Lecture 4: Unit root, Cointegration, and Forecasting

Yu Bai

City University of Macau

Unit root and Cointegration

Testing for unit roots

- The simplest approach to testing for a unit root begins with an AR(1) model:

$$y_t = d_t + \rho y_{t-1} + e_t, \quad t = 1, 2, \cdots, \tag{1}$$

  where y_0 is the observed initial value.

- We assume that $(e_t)_t$ has zero mean, given the past values of y:

$$\mathbb{E}\left(e_t | y_{t-1}, y_{t-2}, \cdots, y_0\right) = 0. \tag{2}$$

- This condition is weaker than, but practically identical to $i.i.d.$ (homoscedasticity and no serial correlation).

Testing for unit roots

- The null hypothesis is that $(y_t)_t$ has a unit root:

$$\mathcal{H}_0 : \rho = 1,$$

against the one-sided alternative ($(y_t)_t$ is covariance stationary):

$$\mathcal{H}_1 : \rho < 1.$$

- It is common to let d_t unspecified when forming hypothesis.
  - There are generally three commonly used specifications for d_t: $d_t = \{\}$, $d_t = \alpha$, and $d_t = \alpha + \delta t$.
  - Sadly, this is because asymptotic distributions of test statistics depend on these nuisance parameters.
- We can, of course, alter $\mathcal{H}_0$ and $\mathcal{H}_1$:
  - See the well known "KPSS" unit root test: Kwiatkowski, Phillips, Schmidt, and Shin (1992).

Testing for unit roots

- A convenient equation for carrying out the unit root test is

$$\Delta y_t = \alpha + \theta y_{t-1} + e_t,$$

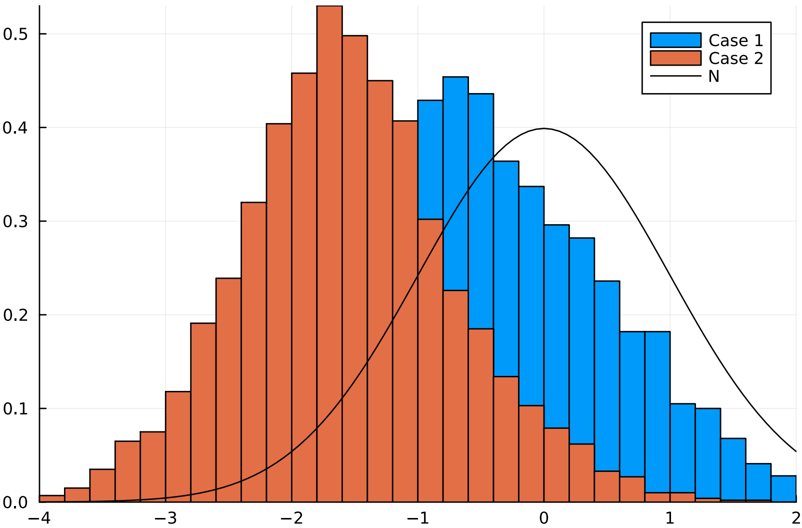
where $\theta = \rho - 1$ and ρ is given as in (1).

- We can use conventional t-test for the null hypothesis:

$$\mathcal{H}_0 : \theta = 0.$$

- However, the asymptotic distribution of such t-statistic is not $\mathcal{N}(0, 1)$ and rather nonstandard.

Testing for unit roots



Testing for unit roots

Table: Asymptotic Critical Values for Dickey-Fuller Test

	1%	2.5%	5%	10%
No time trend	-2.57	-2.22	-1.94	-1.62
Intercept	-3.43	-3.12	-2.86	-2.57
Linear time trend	-3.96	-3.66	-3.41	-3.12

Testing for unit roots

- What to do if we fail to reject the null hypothesis?
- We should only conclude that the data do not provide strong evidence against $\mathcal{H}_0$.
- You have to decide by yourself whether you use the data in levels or in first-difference in the subsequent regression analyses.

Testing for unit roots: Augmented Dickey-Fuller test

- Consider an AR(2) model for (y_t):

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\ &= (\phi_1 + \phi_2)\, y_{t-1} - \phi_2 \Delta y_{t-1} + \varepsilon_t. \end{aligned}$$

- If $(y_t)_t$ has unit root, $\phi_1 + \phi_2 = 1$.
- This suggests we can test for unit root based on the regression:

$$\Delta y_t = d_t + \theta y_{t-1} + \gamma \Delta y_{t-1} + \varepsilon_t.$$

- Can also be extended to models with higher-order lags...
- Interestingly, it can be shown that $\hat{\gamma}_n$ has standard asymptotic distribution, so conventional t-test and F-test apply to those parameters.

Testing for unit roots: other issues

- Of course, apart from adding lags, we can also use simple model, but with HAC standard errors when calculating t-statistic.
- There are many different variants of unit root tests developed in the literature.
  - See BDGH (1993, Section 4-3) for a discussion on other tests.
- We focus on the covariance-stationary alternative (left-tail). The other side (right-tail) may be of independent interests.
  - Google "PSY testing for explosive bubbles" if you are interested in.

Spurious regression

Journal of Econometrics 2 (1974) 111–120. © North-Holland Publishing Company

SPURIOUS REGRESSIONS IN ECONOMETRICS

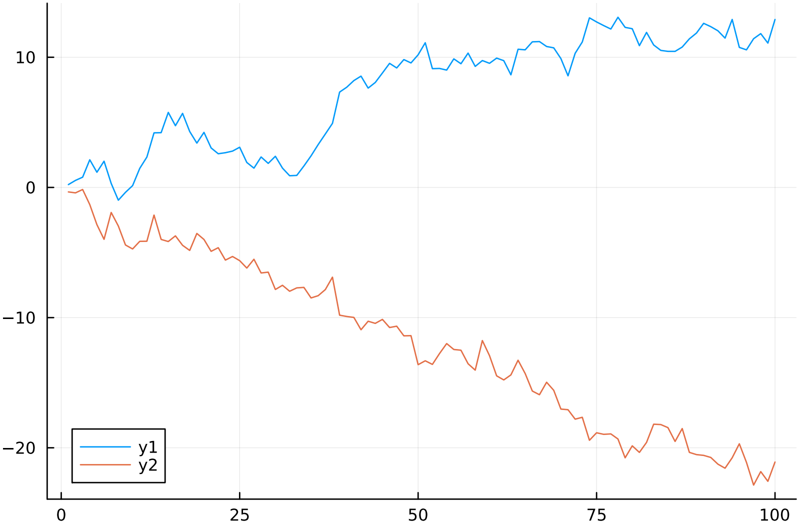
C.W.J. GRANGER and P. NEWBOLD

University of Nottingham, Nottingham NG7 2RD, England

Received May 1973, revised version received December 1973

- See for several (funny) examples at this website.

Spurious regression



Spurious regression

- Suppose we generate two independent random walks:

$$x_t = x_{t-1} + v_t, \quad y_t = y_{t-1} + u_t, \ \ t = 1, 2, \cdots,$$

  where $(v_t)_t$ and $(u_t)_t$ are independent, identically distributed innovations, with mean zero and variances σ_v^2 and σ_u^2, respectively.

- We then run the following regression:

$$y_t = \alpha + \beta x_t + e_t,$$

  and test the null hypothesis $\mathcal{H}_0 : \beta = 0$ by the usual t-statistic:

$$t_{\beta=0} = \frac{\hat{\beta}_n}{\text{se}\left(\hat{\beta}_n\right)}.$$

- What would we expect?

Spurious regression

- Let us look at the simulation results reported in Ganger and Newbold (1974, JoE):

Table 1
Regressing two independent random walks.

S:	0–1	1–2	2–3	3–4	4–5	5–6	6–7	7–8
Frequency:	13	10	11	13	18	8	8	5
S:	8–9	9–10	10–11	11–12	12–13	13–14	14–15	15–16
Frequency:	3	3	1	5	0	1	0	1

- If we use the $\mathcal{N}(0,1)$ critical values, the rejection frequency at the 5% level, is roughly 75% !!!
- Some recent replication results by myself:

```
count_pval_smaller = 6705

count_pval_greater = 3295
```

Spurious regression

- Why it is the case? Let us reconsider the regression model:

$$y_t = \alpha + \beta x_t + e_t.$$

- Under the null hypotheses: $\mathcal{H}_0 : \alpha = \beta = 0$, we thus have $y_t = e_t$.
- In other words, the error $(e_t)_t$ is a random walk process under the null, which clearly violates even the asymptotic version of the Gauss-Markov assumptions from Chapter 11.

Spurious regression: remedies

- Just add lags:

$$y_t = \alpha + \rho y_{t-1} + \beta x_t + \delta x_{t-1} + e_t.$$

- There exist values for the coefficients, specifically $\rho = 1$ and $\alpha = \beta = \delta = 0$, for which $(e_t)_t$ is $I(0)$.
- It can be shown that $\sqrt{n}\left(\hat{\beta}_n - \beta\right)$ converges to a Normal distribution and t-statistic for $\mathcal{H}_0 : \beta = 0$ is asymptotically $\mathcal{N}(0,1)$.

Cointegration

- The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2003 was divided equally between Robert F. Engle III "for methods of analyzing economic time series with time-varying volatility (ARCH)" and Clive W.J. Granger "for methods of analyzing economic time series with common trends (cointegration)"

Cointegration

Definition

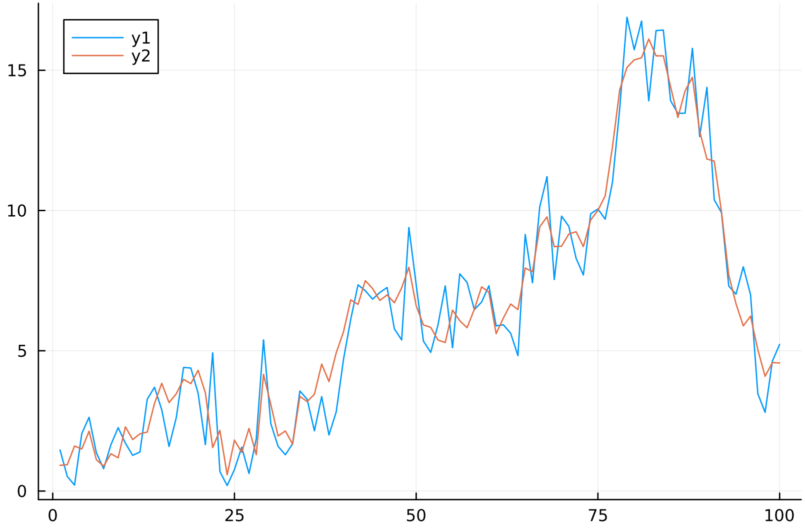
Suppose that both $(x_t)_t$ and $(y_t)_t$ are I(1) processes. We say x and y are *cointegrated* if there exists $\beta \neq 0$, such that $y_t - x_t\beta$ is I(0).

- We call $(1, -\beta)'$ the cointegrating vector.

Going further questions

Let $\{(y_t, x_t) : t = 1, 2, \cdots\}$ be a bivariate time series where each series is I(1) without drift. Explain why, if y_t and x_t are cointegrated, y_t and x_{t-1} are also cointegrated.

Cointegration



Example

Let

- $r6_t$: annualized interest rate for six-month T-bills
- $r3_t$: annualized interest rate for three-month T-bills
- $spr_t = r6_t - r3_t$: term spread

If $r6$ and $r3$ were not cointegrated, the difference between interest rates could become very large, with no tendency for them to come back together. Based on a simple arbitrage argument, this seems unlikely. Therefore, large deviations between $r6$ and $r3$ are not expected to continue: the spread has a tendency to return to its mean value.

Testing for cointegration

- If β is known, very simple, just test whether $u_t = y_t - x_t\beta$ is covariance stationary.
- If β is unknown, the critical values for unit root tests have to be retabulated to account for estimation of β.
- The following is taken from Davidson and MacKinnon (1993, Table 20.2):

Table: Asymptotic Critical Values for Cointegration Test

	1%	2.5%	5%	10%
Intercept	-3.90	-3.59	-3.24	-3.04
Linear time trend	-4.32	-4.03	-3.78	-3.50

- Of course, "business as usual" if we want to add trend, deal with serial correlation, etc.

Estimation of cointegrating vector

- Cointegration only states that $(u_t)_t$ in the model

$$y_t = \beta x_t + u_t,$$

  has to be stationary. It has noting to do with exogeneity of x_t.
- Suppose that $x_t = x_{t-1} + v_t$ with $x_0 = 0$. Then, $x_t = \sum_{s=1}^{t} v_s$.
- Are we willing to assume that

$$\text{cov}\left(x_t, u_t\right) = \text{cov}\left(\sum_{s=1}^{t} v_s, u_t\right) = 0?$$

Example: testing for expectation hypothesis (EH)

$$hy6_t = \beta_0 + \beta_1 hy3_{t-1} + u_t,$$
$$hy3_t = hy3_{t-1} + v_t,$$

where

- $hy6_t$: three-month holding yield (in percent) from buying a six-month T-bill at time $(t-1)$ and selling it at time t (three months hence) as a three-month T-bill
- $hy3_{t-1}$: three month holding yield from buying a three-month T-bill at time $(t-1)$

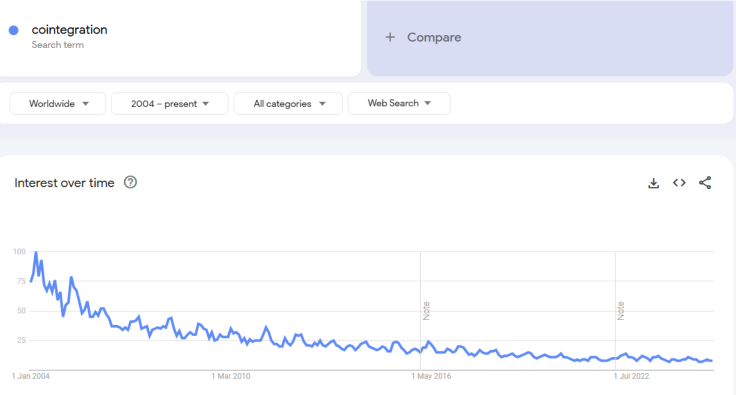
Question

Is it plausible to have $\text{cov}\,(hy3_{t-1}, u_t) = 0$?

Estimation of cointegrating vector

- If $\neq 0$, do we still have consistency of $\hat{\beta}_n$?
- Fortunately, the answer is YES. Thanks to the fact that $(x_t)_t$ is very persistent, we do not need to worry about the usual "Simultaneous Equation Bias" in the cross-sectional or even weakly stationary case.
- However, the asymptotic distribution of $\hat{\beta}_n$ is no longer standard, making usual t-test and F-test asymptotically invalid.
- There are methods developed to obtain valid "$\mathcal{N}$"-type of inference in this context. (*requires advanced treatment so goes beyond the scope of the course...*)

The decline of cointegration...



Error Correction Models

- In general, we can specify an ARDL(1,1) model for $(y_t)_t$ and $(x_t)_t$ when both are I(1):

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + \gamma_0 \Delta x_t + \gamma_1 \Delta x_{t-1} + u_t.$$

- If $s_t = y_t - x_t \beta \sim I(0)$, we may add s_{t-1} to the above equation:

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + \gamma_0 \Delta x_t + \gamma_1 \Delta x_{t-1} + \delta s_{t-1} + u_t,$$

  where $\mathbb{E}\left(u_t | \mathcal{F}_{t-1}\right) = 0$.

- The above is called the **error correction model**.

Error Correction Models

- How should we interpret this model?
- If $s_{t-1} > 0$, then y in the previous period has overshot the equilibrium; because $\delta < 0$, the error correction term works to push y back toward the equilibrium.
- If $s_{t-1} < 0$, then ...
- Estimation of error correction model is fairly straightforward, if β is known.

Example

We consider the error correction model for the holding yields:

$$\Delta hy6_t = \alpha_0 + \gamma_0 \Delta hy3_{t-1} + \delta \left(hy6_{t-1} - hy3_{t-2} \right) + u_t,$$

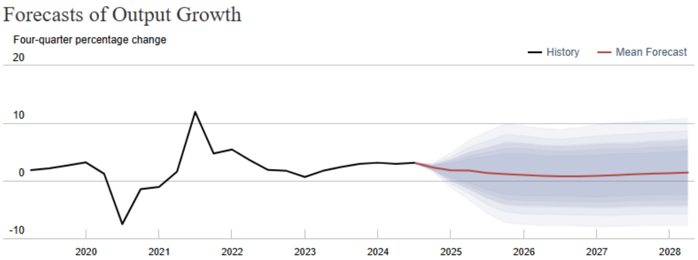
where

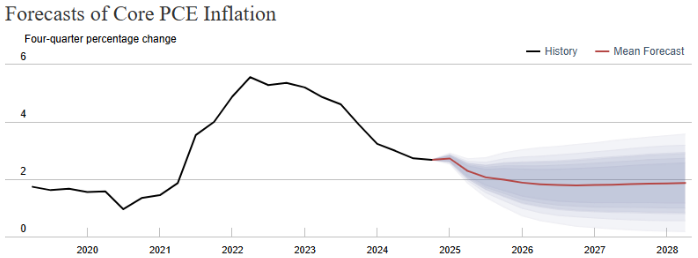
- $hy6_t$: three-month holding yield (in percent) from buying a six-month T-bill at time $t-1$ and selling it at time t as a three-month T-bill;
- $hy3_{t-1}$: three-month holding yield from buying a three-month T-bill at time $t-1$.

The expectations hypothesis implies, at a minimum, $hy6_t$ and $hy3_{t-1}$ are cointegrated. In addition, we expect $\delta < 0$.

Forecasting

Forecasting from The New York Fed DSGE Model





Set up of the environment

- y_{t+1}: target to forecast
- $\mathcal{F}_t$: information set
- f_t: a **one-step-ahead** forecast
- $f_{t,h}$: a *h*-**step-ahead** forecast
- $e_{t+1} = y_{t+1} - f_t$: forecast error
- $\ell(\cdot)$: loss function

Optimal forecast

Optimal forecast f_t^* is defined as one minimizes the conditional expected loss:

$$\hat{f}_t^* := \arg\min \ \mathbb{E}\left[\ell(e_{t+1})|\mathcal{F}_t\right].$$

Optimal forecast under squared error loss

When $\ell\left(e_{t+1}\right) = e_{t+1}^2$, $\hat{f}_t^* = \mathbb{E}\left(y_{t+1}|\mathcal{F}_t\right)$.

Example

We say the process $(y_t)_t$ is a

- martingale difference sequence (**M.D.S.**) if $\mathbb{E}\left(y_{t+1}|\mathcal{F}_t\right) = 0$;
- martingale if $\mathbb{E}\left(y_{t+1}|\mathcal{F}_t\right) = y_t$.

Then, under the squared error loss, the one-step-ahead optimal forecast of

- a M.D.S. is always zero;
- a martingale is always the present value.

Types of Regression Models Used for Forecasting

- Static models:

$$y_t = \beta_0 + \beta_1 z_t + u_t$$

$$\mathbb{E}\left(y_{t+1}|\mathcal{F}_t\right) = \mathbb{E}\left(\beta_0 + \beta_1 z_{t+1} + u_{t+1}|\mathcal{F}_t\right) = \beta_0 + \beta_1 \mathbb{E}\left(z_{t+1}|\mathcal{F}_t\right)$$

- Dynamic models:

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \gamma_1 z_{t-1} + u_t \tag{3}$$

$$\mathbb{E}\left(y_{t+1}|\mathcal{F}_t\right) = \mathbb{E}\left(\delta_0 + \alpha_1 y_t + \gamma_1 z_t + u_{t+1}|\mathcal{F}_t\right) = \delta_0 + \alpha_1 y_t + \gamma_1 z_t$$

- All need to assume that $\mathbb{E}\left(u_{t+1}|\mathcal{F}_t\right) = 0$!!!

Interval forecast

- Suppose that sample runs from $t = 1, 2, \cdots, n$ and we need to forecast y_{n+1} using model (3).
- We obtain the LS estimator $\hat{\beta}_n = \left(\hat{\delta}_{0n}, \hat{\alpha}_{1n}, \hat{\gamma}_{1n}\right)'$ and obtain the optimal forecast

$$\hat{f}_n = \mathbf{X}_n \hat{\beta}_n,$$

  where $\mathbf{X}_n = (1, y_n, z_n)'$.
- $\hat{f}_n$ is a point forecast and a random variable.
- How should we capture the uncertainty around the point forecast?

Interval forecast

- The forecast error is given by

$$\hat{e}_{n+1} = y_{n+1} - \hat{f}_n.$$

- Thus,

$$\text{se}\left(\hat{e}_{n+1}\right) = \sqrt{\left[\text{se}(\hat{f}_n)\right]^2 + \hat{\sigma}^2}.$$

- The (approximate) 95% forecast interval is

$$\hat{f}_n \pm 1.96 \cdot \text{se}\left(\hat{e}_{n+1}\right).$$

- $\text{se}\left(\hat{e}_{n+1}\right)$ can also be obtained by running the augmented regression y_t on $(1, y_{t-1}, z_{t-1}, dnp1_t)'$:
    - $dnp1_t$ equals 1 if $t = n + 1$ and zero otherwise
    - The coefficient on $dnp1_t$ is actually the forecast error.

Comparing One-Step-Ahead Forecasts

- Suppose we specify the following predictive regression model for stock return:

$$r_{t+1} = \alpha + x_t\beta + u_{t+1}$$

- How should we assess the predictability?
- Use in-sample criteria, such as R^2?
- Use out-of-sample (OOS) criteria?

Multiple-Step-Ahead Forecasts

- Suppose we specify an AR(1) model:

$$y_t = \alpha + \rho y_{t-1} + u_t,$$

  where $\mathbb{E}\left(u_t | \mathcal{F}_{t-1}\right) = 0$.

- Under squared error loss, the optimal 1-step ahead forecast is

$$\hat{f}_n = \alpha + \rho y_n,$$

  with forecast error

$$e_{n+1} = u_{n+1}$$

  is also a **M.D.S.**.

- What about 2-step ahead?

Multiple-Step-Ahead Forecasts

- Let us compute the conditional expectation:

$$\mathbb{E}\left(y_{n+2}|\mathcal{F}_n\right) = \alpha + \rho\mathbb{E}\left(y_{n+1}|\mathcal{F}_n\right) = \alpha + \rho\hat{f}_n = \alpha(1+\rho) + \rho^2 y_n,$$

  where we have replaced the unknown y_{n+1} with its optimal forecast.
- The forecast error is given by

$$\begin{aligned} e_{n+2} = y_{n+2} - \hat{f}_{n,2} &= \{\alpha + \rho y_{n+1} + u_{n+2}\} - \left\{\alpha(1+\rho) + \rho^2 y_n\right\} \\ &= \rho\left(y_{n+1} - \alpha - \rho y_n\right) + u_{n+2} \\ &= \rho u_{n+1} + u_{n+2}, \end{aligned}$$

  which is a MA(1) process.

Multiple-Step-Ahead Forecasts

- The h-step ahead optimal forecast from an AR(1) model can be shown to take the form

$$\hat{f}_{n,h} = \left(1 + \rho + \cdots + \rho^{h-1}\right) \alpha + \rho^{h} y_n.$$

- The forecast error e_{n+h} follows a MA(h-1) process.
- Interval forecast can be similarly defined, but perhaps too narrowed or even actually invalid.

Multiple-Step-Ahead Forecasts

- What about AR(2) model?

$$y_t = \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + u_t,$$

  where $\mathbb{E}\left(u_t | \mathcal{F}_{t-1}\right) = 0$.
- Same as before, and we could obtain a recursive formula.

Forecasting with integrated time series

- Because

$$y_{t+1} = \Delta y_{t+1} + y_t$$

- Our forecast of y_{n+1} at time n is just

$$\hat{f}_n = \hat{g}_n + y_n,$$

where $\hat{g}_n$ is a forecast of transitory dynamics Δy_{n+1} at time n. Needs to specify an AR model.

- Multi-step ahead forecasts follow similar, since

$$\begin{aligned} y_{n+h} &= (y_{n+h} - y_{n+h-1}) + (y_{n+h-1} - y_{n+h-2}) + \cdots + (y_{n+1} - y_n) + y_n \\ &= \Delta y_{n+h} + \Delta y_{n+h-1} + \cdots + \Delta y_{n+1} + y_n \end{aligned}$$

Example

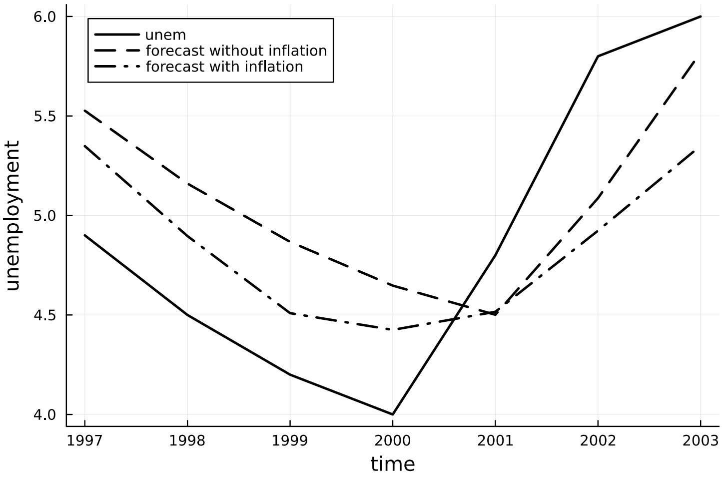
We again use the dataset PHILLIPS, but only for the years 1948 through 1996, to forecast the U.S. unemployment rate from 1997 to 2003. We use two models. The first one is an AR(1) model:

$$unem_t = \alpha + \beta unem_{t-1} + u_t.$$

In the second model, we add lagged inflation as an additional predictor:

$$unem_t = \alpha + \beta unem_{t-1} + \gamma inf_{t-1} + u_t.$$

The OOS RMSE based on the first model is roughly 0.576, while for second model, it is 0.522. If we use MAE, it is 0.542 for the first model and 0.484 for the second model, respectively. Both criteria indicate adding inflation to the model helps to improve the forecast accuracy for the unemployment rate.


unem
forecast without inflation
forecast with inflation
6.0
5.5
5.0
4.5
4.0
unemployment
1997
1998
1999
2000
2001
2002
2003
time