Econometric inference with transitional macroeconomic time series data in autoregressive distributed lag models

Dario Cziráky[a,*], Sanja Tisma[b], Anamarija Pisarovic[b]

[a] Department for Macroeconomic Research, IMO
[b] Department for Resource Economics and Environmental Policy, IMO,
2 Farkasa Vukotinovica, 10000 Zagreb, Croatia

Abstract

Transitional macroeconomic time series data are short and generated by DGPs likely to be variant over time with a known "beginning of time". Several assumption needed for standard econometric inference are thus violated. It is shown that problems caused by data properties cannot be solved by imposing distributional assumptions. Subsequently, quarterly data sets still cannot be used for reliable inference with such data. An example is shown where an apparently small change in the length of the estimation sample can cause drastically differing results. However, it is shown that monthly data, if available, might provide solid basis for reliable inference and model building even if the available series are disaggregated proxies. Furthermore, it is argued that due to the nature of transitional DGPs and lack of firm theory of transitional economies, the general-to-specific approach should be preferred even in light of some theoretically appealing results that apparently prefer simple models. In this regard dynamic mis-specification is discussed and illustrated in a simple simulated example.

JEL classification: C22

Keywords: Data generating process; statistical assumptions; data properties; inference; small samples; asymptotic distributions; ADL models

1. Introduction

For a number of transitional countries large and relatively good quality cross sectional data on firms as well as survey data on households and individuals exist which allows modelling primarily static economic relationships. However, most interesting aspects of the transitional processes are dynamic in nature, requiring time series information. Unfortunately, the availability, length, and quality of macroeconomic aggregate time series data for literally all CEE countries until recently prevented extensive empirical research in dynamic macroeconomics. Nevertheless, after a decade of post-Communist experience, aggregate time

[*] Corresponding author. E-mail: dario@irmo.hr

series data for a number of countries became available thereby enabling, at least in principle, empirical econometric modelling.

The availability of time series long enough to estimate usually simple econometric equations does not, however, guarantee validity of econometric inference made on the basis of such estimation. Quite the contrary, many assumptions implicitly made in the modelling process, while possibly valid for standard western data, cannot be justified for typical transitional time series. As solid economic theory for transitional processes does not exist, and as previous empirical evidence is a scarce and unreliable starting point for applied research we essentially must admit rather weak grounds for any a priory assumptions about such data. Consequently, while we emphasise that empirical time series modelling with transitional data is possible and needed, we also stress that mechanical estimation that does not take care of the specifics of transitional data is likely to result in very poor models and unjustifiable inference.

This paper aims to examine the main characteristics of typical transitional aggregate time series focusing on validity of econometric inference arising from estimation and modelling of dynamic macroeconomic relationships. We illustrate the main issues on two empirical examples of modelling aggregate consumption with Croatian monthly quarterly (1994-2000) and monthly (1992-2000) data.

We start by outlining the known properties of the stochastic processes that generated transitional time series data (i.e., data generating process or DGP) focusing on validity of standard statistical assumptions. Some further theoretical issues, such as small sample bias, Rissanen's theorem and dynamic misspecification are analysed and illustrated with simulated and real-empirical data examples.

2. Characteristics of transitional data generating processes

As far as reliable economic theory goes, very little is known about economic forces behind the transitional processes. However, few important conjectures can be made about these processes. In terms of the generally unknown mechanism that jointly or generated relevant macroeconomic variables, i.e., data generation process (DGP) we can conjecture that:

(i) The transitional DGP might not be invariant over time. As one economic system is changing by converging to another, which is what "transition" essentially means, it cannot be assumed that the DGP is i.i.d. across time.

(ii) The “beginning of time” is known, thus $y_0 = 0$ is not a valid assumption. If we wish to study dynamics of transition then the only relevant span of data are the post-1990 observations. As controlled economies (Communist systems) sharply differ from the western market economies, even if longer span of data is available (pre-1990) we cannot treat the transition period simply as a shock to the system (e.g., by controlling it out by dummies). In other words, we cannot expect to extract more information about transition process by using data from another (previous) period.

Furthermore, data issues arising from problems with national accounts data collection procedures, measurement problems and generally problematic quality of macroeconomic time series present additional problems for modelling transitional economies. We, however, limit our attention here to the issues of theoretical relevance regarding properties of analysed stochastic processes while avoiding data collection problems and measurement error issues which further complicate modelling process thereby calling in questions inference made about estimated econometric models thus requiring additional analysis not attempted in this paper.

We will show that (i) and (ii) markedly affect econometric inference in dynamic macroeconomic models of transitional processes by placing extra requirements on the assumptions about data properties such as sample size, constancy of moments and stationary.

3. An empirical example: Aggregate consumption function

We start with a specific empirical example of modelling aggregate relationship between personal consumption and personal disposable income. Estimation of the consumption function is one of the most common macroeconomic research topics (for a review see Deaton, 1980, 1992; Hendry, 1983), however, it is also one of the least researched ones in case of transitional economies. The consumption/income data is scarce and frequently unavailable for transitional countries—a consequence of which is noticeable lack of this line of research in the literature. Consequently, this application provides an excellent example of pros and cons of time series modelling with transitional macroeconomic data.

Croatia is an example of a CEE country where only limited consumption data exists. Since 1998 Croatian Central Bureau of Statistics carries out annual household consumption surveys harmonised with the EU standards.[1] Presently there are thus two cross sections available. A quarterly time series data set for the 1994-2000 period was recently published in Mikulic and Lovrincevic (2000) who used an expenditure approach methodology to determine consumption and income data. For our purposes we will ignore likely measurement problems with this data set and use it as given to demonstrate several important points.

The simplest linear model is a static relationship between consumption and income of the form

$$cons_t = \alpha + \beta inc_t + \varepsilon_t \qquad (1)$$

where *cons* = natural logarithm of total personal consumption and *inc* = logarithm of GDP. Fig. 1 shows graphs of the levels of both variables together with their accompanying non-parametric (Gaussian kernel) density estimates.[2]

[1] The survey methodology is based on the EU Household Budget Surveys and recommendations for harmonisation 1997, No. 361 used by EUROSTAT for harmonisation of methodologies for EU Member States.

[2] For details on non-parametric density estimation see Hendry and Doornik (1999) and Silverman (1986).

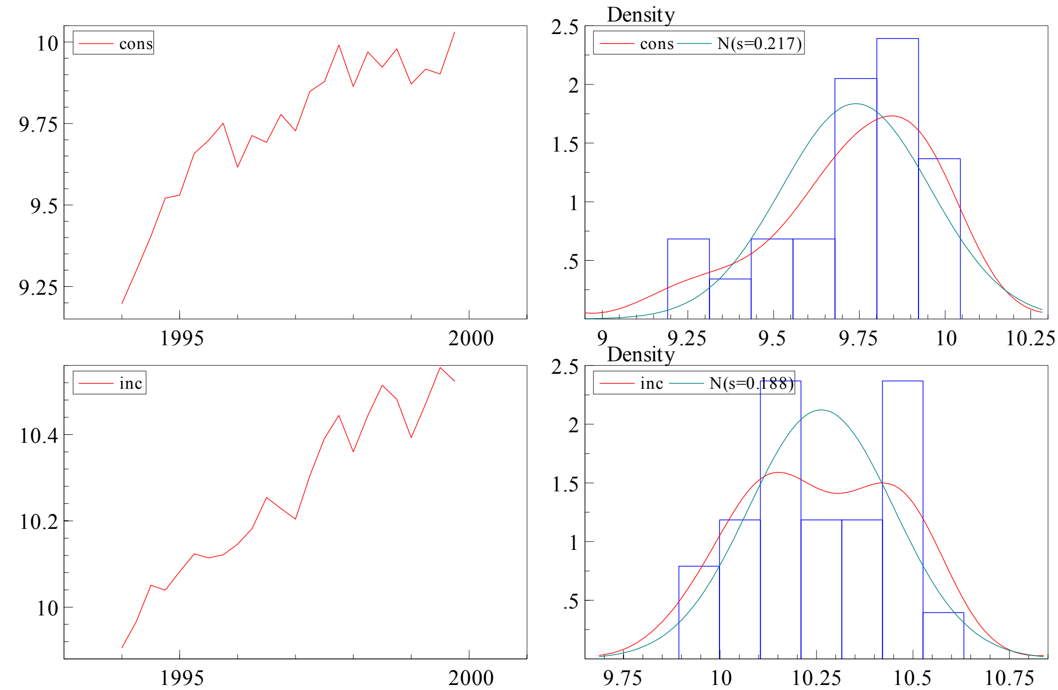

Fig. 1. Levels of $cons_t$ and inc_t with accompanying empirical densities plotted against a Gaussian density with equal first and second moments

Both variables show noted upward trend, apparently deviating from normality. Estimating equation (1) with OLS produces the following results:

$$cons_t = \underset{(0.846)}{1.555} + \underset{(0.0819)}{0.8}\, inc_t + e_t \qquad (2)$$

(SE)

$$R^2 = 0.841 \quad \sigma = 0.0567 \quad DW = 2.12$$

where the first year data was dropped following common modelling practice. It was actually found that both consumption and income variables were $I(1)$ hence nonstationary.[3] However, the model residuals contained no unit root[4] which indicates a possible long-run cointegrating relationship. Furthermore, the Durbin-Watson statistic has an acceptable value of 2.12 further suggesting apparent validity of the model. Fig. 2e and fig. 2f show residual density and QQ plots, respectively, which, to some degree, deviate from normality though the deviation does not appear alarming. Fig. 2c and fig. 2d show residual plot and residual correlogram. While residuals fall within 2 standard deviations limits, the correlogram indicates possible non-modelled dynamics at 3rd and 4th lags.

[3] The augmented Dickey-Fuller (ADF) statistic failed to reject the null of unit root in all settings (with trend, constant, seasonals). Subsequently, the first difference was found to be $I(0)$.

[4] The *t*-ratio was -4.6119 which falls below the –2.5658 ADF value for cointegration tests (see Benerjee, et al. 1993; Harris, 1995).

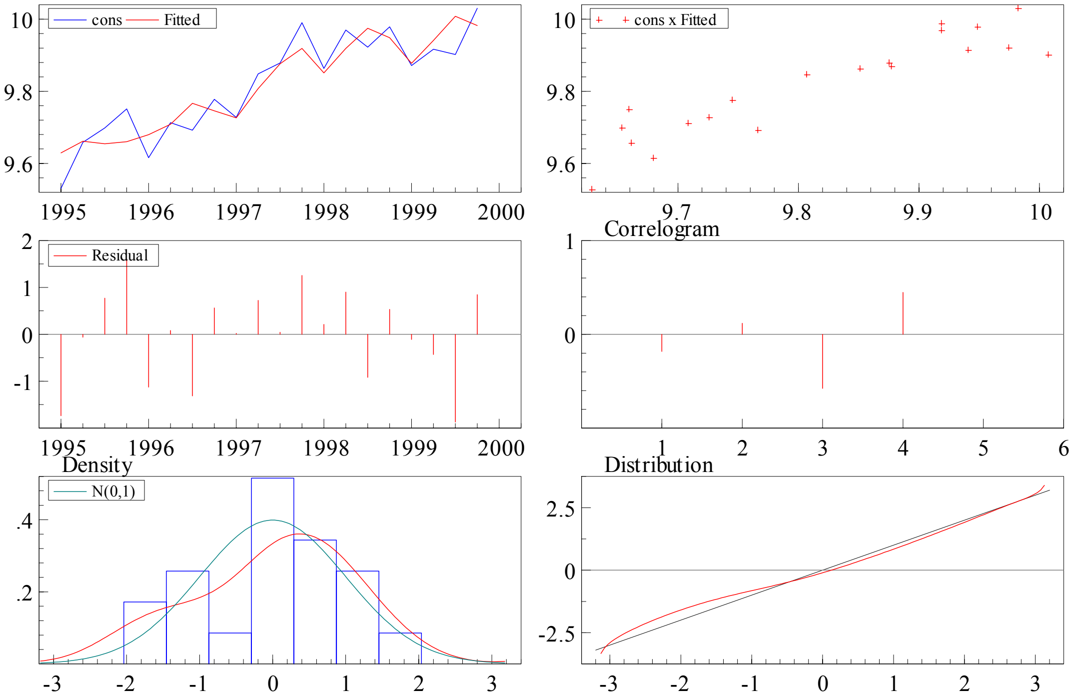

Fig. 2. Post-estimation graphical analysis over the sub-sample 1995 Q1 to 1999 Q4

Should we then accept this as an acceptable model and reject the standard permanent-income hypotheses since the income elasticity is only 0.8? First of all, from Fig. 2a and fig. 2b we can see that the fit is rather poor—most likely calling for modelling additional dynamics and possible deterministic seasonality (e.g., with seasonal dummies).

Now, notice that in start we "cleaned" our data by dropping the first four observations (first year of available data) using for estimation a sample with 20 observations. Re-estimating the same static equation (1) over the entire sample period, i.e., retaining the first year of data, yields:

$$\underset{(SE)}{cons_t} = \underset{(0.9311)}{-1.2924} + \underset{(0.0907)}{1.075}\, inc_t + e_t \qquad (3)$$

$$R^2 = 0.865 \quad \sigma = 0.0567 \quad DW = 1.01$$

It is now found that constant has changed sign from that in equation (2) and that the income elasticity increased to one—suddenly confirming the permanent income hypotheses![5]

[5] Note that we are using national (GDP) instead of personal disposable income and total personal consumption instead of per capita figures.

However, the first alarming sign is a low Durbin-Watson statistic of 1.01. Fig. 3c and fig. 3d further suggest cyclical behaviour of the residuals coupled with a slow-dying correlogram, with poor-looking fit (fig. 3a, 3b).

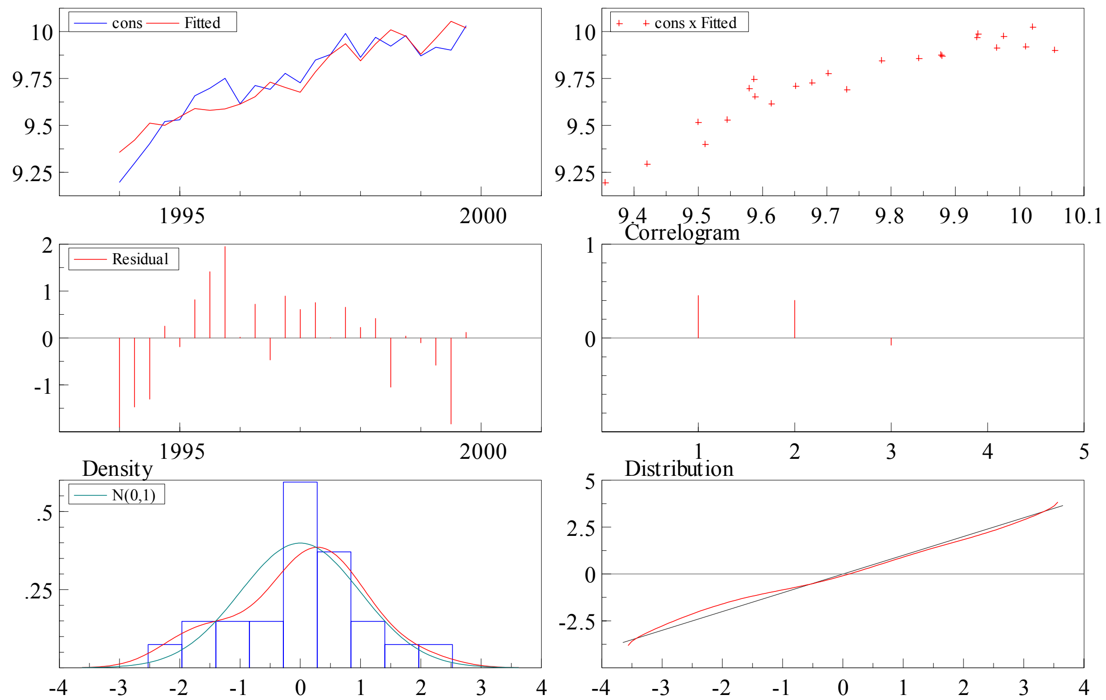

Fig. 3. Post-estimation graphical analysis over the sample 1994 Q1 to 1999 Q4

To make matters worst, the residual unit-root ADF cointegration test no longer rejects the unit-root null suggesting a spurious relationship (see Phillips, 1986; 1987) and hence a seriously problematic model.

This example is rather extreme. Apparently innocuous deletion of several initial observations lead here to unacceptably large changes in the model, ultimately violating the initial conjecture of a static long-run relationship between consumption and income.

We should, first of all, keep in mind the size of this data set (T = 24), as it is clear that degrees of freedom corrections alone limit the model complexity. If we would like to account for possible deterministic seasonality and further allow for, say, four lags in both endogenous and exogenous variables, further losing one observation for differencing (due to nonstationarity and lack of cointegration) together with the constant term, we would need to estimate an equation with three dummy variable coefficients, eight lag coefficients, and a constant. In total this adds up to 12 coefficients to estimate—a rather demanding task for a data set with 24 observations! Solely the 1/(24 – 12) d.f. correction in the estimate of residual variance is likely to lead to insignificance of all coefficients in the model (as it was indeed the case though we are omitting the results from such estimation).

The issue of d.f. will be touched again later in the context of seasonal adjustment and seasonal filters, and theoretical aspects related to simple vs. general formulation will be discussed in relation to the Rissanen (1986, 1987) and Ploberger-Phillips (1998a, 1998b) theorems. It should be pointed out that our previous model evaluation arguments were based on finite sample *t* distribution assumption. In addition to d.f. corrections, the inability to rely on finite sample distributional assumptions further calls in question our small sample results. Therefore, we digress for the moment from the empirical issues to analyse the distribution of estimators in a general bivariate autoregressive distributed lag (ADL) model of a kind used to model the above consumption-income relationship.

4. Analysis of the ADL(*m*, *k*) model

In the above empirical example we estimated a static relationship of the form:

$$y_t = \alpha + \beta x_t + \varepsilon_t,$$

which is actually only a special case of the general autoregressive distributed lag (*m*, *k*) model, i.e., ADL(*m*, *k*), with $m = k = 0$. The general ADL (*m*, *k*) dynamic specification which allows for any number of lags in both endogenous and exogenous variable is given by:

$$\boldsymbol{y}_t = \alpha + \sum_{i=1}^{m} \gamma_i \boldsymbol{y}_{t-i} + \sum_{i=0}^{k} \beta_i \boldsymbol{x}_{t-i} + \boldsymbol{u}_t \qquad (4)$$

where $\boldsymbol{z}_t \equiv (1\ \boldsymbol{y}_{t-1}\ \boldsymbol{y}_{t-2} \ldots \boldsymbol{y}_{t-m}\ \boldsymbol{x}_t\ \boldsymbol{x}_{t-1} \ldots \boldsymbol{x}_{t-k})$ and $\boldsymbol{\beta} \equiv (\alpha\ \gamma_1\ \gamma_2 \ldots \gamma_m\ \beta_0\ \beta_1 \ldots \beta_k)$.

This type of ADL (*m*, *k*) model is one of the most frequently estimated linear dynamic equations in empirical macroeconometrics. It is also one of the simplest models but, nevertheless, sufficiently general to encompass a number of special cases such as autoregression, static regression, leading indicators, common factors, error correction, etc. (see Hendry, 1995; Hendry and Doornik, 1999). In case of modelling transitional economic relationships (with severely limited samples sizes), at the moment, this is perhaps the only type of dynamic model that can, at least to some degree, be practically estimated. Clearly, inclusion of additional exogenous variables and subsequently their lags eats up a tremendous number of degrees of freedom, quickly peoducing practically useless or even inestimable equations. Further popularity of the formulation (1) comes from a known theoretical result that such linear equations can be estimated by OLS. The use of maximum likelihood estimation with Gaussian errors leads again to the same result as OLS and provides further justification for its use.

However, it shouldn't be forgotten that the validity of specification (4) rests on two main assumptions: applicability of asymptotic arguments and exogeneity of the right-hand variables. The exogeneity issue is a well researched one, and hardly specific for transitional data, nevertheless, it is important to keep in mind that single equation estimation can certainly not be justified only on the grounds of data-unavailability and inability to estimate systems or use instrumental variables (IV) estimator (e.g., due to lack of good instruments). Thus, we restrict our attention to the cases with weakly exogenous right-hand variables in the sense of Engle, et. al. (1983).

In order to study applicability of asymptotic results we analyse the asymptotic distribution of the special case used in the above empirical example allowing for *m* lags in the endogenous and *k* lags in the exogenous variable.

Assuming certain properties of the analysed stochastic processes is common in time series analysis. Particularly, analysing asymptotic convergence of stationary processes is commonly considered acceptable for *I*(0) or differenced *I*(1) data[6]. Leaving at this point aside analysis of possibly cointegrated series we proceed analysing our model while making these standard assumptions, subsequently showing on how shaky grounds could any confidence about the nature of analysed stochastic processes rest in case of transitional time series.

We make the following assumptions about the $\{y_t\}$ and $\{x_t\}$ processes:

Assumption 1: For *stochastic processes* $\{y_t\}$ *and* $\{x_t\}$ *suppose that:*

[6] However, it is well known that estimation of models with differenced *I*(1) data might lose the long run information and can thus be considered a form of a misspecification (see Benerjee et. al., 1993; Hendry, 1995). Harris (1995) is a good and rather comprehensive non-technical introduction to the topic. Charemza and Deadman (1997) provide detailed examples of empirical modelling of *I*(1) variables and implications of long-run information.

$$E(y_t) = \mu_y, \quad \forall t$$
$$E(x_t) = \mu_x, \quad \forall t$$
$$E(y_{t-i} - \mu_y)(y_{t-j} - \mu_y) = \gamma_{|i-j|}, \quad \forall t$$
$$E(x_{t-i} - \mu_x)(x_{t-j} - \mu_x) = \delta_{|i-j|}, \quad \forall t$$
$$E(y_{t-i} - \mu_y)(x_{t-j} - \mu_x) = \psi_{|i-j|}, \quad \forall t$$
$$\sum_{k=0}^{\infty} \gamma_k < \infty, \ \sum_{k=0}^{\infty} \delta_k < \infty, \sum_{k=0}^{\infty} \psi_k < \infty$$

Lemma 1: *Let w_t be a covariance-stationary process with finite fourth moments and absolutely summable autocovariances. Then the sample mean satisfies:*

$$(1/T)\sum_{t=1}^{T} w_t \overset{m.s.}{\rightarrow} \mu_y.$$

Proof Omitted. See Hamilton (1994), Proposition 7.5, p. 188.

Lemma 2: *Let $\{y_t\}$ and $\{x_t\}$ be stochastic processes satisfying Assumption 1. Then the following convergence results hold:*

(i) $(1/T)\sum y_{t-i} \overset{p}{\rightarrow} E(y_t) = \mu_y$

(ii) $(1/T)\sum y_{t-i}^2 \overset{p}{\rightarrow} E(y_{t-i} - \mu_y)^2 = \phi_0 + \mu_y^2$

(iii) $(1/T)\sum y_{t-i} y_{t-j} \overset{p}{\rightarrow} E(y_{t-i} y_{t-j}) = \phi_{|i-j|} + \mu_y^2$

(iv) $(1/T)\sum x_t \overset{p}{\rightarrow} E(x_t) = \mu_x$

(v) $(1/T)\sum x_{t-i} \overset{p}{\rightarrow} E(x_t) = \mu_x$

(vi) $(1/T)\sum x_{t-i}^2 \overset{p}{\rightarrow} E(x_{t-i} - \mu_x)^2 = \delta_0 + \mu_x^2$

(vii) $(1/T)\sum x_{t-i} x_{t-j} \overset{p}{\rightarrow} E(x_{t-i} x_{t-j})^2 = \delta_{|i-j|} + \mu_x^2$

(viii) $(1/T)\sum x_{t-i} y_{t-j} \overset{p}{\rightarrow} E(x_{t-i} y_{t-j})^2 = \psi_{|i-j|} + \mu_x \mu_y$

Proof See Appendix A.

We can now state the following theorem about the asymptotic distribution of the OLS estimators of the ADL model (4):

Theorem 1: *For ADL model given by (4) suppose that the roots of the polynomials $(1 - \gamma_1 z - \gamma_2 z^2 - \ldots - \gamma_m z^m) = 0$ and $(1 - \beta_1 z - \beta_2 z^2 - \ldots - \beta_k z^k) = 0$ are outside of the unite circle and with $\{u_t\} \sim i.i.d.(0, \sigma^2)$ and $E(u_t)^4 < \infty$. Then:*

$$\sqrt{T}(\hat{\beta}_{OLS} - \beta) \overset{d}{\rightarrow} N(0, \sigma^2 \Sigma^{-1})$$

with **Σ** *finite, positive definite matrix.*

Proof See Appendix B.

While the above result proves that under stated assumptions the OLS estimates converge to a Gaussian density asymptotically, the coefficients in (4) are not unbiased. In particular, it can be shown that in case of the general ADL(*m*, *k*) model satisfying Assumption 1 for stochastic processes $\{y_t\}$ and $\{x_t\}$ both the OLS coefficient vector **β** and the variance estimate

$$\sqrt{T}(s_T^2 - \sigma^2) \xrightarrow{d} N\left[0, E(u_t^4) - \sigma^4\right]$$

are biased while the *t* and F statistics converge to Gaussian (0, 1) and X^2 variables, respectively[7].

The asymptotic analysis of our simple dynamic model showed that only at rate of √T the OLS estimates converge in probability to a Gaussian density, same holding for the variance distribution. More theoretical details on convergence results could be found in Billingsley (1968; 1986), Davidson (1994) and White (1984).

Subsequently, the main tools of model specification and hypotheses testing, *t* and F statistics do not have regular *t* and F distributions in the small samples but asymptotically converge to a Gaussian and a normalised X^2 variable, respectively. Consequently, this is casting serious doubts about reliability of small-sample results with, e.g., T = 24 as in the above quarterly empirical example.

5. Small sample bias and nonstationarity in dynamic models

The previous asymptotic analysis placed rather heavy demands for the reliability of the ADL estimates requiring a convergence rate of √T. Thus, even if the Assumption 1 would hold, quarterly time series data for transitional economies with T ≤ 40 (for countries where quarterly data is available for 1990-2000 period) are likely to be too short for valid inference about the model parameters. However, we will show that with transitional macroeconomic time series some of the postulates from Assumption 1 regardless of their apparent justifiability might not hold.

[7] The F statistic multiplied by the number of degrees of freedom actually converges to a chi square variable, asymptotically. These results are equivalent to those of the autoregression (see Hamilton, 1994; Gourieroux and Monfort, 1990; Davidson, 2000).

Before returning to the likely violations of the Assumption 1 in light of the characteristics of transitional DGPs, we present a simulation example with artificially generated data aimed to illustrate the relationship between sample size and OLS bias in the simplest case with $m = 1$ and $k = 0$, i.e., an AR(1) process.[8]

For a computer generated $\{\varepsilon_t\}$ process drawn from a Gaussian (0, 1) density, with $\beta_0 = 0$, we estimated the following AR(1) model for different values of the β_1 coefficient:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, 1) \tag{5}$$

Table 1 gives the results of recursive estimation of the equation (5) over different sample sizes using the same $\{\varepsilon_t\}$ series.

Table 1
Recursive estimates of AR(1) coefficients for various sample sizes

T	$\beta_1 = 0.1$	$\beta_1 = 0.3$	$\beta_1 = 0.5$	$\beta_1 = 0.6$	$\beta_1 = 0.8$	$\beta_1 = 0.99$	$\beta_1 = 1.0$
10	0.62344	0.66912	0.72882	0.76646	0.85308	0.91771	0.91888
	(0.29968)	(0.27942)	(0.25817)	(0.24843)	(0.23592)	(0.24259)	(0.24372)
20	0.35633	0.45176	(0.54519	0.59512	0.70403	0.85475	0.86627
	(0.22889)	(0.21692)	(0.20291)	(0.19435)	(0.17072)	(0.12377)	(0.12046)
30	0.42814	0.53419	(0.63206	0.68291	0.79551	0.85811	0.86207
	(0.17405)	(0.16296)	(0.15010)	(0.14262)	(0.12484)	(0.09832)	(0.09419)
40	0.22815	0.38006	0.51163	0.57161	0.68230	0.82876	0.85557
	(0.16084)	(0.15356)	0.14326)	(0.13681)	(0.12063)	(0.08879	(0.08254)
50	0.08789	0.26451	0.42961	0.50884	0.65855	0.80397	0.83767
	(0.14541)	(0.14109)	(0.13264)	(0.12689)	(0.11258)	(0.08278)	(0.07421)
60	0.10617	0.27851	0.43773	0.51249	0.64664	0.78643	0.82990
	(0.13357)	(0.13004)	(0.12306)	(0.11847)	(0.10771)	(0.08056)	(0.06986)
70	0.07285	0.25059	0.42002	0.50226	0.66466	0.81413	0.82764
	(0.12334)	(0.11928)	(0.11152)	(0.10616)	(0.09184)	(0.07113)	(0.06601)
80	0.18062	0.35373	0.51253	0.58702	0.72724	0.86005	0.85557
	(0.11282)	(0.10776)	(0.09973)	(0.09468)	(0.08168)	(0.05738)	(0.05755)
90	0.20513	0.38042	0.54005	0.61437	0.74965	0.86395	0.85805
	(0.10658)	(0.10129)	(0.09297)	(0.08763)	(0.07391)	(0.05513)	(0.05418)
100	0.16414	0.33930	0.50072	0.57667	0.72063	0.84920	0.85561
	(0.10034)	(0.09565)	(0.08799)	(0.08304)	(0.07040)	(0.05288)	(0.05127)

Note: Standard errors in parentheses.

The figs. 4a-4f show a full set of recursive OLS estimates over each possible sample size. One can immediately observe noted upward bias in the smallest sample sizes with $\beta_1 < 0.8$, and a downward bias when $\beta_1 > 0.8$, specially when T = 10 which is admittedly a rather extreme case.

[8] Similar experiment could be performed as a Monte Carlo study. Such an approach would be much more general pointing out to the main characteristics of the model, on average (see Hendry, 1984; 1995).

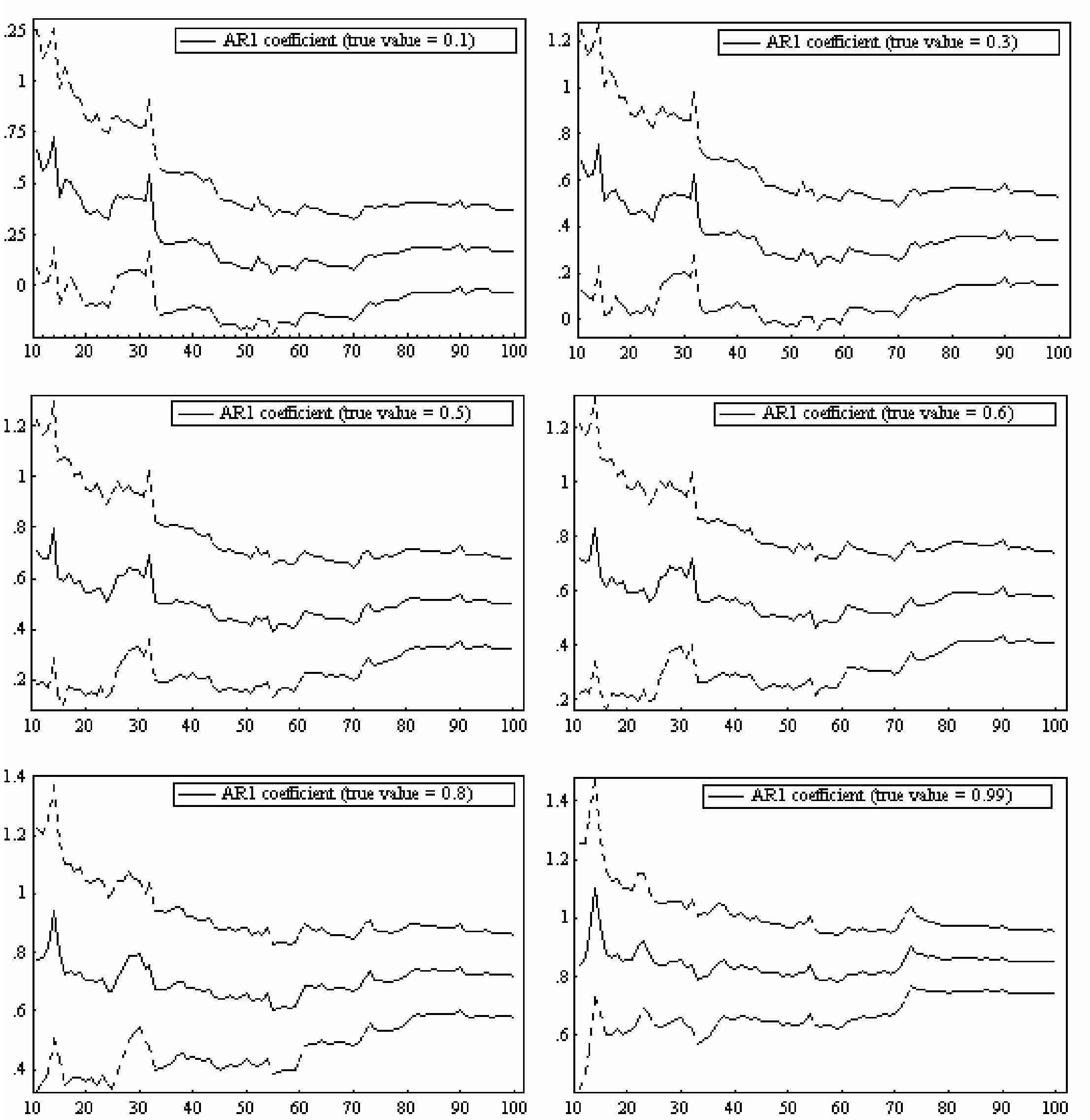

Fig. 4. Recursively estimated AR(1) coefficients for T = 10, 20, …, 100

This simple example shows that in extremely small samples ($T < 50$) it is possible to get a significant *t*-ratio on a coefficient that has an upward bias of more then 500% (e.g., the estimated coefficient for T = 10, $\beta_1 = 0.1$ was a significant 0.62344—an overestimate of more then 600%!). The smallest sample sizes show similar upward bias for values of β_1 less then app. 0.9 above which the bias becomes a downward one, which is most known and researched in the random walk case, i.e., when $\beta_1 = 1$ (fig. 5).

In all cases the bias appears to change with the sample size, starting from a sharp upward bias in samples with app. $T > 40$, turning into a downward bias for app.

40 < T < 80, and finally apparently converging toward the true value for samples with T > 80 for as long as app. $\beta_1 > 0.8$ after which the bias is always downward for all sample sizes.

Properties of estimators, only in regard to sample size, are infrequently researched in extremely short samples as such cases are unconvincing and statistically uninteresting. Unfortunately, the data length of a typical transitional country's macroeconomic time series often dictates estimation with such small samples within a fixed time interval which cannot be extended by simply collecting more data. Therefore, some of insights that might be gained from simulation analyses might be rather useful in detecting potential pitfalls in dynamic econometric models of transitional economies.

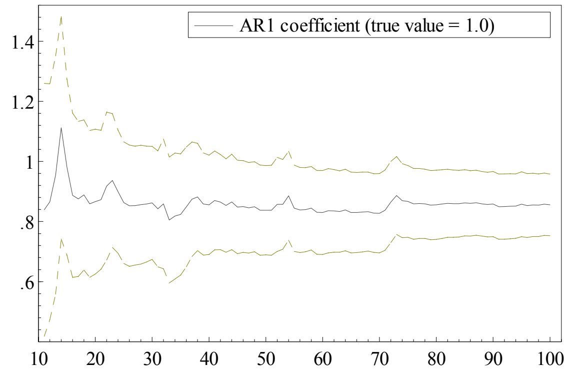

Fig. 5. Recursive estimates of the AR(1) coefficient in a random walk process

So far we considered the statements in the Assumption 1 valid. However, even the simplest assumption of constant expectation, i.e., $E(y_t) = \mu_y \; \forall t$, is unrealistic in very small samples. The constant expectation assumption is based on the infinite order moving average (MA) representation of stable autoregressive or ADL processes. As talking of infinity with strictly limited samples is least said unrealistic, we need to calculate finite sample expectation, which turns out to depend on t and is thus non-constant. For the AR(1) process given in (5) the expectation of y_t is calculated as

$$\begin{aligned} y_t &= \beta_0 \sum_{j=1}^{t-1} \beta_1^j + \sum_{j=1}^{t-1} \beta_1^j \varepsilon_{t-j} + \beta_0^t y_0 \\ \Rightarrow E(y_t) &= E\left(\beta_0 \sum_{j=1}^{t-1} \beta_1^j\right) + E\left(\sum_{j=1}^{t-1} \beta_1^j \varepsilon_{t-j}\right) + E\left(\beta_0^t y_0\right) \\ \therefore \; E(y_t) &= \frac{\beta_0 (1-\beta_1^t)}{1-\beta_1} + \beta_0^t y_0 \end{aligned} \tag{6}$$

and so it follows that for small values of t $E(y_t) \neq \mu_y$, even when $\beta_1 < 1$. From the above derivation it can be noticed that the rate of convergence to constant expectation depends on t, β_0 and β_1. In addition, because in case of transitional countries the usual assumption that $y_0 = 0$ makes little sense since the beginning date of the process is known as proposed in (ii). This is a rather problematic aspect since if $y_0 \neq 0$ and $\beta_0 > 1$ the expectation will actually diverge as $T \to \infty$. In the simplest case with $\beta_0 = 0$ the convergence rates toward the constant expectation are shown in fig. 6.

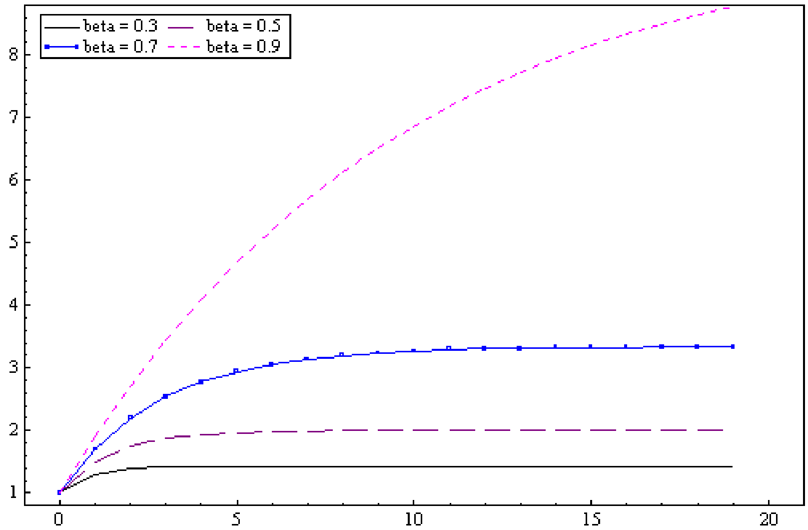

Fig. 6. Convergence of the expected value of y_t toward a constant for various values of β_1 in model (5)

The problem apparent in (6) is that, regardless of the value of β_1 the process $\{y_t\}$ is nonstationary due to the non-constant first moment. It can be subsequently shown than the second moment (i.e., variance) depends on t as well (see Hendry, 1995: 136). Thus, weak stationarity required for the validity of Theorem 1 can be assured either asymptotically (as $t \to \infty$) or by assuming a known density for y_0. The later option does not apply due to the nature of transitional DGPs, particularly because of the assumed properties (i) and (ii). The former option again places heavy requirements on T (sample size) strongly suggesting that the existing data span (app. 10 years = 40 observations) is insufficient for asymptotic arguments with quarterly data.

The main conclusion so far is that usual modelling of quarterly time series macroeconomic data (e.g., standard consumption function estimation) is not appropriate as asymptotic normality does not apply to such quarterly data sets. Needless to say, modelling of annual time series is, from the inferential point of view, completely unacceptable even if mathematically possible.

Since extension of the given data period is unlikely, the only feasible solution is to increase T through increasing the data frequency. This would require monthly data which might not be available for the exactly same variables. Actually, the use of monthly data is likely to require modelling of slightly different structural economic relationships and would need specific attention and investigation. However, up to 10 years of monthly data yields samples with 120 data points—which would not only promise more reliable application of asymptotic arguments, but even allow for estimation of general models that would be less concerned with saving degrees of freedom. Subsequently, such samples would technically allow direct modelling of possible seasonal dynamics and general dynamic specifications of initial models through application of the general-to-specific methodology. Hendry (1987) and Gilbert (1986) provide a concise exposition of this methodology while Hendry (1995) is a more comprehensive reference. Charemza and Deadman (1997) is a good practical guide for general-to-specific modelling and empirical model-reduction procedures.

6. General vs. specific models in small samples and the Rissanen theorem

The methodological debates related to the way of conducting empirical modelling on the lines of general-to-specific vs. specific-to-general (or more often "specific-to-specific") is hardly specific to modelling transitional economies. Actually, these issues are infrequently brought up in the context of empirical modelling of transition. However, the back-bone of the differences between the two approaches is of central importance to this particular discussion. Namely, the model discovery process plays far greater role in a theoretically uncertain world of dynamically evolving new economies of the CEE countries.

As the relevant economic theory that would sufficiently well explain transitional macroeconomic dynamics is rather scarce and unreliable, the empirical approach to model building deriving from the process that actually generated the data (see Hendry, 1987) emerges as the only practical approach to analysing transitional economies. However, in addition to all previously mentioned statistical reasons for needing larger samples, this approach, due to its general nature, places further requirements on the length of data via estimation of general models with larger number of parameters (which are then reduced in the modelling process through model reduction techniques). General models, needless to say, eat up more degrees of freedom but allow for more complicated dynamics then the simple ones (such as the AR(1) example above).

These arguments aim to further support the use of higher-frequency data, namely monthly series, because in the case of CEE countries such methodology could not be used with quarterly data. Up to 10 years of monthly data, on the other hand, would allow for decently general initial specification even if seasonality is modelled with additional dummy variables (thus adding 11 more coefficients to the ADL specification).

In light of limited samples sizes and reluctance to waste too many degrees of freedom on estimation of general models aiming at discovery of the unknown DGP that generated transitional macroeconomic time series it is specially tempting to exploit the apparent upshots of a theorem due to Rissanen (1986, 1987) that recently appeared in theoretical econometric literature. This theorem supports estimation of rather simple models, not on the grounds of firm economic theory, but rather on pure statistical bases via establishment of minimal empirically achievable distance from the true DGP. While not likely to make much impact on standard econometric practices, the Rissanen theorem is likely to be “discovered” by CEE data analysts and possibly even used to justify estimation of overly simple specific models in very small samples.

The original theorem was stated in terms of the Kullback-Leibler information distance defined by E($\ln\partial P_\theta$ - $\ln\partial G_n$), where P_θ denotes the true DGP and G_n is any alternative model. For stationary data the Rissanen theorem establishes that for a subset of parameters Θ with dim Θ = k (i.e., k = number of estimated parameters):

$$\left\{\theta : -E_\theta \ln\frac{\partial G_n}{\partial P_\theta} \leq \frac{1}{2} k \ln T\right\} \xrightarrow{p} 0$$

A corresponding result was subsequently generalised for nonstationary data in Ploberger and Phillips (1998a, 1998b) directly for the log likelihood (not its expectation). The Ploberger-Phillips theorem states that ∀α, ε > 0 the Lebesgue measure of the set

$$\left\{\theta : P_\theta\left[\ln\partial P_\theta - \ln\partial G_n \leq \frac{1-\varepsilon}{2}\ln\det B_n\right] \geq \alpha\right\}$$

converges to zero, where $B_n = -\partial^2 l_n/\partial\theta\partial\theta'$, and l_n is the log likelihood of the estimated model. Thus, in the stationary case it is established that, whatever the model it cannot come closer to the true (unknown) DGP then, on average (due to expectation of the log likelihood), ½ $k$lnT (T = sample size). On the other hand, if the data is nonstationary, then the bound will depend on the information matrix Bn and any empirical model will not be able to come closer to the true DGP (in terms of likelihood ratio) then ½ ln det B_n. Fig. 7 showes Rissanen bounds for the Kullback-Leibler information distance for various sample sizes where can be seen that the distance "from true DGP" indeed increases linearly with k, as claimed in Rissann (1986, 1987) and Ploberger and Phillips (1998a, 1998b), but it also increases linearly with T suggesting that not only models with small parameters should be preferred to more general specifications but that, ceteris paribus, models estimated with smaller data sets (shorter series) will be, on average, closer to the true DGP then models estimated with larger data sets! While the second kind of divergence appears at a slower logarithmic rate, it is still very noticeable and alarming. Solely based on such interpretation it could even be concluded that simplest models estimated with shortest data sets will come closest to the true DGP, which is clearly ridiculous. It is now clear why we are discussing the Rissanen theorem in context of econometric inference with transitional time series data.

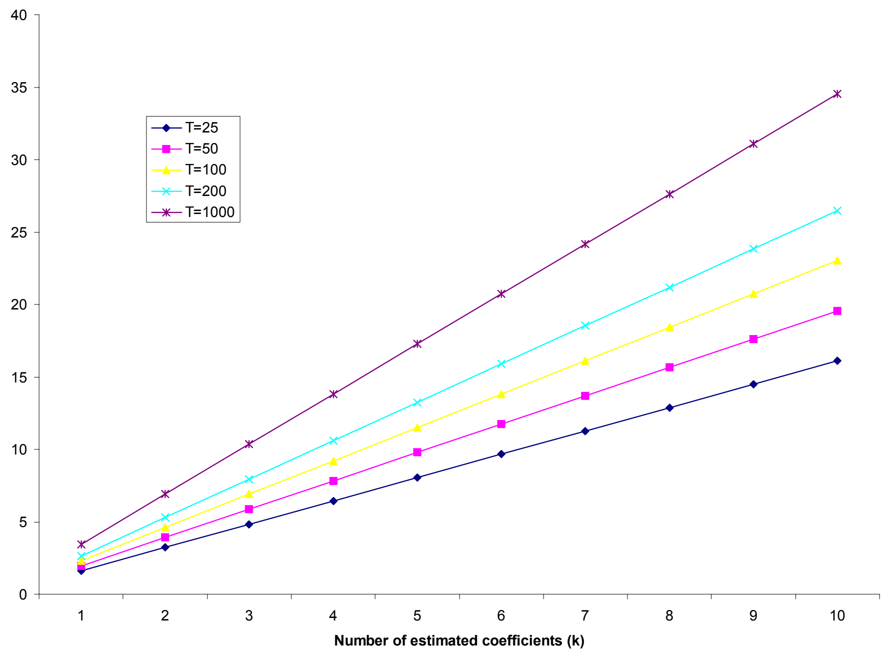

Fig. 7. Rissanen bounds for the Kullback-Leibler information distance for various sample sizes.

As the Rissanen theorem might be interpreted in terms of relative advantage of small, simple (and thus specific) models over the more general ones, it could be misused as a justification for estimation of equations such as (2) and (3) despite of all previously mentioned problems. However, we will show in the following example that such line of reasoning can be very misleading if applied to dynamic models.

Suppose that the true DGP is given by:

$$y_t = 0.8\, y_t + e_t$$
$$z_t = 0.7\, z_t + w_t$$

where we computer generated 100 drawings from a bivariate standard Gaussian density, i.e., $\mathbf{u}_t = (e_t, w_t)' \sim IN(\mathbf{0}, \boldsymbol{I})$. Further, suppose we wish to estimate the following static equation:

$$y_t = \alpha z_t + \varepsilon_t \quad (7)$$

where the true value of α is zero. However, the OLS estimation of (7) with simulated data gave the following estimates:

$$y_t = 0.65047 - 0.16716\, z_t + \varepsilon_t \quad (8)$$
$$\text{(SE)}\ \ (0.1126)\ \ \ (0.0803)$$

Therefore, we obtain a completely spurious relationship since y_t and z_t were independently generated stationary processes. However, a low Durbin-Watson statistic of 0.827 and R^2 of 0.0432 signal possible problems with this static formulation, while residuals appear to be normally distributed (normality X^2 = 0.321 which has probability of 0.852). Fig. 8 shows the Gaussian kernel estimate of the empirical residual density against a N(1, 0) curve.

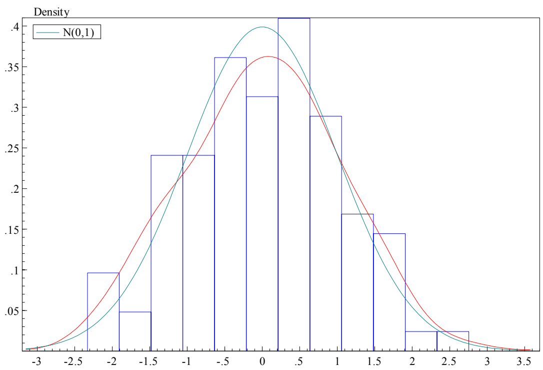

Fig. 8. Gaussian kernel estimate of the residual density

On the basis of *t* statistics (ignoring low DW or correcting for first order autocorrelation in the residuals via. Cochrane-Orcutt procedure) and normally distributed residuals we might accept that y_t and z_t are related. Further using the conclusions based on Rissanen's theorem, we could conjecture that a more general model with, say $k = 6$ would be further away from the true DGP then (8). Yet, we know that (8) is the worst possible model for this case as it is suggesting totally spurious relationship opposite from the true DGP.

Now, re-estimating the model with two lags of each y_t and z_t we obtain the results in table 2. Clearly, the model would need further revisions, and after dropping the insignificant variables we arrive at the model for y_t very similar to the true DGP.

Table 2
OLS results from a more general specification

Variable	$\hat{\alpha}$	$\hat{\sigma}$	**t-value**	**t-prob**
Constant	0.21704	0.11188	1.940	0.0554
y_{t-1}	0.56649	0.10409	5.442	0.0000
y_{t-2}	0.08226	0.10585	0.777	0.4391
z_t	-0.02364	0.08592	-0.275	0.7838
z_{t-1}	-0.09929	0.10535	-0.942	0.3484
z_{t-2}	0.09243	0.08643	1.069	0.2876

$R^2 = 0.397797$ $F(5,92) = 12.154$ [0.0000] $\sigma = 0.901247$ DW = 2.03
RSS = 74.727 $k = 6$, T = 98

The above result is a consequence of dynamic misspecification cased by underestimation of the true variance of α. For theoretical analysis see Hendry (1995: 146-9). Therefore, we have seen that Rissanen theorem, while providing a possible tool for differentiating among models, does not allow progressive model discovery, the crucial process in modelling theoretically unknown relationships hence potentially leading to large dynamic misspecification and false models.

8. An empirical example with monthly disaggregated series

In our initial empirical example we were interested in estimating the relationship between consumption and income. We noted that, at best, in transitional countries these variables might be available at quarterly frequency and concluded that such data is simply too short to be practically useful for econometric modelling. However, the consumption and income variables are themselves constructed. Personal consumption is usually defined to consists of consumption of food, beverages, clothing, housing, energy, transport, etc. Similarly, personal disposable income consists of income from paid employment, self-employment, own business and free lances, property income, etc. Thus, both variables are aggregates of other variables, many of which might be individually available as well.

It is actually the case that in some CEE countries good monthly data exists for some of the variables that comprise constructs such as „personal disposable income" or "personal consumption." From statistical point, of view such variables may very well be better for modelling purposes then their economically justified aggregates. In the matter of fact, aggregation itself can cause a number of statistical problems such as creation of temporal correlations and even artificial nonstationarity. Gourieroux and Monfort (1997: 442-5) and Hendry (1995: 346-7) provide some insides into these issues.

In the specific case of Croatia monthly data is available for retail sales and salary. According to the annual household consumption surveys carried out by the Central Bureau of Statistics since 1998, retail purchases comprises about 60% of the consumption of households. The other 40% goes on expenditures on housing, energy, credits repayment, etc. Salary is defined to include income from paid employment which together with self-employment makes up about 70% of personal disposable income, other parts being property royalties, unemployment benefits, scholarships etc.

While from the economic point of view an attempt to explain part of the consumption (e.g., retail consumption) with part of income (e.g., salary) might be questionable, from statistical point of view modelling such relationship is perfectly legitimate. Furthermore, as the variables are less aggregated, such models might even have better properties then those based on more aggregated series.

In our final example we will estimate a relationship between mean gross salary and retail sales as proxies for income and consumption. The 1992-2000 Croatian data can be found in Artus and Kapur (2000), Dorsey, et. al. (1995), Elkan and Maggi (2000) and Elkan and Temprano-Arroyo (1998). As now eight years of monthly data is available, this gives series with T = 108 which is considerably better the quarterly case.

Estimation of a relatively general ADL model which include 12 dummy variable, after some reductions and dropping of insignificant variables yields the results given in table 3. In addition we can notice that the long run model

$$\underset{(SE)}{} \; cons_t = \underset{(0.0675)}{8.2027} + \underset{(0.0094)}{0.0675}\, inc_t + e_t$$

$$R^2 = 0.987527 \;\; \sigma = 0.144904 \;\; DW = 0.617$$

indicates that the *cons* and *inc* integrate with the appropriate residual ADF t-ratio of - 3.4734 wich rejects the unit-root null in the residuals.[9]

Table 3
OLS estimate of the ADL relationship between retail sales ($cons_t$) and salary (inc_t)

Variable	Coefficient	Std.Error	t-value	t-prob
Constant	1.3044	0.39360	3.314	0.0013
$cons_{t-1}$	0.8740	0.04739	18.443	0.0000
inc	0.5541	0.08551	6.479	0.0000
inc_{t-1}	-0.4696	0.08522	-5.511	0.0000
S_{III}	-0.3638	0.02072	-17.558	0.0000
S_{IV}	-0.0769	0.02036	-3.782	0.0003
S_{VI}	-0.0921	0.01843	-4.995	0.0000
S_{VII}	-0.0708	0.01843	-3.842	0.0002
S_{IX}	-0.0884	0.01915	-4.619	0.0000
S_X	-0.0742	0.01899	-3.905	0.0002
S_{XI}	-0.0471	0.01911	-2.467	0.0155
S_{XII}	-0.1682	0.01946	-8.642	0.0000

$R^2 = 0.998005$ $F(11,92) = 4184.8$ [0.0000] $\sigma = 0.0489629$ DW = 2.10
RSS = 0.2205577028 $k = 12$, T = 104

We have estimated, as most convincing, an ADL(1, 1) model with eight significant seasonal components, R^2 of near unity and Durbin-Watson of about 2.[10] Moreover, the WALD X^2 test for significance of the error correction term has a value of 343.79 which is highly significant further supporting the impression that the variable are cointegrated into a stationary linear combination.[11] The static long-run solution is given by (see Hendry and Doornik, 1999):

$$\underset{(SE)}{cons_t} = \underset{(0.871)}{10.35} + \underset{(0.0804)}{0.6703}\, inc_t - \underset{(1.157)}{2.887}\, S_{II} - \underset{(0.2393)}{0.611}\, S_{III} - \underset{(0.3185)}{0.7307}\, S_{VI} - \underset{(0.263)}{0.5619}\, S_{VII}$$
$$- \underset{(0.339)}{0.702}\, S_{IX} - \underset{(0.2976)}{0.5886}\, S_X - \underset{(0.2329)}{0.3742}\, S_{XI} - \underset{(0.5723)}{1.335}\, S_{XII}$$

[9] The highest significant lag was first, all higher lags were insignificant (see Dickey and Fuller, 1979).
[10] Note that, strictly speaking, DW statistics is inapplicable here due to the presence of lagged endogenous variable.
[11] Both retail sales and salary were found to he *I*(1), hence *I*(0) after first differencing.

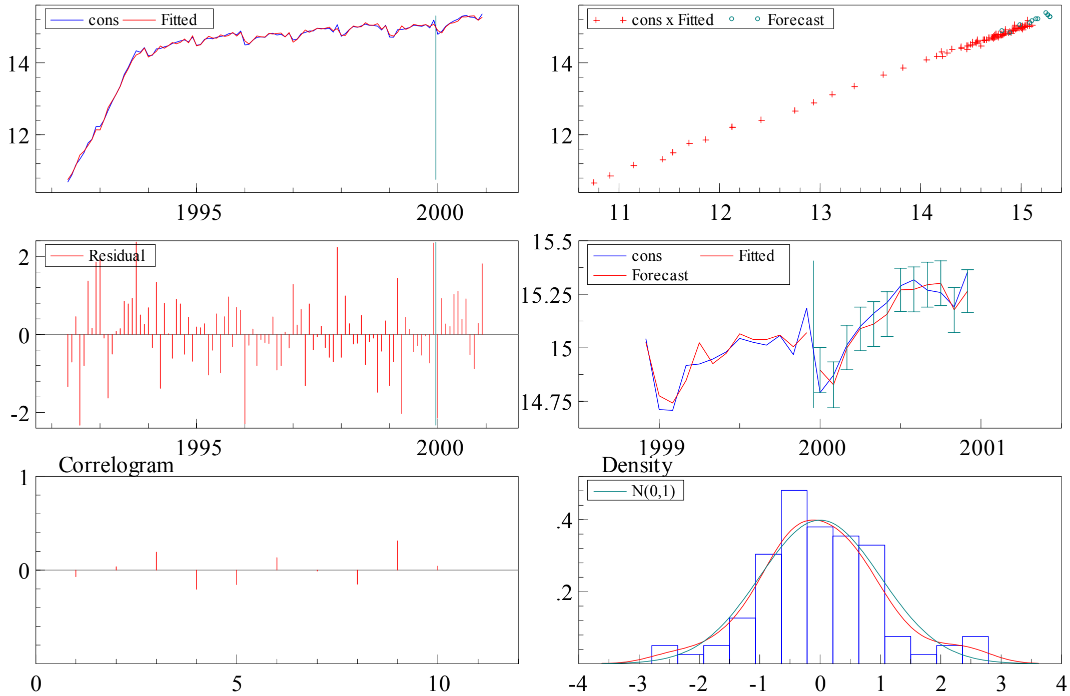

Fig. 8. Model evaluation graphics

Fig. 8 shows that the model fits rather well with nicely behaved, normally distributed, residuals and good forecasts for the last year of the data, falling within 5% confidence bounds. In addition, the residual correlogram and plots indicate white noise residuals, further supporting validity of the model.

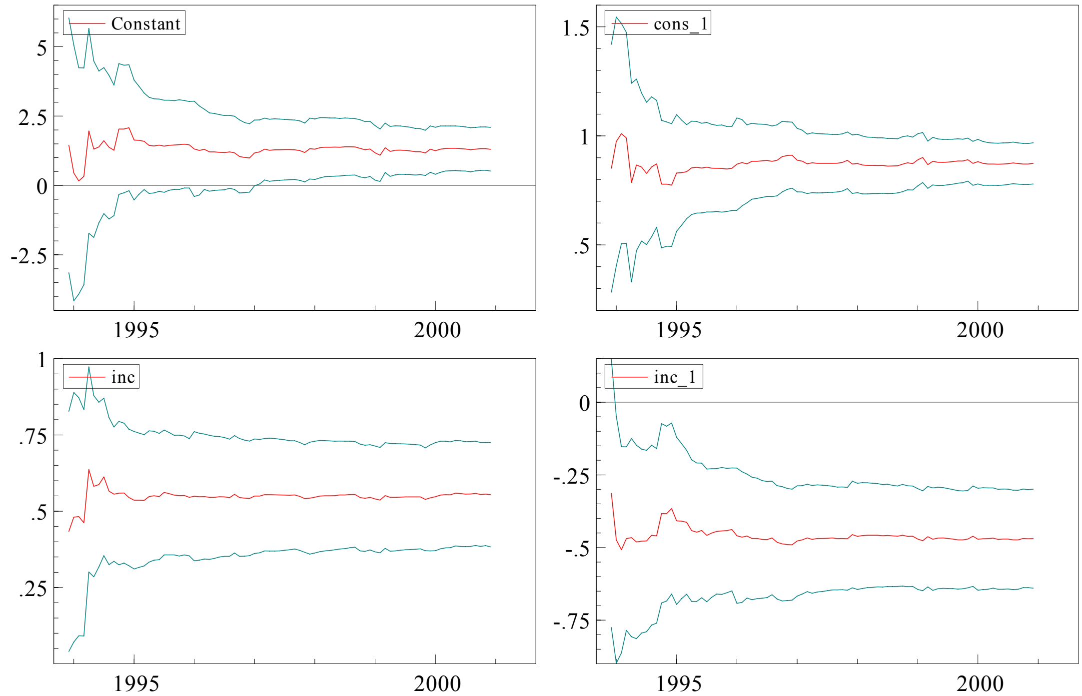

Fig. 9. Recursive estimates of the coefficients

Fig. 9 shows recursively estimated coefficients which show sufficient parameter constancy over the sample period.

Looking at the above calculated static long-run solution, we can observe that the income elasticity is a significant 0.67, thus smaller then the one estimated with quarterly data in the previous examples.

9. Conclusion

Characteristics of transitional time series data and dynamically changing economic systems in the CEE countries prevent estimation of standard macroeconometric models with quarterly data. The use of quarterly time series produces inestimable equations or, at lest, problematic results from the econometric inference point of view. These data are characterised with a known "beginning of time" (end of Communism) and most possibly changing DGP across time. At least, due to the economic regime change from controlled to market economy, it is impossible to assume that the "pre-sample" (pre-1990) data were generated by the same DGP as the transitional data.

Therefore, it was found that the only way to achieve asymptotic stationarity and thus enable inferential validity—even for stable processes—is to let t $\rightarrow \infty$. However, as the sample period cannot be extended back in time, and if one wishes to analyse transitional economies today, the only solution is to increase the frequency of observations, i.e., use monthly data. Monthly data, however, is usually available in more disaggregated form. For example, in case of Croatia, monthly data on personal disposable income does not exist, and it only exists on annual frequency since 1998. Nevertheless, the 1990-2000 data on mean gross (and net) salary is available and is of rather good quality. Similar situation exists for other macroeconomic variables. This brings up the issue of estimation of partial relationship and modelling disaggregated series.

Two data sets were used to illustrate these modelling issues. By using 1994-2000 quarterly consumption-income data we showed that apparently innocent dropping of the first year of observations could lead to completely different outcome, possibly interpretation as well, thus causing huge misspecification. On the other hand, estimation of the consumption-income relationship using disaggregated proxies (retail sales and mean gross salary) yielded a model with excellent properties, fit and forecasts, passing all considered diagnostic evaluations.

Therefore, our primary conclusion is that some dynamic macroeconometric models can indeed be estimated with transitional data from CEE countries with relatively high confidence in the model reliability but only if monthly data is used, even if the modelled relationship would include only disaggregated proxy variables.

Appendix A. Proof of Lemma 2

Proof Results (i), (iv) and (v) follow directly from Assumption 1 and Lemma 1. To derive (ii) and (vi) note that the convergence in mean square in Lemma 1 implies that $E\left[(1/T)\sum(y_t-\mu_y)^2\right]=E\left[(1/T)\sum(y_t^2-2y_t\mu_y+\mu_y^2)\right]=E\left[(1/T)\sum y_t^2\right]-\mu_y^2$, where for $i=j$, this is equal to γ_0, thus $E\left[(1/T)\sum y_t^2\right]=\gamma_0+\mu_y^2$ and similarly for x_t establishing (vi). For (iii) and (vii) the above is easily generalised for higher moments by noting that $E\left[(1/T)\sum(y_{t-i}-\mu_y)(y_{t-j}-\mu_y)\right]=E\left[(1/T)\sum y_{t-i}y_{t-j}\right]-\mu_y^2$ and similarly for x_t. Finally, (viii) is derived directly from the definition in Assumption 1 since $E(y_t\mu_y)=\mu_y^2$ and $E(x_t\mu_x)=\mu_x^2$.

Q.E.D.

Appendix B. Proof of Theorem 1

Proof The OLS estimate of $\boldsymbol{\beta}$ in (1) is given by $\hat{\boldsymbol{\beta}}_{OLS} = \left(\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t'\right)^{-1} \sum_{t=1}^{T} \mathbf{z}_t u_t$. Define the (m + k + 1) × ((m + k +1) matrix $\left(\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t'\right)^{-1} \equiv \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$, where

$$V_{11} \equiv \begin{pmatrix} T & \sum y_{t-1} & \sum y_{t-2} & \cdots & \sum y_{t-m} \\ \sum y_{t-1} & \sum y_{t-1}^2 & \sum y_{t-1} y_{t-2} & \cdots & \sum y_{t-1} y_{t-m} \\ \sum y_{t-2} & \sum y_{t-2} y_{t-1} & \sum y_{t-2}^2 & \cdots & \sum y_{t-2} y_{t-m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum y_{t-m} & \sum y_{t-m} y_{t-1} & \sum y_{t-m} y_{t-2} & \cdots & \sum y_{t-m}^2 \end{pmatrix},$$

$$V_{12} \equiv \begin{pmatrix} \sum x_t & \sum x_{t-1} & \cdots & \sum x_{t-k} \\ \sum y_{t-1} x_t & \sum y_{t-1} x_{t-1} & \cdots & \sum y_{t-1} x_{t-k} \\ \sum y_{t-2} x_t & \sum y_{t-2} x_{t-1} & \cdots & \sum y_{t-2} x_{t-k} \\ \vdots & \vdots & \ddots & \vdots \\ \sum y_{t-m} x_t & \sum y_{t-m} x_{t-1} & \cdots & \sum y_{t-m} x_{t-k} \end{pmatrix},$$

$$V_{21} \equiv \begin{pmatrix} \sum x_t & \sum x_t y_{t-1} & \sum x_t y_{t-2} & \cdots & \sum x_t y_{t-m} \\ \sum x_{t-1} & \sum x_{t-1} y_{t-1} & \sum x_{t-1} y_{t-2} & \cdots & \sum x_{t-1} y_{t-m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{t-k} & \sum x_{t-k} y_{t-1} & \sum x_{t-k} y_{t-2} & \cdots & \sum x_{t-k} y_{t-m} \end{pmatrix},$$

$$V_{22} \equiv \begin{pmatrix} \sum x_t^2 & \sum x_t x_{t-1} & \ldots & \sum x_t x_{t-k} \\ \sum x_{t-1} x_t & \sum x_{t-1}^2 & \ldots & \sum x_{t-1} x_{t-k} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_{t-k} x_t & \sum x_{t-k} x_{t-1} & \ldots & \sum x_{t-k}^2 \end{pmatrix}.$$

From Lemma 2 it follows that

$$T^{-1} \cdot V_{11} \xrightarrow{p} \begin{pmatrix} 1 & \mu_y & \mu_y & \cdots & \mu_y \\ \mu_y & \gamma_0 + \mu_y^2 & \gamma_1 + \mu_y^2 & \cdots & \gamma_{m-1} + \mu_y^2 \\ \mu_y & \gamma_1 + \mu_y^2 & \gamma_0 + \mu_y^2 & \cdots & \gamma_{m-2} + \mu_y^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_y & \gamma_{m-1} + \mu_y^2 & \gamma_{m-2} + \mu_y^2 & \cdots & \gamma_0 + \mu_y^2 \end{pmatrix},$$

$$T^{-1} \cdot V_{12} \xrightarrow{p} \begin{pmatrix} \mu_x & \mu_x & \cdots & \mu_x \\ \psi_1 + \mu_y\mu_x & \psi_0 + \mu_y\mu_x & \cdots & \psi_{|1-k|} + \mu_y\mu_x \\ \psi_2 + \mu_y\mu_x & \psi_1 + \mu_y\mu_x & \cdots & \psi_{|2-k|} + \mu_y\mu_x \\ \vdots & \vdots & \ddots & \vdots \\ \psi_m + \mu_y\mu_x & \psi_{|m-1|} + \mu_y\mu_x & \cdots & \psi_{|m-k|} + \mu_y\mu_x \end{pmatrix},$$

$$T^{-1} \cdot V_{21} \xrightarrow{p} \begin{pmatrix} \mu_x & \psi_1 + \mu_y\mu_x & \psi_2 + \mu_y\mu_x & \cdots & \psi_m + \mu_y\mu_x \\ \mu_x & \psi_0 + \mu_y\mu_x & \psi_1 + \mu_y\mu_x & \cdots & \psi_{|m-1|} + \mu_y\mu_x \\ \mu_x & \vdots & \vdots & \ddots & \vdots \\ \mu_x & \psi_{|1-k|} + \mu_y\mu_x & \psi_{|2-k|} + \mu_y\mu_x & \cdots & \psi_{|m-k|} + \mu_y\mu_x \end{pmatrix},$$

$$T^{-1} \cdot V_{22} \xrightarrow{p} \begin{pmatrix} \delta_0 + \mu_x^2 & \delta_1 + \mu_x^2 & \cdots & \delta_{k-1} + \mu_x^2 \\ \delta_1 + \mu_x^2 & \delta_0 + \mu_x^2 & \cdots & \delta_{k-2} + \mu_x^2 \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{k-1} + \mu_x^2 & \delta_{k-2} + \mu_x^2 & \cdots & \delta_0 + \mu_x^2 \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \xrightarrow{p} \Sigma,$$

where $\mathbf{\Sigma}$ is positive definite since $\forall \mathbf{z}$, $\mathbf{z}`\mathbf{\Sigma}\mathbf{z} > \mathbf{0}$ (see Horn and Johnson, 1999: 396).

Furthermore, note that:

a) $z_t u_t$ is a martingale difference sequence with $\mathrm{Var}(z_t u_t) = \mathrm{E}(u_t^2)\mathrm{E}(z_t z_t`) = \sigma^2\mathbf{\Sigma}$ with $\mathbf{\Sigma}$ as given above;

b) Since $\mathrm{E}(\mathbf{z}_t \mathrm{u})^4$ is bounded by the stationarity assumption and since $\mathrm{E}(\mathrm{u}_t^2) < \infty$ it follows that $\mathrm{E}(\mathrm{y}_t)^4 < \infty$ and $\mathrm{E}(\mathrm{x}_t)^4 < \infty$;

c) Observe that the individual elements of the second moments matrix $(1/T)\sum(z_t u_t)^2$ have the following properties:

$$(1/T)\sum_{t=1}^{T} u_t^2 y_{t-i} y_{t-j} = (1/T)\sum_{t=1}^{T} (u_t^2 - \sigma^2) y_{t-i} y_{t-j} + (\sigma^2/T)\sum_{t=1}^{T} y_{t-i} y_{t-j},$$

$$(1/T)\sum_{t=1}^{T} u_t^2 x_{t-i} x_{t-j} = (1/T)\sum_{t=1}^{T} (u_t^2 - \sigma^2) x_{t-i} x_{t-j} + (\sigma^2/T)\sum_{t=1}^{T} x_{t-i} x_{t-j},$$

$$(1/T)\sum_{t=1}^{T} u_t^2 x_{t-i} y_{t-j} = (1/T)\sum_{t=1}^{T} (u_t^2 - \sigma^2) x_{t-i} y_{t-j} + (\sigma^2/T)\sum_{t=1}^{T} x_{t-i} y_{t-j},$$

where the first terms converge in probability to zero given consistency of u_t^2, i.e.,

$$(1/T)\sum_{t=1}^{T}(u_t^2-\sigma^2)y_{t-i}y_{t-j}\xrightarrow{p}0,$$
$$(1/T)\sum_{t=1}^{T}(u_t^2-\sigma^2)x_{t-i}x_{t-j}\xrightarrow{p}0,$$
$$(1/T)\sum_{t=1}^{T}(u_t^2-\sigma^2)x_{t-i}y_{t-j}\xrightarrow{p}0,$$

and subsequently:

$$(\sigma^2/T)\sum_{t=1}^{T}y_{t-i}y_{t-j}\xrightarrow{p}\sigma^2E(y_{t-i}y_{t-j}),$$
$$(\sigma^2/T)\sum_{t=1}^{T}x_{t-i}x_{t-j}\xrightarrow{p}\sigma^2E(x_{t-i}x_{t-j}),$$
$$(\sigma^2/T)\sum_{t=1}^{T}x_{t-i}y_{t-j}\xrightarrow{p}\sigma^2E(x_{t-i}y_{t-j}),$$

therefore, $(1/T)\sum_{t=1}^{T}(\mathbf{z}_t u)^2\xrightarrow{p}\sigma^2\Sigma$. From a), b) and c) it can be deduced (see White, 1984: 13) that: $(1/\sqrt{T})\sum_{t=1}^{T}\mathbf{z}_t u_t\xrightarrow{d}N(0,\sigma^2\Sigma)$.

Finally, it follows that:

$$\sqrt{T}(\hat{\beta}_{OLS}-\beta)=\left[(1/T)\sum_{t=1}^{T}\mathbf{z}_t\mathbf{z}_t'\right]^{-1}\left[(1/\sqrt{T})\sum_{t=1}^{T}\mathbf{z}_t u_t\right]\xrightarrow{d}N(0,\sigma^2\Sigma^{-1}),$$

which proves asymptotic normality of the OLS estimator.

Q.E.D.

References

Artus, J. and Kapur, I. (2000), *Staff Report for the 1999 Article IV Consultation with the Republic of Croatia*. IMF Staff Country Report No. 00/7. International Monetary Fund, Washington, D.C.

Banerjee, A., Dolado, J., Galbraith, J.W. and Hendry, D.F. (1993), *Co-integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, Advanced Texts in Econometrics. Oxford: Oxford Univ. Press.

Billingsley, P. (1968), *Convergence of Probability Measures*. New York: John Wiley.

Billingsley, P. (1986), *Probability and Measure*. New York: John Wiley.

Charemza, W.W. and Deadman, D.F. (1997), *New Directions in Econometric Practice: General to Specific Modelling, Cointegration and Vector Autoregression*. Cheltenham: Edward Elgar.

Davidson, J.E.H. (1994), *Stochastic Limit Theory*, Advanced Texts in Econometrics. Oxford: Oxford Univ. Press.

Davidson, J.E.H. (2000), *Econometric Theory*. Oxford: Blackwell.

Davidson, R. and MacKinnon, (1993), *Estimation and Inference in Econometrics*. New York: Oxford Univ. Press.

Deaton, A.S. (1992), *Understanding Consumption*. Oxford: Oxford University Press.
Deaton, A.S., ed. (1980), *Essays in the Theory and Measurement of Consumers' Behaviour*. Cambridge: Cambridge University Press.
Dickey, D.A. and Fuller, W.A. (1979), "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, **74**, pp. 427-31.
Dorsey, T, Jarvis, C., Meyermans, E. and Dayal-Gulati, A. (1995), *Republic of Croatia—Recent Economic Developments*. IMF Staff Country Report No. 95/131. International Monetary Fund, Washington, D.C.
Elkan, R. and Maggi, R. (2000), *Republic of Croatia: Selected Issues and Statistical Appendix*. IMF Staff Country Report No. 00/22. International Monetary Fund, Washington, D.C.
Elkan, R. and Temprano-Arroyo, H. (1998), *Republic of Croatia: Selected Issues and Statistical Appendix*. IMF Staff Country Report No. 98/90. International Monetary Fund, Washington, D.C.
Engle, R.F. and Granger, C.W.J. (1991), *Long-Run Economic Relationships: Readings in Cointegration*, Advanced Texts in Econometrics. Oxford: Oxford Univ. Press.
Engle, R.F., Hendry, D.F. and Richard, J.F. (1983), "Exogeneity," *Econometrica*, **51**, pp. 277-304.
Gilbert, C.L. (1986), "Professor Hendry's econometric methodology," *Oxford Bulletin of Economics and Statistics*, **48**/3, pp. 283-307.
Gourieroux, C. and Monfort, A. (1990), *Time Series and Dynamic Models*. Cambridge: Cambridge Univ. Press.
Griliches, Z. and Intriligator, M.D., eds. (1984), *Handbook of Econometrics*, Vol. I-III. Amsterdam: North-Holland.
Hamilton, J.D. (1994), *Time Series Analysis*. Princeton: Princeton Univ. Press.
Harris, R. (1995), *Cointegration Analysis in Econometric Modelling*. London: Prentice Hall.
Hendry, D.F. (1983), "Econometric modelling: The `consumption function` in retrospect," *Scottish Journal of Political Economy*, **30**, pp. 193-220.
Hendry, D.F. (1984), "Monte Carlo experimentation in econometrics," in Griliches and Intriligator, Ch. 16.
Hendry, D.F. (1987), "Econometric methodology: a personal perspective," in Bewley, T.F., pp. 29-48.
Hendry, D.F. (1995), *Dynamic Econometrics*, Advanced Texts in Econometrics. Oxford: Oxford Univ. Press.
Hendry, D.F. and Doornik, J.A. (1999), *Empirical Econometric Modelling Using PcGive: Vol. 1.*, Wickham: TCL.
Hendry, D.F., Pagan, A.R. and Sargan J.D. (1984), "Dynamic specification," in Griliches and Intriligator, Ch. 18.
Horn, R.A. and Johnson, C.R. (1999), *Matrix Analysis*. Cambridge: Cambridge Univ. Press.
Mikulic, D. and Lovrincevic, Z. (2000), "Quarterly GDP estimate of the Republic of Croatia from the first quarter of 1994 to the fourth quarter of 1996 -expenditure approach," *Economski pregled*, **51**, pp. 1006-32.
Phillips, P.C.B. (1986), "Understanding spurious regressions in econometrics," *Journal of Econometrics*, **33**, pp. 311-40.
Phillips, P.C.B. (1987), "Time series regression with a unit root," *Econometrica*, **55**, pp. 277-301.
Ploberger, W. and Phillips, P.C.B. (1998a), "Rissanen's theorem and econometric time series," in Keuzenkamp, H.A., McAleer, M. and Zellner, A.

Ploberger, W. and Phillips, P.C.B. (1998b), "Empirical limits for time series econometric models," (manuscript).
Rissanen, J. (1986), "Stochastic Complexity and Modelling," *Annals of Statistics*, **6**, pp. 461-64.
Rissanen, J. (1987), "Stochastic Complexity," *Journal of the Royal Statistical Society*, **49**, pp. 223-39.
Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
White, H. (1984), *Asymptotic Theory for Econometricians*. Orlando: Academic Press.