# Cointegration in Theory and Practice

# A Tribute to Clive Granger

ASSA Meetings

January 5, 2010

James H. Stock

Department of Economics, Harvard University and the NBER


SMU

# Cointegration: The Historical Setting

Granger and Newbold (*Journal of Econometrics*, 1974)

It is very common to see reported in applied econometric literature time series regression equations with an apparently high degree of fit, as measured by the coefficient of multiple correlation $R^2$ or the corrected coefficient $\bar{R}^2$, but with an extremely low value for the Durbin-Watson statistic. We find it very curious that whereas virtually every textbook on econometric methodology contains explicit warnings of the dangers of autocorrelated errors, this phenomenon crops up so frequently in well-respected work… (p. 111)

Dickey-Fuller (*JASA*, 1979):

$$Y_t = \rho Y_{t-1} + \varepsilon_t$$

The hypothesis that $\rho = 1$ is of some interest in applications because it corresponds to the hypothesis that it is appropriate to transform the time series by differencing. Currently, practitioners may decide to difference a time series on the basis of visual inspection of the autocorrelation function… (p. 427)

Davidson, Hendry, Srba and Yeo (*Economics Journal*, 1978)

$$\Delta_4 c_t = \underset{(.04)}{0.49}\Delta_4 y_t - \underset{(.05)}{0.17}\Delta\Delta_4 y_t - \underset{(.01)}{.06}(c_{t-4} - y_{t-4}) + \underset{(.004)}{0.01}D_t \qquad (41)$$

cf. Hall (1978)

The relationship between error correction models and co-integration was first pointed out in Granger (1981). A theorem showing precisely that co-integrated series can be represented by error correction models was originally stated and proved in Granger (1983). The following version is therefore called the Granger Representation Theorem. Analysis of related but more complex cases is covered by Johansen (1985) and Yoo (1985).

GRANGER REPRESENTATION THEOREM: If the $N \times 1$ vector $x_t$ given in (3.1) is co-integrated with $d = 1$, $b = 1$ and with co-integrating rank $r$, then:

(1) *$C(1)$ is of rank $N - r$.*

(2) *There exists a vector ARMA representation*

$$A(B)x_t = d(B)\varepsilon_t \tag{3.3}$$

*with the properties that $A(1)$ has rank $r$ and $d(B)$ is a scalar lag polynomial with $d(1)$ finite, and $A(0) = I_N$. When $d(B) = 1$, this is a vector autoregression.*

(3) *There exist* $N \times r$ *matrices*, $\alpha$, $\gamma$, *of rank* $r$ *such that*

$$\alpha' C(1) = 0,$$

$$C(1)\gamma = 0,$$

$$A(1) = \gamma\alpha'.$$

(4) *There exists an error correction representation with* $z_t = \alpha' x_t$, *an* $r \times 1$ *vector of stationary random variables*:

$$A^*(B)(1-B)x_t = -\gamma z_{t-1} + d(B)\varepsilon_t \tag{3.4}$$

*with* $A^*(0) = I_N$.

(5) *The vector* $z_t$ *is given by*

$$z_t = K(B)\varepsilon_t, \tag{3.5}$$

$$(1-B)z_t = -\alpha'\gamma z_{t-1} + J(B)\varepsilon_t, \tag{3.6}$$

*where* $K(B)$ *is an* $r \times N$ *matrix of lag polynomials given by* $\alpha' C^*(B)$ *with all elements of* $K(1)$ *finite with rank* $r$, *and* det $(\alpha'\gamma) > 0$.

(6) *If a finite vector autoregressive representation is possible, it will have the form given by* (3.3) *and* (3.4) *above with* $d(B) = 1$ *and both* $A(B)$ *and* $A^*(B)$ *as matrices of finite polynomials.*

# Cointegration: Econometric Theory
# Triumphs and Disappointments

Triangular model – cointegrated (1,1) with $n$=2, $r$=1:

$\Delta x_t = v_t$

$y_t = \theta x_t + u_t$

Main aims of initial econometric theory:

1. Superconsistency of OLS estimator $\hat{\theta}$
2. Use of estimated ECM term $\hat{z}_t = y_t - \hat{\theta} x_t$ as a regressor – without the "generated regressor" problem
3. Testing for cointegration (e.g. EG-ADF test on OLS residual)
4. Efficient (Gaussian) estimation of $\theta$
5. Inference for $\theta$

Simple expositional model:

$\Delta x_t = v_t$

$y_t = \theta x_t + u_t,$

Assume:

$(v_t, u_t) \sim (0,\Sigma)$, serially uncorrelated, and

$$\begin{pmatrix} \dfrac{1}{\sqrt{T}} \sum_{t=1}^{[T.]} v_t \\ \dfrac{1}{\sqrt{T}} \sum_{t=1}^{[T.]} u_t \end{pmatrix} \Rightarrow \begin{pmatrix} B_v(.) \\ B_u(.) \end{pmatrix}, \; B \text{ is BM}(\Sigma)$$

## Superconsistency and distribution

$$T(\hat{\theta}-\theta) = \frac{\frac{1}{T}\sum_{t=1}^{T} x_t u_t}{\frac{1}{T}\sum_{t=1}^{T} x_t^2} \Rightarrow \frac{\sigma_{uv} + \int B_v dB_u}{\int B_v^2}$$

## ECM as a regressor

$$w_t = \beta z_t + \zeta_t = \beta \hat{z}_t + [\beta(z_t - \hat{z}_t) + \zeta_t]$$

$$= \beta z_t + \zeta_t = \beta \hat{z}_t + [(\hat{\theta} - \theta)\beta x_t + \zeta_t]$$

## Efficient estimation and inference

MLE: $f(Y,X|\theta,\gamma) = f(Y|X,\theta,\gamma_1)f(X|\gamma_2)$

so $y_t = \theta x_t + \gamma_1(\mathrm{L})\Delta x_t + u_t^{\perp}$

$$T(\hat{\theta}-\theta) = \frac{\frac{1}{T}\sum_{t=1}^{T} x_t u_t^{\perp}}{\frac{1}{T}\sum_{t=1}^{T} x_t^2} + o_p(1) \Rightarrow \frac{\int B_v dB_{u^{\perp}}}{\int B_v^2} \sim \int N(0, \frac{\sigma_u^{\perp 2}}{\int B_v^2}) dF_{\int B_v^2}$$

Which of these results are robust to changes in assumptions about long run properties?

$x_t = \alpha x_{t-1} + v_t$, $\alpha = 1 + c/T$ (local to unity model)

$y_t = \theta x_t + u_t$,

($v_t$, $u_t$) satisfy same assumptions; $c$ is unknown

<u>Superconsistency and distribution</u>

$$T(\hat{\theta}-\theta) = \frac{\frac{1}{T}\sum_{t=1}^{T} x_t u_t}{\frac{1}{T^2}\sum_{t=1}^{T} x_t^2} \Rightarrow \frac{\sigma_{uv} + \int J_v dB_u}{\int J_v^2}, \; dJ_v = cJ_v + dW_v$$

<u>ECM as a regressor</u>

$$w_t = \beta z_t + \zeta_t = \beta \hat{z}_t + [\beta(z_t - \hat{z}_t) + \zeta_t]$$

$$= \beta z_t + \zeta_t = \beta \hat{z}_t + [(\hat{\theta} - \theta)\beta x_t + \zeta_t]$$

# Efficient estimation and inference

MLE: $f(Y,X|\theta,\gamma) = f(Y|X,\theta,\gamma_1)f(X|\gamma_2)$

so $y_t = \theta x_t + \gamma_1(\mathrm{L})(x_t - \alpha x_{t-1}) + u_t^{\perp}$

$= \theta x_t + \gamma_1(\mathrm{L})\Delta x_t + [\gamma(\mathrm{L})(1-\alpha)x_{t-1} + u_t^{\perp}]$

$= \theta x_t + \gamma_1(\mathrm{L})\Delta x_t + [T^{-1}\gamma(\mathrm{L})cx_{t-1} + u_t^{\perp}]$

so (Elliott (1998))

$$T(\hat{\theta}-\theta) = \frac{\frac{1}{T}\sum_{t=1}^{T} x_t\left(T^{-1}\gamma(1)cx_t + u_t^{\perp}\right)}{\frac{1}{T^2}\sum_{t=1}^{T} x_t^2} + o_p(1)$$

$$= \gamma(1)c + \frac{\frac{1}{T}\sum_{t=1}^{T} x_t u_t}{\frac{1}{T^2}\sum_{t=1}^{T} x_t^2} \Rightarrow \gamma(1)c + \frac{\int J_v dB_{u^{\perp}}}{\int J_v^2} \sim \int N\Big(\gamma(1)c, \frac{\sigma_u^{\perp 2}}{\int J_v^2}\Big) dF_{\int J_v^2}$$

Some recent work on this problem

Jansson and Moreira (2006)

The OLS and MLE distributions are sensitive to other models of long-run behavior e.g. fractional integration – essentially have nuisance parameters that are not estimable. See Müller and Watson (2008)

Testing

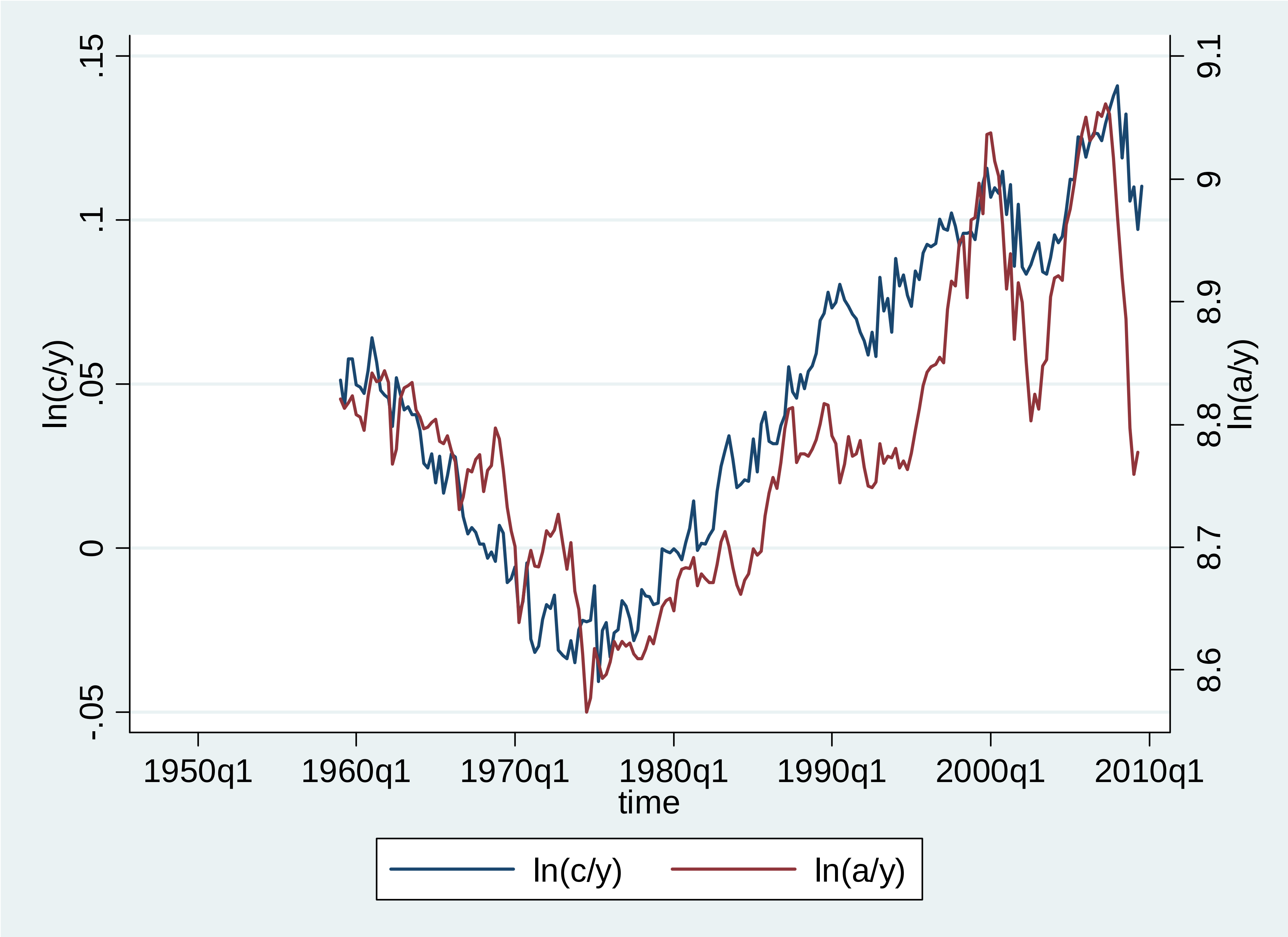
has the same issues

# Cointegration: Empirical Legacy

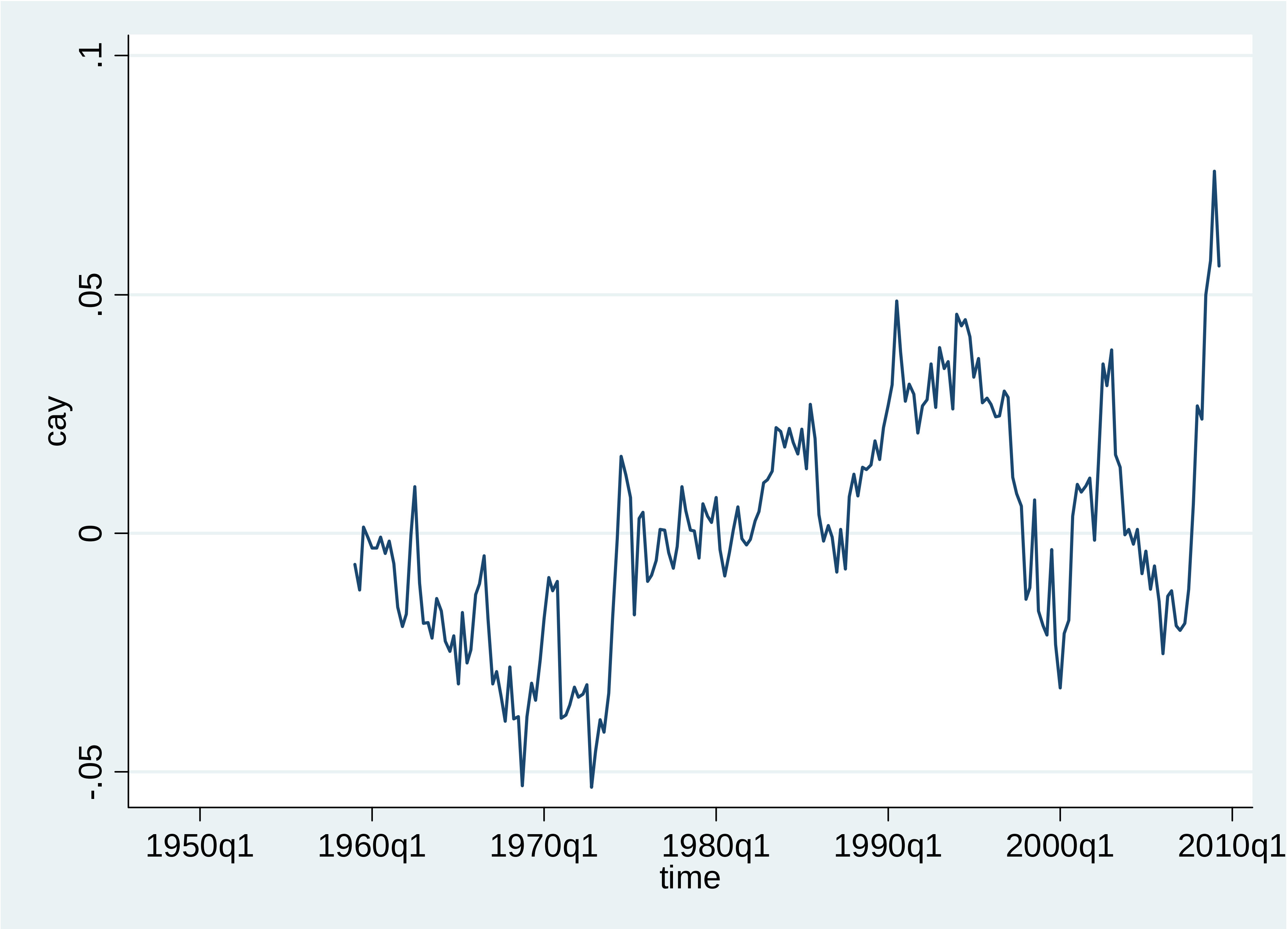
Two examples

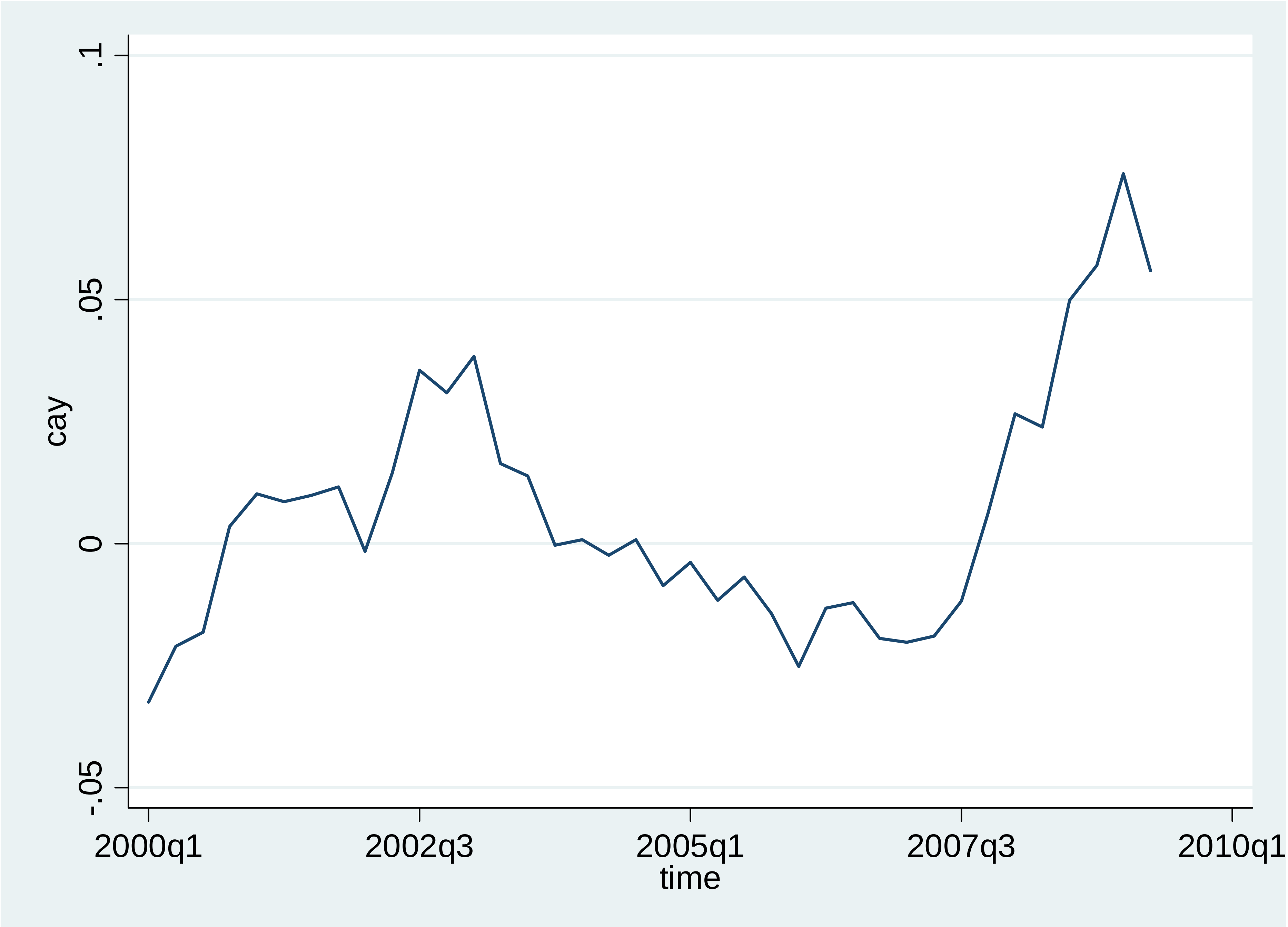
(1) Consumption/income and consumption/income/wealth

DHSY (1978)…. Lettau and Ludvigson (2004)

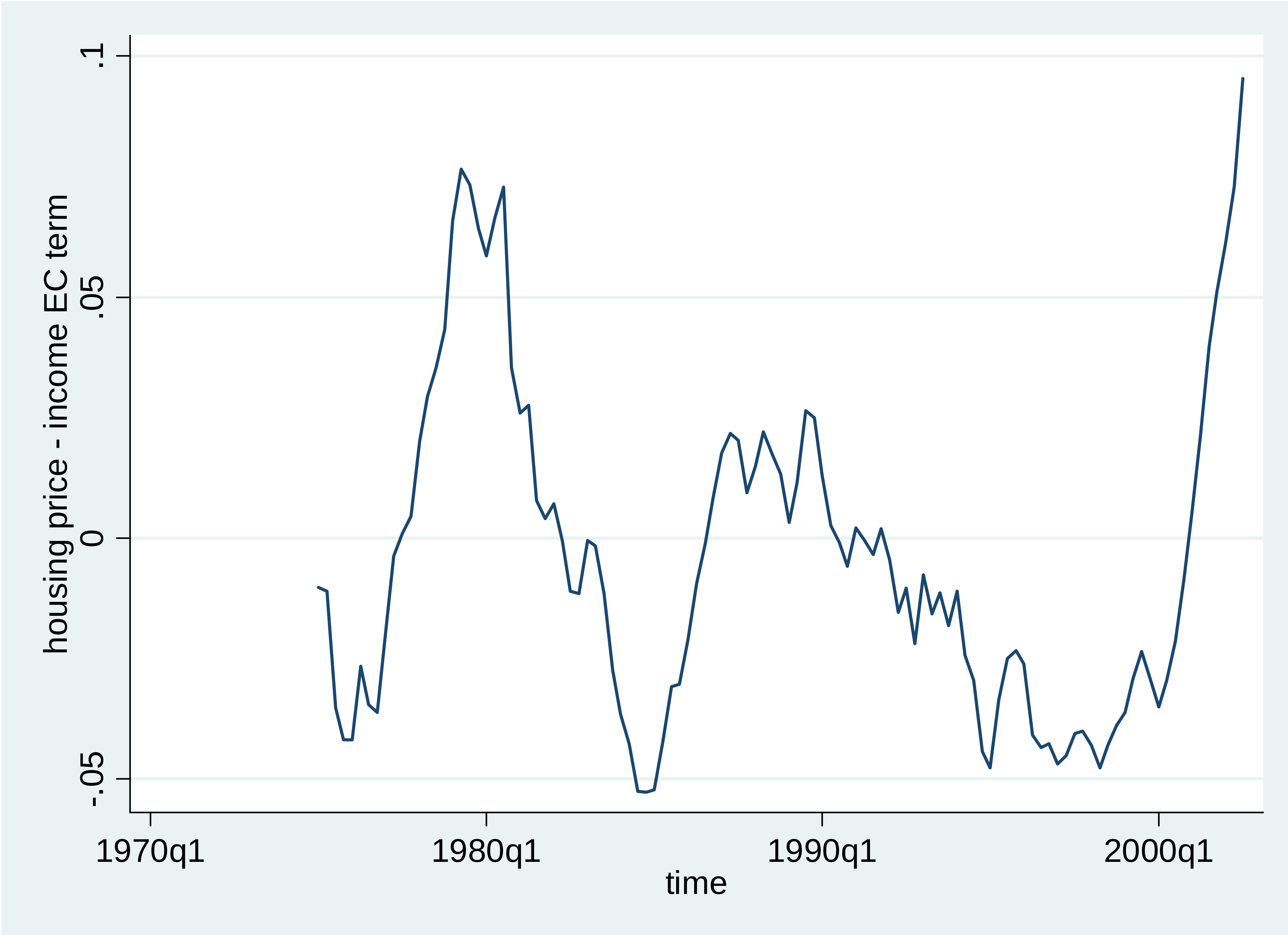


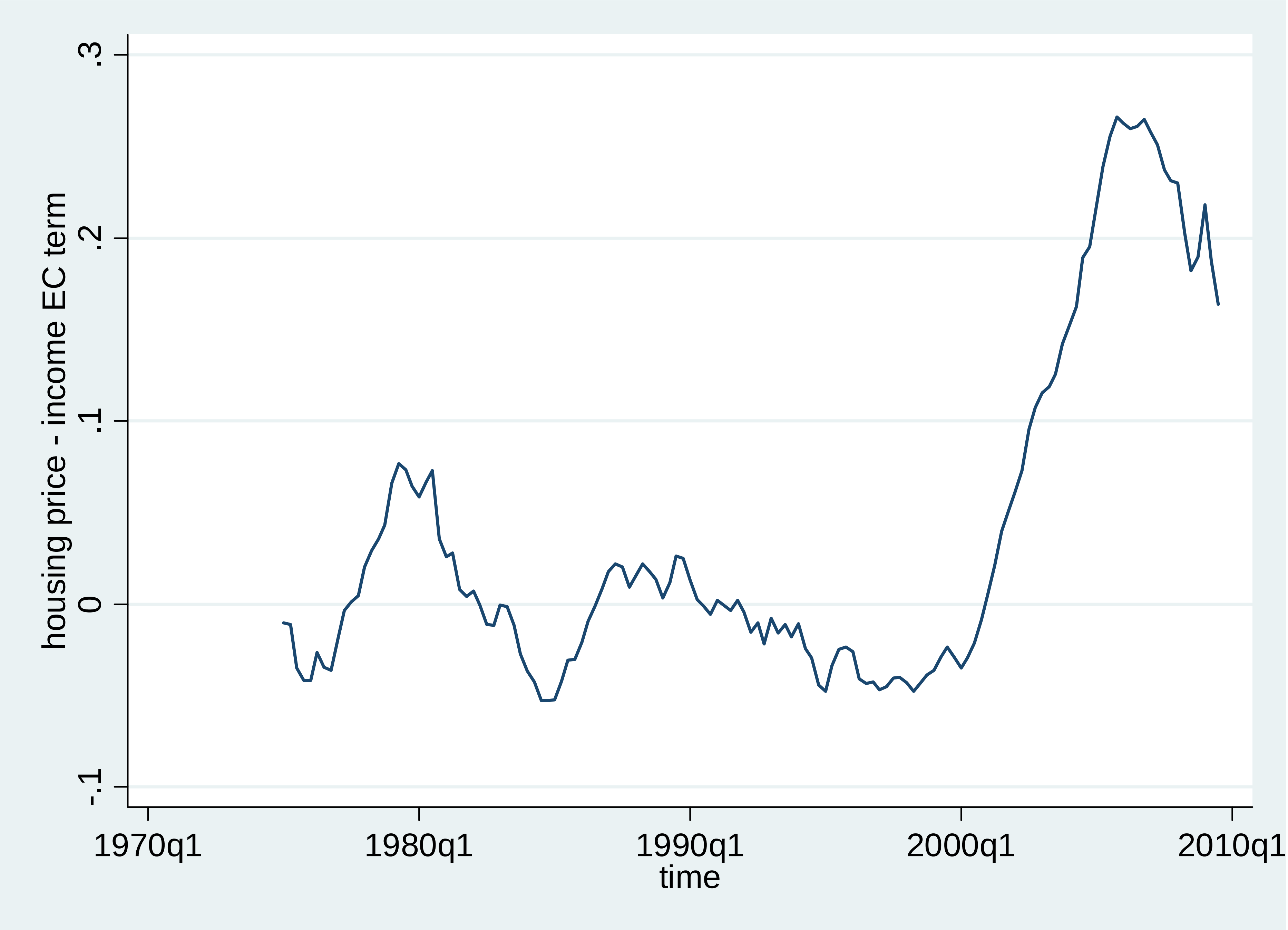
.15
.1
.05
0
-.05
ln(c/y)
9.1
9
8.9
8.8
8.7
8.6
ln(a/y)
1950q1
1960q1
1970q1
1980q1
1990q1
2000q1
2010q1
time
ln(c/y)
ln(a/y)


.1
.05
0
-.05
cay
1950q1
1960q1
1970q1
1980q1
1990q1
2000q1
2010q1
time


.1
.05
0
-.05
cay
2000q1
2002q3
2005q1
2007q3
2010q1
time

# (2) Housing values and median income


housing price - income EC term
.1
.05
0
-.05
1970q1
1980q1
1990q1
2000q1
time


housing price - income EC term
-.1
0
.1
.2
.3
1970q1
1980q1
1990q1
2000q1
2010q1
time


SMU