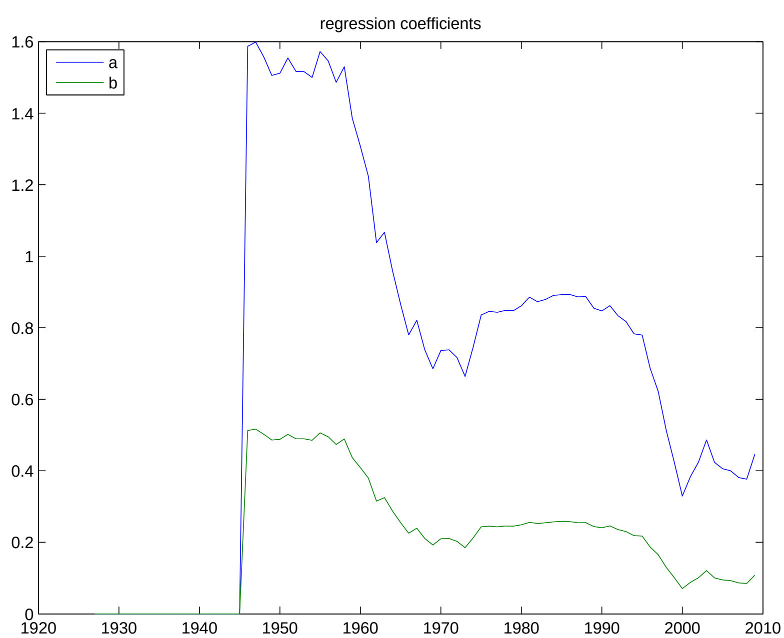
Goyal and Welch think this means "returns aren't predictable." My counterargument is that we should see this all the time even if returns really are predictable. To see why, note that at best were promising a 7% R^2. Using $\sigma(r) = 16\%$, that means we should see a rmse of $\sqrt{(1-0.07)} \times 0.16 = 15.4\%$

for the regression and 16% for the mean even with infinite data! A bit of bad luck in one sample can easily overturn that.

It's more than bad luck – the monte carlo in the dog that didn't bark shows you *expect* to do worse even if returns *really are* predictable. The trouble is that 70 years really is a "short" sample in this business, so we don't really know the regression coefficients. To show this, here are the forecasts and the fitted regression coefficients in this sample,



You can see that the mean is pretty steady, while the dp forecast goes all over the place. Part of that is that the DP varies a lot – this is good – but part is that the regression coefficients wander around a lot. Variation in a is as much of the problem as variation in b. We don't really know what "mean" the dividend yield will revert to.



That's the real problem. Sure returns are predictable, but is the coefficient .05, .1 or .4? You really need to know that to make forecasts!

Thus, my conclusion is not that this exercise shows "returns really are not predictable." It shows that "uncertainties in the regression coefficients make it pointless to *exploit* the predictability in market-timing portfolios." "Bayesian portfolio theory" adds model uncertainty to optimal portfolio calculations and verifies this hunch.