Regression Kink with Multiple Unknown Thresholds

Bruce E. Hansen

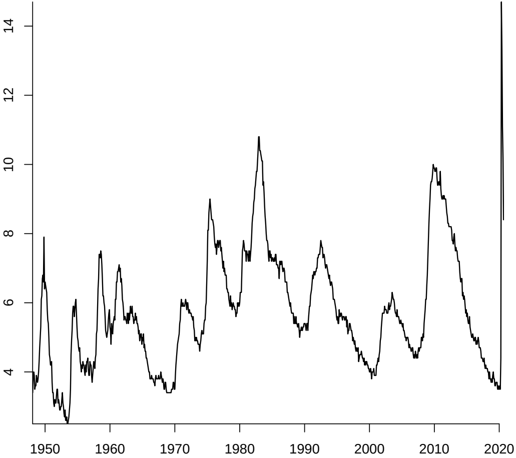
University of Wisconsin-Madison

September 2020

Society for NonLinear Dynamics and Econometrics

- Roots in Complex Dynamics
- Focus on Nonlinear Time Series
- My first SNDE conference: 1991
- My first SNDE keynote: 1996
- Title: "Estimation of Thresholds in Dynamic Models"
- Today's paper:
  - Use regression kinks to capture sharp autoregressive nonlinearity

Empirical Illustration: U.S. Unemployment Rate



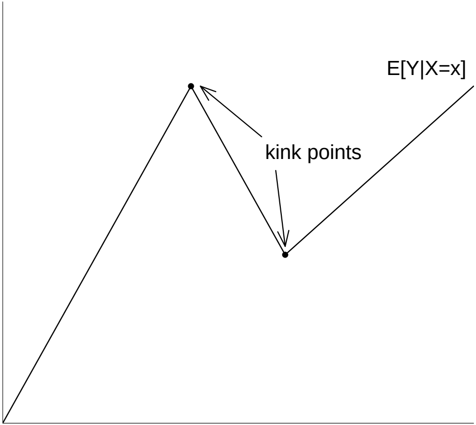
Monthly Unemployment Rate

- Available: 1948-2020
- Asymmetry between expansions and recessions
  - Recessions (increases in unemployment rate) sharp
  - Expansions (decreases in unemployment rate) gradual
- Transitions sharp, especially from recession to expansion

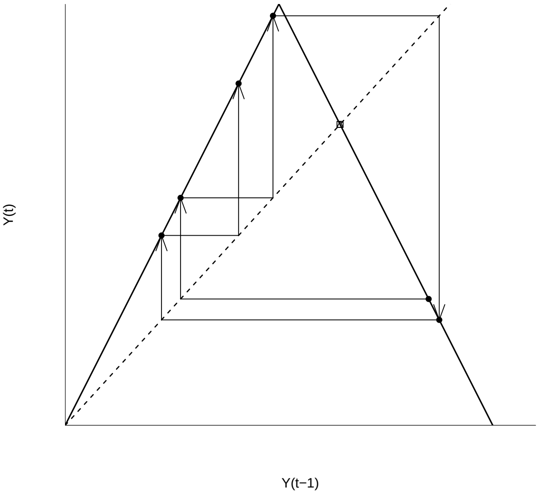
Regression Kink Model

- Piecewise Linear Spline with free knots
- Nonparametric approximation
- Additively separable, continuous in regressors
- Allows sharp nonlinearities
  - This allows for strong nonlinear behavior, including chaotic (self-propagating) dynamics

Kink Regression Function



Tent Map: Chaotic Dynamics



Regression Kink Model

- Observables: $(Y_t, X_{1t}, ..., X_{Kt})$
- Regressors X_{kt} may include lags of Y_t
- Regression Model

$$Y_t = \mu + \sum_{k=1}^{K} \left[\beta_k X_{kt} + \sum_{j=1}^{m_k} \beta_{kj} \left(X_{kt} - \gamma_{kj} \right)_+ \right] + e_t$$

$$\mathbb{E}\left[e_t \mid \mathcal{F}_{t-1} \right] = 0$$

where $(a)_+ = \max[a, 0]$.

Properties of Model

- Additively separable in K regressors X_{kt}
- Continuous linear spline with m_k kinks (knots)
- γ_{kj} are kink points
- β_k are slopes for "leftmost" section of the function
- β_{kj} are changes in the slopes at each kink point
- Function simplifies to linear if $m_0 = 0$
- A regressor can enter linearly, with one kink, or multiple kinks

Regression Function

- Continuous in all regressors
- First derivative is step function, discontinuous at kinks
- Additively separable: No modeling of interaction effects
- $\sum_{k=1}^{K} m_k$ kink parameters
- $\sum_{k=1}^{K} (1 + m_k)$ slope coefficients

Parameterization

- Slope coefficients and kink parameters (β_k, β_{kj}, γ_{kj}) to be estimated.
  - Regression is linear in β_k and β_{kj}
  - Nonlinear in γ_{kj}
  - Nonlinear least squares (NLLS)
- Number of kinds m_k model specification choice.
  - Similar to the number and choice of variables, number of autoregressive lags.
  - Model selection decision
  - Aided by cross-validation

Parameterization

$$
\begin{aligned}
Y_t &= \mu + \sum_{k=1}^{K} \left[\beta_k X_{kt} + \sum_{j=1}^{m_k} \beta_{kj} \left(X_{kt} - \gamma_{kj} \right)_+ \right] + e_t \\
&= \beta' X_t(\gamma) + e_t
\end{aligned}
$$

- β is vector of all slope coefficients
- γ is vector of all kink parameters
- $X_t(\gamma)$ is vector containing X_{kt} and $\left(X_{kt} - \gamma_{kj}\right)_+$
- Nonlinear regression in β and γ

NLLS Estimation

Average squared error function:

$$S_n(\beta, \gamma) = \frac{1}{n} \sum_{t=1}^{n} \left(Y_t - \beta' X_t(\gamma)\right)^2 .$$

NLLS estimator:

$$(\widehat{\beta}, \widehat{\gamma}) = \underset{\beta,\gamma}{\operatorname{argmin}} \, S_n(\beta, \gamma) .$$

Concentration Approach

Given γ, model is linear in β, so latter estimated by least squares

$$\widehat{\beta}(\gamma) = \left(\sum_{t=1}^{n} X_t(\gamma)X_t(\gamma)'\right)^{-1}\left(\sum_{t=1}^{n} X_t(\gamma)Y_t\right)$$

The residuals are $\widehat{e}_t(\gamma) = Y_t - \widehat{\beta}(\gamma)'X_t(\gamma)$.

The concentrated average squared error function is

$$S_n^*(\gamma) = S_n(\widehat{\beta}(\gamma), \gamma) = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \widehat{\beta}(\gamma)'X_t(\gamma)\right)^2$$

Estimation

The minimizer is

$$\widehat{\gamma} = \underset{\gamma}{\operatorname{argmin}}\, S_n^*(\gamma)$$

By concentrated minimization this is the NLLS estimator for γ.

- $S_n^*(\gamma)$ is concentrated NLLS criterion
- β has been profiled out
- $\widehat{\gamma}$ is the global NLLS estimator
- The NLLS slope estimators are $\widehat{\beta} = \widehat{\beta}(\widehat{\gamma})$
- They are obtained by least squares of Y_t on $X_t(\widehat{\gamma})$

Kink Estimation

- Nonlinear minimization of $S_n^*(\gamma)$
- Criterion is continuous in γ, piecewise linear
- First derivative is a step function
- Kinks are at data points
- Similar to quantile regression criterion

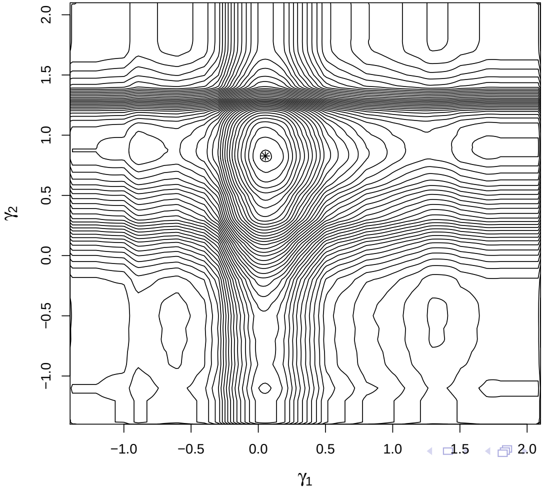
Minimization

- Criterion function can be multi-modal
- If there is a single kink, use golden section search
- For multiple kinks, consider derivative-free methods
  - Nelder-Mead
  - But desirable to impose boundary constraints on γ
- In application (coded in R with optim)
  - Nelder-Mead worked in low dimensional cases
  - Nelder-Mean in R does not allow boundary constraints
  - L-BFGS-B worked well, allows boundary constraints
  - Used L-BFGS-B for reported results

Concentrated NLLS Criterion – Contour plot

From empirical application, 2-kink case

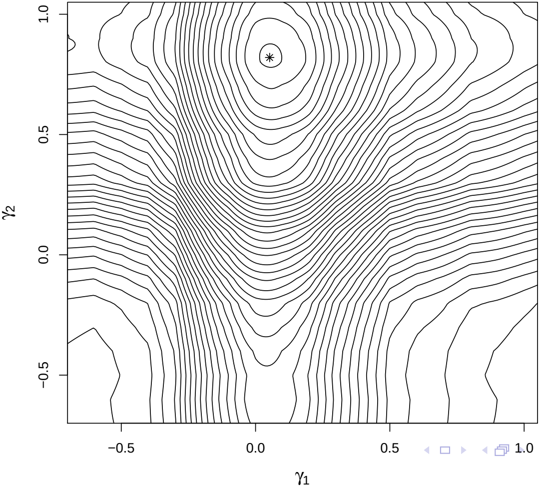
* marks global minimum



Concentrated NLLS Criterion – Contour plot

Previous plot over entire support of variables

This plot is from 0.03 to 0.97 quantiles of marginal distributions



Pseudo-True Coefficients

Squared loss:

$$L(\beta, \gamma) = \mathbb{E}\left[\left(Y_t - \beta' X_t(\gamma)\right)^2\right]$$

Pseudo-true values:

$$(\beta_0, \gamma_0) = \underset{\beta, \gamma}{\operatorname{argmin}} \, L(\beta, \gamma)$$

- If the model is correctly specified then (β_0, γ_0) equal the true regression coefficients.
- Otherwise they are the best-fitting model parameters.

Assumption 1

For some $r > 1$

- $(Y_t, X_{1t}, ..., X_{Kt})$ is strictly stationary, ergodic, and absolutely regular with mixing coefficients $\eta(m) = O(m^{-A})$ for some $A > r/(r-1)$
- $\mathbb{E}\,|Y_t|^{4r} < \infty$ and $\mathbb{E}\,|X_{kt}|^{4r} < \infty$ for $k = 1, ..., K$
- $\inf_{\gamma \in \Gamma} \det Q(\gamma) > 0$ where $Q(\gamma) = \mathbb{E}\left[X_t(\gamma) X_t(\gamma)'\right]$
- X_{jt} has a density function $f_j(x)$ satisfying $f(x) \leq \overline{f} < \infty$
- $\theta = (\beta, \gamma) \in \Gamma$ where Γ is compact.
- The squared loss $L(\beta, \gamma)$ has a unique mimimum.

Theorem 1

$$\sqrt{n}\left(\widehat{\theta}-\theta_0\right) \underset{d}{\longrightarrow} \mathrm{N}\left(0, V\right)$$

where

$$V = Q^{-1} S Q^{-1}$$
$$Q = \mathbb{E}\left[H_t H_t'\right]$$
$$S = \sum_{j=-\infty}^{\infty} \mathbb{E}\left[H_t H_{t+j}' e_t e_{t+j}\right]$$

The vector H_t are the derivatives of the regression function with respect to the parameters.

Asymptotic Distribution

- Coefficient estimates are asymptotically normal
- Conventional convergence rate
- Does not require correct specification – model can be an approximation
- Asymptotic variance takes a HAC/HAR form.
  - Simplifies under correct specification
- Asymptotic distribution obtained by empirical process argument

Covariance Matrix Estimation

For simplicity, focus on non-HAR estimator

$$\widehat{V} = \widehat{Q}^{-1}\widehat{S}\widehat{Q}^{-1}$$

$$\widehat{Q} = \frac{1}{n}\sum_{t=1}^{n}\widehat{H}_t\widehat{H}_t'$$

$$\widehat{S} = \frac{1}{n}\sum_{t=1}^{n}\widehat{H}_t\widehat{H}_t'\widehat{e}_t^2$$

The vector $\widehat{H}_t$ is a plug-in estimator of the derivatives of the regression function.

Model Selection

- Selection choices include the list of regressors (AR order) & the number of kinks
- Old-school approach: Hypothesis testing
- Modern approach: cross-validation (CV)
- The CV criterion is

$$\mathrm{CV} = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \widehat{\beta}'_{-t}X_t(\widehat{\gamma}_{-t})\right)^2$$

where the estimators $\left(\widehat{\beta}_{-t}, \widehat{\gamma}_{-t}\right)$ are calculated omitting observation t.

CV Computation

- In linear regression there is a simple and fast computation
- Nonlinear regression requires n separate estimations
- I do not know a computational shortcut
- For application I used nonlinear optimizer, with starting value equal to NLLS estimator
- Computation cost – illustrated by application ($n = 851$)
- Application has $n = 851$.
  - CV for 2 kinks took 5 seconds
  - CV for 4 kinks took 30 seconds
  - CV for 5 kinks one minute

Inference on Regression function

- In nonlinear and nonparametric regression it is useful to display estimated regression function with confidence bands
- Our regression function is $g(\theta) = \beta' x(\gamma)$
- $g(\theta)$ is continuous but not continuously differentiable in θ
- The derivative is kinked at the kink points
- Non-standard distribution theory
- $g(\theta)$ is **directionally differentiable**

Theorem 2

$$\sqrt{n}\left(g\left(\widehat{\theta}\right)-g\left(\theta_0\right)\right)\underset{d}{\longrightarrow} g_\theta\left(Z\right)$$

- $Z \sim \mathrm{N}\left(0, V\right)$
- $g_\theta\left(h\right)$ is the directional derivative of $g\left(\theta\right)$
- Shapiro (1990), Hong and Li (2018), Fang and Santos (2019)

Algorithm: Wild bootstrap/Numerical Delta Method

- Generate n i.i.d. Rademacher draws u_t
- Set $e_t^* = \widehat{e}_t u_t$
- Set $Y_t^* = \widehat{\beta}' X_t(\widehat{\gamma}) + e_t^*$
- Use $(Y_t^*, X_{1t}, ..., X_{Kt})$ to estimate the model
- Set $x(\widehat{\gamma})$
- Set $r^* = x(\widehat{\gamma})' \left(\widehat{\beta}^* - \widehat{\beta}\right) + g(\widehat{\beta}, \widehat{\gamma}^*) - g(\widehat{\beta}, \widehat{\gamma})$
- Repeat B times, so as to obtain a sample of simulated estimates r^*
- Calculate $q_{1-\alpha}^*$, the $(1-\alpha)^{th}$ quantile of the simulated $|r^*|$
- Confidence Interval: $[g(\widehat{\theta}) - q_{1-\alpha}^*, g(\widehat{\theta}) + q_{1-\alpha}^*]$

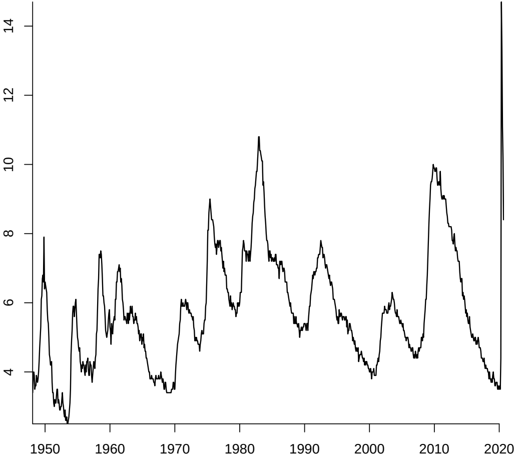
Bootstrap Computation

- Requires B numerical optimizations
- Set initial value for optimization equal to sample estimate
- In application with $B = 1000$, two kinks took about 15 seconds
- Four kinks took 30 seconds, five kinks took two minutes

Unemployment Rate Application

- Monthly data, available 1948m1-2020m8
- Estimate on 1949-2019
  - Reserve 1948 for initial conditions
  - Exclude 2020 due to unusual values
- Autoregressive Framework

U.S. Unemployment Rate



Linear Baseline

$$\Delta Y_t = \beta_1 Y_{t-1} + \beta_2 \Delta Y_{t-1} + \beta_3 \Delta_3 Y_{t-1} + \beta_4 \Delta_3 Y_{t-4} + \mu + e_t$$

- Simple, captures main features
- Constrained AR(7)
- $\Delta_3 Y_{t-1}$ and $\Delta_3 Y_{t-4}$ are quarterly changes
- $\beta_1 = 0$ implies unit root
- $\beta_1 < 0$ allows stationarity

Linear Estimates

$$\Delta Y_t = \underset{(.004)}{-.017}\; Y_{t-1} \underset{(.09)}{-.21}\; \Delta Y_{t-1} \underset{(.03)}{+.20}\; \Delta_3 Y_{t-1} \underset{(.03)}{+.07}\; \Delta_3 Y_{t-4}$$

- β_1 small but unit root rejected
- Coefficients on quarterly changes indicate positive serial correlation for changes

Regression Kink Model

- Four regressors: Y_{t-1}, ΔY_{t-1}, $\Delta_3 Y_{t-1}$, $\Delta_3 Y_{t-4}$
- Allow kinks in each
- Allow multiple kinks per regressor
- Select model by minimizing cross-validation criterion

Cross Validation

Four variables: Y_{t-1}, ΔY_{t-1}, $\Delta_3 Y_{t-1}$, $\Delta_3 Y_{t-4}$

	0	1	2	3	4	5
Variables		4	$3,4$	$2,3,4$	$2,3,3,4$	$2,3,3,4,4$
CV	3.766	3.699	3.620	3.609	3.601	3.597

- CV criterion decreases up to five kinks
- Most decrease obtained with two kinks
- Most nonlinearity in variables 3 & 4 : $\Delta_3 Y_{t-1}$, $\Delta_3 Y_{t-4}$
    - Two kinks in each

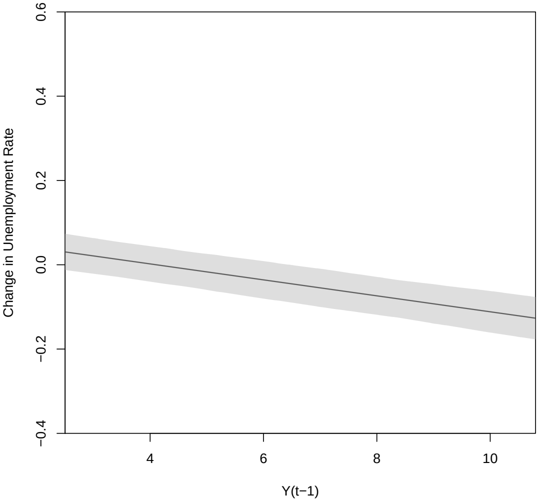
Regression Kink Estimates

$$\Delta Y_t = \underset{(.004)}{-.019}\; Y_{t-1} \underset{(.10)}{+.04}\; \Delta Y_{t-1} \underset{(.09)}{+.21}\; \Delta_3 Y_{t-1} \underset{(.07)}{+.09}\; \Delta_3 Y_{t-4}$$

$$\underset{(.17)}{-.32}\; (\Delta Y_{t-1} \underset{(.10)}{+.20}) \underset{(.14)}{-.28}\; (\Delta_3 Y_{t-1} \underset{(.12)}{+.40}) \underset{(.11)}{+.39}\; (\Delta_3 Y_{t-1} \underset{(.05)}{+.04})$$

$$\underset{(.08)}{+.05}\; (\Delta_3 Y_{t-4} \underset{(.39)}{+.20}) \underset{(.18)}{-.48}\; (\Delta_3 Y_{t-4} \underset{(.17)}{-.80})$$

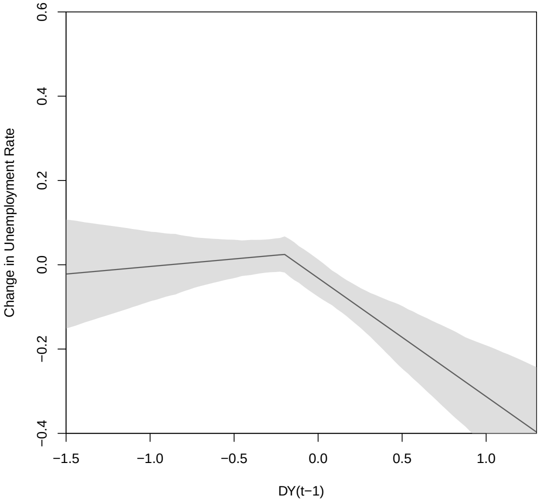
Effect of Y(t-1)



Effect of Y(t-1)

- Displayed is point estimate of regression function
- Plus 95% Numerical Delta method wild bootstrap intervals
- Effect is linear and of small magnitude

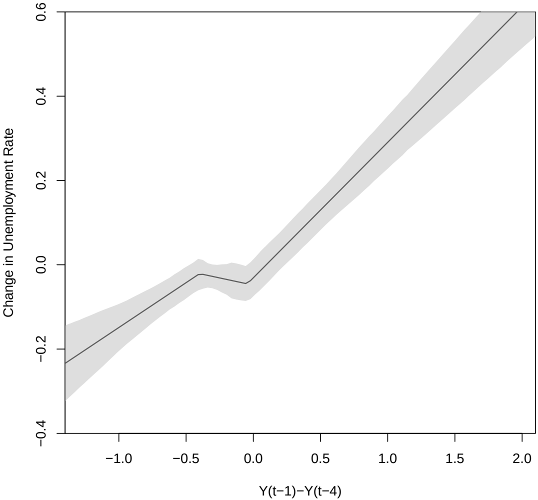
Effect of DY(t-1)



Effect of DY(t-1)

- Kink at $\Delta Y_{t-1} = 0$
- Effect zero for negative values
- Effect negative for positive values
- Implication: Asymmetric negative serial correlation

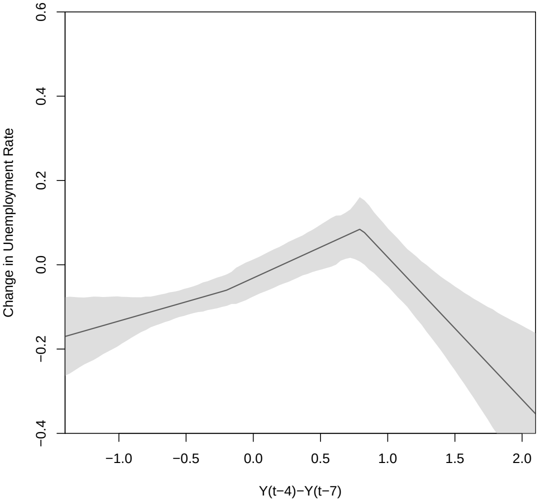
Effect of Y(t-1)-Y(t-4)



Effect of Y(t-1)-Y(t-4)

- Strongly positive
- With flat region near zero
- Implication: Positive serial correlation for "large" changes
- Zero serial correlation for "small" changes

Effect of Y(t-4)-Y(t-7)



Effect of Y(t-4)-Y(t-7)

- There are two kinks (at -0.2 and 0.8) but first hard to see
- Effect is small for quarterly changes which are negative or and positive and less than 1%
- This includes most observations (95%)
- Effect is steep negative for quarterly change above 1%
  - In strong contractions (ur strongly increasing) partial reversal with a lag
  - Kink is at 0.96 quantile: few data points in reversal region

Application Summary

- Nonlinear Regression Kink model “fits” better than linear AR model by CV criterion
- But only a modest improvement – This will not lead to a meaningful reduction in MSFE
- May lead to an improved understanding of nonlinear dynamics
- Produces a model with an improved description of asymmetric dynamics

Summary

- We need nonlinear regression models for nonlinear dynamics
- Models need to be nonparametric, parsimonious, yet allow sharp nonlinearity
- Feasible class: Regression Kink Autoregressive Model
- Estimation: NLLS
- Normal asymptotic distribution of coefficient estimates
- Nonstandard dsymptotic distribution of regression estimates; bootstrap method proposed
- Illustrated by application to unemployment rates