SNP: Nonparametric Time Series Analysis

by

A. Ronald Gallant
Department of Economics
University of North Carolina
Chapel Hill NC 27599-3305 USA

Application

S&P 500, 1928–1987

Price and Volume, 16127 observations

Files: nyse.doc, nyse.dat

References

Gallant, A. Ronald, and George Tauchen (1998), *SNP: A Program for Nonparametric Time Series Analysis, User's Guide* At ftp.econ.duke.edu in pub/arg/snp.

Gallant, A. Ronald, Peter E. Rossi, and George E. Tauchen (1992), "Stock Prices and Volume," *The Review of Financial Studies* 5, 199–242.

Gallant, A. Ronald, Peter E. Rossi, and George E. Tauchen (1993), "Nonlinear Dynamic Structures," *Econometrica* 61, 871–907.

Topics

- Application
- Hermite Expansions
- SNP Density
  - Performance
  - Model Selection
- Extension to Time Series
  - VAR-type Location Function
  - ARCH- or GARCH-type Scale Function
  - Non-homogeneous Innovations
  - Miscellany

Assumptions

1. Stationary, multivariate

$$y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Mt} \end{pmatrix} \qquad M \times 1$$

   Stationarity is assumed so that densities for a stretch of data are time invariant. That is, they are of the form $f(y_{t-L}, .., y_t)$ rather than $f_t(y_{t-L}, .., y_t)$.

2. Markovian

   The conditional density of y_t given the entire past depends only on a finite number of lags That is, $f(y_t|y_{t-\tau}, .., y_{t-1}) = f(y_t|x_{t-1})$ for every $\tau \geq L$, where

$$x_{t-1} = (y_{t-L}, .., y_{t-1})' \quad ML \times 1$$

3. Smooth

   The density $f(y_{t-L}, .., y_t)$, which is the same as $f(x_{t-1}, y_t)$ in the notation above, must have derivatives to the order ML/2 or higher and have tails that are bounded above by $\mathcal{P}(y_{t-L}, .., y_t) \exp\left(\frac{1}{2}\sum_{\tau=0}^{L} y_{t-L}^2\right)$ where $\mathcal{P}$ is a polynomial of large but finite degree.

Transition Density

The transition density of a Markov process is the conditional density

$$f(y_t|x_{t-1}) = f(y_t|y_{t-L}, \ldots, y_{t-1})$$

Given the functional form $f(x,y) = f(y_{-L}, \ldots, y_{-1}, y_0)$ of the joint density the transition density can be obtained from

$$f(y|x) = \frac{f(y,x)}{\int f(y,x)\,dy}.$$

Conversely, given the functional form of a transition density $f(y|x) = f(y_0|y_{-L}, \ldots, y_{-1})$ the marginal density can be recovered by solving the equation

$$f(x) = \int f(y|x) f(x)\,dy$$

for $f(x)$ and the joint density can be obtained from this solution using

$$f(x,y) = f(y_{-L}, \ldots, y_{-1}, y_0) = f(y|x)f(x)$$

Thus, either $f(x,y)$ or $f(y|x)$ can be regarded as containing all the probabilistic information about a Markovian process $\{y_t\}$ and either is a proper focus of nonparametric interest. We shall focus on estimation of the transition density.

Application

The application used for illustration is the S&P 500 price and volume series from 1928–1987 used in Gallant, Rossi, and Tauchen (1992, 1993). The data are in file nyse.dat have been adjusted to remove calendar effects as described in nyse.doc. The multivariate series used for analysis is

$$y_t = \begin{pmatrix} 100 * [\log(P_t) - \log(P_{t-1})] \\ \log(V_t) \end{pmatrix}$$

where P_t is the closing Standard and Poors price index and V_t is the daily volume on the New York Stock Exchange. We abbreviate as

$$y_t = \begin{pmatrix} \Delta p_t \\ v_t \end{pmatrix}$$

Interpretation

Using the GRT nonparametric estimate $\hat{f}_n(y|x)$ of the transition density, we will illustrate some analyses that are possible once a nonparametric transition density estimate has been obtained because it seems reasonable to be sure that having an estimate is of some practical value before going to the bother of derivation and computation.

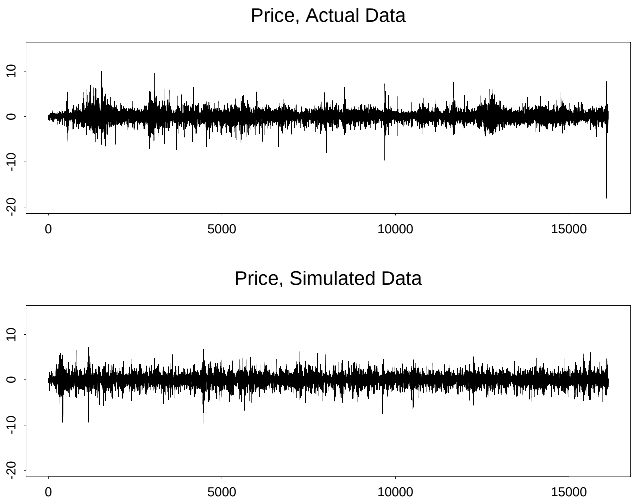
The GRT fit to the S&P 500 that we shall use to illustrate the interpretation of a nonparametric fit has $L = 16$:

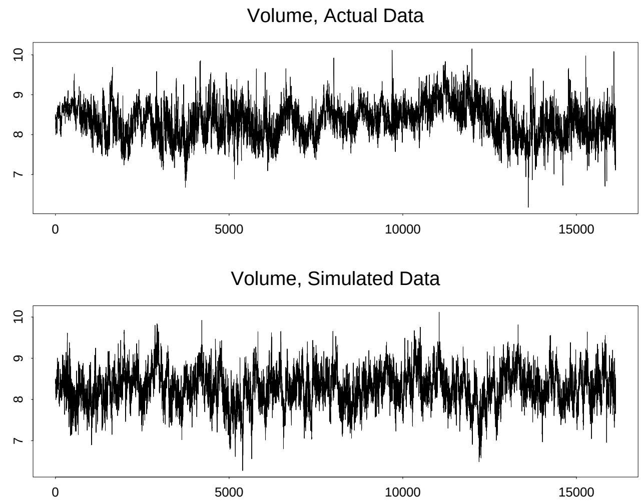
$$\hat{f}(y|x) = \hat{f}_n(\Delta p_0, v_0 \,|\, \Delta p_{-16}, v_{-16}, \ldots, \Delta p_{-1}, v_{-1})$$

$$\hat{f}(y_t|x_{t-1}) = \hat{f}_n(\Delta p_t, v_t \,|\, \Delta p_{t-16}, v_{t-16}, \ldots, \Delta p_{t-1}, v_{t-1})$$

Simulation

One important application is simulation. From a simulation, one can asses the reasonableness of a fit by comparing simulated data to actual data. Also, one can compute both conditional and unconditional expectations of nonlinear functions by simulating and averaging.





Visualization

A visual impression of the conditional density is of interest. Shown next are surface and contour plots of

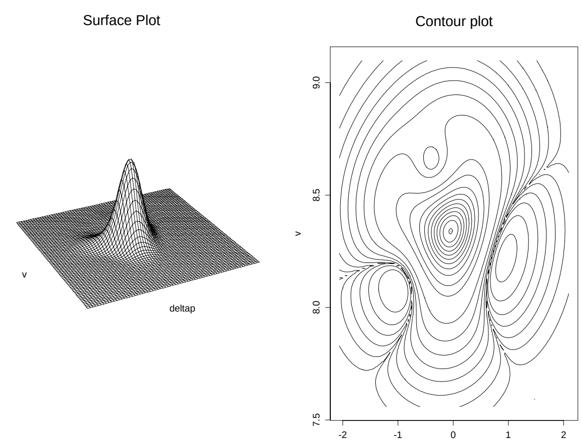
$$\widehat{f}_n(y,x)$$

in the variable

$$y = \begin{pmatrix} \Delta p \\ v \end{pmatrix}$$

with the elements of x set to the unconditional mean of the data. That is,

$$x = (y_{-16}, \ldots, y_{-1})' = (\bar{y}, \cdots, \bar{y})' \quad 32 \times 1$$



One-step ahead dynamics

Density:

$$\hat{f}_n(\Delta p_0, v_0 | \Delta p_{-16}, v_{-16}, \dots, \Delta p_{-1}, v_{-1})$$

Held fixed:

$\Delta p_t =$ sample mean for $t = -16, \dots, -2$
$v_t =$ sample mean for t=-16,...,-1

Vary:

Δp_{-1} over -15 to +15 sample std. devs. from the sample mean

Examine:

$\mathcal{E}(v|x) = \iint v \, \hat{f}_n(p, v|x) \, dp \, dv$

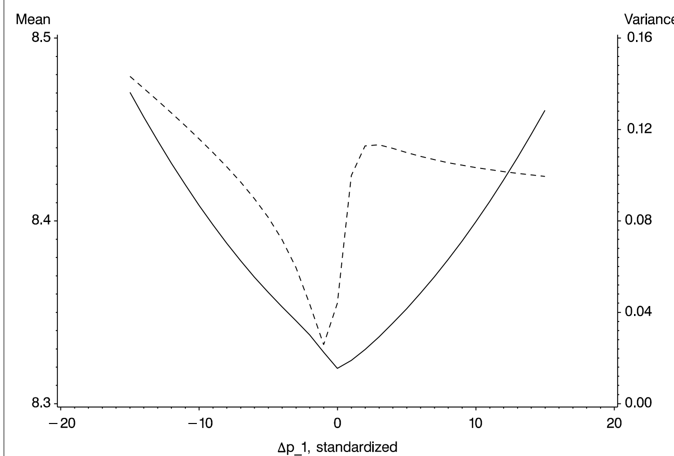
$\text{Var}(v|x) = \iint [v - \mathcal{E}(v|x)]^2 \, \hat{f}_n(p, v|x) \, dp \, dv$

where $x = (\Delta p_{-16}, v_{-16}, \dots, \Delta p_{-1}, v_{-1})$

Conclusion:

Large price movements are followed by high and volatile volume.

GRT (1992) demonstrated that the conditional variance of Δp and the conditional mean of v are nearly the same thing. Thus, the conclusion applies to price volatility as well.



Δp_1 is standardized with mu = 0.016, sigma = 1.15
Dashed: Conditional Mean Solid: Conditional Variance

Multi-step ahead dynamics

Density:

$\hat{f}_n(y_j | x_{j-1})$ $\quad \hat{\mathcal{E}}$, $\widehat{\text{Var}}$ computed wrt this density
$y_j = (\Delta p_j, v_j)' \quad x_{j-1} = (\Delta p_{j-16}, v_{j-16}, \dots, \Delta p_{j-1}, v_{j-1})'$

A Mean Profile:

$\hat{y}_j(x) = \hat{\mathcal{E}}[\hat{\mathcal{E}}(y_j | x_{j-1}) \,|\, x_0 = x]$
$j = 0, 1, 2, \dots, J$

A Volatility Profile:

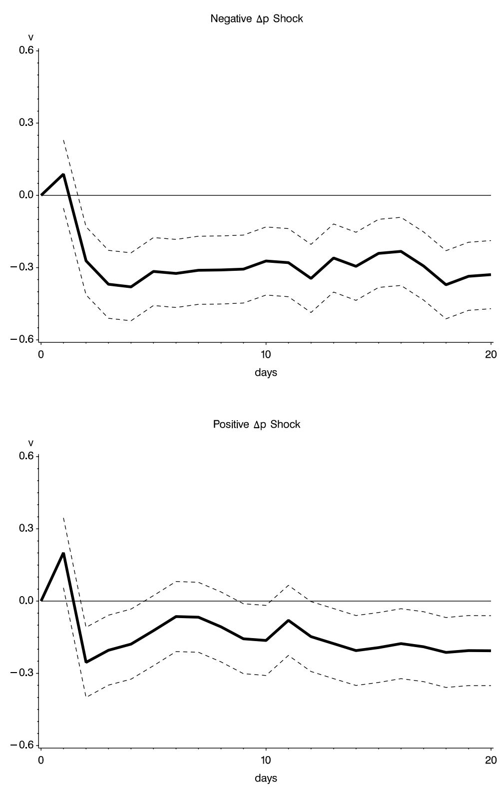
$\hat{\mathcal{V}}_j(x) = \hat{\mathcal{E}}[\widehat{\text{Var}}(y_j | x_{j-1}) \,|\, x_0 = x]$
$j = 1, 2, \dots, J$

A Shock:

$$\begin{array}{llll} x^+ & & \bullet & \text{positive shock} \\ x^0 & \underbrace{\bullet \; \bullet \; \bullet \; \bullet \; \bullet \; \bullet \; \bullet \; \bullet \; \bullet}_{x} & \bullet & \text{sample mean} \\ x^- & & \bullet & \text{negative shock} \end{array}$$

A Differential Response:

Mean: $\hat{y}_j(x+) - \hat{y}_j(x^o) \quad j = 0, 1, \dots, J$
Volatility: $\hat{\mathcal{V}}_j(x+) - \hat{\mathcal{V}}_j(x^o) \quad j = 1, \dots, J$



Sup-Norm Bands

The sup-norm bands shown in the previous plots were constructed as follows:

Bootstrap:

Using the initial conditions from the data and the estimated density, generate 500 simulated data sets. Estimate a density and compute a profile for each of the simulated data sets.

Sup-norm confidence bands:

ϵ-bands are plotted about the profile computed from the data that are just wide enough to contain 95% of the profiles computed from the simulated data sets.

Profile Bundles

A visual method for assessing persistence. One can fit an exponential curve to the bundles and compute a half-life to get a quantitative measure.

Price Profile:

$$\widehat{\Delta p}_j(x) = \mathcal{E}\Big[\mathcal{E}(\Delta p_j | x_{j-1}) \,\Big|\, x\Big] \qquad j = 1, \ldots, 100$$

Volume Profile:

$$\hat{v}_j(x) = \mathcal{E}\Big[\mathcal{E}(v_j | x_{j-1}) \,\Big|\, x\Big] \qquad j = 1, \ldots, 100$$

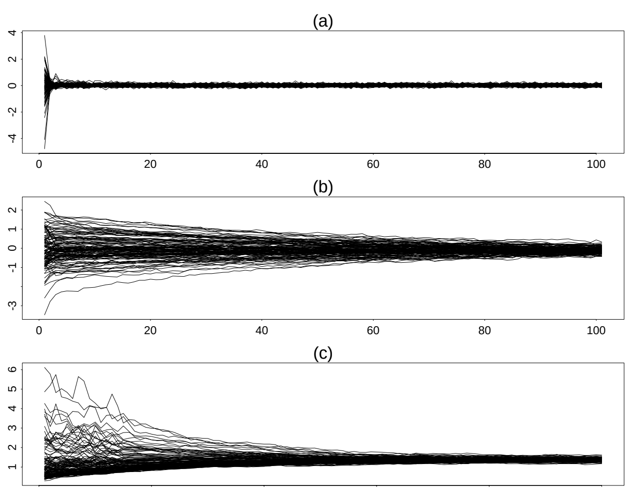
Price Volatility Profile:

$$\hat{\mathcal{V}}_j(x) = \mathcal{E}\Big[\mathrm{Var}(\Delta p_j | x_{j-1}) \,\Big|\, x\Big] \qquad j = 1, \ldots, 100$$

Profile Bundles: Evaluate at the data points

$$x = x_s, \qquad s = 28, 156, 258, \ldots, 16028$$

every 128th, 125 profiles in total.



(a) Price profile bundle, (b) Volume profile bundle, (c) Price volatility profile bundle.

Topics

- Application
- Hermite Expansions
- SNP Density
  - Performance
  - Model Selection
- Extension to Time Series
  - VAR-type Location Function
  - ARCH- or GARCH-type Scale Function
  - Non-homogeneous Innovations
  - Miscellany

Notation for a Multivariate Polynomial

Degree K, dimension M

$$\mathcal{P}(z) = \sum_{|\alpha|=0}^{K} a_\alpha z^\alpha$$

where

$$z^\alpha = (z_1)^{\alpha_1} \cdot (z_2)^{\alpha_2} \cdots (z_M)^{\alpha_M}$$

$$|\alpha| = |\alpha_1| + |\alpha_2| + \ldots + |\alpha_M|$$

Example, K=2, M=2

$$\begin{aligned} \mathcal{P}(z) &= a_{(0,0)} + \underbrace{a_{(1,0)}z_1 + a_{(0,1)}z_2}_{\text{linear terms}} \\ &\quad + \underbrace{a_{(1,1)}z_1 z_2 + a_{(2,0)}z_1^2 + a_{(0,2)}z_2^2}_{\text{quadratic terms}} \end{aligned}$$

Hermite Expansions: Rationale (1)

An unnormalized Hermite polynomial has the form

$$\mathcal{P}(z)\sqrt{\phi(z)}$$

where

$$\phi(z) = N_M(0, I) = (2\pi)^{-\frac{1}{2}M} e^{-\frac{1}{2}z'z}$$

A function $g(z)$ that satisfies

$$\|g\|_2 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g^2(z)\, dz_1 \cdots dz_M < \infty$$

is called an L_2 function and the collection of such functions is denoted by $L_2(-\infty, \infty)$.

The Hermite polynomials are dense in $L_2(-\infty, \infty)$ which means that

$$\lim_{K\to\infty} \left\| g(z) - \mathcal{P}(z)\sqrt{\phi(z)} \right\|_2 = 0$$

where the coefficients $\{a_\alpha\}_{|\alpha|\le K}$ of $\mathcal{P}(z)$ are those that minimize the approximation error $\|g(z) - \mathcal{P}(z)\sqrt{\phi(z)}\|_2$.

Hermite Expansions: Rationale (2)

Let $h(z)$ be a density function. Because

$$\int h(z)\, dz = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(z)\, dz_1 \cdots dz_M = 1,$$

$\sqrt{h(z)}$ is in $L_2(-\infty, \infty)$ and can therefore be approximated by $\mathcal{P}(z)\sqrt{\phi(z)}$ as accurately as desired by taking K large enough.

This fact motivates using

$$h_K(z) = \frac{\mathcal{P}^2(z)\phi(z)}{\int \mathcal{P}^2(s)\phi(s)\, ds}$$

to approximate $h(z)$, where the division is to guarantee that $h_K(z)$ integrates to one.

The Main Idea

Take $h_K(z)$ as the parent density and use a location-scale transform

$$y = Rz + \mu$$

to generate a location-scale family of densities

$$f(y|\theta) = \frac{\left\{\mathcal{P}\left[R^{-1}(y-\mu)\right]\right\}^2 \phi\left[R^{-1}(y-\mu)\right]}{|\det(R)| \int \mathcal{P}^2(s)\phi(s)\, ds}$$

which can be estimated from data $\{y_t\}_{t=1}^n$ by quasi maximum likelihood

$$\hat{\theta} = \underset{\theta}{\text{argmax}} \prod_{t=1}^{n} f(y_t|\theta)$$

The density estimate is

$$\hat{f}(y) = f(y|\hat{\theta})$$

The consistency of the estimator was established by Gallant, A. Ronald, and Douglas W. Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica* 55, 363–390.

Some Remarks

$$f(y|\theta) = \frac{\left\{\mathcal{P}\left[R^{-1}(y-\mu)\right]\right\}^2 \phi\left[R^{-1}(y-\mu)\right]}{|\det(R)| \int \mathcal{P}^2(s)\phi(s)\,ds}$$

Note that $\mathcal{P}^2(z)/\int \mathcal{P}^2(s)\phi(s)\,ds$ is homogeneous of degree zero in the coefficients $\{a_\alpha\}_{\alpha=0}^K$. To achieve identification set $a_0 = 1$.

Note also that

$$N_M(y|\mu,\Sigma) = \frac{\phi\left[R^{-1}(y-\mu)\right]}{|\det(R)| \int \mathcal{P}^2(s)\phi(s)\,ds}$$

where $\Sigma = RR'$ so that

$$f(y|\theta) \propto \mathcal{P}^2\Big[R^{-1}(y-\mu)\Big]\, N(y|\mu,\Sigma)$$

Therefore, $f(y|\theta)$ with $K = 0$ is the normal density.

The constant of proportionality is $1/\int \mathcal{P}^2(s)\phi(s)\,ds$ above.

Topics

- Application
- Hermite Expansions
- SNP Density
  - Performance
  - Model Selection
- Extension to Time Series
  - VAR-type Location Function
  - ARCH- or GARCH-type Scale Function
  - Non-homogeneous Innovations
  - Miscellany

SNP Density: IID Data

Location-scale transform:

$$y = Rz + \mu \quad R \text{ upper triangular}$$

Density:

$$f(y|\theta) \propto \mathcal{P}^2\Big[R^{-1}(y-\mu)\Big]\, N(y|\mu,RR')$$

$$K = 0 \Rightarrow y \sim N_M(\mu, RR')$$

Example: $K = 2$, $M = 2$

$$R = \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}$$

$$\begin{aligned} \theta &= (a_{(0,0)}, a_{(1,0)}, a_{(0,1)}, \\ &\quad a_{(1,1)}, a_{(2,0)}, a_{(0,2)}, \\ &\quad \mu_1, \mu_2, R_{11}, R_{12}, R_{22})' \\ a_{(0,0)} &= 1 \end{aligned}$$

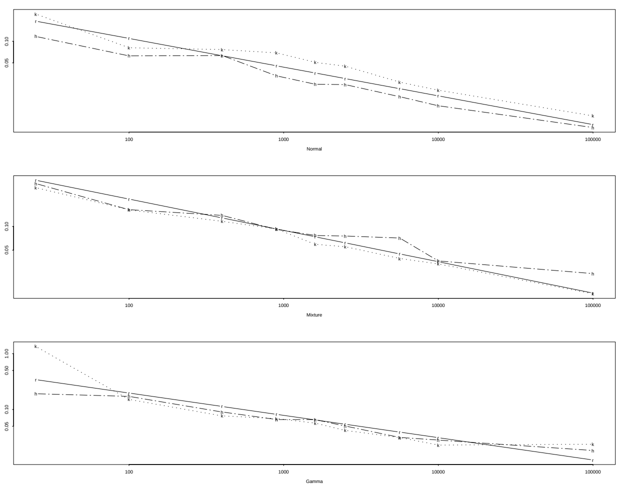
How well does SNP do?

Rate results:

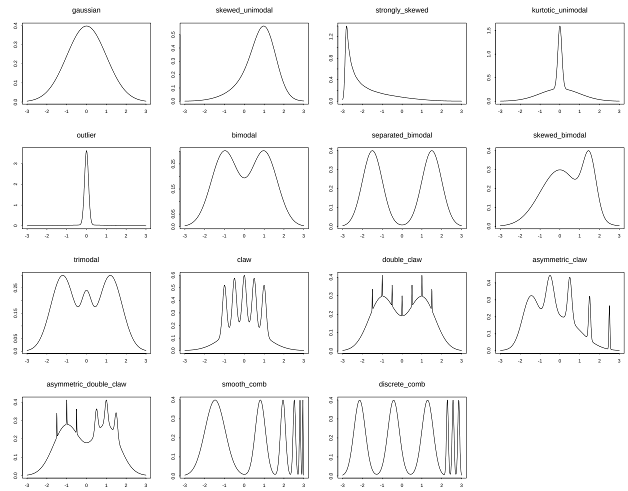
Fenton, Victor M., and A. Ronald Gallant (1996), "Convergence Rates of SNP Density Estimators," *Econometrica* 64, 719–727.

Qualitative comparison:

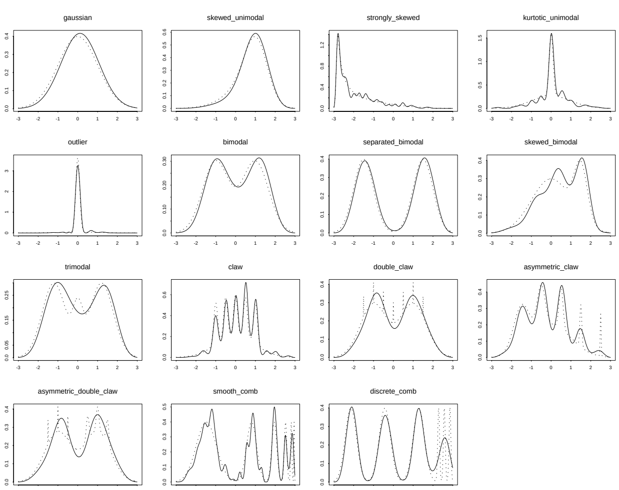
Fenton, Victor M., and A. Ronald Gallant (1996), "Qualitative and Asymptotic Performance of SNP Density Estimators," *Journal of Econometrics* 74, 77–118.



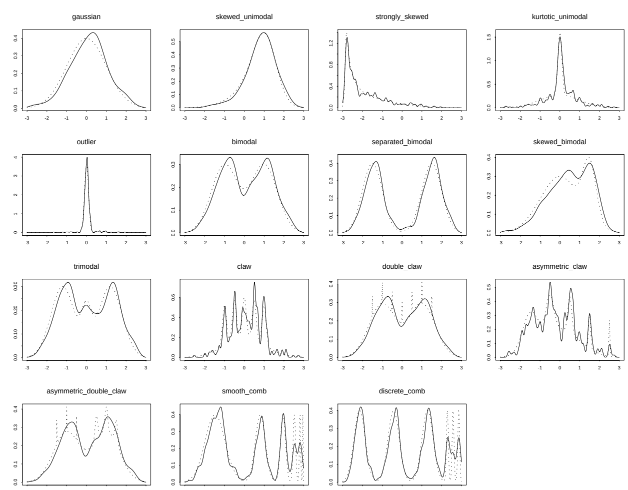
Theoretical, Kernel, and SNP $L_1(-\infty,\infty)$ Error Rates. The figure shows Monte Carlo estimates of $\int_{-\infty}^{\infty} |\hat{f}_n - f_o|\,dx$ based on ten repetitions. In each of the three panels, in the curve marked *k* $\hat{f}_n$ is a normal kernel estimator with bandwidth $Bn^{-1/5}$ where B is optimal for f_o with respect to $\mathcal{E}\int_{-\infty}^{\infty} (\hat{f}_n - f_o)^2\,dx$, in the curve marked *h* $\hat{f}_n$ is an SNP estimator with $p_n = n^{1/5}$ and μ and σ estimated, and in the curve marked *r* the theoretical rate $An^{-2/5}$ is plotted with A chosen to give the best least squares fit to the average of the curves marked *h* and *k*. In the panel marked *Normal*, the data are simulated from a standard normal distribution. In the panel marked *Mixture* the data are simulated from a mixture of a N(-3,1) with probability 0.3 and a N(4,1) with probability 0.7. In the panel marked *Gamma*, the data are simulated from a gamma distribution with parameters $\alpha = 7$ and $\beta = 1$.



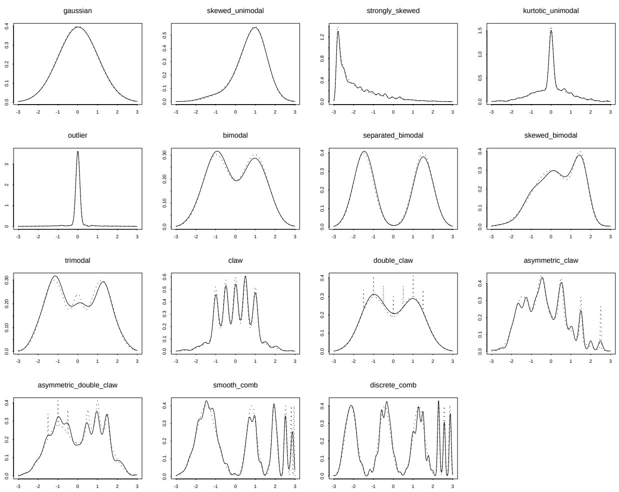
The Marron-Wand Test Suite. In each panel is plotted one of the fifteen densities proposed by Marron and Wand (1992) as a battery for use in evaluating nonparametric density estimators. The labels at the top of each panel are as in Marron and Wand.



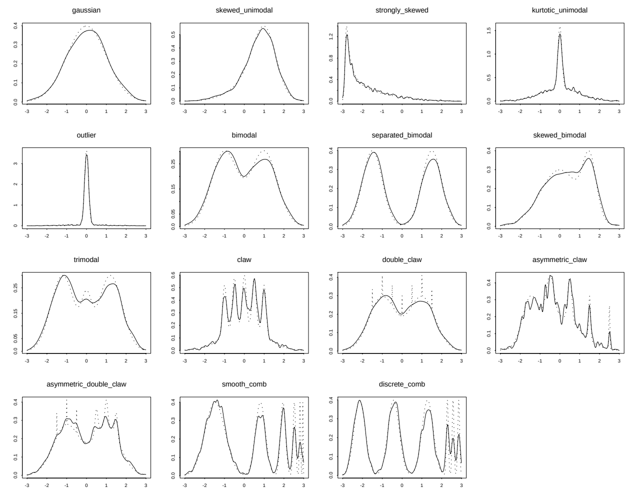
Plots of SNP Estimates, n=400, Marron-Wand Test Suite. In each panel the SNP estimate is plotted as a solid line and the density that was sampled is plotted as a dotted line. For each density, the degree p that gives the smallest value for $\sqrt{\int_{-3}^{3} (\hat{f}_p - f_o)^2\,dx}$ is selected.



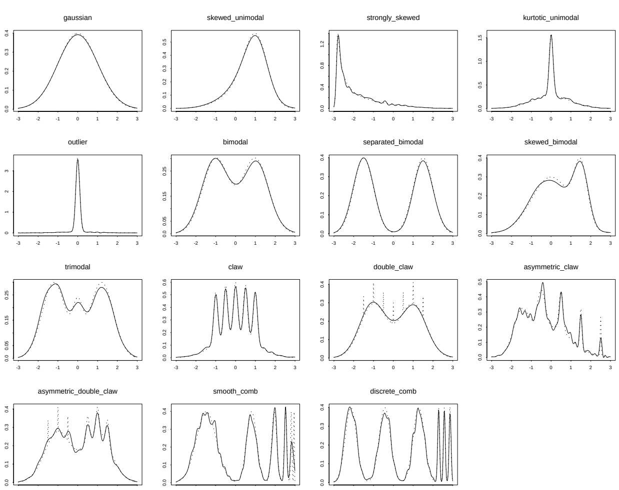
Plots of Kernel Estimates, n=400, Marron-Wand Test Suite. In each panel the kernel estimate is plotted as a solid line and the density that was sampled is plotted as a dotted line. Bandwidth selection is by least-squares cross-validation within the limits of 0.25 to 1.5 times Silverman's rule-of-thumb bandwidth.



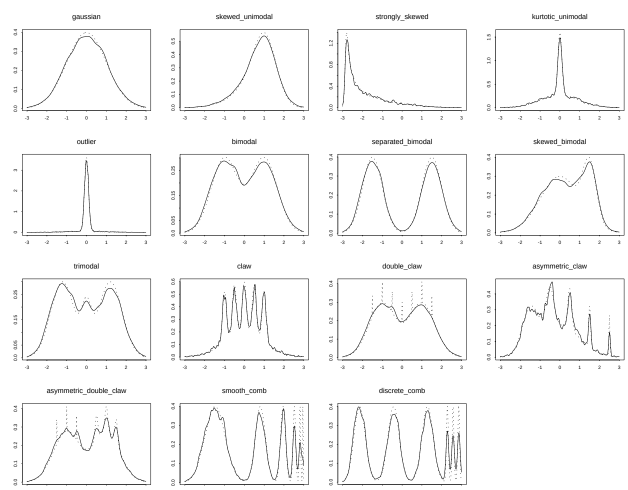
Plots of SNP Estimates, n=1600, Marron-Wand Test Suite. In each panel the SNP estimate is plotted as a solid line and the density that was sampled is plotted as a dotted line. For each density, the degree p that gives the smallest value for $\sqrt{\int_{-3}^{3} (\hat{f}_p - f_o)^2\,dx}$ is selected.



Plots of Kernel Estimates, n=1600, Marron-Wand Test Suite. In each panel the kernel estimate is plotted as a solid line and the density that was sampled is plotted as a dotted line. Bandwidth selection is by least-squares cross-validation within the limits of 0.25 to 1.5 times Silverman's rule-of-thumb bandwidth.



Plots of SNP Estimates, n=5625, Marron-Wand Test Suite. In each panel the SNP estimate is plotted as a solid line and the density that was sampled is plotted as a dotted line. For each density, the degree p that gives the smallest value for $\sqrt{\int_{-3}^{3} (\hat{f}_p - f_o)^2\,dx}$ is selected.



Plots of Kernel Estimates, n=5625, Marron-Wand Test Suite. In each panel the kernel estimate is plotted as a solid line and the density that was sampled is plotted as a dotted line. Bandwidth selection is by least-squares cross-validation within the limits of 0.25 to 1.5 times Silverman's rule-of-thumb bandwidth.

Choice of K

Coppejans, Mark, and A. Ronald Gallant (2000), "Cross Validated SNP Density Estimates," Working paper. Duke University, Durham NC. File: an.ps, an.pdf.

Bottom line: BIC seems to work well.

Estimation: Equivalent to maximum likelihood, but more stable numerically is to minimize the negative of the average log likelihood.

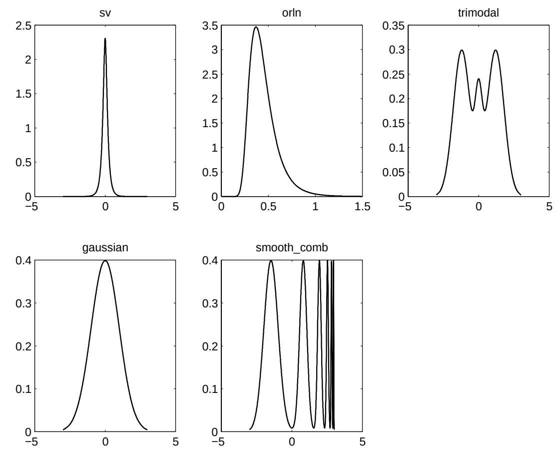
$$\hat{\theta} = \underset{\theta}{\text{argmin}}\ s_n(\theta)$$

$$s_n(\theta) = -\frac{1}{n}\sum_{t=1}^{n} \log\left[f(y_t|\theta)\right]$$

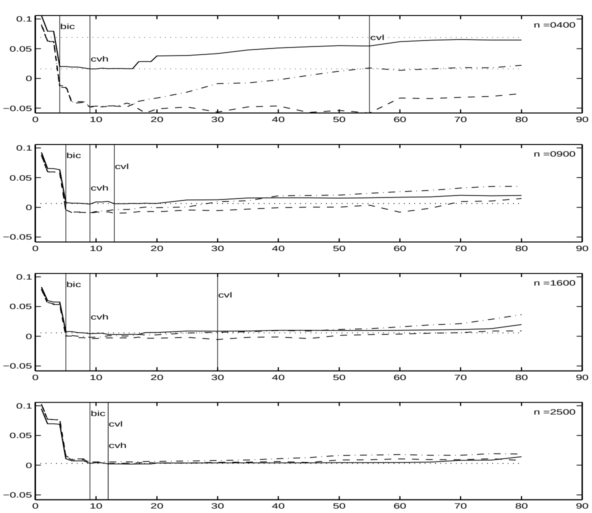
Schwarz criterion: Choose K to minimize

$$\text{BIC}(K) = s_n(\hat{\theta}) + \frac{p}{2n}\log(n)$$

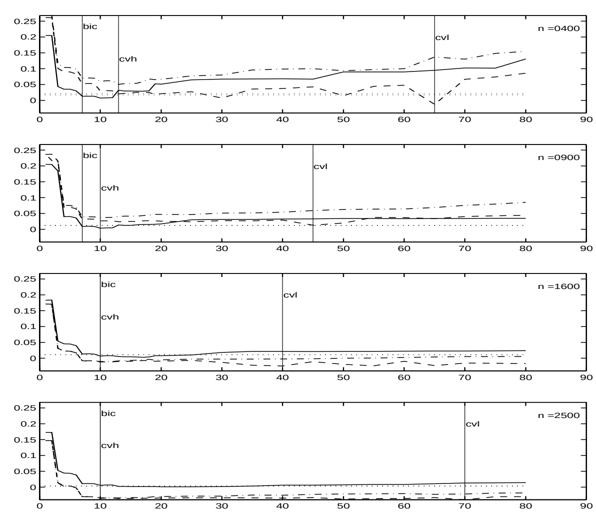
where p is the number of free parameters in θ.



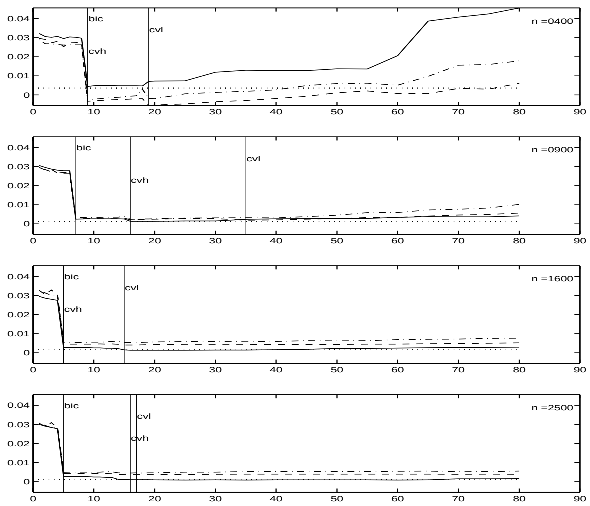
Densities considered. The plot labeled sv is the density of a scale mixture of normals with parameters chosen such that the density has mean 0, variance 1/4, and raw kurtosis 8; orln is the density of the second largest order statistic in a sample of size 100 from the log normal with location parameter -3 and scale parameter 1. The densities trimodal, gaussian, and smooth_comb are densities from the Marron-Wand test suite.



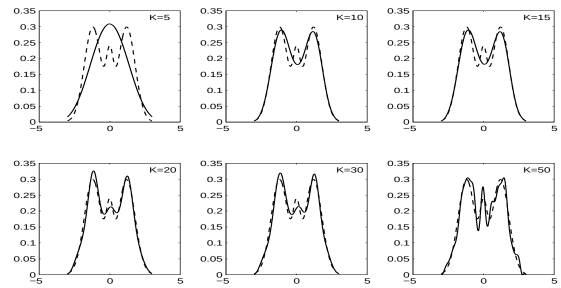
Scale Mixture of Normals. Plotted is the mean squared error (MSE) and its cross validated estimate (CV) for a realization of size n, as shown in each plot, from the density $p(y|\rho) = \int_{-\infty}^{\infty} n(y|\rho_1, e^{2u})\, n(u|\rho_2, \rho_3^2)\, du$ with ρ chosen so that the density has mean 0, variance 1/4, and raw kurtosis 8. Solid line is MSE, dashed line is its leave-one-out CV estimate (CVL), and dashed and dotted line is the average of ten, 10% hold-out-sample CV estimates (CVH). Upper dotted horizontal line is MSE achieved by a cross-validated kernel estimate and lower dotted line is best kernel MSE for this realization. Vertical lines indicate BIC, CVL, and CVH choices of K, as marked.



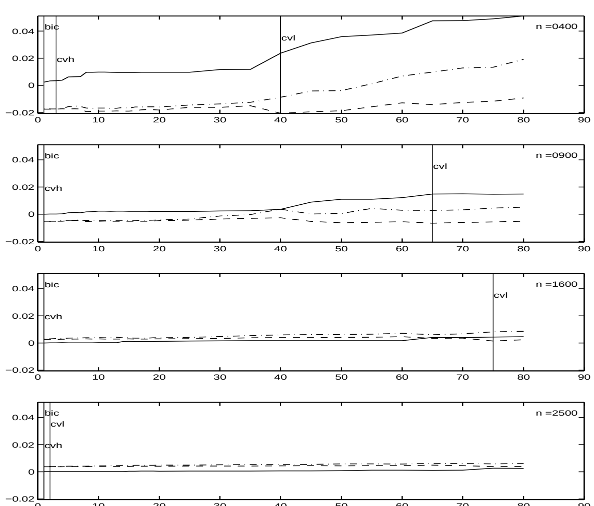
Second Largest Order Statistic of the Lognormal. Plotted is the mean squared error (MSE) and its cross validated estimate (CV) for a realization of size n, as shown in each plot, from the density $p(y|\rho) = \frac{N(N-1)}{y}\left[\Phi\left(\frac{\log y-\rho_2}{\rho_3}\right)\right]^{N-2}\left[1-\Phi\left(\frac{\log y-\rho_2}{\rho_3}\right)\right]\phi\left(\frac{\log y-\rho_2}{\rho_3}\right)$ where $y > 0$, ϕ and Φ denote the standard normal density and distribution functions, respectively, and $(N, \rho_2, \rho_3) = (100, -3, 1)$. Solid line is MSE, dashed line is its leave-one-out CV estimate (CVL), and dashed and dotted line is the average of ten, 10% hold-out-sample CV estimates (CVH). Upper dotted horizontal line is MSE achieved by a cross-validated kernel estimate and lower dotted line is best kernel MSE for this realization. Vertical lines indicate BIC, CVL, and CVH choices of K, as marked.



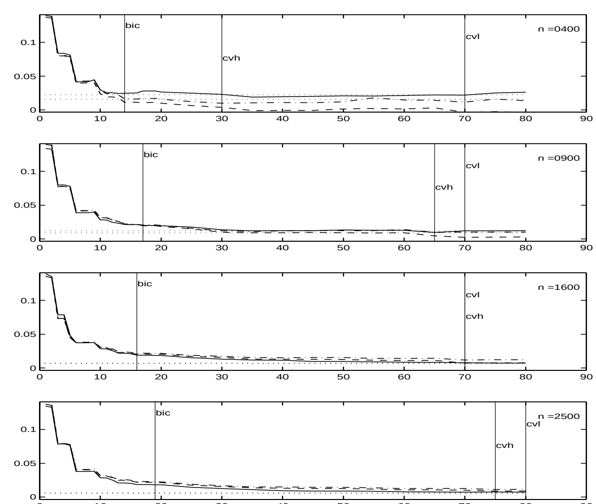
Trimodal. Plotted is the mean squared error (MSE) and its cross validated estimate (CV) for a realization of size n, as shown in each plot, from the trimodal density of the Marron-Wand test suite. Solid line is MSE, dashed line is its leave-one-out CV estimate (CVL), and dashed and dotted line is the average of ten, 10% hold-out-sample CV estimates (CVH). Upper dotted horizontal line is MSE achieved by a cross-validated kernel estimate and lower dotted line is best kernel MSE for this realization. Vertical lines indicate BIC, CVL, and CVH choices of K, as marked.



Trimodal. Plotted are SNP density estimates a realization of size 900 and values of K as shown in each plot, from the trimodal density of the Marron-Wand test suite. Solid line is the estimate, dashed line is true density.



Gaussian. Plotted is the mean squared error (MSE) and its cross validated estimate (CV) for a realization of size n, as shown in each plot, from the gaussian density of the Marron-Wand test suite. Solid line is MSE, dashed line is its leave-one-out CV estimate (CVL), and dashed and dotted line is the average of ten, 10% hold-out-sample CV estimates (CVH). Vertical lines indicate BIC, CVL, and CVH choices of K, as marked.



Smooth Comb. Plotted is the mean squared error (MSE) and its cross validated estimate (CV) for a realization of size n, as shown in each plot, from the smooth comb density of the Marron-Wand test suite. Solid line is MSE, dashed line is its leave-one-out CV estimate (CVL), and dashed and dotted line is the average of ten, 10% hold-out-sample CV estimates (CVH). Upper dotted horizontal line is MSE achieved by a cross-validated kernel estimate and lower dotted line is best kernel MSE for this realization. Vertical lines indicate BIC, CVL, and CVH choices of K, as marked.

Topics

- Application
- Hermite Expansions
- SNP Density
  - Performance
  - Model Selection
- Extension to Time Series
  - VAR-type Location Function
  - ARCH- or GARCH-type Scale Function
  - Non-homogeneous Innovations
  - Miscellany

SNP Density: IID Data

Location-scale transform:

$$y = Rz + \mu \quad R \text{ upper triangular}$$

Density:

$$f(y|\theta) \propto \mathcal{P}^2\Big[R^{-1}(y-\mu)\Big]\, N(y|\mu, RR')$$

$K_z = 0 \Rightarrow$ Gaussian, homogeneous z
$K_z > 0 \Rightarrow$ non-Gaussian, homogeneous z

Extension to Time Series

The idea is to modify the location and scale transforms of the SNP density for iid data to be functions of the past, which is the standard method of modifying a model for iid data for application to time series data. Lastly, the S-NP density itself is modified to accommodate non-homogeneous innovations. We shall proceed step-by-step.

SNP Transition Density for Time Series Data (1)

VAR location function:

$$\begin{aligned} y &= Rz + \mu_{x_{t-1}} && R \text{ upper triangular} \\ \mu_{x_{t-1}} &= b_0 + Bx_{t-1} && \text{linear in the past} \\ x_{t-1} &= (y_{t-L_u}, \dots, y_{t-1})' \end{aligned}$$

b_0 is $M \times 1$, B is $M \times L_u$,

Density:

$$f(y|\theta) \propto \mathcal{P}^2\big[R^{-1}(y-\mu_{x_{t-1}})\big]\, N(y|\mu_{x_{t-1}}, RR')$$

$K_z = 0 \Rightarrow$ Gaussian VAR, homogeneous z
$K_z > 0 \Rightarrow$ non-Gaussian VAR, homogeneous z

Example: $K = 2$, $M = 2$, $L_u = 1$

$$\begin{aligned} \theta = (&a_{(0,0)}, a_{(1,0)}, a_{(0,1)}, a_{(1,1)}, a_{(2,0)}, a_{(0,2)}, \\ &b_{01}, b_{02}, B_{11}, B_{21}, B_{12}, B_{22}, \\ &R_{11}, R_{12}, R_{22})' \end{aligned}$$

SNP Transition Density for Time Series Data (2)

ARCH-type scale function:

$$\begin{aligned} y &= R_{x_{t-1}} z + \mu_{x_{t-1}} && R_{x_{t-1}} \text{ upper triangular} \\ \mu_{x_{t-1}} &= b_0 + Bx_{t-1} \\ x_{t-1} &= (y_{t-L_u}, \dots, y_{t-1})' \\ \text{vech}(R_{x_{t-1}}) &= \rho_0 + \sum_{i=1}^{L_r} P_{(i)} |y_{t-1-L_r+i} - \mu_{x_{t-2-Lr+i}}| \\ S_{t-1} &= (y_{t-L_u-L_r}, \dots, y_{t-1})' && \text{state vector} \end{aligned}$$

vech(R) means columnwise storage of the nonzero elements of an upper triangular matrix, S_{t-1} is the information required to move the system forward one step, ρ_0 is $M(M+1)/2 \times 1$, $P = \big[P_{(1)}|\cdots|P_{(L_r)}\big]$ is $M(M+1)/2 \times L_r$,

Density:

$$f(y|\theta) \propto \mathcal{P}^2\big[R^{-1}_{x_{t-1}}(y-\mu_{x_{t-1}})\big]\, N(y|\mu_{x_{t-1}}, R_{x_{t-1}}R'_{x_{t-1}})$$

$K_z = 0 \Rightarrow$ Gaussian ARCH, homogeneous z
$K_z > 0 \Rightarrow$ non-Gaussian ARCH, homogeneous z

Example: $K = 2$, $M = 2$, $L_u = 1$, $L_r = 2$

$$\text{vech}(R) = (R_{11}, R_{21}, R_{22})'$$

$$\begin{aligned} \theta = (&a_{(0,0)}, a_{(1,0)}, a_{(0,1)}, a_{(1,1)}, a_{(2,0)}, a_{(0,2)}, \\ &b_{01}, b_{02}, B_{11}, B_{21}, B_{12}, B_{22}, \\ &\rho_{01}, \rho_{02}, \rho_{03}, P_{11}, P_{21}, P_{31}, P_{12}, P_{22}, P_{32})' \end{aligned}$$

SNP Transition Density for Time Series Data (3)

GARCH-type scale function:

$$
\begin{aligned}
y &= R_{x_{t-1}} z + \mu_{x_{t-1}} \\
\mu_{x_{t-1}} &= b_0 + B x_{t-1} \\
\text{vech} R_{x_{t-1}} &= \rho_0 + \sum_{i=1}^{L_r} P_{(i)} |y_{t-1-L_r+i} - \mu_{x_{t-2-Lr+i}}| \\
&\quad + \sum_{i=1}^{L_g} \text{diag}(G_{(i)}) R_{x_{t-2-L_g+i}} \\
S_{t-1} &= (y_{t-L_u-L_r}, \ldots, y_{t-1}, \text{vech} R_{x_{t-2-L_g}}, \ldots, \text{vech} R_{x_{t-2}})'
\end{aligned}
$$

ρ_0 is $M(M+1)/2 \times 1$, $P = \left[P_{(1)} | \cdots | P_{(L_r)}\right]$ is $M(M+1)/2 \times L_r$, $G = \left[G_{(1)} | \cdots | G_{(L_g)}\right]$ is $M(M+1)/2 \times L_g$

Density:

$$
f(y|\theta) \propto \mathcal{P}^2 \left[R_{x_{t-1}}^{-1}(y - \mu_{x_{t-1}})\right] N(y | \mu_{x_{t-1}}, R_{x_{t-1}} R'_{x_{t-1}})
$$

$K_z = 0 \Rightarrow$ Gaussian GARCH, homogeneous z
$K_z > 0 \Rightarrow$ non-Gaussian GARCH, homogeneous z

Example: $K = 2$, $M = 2$, $L_u = 1$, $L_r = 1$, $L_g = 1$

$$
\begin{aligned}
\text{vech}(R) &= (R_{11}, R_{21}, R_{22})' \\
\theta &= (a_{(0,0)}, a_{(1,0)}, a_{(0,1)}, a_{(1,1)}, a_{(2,0)}, a_{(0,2)}, \\
&\quad b_{01}, b_{02}, B_{11}, B_{21}, B_{12}, B_{22}, \\
&\quad \rho_{01}, \rho_{02}, \rho_{03}, P_{11}, P_{21}, P_{31}, \\
&\quad G_{11}, G_{21}, G_{31},)'
\end{aligned}
$$

SNP for Non-homogeneous Innovations (1)

The Past:

$$
x = (x_{t-L_p}, \ldots, x_{t-1})'
$$

Polynomial Part:

Non-homogeneous innovations are accommodated by letting the polynomial part of the SNP model

$$
\mathcal{P}(z) = \sum_{|\alpha|=0}^{K_z} a_\alpha z^\alpha
$$

have coefficients a_α that are polynomials in x

$$
a_\alpha(x) = \sum_{|\beta|=0}^{K_x} a_{\alpha\beta} x^\beta
$$

It is denoted by $\mathcal{P}(z, x)$.

SNP for Non-homogeneous Innovations (2)

The SNP density for non-homogeneous innovations is Hermite polynomial in z whose coefficients are polynomials in x

Polynomial Part:

$$
\mathcal{P}(z, x) = \sum_{|\alpha|=0}^{K_z} \sum_{|\beta|=0}^{K_x} a_{\alpha\beta} x^\beta z^\alpha
$$

SNP density:

$$
h_K(z|x) = \frac{\mathcal{P}(z, x)\, \phi(z)}{\int \mathcal{P}(s, x)\, \phi(s)\, ds}
$$

Remarks:
Here $K = (K_z, K_x)$. $\mathcal{P}(z, x)$ is a polynomial in (z, x) of degree $|K| = K_z + K_x$ of a type known as a rectangular expansion. A radial expansion of degree K has the form

$$
\sum_{|\gamma|=0}^{|K|} a_\gamma (z, x)^\gamma = \sum_{|\alpha|+|\beta|=0}^{K_z+K_x} a_{\alpha\beta} x^\beta z^\alpha
$$

SNP Transition Density for Time Series Data (4)

Non-homogeneous innovations:

$$
\begin{aligned}
\mu_{x_{t-1}} &= b_0 + B x_{t-1} \\
\text{vech} R_{x_{t-1}} &= \rho_0 + \sum_{i=1}^{L_r} P_{(i)} |y_{t-1-L_r+i} - \mu_{x_{t-2-Lr+i}}| \\
&\quad + \sum_{i=1}^{L_g} \text{diag}(G_{(i)}) R_{x_{t-2-L_g+i}} \\
S_{t-1} &= (y_{t-\max(L_u+L_r, L_p)}, \ldots, y_{t-1}, \text{vech} R_{x_{t-2-L_g}}, \ldots, \text{vech} R_{x_{t-2}})'
\end{aligned}
$$

Density:

$$
\begin{aligned}
f(y|\theta) &\propto \mathcal{P}^2 \left[R_{x_{t-1}}^{-1}(y - \mu_{x_{t-1}}, x_{t-1})\right] N(y | \mu_{x_{t-1}}, R_{x_{t-1}} R'_{x_{t-1}}) \\
x &= (x_{t-L_p}, \ldots, x_{t-1})'
\end{aligned}
$$

Example: $K = 2$, $M = 2$, $L_u = 1$, $L_r = 1$, $L_g = 1$, $L_p = 1$

$$
\begin{aligned}
\text{vech}(R) &= (R_{11}, R_{21}, R_{22})' \\
\theta &= (a_{(0,0),(0,0)}, a_{(1,0),(0,0)}, a_{(0,1),(0,0)}, \\
&\quad a_{(1,1),(0,0)}, a_{(2,0),(0,0)}, a_{(0,2),(0,0)}, \\
&\quad a_{(0,0),(1,0)}, a_{(1,0),(1,0)}, a_{(0,1),(1,0)}, \\
&\quad a_{(1,1),(1,0)}, a_{(2,0),(1,0)}, a_{(0,2),(1,0)}, \\
&\quad a_{(0,0),(0,1)}, a_{(1,0),(0,1)}, a_{(0,1),(0,1)}, \\
&\quad a_{(1,1),(0,1)}, a_{(2,0),(0,1)}, a_{(0,2),(0,1)}, \\
&\quad b_{01}, b_{02}, B_{11}, B_{21}, B_{12}, B_{22}, \\
&\quad \rho_{01}, \rho_{02}, \rho_{03}, P_{11}, P_{21}, P_{31}, \\
&\quad G_{11}, G_{21}, G_{31},)'
\end{aligned}
$$

Consistency

If the parameters of $f(y|x,\theta)$ are estimated by quasi maximum likelihood, viz.

$$\hat{\theta}_n = \underset{\theta}{\text{argmin}}\; s_n(\theta)$$

$$s_n(\theta) = -\frac{1}{n}\sum_{t=1}^{n} \log f(y_t|x_{t-1},\theta),$$

and $K = (K_z, K_x)$ grows with sample size, then the estimator

$$\hat{f}_n(y|x) = f(y|x,\hat{\theta}_n)$$

converges almost surely to the true transition density $f(y|x)$ in Sobolev norm as sample size increases. Moreover, K can depend on the data.

Reference:
Gallant, A. Ronald, and Douglas W. Nychka (1987), "Seminonparametric Maximum Likelihood Estimation," *Econometrica* 55, 363–390.

Restrictions Implied by Settings of the Tuning Parameters

Parameter setting	Characterization of $\{y_t\}$
$L_u\!=\!0,\, L_g\!=\!0,\, L_r\!=\!0,\, L_p\geq 0,\, K_z\!=\!0,\, K_x\!=\!0$	iid Gaussian
$L_u>0,\, L_g\!=\!0,\, L_r\!=\!0,\, L_p\geq 0,\, K_z\!=\!0,\, K_x\!=\!0$	Gaussian VAR
$L_u>0,\, L_g\!=\!0,\, L_r\!=\!0,\, L_p\geq 0,\, K_z>0,\, K_x\!=\!0$	semiparametric VAR
$L_u\geq 0,\, L_g\!=\!0,\, L_r>0,\, L_p\geq 0,\, K_z\!=\!0,\, K_x\!=\!0$	Gaussian ARCH
$L_u\geq 0,\, L_g\!=\!0,\, L_r>0,\, L_p\geq 0,\, K_z>0,\, K_x\!=\!0$	semiparametric ARCH
$L_u\geq 0,\, L_g>0,\, L_r>0,\, L_p\geq 0,\, K_z\!=\!0,\, K_x\!=\!0$	Gaussian GARCH
$L_u\geq 0,\, L_g>0,\, L_r>0,\, L_p\geq 0,\, K_z>0,\, K_x\!=\!0$	semiparametric GARCH
$L_u\geq 0,\, L_g\geq 0,\, L_r\geq 0,\, L_p>0,\, K_z>0,\, K_x>0$	nonlinear nonparametric

Standard Data Transformation

Sample mean and variance

$$\bar{y} = \frac{1}{n}\sum_{t=1}^{n} \tilde{y}_t$$

$$S = \frac{1}{n}\sum_{t=1}^{n} (\tilde{y}_t - \bar{y})(\tilde{y}_t - \bar{y})'$$

$\tilde{y}_t$ denotes the raw data

Apply the methods above to

$$y_t = S^{-1/2}(\tilde{y}_t - \bar{y})$$

where $S^{-1/2}$ denotes the Cholesky factor of the inverse of S.

Suggest taking S diagonal.

Problem

If the true density $f(y|x)$ is heavy tailed, then x_{t-1} will contain extreme observations which have a strong and undesirable influence on estimates when $L_r > 0$.

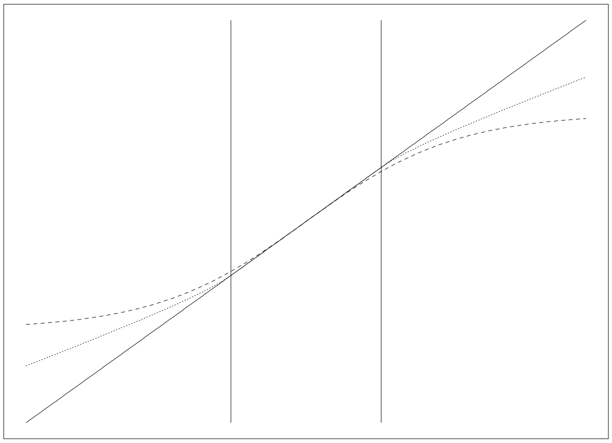
Cure

Replace each component of x by its log spline transform

$$\hat{x}_i = \begin{cases} (1/2)[x_i - \sigma_{\text{tr}} - \log(1 - x_i - \sigma_{\text{tr}})] & x_i < -\sigma_{\text{tr}} \\ x_i & -\sigma_{\text{tr}} < x_i < \sigma_{\text{tr}} \\ (1/2)[x_i + \sigma_{\text{tr}} + \log(1 - x_i - \sigma_{\text{tr}})] & \sigma_{\text{tr}} < x_i. \end{cases}$$

The consistency result is not affected by this transform.

A logistic transform can also be used for this purpose. It is a more aggressive solution to the problem but has poor properties with persistent data such as interest rates. It does work well with strongly mean reverting data such as stock returns.



The logistic and logarithmic spline transforms. The dashed line shows the logistic transform. The dotted line shows the logarithmic spline transformation. The solid line is a 45 degree line, which represents no transformation. The two vertical lines are at $x = -\sigma_{\text{tr}}$ and $x = \sigma_{\text{tr}}$, respectively.

Order in Which Transformations are Applied

$\tilde{y}_t$	$\rightarrow$	y_t	$\rightarrow$	x_{t-1}	$\rightarrow$	$\hat{x}_{t-1}$	$\rightarrow$	$\mu_x,\ R_x$
ra w data		centered, scaled data		lagged data		logistic data		location, scale

Problem

The SNP density $f(y|x,\theta)$, or any density for that matter, can become very small at extreme values of y or x. This can be a nuisance in some applications such as EMM that require taking logarithms of the SNP density when evaluated over data generated by structural models at trial values of structural parameters. The simulated data can be absurd relative to the actual data, but the computations must be performed nonetheless.

Cure

Replace the SNP density throughout all the discussion above by

$$f^*(y|x,\theta) = \frac{\left\{\mathcal{P}^2\left[\,R_x^{-1}(y-\mu_x),\ x\,\right] + \epsilon_0\right\} n_M(y|\mu_x, R_x R_x')}{\int [\mathcal{P}(s,x)]^2 \phi(s)\, ds + \epsilon_0}$$

where ϵ_0 is some small value such as 10^{-3}. The consistency result is not affected.

Problem

Large M or large L_p can generate a large number of interactions for even modest settings of K_z or K_x .

Cure

Filters with parameters I_z and I_x that automatically delete high order interactions from $\mathcal{P}(z|x)$. $I_z = 0$ means no interactions are suppressed, $I_z = 1$ means the highest order interactions are suppressed, namely those of degree K_z. In general, a positive I_z means all interactions of order larger than $K_z - I_z$ are suppressed; similarly for $K_x - I_x$.

Example:
$K_z = 4$ and $I_z = 2$ imply all interaction terms of degree 3 and higher are deleted. For $M = 2$, These would be

$$\alpha = (3,1), (1,3), (2,2), (2,1), (1,2)$$

Tuning Parameters

Major:

$$(L_u, L_g, L_r, L_p, K_z, I_z, K_x, I_x)$$

Recommended Minor:

Diagonal S, diagonal GARCH, logarithmic s-pline transform with $\sigma_{\text{tr}} = 2$, $\epsilon_0 = 10^{-3}$.

Availability

Fortran code and a User's Guide are available by anonymous ftp at host ftp.econ.duke.edu in directory pub/arg/snp or click on "browse ftp site" at http:/www.unc.edu/~arg.