Modern Sampling Methods

Class 10: Window Selection

January 11, 2022

Outline

- Introduction
- Simple Model with Time-Varying Parameters
- Plug-in, cross-validation, and related methods
- Analyzing Window Selection Methods

Introduction

In macro/time series applications, it's common to select a *window* of time over which to estimate model parameters or form current forecasts.

Why restrict analysis to a subset of the data?

- Possibility of structural breaks,
- or a more gradual evolution of underlying parameters.

Choice of window is often done informally – with the danger of cherry-picking/data-snooping.

Simple Model with Time-Varying Parameters

Switch to time-index notation: $t = 1, \ldots, T$.

$$Y_t = \beta_t + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0, \sigma^2),$$

The means β_t could change over time in different ways, e.g.

1. Discrete breaks:

$$\beta_t = \begin{cases} \beta^{(0)} & t < T_1 \\ \beta^{(1)} & t \geq T_1 \end{cases}$$

2. Random-walk parameters:

$$\beta_t = \beta_{t-1} + v_t, \quad v_t \overset{iid}{\sim} N(0, \sigma_v^2).$$

For discrete breaks, can estimate and test the locations of the breaks, e.g. Andrews (1993), Bai & Perron (1998), Elliott & Müller (2007).

- Naive use of Chow test (t-test for size of break) may be misleading if we searched over possible break dates.
- Andrews: search over a set of possible break dates; adjust critical values to reflect this search.

For random-walk and related time-varying parameter models, could estimate:

- Parametric MLE or other methods (can have nonstandard distributions, see e.g. Shephard (1993), Davis & Dunsmuir (1996));
- Nonparametrically using kernel or other flexible regression methods (see e.g. Robinson (1989)):

  $\hat{\beta}_t$ is a weighted average of Y_s for s close to t.

Our focus will not be on the breaks/TVP parameters themselves, but on end-of-sample estimation and forecasting:

$$\text{MSE} = E\left[(\hat{\beta}_T - \beta_T)^2\right].$$

One way is to average Y_t over a *window* $[W_1, T]$:

$$\hat{\beta}_{W_1,T} = \frac{1}{T - W_1 + 1} \sum_{t=W_1}^{T} Y_t.$$

More generally, could weight observations $w = (w_1, \ldots, w_T)$ and use

$$\sum_{t=1}^{T} w_t Y_t.$$

Let's consider the choice of W_1 under the discrete break model where

$$\beta_t = \beta^{(0)} 1(t < T_1) + \beta^{(1)} 1(t \geq T_1).$$

If we set $W_1 = T_1$, then

$$\mathrm{Bias}(\hat{\beta}_{W_1,T}) = 0.$$

$$\mathrm{Var}(\hat{\beta}_{W_1,T}) = \frac{\sigma^2}{T - T_1 + 1}.$$

(Recall MSE $=$ Var $+$ Bias^2.)

Is this optimal for MSE? Suppose we instead choose $W_1 = T_1 - 1$. Then

$$|\text{Bias}| = \frac{1}{T - T_1 + 2} \left| \beta^{(0)} - \beta^{(1)} \right|$$

So $\text{Bias}^2 \uparrow$.

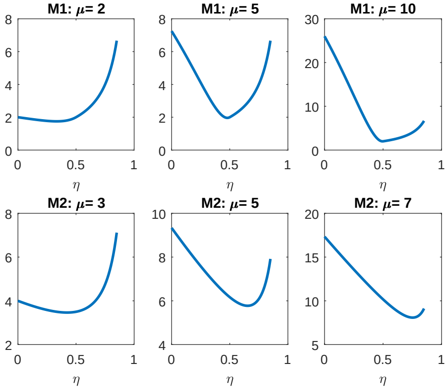
But variance may go *down*, because we are averaging over more observations.

The (infeasible) optimal window could have $W_1 < T_1$.

Optimal window depends on size of break $\beta^{(1)} - \beta^{(0)}$, σ^2, T_1, and T. (Don't want to set $W_1 > T_1$)

See Pesaran & Timmermmann (2007) for specific results and extensions to regression models.

MSE as a function of Window



M1: single break at 0.5; μ indicates size of break
M2: random walk model; μ indicates variability

Feasible Window Selection

1. **Naive Plug-in**: estimate T_1 and $(\beta^{(0)}, \beta^{(1)})$ by least squares

$$\min_{T_1, \beta^{(0)} \beta^{(1)}} \sum_{t=1}^{T} \left(Y_t - \beta^{(0)} 1(t < T_1) - \beta^{(1)} 1(t \geq T_1) \right)^2 ,$$

and set

$$\hat{W}_1 = \hat{T}_1.$$

2. **MSE Plug-in (Pesaran and Timmermann)**: using LS estimates, find $\hat{W}_1$ that minimizes implied MSE.

3. Cross-Validation (Pesaran and Timmermann):

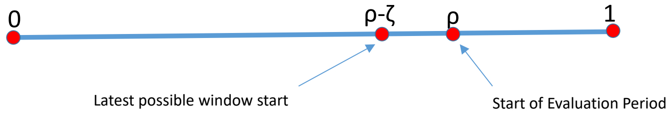
Hold-out sample: $r, r+1, \ldots, T$.

Set $W_{max} < r$.

Solve:

$$\min_{W \leq W_{max}} \sum_{t=r}^{T-1} \left(Y_{t+1} - \hat{\beta}_{W,t} \right)^2 ,$$

where $\hat{\beta}_{W,t}$ is the sample average using observations $Y_W, \ldots, Y_t$.



4. Inoue, Jin and Rossi (2017):

Hybrid method based on Robinson (1989) nonparametric estimation idea.

1. Test for constant parameter using Bai and Perron (1998).
2. Obtain preliminary cross-validated estimate of the window $\tilde{W}$.
3. Obtain preliminary local linear estimate at the end of the sample, $\tilde{\beta}_T$ (weighted regression locally around T).
4. Pick the window to solve

$$\min_{\tilde{W} \leq W \leq T} \left(\hat{\beta}_{W,T} - \tilde{\beta}_T \right)^2 .$$

5. Laplace cross-validation (Hirano & Wright):

Take cross-validation criterion function:

$$C(w) = \sum_{t=r}^{T-1} \left(Y_{t+1} - \hat{\beta}_{w,t} \right)^2 .$$

Let

$$L(W) = \exp\left(-0.5C(W)/\hat{\sigma}^2\right),$$

and treat this as a pseudo-likelihood for W.

Pseudo-posterior mean

$$\overline{W} = \frac{\sum wL(w)}{\sum L(w)}.$$

Remarks

- Cross-validation (and related methods) do not rely on a specific model for the evolution of β_t, so they are potentially more robust.
- However, the time-series nature of the exercise makes it different from CV with cross-section data: fewer out-of-sample cases, so it can be more "noisy."
- All of these methods are challenging to analyze formally with more general specifications.
- Asymptotic approximations that imply that we learn T_1 perfectly in the limit may miss a key source of risk.

Local Asymptotics for Window Selection

Hirano & Wright (2021), building on Elliott & Müller (2007):

$$Y_t = \beta_t + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} (0, \sigma^2).$$

Rescale time: $t = rT \quad r \in [0, 1]$.

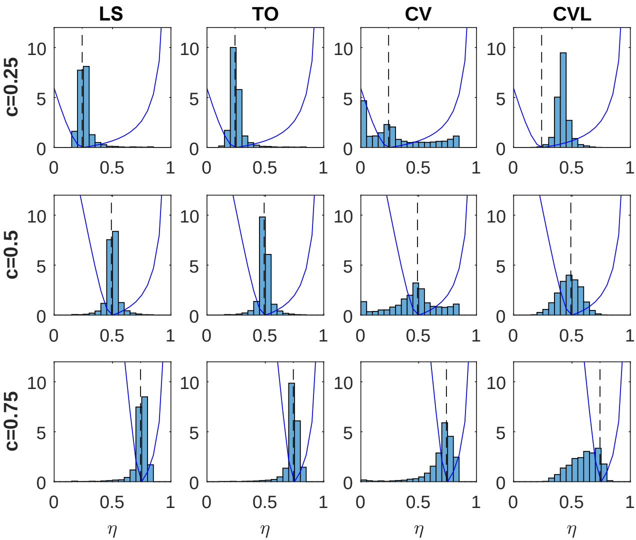
Make β_{rT} local to zero:

$$\beta_{rT} = \frac{H(r)}{\sqrt{T}},$$

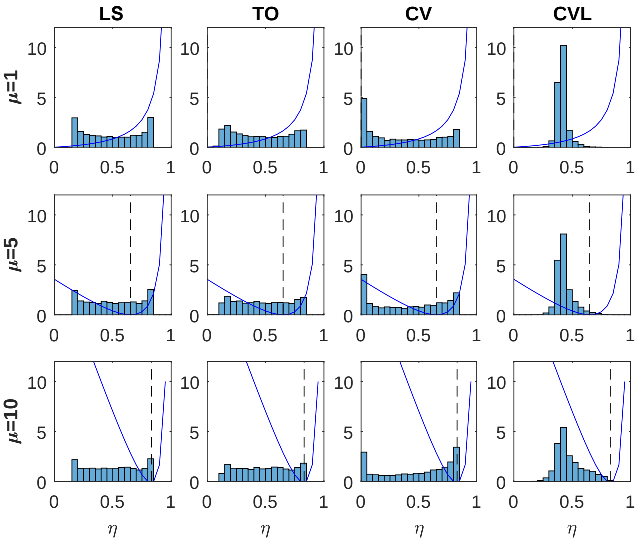
where $H(r)$ is a deterministic or stochastic process.

(M1) single break; (M2) random walk; (M3) Poisson jumps.

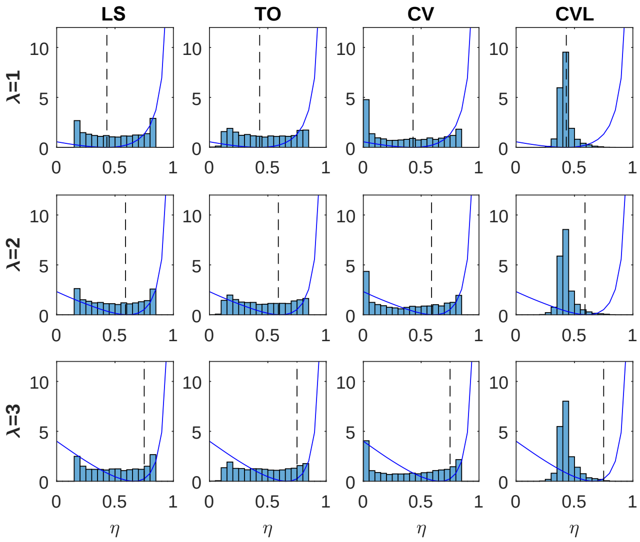
Window Selection under Model M1



Window Selection under Model M2



Window Selection under Model M3



Asymptotic RMSE

Rolling Window	ML	TO	CV	CVL
Model M1				
$c = 0.25$	1.9	1.8	5.8	1.9
$c = 0.5$	2.9	2.8	8.0	3.8
$c = 0.75$	6.8	6.5	10.3	16.9
Model M2				
$\mu = 1$	7.8	4.7	5.3	2.1
$\mu = 5$	9.4	7.7	8.5	6.3
$\mu = 10$	17.1	16.2	14.0	11.6
Model M3				
$\lambda = 1$	7.9	5.3	6.0	3.2
$\lambda = 2$	8.5	6.3	7.2	4.8
$\lambda = 5$	9.0	7.2	7.9	6.2

Application: Phillips Curve

$$\pi_{t+1}^{(k)} - \pi_t = \beta_0 + \beta_1(\pi_t - \pi_{t-1}) + \beta_2(u_t - \bar{u}_t) + \varepsilon_t$$

- Data: 1959:Q1 to 2018:Q2
- π_t: total or core PCE index
- $\pi^{(k)}$: cumulative inflation over k periods
- u_t = civilian UR, $\bar{u}_t$ = CBO NAIRU
- Assess methods using out-of-sample forecast accuracy

Inflation RMSE

	ML 1B	ML BIC	TO	CV All	CV Pre	CVL All	CVL Pre	IJR
Total PCE								
$k = 1$	1.882	1.767	1.834	1.808	1.787	1.784	1.777	1.824
$k = 2$	1.685	1.570	1.670	1.666	1.610	1.597	1.589	1.642
$k = 3$	1.679	1.490	1.650	1.590	1.586	1.475	1.489	1.513
$k = 4$	1.612	1.572	1.567	1.539	1.521	1.411	1.425	1.428
Core PCE								
$k = 1$	0.806	0.814	0.800	0.809	0.812	0.810	0.807	0.804
$k = 2$	0.706	0.728	0.710	0.718	0.705	0.714	0.707	0.706
$k = 3$	0.726	0.727	0.743	0.773	0.708	0.698	0.685	0.733
$k = 4$	0.770	0.878	0.769	0.843	0.736	0.712	0.732	0.763

Application: Large dataset

- Take 210 series from the large dataset of McCracken and Ng (2016), 1959q1-2017q4
- Forecast each by an AR(1) with rolling window
- Evaluate by out-of-sample RMSE

McCracken-Ng Data

RMSE of AR(1) Forecasts with Alternative Windows:

	LS-1	LS-BIC	TO	CV All	CV Pre	CVL All	CVL	IJR
Ave. RRMSE	1.0368	1.0003	1.026	1.0169	1.0085	0.9951	0.9989	1.0032
25th % RRMSE	1.0065	1.000	1.0041	1.0060	1.0019	0.9919	0.9964	0.9954
50th % RRMSE	1.0322	1.000	1.0189	1.0209	1.0153	1.0036	1.0030	1.0107
75th % RRMSE	1.0569	1.000	1.0415	1.0406	1.0261	1.0088	1.0070	1.0198
P(Min RRMSE)	0.0286	0.4762	0.019	0.0667	0.0476	0.1333	0.1238	0.1048
P(RRMSE=1)	0	0.681	0	0	0	0	0	0

(RRMSE: relative to AR(1) estimated on whole sample)
(P(RRMSE=1) corresponds to no-break estimate)