Predictive Regressions: What's Stationarity Got to Do With It

Robert Erbe Gregor Reich

University of Zurich

Structural retreat, Bonn

September 13, 2019

Are stock returns predictable: simply run forecasting regression

$$
\begin{aligned}
& r_{t+1} = a_r + b_r x_t + \epsilon^r_{t+1} \\
\rightarrow \quad & E_t(r_{t+1}) = a_r + b_r x_t
\end{aligned}
$$

- Investment opportunity vs. efficient markets
- The *right* forecasting variable: most prominently price-dividend ratio, term-structure spreads, consumption-to-wealth ratio

Statistical features of usual forecasting variables x_t

- Very persistent, slow-moving: often modeled as AR(1) process

$$x_{t+1} = a_x + \phi x_t + \epsilon^x_{t+1}$$

- Good reasons to confine x_t to be *stationary*, i.e. $|\phi| < 1$
  Lewellen (2004), Cochrane (2007)
- *BUT* ϕ subject to *finite-sample bias* that can inflate b_r

The model we look at: vector autoregression
Stambaugh (1986, 1999), Mankiw and Shapiro (1986)

$$\begin{pmatrix} r_{t+1} \\ x_{t+1} \end{pmatrix} = \begin{pmatrix} a_r \\ a_x \end{pmatrix} + \begin{pmatrix} 0 & b_r \\ 0 & \phi \end{pmatrix} + \begin{pmatrix} \epsilon^r_{t+1} \\ \epsilon^x_{t+1} \end{pmatrix}$$

$$\rightarrow \quad \boldsymbol{x}_{t+1} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{x}_t + \boldsymbol{\epsilon}_{t+1}$$

Our Paper

Key take-away from literature: ϕ matters when asking about b_r

- "Full" likelihood assures stationarity of VAR, least squares not
- MLE and inference: constrained optimization
- Power gains from stationarity
- Informal argument on power loss due to finite-sample bias
- Relate to more-powerful test proposed by Cochrane (2007)
- Size comparison

Joint Distribution and Stationarity

The log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{a}, \boldsymbol{B}, \Sigma\epsilon)$

$$L(\boldsymbol{\theta}) = \log f_{\boldsymbol{\theta}}(\boldsymbol{x}_1, ..., \boldsymbol{x}_t) \quad \rightarrow \quad \hat{\boldsymbol{\theta}} = \text{argmax } L(\boldsymbol{\theta})$$

Suppose time series $\boldsymbol{x}_1, ..., \boldsymbol{x}_t$ follows first-order Markov chain

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}_1, ..., \boldsymbol{x}_t) = f_{\boldsymbol{\theta}}(\boldsymbol{x}_1) \prod_{t=2}^{T} f_{\boldsymbol{\theta}}(\boldsymbol{x}_t | \boldsymbol{x}_{t-1})$$

BUT in routine data analysis, $f_{\boldsymbol{\theta}}(\boldsymbol{x}_1)$ is taken to be fixed

- Likelihood estimation yields least squares estimates
- Well-known that estimator allows for unit roots
  Hamilton (1994), Engsted and Pedersen (2014), Wilson, Reale, and Haywood (2015)

Figure: How big is the difference

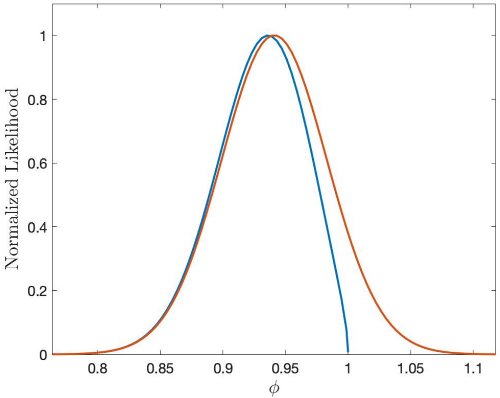

Figure: Conditional and unconditional likelihood for the data as in Cochrane (2007)

Hypothesis Testing

Likelihood-based inference: likelihood ratio

$$2\left(L(\hat{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}_0)\right) \sim \chi^2_{\dim(\boldsymbol{\theta}_0)}$$

In practice often composite hypothesis that fixes $\boldsymbol{\theta}_0^S$ but leaves $\boldsymbol{\theta}^N$ unspecified, where $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}^S, \boldsymbol{\theta}^N)$

How to deal with nuisance parameters $\boldsymbol{\theta}^N$?

$$\text{Profile Likelihood:} \quad L_P(\boldsymbol{\theta}_0^S) = \max_{\boldsymbol{\theta}^N} L(\boldsymbol{\theta}_0^S, \boldsymbol{\theta}^N)$$

$$\text{Estimated Likelihood:} \quad L_E(\boldsymbol{\theta}_0^S) = L(\boldsymbol{\theta}_0^S, \hat{\boldsymbol{\theta}}^N)$$

Profile likelihood to be treated like regular likelihood, thus

$$2\left(L(\hat{\boldsymbol{\theta}}) - L_P(\boldsymbol{\theta}_0^S)\right) \sim \chi^2_{\dim(\boldsymbol{\theta}_0^S)}$$

Maximum Likelihood Estimation

As is standard, suppose $\boldsymbol{x}_t | \boldsymbol{x}_{t-1} \sim N(\boldsymbol{a} + \boldsymbol{B}\boldsymbol{x}_{t-1}, \Sigma(\epsilon))$ and $\boldsymbol{x}_1 \sim N(\mu_0(\boldsymbol{a}, \boldsymbol{B}), \Sigma_0(\Sigma(\epsilon), \boldsymbol{B}))$

BUT normal density known to be unbounded, if $\Sigma \not\succ 0$ or $\Sigma_0 \not\succ 0$

Non-linear constrained optimization

$$
\begin{aligned}
& \max_{\boldsymbol{a},\boldsymbol{B},\boldsymbol{L}} L(\boldsymbol{\theta}) \\
s.t. \quad & \operatorname{diag}(\boldsymbol{L}) \geq 0 \\
& |\lambda(\boldsymbol{B})| \leq 1
\end{aligned}
$$

- Positive-semidefiniteness of Σ: diagonal elements of Cholesky decomposition $\Sigma = LL'$ to be $\operatorname{diag}(L) \geq 0$
- Finiteness of Σ_0: characteristic roots of $\boldsymbol{B}$ to be $|\lambda(\boldsymbol{B})| \leq 1$

Estimates

Estimation	b_r	ϕ	σ_r	σ_ϕ	$\rho_{r,\phi}$	p-value $(b_r = 0)$
Unconditional-Likelihood	0.100 (0.051)	0.937 (0.038)	0.193	0.152	-0.698	4.38%
Conditional-Likelihood	0.096 (0.053)	0.941 (0.042)	0.194	0.152	-0.700	7.50%

Table: Sample estimates for data as in Cochrane (2007): annual data, aggregate US stock market, CRSP, 1927-2004. Standard errors in parentheses

Figure: How big is the power gain?

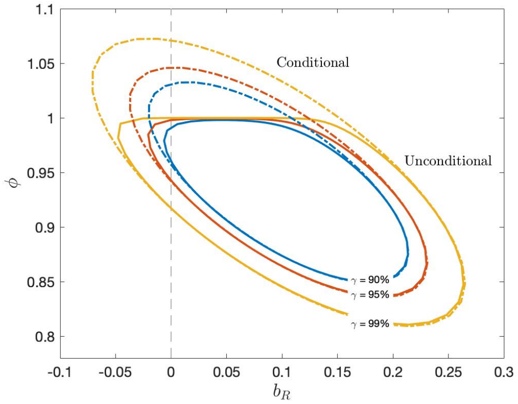

Figure: Joint confidence sets of (un-)conditional likelihood

Figure: What about the bias?

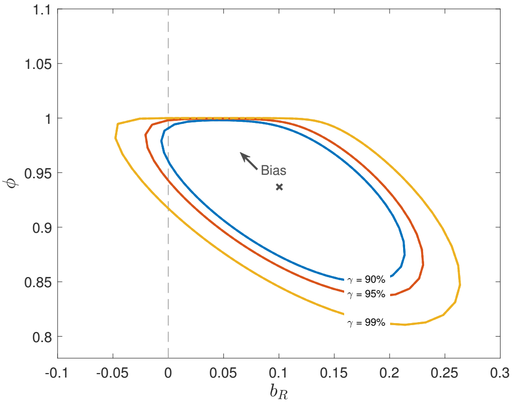

Figure: Effect of finite sample bias with respect to "center" point of joint confidence set.

Returns, Dividend Growth, and Dividend Yields

Much noticed paper by Cochrane (2007)

> *"if returns are not predictable, dividend growth must be predictable"*

- Notion of dividend-growth based test for return predictability Hjalmarsson and Kiss (2019), Leroy and Sinhania (2019)
- Present-value model: dividend growth other side of the coin
- Cochrane (2007) reports overall *p-value* of 1-2%

In a 3-variable VAR with returns r_{t+1}, dividend-growth Δd_{t+1}, and dividend yields dp_{t+1}, the regression coefficients are tied together

$$1 = \rho\phi + b_r - b_{dg}, \quad \text{with} \quad \rho < 1$$

Null Hypothesis

Given stationarity $|\phi| < 1$, b_r and b_{dg} cannot both be zero

$$1 - \rho\phi = b_r - b_{dg} \neq 0$$

Null hypothesis: $b_r = 0 \rightarrow b_{dg} = 1 - \rho\phi$

- What about the rest?
- Cochrane (2007) uses OLS and fixes ϕ to point estimate
- BUT recall argument on profile likelihood
- Simulation based tests also call for coherent $\boldsymbol{\theta}^N$
  Dufour, Khalaf, et al. (1999), Dufour (2006)

Full Picture: Dividend-growth Based Test

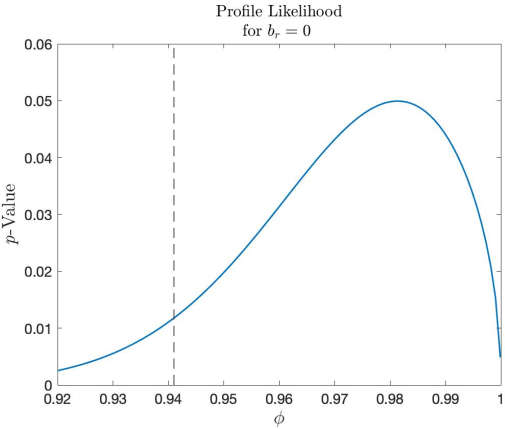

Figure: Dividend growth based test with likelihood for $b_r = 0$ and different values of ϕ, dashed line indicates $\hat{\phi} = 0.941$ as used in Cochrane (2007)

Type I Error

How often do we falsely reject the null?

Controlled environment: simulation-based size assessment

1. Simulate, say 5,000 times, under $\hat{\boldsymbol{\theta}}^N = \text{argmax } L(b_r = 0, \boldsymbol{\theta}^N)$
2. Estimate model parameters $\boldsymbol{\theta}_i$
3. Test for $b_r = 0$ and store *p-values*$_i$
    - Likelihood ratio
    - Dividend growth based test
4. Count rejections for different significance levels α

Test / α Levels	0.01	0.05	0.10
Likelihood ratio	0.033	0.106	0.177
Cochrane (2007)	0.109	0.281	0.404

Concluding remarks

- Likelihood provides natural way to preserve stationarity (both for estimation and inference)
- Computation is not prohibitive
- Our test has too much bite: bias correction might resolve this