Specification Tests for Nonlinear Time Series Models

Igor Kheifets

Department of Economics
Universidad Carlos III de Madrid

January 15, 2010
The Null

- Data \( Y_1, Y_2, \ldots, Y_n \).
- Information set \( \Omega_t = (Y_{t-1}, Y_{t-2}, \ldots) \).
- Family of conditional cdf \( F_t(y|\Omega_t, \theta), \theta \in \Theta \subseteq R^L \).

\[ H_0 : \exists \theta_0 \in \Theta \text{ s.t. } Y_t|\Omega_t \sim F_t(\cdot|\Omega_t, \theta_0). \]
Conditional probability integral transform (CPIT)

**Assumption A1.** The conditional cdf $F_t(y | \Omega_t, \theta)$ are continuous and strictly increasing in $y$.

**Proposition** Suppose A1 holds. Then under $H_0$ random variables

$$U_t = F_t(Y_t | \Omega_t, \theta_0)$$

are i.i.d. uniform.

- The idea goes back to at least Rosenblatt (1952).
- Bai (2003), specification test.
Distribution and Independence tests literature

1. $\hat{F}_X(x) - F_X(x)$, parametric $\hat{F}_X(x) - F_X(x, \hat{\theta})$
   - CPIT, $\hat{F}_X(x) - x$: Bai (2003).

2. $\hat{F}_{X,Y}(x, y) - \hat{F}_X(x)\hat{F}_Y(y)$
   - Empirical df, 2 rv’s: Hoeffding (1948), Blum et al. (1961).

This paper

\[ \implies \hat{F}_{X,Y}(x, y) - xy \]

- Hong and Li (2005) with kernel density estimates.
- Delgado and Stute (2008) in i.i.d setup with covariates.
A new statistics

- Bi-parameter process, for $U_t = F_t(Y_t|\Omega_t, \theta_0)$, $r \in [0, 1]^2$

\[ V_{2n}(r) = \frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} [I(U_t \leq r_1)I(U_{t-1} \leq r_2) - r_1r_2]. \]

- If we do not know $\theta_0$ either $\{Y_t, t \leq 0\}$, we approximate $U_t$ with $\hat{U}_t = F_t(Y_t|\tilde{\Omega}_t, \hat{\theta})$ where $\hat{\theta}$ is an estimator of $\theta_0$ and truncated information is $\tilde{\Omega}_t = (Y_{t-1}, Y_{t-2}, \ldots, Y_1)$

\[ \hat{V}_{2n}(r) = \frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} [I(\hat{U}_t \leq r_1)I(\hat{U}_{t-1} \leq r_2) - r_1r_2]. \]

- The test statistics is $D_{2n} = \Gamma(\hat{V}_{2n}(r))$. For example

\[ D_{2n}^{CvM} = \int_{[0,1]^2} \hat{V}_{2n}(r)^2 \, dr \quad \text{or} \quad D_{2n}^{KS} = \max_{[0,1]^2} \left| \hat{V}_{2n}(r) \right|. \]
Generalizations: $p$-wise and other lags

- More parameters, as Delgado (1996), for $r \in [0, 1]^p$

\[
\hat{V}_{pn}(r) = \frac{1}{\sqrt{n - (p + 1)}} \sum_{t=p+1}^{n} \left[ \prod_{j=1}^{p} I(\hat{U}_{t-j} \leq r_j) - r_1 r_2 \cdots r_p \right]
\]

\[
D_{pn}^{CvM} = \int_{[0,1]^p} \hat{V}_{pn}(r)^2 \, dr \quad \text{or} \quad D_{pn}^{KS} = \max_{[0,1]^p} |\hat{V}_{pn}(r)|.
\]

- Different lags

\[
\hat{V}_{2n,j}(r) = \frac{1}{\sqrt{n - j}} \sum_{t=j+1}^{n} \left[ I(\hat{U}_t \leq r_1) I(\hat{U}_{t-j} \leq r_2) - r_1 r_2 \right].
\]

\[
D_{2n,j}^{CvM} = \int_{[0,1]^2} \hat{V}_{2n,j}(r)^2 \, dr \quad \text{or} \quad D_{2n,j}^{KS} = \max_{[0,1]^2} |\hat{V}_{2n,j}(r)|.
\]
Generalizations: aggregate

- We can aggregate across $p$ or $j$, so that for $k = 1, ..., n - 1$ we get test statistics

\[ ADJ_{kn} = \sum_{j=1}^{k} D_{2n,j}^{CvM} \text{ and } MDJ_{kn} = \max_{j=1,\ldots,k} D_{2n,j}^{KS}. \]

- Adding information from $D_{1n}$, i.e. additional explicit account for unconditional distribution misspecification, we get

\[ ADJ_{kn}^0 = D_{1n}^{CvM} + ADJ_{kn} \text{ and } MDJ_{kn}^0 = \max \left( D_{1n}^{KS}, MDJ_{kn} \right). \]
Asymptotic Properties under the null

**Proposition**  Suppose A1 hold. Then under $H_0$, $V_{2n}(r) \Rightarrow V_{2\infty}(r)$, where $V_{2\infty}(r)$ is a zero mean Gaussian process with covariance

$$
\text{Cov}_{V_{2\infty}}(r, s) = (r_1 \wedge s_1)(r_2 \wedge s_2) + (r_1 \wedge s_2)r_2s_1 + (r_2 \wedge s_1)r_1s_2 - 3r_1r_2s_1s_2.
$$

It is not a bivariate Brownian bridge!

$$
\lim_{n \to \infty} \text{Cov}_{S_n}(r, s) = ((r_1 \wedge s_1) - r_1s_1)((r_2 \wedge s_2) - r_2s_2).
$$
Bivariate Brownian bridge and our process
Assumptions on smoothness of $F(\cdot|\cdot, \theta)$

**Assumption A2.** For every $M > 0$

$$\sup_{u,v \in B(\theta_0, Mn^{-1/2})} \sup_{r \in [0,1]^2} \max_{t=1,\ldots,n} \left| F_t \left( F_t^{-1} (r|u) | v \right) - r \right| = o_p(1).$$

$$\sup_{u,v \in B(\theta_0, Mn^{-1/2})} \sup_{r \in [0,1]^2} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left| F_t \left( F_t^{-1} (r|u) | v \right) - r \right| = O_p(1).$$

There exists a uniformly continuous (vector) function $h(\cdot) : [0,1]^2 \to R^L$ such that for every $M > 0$

$$\sup_{u,v \in B(\theta_0, Mn^{-1/2})} \sup_{r \in [0,1]^2} \left\| \frac{1}{\sqrt{n}} \sum_{t=2}^{n} \left[ F_{t-1} \left( F_{t-1}^{-1} (r_2|u) | v \right) - r_2 \right] r_1 
+ \left( F_t \left( F_t^{-1} (r_1|u) | v \right) - r_1 \right) I(U_{t-1} \leq r_2) 
- h(r)' \sqrt{n} (u - v) \right\| = o_p(1).$$
Assumptions on $F(\cdot|\cdot, \theta)$, if differentiable

**Assumption A2’.** There exists a uniformly (in $x$ and $t$) continuous (wrt $\theta$) gradient $\nabla_\theta F_t (x|\theta)$ which is also bounded, i.e. for some $M > 0$:

$$\sup_{\theta \in B(\theta_0, Mn^{-1/2})} \sup_x \| \nabla_\theta F_t (x|\theta) \| = O_p(1).$$

There exists a uniformly continuous (vector) function $h_1(\cdot) : [0, 1] \to R^L$ and $h_2(\cdot) : [0, 1]^2 \to R^L$ such that for every $M > 0$ (they are related $h(r) = h_1(r_2)r_1 + h_2(r)$)

$$\sup_{u,v \in B(\theta_0, Mn^{-1/2})} \sup_{r \in [0,1]} \left\| \frac{1}{n} \sum_{t=1}^n \nabla_\theta F_t \left( F_t^{-1} (r|u) |v \right) - h_1(r) \right\| = o_p(1).$$

$$\sup_{u,v \in B(\theta_0, Mn^{-1/2})} \sup_{r \in [0,1]^2} \left\| \frac{1}{n} \sum_{t=2}^n \nabla_\theta F_t \left( F_t^{-1} (r_1|u) |v \right) I (U_{t-1} \leq r_2) - h_2(r) \right\| = o_p(1).$$
Under $H_0$

Proposition Under Assumptions A1-A2, the following asymptotic representation holds under $H_0$, uniformly in $r$,

$$\hat{V}_{2n}(r) = V_{2n}(r) + h(r)\sqrt{n}(\hat{\theta} - \theta_n) + o_p(1).$$
Under local alternative

**Assumption A3.** The conditional df’s $H_t(y|\Omega_t)$ are continuous and strictly increasing in $y$, and differs from $F_t$ on a positive measure set.

Local parametric alternatives DGP:

$$G_{nt}(y|\Omega_t, \theta) := \left(1 - \frac{\delta}{\sqrt{n}}\right) F_t(y|\Omega_t, \theta) + \frac{\delta}{\sqrt{n}} H_t(y|\Omega_t).$$

**Proposition** Under Assumptions A1-A3, the following asymptotic representation holds under $G_{nt}$, uniformly in $r$,

$$\hat{V}_{2n}(r) = V_{2n}(r) + \delta(k_1(r) + k_2(r)) + h(r)' \sqrt{n}(\hat{\theta} - \theta_n) + o_p(1).$$
Drifts

Bai (2003)

\[ k_{01}(r_1) := \text{plim} \frac{1}{n} \sum_{t=1}^{n} \left[ H_t(F_t^{-1}(r_1 | \Omega_t, \theta_0) | \Omega_t) - r_1 \right]. \]

Our process

\[ k_1(r) := \text{plim} \frac{1}{n} \sum_{t=2}^{n} \left[ k_{01}(r_1) r_2 + k_{01}(r_2) r_1 \right], \]

\[ k_2(r) := \text{plim} \frac{1}{n} \sum_{t=2}^{n} \left[ (H_t(F_t^{-1}(r_1 | \Omega_t, \theta_0) | \Omega_t) - r_1) (I(U_{t-1} \leq r_2) - r_2) \right]. \]

AR(1) vs AR(2) example: \( k_{01}(r_1) = 0 \), and hence \( k_1(r) = 0 \) but \( k_2(r) \neq 0 \).
Monte Carlo 1. GARCH(1,1) vs AR(1)-GARCH(1,1)

- **DGP**: 
  \[ Y_t = \alpha_1 Y_{t-1} + h_t \varepsilon_t \]
  with \[ h_t^2 = 0.1 + 0.1Y_{t-1}^2 + 0.8h_{t-1}^2 \]
  and \( \alpha_1 = -0.8, ..., -0.2, 0, 0.2, ..., 0.8 \).
- **\( H_0 \)**: GARCH(1,1), i.e. with \( \alpha_1 = 0 \).
- Innovations in both models are independent Gaussian or \( t_5 \).
- Sample sizes \( n = 100 \) and \( n = 300 \).
GARCH(1,1)-N vs AR(1)-GARCH(1,1)-N; \( n = 100 \); \( ADJ^0 \)
GARCH(1,1) vs AR(1)-GARCH(1,1), N; n=100, 300; ADJ
GARCH(1,1) vs AR(1)-GARCH(1,1), $t_5$; $n=100, 300$; ADJ

Specification Tests for Nonlinear Time Series
Igor Kheifets (Carlos III)
Model for NYSE index

- NYSE monthly equal weighted returns, 01/1926-12/1999
- Bai (2003) rejects GARCH(1,1)-N at 1% significance level, while can not reject GARCH(1,1)-t_5 at 5%.
Model for NYSE index: GARCH(1,1)-N; CvM
Model for NYSE index: GARCH(1,1)-N,t₅; CvM, KS
Model for NYSE index: AR(1)-GARCH(1,1)-N, t5; CvM, KS
Model for NYSE index

<table>
<thead>
<tr>
<th></th>
<th>p-Values, CvM statistics</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(H_0)</td>
<td>(D_{1n}^{CvM})</td>
<td>(ADJ_{1n})</td>
<td>(ADJ_{5n})</td>
</tr>
<tr>
<td>1</td>
<td>GARCH(1,1)</td>
<td>0.01***</td>
<td>0.01***</td>
<td>0.01***</td>
</tr>
<tr>
<td>2</td>
<td>GARCH(1,1)-(t_5)</td>
<td>0.38</td>
<td>0.03**</td>
<td>0.03**</td>
</tr>
<tr>
<td>3</td>
<td>AR(1)-GARCH(1,1)</td>
<td>0.01***</td>
<td>0.01***</td>
<td>0.01***</td>
</tr>
<tr>
<td>4</td>
<td>AR(1)-GARCH(1,1)-(t_5)</td>
<td>0.55</td>
<td>0.38</td>
<td>0.20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>p-Values, K-S statistics</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(H_0)</td>
<td>(D_{1n}^{KS})</td>
<td>(MDJ_{1n})</td>
<td>(MDJ_{5n})</td>
</tr>
<tr>
<td>1</td>
<td>GARCH(1,1)</td>
<td>0.01***</td>
<td>0.01***</td>
<td>0.01***</td>
</tr>
<tr>
<td>2</td>
<td>GARCH(1,1)-(t_5)</td>
<td>0.16</td>
<td>0.06*</td>
<td>0.12</td>
</tr>
<tr>
<td>3</td>
<td>AR(1)-GARCH(1,1)</td>
<td>0.01***</td>
<td>0.01***</td>
<td>0.01***</td>
</tr>
<tr>
<td>4</td>
<td>AR(1)-GARCH(1,1)-(t_5)</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Conclusion

- We show analytically and with simulations that Kolmogorov type tests, such as Bai (2003) are inconsistent.
- A new specification procedure is proposed.
  - Weak convergence of the underlying bivariate stochastic process is established, under the null and local alternative, taking into account the parameter estimation effect.
  - Parametric bootstrap approximation is justified.
- The test has a nontrivial power under a wide range of alternatives.
- Important for the efficient likelihood, intensity and hazard specification, risk measurement, forecasting.
- Can be applied to location scale (ARMA, GARCH) and diffusion models.
- Future agenda: study latent variable effect (SV) and multivariate models (DCC).