RESEARCH REPORT SERIES
(Statistics #2006-11)

The Error in Business Cycle Estimates Obtained from Seasonally Adjusted Data

Tucker McElroy

Statistical Research Division
U.S. Census Bureau
Washington, DC 20233

Report Issued: November 14, 2006

The Error in Business Cycle Estimates Obtained from Seasonally Adjusted Data

Tucker McElroy
U.S. Census Bureau

Abstract

Business cycle estimates are typically the output of a two-stage filtering process: a statistical agency first publishes seasonally adjusted data, and from this an econometrician estimates the cycle. In many cases the two filtering procedures used are not compatible, because two different agents are acting on the data independently. This paper derives formulas to state the signal extraction Mean Squared Error (MSE) that results from such two-stage filtering, assuming an $ARIMA$ model-based framework for a finite sample of data. We also look at the "mixed" and "direct" techniques of Kaiser and Maravall (2005) for obtaining implied models for the cycle, and show that the direct approach can generate optimal estimates in the finite-sample context as well. Several two-stage filtering procedures are analyzed theoretically, and the methods are demonstrated and compared on a simulated time series.

Keywords. Filtering, nonstationary time series, seasonality, signal extraction.

1 Introduction

Many economic time series undergo several distinct patterns of stochastic variation. Apart from trading-day and holiday effects, the major stochastic components are the trend, the seasonal, and the cycle (see Peña, Tiao, Tsay (2001) and Durbin and Koopman (2001)). It is common practice to include a white noise irregular as well (some series also require a sampling error component). In this paper, we focus on an observed series Y that can be viewed as a sum of unobserved component series, consisting of cycle C, trend T, seasonal S, and irregular I. In particular we assume

$$Y_t = C_t + T_t + S_t + I_t \tag{1}$$

for times $t = 1, 2, \cdots, n$. This basic model of economic time series plays a prominent role in $ARIMA$ model-based signal extraction methods such as $SEATS$ (Gómez and Maravall, 1997) as well as structural time series models (Harvey, 1989). We will focus on the $ARIMA$ model paradigm in this paper, since this will be the most natural framework in which to develop finite-sample results.

In practice, two-stage filtering for cycle or trend estimation is the most common. This is an artifact of the way in which economic data becomes publicly available: typically, official agencies (such as the U.S. Census Bureau) publish seasonally adjusted data, and an economic practitioner interested in estimating the business cycle may not have access to the raw data. So the first stage involves government statisticians, who construct seasonal adjustment filters and apply them to the data, and the second stage involves econometricians (business-cycle analysts, professors, etc.) who take this filtered data, and in turn construct their own cycle filters and corresponding cycle estimates. Clearly, it would be more optimal to estimate the cycle directly using just one filter that appropriately considers all the dynamics of the data (assuming that a model that reflects all the dynamics could be fitted; Kaiser and Maravall (2005) discuss some of the practical difficulties with obtaining such a filter). However, so long as the original data Y is not made available, this two-stage procedure will continue to be the default situation. Thus the appropriate question is: how much additional error is induced by a two-stage procedure? A related question is: can this error be corrected somehow? We address this topic by explicitly quantifying the error and discussing two approaches for error correction.

In order to provide rigorous, exact formulas for the Mean Squared Error (MSE) of the two-stage procedures, we take the perspective of the $ARIMA$ model-based signal extraction theory. Although many practitioners prefer the ease and elegance of state space models, with the accompanying efficiency of the state space smoother, we find it preferable to use the $ARIMA$ formulation. Firstly, most state space models can be put in $ARIMA$ form. But most importantly, in the $ARIMA$ formulation explicit matrix formulas for the signal extraction linear transformation can easily be derived (McElroy, 2005), from which basic properties of MSE linear optimal signal extraction become transparent. As will become apparent from the mathematics discussed below, only these matrix formulas permit a precise analysis of two-stage procedures and their resultant error covariance matrices. In particular, such formulas *cannot* be obtained as output from the state space smoother. Although these matrix formulas involve inversions of n-dimensional matrices, the computer code is extremely simple to write and implement on a computer. In addition, the matrix approach is not limited to homogeneous nonstationarity, but can be extended to heteroscedastic components as well. For these reasons, the $ARIMA$ formulation will be the most natural and provide the easiest exposition.

Model-based ($ARIMA$ or otherwise) seasonal adjustment and cycle estimation is accomplished through a variety of methods. Model estimation is not our focus here, but for completeness we will include a brief discussion. If it is thought that (1) describes the important unobserved components in the data, we might begin by fitting a correctly-specified $ARIMA$ model to Y, and attempt to derive $ARIMA$ models for the four components via the canonical decomposition method of Hillmer and Tiao (1982), implemented for example in the popular seasonal adjustment program $SEATS$ (Gómez and Maravall, 1997). Although appealing from the standpoint that the component models "add up to" a reasonable model for Y, it sometimes occurs that the partial fraction decomposition method has no solution, i.e., the component models cannot be mathematically determined. In our experience this problem becomes worse when a mixture of auto-regressive and moving average terms are present in the model for Y, which is typically the case if we have both a cycle present in addition to trend and seasonality. This practical difficulty is part of the motivation for the hybrid approach discussed in Kaiser and Maravall (2005); there an *ad hoc* filter that has a model-based interpretation (e.g., the Hodrick-Prescott (HP) filter (Hodrick and Prescott, 1997)) is used to implicitly define a model for the cycle.

Alternatively, one may be interested in setting up models for the unobserved components and directly estimating them from the data through their aggregate Y (see Harvey, 1989). This can be done by postulating $ARIMA$ models for the components, though the implementation by the program $STAMP$ utilizes so-called structural models, which are essentially constrained-coefficient $ARIMA$ models. This latter approach is attractive, in that it incorporates our prior beliefs about the existence of the unobserved components directly into the model estimation, and the mathematical difficulties of the decomposition approach is avoided. However, in some cases we have found it is difficult to obtain good parameter estimates unless the samples are quite large (e.g., 30 years of monthly data). This is especially true of the cycle parameters (see Harvey and Jaeger (1993) and Gómez (2001)), since in the pseudo-spectrum the cycle is dominated by the trend's low frequency spectral peak. We have also explored a hybrid between the structural and decomposition approaches that seems to do reasonably well, and which will be further discussed in Section 4.

Whichever approach is adopted, it is true that the pseudo-spectrum for Y is given as the sum of the pseudo-spectrums of the four component processes. We can write down the $ARIMA$ models for all the components, and the MSE linear optimal signal extraction matrix (where the signal is any combination of the four components, and the noise consists of the remaining components) can be easily calculated. As alluded to above, this matrix applied to the data Y yields the linear signal extraction estimate that is MSE optimal under Assumption A on the components (see Bell, 1984), and is identical to the output of the state space smoother. In this way, $ARIMA$ model-based signal extraction from a finite-sample can always be reduced to the operation of a filter matrix on the data, whether we take a decomposition or a structural approach to component model estimation. Hence, our questions about MSE's for two-stage procedures will be answered by reading off the diagonal entries of the appropriate error covariance matrices. This provides the motivation and the framework for the mathematical developments below.

In this paper, we first set out certain mathematical preliminaries in Section 2. We set up the basic signal extraction notation, and extend certain results of McElroy (2005) and McElroy and Sutcliffe (2006). Our new results are relevant, as they demonstrate that certain idealized two-stage filtering procedures can produce optimal estimates. In Section 3 we discuss our main questions in more detail, applying the mathematical techniques of Section 2. We demonstrate that the common approach to two-stage filtering results in a heightened error that is precisely quantifiable in terms of the models for the components. If instead a cycle filter is used that essentially implies a cycle model, as in Kaiser and Maravall (2005), then the resultant two-stage filter will be optimal under the assumption that this implied cycle model is correct; this result confirms and expands the Kaiser and Maravall (2005) result to the finite sample context. Finally, and perhaps of greatest interest, we illustrate a fairly simple technique for producing an optimal filter out of two sub-optimal filters; our method is essentially the iterative procedure of McElroy and Sutcliffe (2006) extended to the case where all the components may be nonstationary. Next in Section 4 we illustrate the ideas of Section 3 on simulated data, which gives the sense of how our results can be utilized. Proofs are contained in a separate Appendix.

2 Basic results on two-stage signal extraction

We will first give a short preliminary discussion of the two-stage filtering procedures that we consider in this paper. Assuming the underlying component decomposition (1), our *basic* two-stage process involves an initial seasonal adjustment, followed by a cycle estimation phase. The seasonal adjustment basically involves viewing the seasonal S as noise, and the cycle-trend-irregular $C+T+I$ as the signal. We note that in practice, the so-called "trend-cycle" $C + T$ may be treated (and modeled) as a single component. For example, the original approach of Hillmer and Tiao (1982) specified seasonal, trend, and irregular components; the cycle is either ignored completely, or may be viewed as rolled up in the trend component. In the second phase of the basic two-stage, a cycle filter is applied, where the noise process is now trend-irregular $T + I$ – this assumes that only $C+T+I$ is left after the seasonal adjustment. In practice, of course only nonstationary seasonality is removed by the seasonal adjustment filter, as is clear from the matrix formulation of McElroy (2005).

In contrast to this, the so-called $SEATS$ two-stage approach (which corresponds to the "mixed" approach of Kaiser and Maravall, 2005) involves first producing a trend-cycle estimate $C+T$; again, the separate dynamics of cycle and trend are not modeled at this stage. Only the model of their sum, the trend-cycle, is important. Then in the second stage we apply a signal extraction filter, where the noise is the trend T and the signal is the cycle C. Even if we use an *ad hoc* filter to do this, e.g., the HP filter, we can retrospectively define models for the cycle and trend such that the given cycle filter is an MSE optimal filter. This is the idea presented in Kaiser and Maravall (2005) (though not in a finite-sample context), and is currently implemented in $SEATS$. This represents a resolution of the error implied by a two-stage procedure, to the extent that it is believed that the implied model for the cycle correctly models the true underlying cycle dynamics.

Finally, we will also consider the *truncated* two-stage process, where the first step of seasonal adjustment is performed in complete ignorance of the presence of a cycle – we may conceive of a cycle being present, but we do not attempt to model it in any way during the seasonal adjustment phase, and it is not considered to be part of the trend component. So we perform signal extraction, where the noise is the seasonal S and the signal is the trend-irregular $T+I$. In the second phase, we extract the cycle C as the signal from trend-irregular $T+I$ noise, as in the basic two-stage. The term "truncated" refers to the absence of cycle noise being accounted for in the first stage of seasonal adjustment. For many applications, this may be considered a more realistic representation of what happens in practice, versus the basic two-stage.

Now we discuss some of the basic mathematical notation needed for our signal extraction results. Let us write the decomposition (1) in vector form

$$Y = C + S + T + I. \tag{2}$$

So $Y = \{Y_1, Y_2, \cdots, Y_n\}'$, and similarly for the components. For most applications, the seasonal and trend components are nonstationary, with associated differencing polynomials $\delta^S(z)$ and $\delta^T(z)$ respectively, whereas the cycle and irregular are stationary. However, Theorems 1 and 2 of this section are proved generally, in that any of the components may be stationary or nonstationary; the one assumption is that their differencing polynomials are relatively prime, i.e., share no common zeroes. The differencing polynomial for Y_t is $\delta(z) = \delta^S(z)\delta^T(z)$, and $W_t = \delta(B)Y_t$ is stationary, where B is the backshift operator. These polynomials δ, δ^S, and δ^T all have their zeroes located on the unit circle of the complex plane, and their leading coefficient is one by convention. We let the differenced components be defined as

$$U_t^S = \delta^S(B)S_t \qquad U_t^T = \delta^T(B)T_t,$$

which are mean zero weakly stationary time series. Let d be the order of δ, and let d_S and d_T be the orders of δ^S and δ^T respectively. Clearly $d = d^S + d^T$. For example, the seasonal operator for monthly data might be $\delta^S(z) = 1 + z + z^2 + \cdots + z^{11}$, and $\delta^T(z) = (1-z)^2$ is appropriate for a second-order trend.

We assume Assumption A of Bell (1984) holds on the component decomposition, appropriately generalized to four components. Assumption A states that the initial values, i.e., the variables $Y_* = (Y_1, Y_2, \cdots, Y_d)$, are independent of the differenced component series $\{U_t^S\}$, $\{U_t^T\}$, C_t, and I_t. Bell (1984), Bell and Hillmer (1988), McElroy (2005), and McElroy and Sutcliffe (2006) all discuss the implications of this assumption. Note that mean square optimal signal extraction filters derived under Assumption A agree exactly with the filters implicitly used by a properly initialized state space smoother, see Bell and Hillmer (1991, 1992). We will also assume that the differenced components $\{U_t^S\}$, $\{U_t^T\}$, C_t, and I_t are uncorrelated with one another.

Next, we formulate these notations in matrix form. Let Δ be a $n-d \times n$ matrix with entries given by $\Delta_{ij} = \delta_{i-j+d}$ (the convention being that $\delta_k = 0$ if $k < 0$ or $k > d$). The matrices Δ_S and Δ_T have entries given by the coefficients of $\delta^S(z)$ and $\delta^T(z)$, but are $n - d_S \times n$ and $n - d_T \times n$ dimensional respectively. This means that each row of these matrices consists of the coefficients of the corresponding differencing polynomial, horizontally shifted in an appropriate fashion. For example, Δ is given by

$$\Delta = \begin{bmatrix} \delta_d & \cdots & \delta_1 & 1 & 0 & 0 & \cdots \\ 0 & \delta_d & \cdots & \delta_1 & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \delta_d & \cdots & \delta_1 & 1 \end{bmatrix}$$

Hence we can write

$$W = \Delta Y, \qquad U^S = \Delta_S S, \qquad U^T = \Delta_T T.$$

Now to express the important equation

$$W_t = \delta^S(B) U_t^T + \delta^T(B) U_t^S + \delta(B) C_t + \delta(B) I_t$$

in matrix form we need to define further differencing matrices $\underline{\Delta}_T$ and $\underline{\Delta}_S$, which have row entries given by the coefficients of $\delta^T(z)$ and $\delta^S(z)$ respectively, but are $n-d \times n-d_S$ and $n-d \times n-d_T$ dimensional. Then the equation that relates the components is

$$W = \underline{\Delta}_T U^S + \underline{\Delta}_S U^T + \Delta C + \Delta I, \tag{3}$$

which uses the relationship

$$\Delta = \underline{\Delta}_T \Delta_S = \underline{\Delta}_S \Delta_T \tag{4}$$

proved in McElroy and Sutcliffe (2006). We will be principally interested in making estimates of C. For each $1 \leq t \leq n$, the minimum mean squared error signal extraction estimate is $\hat{C}_t = \mathbb{E}[C_t | Y]$.

This can be expressed as a certain linear function of the data vector Y when the data are Gaussian. This estimate is also the minimum mean squared error *linear* estimate when the data is non-Gaussian. For the remainder of the paper, we do not assume Gaussianity, and by optimality we always refer to the minimum mean squared error *linear* estimate. Writing $\hat{C} = (\hat{C}_1, \hat{C}_2, \cdots, \hat{C}_n)^{'}$, the coefficients of these linear functions form the rows of a matrix F:

$$\hat{C} = FY$$

The various rows of F differ, since only a finite number of Y_t's are available for filtering. The last row of F, for example, corresponds to the concurrent filter. Letting Σ_X denote the covariance matrix of any random vector X, the formula for F is

$$F = \Sigma_C \Delta^{'} \Sigma_W^{-1} \Delta, \tag{5}$$

see, e.g. McElroy (2005). From this formula, it is clear that we only need a model for C and a model for the differenced data W; it is not necessary to compute models for S, T, and I, but only for their aggregate.

Now since we will be analyzing two-stage filtering procedures, we require a more flexible notation, which generalizes ideas in McElroy and Sutcliffe (2006). Since we wish to state our results more generally, we will consider a scenario with three components $\{\alpha_t\}$, $\{\beta_t\}$, and $\{\gamma_t\}$ that are possibly nonstationary, with relatively prime differencing polynomials δ^α, δ^β, and δ^γ; let $\{U_t^\alpha\}$, $\{U_t^\beta\}$, and $\{U_t^\gamma\}$ be the differenced components. Denote by F_{XZ}^X the filter matrix such that $F_{XZ}^X(X+Z)$ is the optimal linear estimate of X from observed $X+Z$, given that Assumption A holds on the decomposition $X_t + Z_t$, where X and Z are random vectors of unobserved components. In this fashion we define $F_{\alpha\beta}^\alpha$, $F_{\alpha\beta\gamma}^\alpha$, $F_{\alpha\beta\gamma}^{\alpha\beta}$, etc. The general formulas for these filter matrices can be deduced from Theorem 1 of McElroy (2005). The following theorem shows how the filter matrices are related:

Theorem 1 *Let Assumption A hold on the decomposition $\alpha+\beta+\gamma$, and assume that $\{U_t^\alpha\}$, $\{U_t^\beta\}$, and $\{U_t^\gamma\}$ are independent and purely nondeterministic. Then*

$$F_{\alpha\beta\gamma}^\alpha = F_{\alpha\beta}^\alpha F_{\alpha\beta\gamma}^{\alpha\beta}.$$

Letting α, β, and γ denote trend, cycle, seasonal, irregular, or combinations thereof as appropriate, we have the following corollary.

Corollary 1 *Under the same assumptions as Theorem 1, we have the following relations:*

$$F_{CSTI}^C = F_{CT}^C F_{CSTI}^{CT} = F_{CTI}^C F_{CSTI}^{CTI} = F_{CSI}^C F_{CSTI}^{CSI}$$

Similar relations hold for all of the other components.

The first expression in Corollary 1 is the optimal cycle filter in a one-stage process, while the other expressions all represent two-stage processes, that nevertheless result in the optimal filter. The second expression corresponds to the $SEATS$ approach mentioned in the beginning of this section, where trend-cycle estimation in stage one is followed by cycle estimation from a trend noise. The third expression corresponds to the *basic* two-stage; for the *truncated* two-stage we need a different result. The following theorem provides an explicit formula for the filter $F^{\alpha}_{\alpha\beta\gamma}$ in terms of "two-component" filters $F^{\alpha}_{\alpha\beta}$, $F^{\beta}_{\beta\gamma}$, etc. It is a generalization of Theorem 1 of McElroy and Sutcliffe (2006) to the case where all 3 unobserved components are nonstationary. We let 1 be a shorthand for the $n \times n$ identity matrix.

Theorem 2 *Under the same assumptions as Theorem 1,* $1 - F^{\alpha}_{\alpha\gamma} F^{\beta}_{\beta\gamma}$ *and* $1 - F^{\alpha}_{\alpha\beta} F^{\gamma}_{\beta\gamma}$ *are invertible, and*

$$F^{\alpha}_{\alpha\beta\gamma} = \left(1 - F^{\alpha}_{\alpha\gamma} F^{\beta}_{\beta\gamma}\right)^{-1} F^{\alpha}_{\alpha\gamma}(1 - F^{\beta}_{\beta\gamma}) = \left(1 - F^{\alpha}_{\alpha\beta} F^{\gamma}_{\beta\gamma}\right)^{-1} F^{\alpha}_{\alpha\beta}(1 - F^{\gamma}_{\beta\gamma}).$$

The invertibility claimed in Theorem 2 depends upon the assumption that the differencing operators are relatively prime. This result can be applied to cycle estimation through the following corollary:

Corollary 2 *Under the same assumptions as Theorem 1, we have the following expressions:*

$$F^{C}_{CSTI} = \left(1 - F^{C}_{CTI} F^{S}_{STI}\right)^{-1} F^{C}_{CTI}(1 - F^{S}_{STI}) = \left(1 - F^{C}_{CT} F^{SI}_{STI}\right)^{-1} F^{C}_{CT}(1 - F^{SI}_{STI})$$

Similar expressions hold for all of the other components.

In the next section we will see various applications of Corollaries 1 and 2. We note that the *truncated* two-stage corresponds to the filters $F^{C}_{CTI}(1 - F^{S}_{STI})$, which are found on the right hand side of the second expression in Corollary 2; hence, multiplication by the inverse of the matrix $(1 - F^{C}_{CTI} F^{S}_{STI})$ provides a "correction factor," yielding the optimal filter matrix F^{C}_{CSTI}. The next section develops exact error formulas for several cycle estimation procedures, taking into account the fact that model parameter estimates in the first and second stages of a two-step procedure may well differ from one another, and also differ from the true Data Generation Process (DGP).

3 Errors in Cycle Estimation

In this section, we consider the various cycle estimation methods discussed in the introduction, and provide formulas for the MSE in each case. In general, the MSEs are the diagonal entries of the covariance matrix of $FY - \xi$, where ξ is the signal of interest. For convenience of exposition, we will refer to this covariance matrix as the MSE matrix. Now if $Y = \xi + \eta$ is a signal plus noise decomposition, then

$$FY - \xi = (F - 1)\xi + F\eta$$

gives a breakdown of the error process. It is desirable that this expression be mean zero and a function of only the differenced components U^{ξ} and U^{η}. A process that is a linear transformation of stationary random vectors, or sums thereof, will be called "broadly stationary." We require that the error process $FY - \xi$ be broadly stationary, which is the finite-sample analogue of the bi-infinite sample requirement that the error process be stationary. Without this requirement, the MSE will depend on the variances of nonstationary processes, which may grow in an unbounded fashion with increasing sample size. A sufficient condition for broad stationarity is that

$$(F - 1) = P\Delta_{\xi}, \qquad F = Q\Delta_{\eta} \tag{6}$$

for some (non-square) matrices P and Q. Here Δ_{ξ} and Δ_{η} are the differencing matrices associated with the possibly nonstationary ξ and η. All of the model-based filters have this property, as is easily verified from McElroy (2005). However, for a two-stage procedure resulting in a filter F, broad stationarity of the error process need not hold. This will be explored further below.

We employ the following notation. The true covariance matrix of a random vector X is denoted by Σ_X, while a proxy for this quantity – perhaps based upon some model for $\{X_t\}$ – is denoted by $\dot{\Sigma}_X$. If there is a second proxy, it is denoted by $\ddot{\Sigma}_X$. Below, we will consider proxies $\dot{\Sigma}_X$ for Σ_X used in the first stage of the filtering, obtained perhaps by fitting a model to the data; in the second stage we have different proxies $\ddot{\Sigma}_X$ that arise from different models. Note that even though the parameter estimates that enter into these proxies are data-dependent, and hence random, we treat them as deterministic in the formulas developed below. Thus, our formulas do *not* take parameter uncertainty into account. Finally, $\dot{F}$ and $\ddot{F}$ will denote filter matrices utilized in the first and second stages, respectively.

3.1 One-stage as a Benchmark

Now we are only interested in the optimal, one-stage procedures as a sort of benchmark for the two-stage procedures. We can construct the filter $\dot{F}^{C}_{CSTI}$ by plugging in proxy covariance matrices for the true matrices. With $U^{STI} = \Delta(S + T + I)$, we have

$$\dot{F}^{C}_{CSTI} = \left(\dot{\Sigma}_C^{-1} + \Delta'\dot{\Sigma}^{-1}_{U^{STI}}\Delta\right)^{-1}\Delta'\dot{\Sigma}^{-1}_{U^{STI}}\Delta = \dot{\Sigma}_C^{-1}\Delta'\dot{\Sigma}_W^{-1}\Delta,$$

which follows from the general formulas in McElroy (2005). As noted in that paper, $\dot{\Sigma}_{U^{STI}}$ can be constructed from the component covariance matrices $\dot{\Sigma}_{U^T}$, $\dot{\Sigma}_{U^S}$, and $\dot{\Sigma}_I$ via the equation

$$\dot{\Sigma}_{U^{STI}} = \underline{\Delta}_S\dot{\Sigma}_{U^T}\underline{\Delta}_S' + \underline{\Delta}_T\dot{\Sigma}_{U^S}\underline{\Delta}_T' + \Delta\dot{\Sigma}_I\Delta'.$$

Then the error process is given by

$$\dot{F}^{C}_{CSTI}Y - C = -\left(\dot{\Sigma}_C^{-1} + \Delta'\dot{\Sigma}^{-1}_{U^{STI}}\Delta\right)^{-1}\dot{\Sigma}_C^{-1}C + \left(\dot{\Sigma}_C^{-1} + \Delta'\dot{\Sigma}^{-1}_{U^{STI}}\Delta\right)^{-1}\Delta'\dot{\Sigma}^{-1}_{U^{STI}}U^{STI},$$

which has covariance matrix

$$\left(\dot{\Sigma}_C^{-1} + \Delta'\dot{\Sigma}_{USTI}^{-1}\Delta\right)^{-1} \left(\dot{\Sigma}_C^{-1}\Sigma_C\dot{\Sigma}_C^{-1} + \Delta'\dot{\Sigma}_{USTI}^{-1}\Sigma_{USTI}\dot{\Sigma}_{USTI}^{-1}\Delta\right) \left(\dot{\Sigma}_C^{-1} + \Delta'\dot{\Sigma}_{USTI}^{-1}\Delta\right)^{-1}. \quad (7)$$

Typically, we just use the proxies $\Sigma_C = \dot{\Sigma}_C$ and $\Sigma_{USTI} = \dot{\Sigma}_{USTI}$, which results in a reduction of (7) to the familiar

$$\left(\dot{\Sigma}_C^{-1} + \Delta'\dot{\Sigma}_{USTI}^{-1}\Delta\right)^{-1}.$$

3.2 Basic two-stage

Now we turn to the analysis of the two-stage procedures delineated in Section 2. First, we consider the *basic* two-stage procedure, formulated generally. For generic nonstationary components α, β, and γ, the basic two-stage filter is

$$\ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\alpha\beta}_{\alpha\beta\gamma},$$

where $\ddot{F}^{\alpha}_{\alpha\beta}$ is a filter constructed from component models estimated from $\dot{F}^{\alpha\beta}_{\alpha\beta\gamma}Y$. The idea is to first estimate models for components α, β, γ, and apply $\dot{F}^{\alpha\beta}_{\alpha\beta\gamma}$. This produces a process with nonstationary α and β components, but only a stationary γ, since $\dot{F}^{\alpha\beta}_{\alpha\beta\gamma}$ accomplishes δ^{γ} differencing. So the resulting process can conceivably be modelled as having nonstationary components α and β; the residual $\dot{F}^{\alpha\beta}_{\alpha\beta\gamma}\gamma$ content from the γ component is broadly stationary, and gets lumped in with the nonstationary $\dot{F}^{\alpha\beta}_{\alpha\beta\gamma}\alpha$ and $\dot{F}^{\alpha\beta}_{\alpha\beta\gamma}\beta$. So $\dot{F}^{\alpha\beta}_{\alpha\beta\gamma}Y$ can be given (and is given, in practice) a two-component model, and the filter for α is called $\ddot{F}^{\alpha}_{\alpha\beta}$, which is estimated from models of α and β in $\dot{F}^{\alpha\beta}_{\alpha\beta\gamma}Y$. Thus our final filter is $\ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\alpha\beta}_{\alpha\beta\gamma}$ applied to Y.

For the following theorem, we use the notations employed in the proof of Theorem 2; also let $\overline{\Delta}_{\alpha}$ be an $n - (d_{\alpha} + d_{\beta} + d_{\gamma}) \times n - (d_{\beta} + d_{\gamma})$ dimensional matrix that accomplishes δ^{α} differencing.

Theorem 3 *Under the same assumptions as Theorem 1,*

$$\begin{aligned} \ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\alpha\beta}_{\alpha\beta\gamma} &= \dot{F}^{\alpha}_{\alpha\beta\gamma} + M^{-1}P\dot{\Sigma}_W^{-1}\Delta \\ M &= \Delta_{\alpha}'\ddot{\Sigma}_{U^{\alpha}}^{-1}\Delta_{\alpha} + \Delta_{\beta}'\ddot{\Sigma}_{U^{\beta}}^{-1}\Delta_{\beta} \\ P &= \Delta_{\beta}'\ddot{\Sigma}_{U^{\beta}}^{-1}\dot{\Sigma}_{U^{\beta}}\underline{\Delta}_{\alpha\gamma}' - \Delta_{\alpha}'\ddot{\Sigma}_{U^{\alpha}}^{-1}\dot{\Sigma}_{U^{\alpha}}\underline{\Delta}_{\beta\gamma}'. \end{aligned} \quad (8)$$

Letting $Q = \Delta_{\alpha}'\dot{\Sigma}_{U^{\alpha}}^{-1}\Delta_{\alpha} + \Delta_{\beta\gamma}'\dot{\Sigma}_{U^{\beta\gamma}}^{-1}\Delta_{\beta\gamma}$ *and* $R = \overline{\Delta}_{\alpha}\Sigma_{U^{\beta\gamma}}\dot{\Sigma}_{U^{\beta\gamma}}^{-1}\Delta_{\beta\gamma} - \underline{\Delta}_{\beta\gamma}\Sigma_{U^{\alpha}}\dot{\Sigma}_{U^{\alpha}}^{-1}\Delta_{\alpha}$, *the MSE matrix is given by*

$$\begin{aligned} &M^{-1}P\dot{\Sigma}_W^{-1}\Sigma_W\dot{\Sigma}_W^{-1}P'M^{-1} \\ &+ M^{-1}P\dot{\Sigma}_W^{-1}RQ^{-1} + Q^{-1}R'\dot{\Sigma}_W^{-1}P'M^{-1} \\ &+ Q^{-1}\Delta_{\alpha}'\dot{\Sigma}_{U^{\alpha}}^{-1}\Sigma_{U^{\alpha}}\dot{\Sigma}_{U^{\alpha}}^{-1}\Delta_{\alpha}Q^{-1} + Q^{-1}\Delta_{\beta\gamma}'\dot{\Sigma}_{U^{\beta\gamma}}^{-1}\Sigma_{U^{\beta\gamma}}\dot{\Sigma}_{U^{\beta\gamma}}^{-1}\Delta_{\beta\gamma}Q^{-1}. \end{aligned}$$

Remark 1 The form of (8) says that the two-stage filter equals the benchmark filter $\dot{F}^{\alpha}_{\alpha\beta\gamma}$ plus some error given by $M^{-1} P \dot{\Sigma}_W^{-1} \Delta$. If the DGP matches the first-stage component model estimates, then $R = 0$ and terms two and three in the MSE formula vanish; also, terms four and five consolidate to Q^{-1}. One then has the simpler formula

$$Q^{-1} + M^{-1} P \dot{\Sigma}_W^{-1} P' M^{-1}.$$

Typically, $P = 0$ only if the first and second estimates of the component models match. However, this would be a strange assumption in practice. Now Q^{-1} is the MSE matrix of the optimal filter, so the second term in the above expression represents the additional error incurred by utilizing a two-stage approach.

To apply this result to the *basic* two-stage method, we let α be the cycle, β be the trend-irregular, and γ be the seasonal. Then the filter is given by

$$\ddot{F}^{C}_{CTI} \dot{F}^{CTI}_{CSTI} = \dot{F}^{C}_{CSTI} + \left(\ddot{\Sigma}_C^{-1} + \Delta_T' \ddot{\Sigma}_{U^{TI}}^{-1} \Delta_T \right)^{-1} \left[\Delta_T' \ddot{\Sigma}_{U^{TI}}^{-1} \dot{\Sigma}_{U^{TI}} \underline{\Delta}_S' - \ddot{\Sigma}_C^{-1} \dot{\Sigma}_C \Delta' \right] \dot{\Sigma}_W^{-1} \Delta$$

with $U^{TI} = \Delta_T (T + I)$ here. This formula tells us that the additional error essentially arises from having different models for the cycle and trend-irregular in the first and second stages. The MSE matrix is

$$\begin{aligned}
&\left(\ddot{\Sigma}_C^{-1} + \Delta_T' \ddot{\Sigma}_{U^{TI}}^{-1} \Delta_T \right)^{-1} \left[\Delta_T' \ddot{\Sigma}_{U^{TI}}^{-1} \dot{\Sigma}_{U^{TI}} \underline{\Delta}_S' - \ddot{\Sigma}_C^{-1} \dot{\Sigma}_C \Delta' \right] \dot{\Sigma}_W^{-1} \Sigma_W \dot{\Sigma}_W^{-1} \\
&\quad \left[\underline{\Delta}_S \dot{\Sigma}_{U^{TI}} \ddot{\Sigma}_{U^{TI}}^{-1} \Delta_T - \Delta \dot{\Sigma}_C \ddot{\Sigma}_C^{-1} \right] \left(\ddot{\Sigma}_C^{-1} + \Delta_T' \ddot{\Sigma}_{U^{TI}}^{-1} \Delta_T \right)^{-1} \\
&+ \left(\ddot{\Sigma}_C^{-1} + \Delta_T' \ddot{\Sigma}_{U^{TI}}^{-1} \Delta_T \right)^{-1} \left[\Delta_T' \ddot{\Sigma}_{U^{TI}}^{-1} \dot{\Sigma}_{U^{TI}} \underline{\Delta}_S' - \ddot{\Sigma}_C^{-1} \dot{\Sigma}_C \Delta' \right] \dot{\Sigma}_W^{-1} \\
&\quad \left(\Sigma_{U^{STI}} \dot{\Sigma}_{U^{STI}}^{-1} \Delta - \Delta \Sigma_C \dot{\Sigma}_C^{-1} \right) \left(\dot{\Sigma}_C^{-1} + \Delta' \dot{\Sigma}_{U^{STI}}^{-1} \Delta \right)^{-1} \\
&+ \left(\dot{\Sigma}_C^{-1} + \Delta' \dot{\Sigma}_{U^{STI}}^{-1} \Delta \right)^{-1} \left(\Delta' \dot{\Sigma}_{U^{STI}}^{-1} \Sigma_{U^{STI}} - \dot{\Sigma}_C^{-1} \Sigma_C \Delta' \right) \dot{\Sigma}_W^{-1} \\
&\quad \left[\underline{\Delta}_S \dot{\Sigma}_{U^{TI}} \ddot{\Sigma}_{U^{TI}}^{-1} \Delta_T - \Delta \dot{\Sigma}_C \ddot{\Sigma}_C^{-1} \right] \left(\ddot{\Sigma}_C^{-1} + \Delta_T' \ddot{\Sigma}_{U^{TI}}^{-1} \Delta_T \right)^{-1} \\
&+ \left(\dot{\Sigma}_C^{-1} + \Delta' \dot{\Sigma}_{U^{STI}}^{-1} \Delta \right)^{-1} \dot{\Sigma}_C^{-1} \Sigma_C \dot{\Sigma}_C^{-1} \left(\dot{\Sigma}_C^{-1} + \Delta' \dot{\Sigma}_{U^{STI}}^{-1} \Delta \right)^{-1} \\
&+ \left(\dot{\Sigma}_C^{-1} + \Delta' \dot{\Sigma}_{U^{STI}}^{-1} \Delta \right)^{-1} \Delta' \dot{\Sigma}_{U^{STI}}^{-1} \Sigma_{U^{STI}} \dot{\Sigma}_{U^{STI}}^{-1} \Delta \left(\dot{\Sigma}_C^{-1} + \Delta' \dot{\Sigma}_{U^{STI}}^{-1} \Delta \right)^{-1}
\end{aligned} \tag{9}$$

This formula (9) can be used when the true DGP is known (as in the illustration of Section 4). Otherwise, proxies are used for Σ_C and $\Sigma_{U^{STI}}$, e.g., $\dot{\Sigma}_C$ and $\dot{\Sigma}_{U^{STI}}$.

3.3 SEATS two-stage

As alluded to in Section 2, the so-called *SEATS* two-stage can be described via

$$\tilde{F}^{C}_{CT} \dot{F}^{CT}_{CSTI},$$

where $\tilde{F}$ denotes a filter matrix chosen in some *ad hoc* fashion, e.g., using model-based HP filters with pre-determined parameter values. We use the "tilde" notation rather than a double dot, in order to distinguish between filters chosen according to *a priori* design principles and filters that are based on fitted models. Now the above two-stage filter can be made to equal the benchmark, as long as we define the models for C and T appropriately. In fact, the models for C and T will be "implied" by the choice of the given *ad hoc* filter. This is described in Kaiser and Maravall (2005), but assuming a bi-infinite sample; some specific details are supplied in McElroy (2006b).

We first note that a given *ad hoc* symmetric filter can be written as $H(B)H(F)$ for some power series $H(z)$; in order for this cycle filter to be given a model-based interpretation, it is necessary that it accomplish trend-differencing. In fact, its frequency response should contain $|\delta^T(e^{-i\lambda})|^2$ as a factor. Also, the corresponding signal extraction matrix $\tilde{F}^C_{CT}$ must have the following form:

$$\tilde{F}^C_{CT} = \tilde{\Sigma}_C \Delta_T' \tilde{\Sigma}^{-1}_{UCT} \Delta_T,$$

for some covariance matrices $\tilde{\Sigma}_C$ and $\tilde{\Sigma}_{UCT}$ to be determined. In other words, in order for $\tilde{F}^C_{CT}$ to be a model-based finite-sample analogue of $H(B)H(F)$, it must have the form given by the above equation. Since the pseudo-spectral density $\dot{f}_{CT}(\lambda) = \dot{f}_{UCT}(\lambda)/|\delta^T(e^{-i\lambda})|^2$ is known from the first stage, we can use it to define the "implied" models for C and T. Following Kaiser and Maravall (2005) yields:

$$\begin{aligned}
\tilde{f}_C(\lambda) &:= |H(e^{-i\lambda})|^2 \dot{f}_{CT}(\lambda) \\
\tilde{f}_T(\lambda) &:= \left(1 - |H(e^{-i\lambda})|^2\right) \dot{f}_{CT}(\lambda) \\
\tilde{f}_{UT}(\lambda) &:= \left(1 - |H(e^{-i\lambda})|^2\right) \dot{f}_{UCT}(\lambda)
\end{aligned}$$

Note that in order for $\tilde{f}_C$ to correspond to a stationary process, it is necessary that the filter $H(B)H(F)$ contain the trend-differencing factor, which will cancel the $|\delta^T(e^{-i\lambda})|^2$ factor in the denominator of $\dot{f}_{CT}$. Now $\tilde{f}_C$ can be calculated using the results in McElroy (2006b), which will be further discussed in Section 4 below. Once these spectral densities are defined, we obtain the covariance matrices $\tilde{\Sigma}_C$ and $\tilde{\Sigma}_{UCT}$ by the simple formulas

$$\begin{aligned}
\tilde{\Sigma}_C &:= \Sigma(\tilde{f}_C) \\
\tilde{\Sigma}_{UT} &:= \Sigma(\tilde{f}_{UT}),
\end{aligned}$$

where in general $\Sigma_{jk}(g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\lambda) e^{i\lambda(j-k)}\, d\lambda$, i.e., the autocovariance at lag $j-k$ corresponding to the spectral density g. So our filter matrix $\tilde{F}^C_{CT}$ is in turn defined, using these matrices. Now note that, using Lemma 1 of McElroy (2006a), we have

$$\tilde{\Sigma}_{UCT} = \Delta_T \tilde{\Sigma}_C \Delta_T' + \tilde{\Sigma}_{UT} = \Sigma\left(|\delta^T(e^{-i\cdot})|^2 \tilde{f}_C\right) + \Sigma(\tilde{f}_{UT}) = \Sigma(\dot{f}_{UCT}) = \dot{\Sigma}_{UCT}. \tag{10}$$

This is an important relation, since it is needed to demonstrate the optimality of $\tilde{F}^{C}_{CT}\dot{F}^{CT}_{CSTI}$. We have

$$\begin{aligned}\tilde{F}^{C}_{CT}\dot{F}^{CT}_{CSTI} &= \tilde{\Sigma}_C\Delta_T^{'}\tilde{\Sigma}^{-1}_{U^{CT}}\Delta_T\dot{F}^{CT}_{CSTI} \\ &= \tilde{\Sigma}_C\Delta_T^{'}\tilde{\Sigma}^{-1}_{U^{CT}}\dot{\Sigma}_{U^{CT}}\underline{\Delta}_S^{'}\dot{\Sigma}^{-1}_W\Delta \\ &= \tilde{\Sigma}_C\Delta_T^{'}\underline{\Delta}_S^{'}\dot{\Sigma}^{-1}_W\Delta \\ &= \tilde{\Sigma}_C\Delta^{'}\dot{\Sigma}^{-1}_W\Delta,\end{aligned}$$

using (10) and (4). This result is the optimal filter, noting that

$$\dot{\Sigma}_W = \underline{\Delta}_S\dot{\Sigma}_{U^{CT}}\underline{\Delta}_S^{'} + \underline{\Delta}_T\dot{\Sigma}_{U^S}\underline{\Delta}_T^{'} + \Delta\dot{\Sigma}_I\Delta^{'}$$

is determined completely by models specified in stage one, while the matrix $\tilde{\Sigma}_C$ obtained in stage two is an implied model for the cycle. So this provides a verification that the "mixed" and "direct" approaches discussed in Kaiser and Maravall (2005) are equivalent in the finite-sample scenario, so long as we define $\tilde{F}^{C}_{CT}$ in the prescribed manner.

It follows at once that the $SEATS$ approach produces a broadly stationary error process, and the MSE matrix is

$$\Sigma_C + \tilde{\Sigma}_C\Delta^{'}\dot{\Sigma}^{-1}_W\Sigma_W\dot{\Sigma}^{-1}_W\Delta\tilde{\Sigma}_C - \tilde{\Sigma}_C\Delta^{'}\dot{\Sigma}^{-1}_W\Delta\Sigma_C - \Sigma_C\Delta^{'}\dot{\Sigma}^{-1}_W\Delta\tilde{\Sigma}_C.$$

This might be computed by plugging in the proxies $\tilde{\Sigma}_C$ and $\dot{\Sigma}_W$ (without much justification for the former substitution) for the DGP values; the result is

$$\tilde{\Sigma}_C - \tilde{\Sigma}_C\Delta^{'}\dot{\Sigma}^{-1}_W\Delta\tilde{\Sigma}_C.$$

3.4 Truncated and iterated two-stage

As mentioned in Section 2, the *truncated* two-stage method corresponds to using the filter

$$\ddot{F}^{C}_{CTI}(1 - \dot{F}^{S}_{STI}).$$

In the first stage, no accounting of the cycle is made, and so the cycle frequencies get incorporated into the other three components (primarily in the trend). Since the cycle is typically stationary, this will not pose much of a problem for the seasonal adjustment procedure. However, in the second stage we now assume that a cycle is present in the seasonally-adjusted data. Hence we obtain a model for it, and construct the filter $\ddot{F}^{C}_{CTI}$. Formulated more generally, the filter is given by $\ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\beta}_{\beta\gamma}$. Here α is not modeled in the first stage. Computation shows that

$$\ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\beta}_{\beta\gamma} = \left(\Delta_\alpha^{'}\ddot{\Sigma}^{-1}_{U^\alpha}\Delta_\alpha + \Delta_\beta^{'}\ddot{\Sigma}^{-1}_{U^\beta}\Delta_\beta\right)^{-1}\Delta_\beta^{'}\ddot{\Sigma}^{-1}_{U^\beta}\dot{\Sigma}_{U^\beta}\underline{\Delta}_\gamma^{'}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}\Delta_{\beta\gamma}. \tag{11}$$

Clearly this filter will transform $\beta + \gamma$ to a broadly stationary series. However,

$$\ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\beta}_{\beta\gamma} - 1 = \left(\Delta'_{\alpha}\ddot{\Sigma}^{-1}_{U^{\alpha}}\Delta_{\alpha} + \Delta'_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\Delta_{\beta}\right)^{-1}\left[\Delta'_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\dot{\Sigma}_{U^{\beta}}\underline{\Delta}'_{\gamma}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}\Delta_{\beta\gamma} - \Delta'_{\alpha}\ddot{\Sigma}^{-1}_{U^{\alpha}}\Delta_{\alpha} - \Delta'_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\Delta_{\beta}\right],$$

which reduces nonstationary α to broad stationarity iff

$$\Delta'_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\left(\dot{\Sigma}_{U^{\beta}}\underline{\Delta}'_{\gamma}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}\underline{\Delta}_{\gamma} - 1\right)\Delta_{\beta} = 0.$$

But this is only satisfied if $\underline{\Delta}_{\gamma}$ is the zero matrix, so this method fails our basic requirement (6). Of course, there is no problem when α is stationary, as in our applications where it represents the cycle. But for other applications of two-stage filtering, this could be an important consideration. For example, we might have α as the seasonal, β as the irregular, and γ as the trend; this roughly corresponds to the iterative method of $X-11$, but in a model-based formulation – see McElroy and Sutcliffe (2006). Since α may well be nonstationary, we consider a modification of the *truncated* two-stage, which essentially amounts to an iteration of the filters $\ddot{F}^{\alpha}_{\alpha\beta}$ and $\dot{F}^{\beta}_{\beta\gamma}$, and hence is called *iterated* two-stage. We take the *truncated* two-stage filters, and left-multiply by a correction filter as follows:

$$\left(1 - \ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\gamma}_{\beta\gamma}\right)^{-1}\ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\beta}_{\beta\gamma}. \tag{12}$$

This is motivated by the formulas in Theorem 2 (note that $\dot{F}^{\gamma}_{\beta\gamma} = 1 - \dot{F}^{\beta}_{\beta\gamma}$). This is well-defined whenever the left-hand matrix is invertible; a sufficient condition for this invertibility (which could be imposed at the second estimation phase) is *compatibility*: namely, that

$$\dot{\Sigma}_{U^{\beta}}\underline{\Delta}'_{\gamma}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}\underline{\Delta}_{\gamma}\ddot{\Sigma}_{U^{\beta}} \tag{13}$$

is non-negative definite. This is termed compatibility, since β is the shared component of the first and second stage estimations. A second motivation for using the *iterated* two-stage, is that when the first and second stage proxies for the covariances matrices are identical, then (12) yields the optimal filter by Theorem 2. Even when the first and second proxies are not the same, it turns out that the error process for the *iterated* two-stage filter matrix is always broadly stationary, and the following theorem provides the MSE matrix.

Theorem 4 *Assume compatibility (13) as well as the same assumptions as Theorem 1; then*

$$\left(1 - \ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\gamma}_{\beta\gamma}\right)^{-1}\ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\beta}_{\beta\gamma} = M^{-1}\Delta'_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\dot{\Sigma}_{U^{\beta}}\underline{\Delta}'_{\gamma}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}\Delta_{\beta\gamma}$$

is well-defined, where $M = \Delta'_{\alpha}\ddot{\Sigma}^{-1}_{U^{\alpha}}\Delta_{\alpha} + \Delta'_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\dot{\Sigma}_{U^{\beta}}\underline{\Delta}'_{\gamma}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}\Delta_{\beta\gamma}$. *This filter matrix satisfies (6), and the MSE matrix is given by*

$$M^{-1}\left(\Delta'_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\dot{\Sigma}_{U^{\beta}}\underline{\Delta}'_{\gamma}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}\Sigma_{U^{\beta\gamma}}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}\underline{\Delta}_{\gamma}\dot{\Sigma}_{U^{\beta}}\ddot{\Sigma}^{-1}_{U^{\beta}}\Delta_{\beta} + \Delta'_{\alpha}\ddot{\Sigma}^{-1}_{U^{\alpha}}\Sigma_{U^{\alpha}}\ddot{\Sigma}^{-1}_{U^{\alpha}}\Delta_{\alpha}\right)M'^{-1}.$$

Remark 2 This estimate possesses a form of "mixed optimality." If one equates the first and second stage proxies in the filter formula, one clearly recovers the optimal filter $\dot{F}^{\alpha}_{\alpha\beta\gamma}$, which is in agreement with Theorem 2. Because of this property, and the fact that for nonstationary α (12) provides a broadly stationary error process, this *iterated* two-stage is preferable to the *truncated* two-stage method whenever the correction matrix $1 - \ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\gamma}_{\beta\gamma}$ is invertible (compatibility is a sufficient condition for this, but is not necessary).

Remark 3 The term *iterated* two-stage comes from an iterative approach to signal extraction, as described in McElroy and Sutcliffe (2006). Essentially, we can consider an iterative procedure as follows:

$$\begin{aligned}\hat{\gamma} &\leftarrow \dot{F}^{\gamma}_{\beta\gamma}(Y - \hat{\alpha}) \\ \hat{\alpha} &\leftarrow \ddot{F}^{\alpha}_{\alpha\beta}(Y - \hat{\gamma})\end{aligned}$$

So long as the correction matrix is invertible, this iterative procedure converges quickly to the estimates of α and γ described by the formulas of Theorem 4. As mentioned above, McElroy and Sutcliffe (2006) develops this algorithm with α equal to the seasonal, β equal to the irregular, and γ as the trend.

To apply this result to the two-stage method for cycle estimation, we let α be the cycle, β be the trend-irregular, and γ be the seasonal. Assuming compatibility, the filter is given by

$$\left(1 - \ddot{F}^{C}_{CTI}\dot{F}^{S}_{STI}\right)^{-1}\ddot{F}^{C}_{CTI}\dot{F}^{TI}_{STI} = \left(\ddot{\Sigma}_C^{-1} + \Delta_T^{'}\ddot{\Sigma}^{-1}_{U^{TI}}\dot{\Sigma}_{U^{TI}}\underline{\Delta}_S^{'}\dot{\Sigma}^{-1}_{U^{STI}}\Delta\right)^{-1}\Delta_T^{'}\ddot{\Sigma}^{-1}_{U^{TI}}\dot{\Sigma}_{U^{TI}}\underline{\Delta}_S^{'}\dot{\Sigma}^{-1}_{U^{STI}}\Delta.$$

Note that this filter can be computed solely from a knowledge of the matrices $\dot{F}^{S}_{STI}$ and $\ddot{F}^{C}_{CTI}$. So long as the first-stage filter $\dot{F}^{S}_{STI}$ is made available to the cycle analyst, the correction factor can be computed and applied (assuming it is invertible) completely in the second stage. The MSE matrix is given by

$$\left(\ddot{\Sigma}_C^{-1} + \Delta_T^{'}\ddot{\Sigma}^{-1}_{U^{TI}}\dot{\Sigma}_{U^{TI}}\underline{\Delta}_S^{'}\dot{\Sigma}^{-1}_{U^{STI}}\Delta\right)^{-1}$$
$$\left(\Delta_T^{'}\ddot{\Sigma}^{-1}_{U^{TI}}\dot{\Sigma}_{U^{TI}}\underline{\Delta}_S^{'}\dot{\Sigma}^{-1}_{U^{STI}}\Sigma_{U^{STI}}\dot{\Sigma}^{-1}_{U^{STI}}\underline{\Delta}_S\dot{\Sigma}_{U^{TI}}\ddot{\Sigma}^{-1}_{U^{TI}}\Delta_T + \ddot{\Sigma}_C^{-1}\Sigma_C\ddot{\Sigma}_C^{-1}\right)\left(\ddot{\Sigma}_C^{-1} + \Delta^{'}\dot{\Sigma}^{-1}_{U^{STI}}\underline{\Delta}_S\dot{\Sigma}_{U^{TI}}\ddot{\Sigma}^{-1}_{U^{TI}}\Delta_T\right)^{-1},$$

which can be computed using the proxies $\Sigma_{U^{STI}} = \dot{\Sigma}_{U^{STI}}$ and $\Sigma_C = \ddot{\Sigma}_C$.

Now when α is stationary, the *truncated* two-stage can be used, and its error process is broadly stationary. The MSE matrix in this case is

$$\left(\ddot{\Sigma}_\alpha^{-1} + \Delta_\beta^{'}\ddot{\Sigma}^{-1}_{U^\beta}\Delta_\beta\right)^{-1}\left(\Delta_\beta^{'}\ddot{\Sigma}^{-1}_{U^\beta}\dot{\Sigma}_{U^\beta}\underline{\Delta}_\gamma^{'}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}\Sigma_{U^{\beta\gamma}}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}\underline{\Delta}_\gamma\dot{\Sigma}_{U^\beta}\ddot{\Sigma}^{-1}_{U^\beta}\Delta_\beta + L\Sigma_\alpha L^{'}\right)\left(\ddot{\Sigma}_\alpha^{-1} + \Delta_\beta^{'}\ddot{\Sigma}^{-1}_{U^\beta}\Delta_\beta\right)^{-1}$$
$$L = \left[\Delta_\beta^{'}\ddot{\Sigma}^{-1}_{U^\beta}\dot{\Sigma}_{U^\beta}\underline{\Delta}_\gamma^{'}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}\Delta_{\beta\gamma} - \Delta_\alpha^{'}\ddot{\Sigma}_\alpha^{-1}\Delta_\alpha - \Delta_\beta^{'}\ddot{\Sigma}^{-1}_{U^\beta}\Delta_\beta\right] = -\Delta_\beta^{'}\ddot{\Sigma}^{-1}_{U^\beta}\Delta_\beta\dot{F}^{\gamma}_{\beta\gamma} - \Delta_\alpha^{'}\ddot{\Sigma}_\alpha^{-1}\Delta_\alpha.$$

This can be compared to the MSE in Theorem 4 when α is stationary. In the case of cycle-seasonal-trend-irregular data, this formula becomes

$$\left(\ddot{\Sigma}_C^{-1} + \Delta_T^{'}\ddot{\Sigma}_{UTI}^{-1}\Delta_T\right)^{-1}\left(\Delta_T^{'}\ddot{\Sigma}_{UTI}^{-1}\dot{\Sigma}_{UTI}\underline{\Delta}_S^{'}\dot{\Sigma}_{USTI}^{-1}\Sigma_{USTI}\dot{\Sigma}_{USTI}^{-1}\underline{\Delta}_S\dot{\Sigma}_{UTI}\ddot{\Sigma}_{UTI}^{-1}\Delta_T + L\Sigma_C L^{'}\right)\left(\ddot{\Sigma}_C^{-1} + \Delta_T^{'}\ddot{\Sigma}_{UTI}^{-1}\Delta_T\right)$$
$$L = \Delta_T^{'}\ddot{\Sigma}_{UTI}^{-1}\Delta_T\dot{F}_{STI}^{S} - \ddot{\Sigma}_C^{-1},$$

which can be computed using the proxies $\Sigma_{USTI} = \dot{\Sigma}_{USTI}$ and $\Sigma_C = \ddot{\Sigma}_C$.

4 Illustration

The formulas of the previous section are the main focus of this paper. In order to demonstrate how they can be used in practice, we set out a detailed illustration. The goal is to apply the various methods – one-stage, *basic* two-stage, $SEATS$ two-stage, *truncated* two-stage, and *iterated* two-stage – for cycle signal extraction. We wish to examine not only the cycle estimates from these five methods, but more importantly the MSE plots. However, in order to compute the MSE's it is necessary to have some knowledge of the DGP; as mentioned in the previous section, these MSE's can be calculated by using certain proxies. We choose to illustrate the different MSE's by using a simulation, for which the true DGP is known. Even though the model-based filters are derived from data-dependent parameter estimates, we will treat the filters as if they are fixed for the purpose of calculating the MSE. Clearly, the quality of the cycle estimates depends upon having decent models for the components, and having a reliable estimation procedure. In order to provide a suitable boundary for our study, we will utilize one estimation procedure, which will be applied to all of the five approaches in the same manner. By keeping the model estimation procedure fixed, we will be comparing the filter approaches (this assumes that our model estimation procedure does not inherently favor one filtering approach over another). We utilize the so-called *hybrid* approach, which is described below.

The hybrid approach is designed to be a flexible method for modeling either three-component ($S+T+I$ or $C+T+I$) or four component ($C+S+T+I$) models. The idea is simple: if a cycle is thought to be present, it is given a separate structural model, whereas combinations of the seasonal, trend, and irregular are given an appropriate $SARIMA$ model (Box and Jenkins, 1976) specification. For the one-stage benchmark, this means specifying an Airline model for $S+T+I$, and a separate $ARMA(2,1)$ model for C. Then the models for S, T, and I can be obtained by canonical decomposition (this is why we call the method "hybrid," being a mixture of decomposition and structural approaches). For the *basic* two-stage, we use these same models for the first step. Then we use a hybrid model for the seasonally adjusted data, utilizing a separate component for C and another $SARIMA$ model for $T+I$; the models for T and I can then be found by canonical decomposition (though this is not necessary for the calculations). In the $SEATS$ approach, we omit the cycle model and just use an Airline model on the data, performing a canonical decomposition on $S+TC+I$, where the trend-cycle TC is given an appropriate $ARIMA$ model. Then we obtain implied models for T and C via the method of section 3.3. The fixed cycle filter that we use is an HP filter with signal-noise ratio set at $q = 1/1600$. For the *truncated* two-stage, we again use canonical decomposition on the Airline model, just as in the $SEATS$ two-stage. Then in the second step we model the seasonally adjusted data with a $C + T + I$ model, in the same fashion as the *basic* two-stage. Finally, the *iterated* two-stage requires no additional estimation, being directly implemented from the models obtained from the *truncated* two-stage. This describes the general procedure followed; section 4.1 below gives specific information about the models obtained for each method.

We chose to simulate a time series with cycle, seasonal, and trend behavior. The reason for using a simulation is two-fold: we are guaranteed that a cycle actually exists in the data (whereas this may not be obvious in the case of real data), and we have knowledge of the DGP which allows us to compute the exact MSE's. Moreover, we can control the relative strength of cyclical, seasonal, trend, and irregular components in the simulation, which will facilitate estimation of the models. More details on the simulation are provided in section 4.1; section 4.2 discusses the results, giving a comparison of the cycle estimates and their MSE's.

4.1 Model Estimation

The cycle model that we utilize is the first-order cycle of Harvey and Trimbur (2003), whereas combinations of the trend, seasonal, and irregular are given various $SARIMA$ models. Our simulation was drawn from the following models:

$$(1 - 2(.9)\cos(\pi/24)\,B + .9^2B^2)C_t = (1 - .9\cos(\pi/24)\,B)\epsilon_t^C$$
$$(1 - B)(1 - B^{12})(S_t + T_t + I_t) = (1 - .6B)(1 - .6B^{12})\epsilon_t^{STI}$$

with $\epsilon_t^C \sim WN(0, .01)$ and $\epsilon_t^{STI} \sim WN(0, .001)$. The cycle innovation variance is 10 times that of the airline model; this was determined by trial-and-error to give a reasonable balance of cyclical, seasonal, and trend behavior. (These small values were used so that they would be appropriate for the choice of initial values used for the airline model simulation.) The parameter $\rho = .9$ in the cycle model indicates the persistency ($\rho = 1$ corresponds to a nonstationary cycle), while the frequency $\omega = \pi/24$ corresponds to the peak in the cycle's spectral density. This frequency corresponds to a period of 4 years; cycles are thought to have a period of between 2 and 10 years (see Harvey and Trimbur, 2003 and Kaiser and Maravall, 2005). The Airline model parameters represent moderate values, appropriate for many U.S. Census Bureau time series. A sample of 480 values was drawn, using initial values for the seasonal-trend-irregular drawn from a real series (which were then discarded). The cycle was initialized with a pair of zeroes, but a burn-in period of 120 observations was used. Figure 1 displays the simulated cycle, simulated seasonal-trend-irregular, and the simulated series; note that this is a stochastic cycle, and thus appears more "rough" than the estimated cycle that many are accustomed to seeing.

One-stage We fit the $ARMA(2,1)$ cycle model plus an airline model to the data, obtaining

$$
\begin{aligned}
(1 - 2(.86)\cos(.043\pi)\,B + .86^2 B^2)C_t &= (1 - .86\cos(.043\pi)\,B)\epsilon_t^C \\
(1-B)(1-B^{12})(S_t + T_t + I_t) &= (1 - .52B)(1 - .47B^{12})\epsilon_t^{STI} \\
U(B)S_t &= (1 + 1.24B + 1.19B^2 + 1.03B^3 + 0.80B^4 + 0.55B^5 + 0.31B^6 \\
&\quad + 0.08B^7 - 0.09B^8 - 0.25B^9 - 0.34B^{10} - 0.57B^{11})\epsilon_t^S \\
(1-B)^2 T_t &= (1 + 0.06B - 0.94B^2)\epsilon_t^T,
\end{aligned}
$$

where $U(z) = 1 + z + z^2 + \cdots + z^{11}$. We have included the models for seasonal, trend, and irregular, which were obtained by canonical decomposition. The white noise variances are (respectively for cycle, airline model, seasonal, trend, and irregular):

0.011885	0.000936	0.000065	0.000028	0.000292

All of the estimates seem reasonably good, given the true parameters. The optimal cycle extraction filter is obtained from the models, and used to produce the cycle estimate and its MSE, as described in Section 3.1.

Basic two-stage We take the seasonally adjusted component $\dot{F}_{CSTI}^{CTI} Y$ from the *one-stage* above, and fit a cycle plus $SARIMA(0,2,1)(0,0,1)$ model. At first we tried just an $ARIMA(0,2,2)$ model for the trend-irregular, but there tends to be residual "stationary seasonality" in the seasonally adjusted data, which is adequately modeled by using a seasonal moving average. Moreover the $SARIMA(0,2,1)(0,0,1)$ had an admissible decomposition. The estimated models are

$$
\begin{aligned}
(1 - 2(.84)\cos(.057\pi)\,B + .84^2 B^2)C_t &= (1 - .84\cos(.057\pi)\,B)\epsilon_t^C \\
(1-B)^2(T_t + I_t) &= (1 - .96B)(1 - .65B^{12})\epsilon_t^{TI},
\end{aligned}
$$

with innovation variances for cycle and trend-irregular given by 0.00929 and 0.00154 respectively. It is not necessary to decompose the trend-irregular, so this is omitted. The cycle estimates are quite good, considering they are obtained from filtered data. Note that $.057\pi$ corresponds to a period of 2.92 years. The cycle estimate and MSE are then determined according to Section 3.2.

SEATS two-stage The first step is to fit an airline model to the data, ignoring cyclical behavior. The estimated model is

$$
\begin{aligned}
(1-B)(1-B^{12})(S_t+T_t+I_t) &= (1-.09B)(1-.87B^{12})\epsilon_t^{STI} \\
U(B)S_t &= (1+1.76B+2.05B^2+2.10B^3+1.95B^4+1.68B^5+1.35B^6 \\
&\quad +0.99B^7+0.68B^8+0.36B^9+0.17B^{10}-0.17B^{11})\epsilon_t^S \\
(1-B)^2T_t &= (1+0.01B-0.99B^2)\epsilon_t^T,
\end{aligned}
$$

with innovation variances given by

0.015869 0.000095 0.002895 0.004090

for seasonal-trend-irregular, seasonal, trend, and irregular respectively. The notation T for trend is a bit confusing, since we really think of this as trend-cycle in the $SEATS$ two-stage procedure. The estimates are reasonable; note that the higher innovation variance for seasonal-trend-irregular is due to the fact that it must absorb the contributions to the variance due to the cycle. Now the implied model for the cycle, using the results of McElroy (2006b) and the discussion in Section 3.1, works out to be

$$
(1-1.777091B+.7994438B^2)\tilde{C}_t = (1+0.01B-0.99B^2)\epsilon_t^{\tilde{C}}
$$

with innovation variance 0.002895. The $AR(2)$ polynomial comes from the HP filter with signal-noise ratio 1/1600; this choice of q is a popular *ad hoc* value, and corresponds to a cycle period of 4.688 years (McElroy, 2006b). (The actual value of q corresponding to the given frequency $\pi/24$ is $q=.0018$.) Then we can compute the cycle extraction filter directly from the formula $\tilde{\Sigma}_C\Delta'\dot{\Sigma}_W^{-1}\Delta$, so it is not necessary to actually apply two filters. The MSE is also calculated as described in Section 3.3.

Truncated and Iterated two-stage We start with the same airline model as in the $SEATS$ two-stage, and we compute the seasonally adjusted component. Then an $ARMA(2,1)$ cycle and $SARIMA(0,2,1)(0,0,1)$ trend-irregular is fitted to the seasonally adjusted data, resulting in the models

$$
\begin{aligned}
(1-2(.83)\cos(.044\pi)\,B+.83^2B^2)C_t &= (1-.83\cos(.044\pi)\,B)\epsilon_t^C \\
(1-B)^2(T_t+I_t) &= (1-.96B)(1-.64B^{12})\epsilon_t^{TI},
\end{aligned}
$$

with innovation variances for cycle and trend-irregular given by 0.00976 and 0.00134 respectively. The cycle model is encouraging, being even closer than the second-stage estimate in the *basic approach*. With these models, both the truncated and iterated estimates could be computed as described in Section 3.4.

4.2 Results and Summary

In Figure 2 we plot the *one-stage* and $SEATS$ method cycle estimate against the true component, while Figure 3 compares the other estimates to the cycle. The *one-stage* and $SEATS$ methods most closely reproduced the true underlying cycle. The other methods do not seem to match the true component's movements, which is perhaps not surprising. These two-stage methods attempt to estimate the cycle from seasonally adjusted data; the cycle dynamics may be distorted by the seasonal adjustment procedure. It is interesting that the $SEATS$ method seems to do a better job than the others, which may be because it can be reinterpreted as a one-stage method using the implied models for cycle and trend. The *basic* two-stage, *truncated*, and *iterated* methods capture very few of the cyclical movements, and in addition their variance is too small. Of course, this only represents one example.

The MSE's are compared in Figure 4. As expected, the *one-stage* method does the best. These are "true" MSE's in the sense that the true DGP autocovariances were used rather than the proxies that would be used for real data. Some of the error represented here is due to model estimation error, since if the model estimates were exact there would be simplifications in the MSE formulas, resulting in a lower MSE plot. We chose to present these results, which represent the true error when the following methods are applied (viewing the filters as fixed, rather than being data-dependent). The $SEATS$ method is superior to the other two-stage approaches. The *iterated* and *truncated* do a little better than the *basic*; it is surprising that the *truncated* is superior to the *iterated* in this case. In any event, this figure gives some idea of the scale of the additional error induced by two-stage methods.

Some comments on the implementation and computational time are in order. The code that performed the model estimation and signal extraction was written in the econometrics language Ox, utilizing some functions from SsfPack (Koopman, Shephard, Doornik, 1999). Although many matrix inversions of size 480 were required, the whole procedure was completed in a couple of minutes. While prohibitive for mass production, this is quite tolerable for individual data analysis. Whereas matrix-based approaches to signal extraction are slower than the state space smoother, they tend to be easier to implement; moreover, the MSE formulas needed in this work *cannot* be obtained from state space methods.

Returning to the original questions raised in this paper, we see that the error for two-stage methods can indeed be quantified. The $SEATS$ and *truncated* two-stage procedures correspond most closely with current cycle estimation practice, and the latter can be corrected with the *iterated* procedure to handle nonstationary cycles. For the *basic*, *truncated*, and *iterated* procedures, a knowledge of the models used in both stages is necessary. The limited analysis of a single simulation supports the idea that direct cycle estimation given by the *one-stage* procedure is the best (we note that for this simulation, excellent model estimates for all of the dynamics were obtained). However, this is only possible if the original data is made available to cycle analysts. Since the *one-stage* approach is often not practicable, the $SEATS$ method seems to be the next best alternative.

5 Appendix

Proof of Theorem 1. First we note that $F^{\alpha}_{\alpha\beta\gamma} + F^{\beta}_{\alpha\beta\gamma} + F^{\gamma}_{\alpha\beta\gamma} = 1$ implies that $F^{\alpha}_{\alpha\beta\gamma} + F^{\beta}_{\alpha\beta\gamma} = 1 - F^{\gamma}_{\alpha\beta\gamma} = F^{\alpha\beta}_{\alpha\beta\gamma}$. Now the theorem results from a commutativity property of the extraction matrices:

$$F^{\alpha}_{\alpha\beta} F^{\alpha\beta}_{\alpha\beta\gamma} = (1 - F^{\beta}_{\alpha\beta}) F^{\alpha}_{\alpha\beta\gamma} + F^{\alpha}_{\alpha\beta} F^{\beta}_{\alpha\beta\gamma} = F^{\alpha}_{\alpha\beta\gamma} + \left(F^{\alpha}_{\alpha\beta} F^{\beta}_{\alpha\beta\gamma} - F^{\beta}_{\alpha\beta} F^{\alpha}_{\alpha\beta\gamma} \right)$$

which uses $F^{\alpha}_{\alpha\beta} + F^{\beta}_{\alpha\beta} = 1$. The expression in parentheses is similar to a commutator for the matrices; below we show that it is zero, which will prove the theorem. Let the matrices $\underline{\Delta}_{\alpha\beta}$, $\underline{\Delta}_{\alpha\gamma}$, and $\underline{\Delta}_{\beta\gamma}$ be defined similarly to $\underline{\Delta}_T$ and $\underline{\Delta}_S$, but with coefficients of $\delta^{\alpha}\delta^{\beta}$, $\delta^{\alpha}\delta^{\gamma}$, and $\delta^{\beta}\delta^{\gamma}$ respectively, and with appropriate dimensions such that, similar to (4),

$$\underline{\Delta}_{\alpha\beta} \Delta_{\gamma} = \underline{\Delta}_{\alpha\gamma} \Delta_{\beta} = \underline{\Delta}_{\beta\gamma} \Delta_{\alpha} = \Delta$$

is satisfied. Following McElroy (2005), the filter formulas are given by

$$\begin{aligned} F^{\alpha}_{\alpha\beta} &= \left(\Delta^{'}_{\alpha} \Sigma^{-1}_{U^{\alpha}} \Delta_{\alpha} + \Delta^{'}_{\beta} \Sigma^{-1}_{U^{\beta}} \Delta_{\beta} \right)^{-1} \Delta^{'}_{\beta} \Sigma^{-1}_{U^{\beta}} \Delta_{\beta} \\ \Delta_{\beta} F^{\beta}_{\alpha\beta\gamma} &= \Sigma_{U^{\beta}} \underline{\Delta}^{'}_{\alpha\gamma} \Sigma^{-1}_{W} \Delta. \end{aligned}$$

Therefore $F^{\alpha}_{\alpha\beta} F^{\beta}_{\alpha\beta\gamma} = \left(\Delta^{'}_{\alpha} \Sigma^{-1}_{U^{\alpha}} \Delta_{\alpha} + \Delta^{'}_{\beta} \Sigma^{-1}_{U^{\beta}} \Delta_{\beta} \right)^{-1} \Delta^{'} \Sigma^{-1}_{W} \Delta$. Similarly,

$$\begin{aligned} F^{\beta}_{\alpha\beta} &= \left(\Delta^{'}_{\alpha} \Sigma^{-1}_{U^{\alpha}} \Delta_{\alpha} + \Delta^{'}_{\beta} \Sigma^{-1}_{U^{\beta}} \Delta_{\beta} \right)^{-1} \Delta^{'}_{\alpha} \Sigma^{-1}_{U^{\alpha}} \Delta_{\alpha} \\ \Delta_{\alpha} F^{\alpha}_{\alpha\beta\gamma} &= \Sigma_{U^{\alpha}} \underline{\Delta}^{'}_{\beta\gamma} \Sigma^{-1}_{W} \Delta \end{aligned}$$

which produces $F^{\beta}_{\alpha\beta} F^{\alpha}_{\alpha\beta\gamma} = \left(\Delta^{'}_{\alpha} \Sigma^{-1}_{U^{\alpha}} \Delta_{\alpha} + \Delta^{'}_{\beta} \Sigma^{-1}_{U^{\beta}} \Delta_{\beta} \right)^{-1} \Delta^{'} \Sigma^{-1}_{W} \Delta$. This concludes the proof. $\square$

Proof of Theorem 2. We prove the first line of (6). To show the invertibility of $1 - F^{\alpha}_{\alpha\gamma} F^{\beta}_{\beta\gamma}$, we compute

$$1 - F^{\alpha}_{\alpha\gamma} F^{\beta}_{\beta\gamma} = 1 - F^{\alpha}_{\alpha\gamma} (1 - F^{\gamma}_{\beta\gamma}) = F^{\gamma}_{\alpha\gamma} + F^{\alpha}_{\alpha\gamma} F^{\gamma}_{\beta\gamma}.$$

We need to define some additional differencing matrices. Let $\underline{\Delta}_{\beta}$ and $\underline{\Delta}_{\gamma}$ be defined similarly to the matrices in the proof of Theorem 1 above. Also, the matrix $\Delta_{\beta\gamma}$ does $\delta^{\beta}\delta^{\gamma}$ differencing, but is $n - (d_{\beta} + d_{\gamma}) \times n$ dimensional. Then we have the relation $\Delta_{\beta\gamma} = \underline{\Delta}_{\beta} \Delta_{\gamma} = \underline{\Delta}_{\gamma} \Delta_{\beta}$ as in (4). Let $U^{\beta\gamma}$ denote the differenced $\beta + \gamma$ component:

$$U^{\beta\gamma} = \Delta_{\beta\gamma} (\beta + \gamma) = \underline{\Delta}_{\gamma} U^{\beta} + \underline{\Delta}_{\beta} U^{\gamma}$$

Then the following formulas hold:

$$
\begin{aligned}
F^{\alpha}_{\alpha\gamma} &= \left(\Delta^{'}_{\alpha}\Sigma^{-1}_{U^{\alpha}}\Delta_{\alpha} + \Delta^{'}_{\gamma}\Sigma^{-1}_{U^{\gamma}}\Delta_{\gamma}\right)^{-1}\Delta^{'}_{\gamma}\Sigma^{-1}_{U^{\gamma}}\Delta_{\gamma} \\
F^{\gamma}_{\alpha\gamma} &= \left(\Delta^{'}_{\alpha}\Sigma^{-1}_{U^{\alpha}}\Delta_{\alpha} + \Delta^{'}_{\gamma}\Sigma^{-1}_{U^{\gamma}}\Delta_{\gamma}\right)^{-1}\Delta^{'}_{\alpha}\Sigma^{-1}_{U^{\alpha}}\Delta_{\alpha} \\
\Delta_{\gamma}F^{\gamma}_{\beta\gamma} &= \Sigma_{U^{\gamma}}\underline{\Delta}^{'}_{\beta}\Sigma^{-1}_{U^{\beta\gamma}}\Delta_{\beta\gamma} \\
F^{\alpha}_{\alpha\gamma}F^{\gamma}_{\beta\gamma} &= \left(\Delta^{'}_{\alpha}\Sigma^{-1}_{U^{\alpha}}\Delta_{\alpha} + \Delta^{'}_{\gamma}\Sigma^{-1}_{U^{\gamma}}\Delta_{\gamma}\right)^{-1}\Delta^{'}_{\beta\gamma}\Sigma^{-1}_{U^{\beta\gamma}}\Delta_{\beta\gamma} \\
F^{\gamma}_{\alpha\gamma} + F^{\alpha}_{\alpha\gamma}F^{\gamma}_{\beta\gamma} &= \left(\Delta^{'}_{\alpha}\Sigma^{-1}_{U^{\alpha}}\Delta_{\alpha} + \Delta^{'}_{\gamma}\Sigma^{-1}_{U^{\gamma}}\Delta_{\gamma}\right)^{-1}\left[\Delta^{'}_{\alpha}\Sigma^{-1}_{U^{\alpha}}\Delta_{\alpha} + \Delta^{'}_{\beta\gamma}\Sigma^{-1}_{U^{\beta\gamma}}\Delta_{\beta\gamma}\right]
\end{aligned}
$$

The expression in square brackets is invertible, since it is the sum of two non-negative definite matrices whose null spaces' intersection is the zero vector, due to the fact that δ^{α} and $\delta^{\beta}\delta^{\gamma}$ are relatively prime – see Lemma 2 of McElroy and Sutcliffe (2006). This establishes the invertibility of $1 - F^{\alpha}_{\alpha\gamma}F^{\beta}_{\beta\gamma}$, and gives an expression for its inverse. Finally, we compute the result (6):

$$
\begin{aligned}
&\left(1 - F^{\alpha}_{\alpha\gamma}F^{\beta}_{\beta\gamma}\right)^{-1}F^{\alpha}_{\alpha\gamma}(1 - F^{\beta}_{\beta\gamma}) \\
&= \left[\Delta^{'}_{\alpha}\Sigma^{-1}_{U^{\alpha}}\Delta_{\alpha} + \Delta^{'}_{\beta\gamma}\Sigma^{-1}_{U^{\beta\gamma}}\Delta_{\beta\gamma}\right]^{-1}\Delta^{'}_{\gamma}\Sigma^{-1}_{U^{\gamma}}\Delta_{\gamma}F^{\gamma}_{\beta\gamma} \\
&= \left[\Delta^{'}_{\alpha}\Sigma^{-1}_{U^{\alpha}}\Delta_{\alpha} + \Delta^{'}_{\beta\gamma}\Sigma^{-1}_{U^{\beta\gamma}}\Delta_{\beta\gamma}\right]^{-1}\Delta^{'}_{\gamma}\underline{\Delta}^{'}_{\beta}\Sigma^{-1}_{U^{\beta\gamma}}\Delta_{\beta\gamma} = F^{\alpha}_{\alpha\beta\gamma}. \quad \square
\end{aligned}
$$

Proof of Theorem 3. Using the notations of the proof of Theorem 1, we have

$$
\begin{aligned}
\ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\alpha\beta}_{\alpha\beta\gamma} &= \ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\alpha}_{\alpha\beta\gamma} + \ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\beta}_{\alpha\beta\gamma} \\
&= \dot{F}^{\alpha}_{\alpha\beta\gamma} - \ddot{F}^{\beta}_{\alpha\beta}\dot{F}^{\alpha}_{\alpha\beta\gamma} + \ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\beta}_{\alpha\beta\gamma} \\
&= \dot{F}^{\alpha}_{\alpha\beta\gamma} - \left(\Delta^{'}_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\Delta_{\beta} + \Delta^{'}_{\alpha}\ddot{\Sigma}^{-1}_{U^{\alpha}}\Delta_{\alpha}\right)^{-1}\Delta^{'}_{\alpha}\ddot{\Sigma}^{-1}_{U^{\alpha}}\dot{\Sigma}_{U^{\alpha}}\underline{\Delta}^{'}_{\beta\gamma}\dot{\Sigma}^{-1}_{W}\Delta \\
&+ \left(\Delta^{'}_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\Delta_{\beta} + \Delta^{'}_{\alpha}\ddot{\Sigma}^{-1}_{U^{\alpha}}\Delta_{\alpha}\right)^{-1}\Delta^{'}_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\dot{\Sigma}_{U^{\beta}}\underline{\Delta}^{'}_{\alpha\gamma}\dot{\Sigma}^{-1}_{W}\Delta \\
&= \dot{F}^{\alpha}_{\alpha\beta\gamma} + M^{-1}P\dot{\Sigma}^{-1}_{W}\Delta.
\end{aligned}
$$

Next, the error process can be expressed as

$$
\begin{aligned}
\ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\alpha\beta}_{\alpha\beta\gamma}Y - \alpha &= (\dot{F}^{\alpha}_{\alpha\beta\gamma} - 1)\alpha + \dot{F}^{\alpha}_{\alpha\beta\gamma}(\beta + \gamma) + M^{-1}P\dot{\Sigma}^{-1}_{W}\Delta Y \\
&= -Q^{-1}\Delta^{'}_{\alpha}\dot{\Sigma}^{-1}_{U^{\alpha}}U^{\alpha} + Q^{-1}\Delta^{'}_{\beta\gamma}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}U^{\beta\gamma} + M^{-1}P\dot{\Sigma}^{-1}_{W}W.
\end{aligned} \tag{14}
$$

Since $W = \underline{\Delta}_{\beta\gamma}U^{\alpha} + \overline{\Delta}_{\alpha}U^{\beta\gamma}$, we can easily compute the following covariances:

$$
\mathbb{E}[U^{\alpha}U^{\beta\gamma\prime}] = 0 \qquad \mathbb{E}[U^{\beta\gamma}W'] = \Sigma_{U^{\beta\gamma}}\overline{\Delta}_{\alpha}{}^{'} \qquad \mathbb{E}[U^{\alpha}W^{'}] = \Sigma_{U^{\alpha}}\underline{\Delta}_{\beta\gamma}{}^{'}
$$

Using these in (14) yields the stated expression for the error covariance matrix. $\square$

Proof of Theorem 4. We start with (11), to which we left-multiply the inverse of

$$1 - \ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\gamma}_{\beta\gamma} = 1 - \ddot{F}^{\alpha}_{\alpha\beta} + \ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\beta}_{\beta\gamma}$$
$$= \left(\Delta'_{\alpha}\ddot{\Sigma}^{-1}_{U^{\alpha}}\Delta_{\alpha} + \Delta'_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\Delta_{\beta}\right)^{-1} \left[\Delta'_{\alpha}\ddot{\Sigma}^{-1}_{U^{\alpha}}\Delta_{\alpha} + \Delta'_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\dot{\Sigma}_{U^{\beta}}\underline{\Delta}'_{\gamma}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}\Delta_{\beta\gamma}\right],$$

which again uses (11). Under the compatibility condition (13) the matrix in square brackets is invertible, and we at once obtain the formula for $\left(1 - \ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\gamma}_{\beta\gamma}\right)^{-1}\ddot{F}^{\alpha}_{\alpha\beta}\dot{F}^{\beta}_{\beta\gamma}$ by inverting the above expression. The error process (defining M as in Theorem 4) is

$$M^{-1}\Delta'_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\dot{\Sigma}_{U^{\beta}}\underline{\Delta}'_{\gamma}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}\Delta_{\beta\gamma}Y - \alpha = M^{-1}\left(\Delta'_{\beta}\ddot{\Sigma}^{-1}_{U^{\beta}}\dot{\Sigma}_{U^{\beta}}\underline{\Delta}'_{\gamma}\dot{\Sigma}^{-1}_{U^{\beta\gamma}}U^{\beta\gamma} - \Delta'_{\alpha}\ddot{\Sigma}^{-1}_{U^{\alpha}}U^{\alpha}\right),$$

from which the stated MSE formula follows at once. □

References

[1] Bell, W. (1984) Signal extraction for nonstationary time series. *The Annals of Statistics* **12**, 646 – 664.

[2] Bell, W. (2004) "On RegComponent Time Series Models and Their Applications," in *State Space and Unobserved Component Models: Theory and Applications,* eds. Andrew C. Harvey, Siem Jan Koopman, and Neil Shephard, Cambridge, UK: Cambridge University Press.

[3] Bell, W. and Hilmer, S. (1988) A matrix approach to likelihood evaluation and signal extraction for ARIMA component time series models. *SRD Research Report No. RR*−88/22, U.S. Census Bureau.

[4] Bell, W. and Hillmer, S. (1991). Initializing the Kalman filter for nonstationary time series models. *Journal of Time Series Analysis* **12**, 283 – 300.

[5] Bell, W. and Hillmer, S. (1992). Correction to, "Initializing the Kalman filter for nonstationary time series models." *Journal of Time Series Analysis* **13**, 281 – 282.

[6] Box, G. and Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control,* Holden-Day: San Francisco.

[7] Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods,* Oxford: Oxford University Press.

[8] Gómez, V. (2001) The Use of Butterworth Filters for Trend and Cycle Estimation in Economic Time Series. *Journal of Business and Economics Statistics* **19**, 365 – 373.

[9] Gómez, V. and Maravall, A. (1997). Program SEATS signal extraction in ARIMA time series: instructions for the user. *Working Paper 97001, Ministerio de Economía y Hacienda, Dirrectión General de Análysis y Programación Presupuestaria*, Madrid.

[10] Harvey, A. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter,* Cambridge: Cambridge University Press.

[11] Harvey, A. and Jaeger, A. (1993) Detrending, Stylized Facts and the Business Cycle. *Journal of Applied Econometrics* **8**, 231-247.

[12] Harvey, A. and Trimbur, T. (2003) General Model-Based Filters for Extracting Cycles and Trends in Economic Time Series. *Review of Economics and Statistics* **85**, 244 – 255.

[13] Hillmer, S. and Tiao, G. (1982). An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association* **77**, 377, 63 – 70.

[14] Hodrick, R., and Prescott, E. (1997) Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit, and Banking* **29**, 1 – 16.

[15] Kaiser, R., and Maravall, A. (2005) Combining filter design with model-based filtering (with an application to business-cycle estimation). *International Journal of Forecasting* **21**, 691 – 710.

[16] Koopman, S., Shephard, N., and Doornik, J. (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal* **2**, 113 – 166.

[17] McElroy, T. (2005) Matrix Formulas for Nonstationary Signal Extraction. SRD Research Report No. *RRS*2005/04, U.S. Census Bureau. Also www.census.gov/srd/papers/pdf/rrs2005-04.pdf

[18] McElroy, T. (2006a) Statistical Properties of Model-Based Signal Extraction Diagnostic Test. *SRD Research Report No. RRS*2006/06, U.S. Census Bureau. Also www.census.gov/srd/papers/pdf/rrs2006-06.pdf

[19] McElroy, T. (2006b) Exact Spectral Factorization of the Hodrick-Prescott Filter. *mimeo*, U.S. Census Bureau.

[20] McElroy, T. and Sutcliffe, A. (2006) An iterated parametric approach to nonstationary signal extraction. *Computational Statistics and Data Analysis* **50**, 2206 – 2231. Also www.census.gov/srd/papers/pdf/rrs2004-05.pdf

[21] Peña, D., Tiao, G., and Tsay, R. (2001). *A Course in Time Series Analysis,* New York: John Wiley and Sons.

Figure 1: Simulated Cycle, Simulated Seasonal-Trend-Irregular, and their sum.

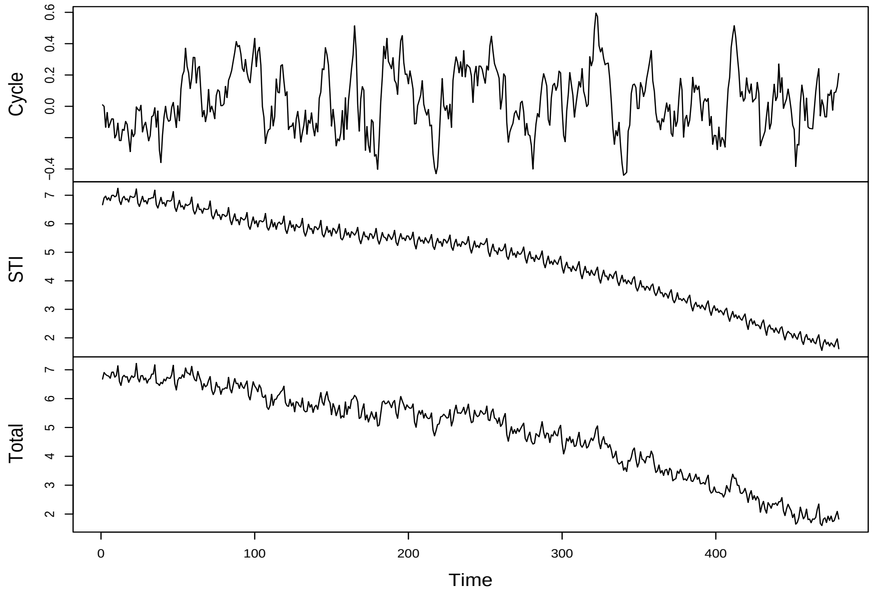

Figure 2: One-stage and SEATS two-stage cycle estimates, together with the true cycle.

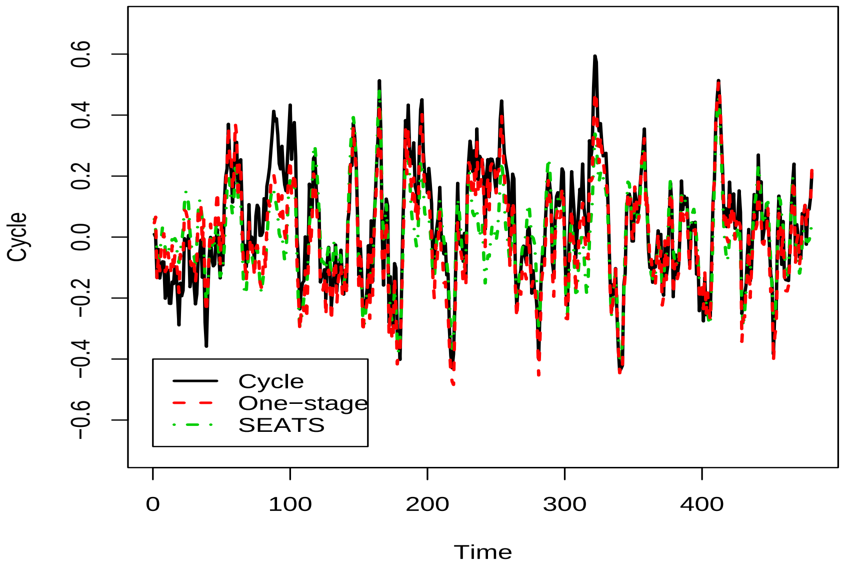

Figure 3: Basic, truncated, and iterated two-stage cycle estimates, together with the true cycle.

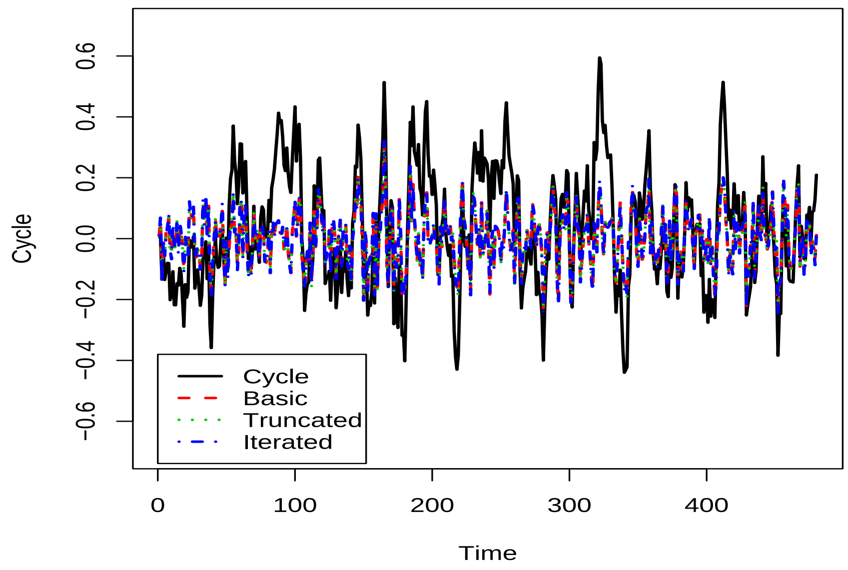

Figure 4: MSE plots for one-stage, SEATS two-stage, basic, truncated, and iterated two-stage methods.