14.384 Time Series Analysis, Fall 2007
Recitation by Paul Schrimpf
Supplementary to lectures given by Anna Mikusheva
September 11, 2008

Recitation 2

HAC

Goal: estimate $\mathcal{J} = \sum_{-\infty}^{\infty} \gamma_k$ (or, more generally, do inference on $\hat{\beta}$, which has asymptotic variance $\mathcal{J}$)

Methods:

1. *Parametric*: estimate ARMA(p,q) for z_t:

$$A(L)z_t = B(L)e_t$$

   Recall the relationship between the spectrum and $\mathcal{J}$. The spectral density is

$$S(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} e^{-i\omega j} \gamma_j$$

   so,

$$2\pi S(0) = \mathcal{J}$$

   Also, remember that the spectrum of an ARMA[1] is:

$$S(\omega) = \frac{1}{2\pi} \sigma^2 \frac{|B(e^{i\omega})|^2}{|A(e^{i\omega})|^2}$$

   so, for an ARMA,

$$\mathcal{J} = \sigma^2 \frac{B(1)^2}{A(1)^2}$$

   Thus, we can estimate $\mathcal{J}$ by estimating $\hat{B}(L)$ and $\hat{A}(L)$ using standard methods (OLS if the ARMA has a finite order AR representation, the Kalman filter otherwise), and then estimate $\mathcal{J}$ as

$$\hat{\mathcal{J}} = \hat{\sigma}^2 \frac{\hat{B}(1)^2}{\hat{A}(1)^2} \tag{1}$$

   - As Anna said, in practice this is often non-parametric since people tend to increase p and q with sample size

---

[1]Anna showed this for the covariance function at the end of lecture 1.

- For multivariate series, this is called VARMA-HAC (or just VAR-HAC) and (1) becomes:

$$\hat{\mathcal{J}} = \hat{A}(1)^{-1}\hat{B}(1)\hat{\Sigma}\hat{B}(1)'\hat{A}(1)^{1'}$$

2. *Non-parametric*: uses a truncated, weighted sum of sample covariances to estimate the long-run variance:

$$\hat{\mathcal{J}} = \sum_{-S_T}^{S_T} k_T(j)\hat{\gamma}_j$$

where S_T grows with T, but slowly so that $\{\hat{\gamma}_j\}_{j=0}^{S_T}$ are consistent, and $k_T(j)$ guarantee that $\hat{\mathcal{J}}$ is positive definite. See lecture 3 for more details.

- *Prewhitening*: nonparametric HAC performs poorly when the series is persistent. Parametric HAC performs poorly if the model is wrong. Prewhitening combines the two. From the above we know that if e_t is white noise with variance Σ, then when $A(L)z_t = B(L)e_t$, the long-run variance of z_t is

$$\mathcal{J}_z = A(1)^{-1}B(1)\Sigma B(1)'A(1)^{1'}$$

Similarly if e_t is not white noise, but has long-run variance $\mathcal{J}_e$, then

$$\mathcal{J}_z = A(1)^{-1}B(1)\mathcal{J}_e B(1)'A(1)^{1'}$$

The prewhitened nonparametric estimate of $\mathcal{J}_z$ is then simply:

$$\hat{\mathcal{J}}_z = \hat{A}(1)^{-1}\hat{B}(1)\hat{\mathcal{J}}_e\hat{B}(1)'\hat{A}(1)^{1'}$$

where $\hat{A}$ and $\hat{B}$ are estimated by OLS or Kalman filtering, and $\hat{J}_e$ is estimated by doing nonparametric HAC hat $\hat{e}_t$.

3. *Keifer-Vogelsang*: set $S_T = T - 1$, which makes $\hat{\mathcal{J}}$ converge to a distribution instead of $\mathcal{J}$. They then calculate the limiting distribution of $t = \frac{\hat{\beta}_i}{\sqrt{\hat{\mathcal{J}}_{ii}}}$. This has a non-normal limiting distribution, which can be used for testing.

- We will see a lot of non-normal limiting distributions in a couple of weeks when we cover unit roots and the functional central limit theorem. It would be a good exercise to come back and try to derive the Keifer-Vogelsang result.

- Müller (2007) takes a related approach that just focuses on low frequency observations.

Practical Advice This summer, Mark Watson gave a lecture on HAC
http://nber15.nber.org/c/2008/si2008/TSE/Lecture9.pdf
and this is a short summary of what he recommended. When doing HAC, you have to
choose which of the three methods to use, and then if you choose ARMA, the lag lengths,
or if you choose nonparametric, the kernel and bandwidth. In this discussion, the goal is to
do inference on $\hat{\beta}$

- Simulations show large size distortions for all methods (reject at 5% level far more
  than 5% of time). Tests work worse when

  - Sample size is smaller

  - Data is more persistent (e.g. an AR(1) with coefficient near one)

- If it is the correct model, parametric ARMA works best. Sometimes theory suggests
  an ARMA (den Haan and Levin 1997).

- Kiefer-Vogelsang leads to smaller size distortions, but has less power than kernel meth-
  ods

- For kernel methods:

  - The theoretically optimal[2] kernel is called the quadratic-spectral (QS) kernel. In
    practice, all common kernels perform similarly.

  - For inference, it is not necessarily best to minimize MSE of $\hat{\mathcal{J}}$

    * See Sun, Philips, and Jin (2008) for a more formal discussion

    * Intuition: suppose $z \sim N(\mu, \sigma^2)$ (think of z as $\sqrt{n}(\beta - \hat{\beta}_0)$) and $\hat{\sigma}^2$ is an
      estimate of σ^2. For testing $H_0 : \mu = 0$ at level α, we would compute a critical
      value, c, from the normal distribution such that $P(|z/\sigma| < c) = \alpha$. If we
      don't know σ, then this how well this test would work depends on how close
      $P\left(\frac{z^2}{\hat{\sigma}^2} < c^2\right)$ is to $P\left(\frac{z^2}{\sigma^2} < c^2\right)$. Very loosely:

$$P\left(\frac{z^2}{\hat{\sigma}^2} < c^2\right) = E\left[\mathbf{1}(z^2 < \hat{\sigma}^2 c^2)\right] = E\left[g(\hat{\sigma}^2)\right] \tag{2}$$

$$\approx E\left[g(\sigma^2) + (\hat{\sigma}^2 - \sigma^2)g'(\sigma^2) + \frac{1}{2}(\hat{\sigma}^2 - \sigma)^2 g''(\sigma^2)\right] \tag{3}$$

$$\approx Eg(\sigma^2) + Bias(\hat{\sigma}^2)g' + \frac{1}{2}MSE(\hat{\sigma}^2)g'' \tag{4}$$

So the error in the test depends on a combination of the bias and MSE of $\hat{\mathcal{J}}$.
The best choice of S_T for testing shouldn't minimize MSE; it should minimize

---

[2]In the sense that it minimizes MSE of $\hat{\mathcal{J}}$

this combination of bias and MSE. Since bias decreases with S_T, the best S_T for testing is greater than the best S_T for MSE.[3]

* Andrews (1991) gives formula for minimal MSE choice of S_T.

  · For inference: use larger S_T

  · For a GMM weighting matrix, minimal MSE seems like a good choice

* Similar reasoning suggests (maybe) using a longer lag length for an ARMA model than suggested by BIC and maybe AIC too (we will cover these in lecture 5)

---

[3]This discussion ignores another issue in testing things like $\sqrt{n}(\beta - \hat{\beta}_0)$. It assumed that $z \sim N$, while in practice we usually only know that z is asymptotically normal. With persistent data, which is common, the finite sample distribution can be far from normal. We will see more of this later.

MIT OpenCourseWare

14.384 Time Series Analysis
Fall 2013