Received: 24 July 2018 | Accepted: 10 August 2018
DOI: 10.1111/stan.12161

SPECIAL ISSUE ARTICLE

WILEY

Seasonal adjustment subject to accounting constraints

Tucker McElroy

Research and Methodology Directorate, U.S. Census Bureau, Washington, DC

Correspondence
Tucker McElroy, Research and Methodology Directorate, U.S. Census Bureau, Washington, DC 20233-9100
Email: tucker.s.mcelroy@census.gov

The indirect seasonal adjustment obtained by aggregating component seasonal adjustments may be inadequate, whereas the direct adjustment of the aggregate can typically be ensured to be adequate by adjusting the statistical model. Reconciliation techniques can be used to allocate the discrepancies between the direct and indirect adjustments of the aggregate unto the various component series, essentially enforcing that the indirect procedure yields the same outcome as the adequate direct procedure. This paper proposes utilizing adequacy of the component seasonal adjustments—given the modifications entailed by reconciliation—as an additional constraint to the accounting problem. We focus on seasonal adjustments arising from X-13ARIMA-SEATS and apply this constrained reconciliation procedure to copper imports, a component of gross domestic product.

KEYWORDS
benchmarking, indirect adjustment, reconciliation

1 | INTRODUCTION

The publication of seasonally adjusted data is an important function of federal statistical agencies. While many time series, such as construction, retail, and import data published by the U.S. Census Bureau, are available in both raw and adjusted formats, other important aggregates, such as gross domestic product (GDP), are only published in an adjusted form. It is crucial that adjustments are adequate, that is, all stable and dynamic seasonality has been suppressed, while causing minimal distortion to other dynamics, such as trend and business cycle. Moreover, for series that are aggregates of other published components, the adjustments should satisfy the same accounting constraints, that is, the aggregation of the component adjustments should equal the adjustment of the raw aggregate. This paper addresses this problem, being motivated by recent concerns about residual seasonality in GDP.

GDP is the most heavily scrutinized economic time series, and beginning in 2015, there were registered concerns about the adequacy of the seasonal adjustment, given that the first quarter appeared to be systematically lower (Furman, 2015; Gilbert, Morin, Paciorek, & Sahm, 2015; Groen & Russo, 2015; Rudebusch, Wilson, & Mahedy, 2015). Public concerns generated action from the Bureau of Economic Analysis (BEA); preliminary findings determined that residual seasonality could arise from the conversion of monthly data to quarterly data, and the more subtle issue that aggregation of nonseasonal data can generate seasonal patterns (Moulton & Cowan 2016). The topic of direct versus indirect adjustment, which has been studied at least as far back as in the work Dagum (1979), was debated; because the direct approach violates accounting constraints (aggregation), further work was focused upon indirect approaches.

Similar challenges have been encountered in other economic databases encountered across the world, as discussed in the works of Hood and Findley (2001) and Astolfi, Ladiray, and Mazzi (2001). Given the pervasiveness of the problem, there have been prior efforts at a solution, some based upon attempts to improve the modeling of individual components. However, this is insufficient to address the core problem (as we demonstrate in Section 2); recognizing this, some authors have advocated multivariate time series analysis to correctly understand linkages in signal content across time series (Birrell, Steel, & Lin, 2011; McElroy, 2017). Others, such as Maravall (2006), have made a forceful case for the direct approach, thereby sacrificing the goal of satisfying accounting constraints. This paper instead adopts a univariate approach based on ideas from the extensive literature on reconciliation.

Benchmarking problems (ensuring that certain temporal aggregation constraints are valid) were addressed by Denton (1971), Cholette (1984), and later authors (summarized in Dagum & Cholette, 2006), whereas methods that ensure aggregation relations hold across different time series have long been available (Stone, Champernowne, & Meade, 1942). Recent works by Di Fonzo and Marini (2011), Quenneville and Fortier (2012), and Chen (2012) have addressed the temporal and contemporaneous facets simultaneously. Hyndman, Ahmed, Athanasopoulos, and Shang (2011) studied the extension of these approaches to forecasting. However, these methods primarily are useful for variables themselves (or their forecasts), not for latent signals of variables, such as seasonal adjustments; in fact, a straightforward application of reconciliation methods can result in a seasonal adjustment that satisfies accounting constraints but has residual seasonality and, therefore, must be rejected. Hence, it is desirable to extend the simplicity of reconciliation methodology to a context where latent signal content is appropriately handled.

This paper addresses this important gap in the reconciliation literature, focusing on the case of accounting constraints. Note that we do not treat the problem of benchmarking, whereby one wishes to ensure that temporal aggregations of a seasonally adjusted time series retain the property of seasonal adjustment adequacy. Such a problem—in passing from monthly to quarterly data—does afflict the components of U.S. GDP, but we aim to address it in future work. Here, we adapt the basic approach discussed in the work of Quenneville and Fortier (2012), namely, to modify good adjustments of component parts as little as possible so as to ensure accounting constraints are satisfied by including the facet of assessing and ensuring the adequacy of seasonal adjustments. We first demonstrate the subtlety of this cross-aggregation phenomenon in Section 2, showing that it cannot be provably resolved simply through larger sample sizes or through more nuanced multivariate modeling.

Our proposed solution (Section 3) is motivated by pragmatic considerations: We do not require additional modeling and human intervention but, instead, utilize seasonal adjustments of the component series and various partially cumulated series; by adding one variable at a time, we can isolate the impact of offending components. Section 4 contains a simulation study that palpably demonstrates the cross-aggregation phenomenon, and the efficacy of the proposed method. In Section 5, we apply this method to real copper imports, a component of GDP, and Section 6 concludes.

2 | THE CROSS-AGGREGATION PHENOMENON

The phenomenon of cross-aggregation refers to the possibility that the aggregation of many series deemed to be nonseasonal (either those for which no seasonality was detected or those that are the outputs of a seasonal adjustment procedure) may exhibit seasonality. As the aggregate can of course be adjusted—called the direct approach—this will not in turn be an aggregate of the component series. However, adopting the aggregate of the component series as the adjustment—called the indirect approach—is also not satisfactory because of its apparent seasonality.

Such phenomena have been empirically observed in economic aggregates, such as U.S. imports data. An economic explanation for the phenomenon can be rendered for component series belonging to the same sector, or region, of measurement for a variable. Perhaps, the aggregate activity over the sector or the region exhibits a seasonal pattern over time, but in each month or quarter, the distribution of that activity over various components may be greater or lesser. A possible illustration (suggested by Ben Cowan of BEA) states that a household or firm may allocate a fixed portion of their budget to a particular sector (e.g., utilities) but allocate the actual funds to different vendors in different months; whereas the overall budget allocation may be subject to seasonal vagaries (energy usage being dictated by climate changes), particular vendors may receive essentially random chunks of that allocation and, hence, do not exhibit seasonal patterns themselves.

This explanation actually suggests a stochastic mechanism for generating the cross-aggregation phenomenon. (We credit Jonathan Wright, who suggested this construction to the authors at the December 9, 2016, FESAC meeting.) We suppose that component time series $\{X_{t,i}\}$ for $1 \leq i \leq n$ and $t \in \mathbb{Z}$ can be written in terms of latent processes $\{S_t\}$ and $\{N_{t,i}\}$, representing unobserved seasonal and nonseasonal dynamics, respectively, as follows:

$$X_{t,i} = S_t + N_{t,i}.$$

Note that the seasonal process is common to all the components. (Actually, this could be substantially relaxed to the assumption that each seasonal $S_{t,i}$ is cross-correlated across various series i, and the same arguments apply, albeit less forcefully.) For simplicity, we suppose that all these time series are stationary, although the case that the nonseasonal processes are $I(1)$ is more realistic for economic data. Assuming that $\{S_t\}$ and $\{N_{t,i}\}$ are uncorrelated with each other, the relationship of autocovariances is

$$\gamma_{X,i}(h) = \gamma_S(h) + \gamma_{N,i}(h)$$

for $h \in \mathbb{Z}$, where each γ_Z denotes the autocovariance function of a corresponding process $\{Z_t\}$. The aggregate process is defined through some weighted sum

$$\underline{X}_{t,n} = \sum_{i=1}^{n} w_i X_{t,i},$$

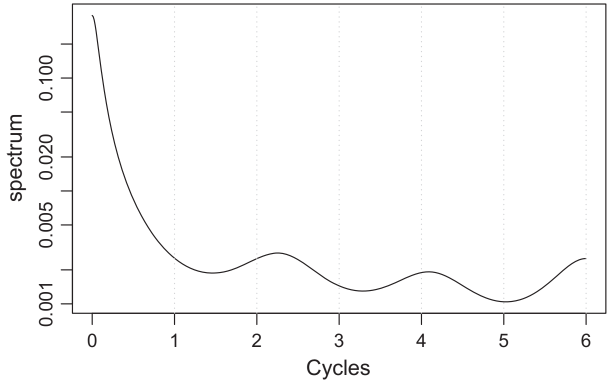

FIGURE 1 Autoregressive spectral estimator of a simulated monthly series with no apparent seasonality. The gray dotted vertical lines indicate the seasonal frequencies

and in order to see the impact of averaging on seasonality, we impose that $\sum_{i=1}^{n} w_i = 1$. Finally, we suppose that all the nonseasonal processes are uncorrelated with one another; hence,

$$\gamma_{\underline{X}_n}(h) = \gamma_S(h) + \sum_{i=1}^{n} w_i^2 \gamma_{N,i}(h).$$

Clearly, the contribution of $\gamma_S(h)$ to $\gamma_{\underline{X}_n}(h)$ is much greater than was its contribution to $\gamma_{X,i}(h)$. If the weights are all equal to $1/n$, then the relative contribution of the seasonality has an n-fold increase. In other words, it will be much more apparent in the aggregate autocovariance function, whereas it may be indiscernible in the case of a single component, if $\gamma_S(h)$ is small relative to $\gamma_{N,i}(h)$.

By simple algebra, we can associate an interpretation to this decomposition as follows: $S_t = \underline{X}_{t,n} - \underline{N}_{t,n}$ and

$$X_{t,i} = \underline{X}_{t,n} - (\underline{N}_{t,n} - N_{t,i}),$$

which expresses a component series in terms of the aggregate minus a correction term. In the example of a budget $\{\underline{X}_{t,n}\}$ allocated to utility usage, the funds allocated to a particular vendor would deviate from the utility budget by $\underline{N}_{t,n} - N_{t,i}$, which essentially cancels out the observable seasonality.

Such processes are easy to simulate. In Figures 1 and 2, we present spectral density estimates from a single such process $\{X_{t,1}\}$ and the aggregate $\{\underline{X}_{t,10}\}$ of ten such processes. Although no peaks are apparent in the first plot at the seasonal frequencies (marked with the numbers 1 through 6 on the x-axis), there are peaks apparent in the second plot at seasonal frequencies 1, 2, 3, 4, and 5. This example demonstrates the cross-aggregation phenomenon.

Standard diagnostics tests for seasonality—such as those based on autocovariances, the spectral density, or seasonal regressors—do not adequately detect seasonality in the component series when the variability in the noise $N_{t,i}$ is sufficiently strong to bury the signal S_t. This is more than an issue of sample size; by suitably increasing $\gamma_{N,i}(h)$, we can guarantee that there is no seasonality present in each $X_{t,i}$.

To see why this is true, suppose that we declare that a time series is seasonal whenever its autocorrelation function at lag p, the seasonal period (12 for monthly and 4 for quarterly), exceeds a threshold κ. We can set κ close to one, in order to demand high persistence year to year to qualify as seasonality. (Such a diagnostic forms the basis for the commonly used QS statistics described

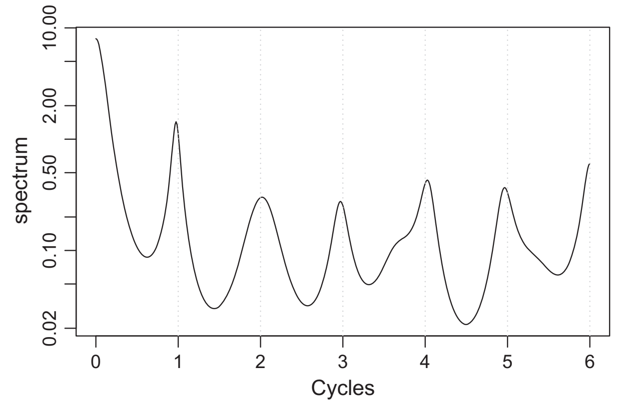

FIGURE 2 Autoregressive spectral estimator of the aggregate of 20 simulated monthly series, each with no apparent seasonality. The gray dotted vertical lines indicate the seasonal frequencies

in the work of Maravall (2012).) Supposing that $N_{t,i}$ is white noise of variance σ^2, the seasonal autocorrelation of the component process is

$$\frac{\gamma_{X,i}(p)}{\gamma_{X,i}(0)} = \frac{\gamma_S(p)}{\gamma_S(0) + \sigma^2} = \frac{\rho_S(p)}{1 + \sigma^2/\gamma_S(0)}.$$

As the signal-to-noise ratio $\gamma_S(0)/\sigma^2$ decreases, the seasonality vanishes. Even a weak seasonality, where κ is arbitrarily close to zero, will fail to result in a declaration of seasonality so long as the signal-to-noise ratio is less than $\kappa/(\rho_S(p) - \kappa)$. Because this analysis pertains to the stochastic process, quite apart from issues of statistical uncertainty arising from finite samples of time series, it is clear that the cross-aggregation phenomenon cannot be resolved merely with longer samples.

The lag-p cross-correlation of two component series has the same sort of expression as the seasonal autocorrelation, which indicates that multivariate analyses will also fail to detect the seasonality. Given this predicament, this paper makes no attempt to capture such hidden seasonality; instead, we propose to modify seasonally adjusted aggregates—where seasonality has been detected—by a reconciliation procedure that balances the direct and indirect approaches.

3 | RECONCILIATION METHODOLOGY

3.1 | Framework and background

Following the treatment of Quenneville and Fortier (2012), we wish to solve a reconciliation problem with the following structure. We say a seasonal adjustment is *adequate* if, according to a seasonality diagnostic, there is no seasonality. There are component series $\{X_{t,i}\}$ for $1 \leq i \leq n$ that aggregate to a composite $\{\underline{X}_{t,n}\}$ via some operators, such as addition and multiplication. We can individually adjust each component series, apply the same aggregation, and hope that the resulting indirect adjustment of $\{\underline{X}_{t,n}\}$ is adequate; if not, we may prefer a direct adjustment of $\{\underline{X}_{t,n}\}$, although the aggregation relationship will surely no longer hold. In such a case, we wish to modify the component adjustments so that they aggregate to the direct adjustment, while still being adequate themselves. To parse this problem with notation, let the aggregation be denoted

$$\underline{X}_{t,n} = \oplus_{i=1}^{n} X_{t,i},$$

which holds true for each time t. We use the symbol $\oplus$ to denote a general form of aggregation; in many applications, this is a straight summation, but might also be a weighted sum, or could even involve multiplication or division. We presume that the associative property holds for $\oplus$. This definition also holds for partial aggregates $\underline{X}_{t,i}$ for $1 \leq i \leq n$, constituting an aggregation of only the first i component series. The component adjustments are denoted

$$\widehat{N}_{t,i} \qquad \underline{\widehat{N}}_{t,i} \tag{1}$$

for $1 \leq i \leq n$, denoting the adjustments of component series and partial aggregates respectively. Each adjustment is purely a function of the corresponding $\{X_{t,i}\}$ or $\{\underline{X}_{t,i}\}$, generated by some seasonal adjustment procedure. This procedure presumes there exists some algebraic operator $\cdot$ that defines the seasonal $S_{t,i}$ such that

$$X_{t,i} = S_{t,i} \cdot N_{t,i}$$

for $1 \leq i \leq n$. This operator could be addition or multiplication, or something more exotic; it may vary depending on i. We will assume that the adjustments (1) are all adequate; either the component series are deemed nonseasonal according to the seasonality diagnostic—in which case $\widehat{N}_{t,i} = X_{t,i}$—or they require adjustment and this has been performed satisfactorily. (Any rejection of adequacy at this stage would require tuning the procedure, perhaps by adjusting the underlying models or parameters, and ultimately by such modifications, we presume that an adequate adjustment is eventually obtained.) Given this notation, the direct adjustment is

$$\underline{\widehat{N}}_{t,n},$$

obtained by applying the seasonal adjustment procedure to the full aggregate $\{\underline{X}_{t,n}\}$. The indirect adjustment is

$$\oplus_{i=1}^{n} \widehat{N}_{t,i},$$

obtained by applying the accounting constraints directly to the component adjustments of each $\{X_{t,i}\}$. Neither the direct nor indirect solutions are satisfactory, as the former violates the accounting constraints and the latter may not be adequate. We propose to determine modifications of the basic adjustments (1), for example, $\widetilde{N}_{t,i}$ and $\underline{\widetilde{N}}_{t,i}$, that have the properties of *fidelity*, *accountability*, and *adequacy*.

- **Fidelity**: $\widetilde{N}_{t,i}$ is close to $\widehat{N}_{t,i}$, and $\underline{\widetilde{N}}_{t,i}$ is close to $\underline{\widehat{N}}_{t,i}$.
- **Accountability**: $\underline{\widetilde{N}}_{t,i} = \oplus_{j=1}^{i} \widetilde{N}_{t,i}$.
- **Adequacy**: $\widetilde{N}_{t,i}$ and $\underline{\widetilde{N}}_{t,i}$ are nonseasonal for $1 \leq i \leq n$.

As to *fidelity*, we know that $\widehat{N}_{t,i}$ is an adequate seasonal adjustment of $X_{t,i}$, so we would like our modification $\widetilde{N}_{t,i}$ to involve as little change as possible. As to *accountability*, this is the aggregation constraint that the final batch of series ($1 \leq i \leq n$) should satisfy; we are building up to that final constraint (given by $i = n$) and seek to impose this subconstraint each step of the way. As to *adequacy*, we need the components $\widetilde{N}_{t,i}$ to be nonseasonal, and imposing that the subaggregates $\underline{\widetilde{N}}_{t,i}$ to be nonseasonal as well will assist us to ensure that the main aggregate is adequate.

To see why it is important to consider all partial aggregates and proceed recursively, consider the approach discussed in the work of Quenneville and Fortier (2012); assuming that all time series are positive (if series are negative, the criterion can be altered in the manner described in Section 5), they propose the relative quadratic loss criterion to minimize deviations from the original adequate component adjustments:

$$\mathcal{L}(y_{t,1}, \ldots, y_{t,n}) = \sum_{i=1}^{n} \left(\widehat{N}_{t,i} - y_{t,i} \right)^2 / \widehat{N}_{t,i}.$$

We seek to minimize this function with each $\{y_{t,i}\}$ belonging to the space of adequate adjustments subject to the aggregation constraint

$$\oplus_{i=1}^{n} y_{t,i} = \underline{\widehat{N}}_{t,n},$$

that is, using the adequate direct adjustment of the aggregate as a benchmark. If the accounting constraint involves straight aggregation (i.e., $\oplus$ is $+$), then the constrained optimization problem can be solved using Lagrangian multipliers and the solution (Quenneville & Fortier, 2012) is

$$\begin{aligned} \widetilde{N}_{t,i} &= \frac{\underline{\widehat{N}}_{t,n}}{\sum_{j=1}^{n} \widehat{N}_{t,j}} \widehat{N}_{t,i} \quad 1 \leq i \leq n \\ \widetilde{N}_{t,n} &= \underline{\widehat{N}}_{t,n}. \end{aligned}$$

This solution rescales each component adjustment by a ratio of direct to indirect adjustment of the aggregate. The potential difficulties with such a solution are clear: If the indirect adjustment has seasonality, this will still be present in the ratio and, hence, will be introduced into $\widetilde{N}_{t,i}$ when the ratio multiplies the nonseasonal $\widehat{N}_{t,i}$.

3.2 | A new approach

Although the above approach may fail, it can be modified to be effective by proceeding recursively. The recursive nature is the following: Supposing that fidelity, accountability, and adequacy hold up to some i, we then seek conditions under which it holds for the case $i + 1$ as well. The compatibility condition then becomes (for additive aggregations)

$$\underline{y}_{t,i+1} = \underline{\widetilde{N}}_{t,i} + y_{t,i+1}, \tag{2}$$

where the equality follows from the recursive principle. The fidelity attribute is imposed by seeking minimizers $y_{t,i+1}$ and $\underline{y}_{t,i+1}$ to

$$\mathcal{L}_{i+1}\left(y_{t,i+1}, \underline{y}_{t,i+1}\right) = \alpha_{i+1}\left(\widehat{N}_{t,i+1} - y_{t,i+1}\right)^2 / \widehat{N}_{t,i+1} + \beta_{i+1}\left(\underline{\widehat{N}}_{t,i+1} - \underline{y}_{t,i+1}\right)^2 / \underline{\widehat{N}}_{t,i+1}.$$

Fidelity is built directly in through relative squared loss. The nonnegative parameters $\alpha_{i+1}, \beta_{i+1}$ allow for unequal weighting of the two portions and grant additional flexibility in the fidelity criterion. Finally, the adequacy attribute is checked for the minimizers. Note that, if we had adequacy at iteration i but then it fails at iteration $i + 1$, then we have isolated the problem to the inclusion of the new series $X_{t,i+1}$, and we can focus our attention there.

Adequacy is checked by assembling solutions over all times t and computing a diagnostic δ that is compared to a threshold τ. In the case of Maravall's (2012) QS diagnostic, δ is computed from the sample autocovariances of the time series, and τ is one minus a p value corresponding to an asymptotic distribution—low p values indicate inadequacy. At the ith stage, we seek time series $\{y_{t,i+1}\}$ and $\{\underline{y}_{t,i+1}\}$ such that

(i) $\mathcal{L}_{i+1}(y_{t,i+1}, \underline{y}_{t,i+1})$ is minimized, for each t;
(ii) Equation (2) holds for each t; and
(iii) $\delta(\{y_{t,i+1}\}) \geq \tau$ and $\delta(\{\underline{y}_{t,i+1}\}) \geq \tau$.

This problem can be solved by using (2) to re-express $\mathcal{L}_{i+1}$ in terms of $y_{t,i+1}$ alone and solving

$$\min_{\{y_{t,i+1}\}} \sum_t \mathcal{L}_{i+1}(y_{t,i+1}, \underline{\widetilde{N}}_{t,i} + y_{t,i+1}) \quad \text{such that} \quad \delta(\{y_{t,i+1}\}) \geq \tau,\ \delta(\{\underline{y}_{t,i+1}\}) \geq \tau. \tag{3}$$

A solution is sought to (3) by searching over all possible time series $\{y_{t,i+1}\}$ and, in principle, can be obtained via Lagrangian techniques with inequality constraints (cf. Kuhn & Tucker, 1951). That a solution exists is obvious, as $y_{t,i+1} \equiv 0$ for all t is nonseasonal, also yielding $\underline{y}_{t,i+1} = \underline{\widetilde{N}}_{t,i}$, which is likewise nonseasonal; such a solution, however, will likely produce a high value in $\mathcal{L}_{i+1}$. While we have experimented with such a nonlinear optimization technique, the dimension is equal to the length of the time series, and hence, such an approach is infeasible for production.

Instead, we advocate seeking an analytical solution to (i) and (ii), followed by checking that (iii) holds. To further facilitate finding a solution rapidly, we allow α_i and β_i to vary at each stage. Rewriting the compatibility condition as a functional constraint and minimizing $\mathcal{L}$ subject to this constraint, the method of Lagrange multipliers yields the solution

$$\widetilde{N}_{t,i+1} = \frac{\widehat{N}_{t,i+1}\left(\underline{\widehat{N}}_{t,i+1} - \gamma_{i+1}\underline{\widetilde{N}}_{t,i}\right)}{\gamma_{i+1}\widehat{N}_{t,i+1} + (1-\gamma_{i+1})\underline{\widehat{N}}_{t,i+1}} \tag{4}$$

$$\underline{\widetilde{N}}_{t,i+1} = \frac{\underline{\widehat{N}}_{t,i+1}\left(\widehat{N}_{t,i+1} + (1-\gamma_{i+1})\underline{\widetilde{N}}_{t,i}\right)}{\gamma_{i+1}\widehat{N}_{t,i+1} + (1-\gamma_{i+1})\underline{\widehat{N}}_{t,i+1}}, \tag{5}$$

where $\gamma_{i+1} = \beta_{i+1}/(\beta_{i+1} + \alpha_{i+1})$. Clearly these solutions satisfy the accountability property, but they may or may not be adequate. When $\gamma_{i+1} = 0$ (or $\beta_{i+1} = 0$), we have the solution

$$\widetilde{N}_{t,i+1} = \widehat{N}_{t,i+1} \qquad \underline{\widetilde{N}}_{t,i+1} = \widehat{N}_{t,i+1} + \underline{\widetilde{N}}_{t,i},$$

which is indirect adjustment, that is, do not alter the original component adjustment, and define the aggregate adjustment by straight summation. On the other extreme, when $\gamma_{i+1} = 1$ (or $\alpha_{i+1} = 0$), we obtain the solution

$$\widetilde{N}_{t,i+1} = \underline{\widehat{N}}_{t,i+1} - \underline{\widetilde{N}}_{t,i} \qquad \underline{\widetilde{N}}_{t,i+1} = \widehat{N}_{t,i+1},$$

which imposes a direct adjustment on the aggregate, solving for the component adjustment by subtraction. This has the demerit of potentially yielding negative values.

Therefore, with $\gamma_{i+1} = 0$, the component adjustment is adequate, but the aggregate is determined indirectly and, hence, may be inadequate. Conversely, with $\gamma_{i+1} = 1$, we obtain a direct adjustment of the aggregate, but the component adjustment is determined by a subtraction and may be inadequate. We seek some $\gamma_{i+1} \in [0, 1]$ such that both the component adjustment and the aggregate adjustment are adequate; we propose to search for such a γ_{i+1}, beginning with the indirect adjustment ($\gamma_{i+1} = 0$). In this formulation, there is no optimal choice of γ_{i+1}, but we favor lower values *a priori* (because in the base case that indirect adjustment is adequate, we should obtain $\gamma_{i+1} = 0$ if we were to spuriously run the procedure) and, hence, seek the smallest value of γ_{i+1} such that (i), (ii), and (iii) are satisfied. Our proposed algorithm is as follows:

1. Determine $\widehat{N}_{t,i}$ and $\underline{\widehat{N}}_{t,i}$ for all $1 \leq i \leq n$. (These are adequate.)
2. Set $\widetilde{N}_{t,1} = \underline{\widetilde{N}}_{t,i}$ equal to $\hat{N}_{t,1}$, and let $i = 1$.
3. Recursively determine $\widetilde{N}_{t,i+1}$ and $\underline{\widetilde{N}}_{t,i+1}$ from $\widetilde{N}_{t,j}$ and $\underline{\widetilde{N}}_{t,j}$ for $1 \leq j \leq i$ utilizing the Lagrange solution (4) and (5), based on some choice of γ_{i+1}.
4. Check for adequacy of component adjustment and partial aggregate adjustment.
5. If inadequate, return to step 3 and use a different choice of γ_{i+1}; else, increment i and go to step 3.
6. Terminate when $i = n$.

Note that series are prescreened such that if the indirect adjustment is adequate, no further action is needed. (Equivalently, we obtain $\gamma_{i+1} = 0$ immediately.) When we reach $i = n$, step 4 has already checked our final aggregate for adequacy; if this diagnostic is satisfactory, we are finished.

4 | SIMULATION EXPERIMENTS

Our implementation of the reconciliation algorithm uses the X-11 (Findley, Monsell, Bell, Otto, & Chen, 1998) option of the X-13ARIMA-SEATS software program to adjust each input series, using the QS diagnostic (Maravall, 2012) applied to the trend-differenced seasonal adjustment. In particular, we compute the original adjustments using X-11 by first pretesting each component series or partial aggregate with QS and only proceed to adjustment if there is a detection of seasonality. A posttest is also applied to ensure our assumption that the initial adjustments are adequate. (We also investigated the performance using different seasonal diagnostics, with similar results being obtained.) Then, we proceed to steps 2 through 6 of the algorithm in Section 3, in step 5 gradually incrementing γ_{i+1} from an initial value of zero by steps of size $1/n$ to a final value of one.

We comment briefly on the QS criterion. It can be applied to the irregular or the trend-differenced seasonally adjusted component, and in either case, outlier regression effects (e.g., additive outliers and level shifts) should be first removed; in our applications, we only apply QS to the seasonally adjusted component. QS is defined as

$$\frac{\max\{0, \widehat{\rho}(p)\}^2}{T-p} + \frac{\max\{0, \widehat{\rho}(2p)\}^2}{T-2p},$$

for sample size T, and the null hypothesis is that $\rho(p) \leq 0$ and $\rho(2p) \leq 0$ (see Findley, Lytras, & McElroy, 2017). Because the diagnostic is based upon autocorrelations alone, this criterion is not able to distinguish between genuinely seasonal processes and those which are nonseasonal and yet happen to have strong seasonal correlation arising through tight linkages between the intervening months. For instance, a seasonal autoregression of order one with parameter ϕ_p will have $\rho(p) = \phi_p$, whereas a regular AR(1) process of parameter ϕ_1 will have $\rho(p) = \phi_1^p$. Clearly, the seasonal autocorrelation $\rho(p)$ is identical for these two processes whenever $\phi_1^p = \phi_p$. While the null hypothesis implies that no seasonality is present, the converse need not be true.

For example, with $\phi = 0.98$ and $p = 4$, we obtain $\rho(4) = 0.92$. To see the impact in simulation, consider 10^5 repetitions of a 20-year quarterly Gaussian AR(1) process with $\phi = 0.98$. The empirical type-I error rate is 0.975 based on the nominal of 0.05, that is, a 97.5% chance of falsely indicating seasonality. Repeating the study with $\phi = 0.1$, the type-I error rate drops to

TABLE 1 Values of the QS statistic p value for $n = 20$ components and their partial aggregates

i	$X_{t,i}$	$\underline{X}_{t,i}$	$\widehat{N}_{t,i}$	$\underline{\widehat{N}}_{t,i}$	$\widetilde{N}_{t,i}$	$\underline{\widetilde{N}}_{t,i}$
1	0.009	0.009	1.000	1.000	1.000	1.000
2	0.635	0.000	0.635	1.000	0.635	0.890
3	0.293	0.000	0.293	1.000	0.293	0.899
4	0.052	0.001	0.052	1.000	0.052	0.227
5	0.021	0.000	0.021	1.000	0.341	0.015
6	0.226	0.000	0.226	0.930	1.000	0.016
7	0.004	0.000	1.000	1.000	1.000	0.365
8	0.133	0.000	0.133	1.000	0.133	0.040
9	0.647	0.000	0.647	1.000	0.677	0.010
10	0.067	0.000	0.067	1.000	0.927	0.022
11	0.489	0.000	0.489	1.000	1.000	0.017
12	0.028	0.000	0.028	1.000	1.000	0.012
13	0.009	0.000	1.000	1.000	1.000	0.035
14	0.085	0.000	0.085	1.000	0.782	0.010
15	0.009	0.000	0.898	1.000	0.898	0.035
16	0.502	0.000	0.502	1.000	0.502	0.011
17	0.159	0.000	0.159	1.000	1.000	0.015
18	0.002	0.000	1.000	1.000	1.000	0.027
19	0.149	0.000	0.149	1.000	0.611	0.010
20	0.016	0.000	0.016	1.000	0.908	0.021

Note. Columns 2 and 3 pertain to raw data, columns 4 and 5 pertain to the initial adjustments, and columns 6 and 7 pertain to the reconciled adjustments.

0.015. However, for the null that $\rho(p) \leq 0$ and $\rho(2p) \leq 0$, the test is approximately correctly sized. Therefore, we interpret the QS results as follows: Low p values provide evidence that there is high seasonal autocorrelation (though this does not necessarily entail the presence of seasonality), whereas high p values indicate the absence of positive seasonal autocorrelation, and hence, the series cannot be seasonal. In our simulations, we set the threshold of adequacy at $\tau = 0.01$.

For a simulation experiment, we follow the broad prescription described in Section 2: A single seasonal pattern is randomly generated, to which independent simulations of white noise and random walk are added. The variability in the random walk, which represents a trend effect, is chosen so that it does not dominate the seasonality; the white noise variance is set sufficiently high so that seasonality will rarely be detected by QS for any component series. All of these simulations are exponentiated because the reconciliation method presumes positive input series. We then aggregate up $n = 20$ such simulations, expecting that the seasonal behavior will become more apparent as the degree of partial aggregation increases. Table 1 records the QS p values in columns 2 and 3: Observe that, as most of the component series are nonseasonal (excepting series 1, 7, 13, 15, and 18), all of the partial aggregates ($i \geq 2$) indicate strong seasonality according to the QS criterion. (Although QS can wrongly flag trend-persistent series as seasonal, as mentioned above, independent confirmation by other seasonal diagnostics confirms the general pattern evident in Table 1.) Figures 1 and 2 tell the same story: The component series are nonseasonal, but the aggregates display extremely salient seasonality.

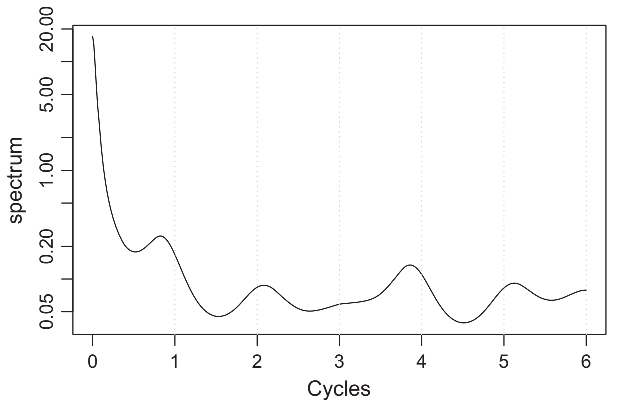

FIGURE 3 Autoregressive spectral estimator of the reconciled aggregate of 20 simulated monthly series, each with no apparent seasonality. The gray dotted vertical lines indicate the seasonal frequencies

From these simulations, the initial adjustments are obtained, to which the QS diagnostic is applied, with QS p values given in columns 4 and 5 of Table 1. All of these are adequate; note that partial aggregates for $2 \leq i \leq 20$ required adjustment. The final step is the reconciliation, which proceeds through 20 iterations: The values of γ ranged from zero in several cases (the indirect adjustment) to as high as 0.75, which comes fairly close to direct adjustment. The reconciled series have QS p values provided in columns 6 and 7 of Table 1 and display adequacy—although for partial aggregates 9, 14, and 19, the degree of seasonality present is the maximal amount permissible. We can visually assess the adequacy of the final aggregate ($i = 20$) by the spectral density plot of Figure 3: The seasonal peaks are no longer evident, in comparison with Figure 2. (There was no visual discrepancy between the initial and reconciled adjustments, so these are not displayed.)

5 | NOMINAL AND REAL NIPA COPPER IMPORTS

While a complete analysis of the thousands of component series constituting GDP is beyond the scope of this article, we here provide an analysis of one constituent part and the solution given by the reconciliation methodology of Section 3. Earlier work by economists at BEA, documented in the work of Moulton and Cowan (2016), described various problematic series, and we here focus on the National Income and Product Account (NIPA) copper imports data, which is available for quarters from 2006 through the end of 2015. (This is admittedly easier than reconciling monthly series.) As the nominal estimates are based upon monthly trade data published by the U.S. Census Bureau, the corresponding *real* series is obtained by deflating via a price series available from Industrial Production Index data (available from Bureau of Labor Statistics); see Figure 4.

Although there is no evidence of seasonality in the nominal or price data, there is evidence of weak seasonality in the real copper imports. The p values for QS are 0.715 and 0.700 for nominal copper imports and price but 0.143 for real copper imports. (These results are based off first differences of the logged data, although arguably these are not unit root processes.) The spectral density estimates (not shown) broadly support these results, as the autoregressive spectral estimator for real copper has a global maximum at frequency π (a seasonal frequency)—although we cannot be sure whether this is a real peak, as it occurs on the boundary of $[-\pi, \pi]$. The weak seasonality indicated by the 0.143 p value seems to be driven by a somewhat substantial lag-8 sample

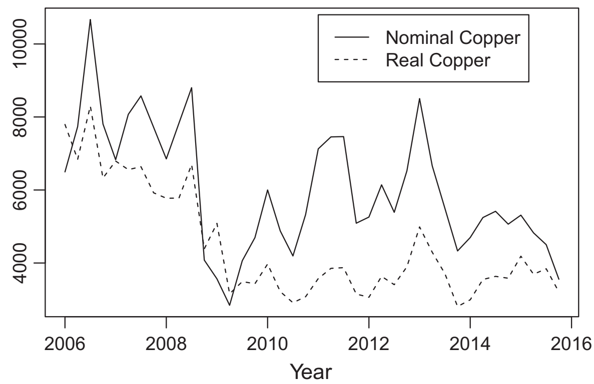

FIGURE 4 Nominal and real copper imports (2006.Q1 through 2016.Q4)

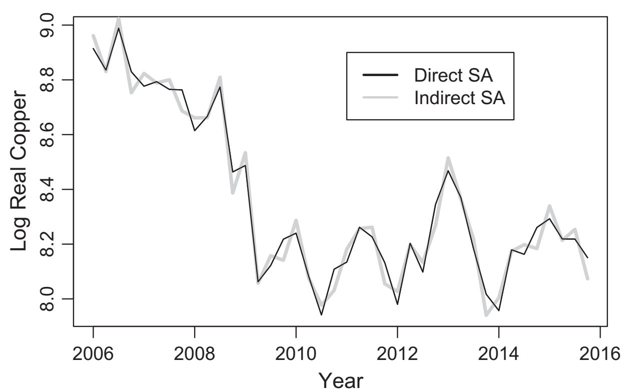

FIGURE 5 Direct and indirect seasonal adjustments (SAs) of real copper imports (2006.Q1 through 2016.Q4)

autocorrelation in log difference real copper. Therefore, to corroborate the presence of weak seasonality, we fit a quarterly SARIMA(0,1,0)(0,0,2) model to log real copper and find that the first and second seasonal moving-average coefficients are 0.037 (not significant) and 0.249 (standard error of 0.150). (This model was superior, according to AIC, to several other SARIMA contenders.) As the second seasonal moving-average coefficient generates a lag-8 effect in the autocorrelation and is weakly significant (*p* value for upper one-sided test is 0.048), we may be justified in having concerns about the presence of seasonality.

Assuming that one wants to take action in such a scenario, without a reconciliation methodology, it is unclear how to proceed; the nominal and price components are deemed to be nonseasonal so that the indirect adjustment is identical to real copper imports. On the other hand, the direct adjustment is adequate (obtained using automatic settings with a log transformation) with a QS *p* value of 0.961. Unfortunately, it differs slightly from the indirect adjustment, as shown in Figure 5.

This minor discrepancy indicates that a slight reconciliation can achieve fidelity, accountability, and adequacy. With regard to the methodology of Section 3, we take log nominal copper and the negative of log price (divided by 100) as the input series, which are enforced to sum up to log real copper. Because we are in log scale for the aggregation relationships, a slight modification to the methodology is needed, as the input series are no longer positive. Hence, we will instead use the objective function

$$\mathcal{L}_{i+1}\left(y_{t,i+1}, \underline{y}_{t,i+1}\right) = \alpha_{i+1}\left(\widehat{N}_{t,i+1} - y_{t,i+1}\right)^2 + \beta_{i+1}\left(\widehat{\underline{N}}_{t,i+1} - \underline{y}_{t,i+1}\right)^2.$$

Then, imposing (2), the method of Lagrange multipliers yields the solution

$$\widetilde{N}_{t,i+1} = (1-\gamma_{i+1})\widehat{N}_{t,i+1} + \gamma_{i+1}\left(\widehat{\underline{N}}_{t,i+1} - \widetilde{\underline{N}}_{t,i}\right)$$
$$\widetilde{\underline{N}}_{t,i+1} = (1-\gamma_{i+1})\left(\widehat{N}_{t,i+1} + \widetilde{\underline{N}}_{t,i}\right) + \gamma_{i+1}\widehat{\underline{N}}_{t,i+1},$$

which replace (4) and (5) in the algorithm of Section 3. For our results, we set the QS threshold to a p value of $\tau = .50$, which will screen out even a low degree of seasonality. The input series for nominal and price are both deemed nonseasonal according to QS, and hence, they are not adjusted. (Note that running nonseasonal series through X-13ARIMA-SEATS could result in small changes.) Running the method, a value of $\gamma_2 = 0.32$ is selected, indicating that the reconciled aggregate is more similar to the indirect adjustment than the direct adjustment. As a byproduct, modifications to the adjustments for nominal and price are calculated, although the changes are small. The resulting QS p values for first differences of log nominal ($\widetilde{N}_1$), price ($\widetilde{N}_2$), and real ($\widetilde{\underline{N}}_2$) are 0.715, 0.619, and 0.503. There was no change to nominal copper, but the price

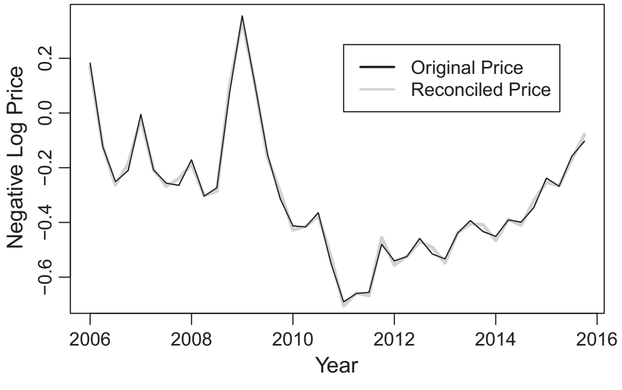

FIGURE 6 Price series with and without reconciliation (2006.Q1 through 2016.Q4)

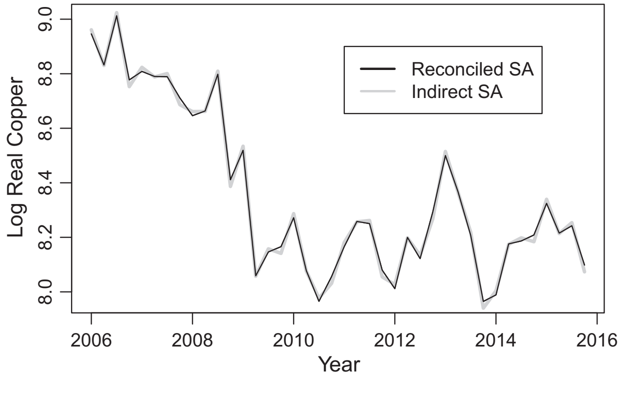

FIGURE 7 Reconciled and indirect seasonal adjustments of real copper imports (2006.Q1 through 2016.Q4)

was adjusted slightly (Figure 6). A comparison of the indirect adjustment for real copper imports (which was identical to real copper imports) to the reconciled adjustment is given in Figure 7.

Further experimentation reveals that altering the QS threshold has an impact on γ_2, as expected; increasing the QS critical value demands a lower degree of seasonality, with more recourse to the direct adjustment through a higher value of γ_2. A second observation is that nominal copper suffers no modifications merely because it was listed first (index $i = 1$) in our method, but reversing the order of the input series would instead place reconciliation on nominal copper instead of price. Either way, the modifications are quite small in this example.

6 | CONCLUSION

The problem of direct versus indirect seasonal adjustment has existed for many decades, and nuanced approaches involving multivariate modeling have been advanced; these have the limitation, thus far, of relying on linear accounting constraints (that is, weighted sums). The extra resources required for multivariate modeling may yield dividends in terms of more nuanced understanding of joint dynamics and a better quantification of uncertainty. Because cross-series dependence is measured by multivariate models, in principle, the cross-aggregation phenomenon can be assessed and accounted for, as argued in the work of Findley et al. (2017). Nevertheless, in practice, the success of such an approach is contingent upon the ability of multivariate time series models to capture such cross-sectional patterns.

This paper instead aims to satisfy accounting constraints while maintaining fidelity to adjustments deemed optimal when a time series is analyzed in isolation and avoids extensive joint modeling. Such an approach, which is motivated by the reconciliation literature, can be adopted under a wider array of accounting constraints—one only needs to modify the constraint function in the Lagrangian method. A key challenge in adopting reconciliation techniques for seasonal adjustment is ensuring the adequacy of all outputs. This was not addressed in previous literature (Denton, 1971); for forecasting reconciliation (Hyndman et al., 2011), it is not important. This paper sets forth a novel solution to this problem, by incorporating a quality diagnostic as an inequality constraint in the optimization.

To our knowledge, this is the first article to demonstrate through theory and simulation that the cross-aggregation phenomenon (as suggested by Jonathan Wright) is fundamental, that is, it is a phenomenon associated with the data process and not merely the inadequacies of our models or our information sets. Our motivation stems from a composite indicator consisting of series corresponding to subclassifications, although extensions to spatial aggregation (e.g., regional GDP summing to national GDP) can be envisioned as another application. In such a scenario, we forgo the attempt to model spatial correlation patterns in regional latent seasonal processes and directly apply the reconciliation methodology. While we have stated a preference here for indirect adjustments (lower values of the γ weights orient our solution in this direction) when possible, this could be altered in the algorithm by adopting other values of γ.

We have found—in other case studies not reported here—that many apparent instances of the cross-aggregation phenomenon can indeed be resolved through either careful modeling or appropriate interpretation of diagnostics. Therefore, one needs to adopt a proper measure of skepticism toward claims of such phenomena being detected in production databases. Given this proviso, the example of real copper imports gives support to the belief that the cross-aggregation phenomenon exists empirically, and not just in simulation. For such instances, the methods of this paper will be useful.

DISCLAIMER

This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the author and not necessarily those of the U.S. Census Bureau.

ORCID

Tucker McElroy http://orcid.org/0000-0002-2991-9067

REFERENCES

Astolfi, R., Ladiray, D., & Mazzi, G. L. (2001). *Seasonal adjustment of European aggregates: Direct versus indirect approach* (Working Paper). Luxembourg: Eurostat.

Birrell, C. L., Steel, D. G., & Lin, Y.-X. (2011). Seasonal adjustment of an aggregate series using univariate and multivariate basic structural models. *Journal of Statistical Theory and Practice*, *5*, 179–205.

Chen, B. (2012). A balanced system of U.S. industry accounts and distribution of the aggregate statistical discrepancy by industry. *Journal of Business and Economic Statistics*, *30*, 202–221.

Cholette, P.-A. (1984). Adjusting sub-annual series to yearly benchmarks. *Survey Methodology*, *10*, 35–49.

Dagum, E. B. (1979). On the seasonal adjustment of economic time series aggregates: A case study of the unemployment rate. Counting the labor force. National Commission on Employment and Unemployment Statistics, Appendix, 2, pp. 317–339.

Dagum, E. B., & Cholette, P. A. (2006). *Benchmarking, temporal distribution and reconciliation methods for time series*. New York, NY: Springer-Verlag.

Denton, F. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of the American Statistical Association*, *66*, 99–102.

Di Fonzo, T., & Marini, M. (2011). Simultaneous and two-step reconciliation of systems of time series: Methodological and practical issues. *Journal of the Royal Statistical Society C*, *60*, 143–164.

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., & Chen, B.-C. (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *Journal of Business and Economics Statistics*, *16*, 127–152.

Findley, D. F., Lytras, D., & McElroy, T. S. (2017). *Detecting seasonality in seasonally adjusted monthly times series* (Research Report 2017-03). Washington, DC: U.S. Census Bureau.

Furman, J. (2015, May 29). Second estimate of GDP for the first quarter of 2015 [Council of Economic Advisers Blog]. https://obamawhitehouse.archives.gov/blog/2015/05/29/second-estimate-gdp-first-quarter-2015

Gilbert, C. E., Morin, N. J., Paciorek, A. D., & Sahm, C. R. (2015). Residual seasonality in GDP. FEDS Notes.

Groen, J., & Russo, P. (2015). The myth of first-quarter residual seasonality. Liberty Street Economics.

Hood, C. C., & Findley, D. F. (2001). Comparing direct and indirect seasonal adjustments of aggregate series. In *American Statistical Association Proceedings of the Business and Economics Statistics Section*.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis*, *55*, 2579–2589.

Kuhn, H. W., & Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability*, pp. 481–492.

Maravall, A. (2006). An application of the TRAMO-SEATS automatic procedure; direct versus indirect adjustment. *Computational Statistics and Data Analysis*, *50*, 2167–2190.

Maravall, A. (2012). Update of seasonality tests and automatic model identification in TRAMO-SEATS. Bank of Spain.

McElroy, T. (2017). Multivariate seasonal adjustment, economic identities, and seasonal taxonomy. *Journal of Business and Economics Statistics*, *35*, 611–625.

Moulton, B. R., & Cowan, B. D. (2016). Residual Seasonality in GDP: Findings and Next Steps. *Survey of Current Business 96 (July 2016)*.

Quenneville, B., & Fortier, S. (2012). Restoring accounting constraints in time series methods and software for a statistical agency. In *Economic time series: Modelling and seasonality* (pp. 231–253). Boca Raton, FL: CRC Press.

Rudebusch, G. D., Wilson, D., & Mahedy, T. (2015). The puzzle of weak first-quarter GDP growth. *FRBSF Economic Letter*, *16*.

Stone, R., Champernowne, D. G., & Meade, J. E. (1942). The precision of national income estimates. *Review of Economic Studies*, *9*, 111–125.

How to cite this article: McElroy T. Seasonal adjustment subject to accounting constraints. *Statistica Neerlandica*. 2018;1–16. https://doi.org/10.1111/stan.12161