Regresión y Cointegración No Paramétricas para procesos recurrentes de Harris

por

Guillermo Moloche

Introducción y Motivación

- La mayoría de variables macroeconómicas nominales presentan (o se sospecha que presentan) no-estacionariedad.
- ¿Qué es no-estacionariedad?
  - En Econometría, generalmente se denomina NE a la inexistencia de la distribución incondicional de probabilidad. ¿Qué es la D.I.P.? Estadística I.
  - Nota: Para la mayoría de autores en estadística y matemática, E es homogeneidad temporal, se refiere a las probabilidades condicionales o de transición.
  - Muchas causas de NE: tendencias, persistencia, estacionalidad, cambio estructural, de régimen, etc. etc.

Introducción (cont.)

- Conocer qué es NE es muy importante porque:
  - La técnica principal del econometrista: mínimos cuadrados, o no funciona, o no funciona igual.
  - El econometrista que desconoce la NE puede "descubrir", y usar, resultados econométricos espúreos con resultados catastróficos en política económica, estrategia empresarial, finanzas, etc.
  - Transformar los datos para volverlos estacionarios implica apartarnos del modelo teórico o desperdiciar información económica interesante.

Introducción (cont.)

- Un ejemplo ya clásico de NE: Raíces Unitarias y Cointegración.
- ¿Qué es una Raíz Unitaria?
  - El elemental modelo autoregresivo:

$$X_{t+1} = a + bX_t + \varepsilon_t$$

es estacionario si $0 < b < 1$

es no-estacionario si $b = 1$

Introducción (cont.)

- ¿Qué es una Raíz Unitaria? (cont.)
  - Si el modelo es NE se dice que es un "camino aleatorio" o "variable integrada *I(1)*".
  - Si adicionalmente

$$a = 0$$

se dice que X_t es una "martingala"

Introducción (cont.)

- Los caminos aleatorios y las martingalas son omnipresentes en teoría económica, finanzas, y en trabajo aplicado.
  - Ej: la hipótesis de los mercados eficientes (EMH) consiste simplemente en afirmar que los precios de las acciones bursátiles siguen un camino aleatorio.
- ¿Qué pasa si dos variables *I(1)* están relacionadas económicamente?

Introducción (cont.)

- Supongamos que

$$Y_t = \alpha + \beta X_t + u_t$$

- Si u_t es estacionario, se dice que X_t y Y_t están cointegradas. Tienen una “relación de equilibrio” o “de no arbitraje”.
- Test de cointegración : test de estacionariedad para u_t

Introducción (cont.)

- Supongamos ahora que la relación de cointegración es de forma funcional desconocida: lineal? no-lineal?

$$Y_t = f\left(X_t\right) + u_t$$

- Las técnicas econométricas que no asumen formas funcionales *a priori* se llaman "no-paramétricas"
- ¿Por qué asumimos formas funcionales? No siempre por teoría económica

Introducción (cont.)

- ¿Es relevante este caso de cointegración? ¿Para qué necesito la cointegración noparamétrica? ¿Cointegración no lineal?
- Demanda de dinero
- ¿Curvas de Phillips? ¿Ecuaciones de Euler?
- PPP, cotizaciones de divisas
- Acciones bursátiles
- Curvas de rendimiento, tasas de interés

Introducción (cont.)

- ¿Qué se hace en este trabajo?
- Se estudia el caso de (posible) cointegración no-paramétrica con una teoría tan general como sea posible. El método no-paramétrico que estudiamos es la estimación por núcleos (kernel regression).
- Se demuestra matemáticamente que:

  el estimador de regresión de núcleos (kernel regression) funciona en situaciones NE, además de las E.

Sumario

- El estimador
- Resultados principales
  - Consistencia
  - Normalidad asintótica (mezclada).
- Supuestos teóricos
- Otros enfoques en regresión y cointegración NP
- Potenciales aplicaciones
- Cointegración NP en la práctica

El Estimador

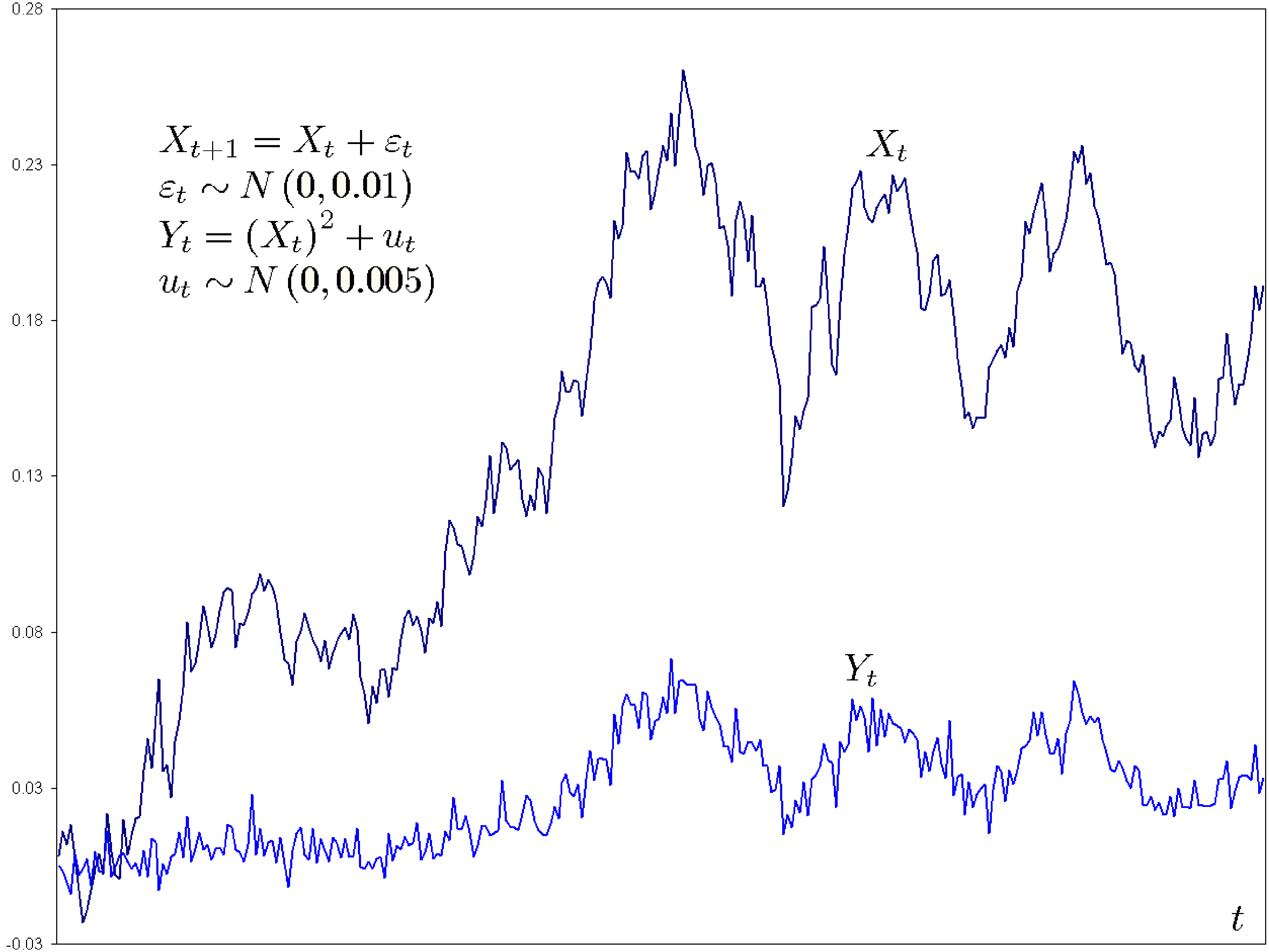
- Estimador de Nadaraya y Watson
- Regresión de Kernel

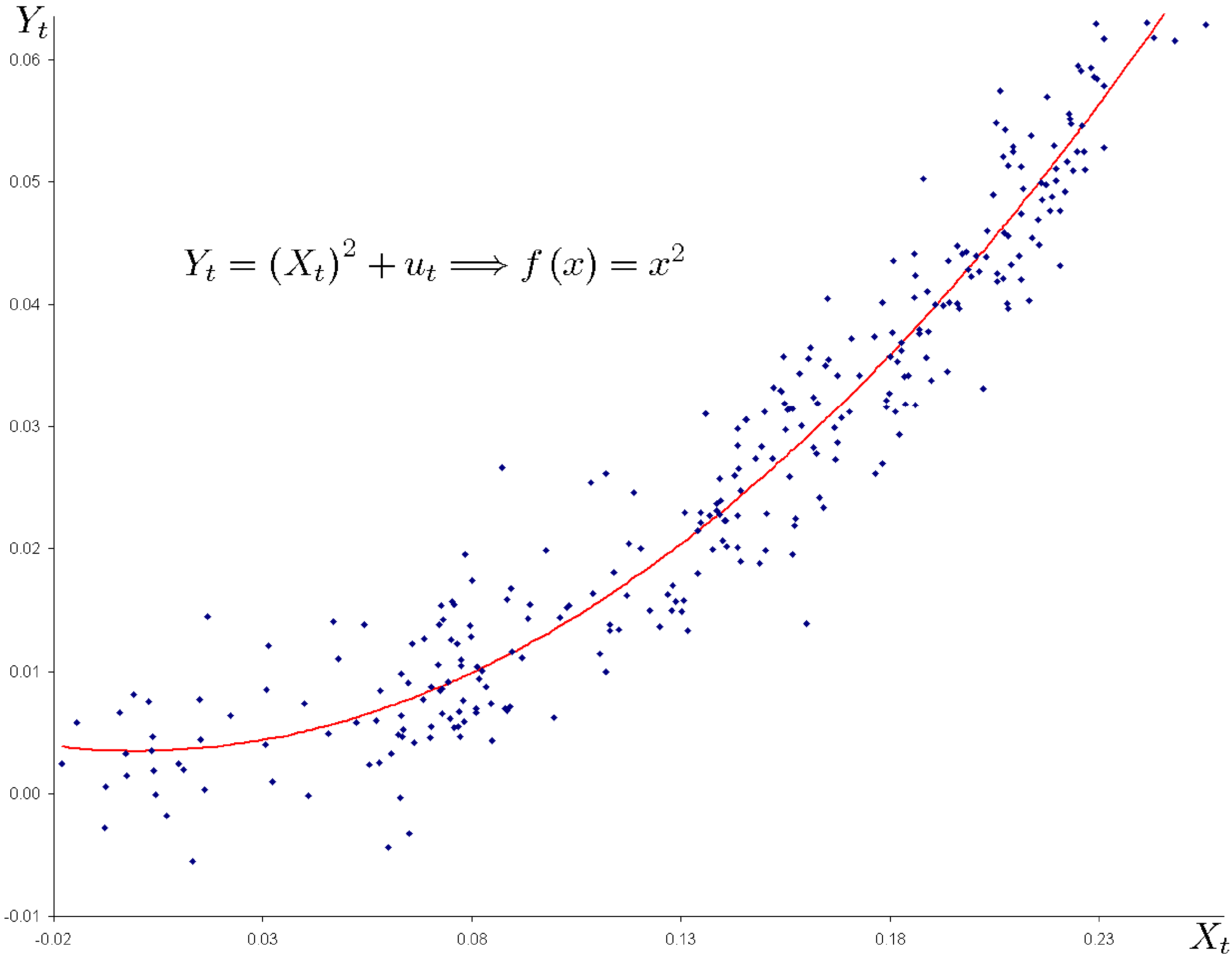
$$\widehat{f}(x) = \frac{\sum_{t=1}^{T} K((X_t - x)./h_T)Y_t}{\sum_{t=1}^{T} K((X_t - x)./h_T)}$$

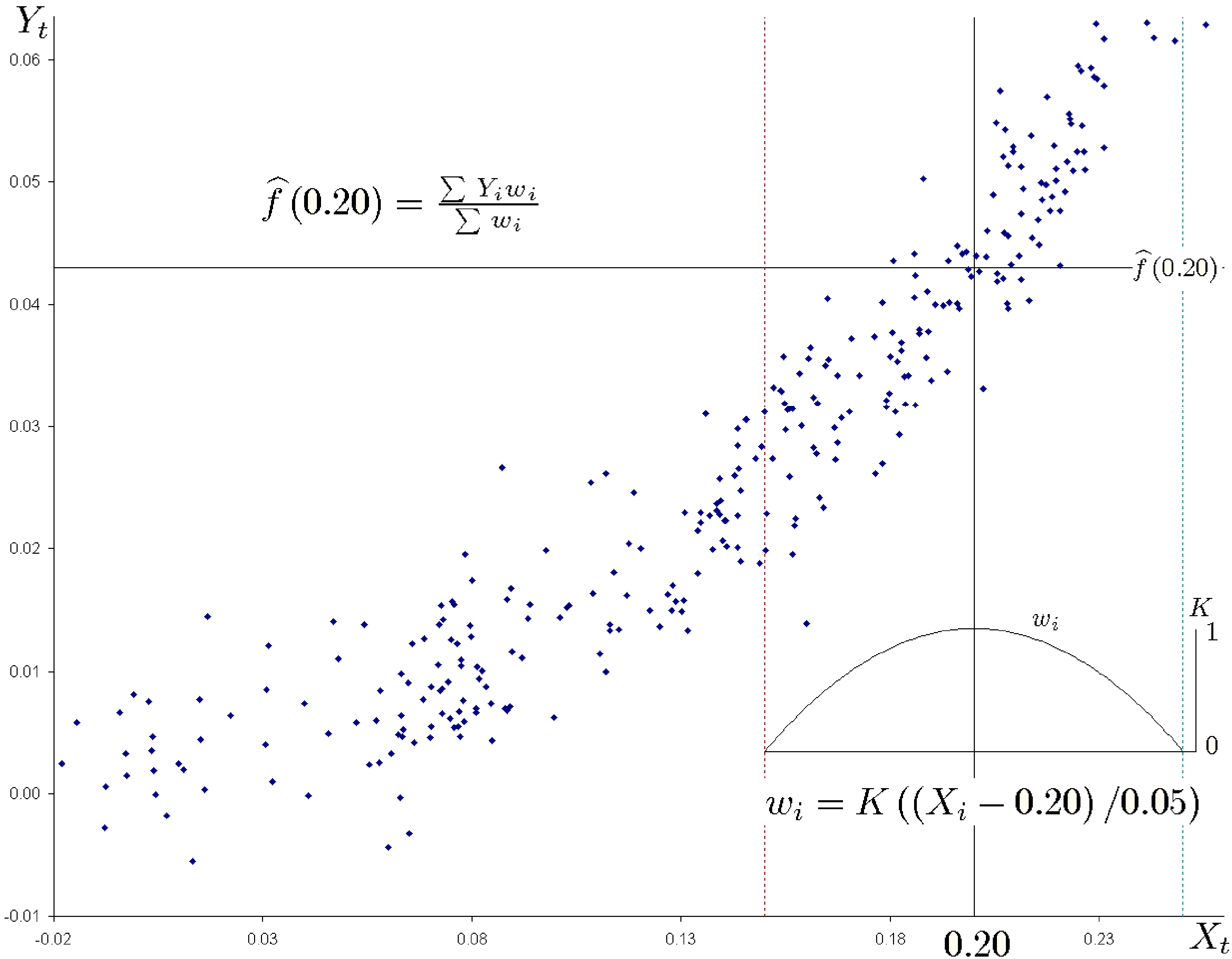
- h_T es el "ancho de banda" o bandwidth.

Intuición

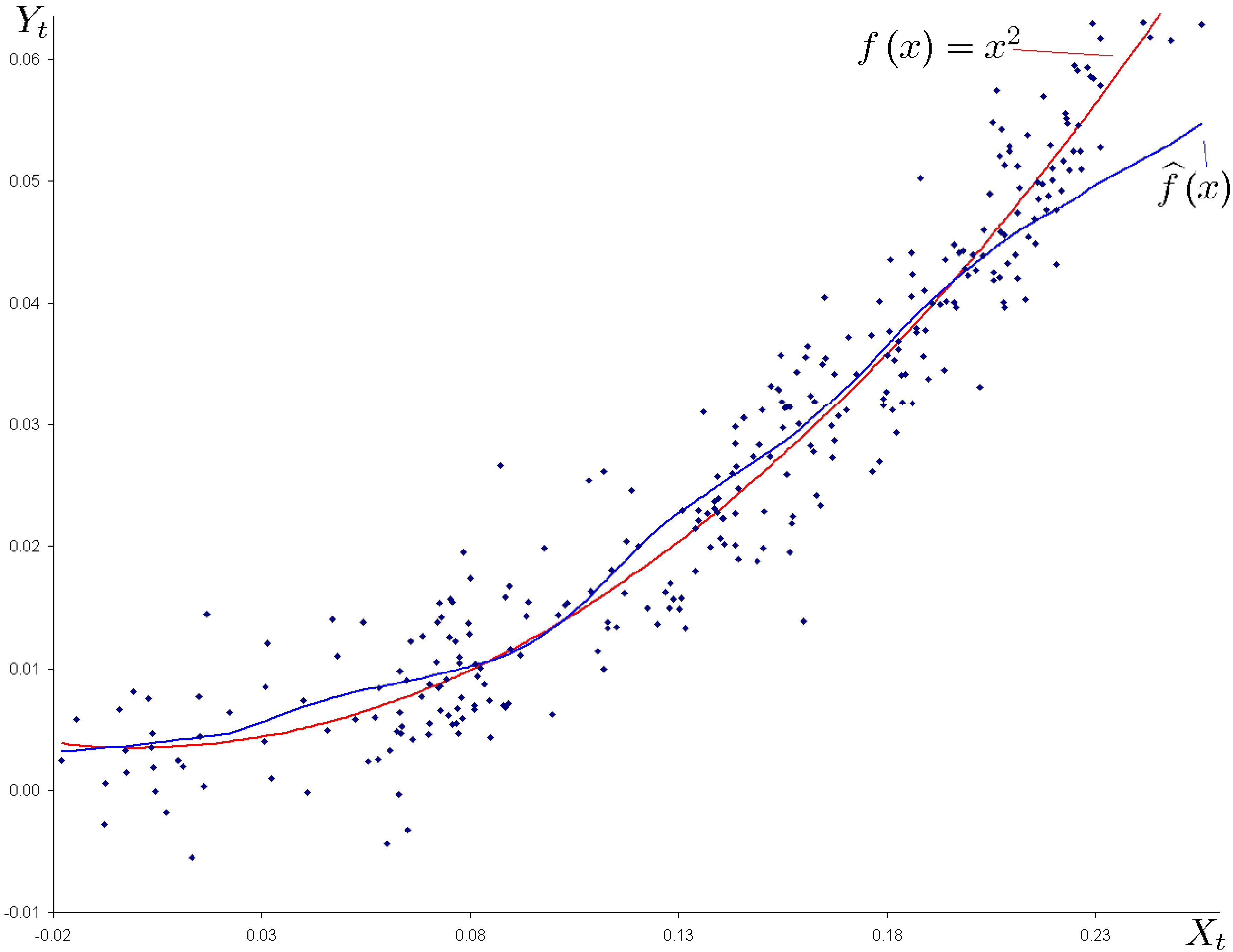
- La “regresión” de núcleos es simplemente un promedio ponderado de los valores de la variable “de la izquierda”, pero sólo de aquellos en los que la variable “de la derecha” es cercano al valor que se quiere estimar
- Más sencillo que mínimos cuadrados


0.28
0.23
0.18
0.13
0.08
0.03
-0.03
$X_{t+1} = X_t + \varepsilon_t$
$\varepsilon_t \sim N(0, 0.01)$
$Y_t = (X_t)^2 + u_t$
$u_t \sim N(0, 0.005)$
X_t
Y_t
t

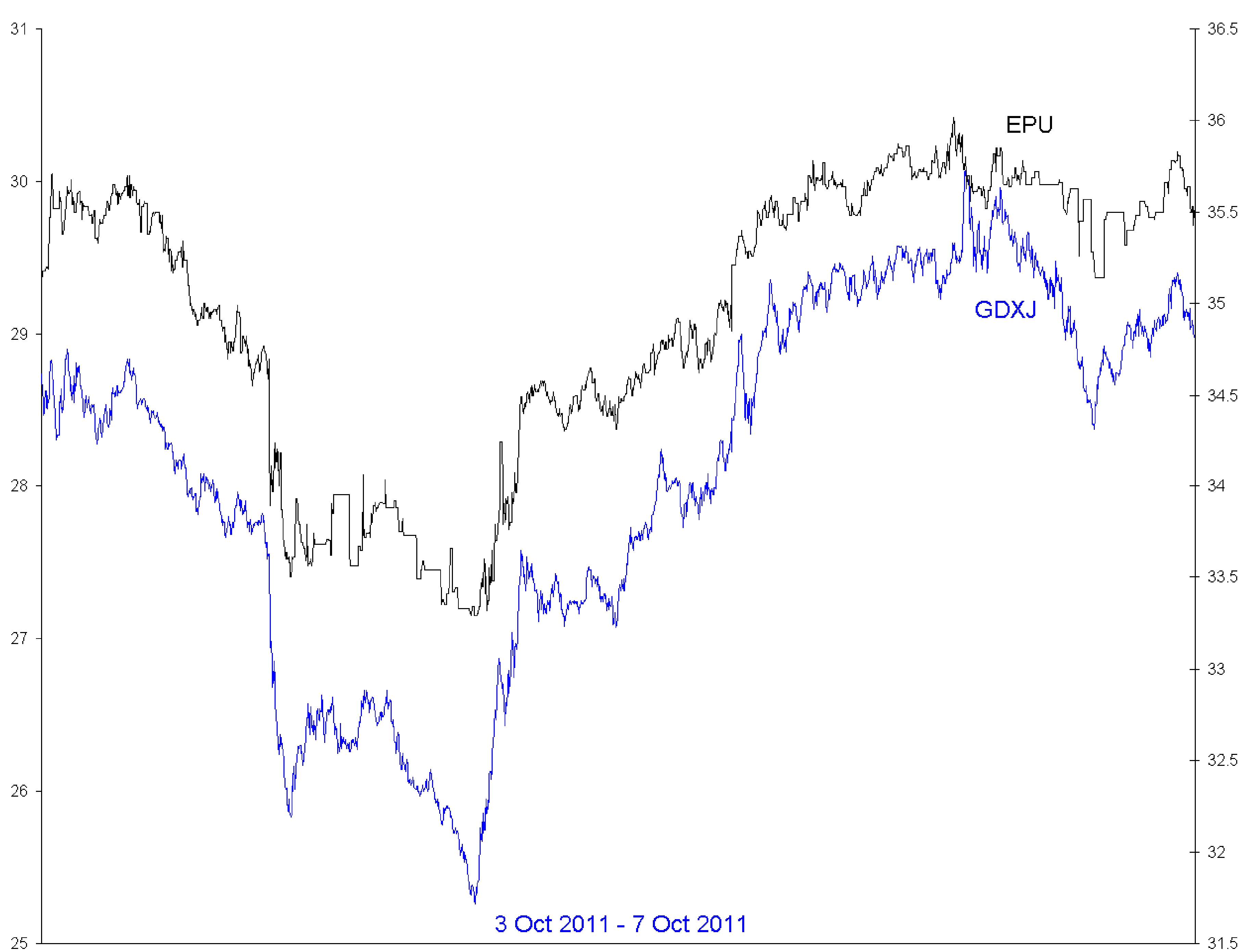

Y_t
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
$Y_t = (X_t)^2 + u_t \Longrightarrow f(x) = x^2$
-0.02
0.03
0.08
0.13
0.18
0.23
X_t



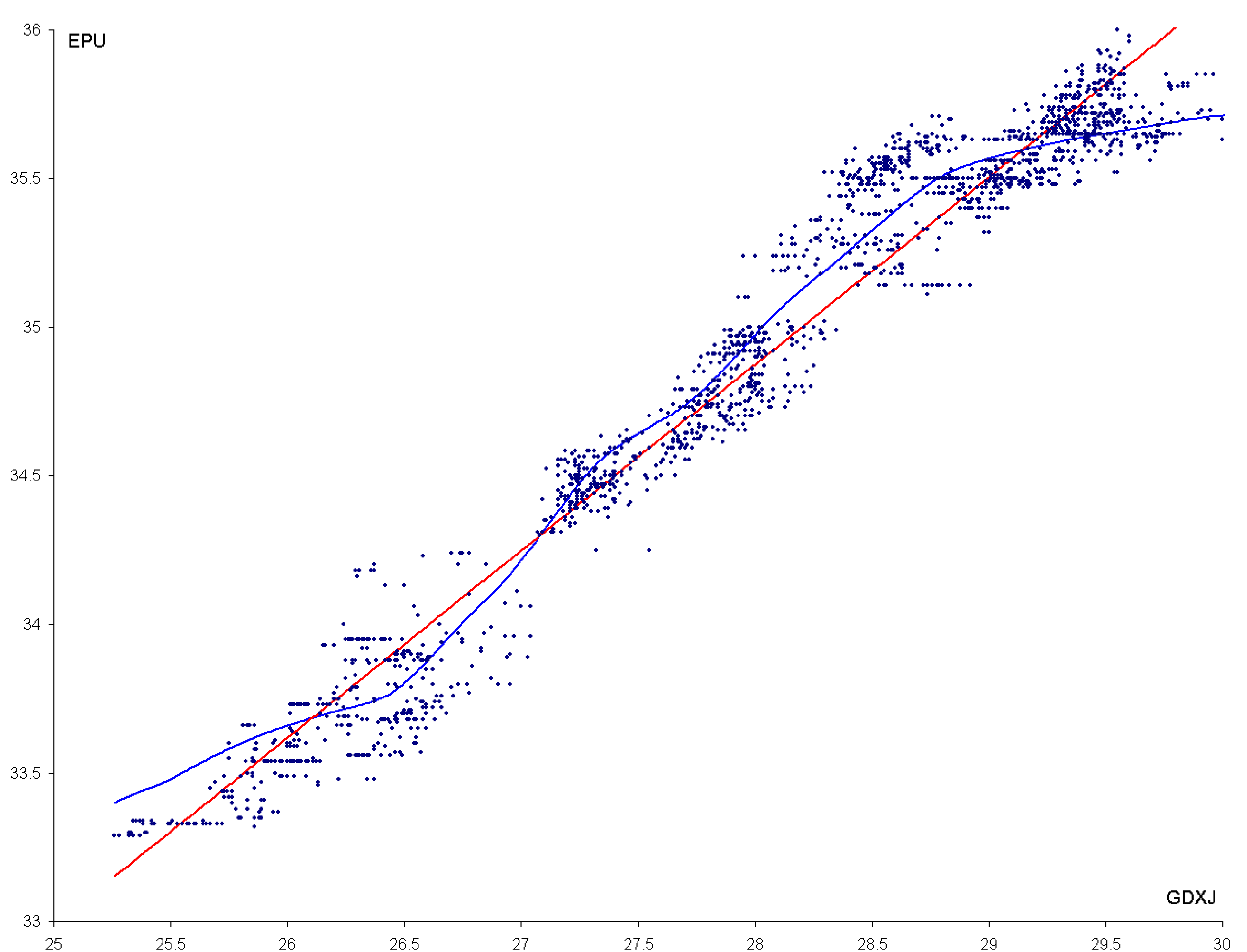
Y_t
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
$\widehat{f}(0.20) = \frac{\sum Y_i w_i}{\sum w_i}$
$\widehat{f}(0.20)$
w_i
K
1
0
$w_i = K\left(\left(X_i - 0.20\right)/0.05\right)$
-0.02
0.03
0.08
0.13
0.18
0.20
0.23
X_t



Y_t
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
0.03
0.08
0.13
0.18
0.23
X_t
$f(x) = x^2$
$\widehat{f}(x)$



EPU
GDXJ
3 Oct 2011 - 7 Oct 2011
31
30
29
28
27
26
25
36.5
36
35.5
35
34.5
34
33.5
33
32.5
32
31.5



EPU
GDXJ
36
35.5
35
34.5
34
33.5
33
25
25.5
26
26.5
27
27.5
28
28.5
29
29.5
30

Resultados Principales

- Consistencia

$$\widehat{f}(x) \to f(x)$$

- Normalidad Mezclada Asintótica

(Normal con matriz de variancias y cov. aleatoria)

$$\widehat{f}(x) \Rightarrow N\left(f(x), \frac{\int K^2(u)du}{h_T \widehat{L}_{T,h_T}(x)} E_x(u \cdot u')\right)$$

$$\widehat{L}_{T,h_T}(x) = \frac{1}{h_T} \sum_{t=1}^{T} K((X_t - x)./h_T)$$

Supuestos Teóricos

- Recurrencia de Harris (posible no-estacionariedad)
- Propiedad de Markov
- P.T. homogéneas (estacionarias) con la condición de Feller (continuidad)
- Condiciones sobre los núcleos (funciones *K*)
- Errores tienen variancia *condicional* finita (b)
- Condiciones sobre los anchos de banda (b)
- Condición de Darling-Kac (tne)

Recurrencia de Harris

- Intuitivamente: “la Historia se repite…”
- Incluye series estacionarias y no-estacionarias
- Incluye caminos aleatorios sin tendencia y de hasta dimensión dos. En dimensiones mayores a dos la recurrencia es posible dependiendo de si hay heteroskedasticidad condicional no-lineal.

Condición de Darling-Kac

- Parámetro de regularidad α. α en $[0,1]$.
- Intuición: la lentitud de convergencia de las sumas parciales (estimadores), el grado de recurrencia no-estacionaria.
  - $\alpha=1$: Todas las series estacionarias $(O(T))$
  - $\alpha=1/2$: Camino aleatorio de dimensión 1 $(O(T^{0.5}))$
  - $\alpha=0$: Camino aleatorio de dimensión 2 $(O(\log T))$
- Caracterización de E/NE alternativa a $I(d)$.
- Es muy fuerte esta condición?
  - No se sabe con certeza, aparentemente no lo es.
  - Conjetura de Harris: la recurrencia de Harris implica la condición de Darling-Kac.

¿Cómo es la prueba matemática?

- Teoría de las cadenas de Markov: hay teoremas muy generales que no requieren estacionariedad
- Teorema de Chacón-Ornstein
- Teorema de Darling-Kac
- Teoremas de convergencia en distribución para funcionales aditivos centrados
- Convergencia uniforme

Otros enfoques en regresión y cointegración NP

- Park y Phillips (wp 1998), Karlsen y Tjøstheim (AoS 2001)
  - Auto-regresión NP (posiblemente) no estacionaria univariada.
  - KT: caso aperiódico, no estacionario $0<\alpha<1$
- Wang y Phillips (EctTh 2009a, Ecta 2009b), Karlsen, Myklebust y Tjøstheim (AoS 2007)
  - (Posible) Cointegración NP caso univariado
  - KMT: Caso aperiódico, no estacionario $0<\alpha<1$

Otros enfoques (cont.)

- Guerre (wp2004), Schienle (wp 2008a,b,2011)
- Sancetta (JMultivAn 2009)
  - Nearest-neighbor nonparametric regression

    (Regresión noparamétrica por vecinos más próximos)
  - Recurrencia Harris (posible no estacionariedad)
  - Funciona porque al igual que los núcleos, restringe la estimación a un subconjunto de los datos, en este caso a los *k*-vecinos más próximos.

Otras Posibles Aplicaciones de esta Teoría

- Cointegración No Paramétrica
- Regresión con variables estacionarias y no-estacionarias (regresores casi integrados) (identificación, problemas inversos, vector de integración no es único).
- Autoregresión NP No Estacionaria
  - En este caso es más interesante la estimación de volatilidades (heteroskedasticidad condicional).
  - Es posible incluir caminos aleatorios de dimensión mayor a 2.

Cointegración no paramétrica en la práctica

- Esta teoría no es suficiente
- Cobra relevancia el trabajo en progreso de otros autores
  - Primero: Selección completamente automática (sin subjetividad) del ancho de banda h_T (bandwidth).
  - Segundo: Necesito test de E/NE para verificar si las variables cointegran en forma noparamétrica.

Ancho de Banda No-Estacionario

- Guerre (wp2004) Opt. sesgo-varianza (constante?)
- Gao, King, Lu y Tjøstheim (EcTh2009) Bootstrap (general?)
- Bandi, Corradi y Wilhelm (wp2010) Escogen el ancho de banda mediante dos etapas de contrastes. ¿Qué contrastan? Los supuestos de la teoría.

Tests de Cointegración NP

- Necesitamos tests de estacionariedad/no-estacionariedad robustos a nolinearidades
  - Breitung – Rank Test
  - Aparicio, Escribano – Record Counting Test
  - Bandi y Corradi (wp2011) - Basado en la condición de Darling-Kac: robusto a nolinearidades en nivel y en volatilidad

Regresión y Cointegración No Paramétricas para procesos recurrentes de Harris

por

Guillermo Moloche