Nonstationarity and the Kalman Filter

by

Taner Yiğit
Bilkent Univ.

Burak Eroglu
Bilkent Univ.

Isaac Miller
Univ. of Missouri-Columbia

Motivation

- Caution in time varying parameter estimation in the presence of nonstationarity
- Is "spurious regression" possible when using the KF?
- Suggest a method to avoid these problems in the Kalman Filter framework
- And use Kalman Filter for “time varying cointegration”

Literature

- Theoretical viewpoint:
  - Structural break in the cointegrating vector: Hansen (1992), Quintos & Phillips (1993), Andrews et al. (1996), Seo (1998), Hansen and Johansen (1999), Hansen (2003), Andrade et al. (2005)
  - Switching cointegrating vector: Hall, S.G., Z. Psaradakis and M. Sola (1997), Saikkonen and Choi (2004)
  - Time varying cointegration: Park and Hahn (1999, OLS), Harris et al. (2002, IV), Xiao (2006, kernel and local polynomial), Bierens and Martins (2009)

- Kalman Filter:
  - Kalman Filter: Wolff (1987), Kim and Nelson (1989), Kim (1993,2006), Evans (1991), Stock and Watson (1996), Huang and Hueng (2008)

Setup

Model:

$$Y_t = \alpha + \beta_t X_t + \varepsilon_t \qquad \text{var}(\varepsilon_t) = H$$

$$\beta_t = \mu + T\beta_{t-1} + \eta_t \qquad \text{var}(\eta_t) = Q$$

Prediction Equations:

$$\beta_{t|t-1} = \mu + T\beta_{t-1|t-1}$$

$$P_{t|t-1} = TP_{t-1|t-1}T' + Q$$

$$\omega_{t|t-1} = Y_t - Y_{t|t-1} = Y_t - \alpha_{t|t-1} - \beta_{t|t-1}X_t$$

$$f_{t|t-1} = X_t P_{t|t-1} X_t' + H$$

Setup

Updating Equations:

$$\beta_{t|t} = \beta_{t|t-1} + K_t \omega_{t|t-1}$$

$$P_{t|t} = P_{t|t-1} - K_t X_t P_{t|t-1}$$

Kalman Gain:

$$K_t = P_{t|t-1} X_t f_{t|t-1}^{-1}$$

Intuition: Adjust the TVP by a lot if Q is high, don't if H is high.

Problem

$$Y_t = \alpha + \beta_t X_t + \varepsilon_t \qquad \operatorname{var}(\varepsilon_t) = H$$
$$\beta_t = \mu + T\beta_{t-1} + \eta_t \qquad \operatorname{var}(\eta_t) = Q$$

- In spurious regression, $\varepsilon_t \sim I(1)$, but its increasing variance will be interpreted as "forecast error" by the Kalman filter and the β_t will be "overcorrected" (too much unnecessary variation in the TVP). Even if T is estimated using MLE (Canarella et al., 1990), it will converge to 1 and β_t will still have too much variation.
- In cointegration, $\varepsilon_t \sim I(0)$, and the superconsistency will cause the TVP to converge to the "correct" cointegrating vector.

Problem

- Canarella et al. (1990):

$$Y_t = a_{1t} + a_{2t} X_t + u_t \qquad u_t \sim N\left(0, \sigma_u^2\right)$$
$$a_{1t} = \beta_0 + \beta_1 a_{1t-1} + e_{1t} \qquad e_{1t} \sim N\left(0, \sigma_{e1}^2\right)$$
$$a_{2t} = \gamma_0 + \gamma_1 a_{2t-1} + e_{2t} \qquad e_{2t} \sim N\left(0, \sigma_{e2}^2\right)$$

$\beta_0, \beta_1, \gamma_0, \gamma_1, \sigma_u^2, \sigma_{e1}^2, \sigma_{e2}^2$ are estimated using MLE.

- Honohan (1993):

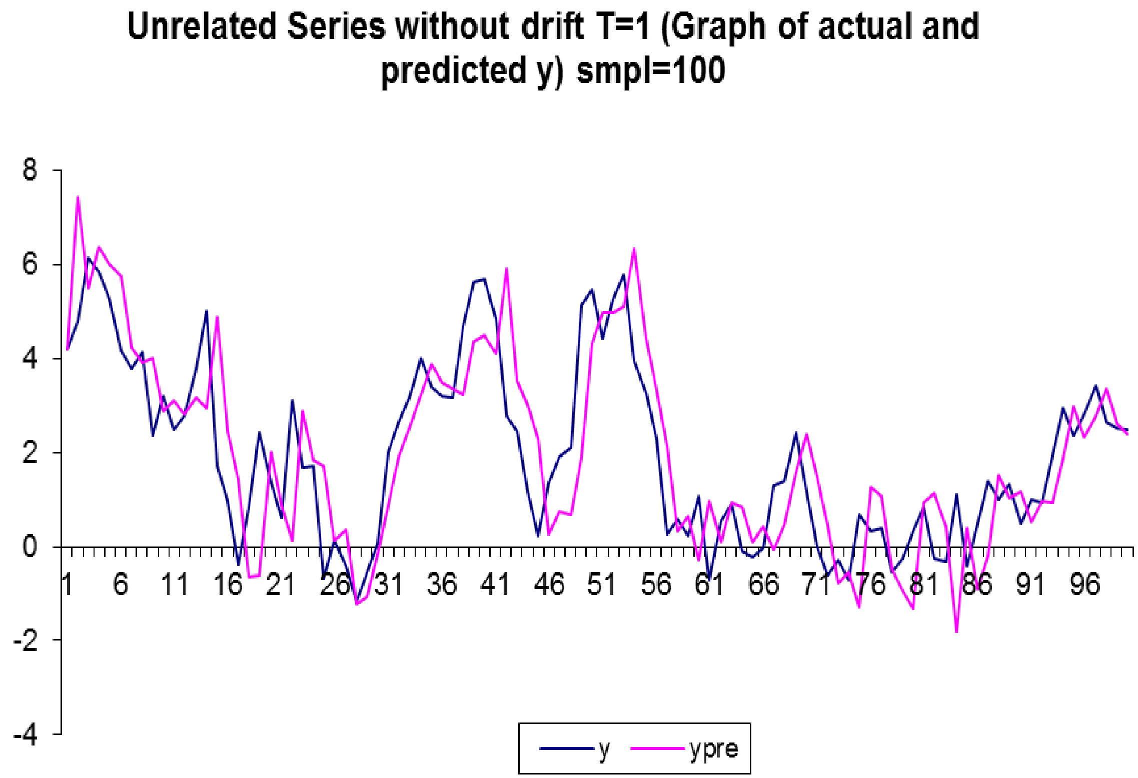
$$Y_t = a_{1t} + a_2 X_t + u_t$$

When $X_t = X_{t-1} + w_t$ and $Y_t = Y_{t-1} + v_t$, one can easily define a series $a_{1t} = Y_{t-1} - X_{t-1}$ such that the linear combination $Y_t - X_t - a_{1t} = v_t - w_t$ and we consequently (but spuriously) conclude that X and Y are cointegrated.

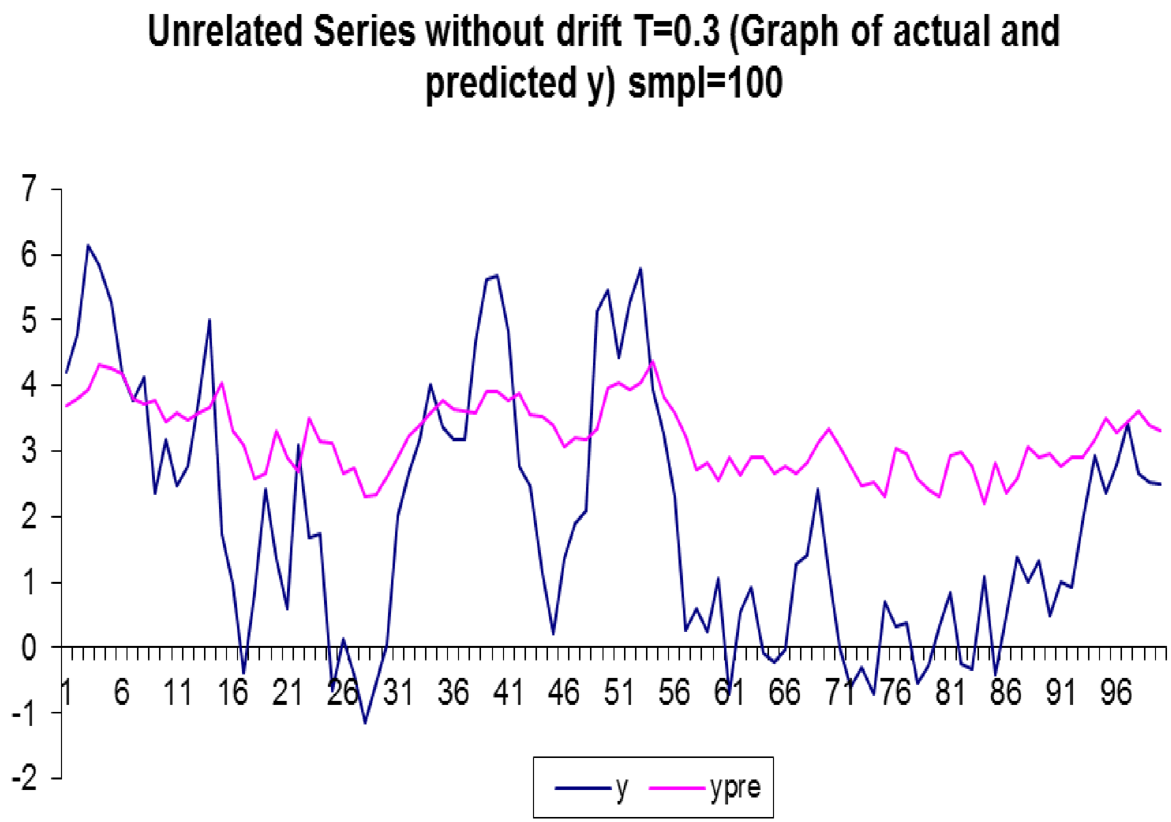
Problem

$$Y_t = \alpha + \beta_t X_t + \varepsilon_t \qquad \operatorname{var}(\varepsilon_t) = H$$

$$\beta_t = \mu + T\beta_{t-1} + \eta_t \qquad \operatorname{var}(\eta_t) = Q$$



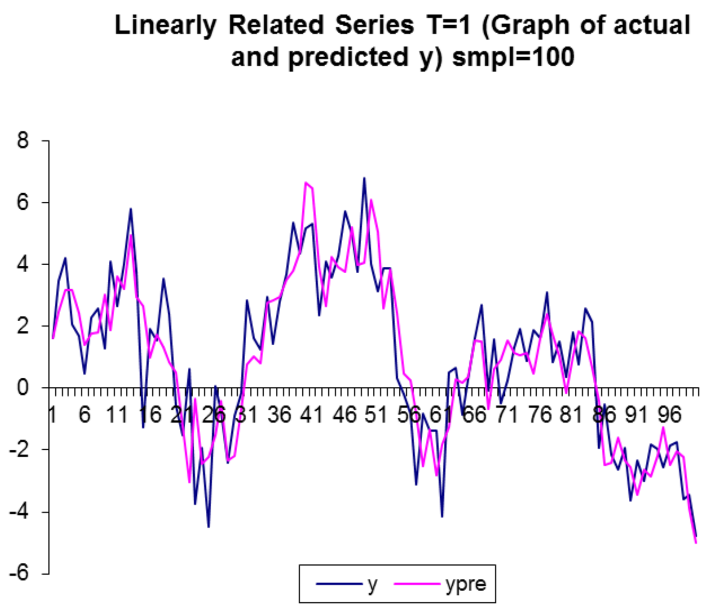
Problem

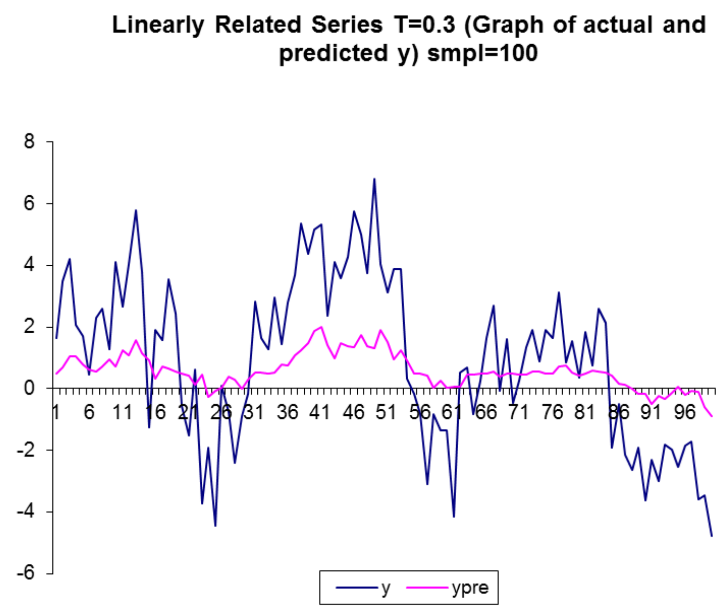


Unrelated Series without drift T=0.3 (Graph of actual and predicted y) smpl=100
7
6
5
4
3
2
1
0
-1
-2
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96
y
ypre

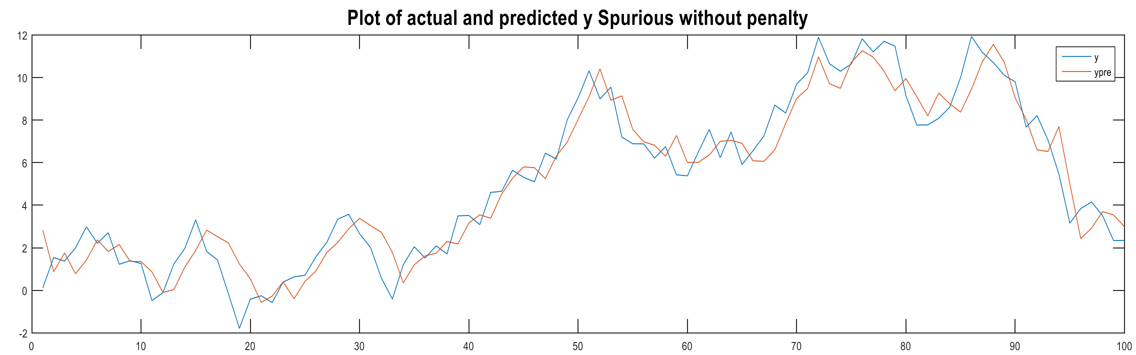
Problem

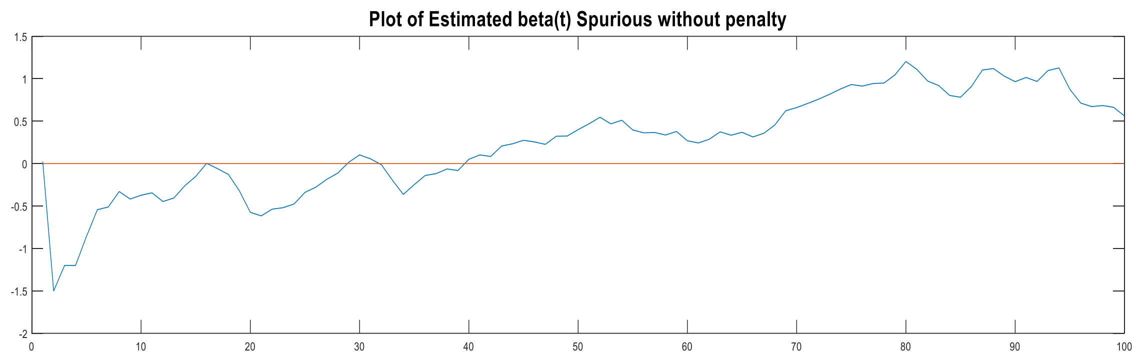
Is it a good way?



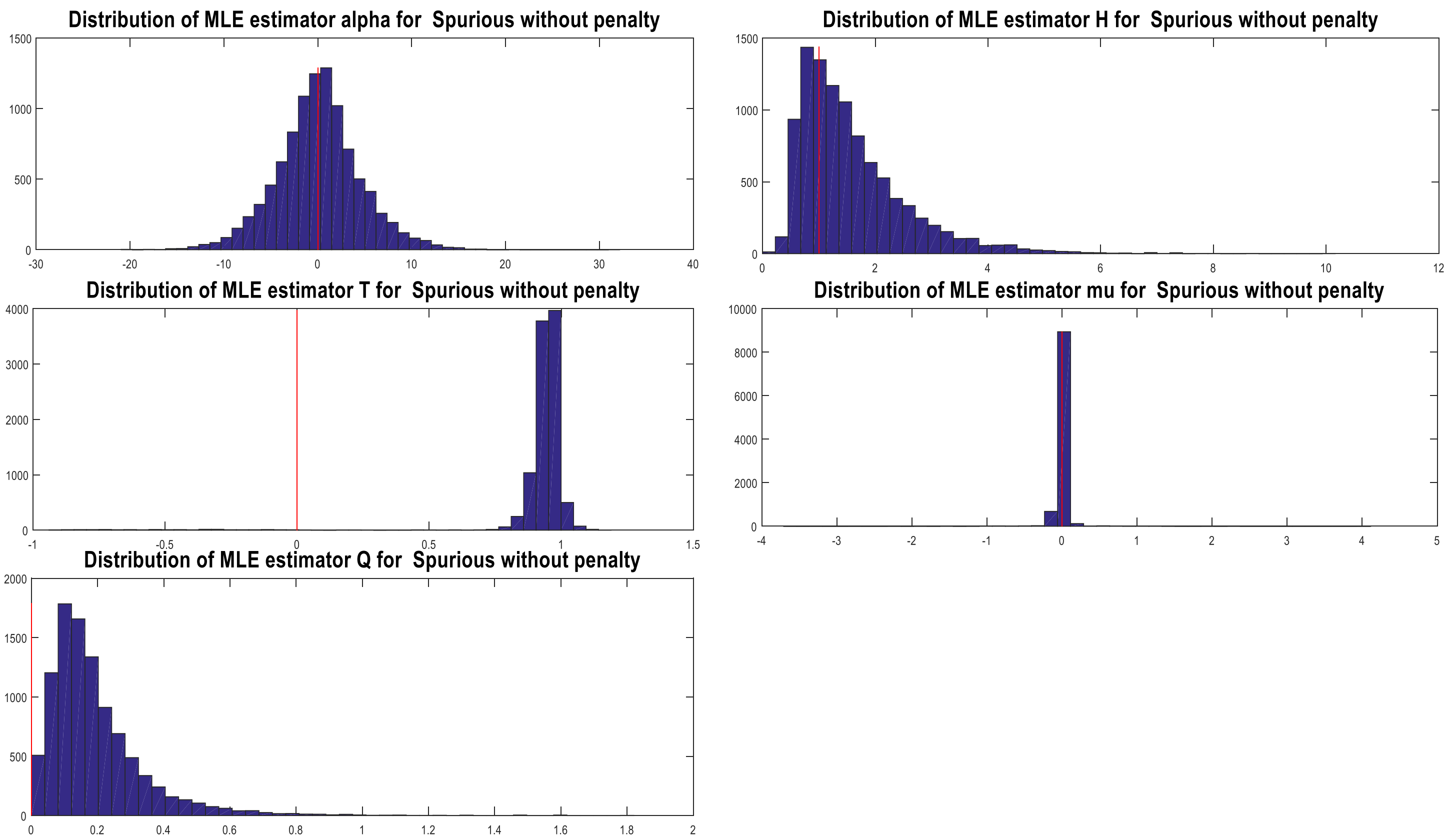


Does MLE change anything?





Does MLE change anything?



	Sample Size		α	H	T	μ	Q
Spurious	100	mean	0.094	1.573	0.919	0.000	0.113
		std	4.434	0.973	0.194	0.151	0.215

Solution

$$y_t = \alpha + \beta_t x_t + w_t$$

$$w_t = \theta w_{t-1} + \nu_t \qquad \nu_t \sim N(0, H)$$

$$\beta_t = \mu + T\beta_{t-1} + \eta_t \qquad \eta_t \sim N(0, Q)$$

- When $\theta \to 1$, Y_t and X_t are not cointegrated (hence can be devised as a cointegration test)
- When $\theta \ll 1$, "$T = 0$ and $Q = 0$" <u>or</u> "$T \ll 1$ and $Q = 0$", we have constant cointegration
- When $\theta \ll 1$, $T = 1$ with very small Q, we have a smooth time varying cointegration
- When $\theta \ll 1$, "$T \ll 1$ and $Q \neq 0$", we have stochastic cointegration

Finite Sample Performance (DGP)

Spurious regression:

$$x_t = x_{t-1} + e_t^x \quad \begin{bmatrix} e_t^x \\ e_t^y \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right]$$
$$y_t = y_{t-1} + e_t^y$$

Constant cointegration:

$$y_t = 1.5 + 0.9 x_t + \varepsilon_t \quad \varepsilon_t \sim iid(0,1)$$

Stochastic cointegration:

$$y_t = 1.5 + \beta_t x_t + \varepsilon_t \quad \varepsilon_t \sim iid(0,1)$$
$$\beta_t = 0.5 + 0.7\beta_{t-1} + \eta_t \quad \eta_t \sim iid(0, 0.14)$$

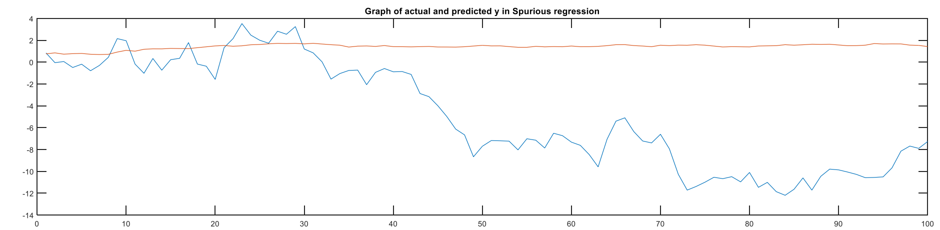
Smooth cointegration:

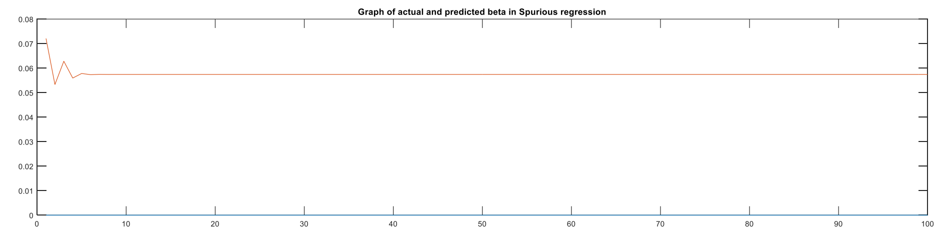
$$y_t = 1.5 + \beta_t x_t + \varepsilon_t; \ \varepsilon_t \sim iid(0,1)$$
$$\beta_t = 2\left(\frac{t}{N}\right)^3 - 3\left(\frac{t}{N}\right)^2 + 1.5$$

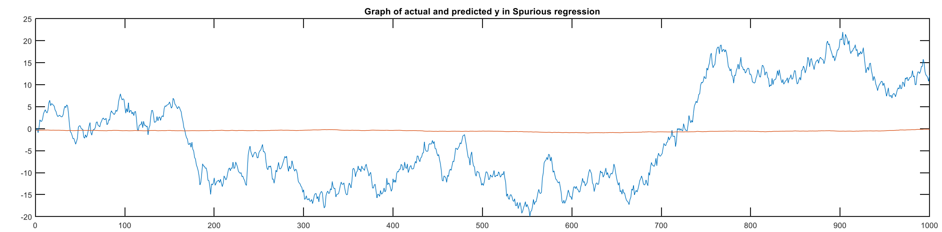
Finite Sample Performance

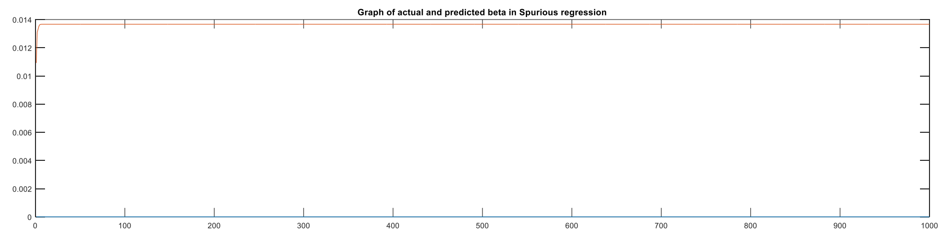
	Sample Size		α	H	T	μ	Q	θ
Spurious	100	mean	0.045	0.953	**-0.003**	**0.000**	0.014	**0.974**
		std	1.203	0.084	0.342	0.113	0.033	0.046
	1000	mean	0.081	0.990	**0.006**	**0.000**	0.003	**0.998**
		std	1.065	0.026	0.331	0.034	0.006	0.004
Constant	100	mean	**1.499**	0.948	-0.001	**0.901**	0.016	-0.022
		std	0.177	0.091	0.381	0.344	0.042	0.117
	1000	mean	**1.502**	0.988	0.008	**0.893**	0.003	-0.002
		std	0.055	0.027	0.381	0.343	0.008	0.035
Smooth	100	mean	**1.503**	0.984	**0.927**	-0.091	**0.001**	-0.015
		std	0.258	0.075	1.013	1.296	0.009	0.108
	1000	mean	**1.498**	0.993	**1.000**	-0.001	**0.003**	-0.010
		std	0.115	0.023	0.000	0.000	0.002	0.033
Stochastic	100	mean	**1.500**	0.904	**0.642**	**0.595**	0.370	-0.046
		std	0.383	0.264	0.125	0.211	0.057	0.276
	1000	mean	**1.502**	0.969	**0.696**	**0.507**	0.374	-0.022
		std	0.204	0.141	0.025	0.043	0.011	0.166

Spurious Regression Example

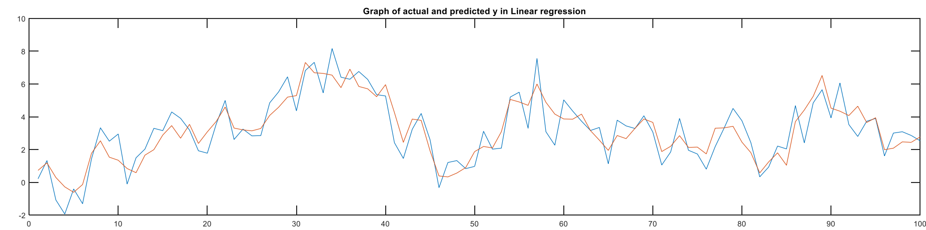


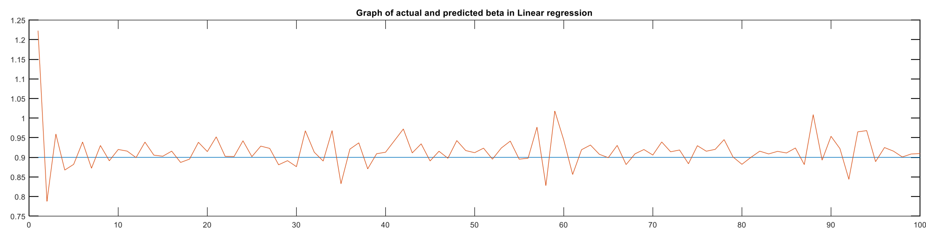


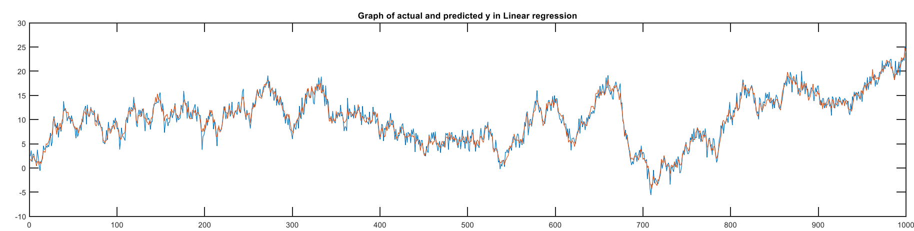


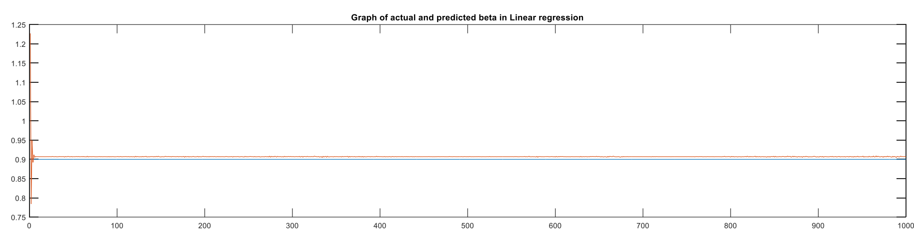


Constant Cointegration Example

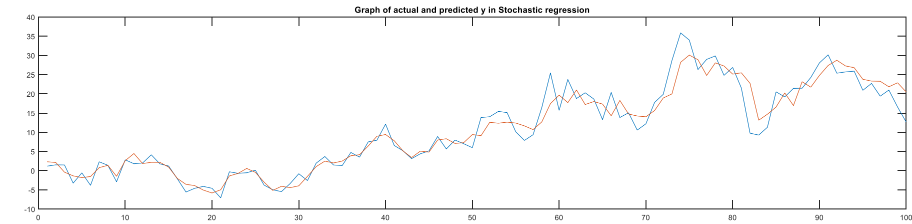


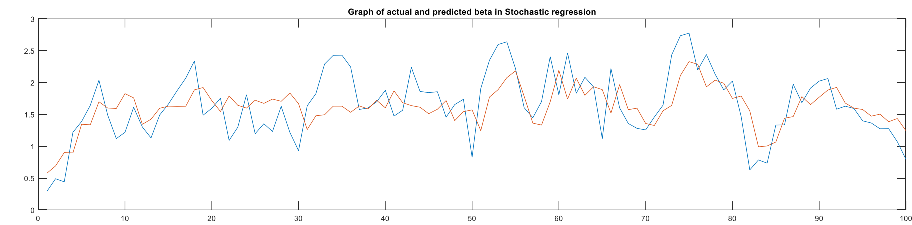


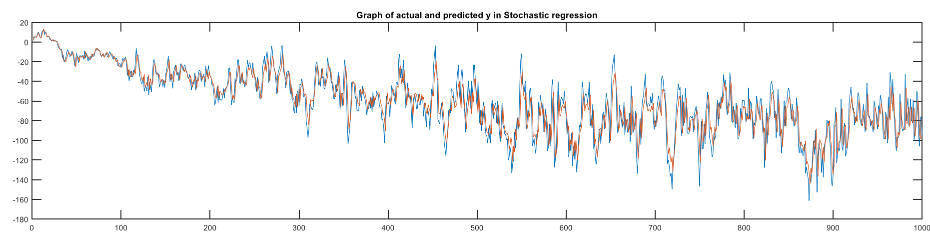


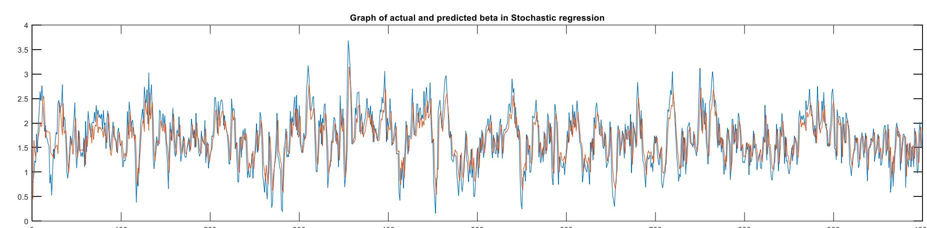


Stochastic Cointegration Example

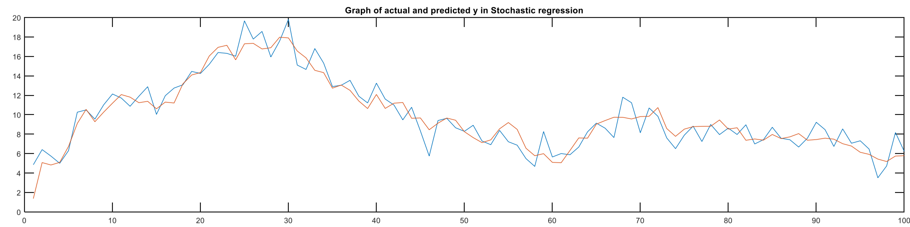


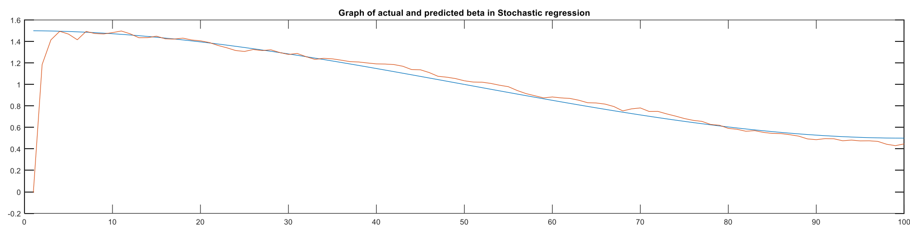


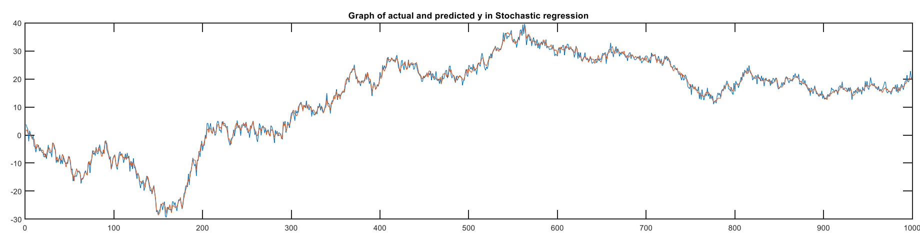


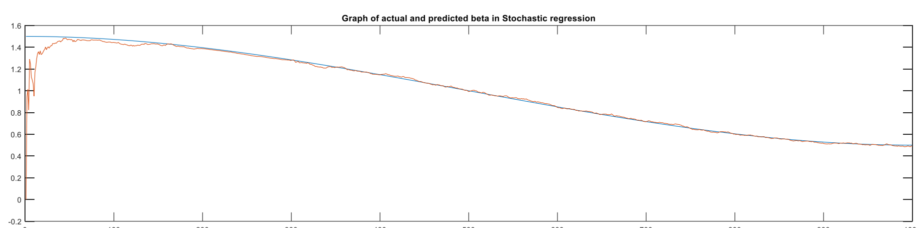


Smooth Varying Cointegration Example

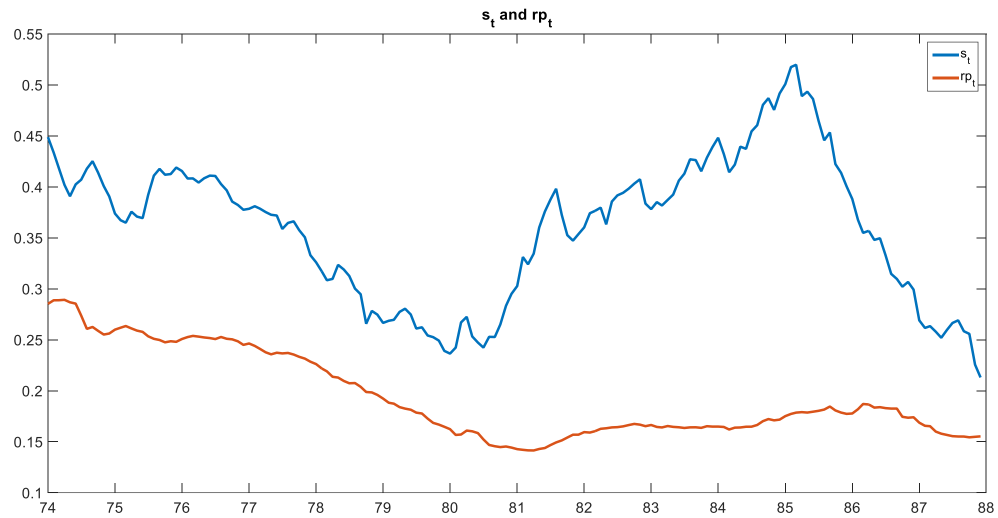








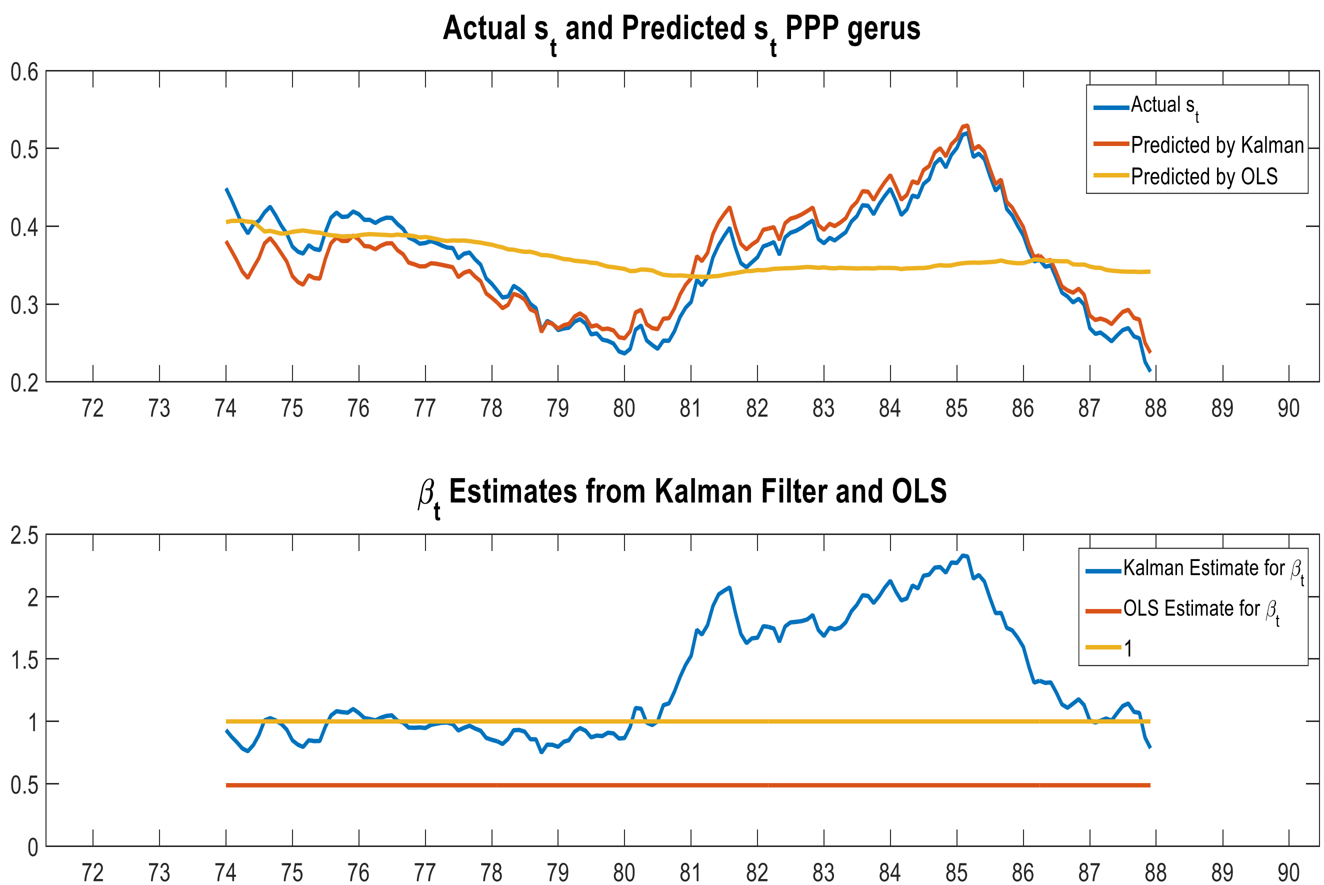
Application (PPP)



$$s_t = \alpha + \beta_t rp_t + \mu rp_t + w_t$$

$$\begin{bmatrix} \beta_t \\ w_t \end{bmatrix} = \begin{bmatrix} T_\beta & 0 \\ 0 & \theta \end{bmatrix} \begin{bmatrix} \beta_{t-1} \\ w_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \ \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} H^2 & 0 \\ 0 & Q^2 \end{bmatrix}\right)$$

Fitted values and Parameter Estimate



Conclusion

- There is a risk of spurious regression in TVP models. It cannot be avoided with the standard practices.

- Our solution seems to work very well in finite samples for a wide variety of cointegration models, time varying or not.

Extensions

- Theory.

- Extensions on testing.

Background information

- Integration:
  - Definition of nonstationarity
  - Properties of integrated variables (variance, mean, etc.)
    - Ex:

$$Y_t = \alpha + Y_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim iidN(0,1)$$

$$Y_t = Y_0 + \alpha t + \sum_{i=0}^{t-1} \varepsilon_{t-i}$$

$$E(Y_t) = Y_0 + \alpha t \qquad \operatorname{var}(Y_t) = t$$

  - Rates of convergence of estimates (T instead of $\sqrt{T}$)
    - Ex: $\sigma^2 (X'X)^{-1}$

Background information

- Cointegration:
  - Definition: $X_t \sim I(1); Y_t \sim I(1); Y_t - \beta X_t \sim I(0)$
  - Superconsistent if it exists
  - Spurious (regression) if it doesn't
    - Ex: US Export Index (Y) on Australian males' life expectancy (X) - annual data 1960-90

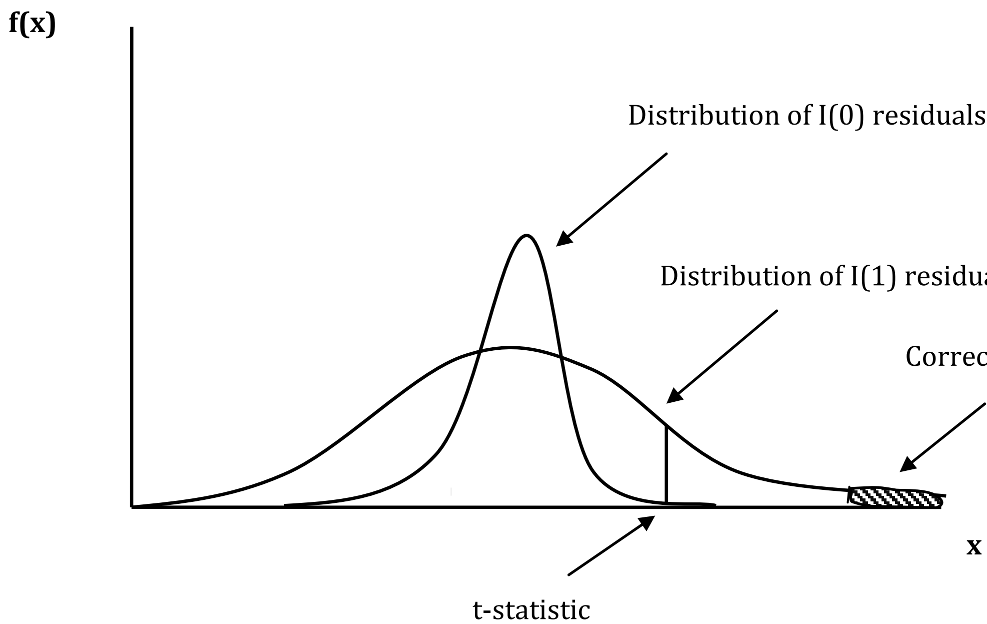
$$\hat{Y} = \underset{(-16.7)}{-2943} + \underset{(17.8)}{45.8X} \qquad R^2 = 0.92 \quad DW = 0.36 \quad F = 315.2 \quad Corr = 0.96$$

    - Ex: Total Crime rates in US (Y) on life expectancy of South Africa (X) - annual data 1971-91

$$\hat{Y} = \underset{(-6.03)}{-24569} + \underset{(9.04)}{628.9X} \qquad R^2 = 0.81 \quad DW = 0.51 \quad F = 81.7 \quad Corr = 0.90$$

Background information

- Reason for spurious regression: noise as strong as the signal
  - $Y_t = \alpha + \beta X_t + \varepsilon_t \qquad Y_t \sim I(1); X_t \sim I(1); \varepsilon_t \sim I(1)$
- Phillips (1986); Entorf (1997); Granger (2001)



Background information

- State space models

- (Special case) Time varying parameters:
$$Y_t = a_t Z_t + \varepsilon_t$$
$$a_t = T a_{t-1} + v_t$$

- Kalman Filter:
  - Recursive procedure for computing the optimal estimator of the state vector at time *t* based on information at time *t*
  - Minimum Mean Square (Linear) Estimator

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \right)$$

The optimal forecast of x_2 having observed x_1 is $N\left(m_2, \Sigma\right)$

$$m_2 = \mu_2 + \Omega_{21}\Omega_{11}^{-1}\left(x_1 - \mu_1\right)$$

$$\Sigma = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$$