No. 9109

DETECTING LEVEL SHIFTS IN TIME SERIES: MISSPECIFICATION AND A PROPOSED SOLUTION

by

Nathan S. Balke*

Research Paper

Federal Reserve Bank of Dallas

No. 9109

DETECTING LEVEL SHIFTS IN TIME SERIES: MISSPECIFICATION AND A PROPOSED SOLUTION

by

Nathan S. Balke*

June 1991

*Assistant Professor of Economics, Southern Methodist University, Dallas, Texas and Visiting Scholar, Research Department, Federal Reserve Bank of Dallas. I would like to thank Tom Fomby for many helpful discussions and much encouragement and Mark Wynne for comments on a previous draft. Ellah Pina helped with the manuscript and tables.

Detecting Level Shifts in Time Series:

Misspecification and a Proposed Solution*

Nathan S. Balke

Assistant Professor of Economics
Southern Methodist University

and

Research Associate
Federal Reserve Bank of Dallas

June 1991
revised September 1991

Abstract This paper demonstrates the difficulty that traditional specification, estimation, detection, and removal outlier detection methods have in identifying level shifts in time series. A simple modification to the well-known outlier/level shift detection algorithm described in Tsay (1988) is proposed that dramatically improves the ability to correctly identify level shifts. This modification involves combining an outlier search that is initialized by an ARMA(0,0) model with an outlier search that employs a traditionally specified ARMA model. The results of the outlier searches are used to specify a single intervention whose final specification is then determined by stepwise reduction. This "combine/reduce" approach is relatively easy to implement and appears to be quite effective in practice.

Key Words: Level shifts, outliers, ARMA models, intervention models

* I would to like thank Tom Fomby for many helpful discussions and much encouragement and Mark Wynne for comments on a previous draft. Shengyi Guo provided excellent research assistance. Ellah Pina helped with the manuscript and tables. The views expressed in this paper are solely those of the author and should not be attributed to the Federal Reserve Bank of Dallas or the Federal Reserve System.

1. Introduction

Anomalies such as outliers and level shifts are quite common in time series data. For example, Balke and Fomby (1991) examined fifteen macroeconomic time series and found outliers and/or level shifts in almost every series. These extraordinary observations (or sequence of extraordinary observations) are often associated with identifiable events such as wars, strikes, and changes in policy regimes. In addition, the presence of outliers and level shifts pose problems for the identification and estimation of ARIMA models, see Chang (1982), Chang, Tiao, and Chen (1988), and Chen and Tiao (1990). If the type and dates of these disturbances were known, their effects could be controlled with intervention analysis, as in Box and Tiao (1975). In practice, however, the type and date of an intervention is seldom known a priori.

As a result, methods of identifying and correcting for outliers and level shifts in time series have recently attracted much interest. Among the several approaches include robust estimation techniques (Martin (1980)), Bayesian analysis (Abraham and Box (1979), McCulloch and Tsay (1991)), and "leave-k-out" diagnostics (Bruce and Martin (1989)). In addition, iterative procedures proposed and employed by Chang (1982), Tsay (1986), and Chang, Tiao, and Chen (1988) have been used with success in the identification of outliers when the number and times of the disturbances are unknown. These iterative procedures have also been adapted for identification of level shifts (Chen and Tiao (1990)). Tsay (1988) provides a unified treatment of outliers and level shifts in the context of these iterative procedures. The procedure outlined by Tsay is particularly useful for applied work because it is quite flexible (it can be used to search for both additive and innovative outliers as well as level shifts) and is relatively easy to implement.

Unfortunately, as I show below, the procedure suggested by Tsay (1988) does not always perform satisfactorily when level shifts are present. The iterative procedure described by Tsay consists of several distinct steps: specifying and estimating an initial ARMA model, detecting outliers based on a prespecified criteria, removing the outliers, and then respecifying and reestimating the ARMA model. This sequence is repeated until no remaining outliers are detected. However, the presence of level shifts causes serious problems in the specification and estimation of the initial ARMA model (Chen and Tiao (1990)) which in turn affects the subsequent detection and removal steps. As a consequence, the procedure described by Tsay may misidentify outliers and level shifts.

To deal with this misspecification problem, I suggest a simple modification to the outlier identification approach of Tsay. The modification is as follows:

(i) In addition to conducting an outlier search based on specifying and estimating an initial ARMA model as originally proposed in Tsay (1988), a separate search employing an ARMA(0,0)--or "white noise"--model as the initial ARMA specification is conducted.

(ii) The results from the two outlier searches are combined to form a single intervention model (Box and Tiao (1975)). The final specification is determined by a stepwise reduction of this comprehensive intervention model.

This "combine/reduce" (hereafter CR) approach appears to be much more capable of handling level shift outliers than is the original Tsay procedure.

The remainder of the paper is organized as follows. In Section 2, I present two examples using actual data that demonstrate the difficulty the Tsay procedure can have in correctly identifying level shifts and the improvement offered by the CR method proposed here. Section 3 describes the outlier search procedure examined in Tsay (1988) in more detail. The analytical and Monte Carlo analyses in Section 4 document the problems that level shifts pose for the traditional outlier search method. In particular, I examine the sensitivity of the outlier search with respect to the choice of the initial ARMA specification. In Section 5, I outline a simple modification to the Tsay (1988) outlier search that was employed on the two data series examined in Section 2. Two seperate Monte Carlo experiments suggests that this modification results in a substantial improvement in the Tsay procedure when level shifts are present.

2. Some Examples

To motivate the problem that level shifts pose for the outlier identification procedure described by Tsay (1988) and the potential effectiveness of the proposed modification, I consider two examples.

First, consider the quarterly compensation per hour for the nonfarm business sector in logarithms (series LBCPU from CITIBASE). The Tsay outlier search procedure results in the following specification for the sample 1958Q1 to 1990Q2:

$$(1)\quad (1 - B)Y_t = \underset{(.003)}{.008}\ AO1960Q1_t$$

$$+ [1/(1 - \underset{(.090)}{.243}B - \underset{(.091)}{.234}B^2 - \underset{(.090)}{.229}B^3 - \underset{(.089)}{.151}B^4)]\ [\ \underset{(.001)}{.002} + a_t]$$

$$\hat{\sigma}_a = .0046,$$

where $B^iX_t = X_{t-i}$, and the standard errors are in parentheses. AO1960Q1t is an additive outlier intervention with AO1960Q1 = 1 when t = 1960Q1 and 0 otherwise. Figure 1a plots compensation per hour, nonfarm business and the outlier component for the Tsay procedure (the unconditional sample mean of the series has been added to the outlier component).

On the other hand, following the combine/reduce (CR) outlier procedure described briefly in the introduction and in more detail in Section 5 leads to the following specification:

$$(2) \quad (1 - B)Y_t = \underset{(.001)}{.007}\, LS1968Q1_t + \underset{(.001)}{.005}\, LS1972Q4_t - \underset{(.001)}{.012}\, LS1982Q2_t$$
$$+ \underset{(.004)}{.013}\, AO1960Q1_t + [1/(1 - \underset{(.089)}{.127}B)]\, [\underset{(.001)}{.009} + a_t],$$

$$\hat{\sigma}_a = .0041,$$

where $LS1968Q1_t$, $LS1972Q4_t$, and $LS1982Q2_t$ are level shift interventions in which $LSdate_t = 1$, $t \geq date$, 0 otherwise. Figure 1b plots compensation and the outlier/level shift component (with unconditional mean added) for the CR procedure. These level shifts correspond to the height of the Vietnam War buildup, the 1972-73 expansion, and the "Volcker disinflation" of 1982. Not only are these level shifts present in the compensation data, but they also appear to be present (at very similar dates) in other inflation series such as the Consumer Price Index and the GNP deflator (Balke and Fomby (1991)).

Second, consider the monthly series for car drivers killed and seriously injured in the United Kingdom (U.K.), in logarithms and with monthly means extracted, from January 1969 to December 1984. Harvey and Durbin (1986) examined the killed and seriously injured series to assess the effects of a U.K. seat belt law that went into effect in February 1983. These data are published in Harvey (1989), Appendix 2. Using an intervention model (the intervention variable also includes an anticipatory effect for January 1983), Harvey and Durbin found that the introduction of a seat belt law in 1983 caused a significant reduction in traffic accident casualties.

The Tsay outlier search implies the following final intervention model

$$(3)\quad Y_t = [1/(1 - \underset{(.071)}{.426}B - \underset{(.074)}{.308}B^2 - \underset{(.070)}{.145}B^3)]\,[\,-\underset{(.073)}{.285}\ \text{IO1983Feb}_t + a_t\,],$$

$$\hat{\sigma}_a = .073,$$

where $\text{IO1983Feb}_t = 1$ if t = February 1983, 0 otherwise. Figure 2a plots the killed and seriously injured series as well as the outlier component implied by the Tsay procedure. In contrast, the CR procedure yields the final intervention model

$$(4)\quad Y_t = \underset{(.014)}{.132}\ \text{LS1970Feb}_t - \underset{(.017)}{.155}\ \text{LS1974Nov}_t - \underset{(.023)}{.199}\ \text{LS1983Jan}_t$$

$$+ [1/(1 - \underset{(.073)}{.208}B - \underset{(.073)}{.167}B^2)]\ a_t,$$

$$\hat{\sigma}_a = .067.$$

Figure 2b plots the killed and seriously injured series and outlier component for the CR procedure.

While both models suggest a sharp reduction in injuries during 1983, the Tsay outlier search identifies this reduction as an innovative outlier and as such represents a temporary reduction, while the CR procedure implies a level shift intervention and a permanent reduction. The combined outlier search suggests two additional important level shifts (LS1970Feb and LS1974Nov) not identified by the Tsay outlier search. The November 1974 level shift probably reflects the effects of the jump in energy prices that occurred during 1974. McCulloch and Tsay (1991) found level shifts very similar to those found here (February 1970, January 1975, January and February 1983) using Bayesian analysis employing a Gibbs Sampler.

Thus, for both data sets, the outlier search procedure described by Tsay

(1988) fails to correctly identify some very plausible level shifts. In the first series, the Tsay outlier search misses what appear to be level shifts, while in the second series the Tsay outlier search identifies a probable level shift as an innovative outlier.

3. The Tsay iterative outlier search procedure

Consider the following outlier model described in Tsay (1988). Let

(5) $Y_t = f(t) + Z_t$,

where $Z_t = (\theta(B)/\phi(B))a_t$, and a_t is a Gaussian variate with zero mean and variance $\sigma^2{}_a$. $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_1 B^q$, and $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_1 B^p$. One can think of Z_t as the regular component of the time series Y_t.

The variable f(t) contains anomalous exogenous disturbances such as outliers and level shifts. Again following Tsay (1988), let

(6) $f(t) = \omega_o\ (\omega(B)/\delta(B))\ \xi_t^{(d)}$,

where $\xi_t^{(d)} = 1$ if $t = d$ and 0 otherwise indicates whether a disturbance occurs at time d. $\omega(B)$ and $\delta(b)$ are backshift polynomials that describe the dynamic effect the disturbance has on Y_t. When $(\omega(B)/\delta(B)) = 1$, the disturbance is an additive outlier (AO); when $(\omega(B)/\delta(B)) = (\theta(B)/\phi(B))$, the disturbance is an innovative outlier (IO); when $(\omega(B)/\delta(B)) = 1/(1-B)$, the disturbance is a level shift (LS).

To help identify the presence of outliers and level shifts, Tsay (1988) proposes several statistics. Define $y_t = (\phi(B)/\theta(B))Y_t$. Define $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \ldots = \phi(B)/\theta(B)$, and $\eta(B) = 1 - \eta_1 B - \eta_2 B^2 - \ldots = \pi(B)/(1-B)$. Tsay suggests the following test statistics for the various types of outliers:

$\lambda_{IO,t} = y_t/\sigma_a$,

$\lambda_{AO,t} = \rho^2_{A,t}(y_t - {}_{i=1}\Sigma^{T-t}\ \pi_i y_{t+i})/(\rho_{A,t}\sigma_a)$, and

$\lambda_{LS,t} = \rho^2_{L,t}(y_t - {}_{i=1}\Sigma^{T-t}\ \eta_i y_{t+i})/(\rho_{L,t}\sigma_a)$,

where $\rho^2_{A,t} = (1 + {}_{i=1}\Sigma^{T-t}\ \pi_i^2)^{-1}$, $\rho^2_{L,t} = (1 + {}_{i=1}\Sigma^{T-t}\ \eta_i^2)^{-1}$, σ^2_a is the variance of a_t, and T is the sample size. Let $\lambda_{max} = \max\ (\lambda_{IO,max}, \lambda_{AO,max}, \lambda_{LS,max})$, where $\lambda_{j,max} = \max_{1\le t\le T}\ (|\lambda_{j,t}|)$, j = IO, AO, LS. If the λ_{max} statistic exceeds a prespecified critical value, then an outlier has occurred.

Tsay suggests a sequential algorithm for identifying outliers. First, estimate an ARMA model and extract the residuals and the residual variance. Second, search for outliers in the residuals using the statistics described above. If an outlier is found, remove the effect of the outlier and recalculate the residuals and residual variance. Continue searching and adjusting until no more outliers are indicated. Reestimate the ARMA model using the adjusted series and extract the residuals. Once again, search for outliers. Stop the algorithm when no additional outliers are found.

4. ARMA specification and outlier identification

Note that the initially estimated ARMA model is the correct specification of the regular dynamics $((\theta(B)/\phi(B))a_t)$ under the null hypothesis of no outliers. If, however, outliers are found, then the initial ARMA model for the regular component will be misspecified. Unfortunately, misspecification of the initial ARMA model can lead to misidentification of outliers. In particular, series in which a level shift is present will exhibit a high degree of serial correlation regardless of the regular dynamics (Chen and Tiao (1990)). In this case, the initial ARMA model for the regular dynamics implies greater serial correlation than is in fact the case; therefore, the residuals from this model will not reflect the true nature of the outlier.

To see the dangers of this misspecification, consider the following example. Let $\phi(B) = 1 - \phi B$ and $\theta(B) = 1$.

$$Y_t = [1/(1-\phi B)]\ a_t, \quad t \leq T_1,$$

$$Y_t = [1/(1-\phi B)]\ a_t + \mu, \quad t > T_1.$$

The size of level shift is given by μ, and it occurs at $T_1 + 1$. The sample size is given by T. Subtracting the sample mean of the series yields

$$Y_t - \bar{Y} = [1/(1-\phi B)]\ (a_t - a) - \mu(T-T_1)/T, \quad t \leq T_1,$$

$$Y_t - \bar{Y} = [1/(1-\phi B)]\ (a_t - a) + \mu T_1/T, \quad t > T_1,$$

where $a = (1-\phi)\ {}_{t=1}\Sigma^T\ [1/(1-\phi B)]\ a_t/T$.

Suppose we estimate an AR(1) model for Y_t. Let $\tilde{Y}_t = Y_t - \bar{Y}$. The least squares estimate of the autoregressive parameter,

$$\hat{\phi} = {}_{t=2}\Sigma^T\ (\tilde{Y}_t\tilde{Y}_{t-1})\ /\ {}_{t=2}\Sigma^T\ \tilde{Y}_{t-1}{}^2,$$

is inconsistent when $\mu \neq 0$. Keeping the proportion of the sample before (and after) the level shift date constant (that is $(T-T_1)/T$ and T_1/T are constant) as the sample size increases,

$$(8)\quad \text{plim}\ \hat{\phi} = \phi + [(1-\phi)\mu^2(T-T_1)T_1/T^2]/\sigma^2,$$

where

$$\sigma^2 = \sigma^2{}_a/(1-\phi^2) + \mu^2(T-T_1)T_1/T^2.$$

Notice that the presence of level shifts causes the autoregressive parameter to be overestimated asymptotically. The degree of the overstatement depends on the variance of the regular component $(\sigma^2{}_a/(1-\phi^2))$ relative to the variance of the level shift component $(\mu^2(T-T_1)T_1/T^2)$. The larger the variance of the level shift component (that is, the larger the level shift), the more the autoregressive estimate is biased. Table 1 provides an indication of how serious this bias can be.

This misspecification of the ARMA model in the presence of level shifts affects the outlier statistics in several ways. First, as mentioned above, the residuals from the ARMA model will not truly reflect the outlier or level shift. Second, the filter used to generate the λ statistics will be misspecified.

In Table 2, a limited Monte Carlo experiment indicates the effects of this misspecification (see Appendix A for a detailed description of this experiment). Here, a level shift occurs halfway through the sample, and the outlier search follows the procedure described in the previous section in which an ARMA model is estimated first; the initial specification of the ARMA was set as an AR(1). Table 2 indicates the number of samples in which the algorithm identified outliers of the various types and whether the identified outliers were on, close to ($T_1+1 \pm 5$), or not close to the actual date in which the level shift occurred. For all experiments, the variance of the regular component is fixed at 1, while the size of the level shifts and the AR(1) coefficient varies. The critical value for the outlier classification was set at 3. Note that it is possible for the outlier procedure to identify several different outliers for a given sample (replication), which explains why adding up the identified outliers may exceed the number of Monte Carlo replications for some parameter settings.

The Monte Carlo experiments described in Table 2 suggest that the traditional approach of estimating an ARMA model first does not always yield satisfactory results when a level shift is present. For small level shifts ($\mu = 1$ and $\mu = 2$), the outlier algorithm might miss the level shift entirely. For larger level shifts ($\mu = 3$ and $\mu = 5$), the outlier algorithm will often misidentify the level shift as an innovative outlier. This is especially true as the persistence of the regular component increases (compare as ϕ moves from 0 to 0.8). For example, when $\mu = 5$ and $\phi = 0.8$, estimating the ARMA model first causes the outlier algorithm to correctly identify the level shift in only 6 percent of the samples while it identifies an innovative outlier instead of a level shift in 94 percent of the samples. In addition, even after the outliers have been identified and removed, the autoregressive parameter of the AR model is still overestimated. This is particularly true when the level shift is large. Additional Monte Carlo work not reported here and the analysis below suggests that the misspecification is even more likely if the level shift occurs early in the sample.

However, suppose that instead of estimating an ARMA model first, no ARMA filter is applied to the data except for extracting the mean of the series (or a time trend if one is present). That is, the initial ARMA model is specified as white noise. Table 3 contains the results of the same Monte Carlo experiment as in Table 2 except the initial ARMA model was specified as white noise. After the first iteration of the outlier search, an ARMA model (here, an AR(1)) was estimated for subsequent iterations as in the traditional approach.

The results in Table 3 imply that when starting with the white noise ARMA model, the outlier algorithm correctly identifies the level shift more times than starting with an estimated ARMA model. Not only is the white noise model more likely to capture small level shifts, but it is also much less likely to misidentify a level shift as another type of outlier. The ability to identify level shifts when starting from the white noise ARMA model is not particularly sensitive to the specification of the regular component (as long as the unconditional variance of the regular component is held constant). Furthermore, comparing Tables 2 and 3 reveals that the final estimated AR coefficient is often closer to the true parameter when the white noise model is used than when an ARMA model is used in the initial iteration; this is especially true for large level shifts. Thus, starting with the white noise ARMA model may improve the ability of the outlier identification algorithm to correctly identify level shifts and to estimate the systematic dynamics.

Unfortunately, there are some drawbacks to initializing the algorithm with the white noise model. First, as Table 3 demonstrates, starting with the white noise model tends to identify too many level shifts. The spurious identification of level shifts is more severe as the persistence in the regular component increases (see $\mu = 5$ and $\phi = 0.8$, for example). Second, the white noise model by construction is incapable of distinguishing between innovative and additive outliers. In fact, as I show below, innovative outliers tend to show up as a sequence of additive outliers.

To further understand the differences between starting the outlier algorithm with an estimated ARMA model and starting with a white noise model, consider the properties of the outlier statistics both under the null hypothesis of no outliers or level shifts and under the alternative hypothesis of a level shift. Denote the innovative and level shift outlier statistics when an AR(1) is estimated by $\lambda^{AR}_{IO,T1+1}$ and $\lambda^{AR}_{LS,T1+1}$ while the statistics for the white noise are denoted by $\lambda^{WN}_{AO,T1+1}$ and $\lambda^{WN}_{LS,T1+1}$.

Consider the example described above. Under the null hypothesis of no outliers or level shifts ($\mu = 0$), $\lambda^{AR}_{IO,T1+1}$, $\lambda^{AR}_{LS,T1+1}$ and $\lambda^{WN}_{AO,T1+1} \sim AN(0,1)$, where AN denotes asymptotic normality. On the other hand, while $E[\lambda^{WN}_{LS,T1+1}]$

= 0, in large samples

$$Var[\lambda^{WN}_{LS,T1+1}] \approx 1 + 2 \sum_{i=1}^{T-T1-1} (T-T_1-i)/(T-T_1)\ \phi^i.$$

Serial correlation in the systematic dynamics prevents the asymptotic distribution of $\lambda^{WN}_{LS,T1+1}$ from being standard normal, because the denominator of $\lambda^{WN}_{LS,T1+1}$ is too low when $\phi > 0$. In other words, the standard error of the white noise, level shift estimate understates the true standard error when $\phi > 0$. This explains why in the Monte Carlo experiment the white noise model found too many level shifts for cases in which ϕ was relatively large.

When a level shift occurs (that is, $\mu \neq 0$), the expected value of the various statistics at the shift date, T_1+1, (taking the estimated parameters of the ARMA model as given) are

$E[\lambda^{AR}_{IO,T1+1}] = [\ \mu T_1/T + \hat{\phi}\mu(T-T_1)/T\ ]/\hat{\sigma}_a$,

$E[\lambda^{AR}_{LS,T1+1}] = \hat{\rho}^2_{L,T1+1}[\mu T_1/T + \hat{\phi}\mu(T-T_1)/T + (1-\hat{\phi})^2(T-T_1-1)\mu T_1/T]/(\hat{\rho}_{L,T1+1}\ \hat{\sigma}_a)$,

$E[\lambda^{WN}_{AO,T1+1}] = (\mu T_1/T)/\hat{\sigma}$, and

$E[\lambda^{WN}_{LS,T1+1}] = (\mu T_1/T)/(\hat{\sigma}^2/(T-T_1))^{1/2}$,

where $\hat{\rho}^2_{L,T1+1} = (1 + (1-\hat{\phi})^2(T-T_1-1))$,

$\text{plim}\ \hat{\sigma}^2_a = \sigma^2_a + (\phi-\hat{\phi})^2\sigma^2_a/(1-\phi^2) + (1-\hat{\phi})^2\mu^2(T-T_1)T_1/T^2$, and

$\text{plim}\ \hat{\sigma}^2 = \sigma^2_a/(1-\phi^2) + \mu^2(T-T_1)T_1/T^2$.

Note that misspecification of the initial ARMA model has a direct effect on the expected value of the outlier statistics of the AR(1) model.

The λ statistics for various level shift sizes are plotted in Figure 3. Figure 3 shows that the white noise level shift statistic is more sensitive to small level shifts than the ARMA model statistics are. In fact, for a critical value of 3, the expected level shift statistics imply that the white noise model can detect a smaller μ than can the search in which an ARMA model is estimated first. Furthermore, when the level shift occurs early in the sample (consider the cases in which $T_1 = 25$ with $T = 100$), $E[\lambda^{AR}_{IO,T1+1}]$ is often greater than $E[\lambda^{AR}_{LS,T1+1}]$. This suggests that starting with an ARMA model will systematically misidentify level shifts as innovative outliers. As ϕ gets larger, this tendency of the ARMA outlier search to misidentify level shifts as innovative outliers increases. For the case in which $T_1 = 50$ (and higher), $E[\lambda^{AR}_{IO,T1+1}]$ approaches $E[\lambda^{AR}_{LS,T1+1}]$ as the size of the level shift increases and as ϕ increases, at least with respect to the expected values of the test statistics. Thus, the search beginning with an estimated ARMA model is not as sensitive to small level shifts as the white noise search is, and it is more likely to misclassify a level shift as an innovative outlier if the level shift occurs early in the sample and if ϕ is relatively high. These results are consistent with the Monte Carlo experiment presented in Tables 2 and 3.

If there is an innovative or additive outlier rather than a level shift, how do the various statistics perform? Tsay (1986, 1988) and Chang, Tiao, and Chen (1988) suggest that the traditional approach does fairly well in identifying and distinguishing innovative and additive outliers. While the white noise initialization is incapable of modeling an innovative outlier in the initial iteration, it is possible that an innovative outlier could be captured by a series of additive outliers. For both start-up models, there appears to be little danger of misclassifying innovative outliers as level shifts (as long as the outlier does not occur near the end of the sample). Figure 4 presents the expected outlier statistics for the case in which there is an innovative outlier at $T_1+1 = 51$ (the size of the innovative outlier is given by μ). From Figure 4, the white noise model is unlikely to classify that outlier as a level shift; it is much more likely to classify that outlier as an additive outlier.

5. Proposed methodology

It is clear that the outlier search algorithm advocated by Tsay (1988) has difficulties in the presence of level shifts. As shown above, starting the outlier search by specifying and estimating an ARMA model can lead to the misclassification of outliers in the presence of level shifts. Yet, as demonstrated by Chang, Tiao, and Chen (1988) and Tsay (1986, 1988), the above algorithm is effective in identifying and classifying additive and innovative outliers. While starting the search by specifying a white noise model leads to more effective identification of level shifts than with an ARMA model, it is incapable of distinguishing between additive and innovative outliers (at least in the first iteration of the outlier search), and it has a tendency to identify spurious level shifts.

In this section, I describe in detail a simple extension to the outlier search procedure outlined by Tsay that attempts to take advantage of the relative strengths of both outlier searches. I also consider two seperate Monte Carlo experiments to evaluate the proposed procedure. The first experiment is a conditional experiment where the number and type of level shifts and outliers is controlled. The second experiment is an unconditional experiment where the number as well as timing of the various outliers are determined randomly. Both experiments suggest that proposed procedure improves upon the basic Tsay methodology when level shifts are present.

5.1 The Combine/Reduce Procedure

The recommended procedure is as follows:

(i) Identify potential outliers and level shifts using the outlier search method suggested in Tsay (1988) but, in addition to starting from an estimated ARMA model, run the outlier search starting with the white noise model as well. These two outlier searches will provide a list of potential outlier and level shift candidates.

(ii) Combine the outliers from the two searches into a single intervention model while letting the specification of the ARMA model for the systematic dynamics of the combined intervention model encompass the ARMA models suggested by the two outlier searches. Estimate this intervention model. Eliminate the intervention dummy with the lowest t statistic if this t statistic is lower than a prespecified critical value. Below I use the same critical value for the t statistic as I use for the outlier search ($|t\ stat| < 3$). Reestimate the intervention model. Continue this stepwise reduction of the intervention model until all intervention variables have t statistics greater than the specified critical value. Once outliers have been identified in this way, reduce the ARMA specification by stepwise reduction (here I use a more strict criteria of elimination: $|t\text{-stat}| < 1$).

(iii) As a final check, determine whether the outliers or level shifts coincide with identifiable historical events that could affect the time series. For example, in Balke and Fomby (1991), we attempt to match outliers and level shifts in fifteen macroeconomic time series with identifiable economic events. In practice, the use of historical introspection can be quite useful in avoiding overparameterizations that may remain after the stepwise reduction of the combined interventions.

The proposed procedure allows the maximum flexibility in the identification of outliers by using the two outlier searches. This procedure lessens the danger of misspecification and misidentification of outliers when level shifts are present. Yet, the stepwise reduction in the intervention model increases the chances of weeding out spurious outliers and level shifts. Of course, the standard errors of the final intervention model are conditional standard errors (as is the case with the Tsay procedure) and, therefore, may understate the true uncertainty surrounding the parameter estimates.

5.2 A Conditional Monte Carlo Experiment

To further evaluate the proposed procedure, I ran two separate Monte Carlo experiments. In the first experiment, I generated twenty-four data sets, each one corresponding to a particular combination of outliers, level shifts, and serial correlation in the regular component (these combinations are listed under the Actual column of Table 4). The sample length was 100 observations. The regular component was specified as an AR(1) with $\phi = 0$, 0.4, and 0.8, and I set the unconditional variance of the regular component, $\sigma^2_a/(1-\phi^2)$, equal to 1. The date at which an outlier or level shift occurs was randomly determined subject to the constraint that it cannot happen in the first ten or last ten observations. The size of the outlier or level shift was drawn from a N(0,3) distribution; draws whose absolute value are less than 3 are rejected and another draw is taken in order to ensure that an outlier of sufficient size is present. The value of the AR parameter, the particular combination of outliers, as well as the date and size of the outliers were unknown at the time the data was analyzed; a colleague generously agreed to randomize the data sets so that I did not know the contents of the data set before the analysis.

Table 4 summarizes the final specifications obtained by the proposed CR intervention modeling approach. I compare actual outlier dates with those indicated by the final intervention model from an ARMA outlier search alone and with those indicated by the final intervention model from the CR procedure. The estimated AR(1) coefficients for the regular component are displayed as well. A "+" indicates an improvement in the overall specification from adding the information in the white noise search. A "-" indicates a case in which using the white noise outlier search adds a spurious outlier or level shift.

In eight of the twenty-four data sets, including the results of the white noise outlier search improves the identification and specification of the final intervention model. In each of these eight cases a level shift was present. As suggested in the previous section, the ARMA outlier search sometimes has difficulty when level shifts are present. In data sets 10, 14, 18, 21, and 22, the ARMA search misidentifies level shifts as innovative outliers. In data sets 6 and 8, the ARMA search misses the level shift altogether. In addition to providing a better specification of the outliers and level shifts, including the outliers from the white noise search dramatically improves the accuracy of the autoregressive parameter estimates in these data sets. Again, this points out the importance of correctly capturing the effect of the level shift when estimating the ARMA model. On the other hand, the white noise outlier search adds spurious outliers or level shifts for only two data sets; the spurious level shift in data set 17 occurs late in the sample and when the AR coefficient is relatively high (ϕ = 0.8). Of course, I am abstracting from the historical introspection phase of the CR approach. Potentially, spurious level shifts such as these can be eliminated after introspection. Nevertheless, estimates of the AR coefficient are only slightly affected by the spurious level shift. At least for this experiment, the benefits of including the results from the white noise outlier search appear to exceed the costs.

To illustrate how the proposed procedure works in practice, consider the analysis of data set 21. Here, $\phi = 0.8$ with a level shift of -3.112 in time period 28 and an innovative outlier of -4.073 in time period 31. Table 5 presents the stepwise reduction of the combined intervention model. It seems very plausible that the white noise search that picks up an AO at time period 31 and an IO at time period 32 is really picking up an IO at time period 31 (as was indicated in the ARMA outlier search). Because the standard error of the regression falls if IO31 is used instead of AO31 and IO32, the intervention model is specified with an IO at 31. The subsequent results are nearly identical if AO31 and IO32 are used instead of IO31. Note that the spurious level shifts suggested by the white noise search (LS30 and LS31) are eliminated during the stepwise reduction. Even the spurious level shift suggested by the ARMA search (LS59) is eliminated. Notice that the spurious outliers, AO96 and IO87, were not eliminated from either the ARMA search intervention model or the combination intervention model.

5.2 An Condition Monte Carlo Experiment

The first Monte Carlo experiment was limited in scope, so that we could determine how well the proposed procedure worked in certain circumstances. The second Monte Carlo experiment consists of a much more general experiment and more closely resembles the problem facing an analyst who has little a priori information about the number or nature of outliers in the sample.

Unlike the previous experiment, the number of level shifts or outliers that occurred as well as their timing and size were determined randomly. As in the conditional Monte Carlo experiment, the unconditional experiment suggests that the combine/reduce approach provides a substantial improvement in the Tsay procedure when level shifts are present.

Rather than taking the number of level shifts or outliers as fixed, in this experiment whether a level shift, innovative outlier, or additive outlier occurred in a given time period was determined by draws from independent Bernoulli distributions. The probability that a level shift or outlier occurring in a given time period was set at 0.01 percent. The only restrictions concerning the timing of outliers were that level shifts could not occur in the last period of the sample nor at the beginning of the sample while innovative outliers could not occur in the last period of the sample. There were no restrictions on the timing of additive outliers. It was even possible for level shifts or different outlier types occur in the same time period. The end period restriction is needed because all three outlier types are observationally equivalent when they occur on the last date in the sample. The size of the level shift or outlier was determined as in the previous experiment. The sample length equaled 100 observations. The regular component was an AR(1) with $\phi = 0$, 0.4, and 0.8, and $\sigma^2_a/(1-\phi^2) = 1$. Each experiment consisted of 100 replications.

Tables 6 and 7 describe the results of the second Monte Carlo experiment. Table 6 displays the degree to which the combine/reduce procedure and the Tsay procedure method correctly classified actual level shifts, innovative outliers, and additive outliers. If an actual outlier's date and type were correctly identified, then that outlier was classified as correct. If a procedure correctly identified the type of the actual outlier and was ± 5 observations from the correct date, then that outlier was classified as close. Actual outliers for which the procedure failed to identify an outlier of any type that was ± 5 observations from the actual outlier were classified as missed. Actual outliers for which the procedure identified an outlier that was close (± 5 observations) but of the wrong type were classified as misidentified.

From Table 6, it is clear that the CR procedure does a much better job at correctly classifying level shifts than the basic Tsay procedure--the CR procedure correctly identifies more level shifts than the Tsay procedure (nearly 30% more). The CR procedure almost always does just as well as the Tsay procedure at identifying additive and innovative outliers. The Tsay procedure as we suggested above has a tendency to misidentify level shifts as innovative outliers and that is reflected in the results of Table 6. Furthermore, the CR prodecure seems to provide better estimates of the regular component--the mean squared error of $\hat{\phi}$ is significantly lower for the CR procedure as compared to the Tsay procedure. This suggests that the CR procedure does a better job at characterizing both the true outlier component and the regular component than the basic Tsay procedure.

As noted before, one potential drawback of the CR procedure was that the white noise outlier search may be prone to identify spurious level shifts. In order to assess how well the CR procedure controls for the possibility of spurious level shifts, Table 7 describes properties of identified level shifts and outliers for the two procedures. The results reported in Table 7 provide a sense of how confident we can be that an identified outlier is actually present in the data. If an identified outlier correctly identifies the type and date of an actual outlier, then the identified outlier is classified as correct. If an identified outlier correctly identified the type and is ± 5 observations from an actual outlier, the identified outlier is classified as close. If an identified outlier is not ± 5 observation from an actual outlier of any type, then identified outlier is classified as spurious. The remaining identified outliers are close (± 5 observations) to an actual outlier but are of the incorrect type.

The results reported in Table 7 suggest that the CR procedure is not appreciably more prone to spurious outliers than the Tsay procedure. Only when there is substantial serially correlation ($\phi = 0.8$) does the percentage of spurious level shifts become worrisome, yet the Tsay procedure also is subject to spurious level shifts at this parameter setting. In fact, for the most part one can be more confident (for both procedures) about identified level shifts being correct than about identified innovative or additive outliers being correct. The reason is that for low values of ϕ there is not much of a difference between additive and innovative outliers and both procedures are more likely to misclassify these outliers. However, as pointed out above, the Tsay procedure often indicates that an innovative outlier has occurred when in fact a level shift is present.

6. Summary

The presence of level shifts poses problems for the iterative outlier search procedure suggested by Tsay (1988) and others. However, the simple adjustment to the Tsay algorithm suggested here can eliminate many of these difficulties. In particular, combining the outlier search based on an estimated ARMA model with the outlier search based on an ARMA(0,0) model and conducting a stepwise reduction of the resulting intervention model appears to be a relatively easy and inexpensive way to deal with the possibility of level shifts at unknown dates. Monte Carlo evidence suggests that the combine/reduce procedure is substantially better at identifying actual level shifts than the Tsay procedure yet is not any more prone to identifying spurious level shifts. How this modification compares with more sophisticated approaches to level shifts, such as the Bayesian analysis of McCulloch and Tsay (1991) is still to be examined. The second example in section 2 suggests that the two methods may yield very similar results.

References

Abraham, B., and Box, G. E. P. (1979), "Bayesian Analysis of Some Outlier Problems in Time Series," _Biometrika_, 66, 220-236.

Balke, N. S., and Fomby, T. B. (1991), "Large Shocks, Small Shocks, and Economic Fluctuations: Outliers in Macroeconomic Time Series," Working Paper No. 9101, Federal Reserve Bank of Dallas.

Box, G. E. P., and Tiao, G. C. (1975), "Intervention Analysis with Applications to Environmental and Economic Problems," _Journal of the American Statistical Association_, 70, 70-79.

Bruce, A. G., and Martin, R. D. (1989), "Leave-k-out Diagnostics for Time Series," _Journal of the Royal Statistical Society_, Series B, 51, 363-424.

Chang, I., (1982), "Outliers in Time Series," unpublished Ph D. dissertation, University of Wisconsin, Dept. of Statistics.

Chang, I., Tiao, G. C., and Chen, C. (1988), "Estimation of Time Series Parameters in the Presence of Outliers," _Technometrics_, 30, 193-204.

Chen, C., and Tiao, G. C. (1990), "Random Level-Shift Time Series Models, ARIMA Approximations, and Level-Shift Detection," _Journal of Business and Economic Statistics_, 8, 83-97.

Harvey, A. C., (1989), _Forecasting, Structural Time Series Models and the Kalman Filter_, Cambridge: Cambridge University Press.

Harvey, A. C., and Durbin, J. (1986), " The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling," _Journal of the Royal Statistical Society_, Series A, 149: 187-227.

Martin, R. D., (1980), "Robust Estimation of Autoregressive Models," in Directions in Time Series, eds. D. R. Brillenger and G. C. Tiao, Hayward, CA.: Institute of Mathematical Statistics, 228-254.

McCulloch, R. E., and Tsay, R. S. (1991), "Bayesian Analysis of Autoregressive Time Series via the Gibbs Sampler," mimeo, Graduate School of Business, University of Chicago.

Tsay, R. S. (1988), "Outliers, Level Shifts, and Variance Changes in Time Series," Journal of Forecasting, 7, 1-20.

Tsay, R. S. (1986), "Time Series Model Specification in the Presence of Outliers," Journal of the American Statistical Association, 81, 132-41.

Table 1

Probability limit of the least squares estimator of the AR(1) coefficient for various level shift sizes

	true value of ϕ		
μ	.0	.4	.8
0.0	.000	.400	.800
1.0	.200	.520	.840
2.0	.500	.700	.900
3.0	.692	.815	.938
4.0	.800	.880	.960
5.0	.862	.917	.972
6.0	.900	.940	.980
7.0	.925	.955	.985
8.0	.941	.965	.988
9.0	.953	.972	.991
10.0	.962	.977	.992

Note: True model:

$$y_t = [1/(1-\phi B)]a_t \qquad 1 \le t \le 50$$

$$= \mu + [1/(1-\phi B)]a_t \qquad 50 < t \le 100$$

where $\sigma_a^2/(1-\phi^2) = 1$.

Table 2

Results of ARMA Outlier Search in the Presence of a Level Shift

Number of samples in which outliers and level shifts are identified

	$\mu = 1$			$\mu = 2$			$\mu = 3$			$\mu = 5$		
		$\phi =$			$\phi =$			$\phi =$			$\phi =$	
Identified	.0	.4	.8	.0	.4	.8	.0	.4	.8	.0	.4	.8
LS correct date	8	0	0	41	6	16	53	32	24	47	30	6
LS close to date	12	0	0	64	6	16	58	33	25	48	31	7
LS not close to date	18	2	1	17	3	2	12	4	4	14	4	1
IO correct date	0	1	5	1	6	34	1	17	65	31	61	94
IO close to date	1	1	5	1	7	35	1	17	65	31	61	94
IO not close to date	10	10	10	15	13	16	9	9	19	8	10	17
AO correct date	1	0	2	1	4	2	4	3	3	3	2	0
AO close to date	1	0	3	2	6	7	5	8	6	10	9	1
AO not close to date	17	21	19	34	26	19	44	29	23	53	41	23
$\hat{\phi}$	0.137	0.495	0.804	0.259	0.670	0.855	0.431	0.713	0.892	0.665	0.837	0.951
	(0.020)	(0.009)	(0.004)	(0.047)	(0.008)	(0.003)	(0.076)	(0.021)	(0.003)	(0.071)	(0.013)	(0.001)
$\hat{\sigma}_a^2$	1.097	0.876	0.358	1.160	0.973	0.364	1.291	0.987	0.360	1.443	1.024	0.373

Notes: 100 replications. T = 100. A single level shift of size μ occurs at time = 51. LS = level shift, IO = innovati outlier, AO = additive outlier. Standard errors are in parentheses.

Level shifts identified as occurring at the beginning of the sample were not counted as level shift outliers beca these are essentially adjustments in the unconditional sample mean of the series.

Table 3

Results of White Noise Outlier Search in the Presence of a Level Shift

Number of samples in which outliers and level shifts are identified

	$\mu = 1$			$\mu = 2$			$\mu = 3$			$\mu = 5$		
		$\phi =$			$\phi =$			$\phi =$			$\phi =$	
Identified	0	.4	.8	.0	.4	.8	.0	.4	.8	.0	.4	.8
LS correct date	23	20	20	70	62	64	93	90	92	100	100	100
LS close to date	45	37	39	99	94	87	100	100	99	100	100	100
LS not close to date	33	52	75	36	57	87	29	58	87	31	55	86
IO correct date	0	1	5	0	3	11	0	3	19	0	5	20
IO close to date	1	2	6	1	4	12	0	4	20	1	6	21
IO not close to date	1	9	15	5	10	18	1	11	15	2	9	16
AO correct date	1	0	2	1	1	3	1	1	5	2	0	5
AO close to date	1	0	4	1	5	6	1	4	10	4	3	8
AO not close to date	27	28	29	30	33	29	25	28	10	29	30	26
$\hat{\phi}$	0.061	0.385	0.688	0.054	0.378	0.683	0.086	0.399	0.676	0.065	0.390	0.686
	(0.012)	(0.010)	(0.008)	(0.011)	(0.013)	(0.008)	(0.010)	(0.010)	(0.009)	(0.010)	(0.009)	(0.007)
$\hat{\sigma}_a^2$	1.023	0.801	0.321	1.009	0.790	0.322	1.048	0.805	0.319	1.013	0.802	0.323

Notes: 100 replications. T = 100. A single level shift of size μ occurs at time = 51. LS = level shift, IO = innovative outlier, AO = additive outlier. Standard errors are in parentheses.

Level shifts identified as occurring at the beginning of the sample were not counted as level shift outliers because these are essentially adjustments in the unconditional sample mean of the series.

Table 4

Final Intervention Model

Data Set		Actual ϕ	Actual Outliers	ARMA Outlier Search $\hat{\phi}$	ARMA Outlier Search Outliers	CR Outlier Search $\hat{\phi}$	CR Outlier Search Outliers
1		0.0	NONE	0.107	A097	0.107	A097
2		0.0	LS55	0.0	LS55	0.0	LS55
3		0.0	A076	-0.170	A076	-0.170	A076
4		0.0	A078	-0.120	I078,A079	-0.123	A078,A079
5		0.0	LS58,A059	0.0	A058,LS60	0.0	A058,LS60
6	+	0.0	LS17,A029	0.702	NONE	0.0	LS18
7		0.0	A035,A041	0.0	A035,A041	0.0	A035,A041
8	+	0.0	LS23,A036,A038	0.813	A022,A038,A082	0.0	LS23,A038
9		0.4	NONE	0.324	NONE	0.324	NONE
10	+	0.4	LS48	0.920	I048	0.293	LS48
11		0.4	I031	0.446	I031,A052	0.446	I031,A052
12		0.4	A056	0.507	I037,A044,A056	0.507	I037,I044,A056
13	+	0.4	I056,LS81	0.181	A056,LS81	0.255	I056,LS81
14	+	0.4	LS23,A061	0.890	A013,I023,A061	0.272	LS23,A061
15		0.4	I018,A052	0.419	I018,A052	0.419	I018,A052
16		0.4	A011,LS44,I078	0.246	A011,LS44,A078	0.246	A011,LS44,A078
17	-	0.8	NONE	0.753	NONE	0.688	LS91
18	+	0.8	LS59	0.964	I059	0.774	LS59
19	+	0.8	I087	0.784	I087	0.784	I087
20		0.8	A057	0.885	I049,A057	0.885	I049,A057
21	+	0.8	LS28,I031	0.942	I029,I031,LS59,I087,A096	0.747	LS28,I031,I087,A096
22	+-	0.8	LS59,A077	0.985	I059,A077	0.826	I09,LS59,A077
23		0.8	I028,A045	0.794	I028,A045,I099	0.794	I028,A045,I099
24		0.8	I034,A055,LS72	0.796	I034,I047,A055,LS72	0.796	I034,I047,A055,LS72

Notes: + indicates an improvement in model specification when using combine/reduce outlier search as opposed to the ARMA outlier search.

\- indicates a spurious outlier or level shift was introduced by the combine outlier search.

AO## indicates an additive outlier in time period ##, IO## indicates an innovative outlier in time period ##, and LS## indicates a level shift in time period ##.

Table 5

Example of Stepwise Reduction of Intervention Model

Data Set 21: ϕ = .8, σ_a = .6, LS28 = -3.112, IO31 = -4.073

Outliers and level shifts identified by ARMA search[a]:
IO28, IO31, LS59, IO87, AO96

Outliers and level shifts identified by white noise search[a]:
LS28, IO28, LS30, LS31, AO31, IO32, IO87

	Intervention Model Iteration					
Variable	(1)	(1′)	(2)	(3)	(4)	Final
Constant	0.247 (2.501)	0.239 (2.489)	0.247 (2.594)	0.236 (2.703)	0.196 (2.477)	0.179 (2.098)
ϕ	0.751 (11.939)	0.761 (13.426)	0.762 (13.486)	0.768 (14.244)	0.781 (15.272)	0.747 (13.228)
LS28	-1.927 (-1.087)	-1.885 (-1.026)	-3.061 (-7.190)	-3.104 (-7.999)	-3.290 (-9.286)	-3.674 (-11.166)
IO28	-1.189 (-0.056)	-1.233 (-0.658)	----	----	----	----
LS30	-0.670 (-1.102)	-0.673 (-1.113)	-0.400 (-0.905)	-0.450 (-1.170)	----	----
LS31	-1.018 (-0.727)	-1.044 (-0.714)	-0.172 (-0.261)	-----	----	----
IO31	----	-3.241 (-2.683)	-3.905 (-5.600)	-4.052 (-9.403)	-4.087 (-9.503)	-3.928 (-8.783)
AO31	-3.263 (-2.829)	----	----	----		----
IO32	-2.603 (-2.639)	----	----	----		----
LS59	-0.769 (-2.195)	-0.792 (-2.226)	-0.795 (-2.237)	-0.844 (-2.562)	-0.958 (-2.913)	----
IO87	-1.498 (-3.415)	(-1.507) (-3.458)	-1.508 (-3.471)	-1.516 (-3.513)	-1.535 (-3.554)	-1.602 (-3.585)
AO96	-1.064 (-3.096)	-1.065 (-3.130)	-1.065 (-3.141)	-1.066 (-3.170)	-1.068 (-3.192)	-1.081 (-3.054)
$\bar{R}^2$	0.955	0.956	0.956	0.956	0.956	0.953
$\hat{\sigma}_a$	0.429	0.427	0.426	0.424	0.425	0.442

Notes: T-statistics are in parentheses.
[a]Also indicated a level shift in initial time period. This represents on adjustment in the mean of the series rather than a level shift.

Table 6.

Classification of Actual Outliers and Estimated ARMA Parameters for the Tsay and Combine/Reduce (CR) Procedures

	$\phi = 0$		$\phi = 0.4$		$\phi = 0.8$	
	Tsay	CR	Tsay	CR	Tsay	CR
Percent of actual LS:						
correctly identified	31	69	36	77	40	66
closely identified	35	76	40	87	43	71
misidentified	35	14	45	7	57	29
missed	30	10	15	6	0	0
Percent of actual IO:						
correctly identified	-	-	52	45	89	88
closely identified	-	-	60	51	91	91
misidentified	-	-	28	39	8	8
missed	-	-	12	10	1	1
Percent of actual AO:						
correctly identified	52	61	77	79	95	94
closely identified	57	64	78	83	95	95
misidentified	24	16	15	11	5	5
missed	19	20	7	6	0	0
ARMA Model Parameters						
mean of ϕ	0.28	0.04	0.56	0.42	0.82	0.78
(st. dev.)	(0.40)	(0.22)	(0.24)	(0.14)	(0.09)	(0.08)
MSE of ϕ (x10)	2.35	0.49	0.81	0.21	0.09	0.06
(st. error) (x10)	(0.31)	(0.12)	(0.11)	(0.04)	(0.01)	(0.01)

Notes: Based on 100 replications for each value of ϕ.
Classification of actual outliers is as follows:
(i) correctly identified - if there is an identified outlier with the correct type and date.
(ii) closely identified - if there is an identified outlier with the correct type and is ± 5 observations from the correct date.
(iii) misidentified - if there is an identified outlier with the incorrect type but is ± 5 observations from the correct date.
(iv) missed - if there is no identified outlier ± 5 observations from the correct date.

Table 7.

Classification of Identified Outliers for the Tsay and Combine/Reduce (CR) Procedures

	$\phi = 0$		$\phi = 0.4$		$\phi = 0.8$	
	Tsay	CR	Tsay	CR	Tsay	CR
Percent of identified LS:						
correct	84	86	81	84	69	63
close	89	95	86	92	74	70
wrong type	3	1	4	3	6	8
spurious	8	4	10	5	20	22
Percent of identified IO:						
correct	0	0	38	51	51	54
close	0	0	50	58	56	60
wrong type	84	76	33	22	30	19
spurious	16	24	17	20	14	21
Percent of identified AO:						
correct	60	65	48	47	75	66
close	63	70	51	50	75	67
wrong type	12	13	31	34	9	11
spurious	25	17	18	16	16	22

Notes: Based on 100 replications for each value of ϕ.
Classification of identified outliers is as follows:

(i) correct – if there is an actual outlier of the same type at the same date.

(ii) closed – if there is an actual outlier of the same type and is $\pm$ 5 observations from the identified date.

(iii) wrong type – if there is an actual outlier $\pm$ 5 observations from the identified date but of a different type from that of the identified outlier.

(iv) spurious – if there are no actual outliers $\pm$ 5 observations from the identified date.

Figure 1a, Compensation per hour nonfarm business

Actual series (log differences) and outlier component from Tsay procedure

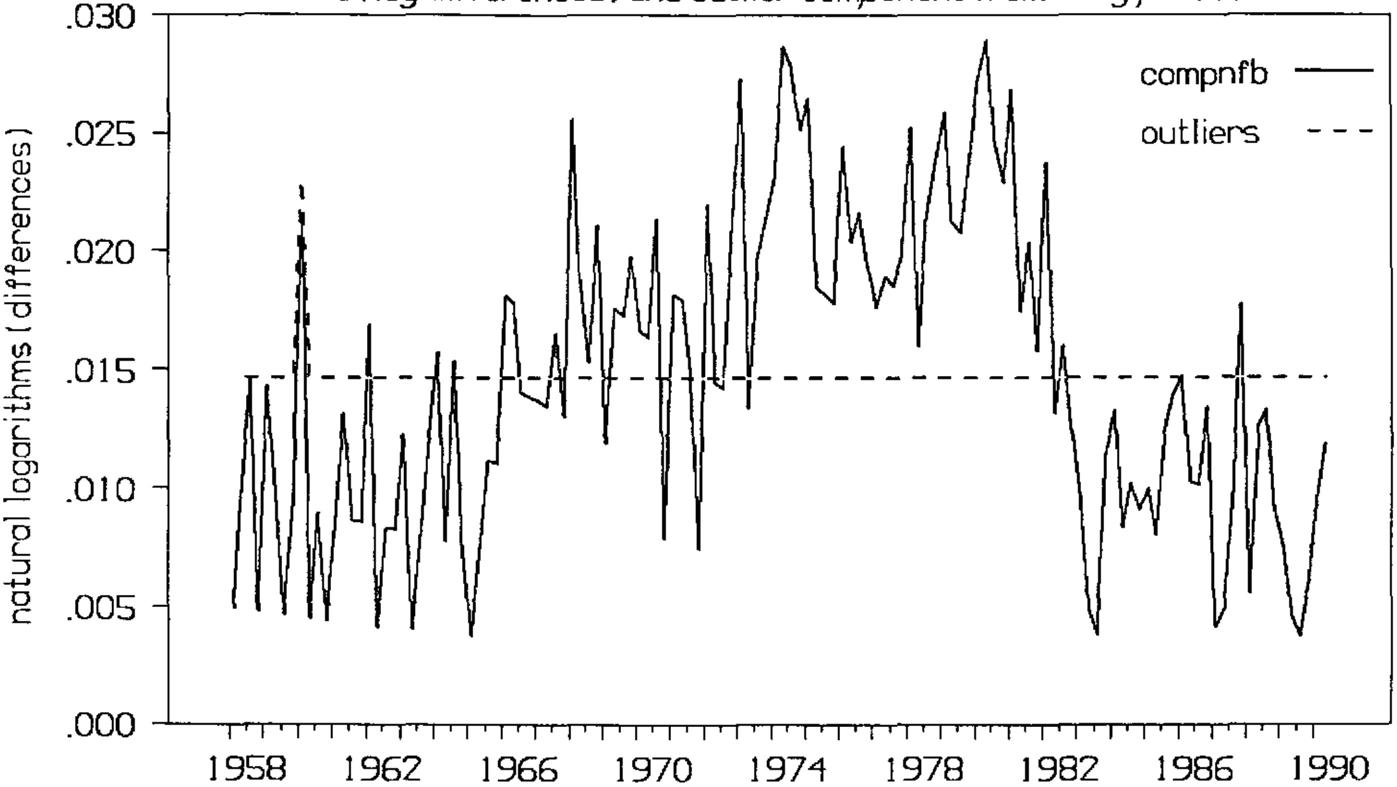


Figure 1b, Compensation per hour nonfarm business

Actual series and outlier component from combine/reduce procedure

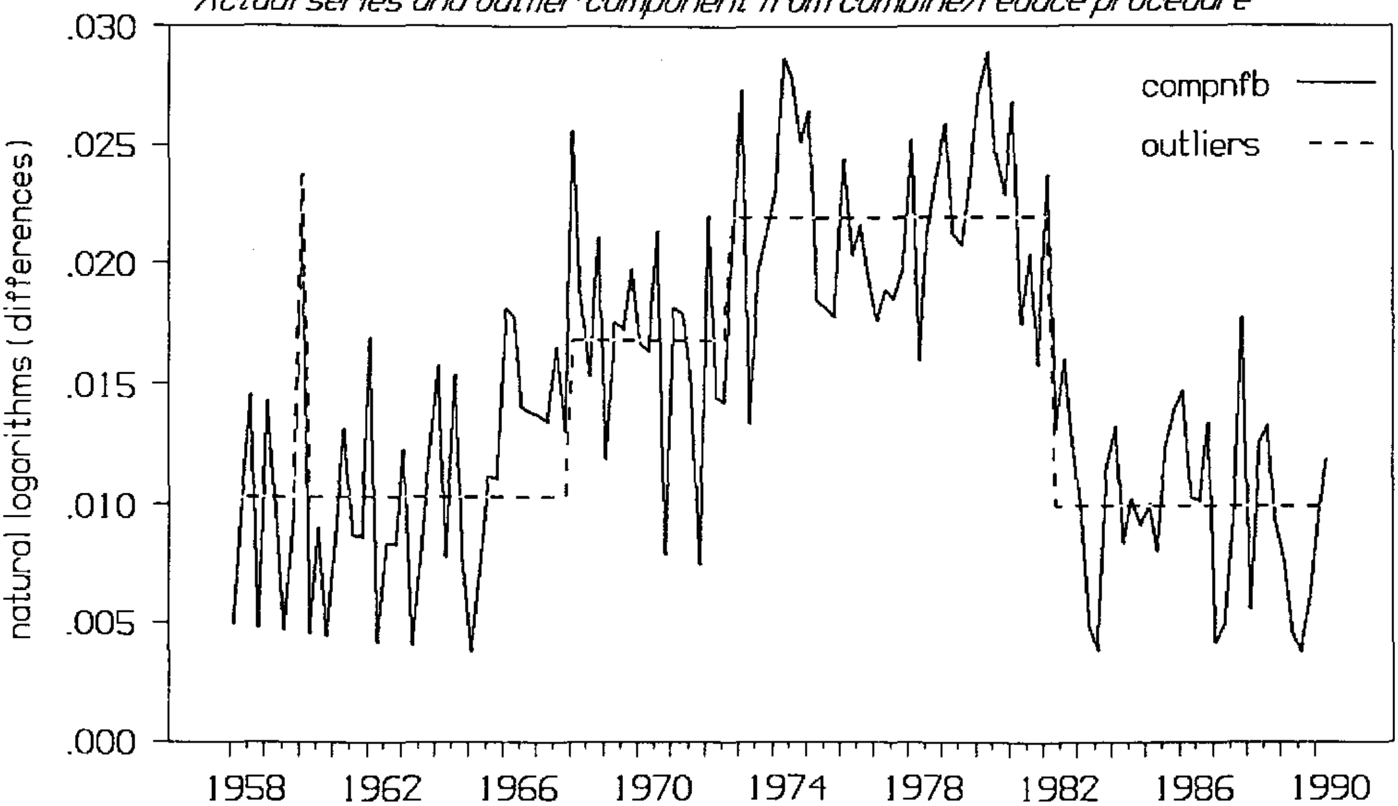


Figure 2a, Car drivers killed and seriously injured

Actual (monthly means removed) and outlier component from Tsay procedure

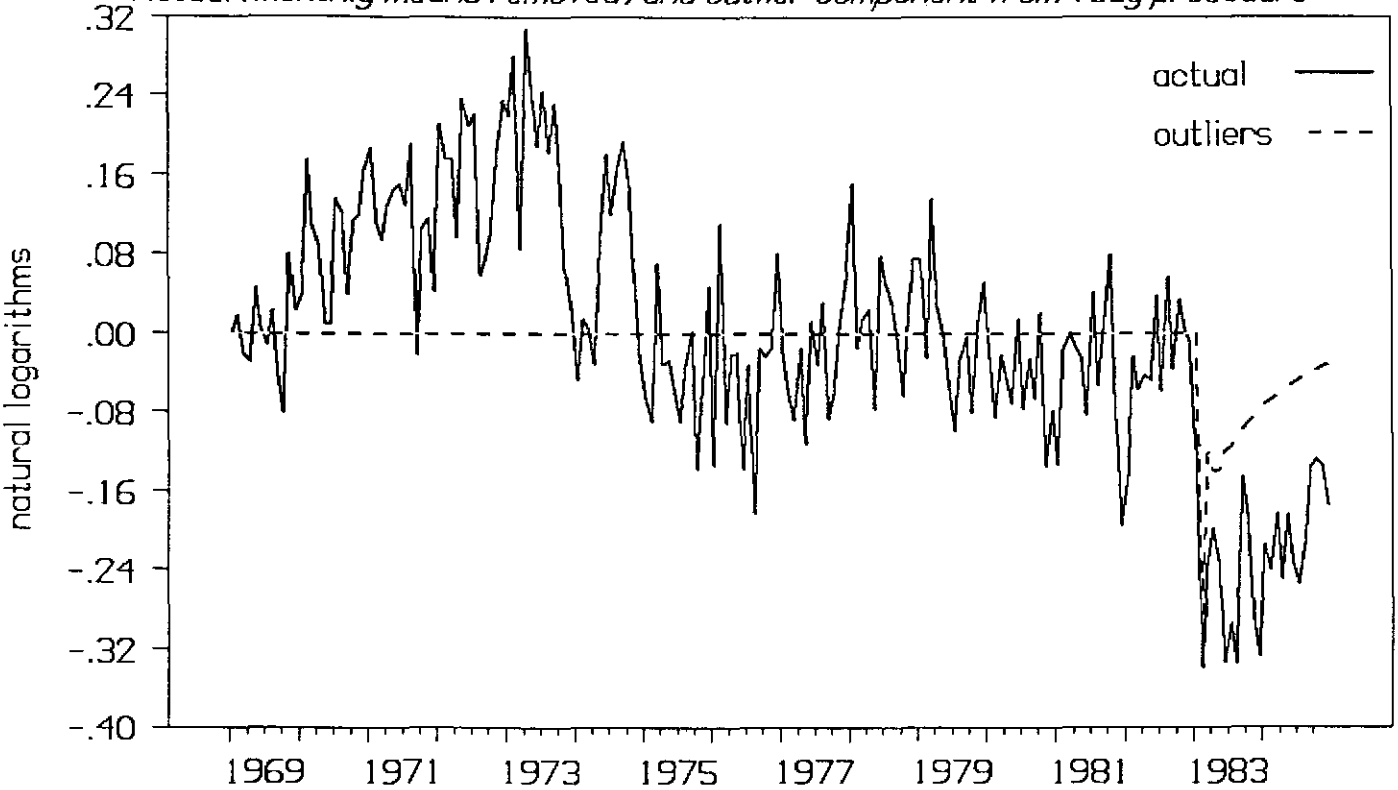


Figure 2b, Car drivers killed and seriously injured

Actual and outlier component from combine/reduce procedure

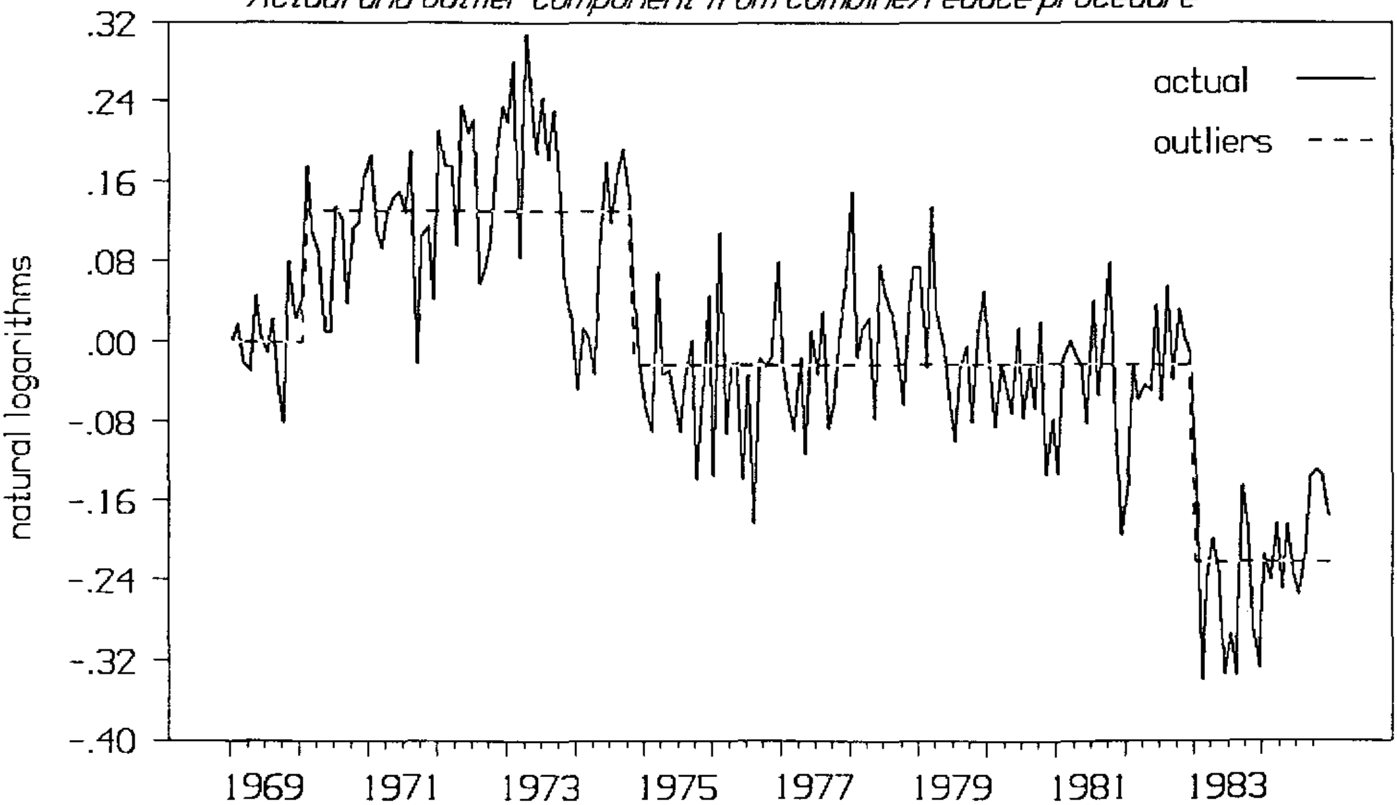


Figure 3
Expected λ statistics for case of level shift

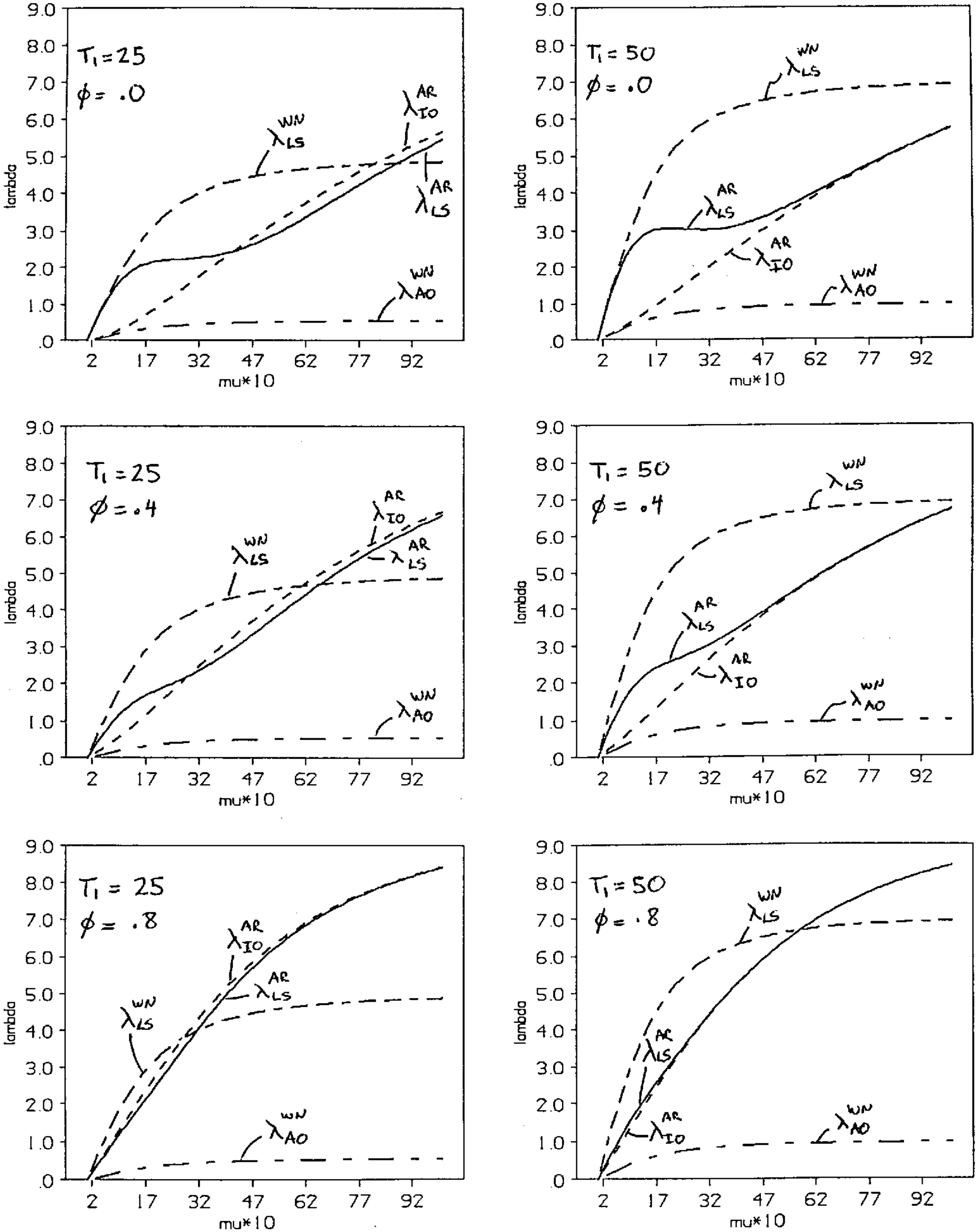
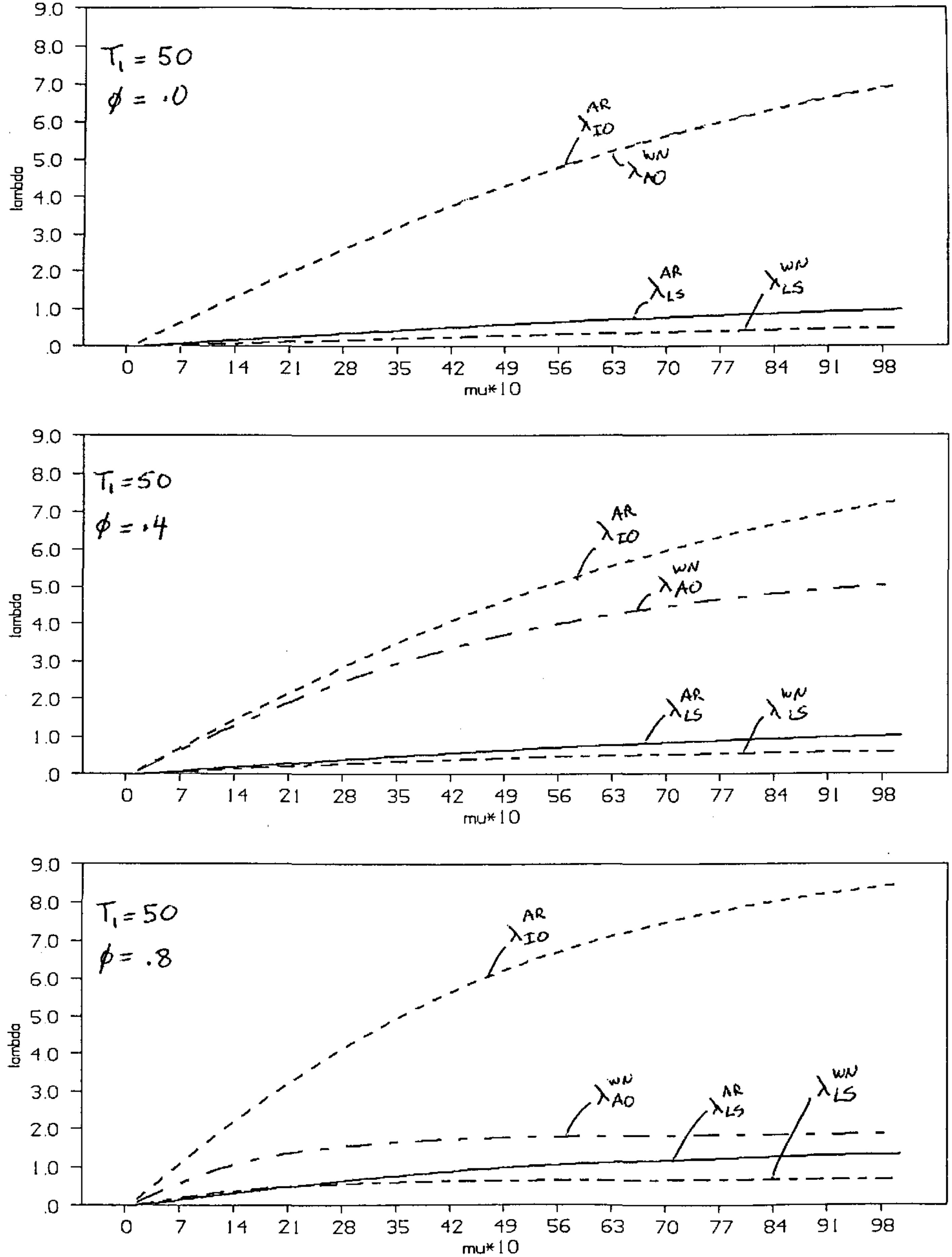


Figure 4
Expected λ statistics for case of an innovative outlier



Appendix for

Detecting Level Shifts in Time Series: Misspecification and a Proposed Solution

Nathan S. Balke

Assistant Professor of Economics
Southern Methodist University

and

Research Associate
Federal Reserve Bank of Dallas

1. Appendix A: Description of Monte-Carlo Experiment in Section 4.

I conducted the Monte Carlo experiment using RATS Version 4.2 (mainframe version) on Southern Methodist University's IBM 3081. The AR(1) model for the systematic dynamics was randomly generated starting from an initial value of zero; that is,

$$Z_t = \phi Z_{t-1} + a_t,$$

with $Z_0 = 0$. The a_t were independently drawn from a $N(0,\sigma^2{}_a)$, where $\sigma^2{}_a = 1/(1-\phi^2)$. The random draws were generated using the RAN function of the MATRIX instruction. The sample runs from t = 1 to t = 100. A level shift of size μ is placed at t = 51.

The data were processed through the Tsay (1988) outlier/level shift identification procedure. For the ARMA search, I estimated an AR(1) as the initial ARMA model to filter the data. For the white noise search, I extracted the sample means before the outlier search was undertaken. Once no more outliers were found in the first pass through the algorithm, an AR(1) model was estimated and used to filter the data for the second iteration of the outlier search. Thus, the only difference between the two procedures is the initial ARMA model.

The experiment was conducted 100 times for both the ARMA and white noise model. The same seed (for the random number generator) was used for both models to keep the intermodel sampling variation to a minimum.

2. Appendix B. Algebraic Appendix

Recall that

$$Y_t - \overline{Y} = [1/(1-\phi B)](a_t - \overline{a}) - \mu(T-T_1)/T \qquad t \le T_1$$

$$Y_t - \overline{Y} = [1/(1-\phi B)](a_t - \overline{a}) + \mu T_1)/T \qquad t > T_1$$

To simplify notation, let $\widetilde{Y}_t = Y_t - \overline{Y}$.

The estimate of the autoregressive parameter in the AR(1) model is

$$\hat{\phi} = \frac{\sum_{t=2}^{T} \widetilde{Y}_t \widetilde{Y}_{t-1}/(T-1)}{\sum_{t=2}^{T} \widetilde{Y}_{t-1}{}^2/(T-1)}.$$

Now,

$$E[\sum_{t=2}^{T} \widetilde{Y}_t \widetilde{Y}_{t-1}/(T-1)] = E[\sum_{t=2}^{T} (\sum_{i=0}^{\infty} \phi^i (a_{t-1} - \overline{a}))(\sum_{i=0}^{\infty} \phi^i (a_{t-1-i} - \overline{a}))]/(T-1)$$

$$+ [\mu^2 (T-T_1)^2 (T_1 - 1) - \mu^2 (T-T_1) T_1 + \mu^2 T_1{}^2 (T-T_1-1)]/[T^2(T-1)]$$

$$= \phi\sigma_a^2/(1-\phi^2) + \mu^2 (T-T_1) T_1/T^2 + O(1/T).$$

Recall that $(T-T_1)/T$ and T_1/T are kept constant as the sample size increases.

$$\mathrm{Var}[\sum_{t=2}^{T} \widetilde{Y}_t \widetilde{Y}_{t-1}/(T-1)] = O(1/T).$$

$$\mathrm{plim} \sum_{t=2}^{T} \widetilde{Y}_t \widetilde{Y}_{t-1}/(T-1) = \phi\sigma_a^2/(1-\phi^2) + \mu^2 (T-T_1) T_1/T^2.$$

Also, $E[\sum_{t=2}^{T} \widetilde{Y}_{t-1}{}^2/(T-1)] = E[\sum_{t=2}^{T} (\sum_{i=0}^{\infty} \phi^i (a_{t-1} - \overline{a}))^2]/(T-1)$

$$+ [\mu^2 (T-T_1)^2 (T_1 - 1) - \mu^2 T_1{}^2 (T-T_1)]/[T^2(T-1)]$$

$$= \sigma_a^2/(1-\phi^2) + \mu^2 (T-T_1) T_1/T^2 + O(1/T).$$

Also,

$$\mathrm{Var}[\ \sum_{t=2}^{T} \tilde{Y}_{t-1}{}^2/(T-1)] = O(1/T).$$

Thus,

$$\mathrm{plim}\ \sum_{t=2}^{T} \tilde{Y}_{t-1}{}^2/(T-1) = \sigma_a^2/(1-\phi^2) + \mu^2(T-T_1)T_1/T^2.$$

Therefore,

$$\mathrm{plim}\ \hat{\phi} = \frac{\phi\sigma_a^2/(1-\phi^2) + \mu^2(T-T_1)T_1/T^2}{\sigma_a^2/(1-\phi^2) + \mu^2(T-T_1)T_1/T^2}$$

$$= \phi + \frac{(1-\phi)\ \mu^2(T-T_1)T_1/T^2}{\sigma_a^2/(1-\phi^2) + \mu^2(T-T_1)T_1/T^2}\ .$$

The $\hat{\lambda}^{AR}$ statistics are calculated with the residuals from an AR(1) model. These are described by:

$$\hat{y}_t = \tilde{Y}_t - \hat{\phi}\tilde{Y}_{t-1} = a_t - \bar{a} + \sum_{i=0}^{\infty} (\phi-\hat{\phi})\phi^i(a_{t-1-i} - \bar{a}) - (1-\hat{\phi})\mu(T-T_1)/T \quad t \le T_1$$

$$\hat{y}_t = \tilde{Y}_t - \hat{\phi}\tilde{Y}_{t-1} = a_t - \bar{a} + \sum_{i=0}^{\infty} (\phi-\hat{\phi})\phi^i(a_{t-1-i} - \bar{a}) - \mu T_1/T + \hat{\phi}\mu(T-T_1)/T \quad t=T_1+1$$

$$\hat{y}_t = \tilde{Y}_t - \hat{\phi}\tilde{Y}_{t-1} = a_t - \bar{a} + \sum_{i=0}^{\infty} (\phi-\hat{\phi})\phi^i(a_{t-1-i} - \bar{a}) - (1-\hat{\phi})\mu/T_1/T \quad t > T_1$$

Note that

$$\hat{\sigma}_a^2 = \sum_{t=2}^{T} \hat{y}_a^2/(T-2)$$

$$E[\ \hat{\sigma}_a^2\ ] = E[\sum_{t=2}^{T} (a_t - \bar{a} + \sum_{i=0}^{\infty} (\phi-\hat{\phi})\phi^i(a_{t-1-i} - \bar{a})\ )^2\ ]/(T-2)$$

$$[\ (1-\hat{\phi})^2\mu^2(T-T_1)^2(T_1-1)/T^2 + \hat{\phi}\mu(T-T_1)+\mu T_1)^2/T^2 + (1-\hat{\phi})^2\mu^2 T_1{}^2/T^2\ ]/(T-2)$$

$$= \sigma_a^2 + (\phi-\hat{\phi})^2\sigma_a^2/(1-\phi^2) + (1-\hat{\phi})^2\mu^2(T-T_1)T_1/T^2 + O(1/T).$$

Also,

$$\text{Var }[\hat{\sigma}_a^2] = O(1/T).$$

Thus,

$$\text{plim } \hat{\sigma}_a^2 = \sigma_a^2 + (\phi-\hat{\phi})^2\sigma_a^2/(1-\phi^2) + (1-\hat{\phi})^2\mu^2(T-T_1)T_1/T^2.$$

Recall that

$$\hat{\lambda}^{AR}_{IO,T_1+1} = \hat{y}_{T_1+1} / \hat{\sigma}_a, \text{ and}$$

$$\hat{\lambda}^{AR}_{LS,T_1+1} = \hat{\rho}^2 \left[\hat{y}_{T_1+1} + \sum_{t=T_1+2}^{T} (1-\hat{\phi})\hat{y}_t \right] /(\hat{\rho}\hat{\sigma}_a),$$

where $\hat{\rho}^2 = (1+(1-\hat{\phi})^2(T-T_1-1))^{-1}$.

Under the alternative hypothesis ($\mu\neq0$) and taking the parameter estimates $\hat{\phi}$ and $\hat{\sigma}_a$ as given,

$$E[\hat{\lambda}^{AR}_{IO,T_1+1}] = \mu T_1/T + \hat{\phi}\mu(T-T_1)/T_1 \quad / \hat{\sigma}_a, \text{ and}$$

$$E[\hat{\lambda}^{AR}_{LS,T_1+1}] = \hat{\rho}^2 \left[\mu T_1/T + \hat{\phi}\mu(T-T_1)/T + (1-\hat{\phi})^2(T-T_1-1)\mu T_1/T \right]/(\hat{\rho}\hat{\sigma}_a).$$

For the white noise model,

$$\hat{\lambda}^{WN}_{AO,T_1+1} = \tilde{Y}_{T_1+1} / \hat{\sigma}, \text{ and}$$

$$\hat{\lambda}^{WN}_{LS,T_1+1} = \left[\sum_{t=T_1+2}^{T} \tilde{Y}_t/(T-T_1) \right]/[\hat{\sigma}^2/(T-T_1)]^{1/2},$$

where

$$\hat{\sigma}^2 = \sum_{t=1}^{T} \tilde{Y}_t^2/T, \text{ with}$$

$$\text{plim } \hat{\sigma}^2 = \sigma^2 = \sigma_a^2/(1-\phi^2) + \mu^2(T-T_1)T_1/T^2.$$

For the white noise model, under the null of $\mu=0$

$$\tilde{y}_t \sim AN(0,\sigma^2). \quad \text{Thus, under the null, } \hat{\lambda}^{WN}_{AO,T_1+1} \sim AN(0,1).$$

Consider the numerator of $\hat{\lambda}^{WN}_{LS,T_1+1}$.

$$E\left[\sum_{t=T_1+1}^{T} \tilde{Y}_t/(T-T_1) \right] = 0.$$

$$VAR\left[\sum_{t=T_1+1}^{T} \tilde{Y}_t/(T-T_1) \right] = \sum_{t=T_1+1}^{T} VAR(\tilde{Y}_t)/(T-T_1)^2$$

$$+ 2 \sum_{t=T_1+1}^{T-1} \sum_{j=t+1}^{T} COV\ (\tilde{Y}_t, \tilde{Y}_j)/(T-T_1)^2$$

$$= [\sigma_a^2/(1-\hat{\phi}) + 2 \sum_{i=1}^{T-T_1-1} (T-T_1-i)/(T-T_1)\phi^i\ \sigma_a^2/(1-\phi^2)\ ]/(T-T_1) + O(1/T^2)$$

If we take $\hat{\sigma}^2$ as given, then under the null

$$E[\ \lambda^{WN}_{LS,T_1+1}\ ] = 0, \text{ and}$$

$$VAR[\ \lambda^{WN}_{LS,T_1+1}\ ] = \frac{1 + 2 \sum_{i=1}^{T-T_1-1} (T-T_1-i)/(T-T_1)\phi^i + O(1/T)}{1 + O(1/T)}.$$

Finally, under the alternative hypothesis that $\mu \neq 0$ and taking $\hat{\sigma}^2$ as given

$$E[\ \hat{\lambda}^{WN}_{AO,T_1+1}\ ] = (\mu T_1/T)/\ \hat{\sigma}, \text{ and}$$

$$E[\ \hat{\lambda}^{WN}_{LS,T_1+1}\ ] = (\mu T_1/T)\ /\ (\hat{\sigma}^2/(T-T_1))^{1/2}.$$

RESEARCH PAPERS OF THE RESEARCH DEPARTMENT
FEDERAL RESERVE BANK OF DALLAS

Available, at no charge, from the Research Department
Federal Reserve Bank of Dallas, Station K
Dallas, Texas 75222

9001 Another Look at the Credit-Output Link (Cara S. Lown and Donald W. Hayes)

9002 Demographics and the Foreign Indebtedness of the United States (John K. Hill)

9003 Inflation, Real Interest Rates, and the Fisher Equation Since 1983 (Kenneth M. Emery)

9004 Banking Reform (Gerald P. O'Driscoll, Jr.)

9005 U.S. Oil Demand and Conservation (S.P.A. Brown and Keith R. Phillips)

9006 Are Net Discount Ratios Stationary?: The Implications for Present Value Calculations (Joseph H. Haslag, Michael Nieswiadomy and D.J. Slottje)

9007 The Aggregate Effects of Temporary Government Purchases (Mark A. Wynne)

9008 Lender of Last Resort: A Contemporary Perspective (George G. Kaufman)

9009 Does It Matter How Monetary Policy is Implemented? (Joseph H. Haslag and Scott E. Hein)

9010 The Impact of Differential Human Capital Stocks on Club Allocations (Lori L. Taylor)

9011 Is Increased Price Flexibility Stabilizing? The Role of the Permanent Income Hypothesis (Evan F. Koenig)

9012 Fisher Effects and Central Bank Independence (Kenneth M. Emery)

9013 Methanol as an Alternative Fuel (Mine K. Yucel)

9101 Large Shocks, Small Shocks, and Economic Fluctuations: Outliers in Macroeconomic Time Series (Nathan S. Balke and Thomas B. Fomby)

9102 Immigrant Links to the Home Country: Empirical Implications for U.S. and Canadian Bilateral Trade Flows (David M. Gould)

9103 Government Purchases and Real Wages (Mark Wynne)

9104 Evaluating Monetary Base Targeting Rules (R.W. Hafer, Joseph H. Haslag and Scott E. Hein)

9105 Variations in Texas School Quality (Lori L. Taylor and Beverly J. Fox)

9106 What Motivates Oil Producers?: Testing Alternative Hypotheses (Carol Dahl and Mine Yucel)

9107 Hyperinflation, and Internal Debt Repudiation in Argentina and Brazil: From Expectations Management to the "Bonex" and "Collor" Plans (John H. Welch)

9108 Learning From One Another: The U.S. and European Banking Experience (Robert T. Clair and Gerald P. O'Driscoll)

9109 Detecting Level Shifts in Time Series: Misspecification and a Proposed Solution (Nathan S. Balke)