A TALE OF THREE SEASONAL ADJUSTMENT PROCEDURES: THE CASE OF SWEDEN'S GDP

Luigi Ermini
Stockholm School of Economics[1]

March 1998

Abstract

The paper compares the search for structural breaks in Sweden's GDP conducted with X-11 seasonally adjusted data, with seasonally undajusted data, and with temporally aggregated data. As a structural break (in 1980) is found only in the X-11 adjusted data, it is plausible to conclude that this break is due to data distortions (particularly, distortions caused by the application of the filter). However, this interpretation is only plausible *a posteriori*: had the seasonally unadjusted data not been available, the break found in the adjusted series could be just as well interpreted as a *break in the economy* and not as a break in the data. The study suggests that seasonally adjusted data should not be used when the unadjusted version is also available.

Keywords: X-11 filter, seasonal adjustment, seasonal unit roots
JEL CLassification: C12, C32.

1. Visiting professor from January to May 1998 at the Department of Economic Statistics. I thank the Stockholm School of Economics for giving me the opportunity to enjoy an excellent research environment.

A Tale of Three Seasonal Adjustment Procedures:
The Case of Sweden's GDP

1. Introduction

Three econometricians receive the task of investigating whether the changes and events occurred in the Swedish economy after 1980 (social conflicts in the early 80's, financial deregulation in the mid 80's, the crash of the property market and the recession of the early 90's, etc.) are reflected in significant structural breaks in Sweden's gross domestic product (GDP). To perform this task, each econometrician is instructed to follow a different procedure. One of the three econometricians is given a series of quarterly GDP already seasonally adjusted by Sweden's statistical office through application of the Census X-11 ARIMA filter. Another econometrician is given the original seasonally unadjusted quarterly series of GDP, and is instructed to seasonally adjust the series through direct modelling. The third econometrician is also given the original seasonally unadjusted series, but is instructed to seasonally adjust the series by temporally aggregating it to (quarterly) annual values. Figure 1 reports the unadjusted series *LY*, its temporal aggregation *LYta*, and the X-11 adjusted series *LYsa* in natural logarithms, and their overlap, from 1970 (1) to 1997 (3).

The purpose of this paper is to report the results obtained by the three econometricians, and to draw some conclusions from their comparison. To avoid possible prejudice (or even secret collusion between the investigators!), the three econometricians were told to follow a pre-specified common modelling strategy. This common strategy is described in section 2. The results of the exercise can be summarized as follows: (*i*) the econometrician working with seasonally adjusted GDP, *LYsa*, finds one unit root, and by taking first differences, finds that the best congruent model is a random walk from the beginning of the sample to 1980.II and an AR(1) with *negative* autoregressive coefficient after 1980; the latter remains stable until the present date; (*ii*) the econometrician working directly with the seasonally unadjusted GDP, *LY*, finds three roots on the unit circle, at + 1 and $\pm i$; by taking the appropriate data transformation to eliminate these three unit-circle roots, the best model is found to be an AR(3) with centered seasonal dummies, and is found to be stable throughout the sample; (*iii*) the econometrician working with the temporally aggregated GDP, *LYta*, finds a unit root in the data, and by taking first differences - which is equivalent to taking fourth differences of the original series - the best congruent model is found to be an AR(1) model with *positive* autoregressive coefficient, and this model is found to be stable throughout the whole sample. Note that, had the second econometrician also found the root -1 in the original data, the two latter procedures would have produced the same final model. A discussion on these three different results, and their implications for economic inference, follows in section 6.

Two main conclusions can be drawn from the exercise: (*i*) as regards the search for structural breaks in GDP, only the procedure based on the X-11 seasonally adjusted data revealed a structural break in the model - apart from three outliers in 1977(1), 1980(2) and 1992(4) highlighted by all three procedures; (*ii*) the manifest recession of the early 90's (see figure 1) is not reflected in any structural break. These findings raise two issues: (*a*) by comparing these three procedures, it seems plausible to conclude that the break in the X-11 adjusted data is due to data distortions (perhaps the change in base year, or a change in the adjustment filter); (*b*) had the seasonally unadjusted data not been available (as it is, for example, the case for most US series), the break found in the adjusted series could be just as well interpreted as a *break in the economy* and not as a break in the data. Overall, the study suggests that seasonally adjusted data should not be used when the unadjusted version is also available.

The paper is organized as follows. Section 2 describes the common modelling strategy. Section 3 reports on the model with seasonally adjusted data. Section 4 reports on the direct modelling of seasonally unadjusted data. Section 5 reports on the modelling of the temporally aggregated series. Section 6 discusses and compares the results.

2. The Modelling Strategy

The modelling strategy was set as follows (for further discussion and details, see among others Hendry [1995] and Ermini [1998]):

(i) search for congruent models: the search is restricted to pure autoregressive univariate models *in levels*, AR(p), with order p high enough to capture the inversion of possible (invertible) moving average components. The purpose of the search is to find an AR representation of the series x_t, such that the residuals

$$\hat{u}_t = \hat{a}(B)\, x_t + \hat{\Phi}\, D_t \qquad (1)$$

behave as close as possible as innovations; here ˆ indicates estimated values, and $a(B)$ is a polynomial of order p in the lag operator B, that is $a(B) = 1 - \sum_{j=1}^{p} a_j B^j$, with B such that $B^j x_t = x_{t-j}$; D_t is a vector of deterministic components (namely, a constant, a trend and dummies, seasonal and nonseasonal) and Φ is a row vector of parameters.

The three defining properties of innovations are zero-mean, orthogonality with the past and constant unconditional variance. An AR(p) model is said to be *congruent* if its residuals (1) satisfy these properties. Given the convenience of normality for standard inference, normality is added to the definition of congruence, though strictly speaking is not a requirement of innovations. Recall that normality implies conditional homoskedasticity; thus, to check the possibility that rejection of normality may be caused by autocorrelated square residuals, a test for autoregressive conditional heteroskedasticity, or ARCH (Engle [1982]) is added to the list.

The tests for congruence (that is, residual autocorrelation, normality, unconditional heteroskedasticity, and ARCH) are described in section 3. Note that, as the polynomial $\hat{a}(B)$ may have unit-circle roots, t-statistics related to the coefficients of the model do not have standard distributions; thus, inference on the model will not be made at this stage. The presence of unit-circle roots, however, does not invalidate the congruence tests on the residuals, as long as these are stationary.

Possible causes of normality rejection are conditionally heteroskedastic innovations (ARCH effects) and the presence of strong outliers. If ARCH effects are not present, the investigator proceeds to verify the possibility that the rejection of normality is directly related to the presence of outliers. Note, however, that the presence of outliers can bias the ARCH test toward rejection; see, for example, Franses and Van Dijk [1997]. If the elimination of outliers through impulse dummies does not restore normality, the investigator may ignore the rejection of normality on the grounds (or hope) that the residuals admit a central limit theorem.

In practice, the search starts with a relatively high AR order p_0; if AR(p_0) is not congruent, one need to re-start with a higher value. If AR(p_0) is congruent, it may be that lower-order AR models are also congruent; in this case, the search continues by sequential reduction of the order p until congruence is lost. When this is the case, the modelling strategy moves on to the next step. The starting value is chosen as $p_0 = 7$.

(ii) select a congruent model: if more than a congruent model is found, the final model is chosen as the one that minimizes a given selection criteria. The following selection criteria will be considered: the ML estimator of the residual standard error ($\hat{\sigma}_u$), the Schwarz criterion (SC), the Hannan-Quinn criterion (HQ), and the Final Prediction Error (FPE) (for further details, see Judge, Griffith, Hill, Luetkephol and Lee [1985], and Hendry and Doornik [1996]). The final model will be chosen as the model that receives the highest consensus from these four selection procedures. Note that, as a consequence of possible unit-roots in the polynomial $\hat{a}(B)$, selection procedures based, for example, on sequences of t-tests applied to the highest-lag coefficient of the congruent models cannot be adopted here.

(iii) test the final model for structural breaks in parameters: firstly, the chosen final model is re-estimated over the full sample to exploit as many degrees of freedom as possible; the sequential search for congruent models, in fact, uses a constant sample size of $T - p_0$, p_0 being the number of observations lost in estimating the initial AR(p_0); as $p^* \leq p_0$, the final model AR(p^*) can be re-estimated by using the larger sample size $T - p^*$. Before re-estimation, the final model is tested for roots on the unit circle (seasonal and non-seasonal), to separate the stable autoregressive component. The re-estimation is then applied only to the latter to benefit from standard inference, and it is done with recursive methods to test possible structural instability in the coefficients of the autoregressive component, as described below.

3. Modelling the X-11 Seasonally Adjusted GDP

Starting with p = 7, the OLS estimation of the X-11 seasonally adjusted series in levels, with the vector D_t including a constant, a trend and three centered seasonal summies, produced the following diagnostics (p-values in square brakets) [2]

$$\begin{aligned} &AR(5): \ F(5,87) = 1.627 \ [0.161] \\ &Normality: \ \chi^2(2) = 14.633 \ [0.000] \\ &UncH: \ F(19,72) = 0.389 \ [0.988] \\ &ARCH(4): \ F(4,84) = 0.771 \ [0.547] \ , \end{aligned}$$

where (for further details, see Hendry and Doornik [1996]):

(i) AR(r): is the the Lagrange-multiplier test for the significance of the AR(r) model for the residuals (1); under the null of orthogonal residuals, the statistic $T\,R^2$, where T is the sample size, is distributed as $\chi^2(r)$ in large samples; for small samples, it is preferable to use the equivalent F-form $R^2\,(T-k-r)/(1-R^2)r$ which is distributed as $F(r,\ T-k-r)$, where k is the number of parameters in the model, i.e. p plus the dimension of Φ;

(ii) Normality $\chi^2(2)$: is the normality test based on the $\chi^2(2)$ statistic of Doornik and Hansen [1994], which is better suited for small samples. Under the null, the test has zero skewness and a kurtosis equal to 3;

(iii) UncH: is the unconditional heteroskedasticity test of White's [1980]. The test is conducted by regressing the squared residuals against the levels, the square values of the regressors, and the cross-products of the regressors. Under the null that all the coefficients are jointly insignificant, the statistic is distributed as $F(n,\ T-p-n)$ with $n = p + p\,(p+1)/2$;

(iv) ARCH(r): is the test for the autocorrelation of the squared residuals, that is a test for the joint significance of the coefficients of the regression of $\hat{u}_t^2$ on its first r lags, $\hat{u}_{t-1}^2,\ldots,\ \hat{u}_{t-r}^2$. Under the null of no ARCH effects, the statistic $T\,R^2$ is asymptotically distributed as $\chi^2(r)$; the equivalent F-form is reported here.

As normality is rejected but ARCH effects are not present, the investigator proceeds to detect possible outliers through recursive estimation. Figure 2 reports: the recursive estimates of the constant term with ± twice the standard error; the estimated residuals with $\pm\, 2\,\hat{\sigma}_{u,t}$; the one-step ahead Chow test (that is, the test that the model estimated with n observations, with $n \le T$, is not statistically different from the model estimated with n - 1 observations); and the break-point Chow test (that is, a test that the model estimated with n observations is not statistically different from the model estimated over the entire sample *T*). Note that, although the latter two graphs include a straight line to denote 5% level of significance, the actual level of significance for these two Chow tests is much higher due to the sequential nature of the recursive tests; thus, these tests should only be used as informative about the *location* of possible breaks rather than informative about their significance.

The recursive estimation reveals the presence of a huge outlier at 1980(2), which also appears to be responsible for the jump in the residual variance at the same date. Based on this observation, an impulse dummy dated 1980(2) is added to the vector D_t in (1); upon re-estimating the AR(7) model, the problem of non-normality disappears with a $\chi^2(2)$ statistic of 2.393 and a *p*-value of 0.302 (for the other diagnostics, see table 1). The tests of congruence for the sequence of AR(*p*) models for $0 \leq p \leq 7$, with D_t including constant, trend, three centered seasonals and an impulse dummy at 1980(2), are reported in table 1.

TABLE 1. Congruence Tests for the X-11 Seasonally Adjusted GDP, *LYsa*

	$p = 7$	$p = 6$	$p = 5$	$p = 4$	$p = 3$	$p = 2$	$p = 1$	$p = 0$
AR(5)	0.282 [0.922]	0.962 [0.446]	1.560 [0.180]	1.209 [0.312]	1.445 [0.215]	1.255 [0.290]	0.771 [0.573]	76.615* [0.000]
Norm $\chi^2(2)$	2.393 [0.302]	2.473 [0.290]	3.646 [0.162]	3.311 [0.191]	2.962 [0.227]	3.108 [0.211]	1.777 [0.411]	0.708 [0.702]
UncH	0.561 [0.927]	0.576 [0.905]	0.680 [0.805]	0.741 [0.727]	0.939 [0.512]	1.143 [0.341]	0.446 [0.890]	2.391* [0.034]
ARCH(4)	0.991 [0.417]	1.365 [0.253]	1.361 [0.254]	1.177 [0.326]	1.149 [0.212]	1.541 [0.197]	0.986 [0.419]	36.88* [0.000]
$\hat{\sigma}_u$	0.0111#	0.0113	0.0114	0.0115	0.0115	0.0114	0.0116	0.0265
SC	-8.549	-8.556	-8.562	-8.582	-8.615	-8.658	-8.674#	-7.056
HQ	-8.747	-8.737	-8.729	-8.734	-8.751	-8.800#	-8.780	-7.147
FPE (10^{-3})	0.139	0.142	0.144	0.145	0.144	0.141#	0.143	0.740

The values in square brackets are the *p*-values of the various statistics; * indicates 5%-level significance and # indicates minimum value. The sample size is $T = 103$. Note that congruence is lost at $p = 0$. As regards the choice of the final model, the AR(2) model receives higher consensus than AR(1), and it is thus selected as the model to be studied next.

---

2. All the estimations were done by using the software package PcGive version 9.0 (Hendry and Doornik [1996]).

The ADF unit root test with one lag of first-differences (corresponding to the AR(2) in levels previously found) does not reject the presence of unit root, with a statistic of -2.235 against an ADF critical value of -3.453. The final model is thus taken to be an AR(1) model in first differences; recall that these are the growth rates of X-11 s.a. GDP.

As in re-estimating the AR(1) model in first differences with the full sample the three centered seasonals and the trend turned out to be highly insignificant, the model was re-estimated with only a constant, an impulse dummy at 1980(2) and lagged first differences; table 2 reports the results. The statistic HCSE is White's [1980] heteroskedastic-consistent standard error of the model parameters; a noticeable difference between HCSE and the usual standard error - as in the case, for example, of $\Delta LYsa_{t-1}$ - may indicate that the associated coefficient is likely to be unstable. The sample size is now T = 109.

TABLE 2. AR(1) model for $\Delta LYsa$ with impulse dummy

Variable	Coefficient	Std. Error	*t*-value	HCSE	Partial Corr.
Constant	0.0053	0.0012	4.388	0.0013	0.154
$\Delta LYsa_{t-1}$	- 0.117	0.0858	- 1.364	0.118	0.017
Impulse 1980(2)	-0.054	0.012	-4.508	0.003	0.161

Model R^2	0.189	$\hat{\sigma}_u$	0.0118	RSS	0.015	DW	1.93
AR $F(5,102)$	0.628 [0.678]	Norm. χ^2	1.948 [0.377]	UncH	1.787 [0.154]	ARCH $F(4,99)$	1.500 [0.208]

Figure 3 reports the corresponding recursive statistics. We see that the elimination of the outlier at 1980(2) makes the variance of residuals noticeably more stable; that there remain two significant outliers at 1977(1) and 1992(4); and that the model is overall stable, even though the coefficient of lagged first differences appears to be unstable but insignificant. Note that the instability of this coefficient is also signalled by the big difference between its HCSE and its usual standard error, as reported in table 2. By adding the two impulse dummies at 1977(1) and 1992(4), the model improves further, as its standard error decreases from 0.0118 to 0.0111, and the R^2 goes from 0.189 to 0.296; moreover, the coefficient of $\Delta LYsa_{t-1}$ appears less stable, but still insignificant. Before eliminating from the model the latter regressor, however, it might be of interest to investigate whether the overall insignificance of its coefficient could be the result of a break of value in early 1980, as suggested by its recursive estimate of figure 3.

To include this break, the following model was re-estimated

$$\Delta LYsa_t = \alpha_0\, \Delta LYsa_{t-1} + \alpha_1\, \Delta LY{*}sa_{t-1} + d + d_1 + d_2 + d_3 + u_t, \qquad (2)$$

where $\Delta LY{*}sa_{t-1}$ is obtained by multiplying $\Delta LYsa_{t-1}$ by the step dummy that is equal to zero up to 1980(2) and equal to one from 1980(3) on; and where d_1, d_2 and d_3 are the three impulse dummies for 1977(1), 1980(2) and 1992(4), respectively. By estimating model (2), the coefficient α_0 was found to be insignificant at the 5% level (p-value of the t-statistic 0.47), but the coefficient α_1 was found to be significant (p-value of the t-statistic 0.022). To obtain a more parsimonious model, the insignificant regressor is eliminated, obtaining the model reported in table 3. This is the model chosen as the final model of the investigation with the X-11 seasonally adjusted series.

TABLE 3. Final Model for $\Delta LYsa$

Variable	Coefficient	Std. Error	t-value	HCSE	Partial Corr.
Constant	0.0059	0.0011	5.497	0.0011	0.225
ΔLY^*sa_{t-1}	- 0.274	0.099	- 2.77	0.111	0.07
d_1	- 0.033	0.011	- 3.01	0.001	0.08
d_2	- 0.058	0.010	- 5.30	0.001	
d_3	- 0.033	0.011	- 3.01	0.001	0.08

Model R^2	0.327	$\hat{\sigma}_u$	0.0109	RSS	0.012	DW	2.06
AR $F(5,99)$	0.760 [0.581]	Norm. χ^2	1.103 [0.576]	ARCH $F(4,96)$	0.916 [0.458]	Unc	0.462 [0.804]

The main conclusion from this investigation is that, apart from three outliers, the X-11 seasonally adjusted series is well fitted by a random walk between 1970 and 1980 and by an AR(1) with *negative* coefficient thereafter.

4. Modelling the Seasonally Unadjusted GDP

The fundamental assumption in modelling seasonality is that whatever the nature of the seasonal component in the series, this component persists forever, in the sense that its expectation in the distant future is not zero. Given persistence, the modelling of seasonality mirrors exactly the same dichotomy that the investigator faces when modelling long-term growth, that is whether the persistence of seasonality is deterministic or stochastic. In the former case, seasonality is modelled by seasonal dummies added to a series which is "stationary" at the seasonal frequencies (i.e., has no unit-circle roots at these frequencies); in the latter case, seasonality is modelled by the presence of unit-circle roots at these frequencies. With quarterly data, the deterministic case entails the addition of four (or three centered) seasonal dummies, whereas the stochastic case entails the presence of unit-circle roots $\pm i$ and -1, corresponding to the seasonal frequencies $\pi/2$ and π, respectively.

To discriminate between the two cases, one can conduct an HEGY test (Hylleberg, Engle, Granger, Yoo [1990]), whose null is the presence of the roots ± 1 and $\pm i$, or equivalently the presence of the factor $1 - B^4$, in the autoregressive component of the series (thus the distributions of the test statistics are non-standard). We should emphasize that this test has low power when the true DGP has factor $1 - \alpha B^4$ with α close to 1; for further discussion on this test, see Franses [1996] and the references therein. Depending on the test outcome, the investigator eliminates from the series whichever unit-circle root has been found by applying appropriate data transformations on the series: if all four roots are found (and none is repeated), these are eliminated by fourth differencing, i.e. by transforming the data as $(1 - B^4)\, x_t$; if only roots at 1 and -1 are found, these are eliminated by second-differencing, i.e. by transforming the data as $(1 + B)(1 - B)\, x_t = (1 - B^2)\, x_t$; and so forth.

As regards the modelling strategy specified in section 2, the HEGY test is performed in place of the ADF test for non-seasonal unit root, and steps (*i*) and (*ii*) are performed as described above. Before entering into the modelling details, it is interesting to look at some sample properties. Figure 4 reports the autocorrelogram, the spectrum and the density function for the seasonally unadjusted series in levels, in first-differences, in second-differences, and in fourth-differences. Without risk of prejudging the outcome of the HEGY test, we observe that the autocorrelograms remain remarkably persistent in all cases except in fourth-differencing, where it appears to attain stationarity.

Table 4 reports the results of the search for congruent models of the series LY (seasonally unadjusted Sweden's GDP in logs), from $p = 7$ to 3; as before, the vector D_t contains a constant, a trend and three centered seasonal dummies. Congruence is lost at $p = 4$, and normality is never rejected; the model chosen for the next step - the HEGY test - is thus the AR(6) model. Note once again that, due to the possible presence of unit-circle roots, inference on the model parameters cannot be done at this stage.

The HEGY tests entails the application of ordinary least squares to the auxiliary regression

$$D(B)\, \Delta_4\, x_t = \Phi\, D_t + \sum_{j=1}^{4} \pi_j\, x_{j,t-1} + \varepsilon_t\,, \qquad (3)$$

where $D(B)$ is an autoregressive polynomial of order p - 4 (in our case, of order 2), and where the auxiliary regressors are defined as:

TABLE 4. Congruence Tests for the Seasonally Unadjusted GDP, *LY*

	$p = 7$	$p = 6$	$p = 5$	$p = 4$	$p = 3$
AR(5)	0.121 [0.987]	0.108 [0.990]	2.277 [0.054]	7.371* [0.000]	9.342* [0.000]
Norm $\chi^2(2)$	5.03 [0.081]	5.188 [0.075]	5.474 [0.065]	4.726 [0.094]	6.816* [0.033]
UncH	0.499 [0.955]	0.414 [0.978]	0.385 [0.979]	0.582 [0.862]	0.773 [0.666]
ARCH(4)	0.381 [0.822]	0.313 [0.868]	0.955 [0.436]	0.778 [0.542]	0.448 [0.774]
$\hat{\sigma}_u$	0.0140	0.0140#	0.0147	0.0164	0.0174
SC	-8.119	-8.162#	-8.098	-7.908	-7.820
HQ	-8.300	-8.328#	-8.249	-8.044	-7.941
FPE (10^{-3})	0.220	0.216#	0.236	0.293	0.328

$$x_{1,t} = (1 + B + B^2 + B^3)\, x_t \tag{4}$$

$$x_{2,t} = -\,(1 - B + B^2 - B^3)\, x_t$$

$$x_{3,t} = -\,(1 - B^2)\, x_t$$

$$x_{4,t} = x_{3,t-1}\ .$$

The test for the root 1 is a one-sided test on the t-statistic of π_1, which is zero under the null, and less than 1 under the alternative of stable root at frequency zero; similarly for the test of the root -1, which is a one-sided test on the t-statistic of π_2. The test for $\pm\, i$ is instead an F-test on the joint significance of both π_3 and π_4. Note that the F-test of the joint significance of all four parameters π_j, j = 1,..., 4 is equivalent to an ADF test (for further details and comments, see Franses [1996]).

Table 5 reports the test results for the two cases where D_t includes a constant (C), a trend (T) and centered seasonal dummies (CD), or only constant and seasonal dummies. The sample size is $T = 105$; * indicates 5% significance, and the values in parenthesis are the critical values taken from Franses [1966, p.67] for a sample size of 80 and 120, respectively. As the test does not reject a unit root at 1 and $\pm\, i$ but it does reject the root at -1, the modelling process moves on to re-estimating the transformed series y_t, obtained as $(1 + B^2)(1 - B)\, x_t$. As this filter is a third-order polynomial and the congruent model for x_t is AR(6), the series y_t is re-estimated with a third-order polynomial $d(B)$; according to the outcome of the HEGY test, the latter supposedly

TABLE 5. HEGY Tests for Seasonally Unadjusted GDP, *LY*

	(C, T, CD)			(C, CD)		
test	statistic	5% (T = 80)	5% (T = 120)	statistic	5% (T = 80)	5% (T = 120)
$t(\pi_1)$	-3.233	-3.37	-3.40	-1.109	-2.81	-2.83
$t(\pi_2)$	-3.722*	-2.81	-2.83	-3.546*	-2.80	-2.82
$F(\pi_3, \pi_4)$	4.995	6.57	6.66	5.227	6.62	6.70
$F(\pi_1,.., \pi_4)$	9.229*	6.47	6.41	6.46*	5.70	5.67

does not contain unit-circle roots. Upon re-estimation, the trend is found to be insignificant, with *p*-value of the *t*-statistic equal to 0.49, but the same three outliers identified in the previous section are significant. Thus, table 6 reports the estimates of the AR(3) model without trend but with the three impulse dummies d_1, d_2 and d_3 defined above.

TABLE 6. AR(3) Model for $y_t = (1 + B^2)(1 - B)\, LY_t$

Variable	Coefficient	Std. Error	t-value	HCSE	Partial Corr.
Constant	0.0055	0.002	2.946	0.002	0.084
y_{t-1}	- 0.294	0.091	- 3.234	0.103	0.10
y_{t-2}	0.580	0.082	7.105	0.080	0.347
y_{t-3}	0.184	0.086	2.082	0.093	0.044
CD_1	- 0.11	0.030	- 3.071	0.035	0.13
CD_2	0.003	0.004	0.82	0.004	0.007
CD_3	- 0.114	0.030	- 3.86	0.035	0.136
d_1	- 0.05	0.015	- 3.37	0.003	.107
d_2	- 0.03	0.015	- 1.95	0.003	0.04
d_3	- 0.04	0.015	- 2.47	0.003	0.06

Model R^2	0.994	$\hat{\sigma}_u$	0.0143	RSS	0.019	DW	2.07
AR F(5,90)	0.488 [0.784]	Norm. χ^2	0.674 [0.714]	UncH	0.969 [0.979]	ARCH F(4,87)	0.110 [0.515]

We see that the HCSE's of the lagged regressors appear close to the usual standard errors, thus signalling structural stability. In fact, by adding to the model the regressor y^*_{t-1} defined to be zero up to 1980(2) and equal to y_{t-1} from 1980(3) on (compare with a similar setting in (2)), the latter was found insignificant (*p*-value of the *t*-statistic of 0.376), as opposed to the finding of the previous section. Thus, with seasonally unadjusted data there is no significant break in the model. The model of table 6 is taken to be the final model of the investigation of the second econometrician.

5. Modelling the Temporally Aggregated GDP

There are two main reasons to conduct this third investigation: one is the low power of HEGY test, and thus the possibility that with a relatively small sample of about 100 observations its outcome is in fact indeterminate. The second reason is rooted on a legacy of the past: before the appearance in the literature of the HEGY test for seasonal unit roots, the common practice in modelling stochastic seasonality was to model the fourth differences of the data (for a discussion, see for example Granger and Newbold [1986]). As fourth-differencing is equivalent to summing up first differences, i.e. $1 - B^4 = S(B)(1 - B)$ where $S(B) = 1 + B + B^2 + B^3$, the third investigator is instructed to model first the temporally aggregated series $LYta = S(B)\, LY$, so to leave open the possibility of a non-seasonal unit root.

Starting with $p = 4$ (i.e. 7 minus the order of $S(B)$ which is 3), the tests for congruence of the AR(p) models for *LYta*, with D_t including constant, trend and centered seasonal dummies, are reported in table 7.

TABLE 7. Congruence Tests for the Temporally Aggregated GDP, *LYta*

	$p = 4$	$p = 3$	$p = 2$	$p = 1$
AR(5)	1.164 [0.333]	0.973 [0.439]	1.939 [0.095]	16.638* [0.000]
Norm $\chi^2(2)$	3.215 [0.200]	3.330 [0.189]	2.964 [0.227]	6.128* [0.047]
UncH	1.019 [0.467]	1.236 [0.236]	1.142 [0.327	0.691 [0.770]
ARCH(4)	0.676 [0.610]	0.695 [0.597]	1.350 [0.258]	7.588* [0.000]
$\hat{\sigma}_u$ SC HQ FPE (10^{-3})	0.0156 -8.016 -8.153 0.263	0.0155# -8.059# -8.180# 0.258#	0.0158 -8.049 -8.155 0.267	0.021 -7.51 -7.600 0.470

Unsurprisingly, the chosen model is AR(3), as it corresponds to the model AR(6) chosen in the previous section; note also that normality is not rejected for congruent models. Furthermore, the ADF test does not reject the hypothesis of unit root with a statistic of -3.102 against a 5%-level critical value of -3.454. As in re-estimating this model with the full sample the three seasonal dummies, the trend and the second-lag regressor turned out to be insignificant, the model was re-estimated recursively with constant and one lagged regressor only, obtaining the values reported in table 8. Figure 5 reports the associated recursive statistics. Note that the model parameters are fairly stable (as also confirmed by the small differences between HCSE’s and standard errors in table 8), and that an outlier at 1984(1) appears which was not present in the

TABLE 8. AR(1) Model for $\Delta_4 LY$

Variable	Coefficient	Std. Error	t-value	HCSE	Partial Corr.
Constant	0.0059	0.0020	2.93	0.0022	0.077
$\Delta_4 LY_{t-1}$	0.629	0.075	8.38	0.086	0.41

Model R^2	0.41	$\hat{\sigma}_u$	0.0163	RSS	0.028	DW	2.1
AR F(5,99)	1.97 [0.09]	Norm. χ^2	0.50 [0.779]	UncH	1.14 [0.28]	ARCH F(4,96)	1.30 [0.325]

X-11 s.a. data. To obtain a better fit, the following three outliers were added, d_1 = 1982(2), d_4 = 1984(1) and d_3 = 1992(4), obtaining the results of table 9.

TABLE 9. Final Model for $\Delta_4 LY$

Variable	Coefficient	Std. Error	t-value	HCSE	Partial Corr.
Constant	0.0061	0.0019	3.27	0.0022	0.096
$\Delta_4 LY_{t-1}$	0.636	0.071	8.96	0.080	0.44
d_2	- 0.45	0.015	- 2.91	0.003	0.08
d_4	- 0.035	0.015	2.61	0.002	0.06
d_3	-0.035	0.015	- 2.30	0.003	0.05

Model R^2	0.51	$\hat{\sigma}_u$	0.0150	RSS	0.023	DW	2.1
AR F(5,96)	0.875 [0.50]	Norm. χ^2	0.08 [0.96]	UncH	0.995 [0.43]	ARCH F(4,93)	0.575 [0.68]

This latter model is taken to be the final model of the investigation of the third econometrician.

6. A Comment and Conclusion

The results of the three investigations can be summarized as follows:

(*i*) using X-11 seasonally adjusted data (*LYsa*), the search for structural breaks reveals three outliers at 1972(1), 1982(2) and 1992(4) and a break in the model at 1980(3); before this date the best fit is obtained with a random walk model, and after this date with an ARI(1,1) model with *negative* coefficient.

(*ii*) using seasonally unadjusted data (*LY*) and performing an HEGY test to detect the presence of unit-circle roots, the search for structural break reveals the same three outliers, but no break in the model. Having found roots at 1 and $\pm i$, and eliminating them by transforming the data as $y_t = (1 - B)(1 + B^2)\, LY_t$, the best fit is found with an AR(3) with three centered seasonal dummies. This model is stable throughout the sample.

(*iii*) using temporally aggregated data (*LYta*), the search for structural break reveals two of the same outliers revealed in (*i*) and (*ii*), namely 1980(2) and 1992(4), and an additional outlier at 1984(1), but no break in the model. The best fit is found with an AR(1) in fourth differences with *positive* coefficient. This model also is stable throughout the sample.

Having received the three separate reports from the three investigators, the natural question to ask is which model should we consider to analyze possible economic implications. The latter two models use the same seasonally unadjusted data, and their difference lies on whether the seasonal persistence manifest at frequency π (figure 2) is deterministic or stochastic in nature. In the former case, the persistence is modelled with seasonally dummies (model (*ii*)); in the latter case with a root at -1 (model (*iii*)). The choice between these two models depends on the credibility of the HEGY test, that is on whether one believes that with 100 observations the HEGY test has enough power to discriminate between these two possibilities. Notice that an accurate analysis of the residuals of both models reveals that nothing is "wrong", although from a theoretical point of view if the root -1 is not present in the data (that is, of the outcome of the HEGY test is right), then the residuals of model (*iii*) should exhibit a significant autocorrelation of 0.5 at lag one, as fourth differencing would introduce the factor $1 + B$ into the moving average component. By noting that model (*iii*) is more compact and thus more interpretable than model (*ii*), and for lack of better criteria, between the two models we prefer to choose model (*iii*) as the better way to model seasonally unadjusted Sweden's GDP.

The next question is the choice between model (*i*) and model (*iii*). The answer to this question differs, depending on the use that is made of the chosen model. If the model is used, say, for forecasting, then model (*i*) could be preferred, having a lower standard deviation (0.011 against 0.015). If the model is used instead to infer the existence of some structural change in the Swedish economy - which is in fact the original purpose of the study - then the two models are no longer comparable, as they offer two completely different results: although both models deny structural change in correspondence with the recession of the early 1990's, model (*i*) reveals an important change in early 1980, whereas model (*iii*) is stable throughout.

Several conclusions can be drawn from this study. Firstly, by comparing the outcomes of the three procedures, it is plausible to conclude that the break in the X-11 adjusted data is due to data distortions (for example, the result of the change of base year, or a change of the X-11 filter). However, this interpretation is only plausible *a posteriori*: had the seasonally unadjusted data not been available (as it is, for example, the case for most US series), the break found in the adjusted series could be just as well interpreted as a *break in the economy* and not as a break in the data. It follows that, as a rule, some caution should be spent in interpreting structural breaks, especially when using seasonally adjusted data; a corollary to this conclusion is that seasonally adjusted data should not be used when the unadjusted version is also available.

References

Dickey D.A., and W.A. Fuller, 1981, "Likelihood ratio statistics for autoregressive time series with a unit root", *Econometrica*, 49, 1057-1072.

Doornik J.A., and H. Hansen, 1994, "A practical test for univariate and multivariate normality", discussion paper, Nuffield College.

Engle R.F., 1982, "Autoregressive conditional heteroskedasticity, with estimates of the variance of United Kingdom inflation", *Econometrica*, 50, 987-1007.

Ermini L., 1998, *Macroeconometric Modelling*, manuscript.

Franses P.H., 1996, *Periodicity and stochastic trends in economic time-series*, Oxford University Press.

Franses P.H., and D. Van Dijk, 1997, "Do we often find ARCH because of neglected outliers?", working paper, Tinbergeb Institute.

Granger C.W.J., and P. Newbold, 1986, *Forecasting Economic Time Series*, Academic Press.
Hendry D.F., 1995, *Dynamic Econometrics*, Oxford University Press.

Hendry D.F., and J.A. Doornik, 1996, *Empirical Econometric Modelling Using PcGive 9.0 for Windows*, International Thompson, London.

Hylleberg S., R.F. Engle, C.W.J. Granger, and B.S. Yoo, 1990, "Seasonal integration and cointegration", *Journal of Econometrics*, 44, 215-38.

Judge G.G., W.E. Griffiths, R.C. Hill, H. Luetkepohl, and T.C. Lee, 1985, *The Theory and Practice of Econometrics*, John Wiley, New York.

White H., 1980, "A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica*, 48, 817-838.