Teaching Undergraduate Econometrics

David F. Hendry

Department of Economics, Oxford University

ESRC Funded under Grant RES-062-23-0061

Introduction

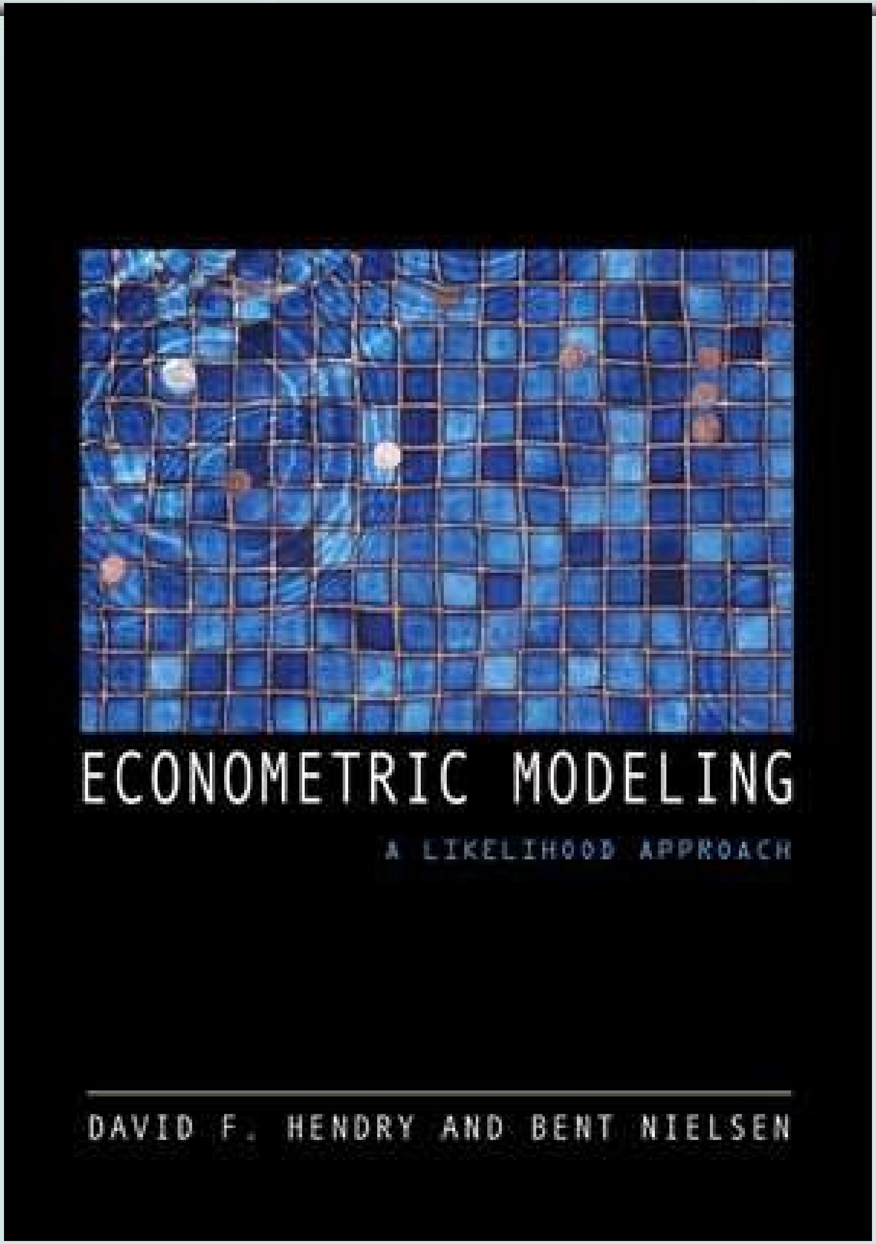
Perfect time to review teaching undergraduate econometrics.

(1) Massive changes in coverage, approach and methods from **maths on blackboards in the 1970s.**
Huge improvements in:
hardware, software, data, graphics and methods.

(2) Almost 25 years since first version of PcGive and 21 years since Hendry (1986)–computer teaching of econometrics via PcGive.

(3) Now use Hendry and Nielsen (2007) as the textbook:
based on PcGive (Doornik and Hendry, 2006)
within OxMetrics (Doornik, 2006)–
half taught in IT room with continuous computer access.

Econometric Modelling



PcGive

Doornik and Hendry (2006) discuss computer-based teaching of econometrics at all levels from very elementary, through intermediate to advanced, loosely based on Hendry (1986).

Cannot cover all the angles here: will rapidly describe early steps in econometrics, commenting on little tricks that help keep student interest.

Then turn to model selection in non-stationary data –yes, for undergraduates!

PPE does not attract most maths oriented:
yet we raised option enrollment from 6 to 48 over 8 years.
Seek to excite student interest.

Background

Assume no elementary statistical theory known.
Probability; distributional concepts; location and spread; elementary notions of randomness and distributions of statistics all needed.

Need to make econometrics exciting, yet comprehensible.

Five central themes:
likelihood;
relevant empirical applications;
mastery of software to implement;
emphasize graphics;
rigorous evaluation of 'findings'.

Introduce derivations as following from need to understand properties: upskill maths to understand empirics, concepts, formulations, and interpretations.

Theory first steps

How do we advance in 16 weeks from IID binary models to selecting cointegrated equations in the face of structural breaks?

First steps introduce elementary statistical theory:
binary events in a Bernoulli model with random draws.
Explain sample and population distributions;
then distribution functions and densities.

Leads to inference in the Bernoulli model,
discuss expectations and variances, and
introduce asymptotic theory and inference.

Next introduce continuous variables (wages)
and regression via simplest case: $y_i = \beta + u_i$.
Builds on estimation of mean, yet leads to logit regression;
and on to bivariate regression models.

Empirical first steps

Simultaneously, teach OxMetrics & conduct empirical work.

Large bank of long historical data series:
offer students choice of modelling one of U_r; Δp; $w - p$; gdp.

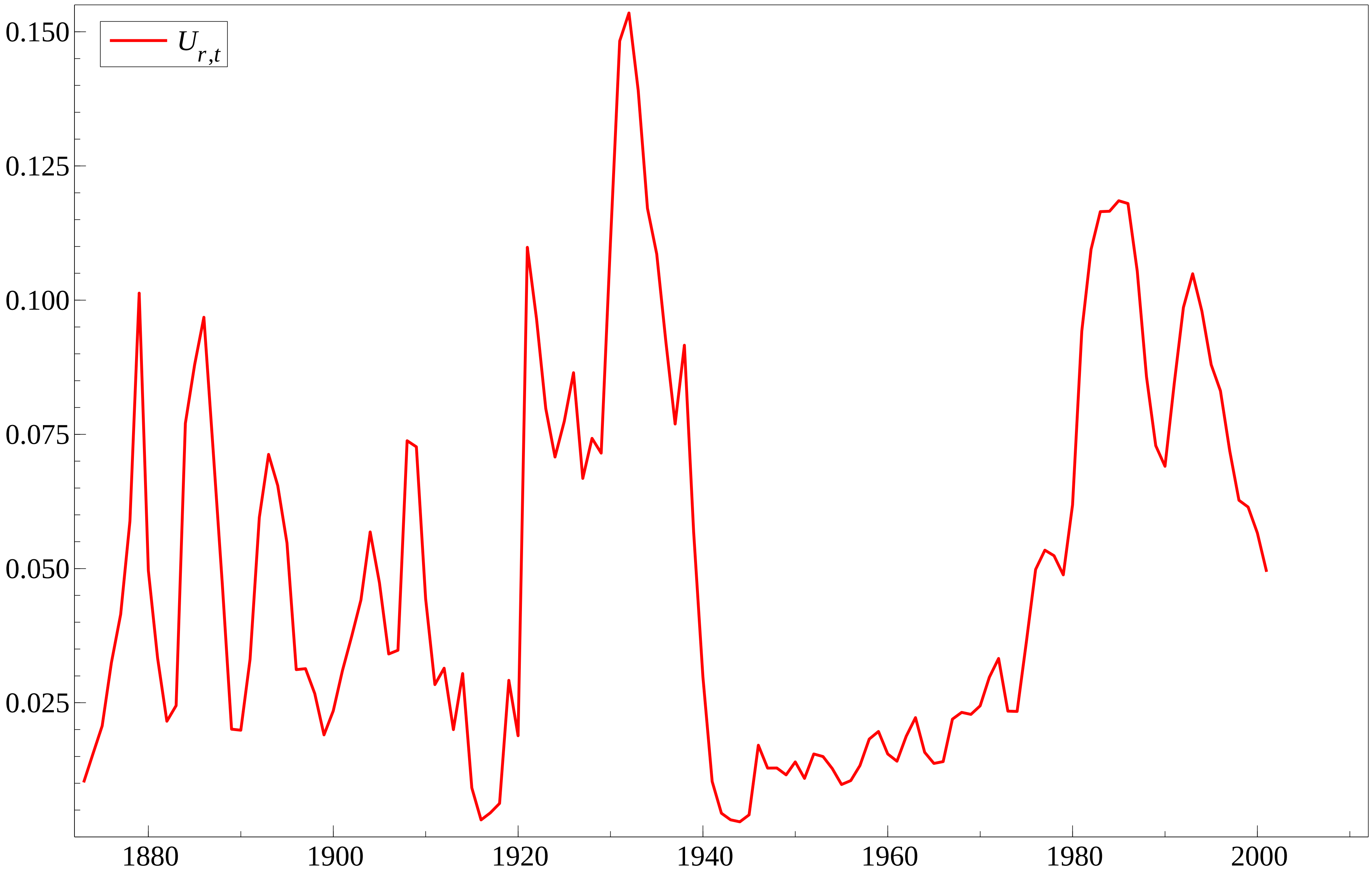
First graph levels: **Assume choose** U_r–see figure 8.

Essential to explain graphs–amazingly poor prior skills.
Discuss axes, units, data transforms including role of logs and their properties.

Then salient features: major events, cycles, trends, and breaks, leading to concept of non-stationarity.

Aim for students who can read published empirical findings, and sensibly conduct and interpret their own empirical research–so cannot finesse difficulties such as non-stationarity and model selection.

Graph of unemployment rate



Empirical second steps

Relate to their life prospects; and those of fellow citizens.

Everyone must make a new comment–however simple–about some aspect, every time.
Many weeks till axes are mentioned first by anyone!

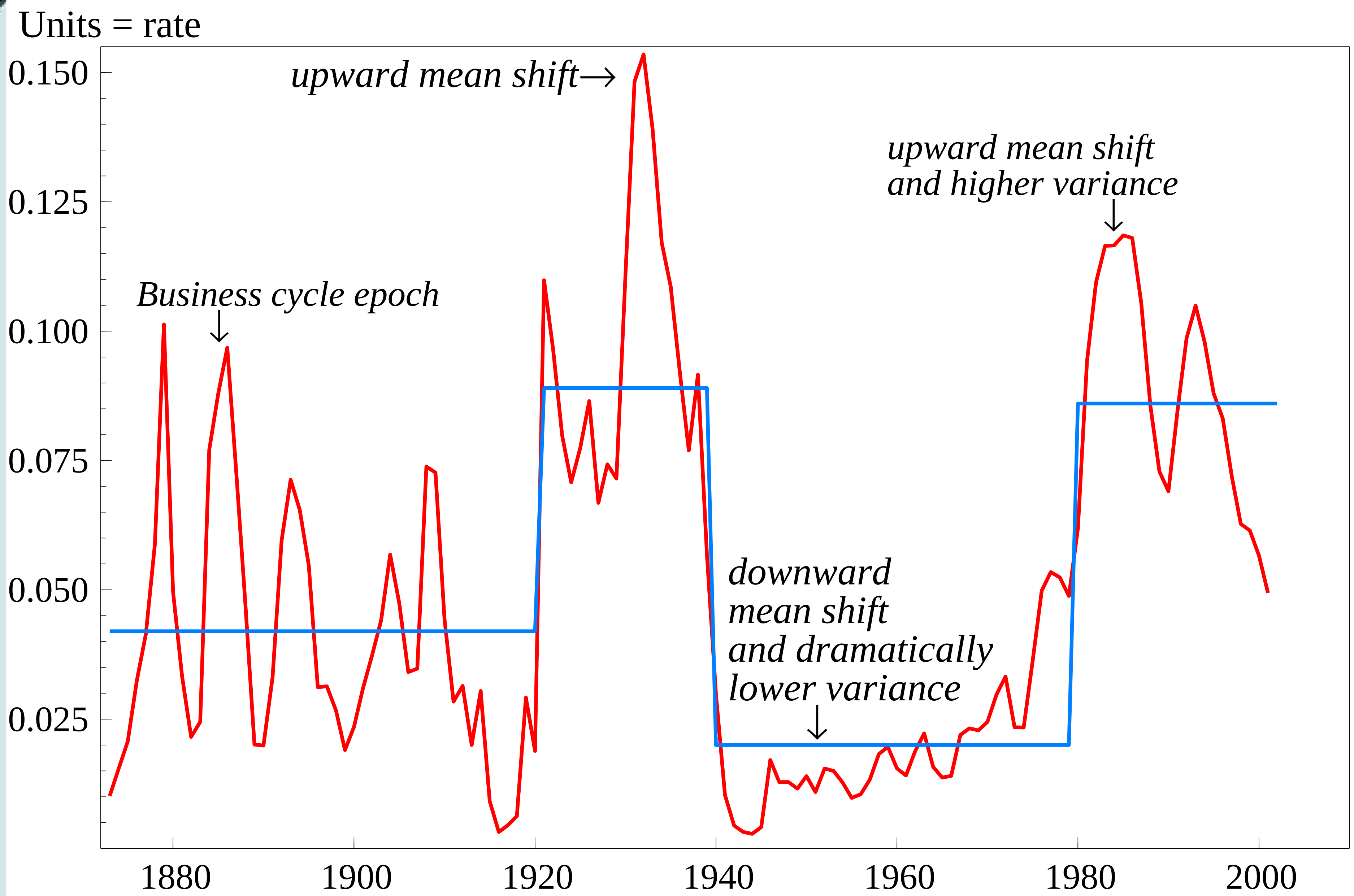
Emphasize manifest non-stationarity.

Figure 10 shows possible *general* comments:
and fig. 11 adds **specific**.

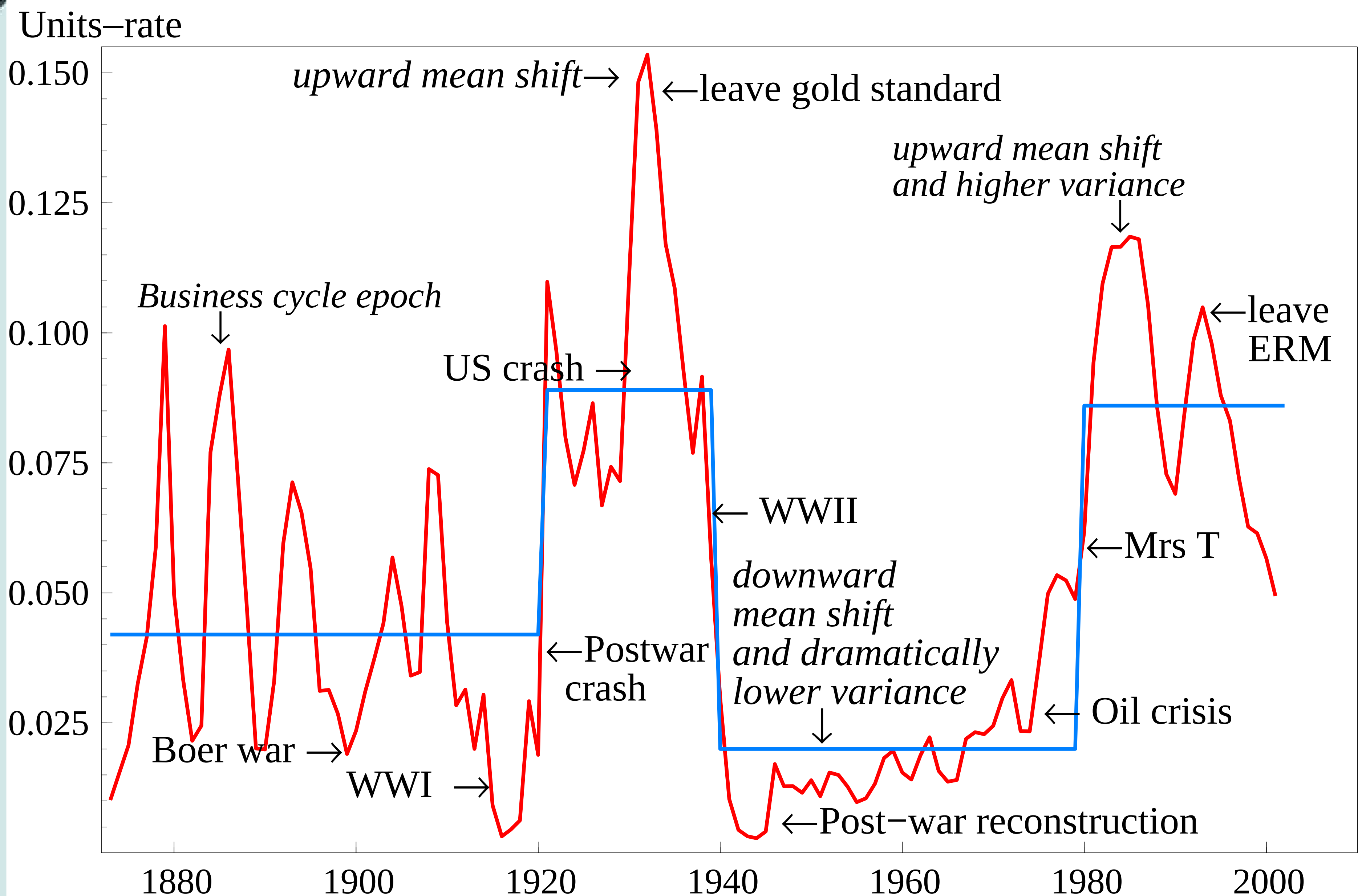
Now discuss economic theory and institutions for chosen series.

Next plot differences graph–discuss: hugely different 'look'.

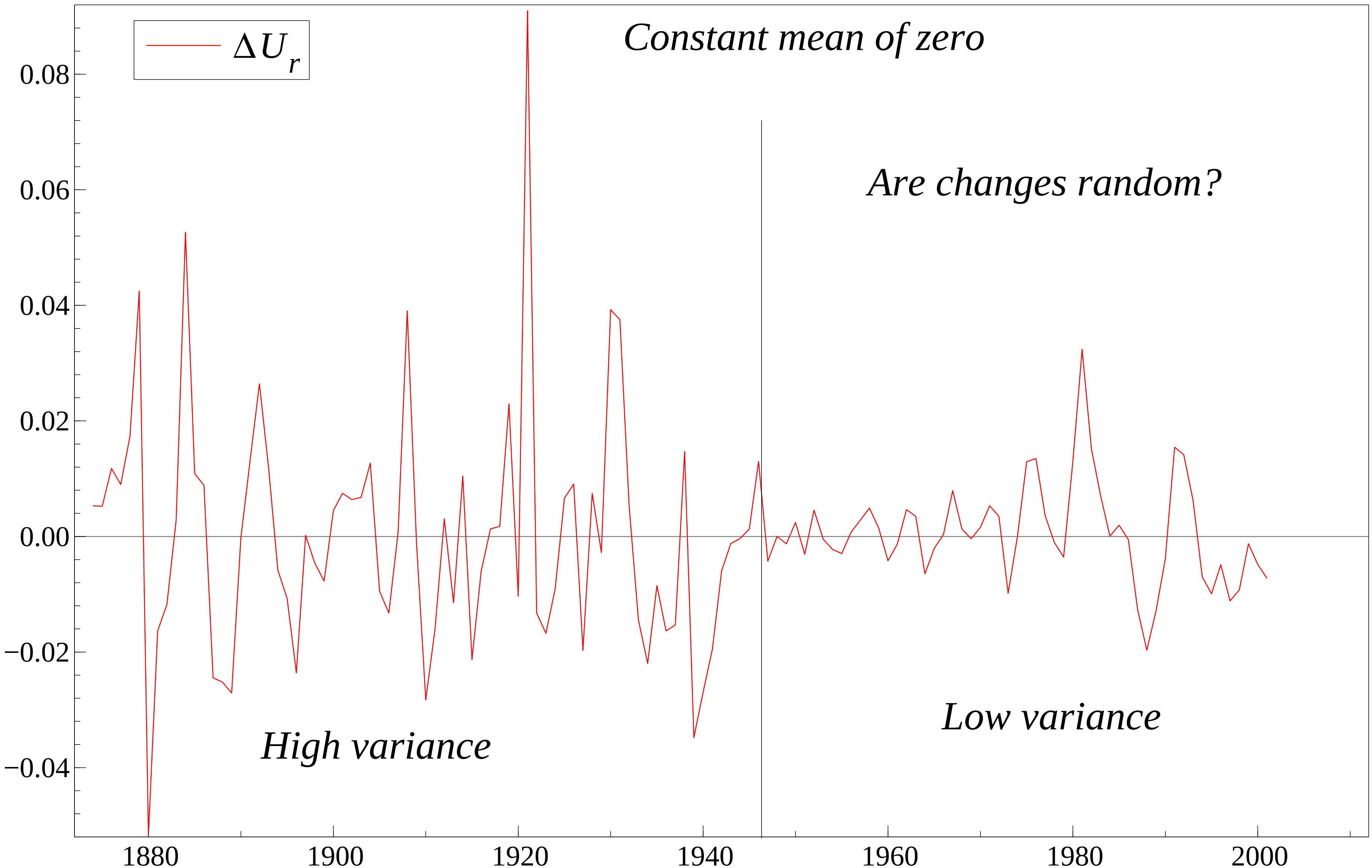
Comments about unemployment rate



All comments about unemployment rate



Changes in unemployment rate



Theory and evidence

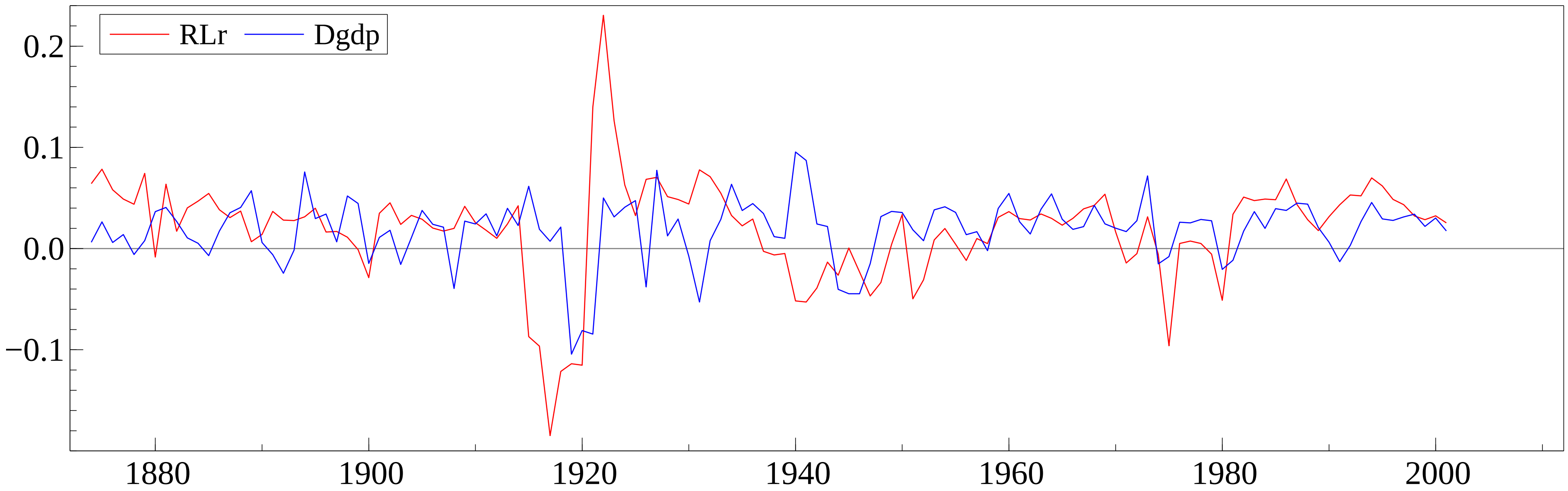
Relate distributional assumptions to model formulation: conditioning in a bivariate normal=>linear regression.

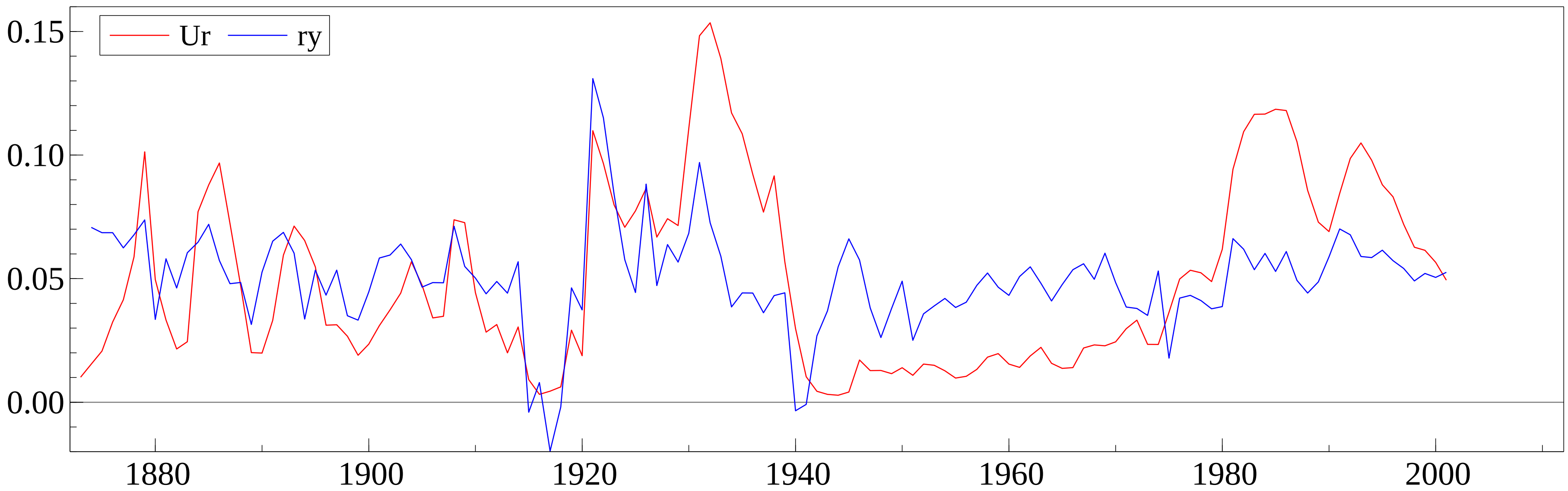
Leads naturally to interpretation of models; modelling and model design; how to judge a model; and hence concepts of **congruence and encompassing**.

Postulate simple model of ‘golden-growth’ to explain deviations of unemployment from historical mean. Measure steady-state equilibrium path by:
$R_{L,t} - \Delta \ln P_t - \Delta \ln Y_t = ry_t$.

Look at graphs of each and their properties.

Graphs for $U_{r,t}$ and ry_t





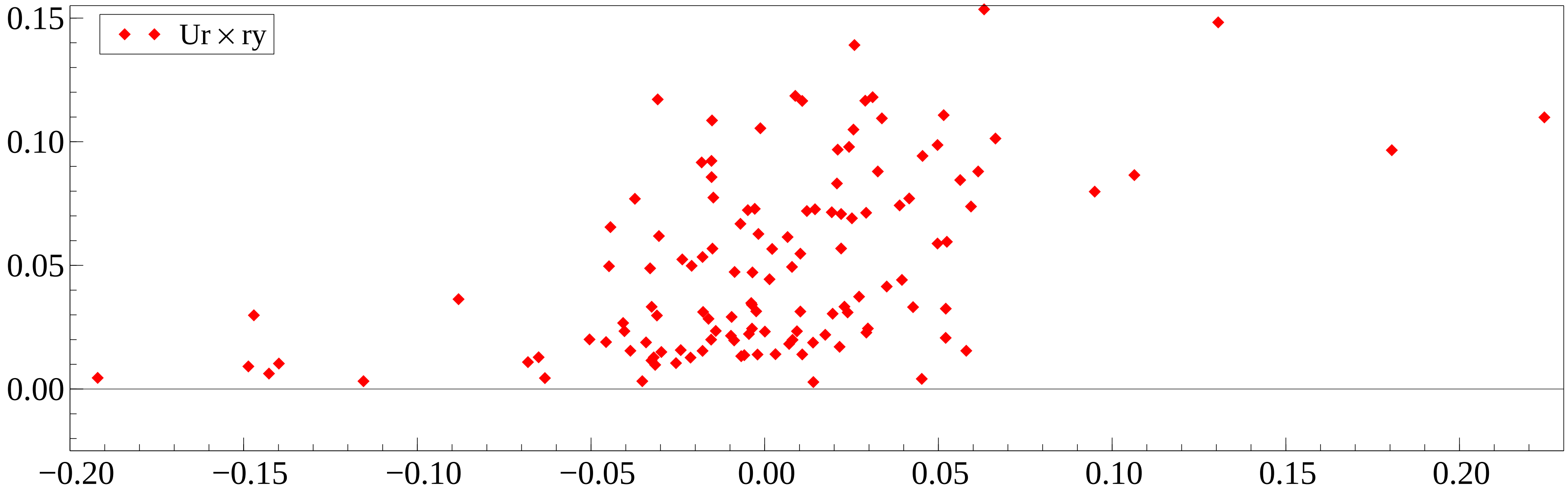
Regression concepts

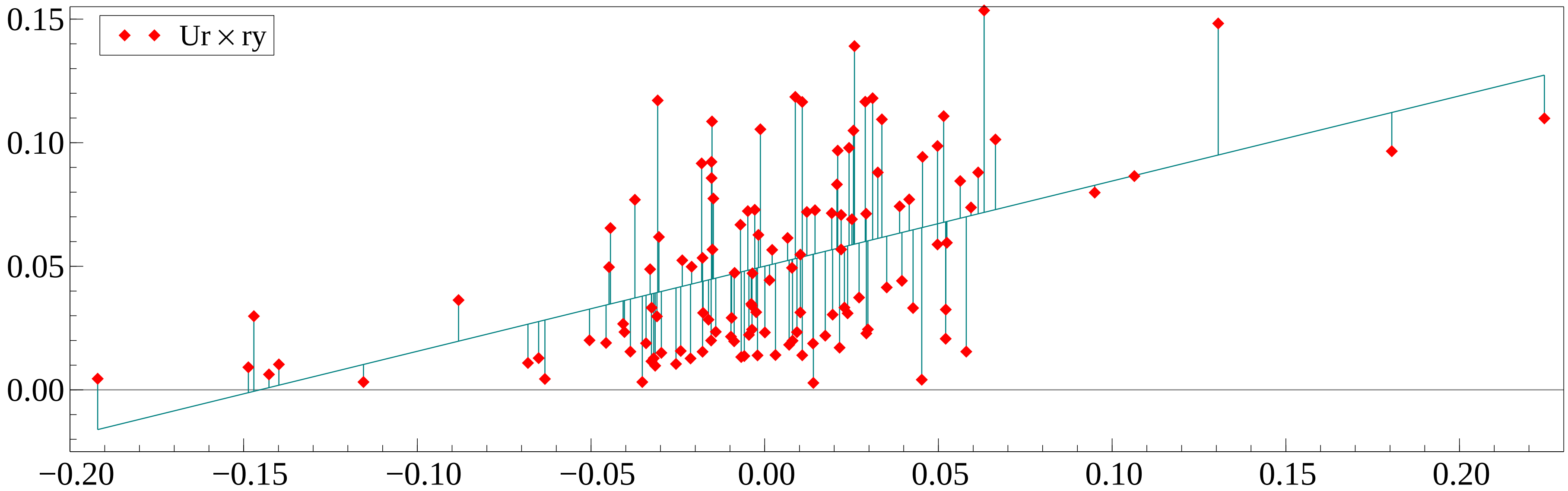
Once econometric theory explained, do a regression plot; add projections–discuss least squares: see fig. 25. Can also explain slope & intercept graphically.

Several key concepts underpin regression: exogeneity; IID errors; normality; functional form; parameter constancy.
All are aspects of model (mis)specification.

Relate maths derivations to graphs as needed to understand properties– but always same principles:
data→suggests putative DGP→model of PDF→
Likelihood function→maximize it→ get statistic→
then its distribution→apply to data→interpret→evaluate.

Regression for $U_{r,t}$ on ry_t





The way ahead

On theory side, reinterpret estimation of a mean as regression on an intercept.

Leads to regression in general;
apply to fish market data **as if cross-section**,
then will reinterpret as time series.

Next treat price and quantity as a system,
leading to identification and simultaneity,
using ‘structural breaks’ as instruments;
and on to cointegration;
picking up model selection en route.

Thus, each topic segues smoothly into the next.

Regression as 'non-parametric'

Can see relation is OK in tails; 'erratic' in middle.

Next write signature and run a regression through it!

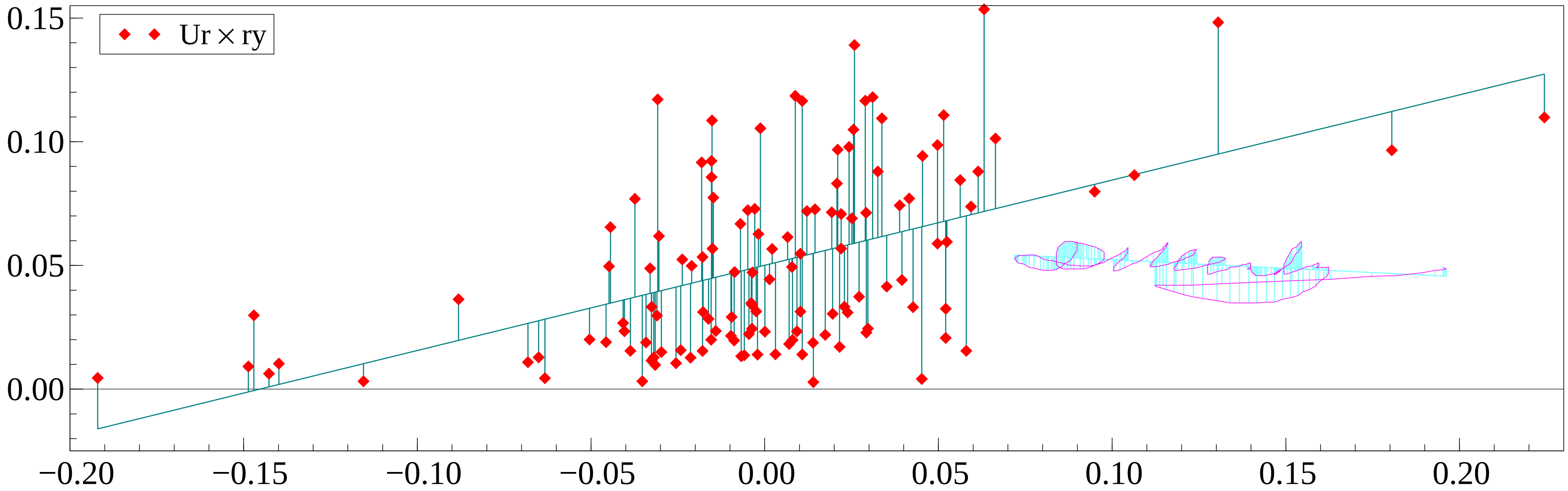
Explain pixels$\rightarrow$world coordinates**=>**data;
so can regress. See fig. 19.
Helps 'demystify' regression analysis: just line fitting.

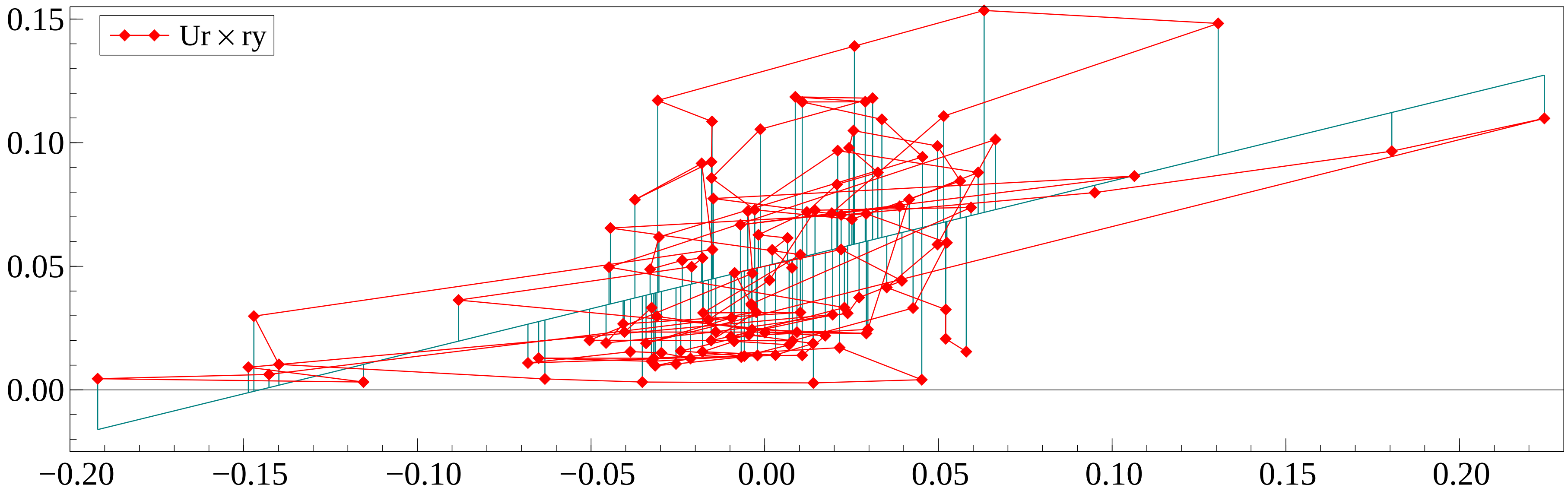
Could join points: "Phillips' loops", leading to dynamics.

Many routes possible–concepts, models, evaluation.
Here will do several sequential regressions–
leads to recursive methods for parameter constancy:
fig. 20.

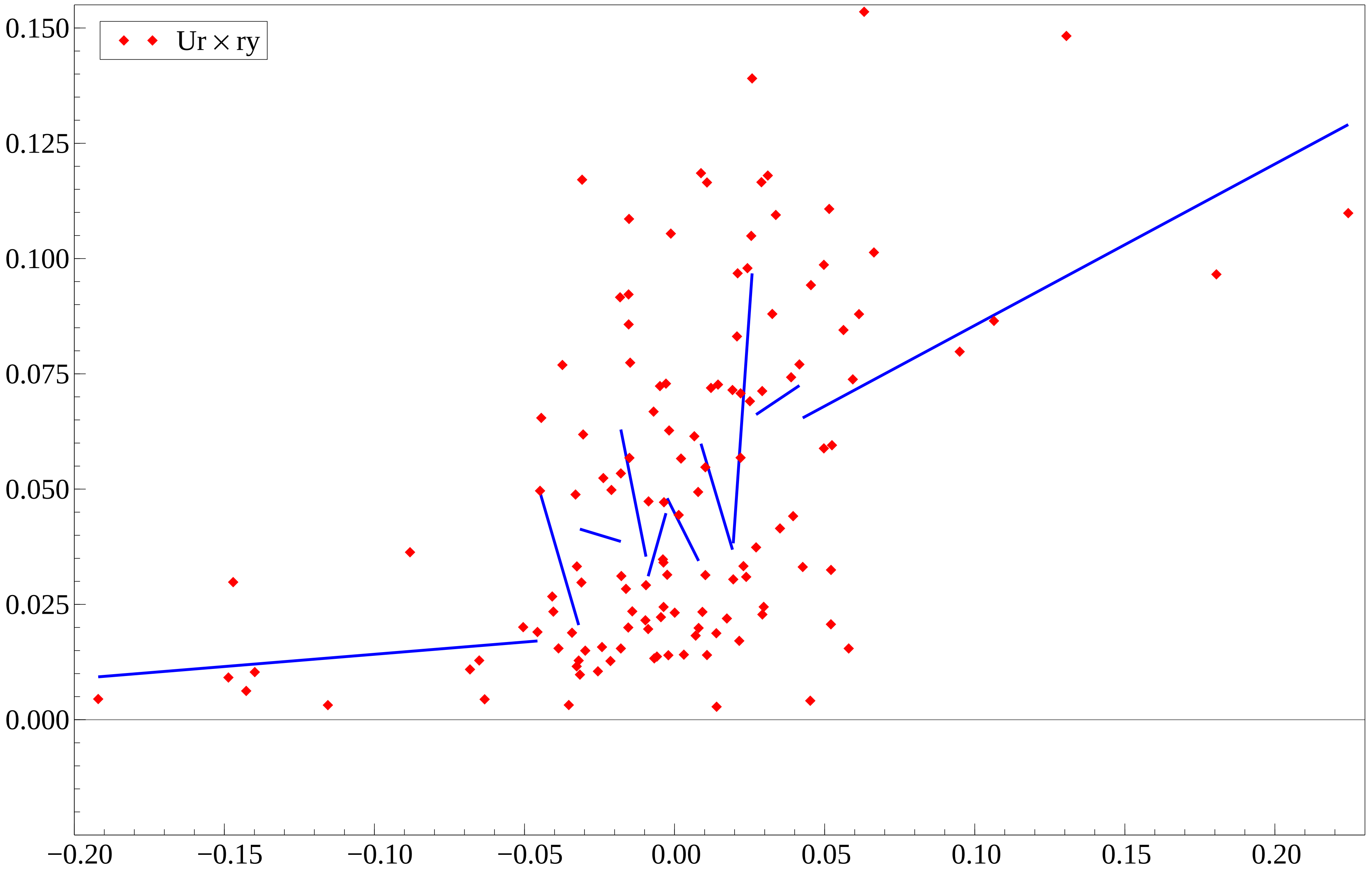
**Note–graphs give opportunity to introduce LaTex:
invaluable later as models can be output that way**.

Regression for a signature





Ten regressions for U_r



Distributions

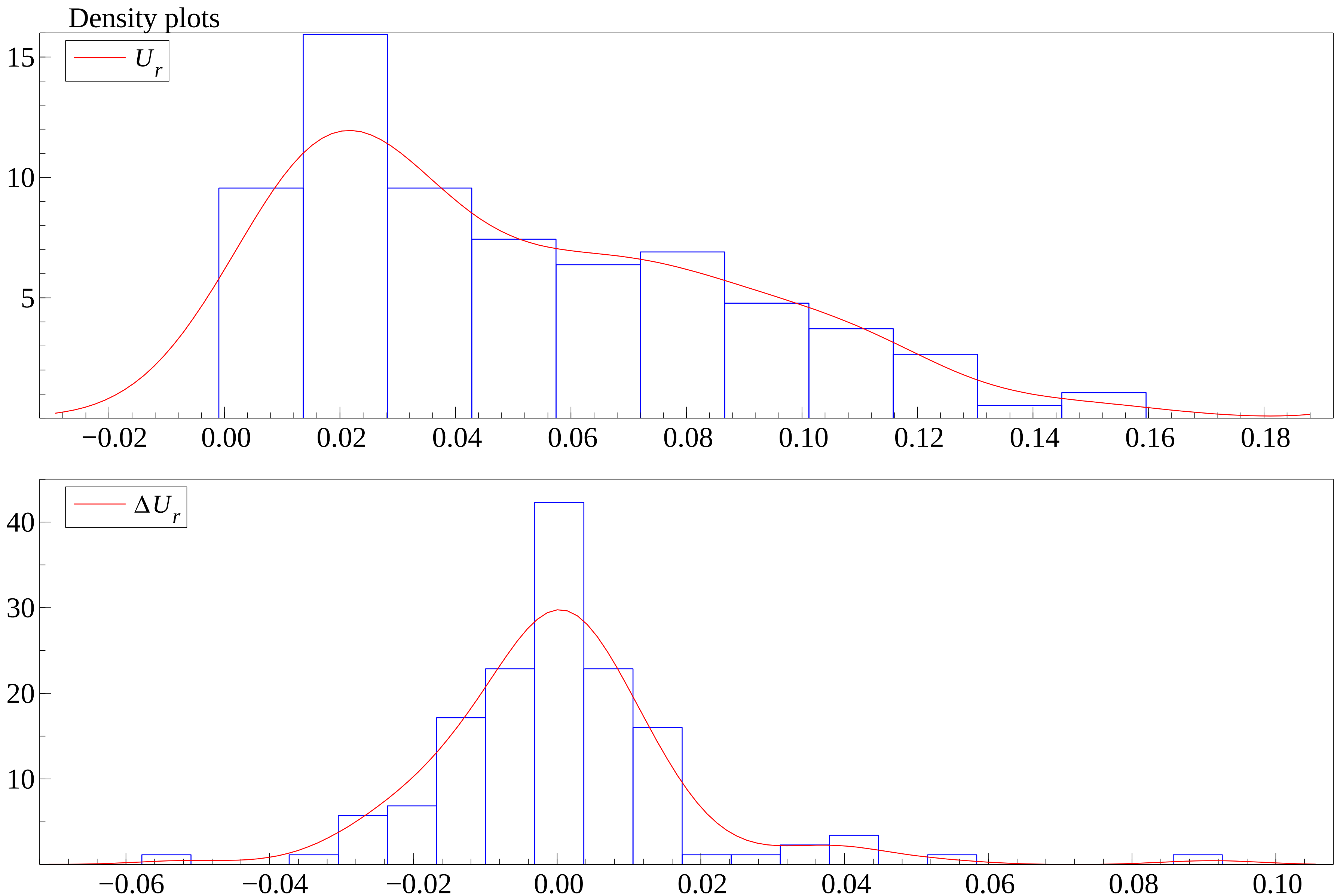
Also can plot histogram and interpolated density: fig. 22. Could use to explain non-parametric/kernel approaches.

Emphasize the very different features:
$U_{r,t}$ like a uniform–many values similarly likely.

$\Delta U_{r,t}$ closer to normal with some outliers.

Thus, differencing changes distibutional shape–
explain as unconditional versus conditional on previous value, so latter is deviation from past.

Distributions for unemployment rate



Time series and randomness

Additional key concept of randomness–can explain here.

Regression subsumes correlation, so can explain correlograms as correlations between successively longer lagged values. See fig. 24.

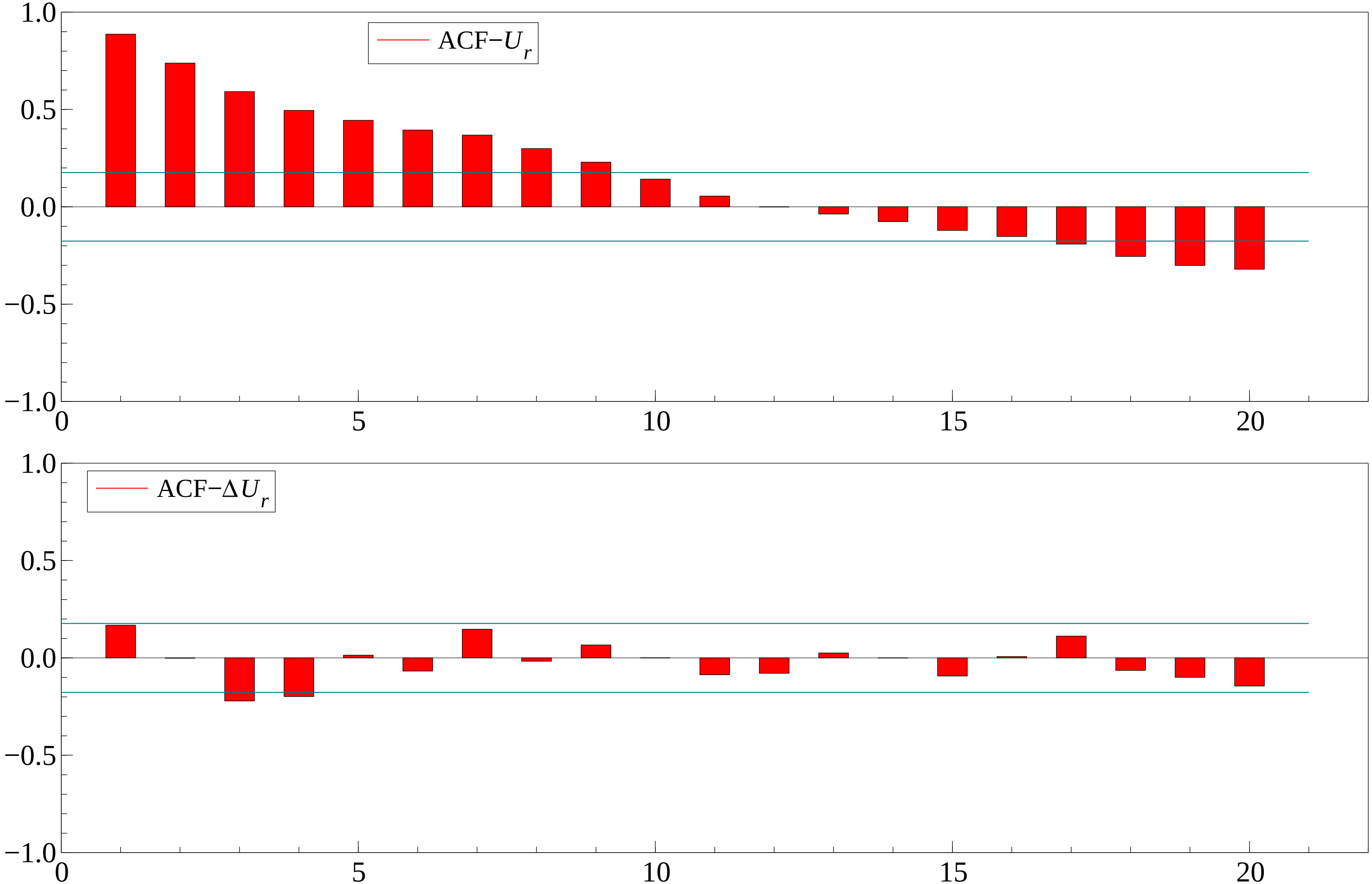
$U_{r,t}$ has many high correlations; almost a trend;
$\Delta U_{r,t}$ has almost no autocorrelation:
changes in $U_{r,t}$ are ‘surprise’ like:
again unconditional versus conditional.

Exploit U_r; ΔU_r comparison: see fig. 25.

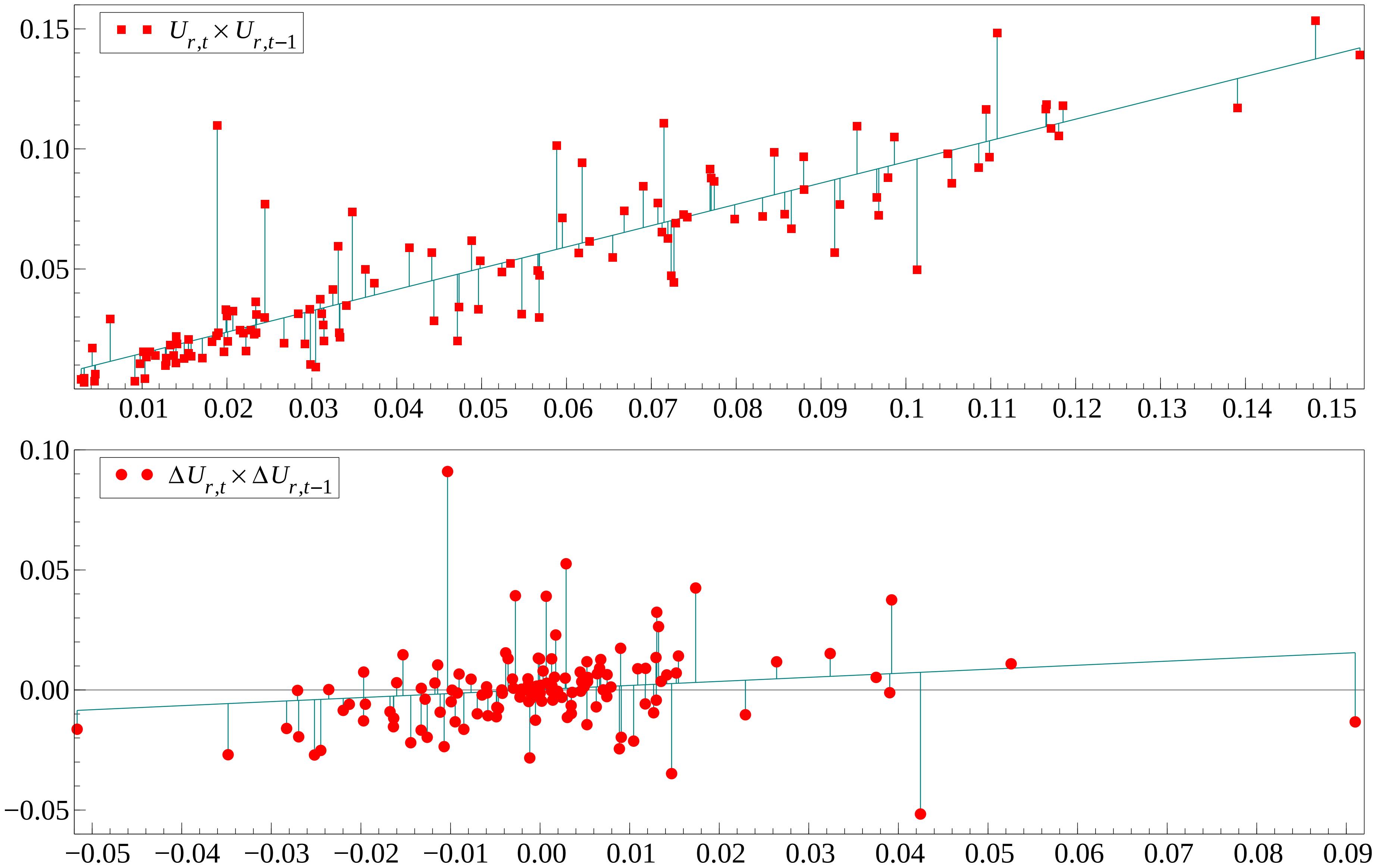
Explain congruence as needing all properties of variables matching in a model.

Now ready for formal estimation of the graph line.

Correlograms for unemployment rate



Unemployment regression on lag



On to model estimation

Long-run (later cointegrated) relation: $U_{r,t} = \beta_0 + \beta_1 ry_t$.

$$\widehat{U}_{r,t} = \underset{(0.0028)}{0.0501} + \underset{(0.052)}{0.345}\ ry_t \tag{1}$$

$\widehat{\sigma} = 0.0315$ $\mathsf{R}^2 = 0.26$ $\mathsf{F}_{\mathsf{GUM}}(1, 126) = 44.64^{**}$

Unemployment rises/falls as real long-run interest rate is above/below real growth rate (i.e., $ry_t \lesseqgtr 0$).

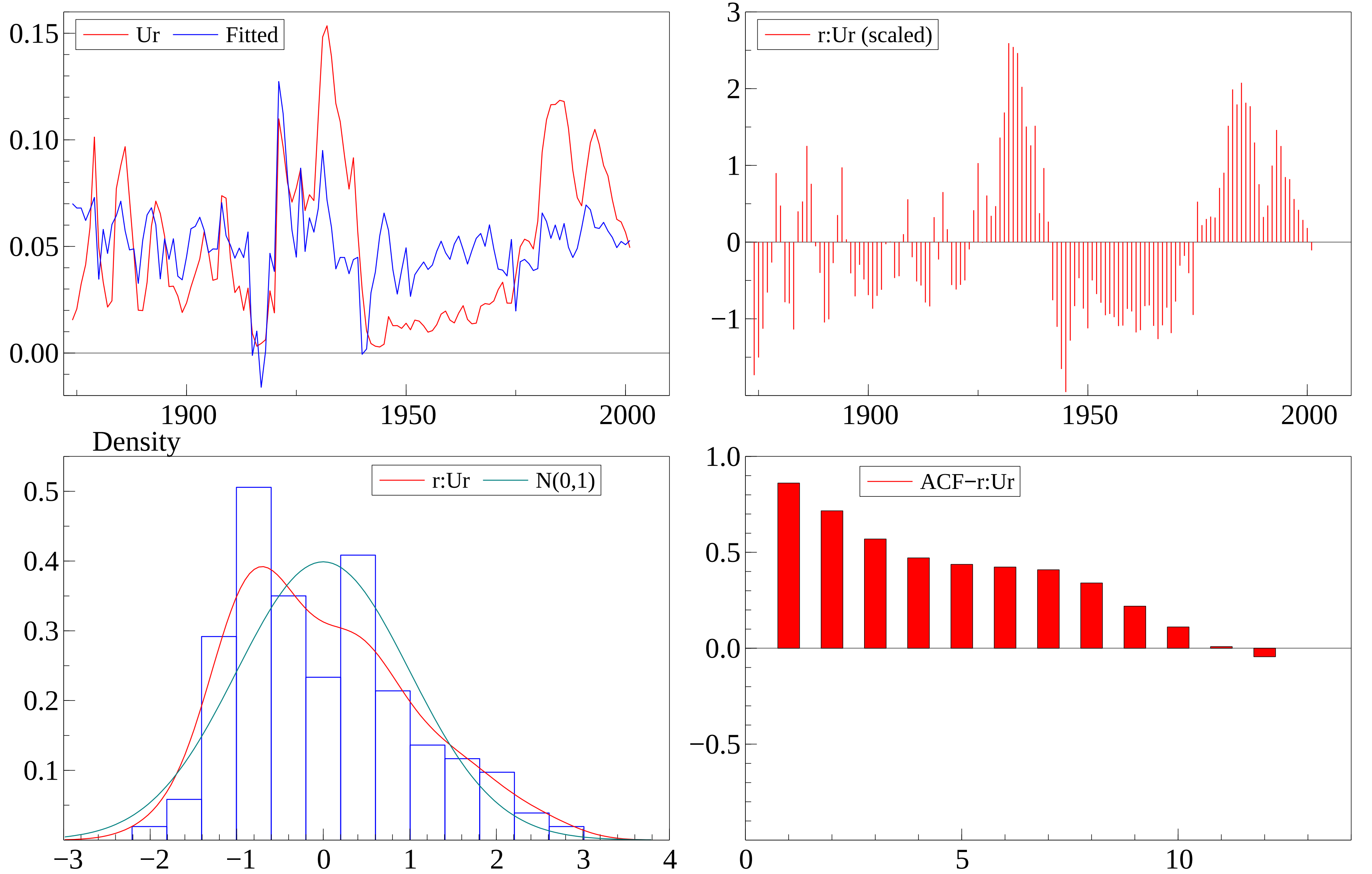
Explain each measure and its invariance;
stress assumptions–easily tested:

$\mathsf{F}_{\mathsf{ar}}(2, 124) = 180.4^{**}$ $\mathsf{F}_{\mathsf{arch}} = 229.9^{**}$ $\chi^2_{nd}(2) = 15.02^{**}$
$\mathsf{F}_{\mathsf{het}} = (2, 123) = 2.62$ $\mathsf{F}_{\mathsf{reset}} = (1, 125) = 0.33$.

Explain each, and stress must do for every formulation.

$\widehat{U}_{r,t}$ graphical output visually confirms the tests:
thus, successfully applied key concepts to *residuals*.

U_r on ry graphical output



Simple dynamic models

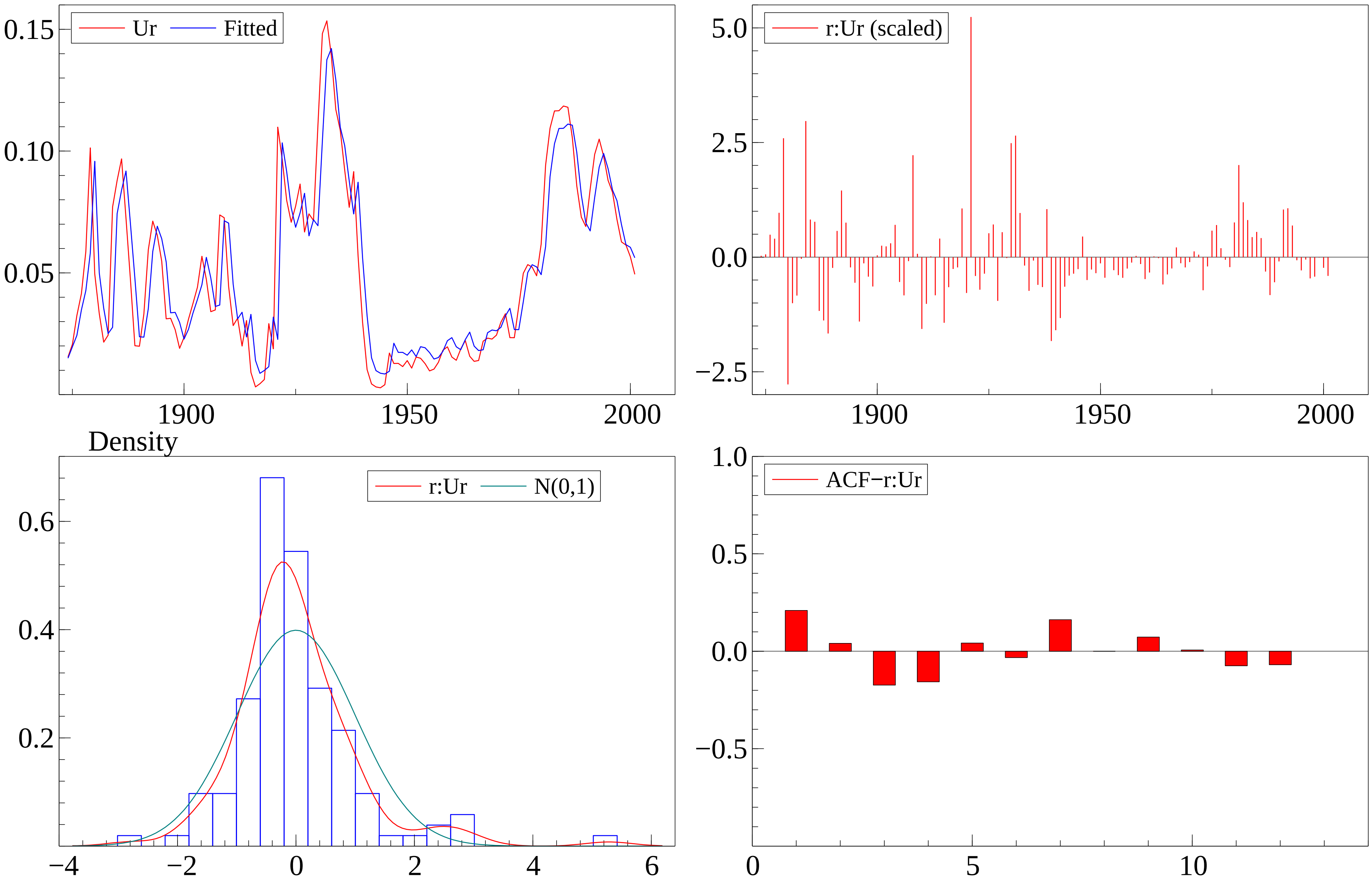
Explain multiple testing concepts now:
each test derived under separate null;
any other rejections contradict assumptions of derivations.
Clear that model is badly mis-specified:
not clear which assumption(s) is invalid,
hence no obvious 'solution'–leading to general-to-simple.

Other simple model of $U_{r,t}$ on $U_{r,t-1}$; $U_{r,t} = \gamma_0 + \gamma_1 U_{r,t-1} + \epsilon_t$
follows from graph 25–and can also relate to ΔU_r.

$$\widehat{U}_{r,t} = \underset{(0.040)}{0.887} \; U_{r,t-1} + \underset{(0.003)}{0.006}$$

$\widehat{\sigma} = 0.017$ $\mathsf{R}^2 = 0.79$ $\mathsf{F}_{\mathsf{GUM}} 1, 126) = 485.7^{**}$
$\mathsf{F}_{\mathsf{ar}}(2, 124) = 3.8^{*}$ $\mathsf{F}_{\mathsf{arch}}(2, 124) = 0.55$ $\chi^2_{nd}(2) = 33.0^{**}$
$\mathsf{F}_{\mathsf{het}} = (2, 123) = 0.42$ $\mathsf{F}_{\mathsf{reset}} = (1, 125) = 0.01$.
Improved, but still mis-specified–obvious outlier in 1920.

$U_{r,t}$ on $U_{r,t-1}$ graphical output



More general models

Discuss lags; measures of lag responses; etc.
Long-run solution is 5.3% unemployment; cannot reject unit root, so explain rudiments of stochastic trends...

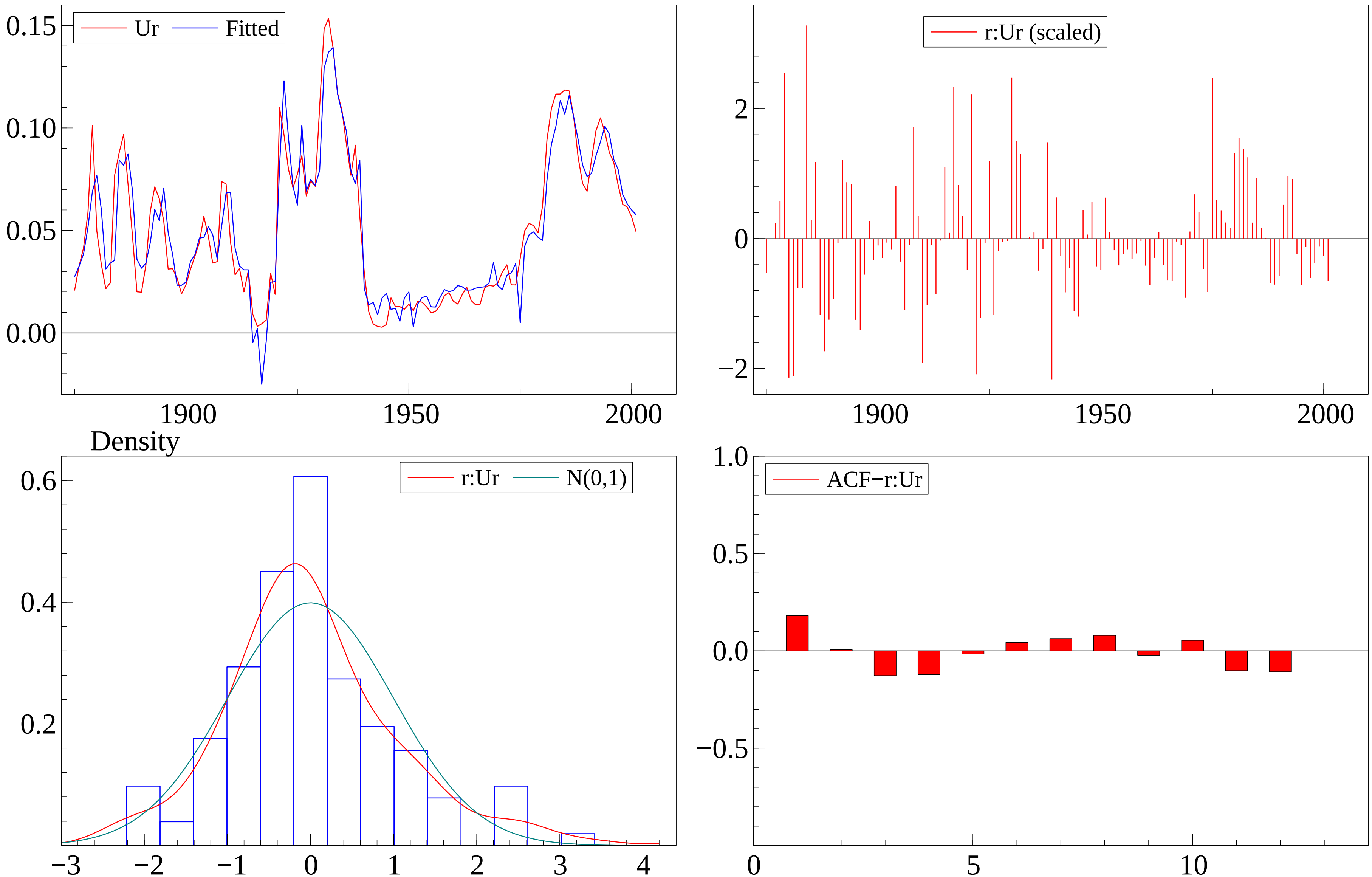
Now move to a dynamic model with regressors:
$U_{r,t} = \beta_0 + \beta_1 ry_t + \beta_2 U_{r,t-1} + \beta_3 ry_{t-1}$.

$$\widehat{U}_{r,t} = \underset{(0.04)}{0.86}\, U_{r,t-1} + \underset{(0.002)}{0.007} + \underset{(0.03)}{0.24}\, ry_t - \underset{(0.03)}{0.10}\, ry_{t-1}$$

$$\begin{aligned} &\widehat{\sigma} = 0.013 \;\; \mathsf{R}^2 = 0.88 \;\; \mathsf{F}_{\mathsf{GUM}}(3,123) = 308.2^{**} \\ &\mathsf{F}_{\mathsf{ar}}(2,121) = 2.5 \quad \mathsf{F}_{\mathsf{arch}}(1,121) = 3.1 \quad \chi^2_{nd}(2) = 7.2^{*} \\ &\mathsf{F}_{\mathsf{het}} = (6,116) = 4.2^{**} \quad \mathsf{F}_{\mathsf{reset}} = (1,122) = 4.2^{*}. \end{aligned} \qquad (2)$$

Not fully congruent, but much better–**progressive research**. Can reject a unit root, so ‘cointegrated’: long-run is $U_r = 0.052 + 1.02ry$.
Contrast to coefficient of 0.35 in (1)–former down biased.

$U_{r,t}$ model graphical output



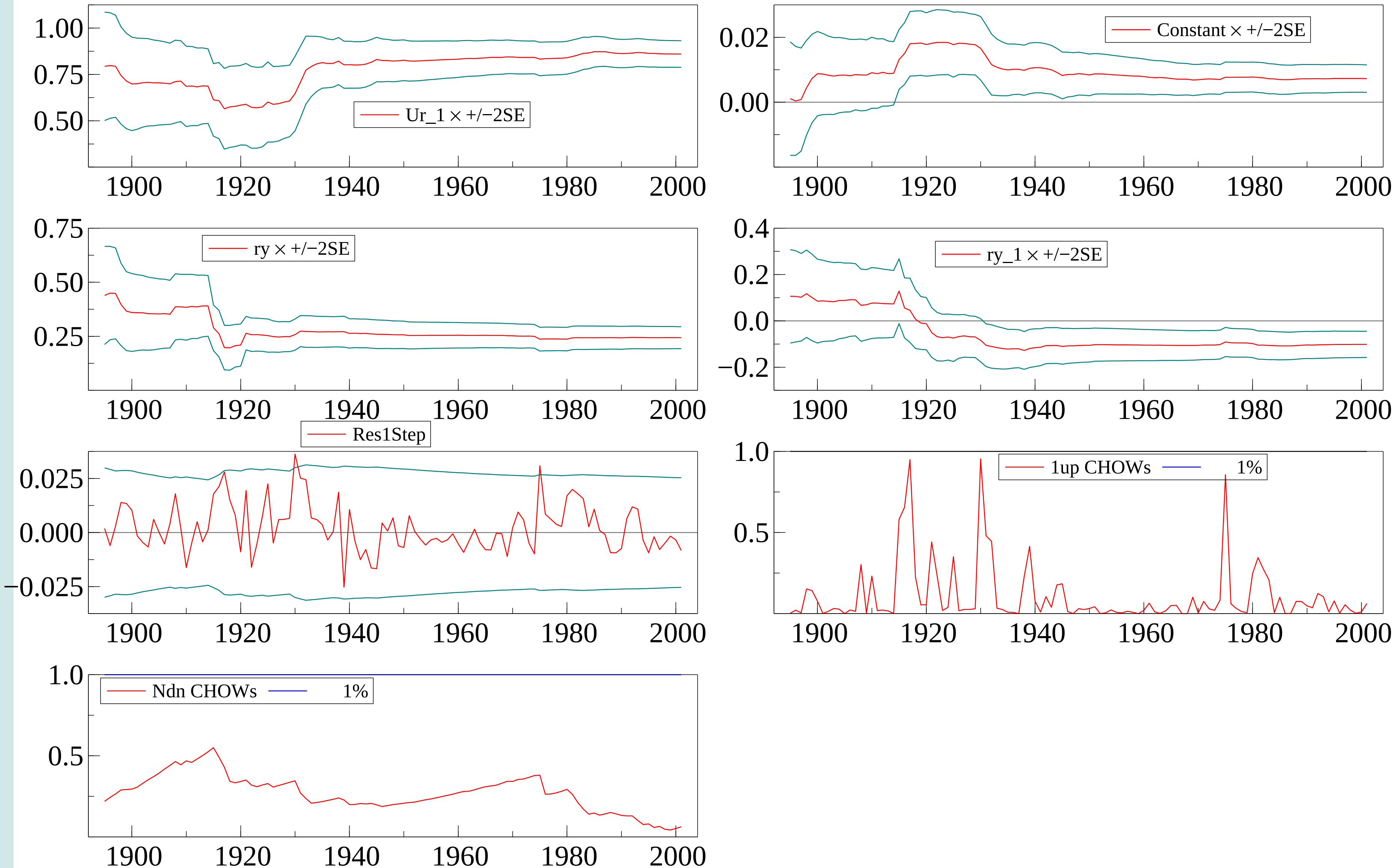
Extensions

Now confront constancy by formal recursive methods, building on earlier graphs.
Figure 33 records: tests do not reject, despite ‘wandering’ estimates–much to discuss as desired.

Few outliers – but some negative fitted values (try logit?). May also alleviate heteroscedasticity...

Challenge students to formulate alternative explanations: test their proposals against the evidence–and (2)!

$U_{r,t}$ model recursive output



Model selection

Seen dangers of simple approaches:
so general-to-specific needs explained.
In a general model, cannot know in advance which variables will matter: some will, but some will not.

Must confront for competent practitioners:
any test+decision entails selection, so ubiquitous.

Sketch theory of model selection in simplest case:
2 decisions (keep/delete) and 2 states (relevant/irrelevant).

Translate ‘retain irrelevant variables’ & ‘exclude relevant’ as akin to probabilities of type I & II errors.

Understanding model selection

Consider a perfectly orthogonal regression model:

$$y_t = \sum_{i=1}^{N} \beta_i z_{i,t} + \epsilon_t \tag{3}$$

where $\mathsf{E}[z_{i,t}z_{j,t}] = \delta_{i,j}\lambda$ and $T >> N$.

Order the N sample t^2-statistics testing $\mathsf{H}_0 : \beta_j = 0$:

$$\mathsf{t}^2_{(N)} \geq \mathsf{t}^2_{(N+1)} \geq \cdots \geq \mathsf{t}^2_{(1)}.$$

Cut-off n between included and excluded variables is:

$$\mathsf{t}^2_{(n)} \geq c_\alpha > \mathsf{t}^2_{(n+1)}.$$

All larger values retained: all others eliminated.
Only one decision needed even for $N = 1000$:
'repeated testing' does not occur.

Path search gives impression of 'repeated testing'.
Confused with selecting from $2^{1000} = 10^{301}$ possible models.

Maintain false null retention at **one variable** by $c_\alpha = 1/N$.

Monte Carlo of model selection

How to persuade?
Everyone generates a PcNaive sample from same DGP which they design as a group,
then all apply Autometrics to their sample.
Pool class results and relate to delete/keep calculations.
Repeat at a looser/tighter selection criterion.

Explain key role of marginal decisions:
empirical t close to critical value is danger zone.

Now they can handle complicated models by automatic modelling; huge improvement in quality of empirical work– and in interpretation of what they find.

Conclusions

Computer-based teaching of econometrics enhances all levels from very elementary, through intermediate to advanced.

Key texts are:
Hendry, D.F. and Nielsen, B. (2007), *Econometric Modeling: A Likelihood Approach*, Princeton University Press; and Doornik, J.A. and Hendry, D.F. (2007), *Empirical Econometric Modelling using PcGive*, Timberlake, which is about to appear.

Undergraduates can progress from binary events in a Bernoulli model with random draws to model selection in non-stationary data in a year-long course.

And learn to build empirical models of non-stationary data by automatic modelling.

References

Doornik, J. A. (2006). *An Introduction to OxMetrics*. London: Timberlake Consultants Press.

Doornik, J. A. (2007). Autometrics. Working paper, Economics Department, University of Oxford.

Doornik, J. A., and Hendry, D. F. (2006). *Empirical Econometric Modelling using PcGive: Volume I*. London: Timberlake Consultants Press.

Hendry, D. F. (1986). Using PC-GIVE in econometrics teaching. *Oxford Bulletin of Economics and Statistics*, **48**, 87–98.

Hendry, D. F., Johansen, S., and Santos, C. (2004). Selecting a regression saturated by indicators. Unpublished paper, Economics Department, University of Oxford.

Hendry, D. F., and Nielsen, B. (2007). *Econometric Modeling: A Likelihood Approach*. Princeton: Princeton University Press.