

Measurement Error:  
A Primer for Macroeconomists

Simon van Norden\*

September 11, 2019

**Keywords:** Measurement Error; Data Revision; National Accounts; Seasonal Adjustment, Business Cycles

---

\*The author thanks Dean Croushore and Jan P.A.M. Jacobs for their insights and collaboration over the years.

## Summary

This chapter is intended as an introduction to and overview of measurement error and revision in macroeconomic data. It will be useful for most applied researchers in macroeconomics who work with official macroeconomic statistics, such as those found in the National Accounts, the Balance of Payments, national government budgets, labour force statistics, etc. Many of the lessons discussed below also apply much more broadly whenever researchers treat data as immutable rather than subject to estimation error and revision.

The first section below reviews sources of measurement error in macroeconomics. On the conceptual side, aggregate measures such as GDP are acknowledged to be poor measures of aggregate welfare as they omit leisure and nonmarket production activity and fail to consider intertemporal issues related to the sustainability of economic activity. Even modest attempts to improve GDP estimates can result in controversial changes, as is illustrated with an Irish example.

The next section clarifies the relationship between data revision and measurement errors. It shows that common assumptions about the actions of the statistical agency imply that data revisions should have properties unlike those commonly imposed in Classical Measurement Error models and in popular Factor Models. Instead, data revisions should behave analogously to forecast errors, and recent papers have shown how multiple vintages of data series may be combined in a flexible way to improve estimates of the underlying true values.

The following section explains how simple data transformations, such as combining series, or converting to growth rates, or removing trends or cycles, can have profound effects on the relative importance of measurement error. As examples, it is shown that the latter may be over one hundred times larger for the change in the rate of unemployment than for its level, or that it may be larger for measured productivity growth than for any of the series used to construct productivity growth. Two common but more complex data transformations, seasonal and cyclical adjustment, are then examined. They have similarities not only in their econometric treatment but also in their implications for mea-

surement error and data revision. Examples from the seasonal adjustment of dividends and the estimation of business cycles are discussed.

For macroeconomists, increasingly the challenge is to (a) recognise the impact that measurement error may have on their analysis, and (b) where possible, minimize its effects.

# Introduction

Macroeconomic series are regularly revised, sometimes substantially. This reflects the fact that published macroeconomic series not direct observations of macroeconomic variables. Rather, they are estimates of those variables, and estimates which are sometimes revised as more information becomes available. Table 1 gives a glimpse of the extent of the revisions for a number of OECD nations.

Table 1: GDP Revisions 1994Q4-2013Q4

Country	Mean Revision		Mean Absolute Revision	
	Q/Q	Y/Y	Q/Q	Y/Y
Australia	0.11	0.35	0.34	0.57
Belgium	0.07	0.11	0.37	0.48
Canada	0.05	0.26	0.21	0.44
Denmark	0.09	0.06	0.63	0.74
Finland	0.13	0.40	0.59	0.84
France	0.02	0.15	0.23	0.53
Germany	0.02	-0.03	0.32	0.47
Italy	0.04	0.11	0.23	0.39
Japan	0.10	0.29	0.57	0.93
Korea	0.07	0.15	0.49	0.52
Netherlands	0.10	0.44	0.39	0.60
New Zealand	0.11	0.34	0.49	0.81
Norway	-0.01	0.15	0.71	0.95
Portugal	0.01	0.34	0.55	0.53
Spain	0.07	0.26	0.22	0.43
Switzerland	0.08	0.34	0.30	0.59
United Kingdom	0.05	0.15	0.23	0.58
United States	-0.07	-0.33	0.28	0.68
Average	0.06	0.20	0.40	0.61

**Notes:**

The above figures reflect the revision in quarter-over-quarter (Q/Q) and year-over-year (Y/Y) GDP growth rates (in percentage points) over the first 3 years after initial publication.

Source: Zwiijnenburg (2015). Table 2.

It shows that, to varying degrees, GDP growth tends to be revised upwards after its initial release. The average absolute revision may be over 0.5% on a quarter-over-quarter rate or closer to 1% on a year-over-year rate. Much of this in turn stems from revisions to gross fixed capital formation or net exports; personal consumption expenditures typically

suffer relatively smaller revisions.<sup>1</sup>

This chapter is intended as an introduction to and overview of measurement error and revision in macroeconomic data. It will be useful for most applied researchers in macroeconomics who work with official macroeconomic statistics, such as those found in the National Accounts, the Balance of Payments, national government budgets, labour force statistics, etc. Many of the lessons discussed below also apply much more broadly whenever researchers treat data as immutable rather than subject to estimation error and revision.

Macroeconomists may be familiar with measurement error in the context of debates about how well official statistics capture productivity growth. That literature examines both the sources of productivity growth and the construction of productivity statistics.<sup>2</sup> In so doing, it investigates the *accuracy* of productivity. In contrast, the focus in this chapter is on the *precision* of these and other macroeconomic series.<sup>3</sup>

The first section below reviews sources of measurement error in macroeconomics, including definitional issues as well as the role of the statistical agency. The following section presents the most widely used statistical models of measurement errors and discusses the relationship between measurement error and data revision. Thereafter, the impact of common data transformations on the importance of measurement errors is reviewed, showing conditions under which highly reliable raw data may give rise to much less reliable derived series. These results are illustrated with a number of popular macroeconomic series. Two widespread but more complex data transformations are then examined, seasonal and cyclical adjustment, which have similarities in their econometric treatment as well as their implications for measurement error and data revision. The conclusions review some of the implications for macroeconomic analysis and forecasting.

## **Measurement Errors and Macroeconomic Statistics**

Understanding the likely importance of macroeconomic measurement errors requires some understanding of their sources and typical characteristics. Errors can result from ambi-

guity or disagreement about both what *should* be measured as well as *how* it should be measured. This section gives examples of both types of problems in contemporary macroeconomic settings.

### ***What Are We Trying to Measure?***

As priorities change, the things economists seek to measure may evolve or be redefined. To the extent that economists in the past used primitive or outmoded concepts, this implies that their work may be seen to suffer from measurement errors caused by conceptual or definitional errors. The potential importance of such problems in macroeconomics may be appreciated by considering its use of Gross Domestic Product (GDP.) Macroeconomists frequently analyse the welfare implications of a policy or other change by examining its impact on GDP (or GDP per capita). Heuristically, this stems from the notion that higher production levels should be potentially Pareto-improving for society overall. But it also ignores widely acknowledged issues with how GDP is commonly measured. Many modifications have been proposed to the system of National Accounts that seek to correct important conceptual shortcomings in GDP and thereby better capture aggregate welfare.<sup>4</sup>

One of the first of these was the replacement of Gross *National* Product (i.e. production by a country's people) by GDP (production within the country's borders.) Nordhaus and Tobin (1972) proposed a Measure of Economic Welfare (MEW) that added corrections for components of consumption that GDP omits, such as leisure and nonmarket production activity. The fact that all of these measures ignore intertemporal trade-offs has also been acknowledged as a problem. For example, one may temporarily boost consumption by reducing investment and running down the capital stock, or boost production by increasing the extraction rate of a non-renewable resource. However, such boosts will not be sustainable, implying that short-run improvements in consumption or production may not be welfare-improving when future repercussions are considered. To address such issues, substantial effort has been made on the one hand to integrate capital accounts in national accounting systems (for example see World Bank (2006),)

and on the other hand to shift the focus from actual to sustainable levels of production or consumption (as in World Bank (2011).) These build on the concept of *genuine savings* proposed by Hamilton and Clemens (1999), which modify the traditional measures of aggregate savings by adding the value of investment in human capital and subtracting the value of resource depletion.<sup>5</sup> Welfare improvements are only sustainable if genuine savings per capita are non-negative.

To date, while the conceptual framework for these adjustments to GDP exists, the detailed series required for their accurate and timely measurement do not, which has limited their incorporation into the quarterly national accounts series typically used by macroeconomists. However, even the limited modifications that have been incorporated to date have sometimes proved problematic.

#### *Irish GDP Growth and Intellectual Property*

One such example may be seen in Figure 1, which shows a substantial change in the character of official Irish GDP growth starting in 2015Q1. Even during the Great Recession of 2008-09, annualised quarterly growth rates prior to 2015 fluctuated in a range of  $\pm 20\%$ . Thereafter growth became substantially more volatile, starting with a rate exceeding 100% in 2015Q1. This extraordinary growth rate (26.3% on a Q/Q basis) was largely “paper-based.” Among other things, it reflected the decision of multinational firms to move their portfolios of intellectual property assets to Ireland to take advantage of favourable local tax treatments. This was duly incorporated into the National Accounts as an inflow of foreign (intangible) capital, an increase in investment and an increase in foreign indebtedness. Taken together with other corporate restructurings and developments in aircraft leasing (for aircraft operated abroad but owned by Irish leasing firms, another “paper-based” activity), this led the Central Bank of Ireland to argue that

“...the 2015 National Income and Expenditure accounts are not reflective of actual economic activity taking place in Ireland. Instead, these developments reflect the statistical on-shoring of economic activity ... As a result, National Accounts data now include a very significant amount of activity carried out

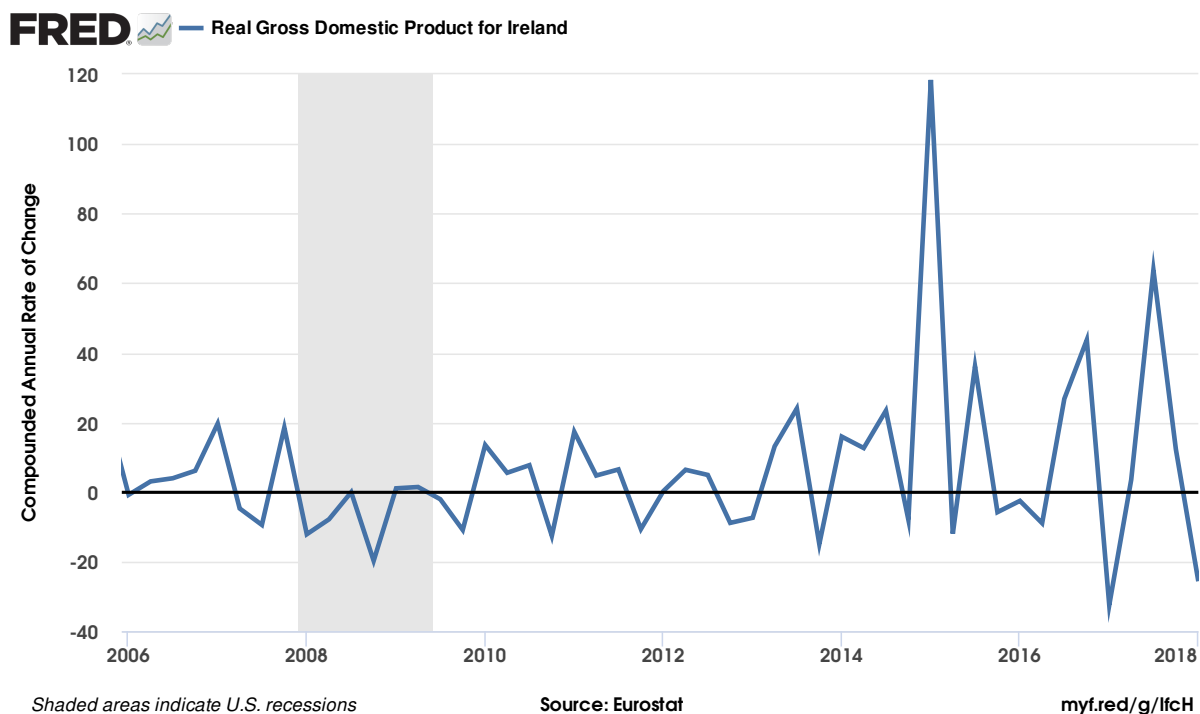


Figure 1: Irish Real GDP Growth

elsewhere, but formally recorded as part of Irish GDP and GNP.”<sup>6</sup>

Despite this assessment, it should be stressed these figures were considered to be logically correct and consistent with contemporary international standards as Bank of Ireland (2016) acknowledges. Their disagreement instead lay in the definition of aggregate economic activity or welfare that was pertinent for macroeconomic analysis. This demonstrates that subtle differences in the definition of aggregate economic output can give rise to what appear to be substantial measurement errors, even in series as widely used and central to macroeconomic analysis as GDP.<sup>7</sup>

Having acknowledged that such conceptual disagreements or confusions may cause substantial measurement errors, the remainder of this chapter focuses instead on more applied problems (the *how* rather than the *what* of measurement), starting with the problem facing the statistical agency.

### ***The Role of the Statistical Agency***

Statistical agencies have a complex mandate. Given limited resources, they try to produce statistics that are timely, precise and relevant to diverse communities of end-users in both



the public and private sectors. Doing so requires them to make judgemental trade-offs between costs, precision, timeliness and accuracy. The degree to which their statistics contain measurement errors is a reflection of some of the choices that they make, and these may vary considerably from one national statistical agency to another.

Sargent (1989) proposes two alternative models of how statistical agencies may respond to known measurement errors. He assumes that in each period they try to estimate an  $n \times 1$  vector of “true” values  $Z_t$  but that they can only observe an error-corrupted version of  $Z_t$ , namely

$$z_t = Z_t + v_t \tag{1}$$

where  $z_t, Z_t$  and  $v_t$  are assumed to be covariance stationary  $n \times 1$  vectors of stochastic processes and  $v_t$  is a mean-zero vector of measurement errors whose realisations are independent of  $Z$ .<sup>8</sup> These assumptions are consistent with what has become known as the “Classical” Measurement Error framework.

In the simplest case the statistical agency simply publishes  $z_t$ . Note that under some circumstances agencies may choose this approach even when they are aware that measurement errors are large and can easily be reduced. For example, when transparency, objectivity, reproducibility and simplicity are key concerns, it may be desirable to simply publish the “raw” data and leave further adjustments to the end-users. Perhaps the most common example of this involves seasonal adjustment. While the concept of interest for most users of macroeconomic series is a seasonally-adjusted value, the availability of seasonally-adjusted figures varies considerably from one national statistical agency to another. End-users understand that the available series contain a serially-correlated measurement error (the seasonal effect) despite the fact that many standardized tools for deseasonalisation are readily available.<sup>9</sup>

Alternatively, the statistical agency may instead report

$$\tilde{z}_t = E(Z_t | z_t, z_{t-1}, \dots) \tag{2}$$

That is, they use a statistical model to estimate the unobserved “true” values  $Z_t$  based on current and past values of their raw series. Sargent (1989) refers to this as the case where the agency “filters” the data, and notes that standard methods for linear state-space models (such as the Kalman filter) naturally lend themselves to such applications. Since in general  $E(Z_t|z_t, z_{t-1}, \dots) \neq E(Z_t|\dots, z_{t+1}, z_t, z_{t-1}, \dots)$ , this suggests revising previously published estimates  $\tilde{z}_t$  as new raw data  $z_{t+1}, z_{t+2}, \dots$  arrive.<sup>10</sup> This also implies that published estimates should become more precise as they are revised. This seems well-suited to describing the common practice of revising recently published estimates at regular intervals after their initial release.

An alternative view of data revision focuses on the occasional and often irregular revision of an entire series, usually due to changes in its definition and/or the methodology used to estimate it. Here such changes will be referred to as “benchmark” revisions. They may simply be due to changes in the units of measure (e.g. the base year used for real value calculations.) In other cases, however, statistical agencies may decide to revise their methodologies in order to better adapt them to a changing economic environment or changing user demands. For example, as new technologies introduce new products and services, there is often a lag before they are incorporated into the samples or benchmarks used by statistical agencies. If those new products and services tend to be the ones with high productivity and output growth, this suggests that initial estimates of overall production and productivity may be biased downwards until benchmark revisions correct the situation. Aghion et al. (2017) provide a formal model in which the imputation of prices for products from new firms creates systemic downward bias in estimates of productivity growth. Using US data, they estimate that this understated annual productivity growth by roughly 0.5% over the 1983-2013 period. Bognanni and Zito (2016) present evidence that benchmark revisions since 1968 have tended to revise initial estimates of US productivity growth rates upwards by an average of 0.5%, and that the average revisions are more positive for the lowest initial estimates. Youll (2008) reviews related arguments and evidence that early estimates of UK GDP growth are biased downwards.<sup>11</sup>

# Data Revision

## *Measurement Errors and Data Revisions*

The fact that macroeconomic series are revised is the clearest evidence that such series are *estimates* and therefore contain measurement errors. That said, there is often unnecessary confusion about the relationship between data revisions and measurement error. Fortunately, the discussion above can to be used to clarify matters.

1. It is sometimes claimed that series which undergo less revision must be more “reliable.” This is false: the case in which the statistical agency simply publishes  $z_t$  and never revises it shows that the absence of data revision tells us nothing about the absence or size of measurement errors.
2. It is sometimes claimed that measurement errors must be “at least as large as” data revisions. This is based on the assumption that the statistical agency’s releases are generated by a process like that in (2). Because the information used to construct the earlier releases is only a subset of that used to construct subsequent releases, the revised estimate must be “no worse” than the earlier release. In the case where the conditional expectation in (2) is formed by linear projection, the expected mean-squared error (MSE) of the later release must be no greater than that of the earlier release.<sup>12</sup>
3. The assumption of a process like that in (2) may be reasonable in the context of revising recently published estimates at regular intervals after their initial release, but it seems untenable in the context of benchmark revisions. Whether series after benchmark revisions are more or less accurate depends on the specific context and the task at hand. The case of Irish GDP discussed above is a clear example where the central bank felt that revised GDP (and GNP) figures were less accurate for the purpose of guiding monetary policy than the series that they replaced.

Establishing a tighter relationship between data revisions and measurement errors requires stronger assumptions. For example, if one assumes that the “final” published

estimate is the true value  $Z_t$ , then the extent to which any estimate is eventually revised would precisely equal its measurement error. Kapetanios and Yates (2010) and Cunningham et al. (2012) propose an alternative model of the statistical agency which replaces Sargent's projection model and (2) with a model in which the agency continues to draw samples from the population of  $z_t$  over time and weighs the results to produce the published estimate  $\tilde{z}_t$ . Together with the assumption that the effective sample size of successive surveys declines at a strictly geometric rate, this enables them to deduce the overall measurement error from the declining volatility of successive data revisions. Jacobs and van Norden (2011) instead assume that measurement errors are a linear combination of news and noise measurement errors (defined below). Together with strict assumptions on the dynamics of  $Z_t$ , this enables them to consistently estimate overall measurement errors from the data revision process. However, neither of these methods have seen wide application.

### ***Modeling Data Revisions***

A considerable literature has studied the characteristics of data revisions in macroeconomic series.<sup>13</sup> It includes examinations of which of the two behavioural models proposed by Sargent (1989) seems to better capture the properties of published data. In particular, measurement errors  $v_t$  are said to be *Noise* when  $Cov(Z_t, v_t) = 0$ . This is consistent with the assumptions of the Classical Measurement Error model and (1). One implication of this is that data revisions should be partially forecastable.<sup>14</sup> It also implies that the true values  $Z_t$  can be no more variable than the observed series  $z_t$ .<sup>15</sup>

Alternatively, measurement errors  $v_t$  are said to be *News* when  $Cov(z_t, v_t) = 0$ , or if the statistical agency reports  $\tilde{z}_t$  instead of  $z_t$ , when  $Cov(\tilde{z}_t, v_t) = 0$ . The latter condition will be satisfied where  $\tilde{z}_t$  is based on (2), and implies that data revisions will be unforecastable.<sup>16</sup> It also implies that the true values  $Z_t$  must be at least as variable as the observed series  $\tilde{z}_t$ .<sup>17</sup>

Empirical work since at least Croushore and Stark (2003) and Faust, Rogers and Wright (2005) has found that, while results vary somewhat from one series to another,

revisions to most macroeconomic series do not fit cleanly into either model. This in turn has led to the development of models in which measurement errors may be a combination of both News and Noise errors. Such models are discussed below, but first a more popular approach should be considered.

### ***Factor Models and Data Revision***

Recent years have seen the widespread adoption of Dynamic Factor Models (DFMs) in macroeconomics, reflecting the usefulness of approaches pioneered by authors such as Forni et al. (2000) and Stock and Watson (2002). While many variations on the basic DFM have been proposed, among the most influential for policy-making institutions and forecasters is the work by Giannone, Reichlin and Small (2008) and Banbura et al. (2013) and others, who adapt the DFM to the problem of integrating and updating estimates from a broad range of data sources with varying reporting frequencies and time lags.<sup>18</sup> The vast majority of such models (including the DFMs of Giannone, Reichlin and Small (2008) and Banbura et al. (2013)) use only the latest available data vintage of each variable on the assumption that the effects of measurement errors should be slight in a data-rich environment.<sup>19</sup> While this might seem a reasonable approach if data revisions are “news”, a “noise” model suggests that additional information can be obtained by including multiple data vintages and attempting to filter the noise.

The simplest approach to do so would be to simply include all data vintages for all series in the above DFM models. However this overlooks the fact that the above DFMs impose structural assumptions on the relationships between variations in the factors and the idiosyncratic errors. To see this, one may write the above DFMs in their state-space representation,

$$y_t = A \cdot \alpha_t + \epsilon_t \tag{3}$$

$$\alpha_{t+1} = T \cdot \alpha_t + R \cdot \eta_{t+1} \tag{4}$$

where  $\epsilon_t \sim i.i.d. N(0, H)$ ,  $\eta_t \sim i.i.d. N(0, Q)$  and  $\epsilon_t \perp \eta_s \forall s, t$ .<sup>20</sup> The orthogonality assumption between  $\epsilon_t$  and  $\eta_t$  is both central to identification in such models and the stumbling block in using them to account for data revision.

To understand why, consider the case where the vector  $y_t$  contains only the various releases of a single series, and let the scalar  $Y_t$  be the unobserved “true” value of the series at time  $t$ . One could use (3) and (4) to estimate  $\alpha_t$ . However, those estimates will be biased for  $Y_t$  unless the measurement errors are pure noise, since that is the only case in which the idiosyncratic errors  $\epsilon_t$  are uncorrelated with the true values  $\alpha_t$  and therefore the condition  $\epsilon_t \perp \eta_s \forall s, t$  is respected.

To solve this problem, Jacobs and van Norden (2011) and Kishor and Koenig (2012) relax the orthogonality conditions. The former’s model is a special case of (3) and (4) where they set

$$A = [1_{l \times 1}, I_{l \times l}] \quad (5)$$

$$\epsilon_t = 0 \quad (6)$$

$$\alpha_t = [Y_t, \nu_t] \quad (7)$$

$$R = \begin{bmatrix} R_1 & R_2 \\ 0_{l \times 1} & -U_1 \cdot \text{diag}(R_2) \end{bmatrix} \quad (8)$$

$$Q = I_{(l+1) \times (l+1)} \quad (9)$$

where  $l$  is the number of releases in  $y_t$ ,  $\nu_t$  is the vector of news measurement errors,  $U_1$  is an  $l \times l$  matrix with zeros below the main diagonal and ones everywhere else, and  $R_1, R_2$  control the variability of the true values and the news shocks  $\nu_t$ .<sup>21</sup> Note that  $A$  defines the  $l$  observed releases to be the sum of the true value  $Y_t$  and news errors  $\nu_t$ . Both are generated from the vector of i.i.d. errors  $\eta_t$  via  $R$ , which imposes the news structure. The first row of  $R$  ensures that all elements of  $\eta_t$  affect  $Y_t$ ; the remaining rows of  $R$  strip away a decreasing number of these to produce the  $l$  different data vintages. The result is that  $R$  imposes the correlation structure between the various data releases and the true values

as well as the variance inequalities that are implied by news errors.<sup>22</sup> This comes at the cost of a much larger state-vector. In the DFM models the state vector for a comparable univariate model would be a scalar containing just the true values  $Y_t$ . Here it contains an additional  $l$  measurement errors.

## The Simple Arithmetic of Measurement Errors

While the above has argued that measurement errors may be important in some situations, it has given no indication of when this is likely to be problem in practice. The remainder of this chapter tries to remedy this by considering common practices which may exacerbate the seriousness of measurement errors. This section considers how simple arithmetic operations affect the economic importance of measurement error, while the section thereafter examines more complex data transformations.

Until now true values and measurement errors have been considered in the context of a vector  $Z$  of arbitrary dimension. Below, several commonly-used univariate and bivariate transformations are examined. To keep the exposition as clear as possible, the notation that has been used to this point must be modified.

### *A Change in Notation*

In this section,  $X_t, Y_t, Z_t$  are used to denote the true values of three distinct time series, and  $x_t, y_t, z_t$  to denote their estimated or published values. For compactness, the time subscript  $t$  (which is assumed to range from 1 to  $T$ ) may be omitted. Each series has its associated measurement error defined as

$$\epsilon_x \equiv x - X \tag{10}$$

$$\epsilon_y \equiv y - Y \tag{11}$$

$$\epsilon_z \equiv z - Z \tag{12}$$

It is not assumed that the  $\epsilon$ 's have means of zero (i.e. estimates could be biased) and they may or may not be independent of their corresponding variable's true values (i.e. they could be news or noise.) The objective is to understand the economic importance of measurement error in  $z$ . To do so, one may focus on its Noise-to-Signal (NS) ratio, defined as

$$\varphi(z) = E(\epsilon_z^2)/Var(z) \quad (13)$$

Note that when  $E(\epsilon_z) <> 0$ , the numerator will be larger than  $Var(\epsilon_z)$ , so that  $\varphi(z)$  captures both the variability of measurement errors as well as any bias.

### *Arithmetic*

Now suppose that  $z$  is not measured directly, but is instead constructed as the sum of two variables  $x$  and  $y$  so that<sup>23</sup>

$$\begin{aligned} z &= x + y \quad \text{and} \quad Z = X + Y \\ \Rightarrow \epsilon_z &= \epsilon_x + \epsilon_y \end{aligned}$$

This in turn implies that

$$\begin{aligned} \varphi(z) &= E([\epsilon_x + \epsilon_y]^2)/Var(z) \\ &= \frac{E(\epsilon_x^2 + \epsilon_y^2 + 2 \cdot \epsilon_x \cdot \epsilon_y)}{Var(z)} \\ &= \frac{E(\epsilon_x^2)}{Var(x)} \cdot \frac{Var(x)}{Var(z)} + \frac{E(\epsilon_y^2)}{Var(y)} \cdot \frac{Var(y)}{Var(z)} + 2 \cdot \frac{E(\epsilon_x \cdot \epsilon_y)}{Cov(x, y)} \cdot \frac{Cov(x, y)}{Var(z)} \\ &= \varphi(x) \cdot \frac{\sigma_x^2}{\sigma_z^2} + \varphi(y) \cdot \frac{\sigma_y^2}{\sigma_z^2} + 2 \cdot \frac{E(\epsilon_x \cdot \epsilon_y)}{\sigma_z^2 - \sigma_x^2 - \sigma_y^2} \cdot \frac{\sigma_z^2 - \sigma_x^2 - \sigma_y^2}{\sigma_z^2} \end{aligned} \quad (14)$$

where  $\sigma_i^2$  is the variance of variable  $i$ . This last equation relates the noise-signal ratio for  $z$  to a weighted average of the noise-signal ratios of its components  $x$  and  $y$ , as well as a term in the cross-moment of their measurement errors. When the latter is zero, the larger  $\sigma_x^2$  (or  $\sigma_y^2$ ) is relative to  $\sigma_z^2$ , the more weight it receives in  $\varphi(z)^2$ , a result which seems intuitive.



Less intuitive perhaps is the case where both  $\sigma_x^2/\sigma_z^2 > 1$  and  $\sigma_y^2/\sigma_z^2 > 1$  (as might occur when  $Cov(x, y) < 0$ , for example.) In this case, one may find  $\varphi(z)^2 > \max(\varphi(x)^2, \varphi(y)^2)$ . This arises simply because  $z$  is less variable than either of its components  $x$  or  $y$ , so the *relative* importance of their respective measurement errors becomes larger.

This simple result leads to two rules of thumb for empirical researchers.

1. Measurement errors may be magnified when working with the first-difference of persistent series.
2. Measurement errors may be magnified when working with the difference of two correlated series.

Each of these cases is now examined in turn.

### ***Measurement Errors in First-Differences***

In the case of first differences, one may set  $y_t = -x_{t-1}$  so that  $z_t = x_t - x_{t-1} = \Delta x_t$ . Assuming that  $\sigma_y^2 = \sigma_x^2$  and  $\varphi(x) = \varphi(y)$ , (14) reduces to

$$\begin{aligned} \varphi(z) &= 2 \cdot \left[ \varphi(x) \cdot \frac{\sigma_x^2}{\sigma_z^2} + \frac{E(\epsilon_{x_t} \cdot \epsilon_{x_{t-1}})}{\sigma_z^2 - 2 \cdot \sigma_x^2} \cdot \frac{\sigma_z^2 - 2 \cdot \sigma_x^2}{\sigma_z^2} \right] \\ \Rightarrow \varphi(\Delta x) &= 2 \cdot \left[ \varphi(x) \cdot \frac{\sigma_x^2}{\sigma_{\Delta x}^2} + \frac{E(\epsilon_{x_t} \cdot \epsilon_{x_{t-1}})}{\sigma_{\Delta x}^2 - 2 \cdot \sigma_x^2} \cdot \frac{\sigma_{\Delta x}^2 - 2 \cdot \sigma_x^2}{\sigma_{\Delta x}^2} \right] \end{aligned} \quad (15)$$

Even when  $E(\epsilon_{x_t} \cdot \epsilon_{x_{t-1}}) = 0$ , the Noise-Signal ratio in first differences will be larger than that in the level of the series by a factor of  $2 \cdot \sigma_x^2/\sigma_{\Delta x}^2$ . This ratio will typically increase with the persistence of  $x$ .<sup>24</sup>

As an example, consider the US unemployment rate (measured in %.) Taking the 70-year period starting in January 1948, the rate had a sample variance of 2.67 while that of monthly changes in the rate was 0.044, implying that  $\frac{\sigma_x^2}{\sigma_{\Delta x}^2} = 60.8$ . One should therefore expect that, even when measurement errors have a mean of zero and are serially independent so that  $E(\epsilon_{x_t} \cdot \epsilon_{x_{t-1}}) = 0$ , the noise-signal ratio for monthly changes in the unemployment rate should be roughly 120 times larger than that for the level of the rate itself. Kozicki and Hoffman (2004) provide evidence of similar problems with US CPI

data which (like the unemployment rate) is typically reported to only one decimal place. While this rounding is relatively innocuous for measuring the level of the series, they document the severe measurement error that this discretization sometimes introduces for inflation rates measured over relatively short time spans.

### ***Measurement Errors in Differences and Ratios***

Now suppose that  $z$  is constructed as the difference of two series so that  $z = y - x$ , as might be found for example in the case of fiscal surplus/deficits, current account balances, or net savings. Alternatively, one might be interested in cases where  $z$  is calculated as the ratio of two series  $z = y/x$ , as would be the case for unemployment rates (unemployment as a fraction of the labour force) or productivity measures (the amount produced per unit input.) In this case  $\ln z = \ln y - \ln x$  so one could work with the NS ratios of the logarithms of the variables. From (14) this gives

$$\begin{aligned}
\varphi(z) &= E([\epsilon_y - \epsilon_x]^2)/Var(z) = \frac{E(\epsilon_x^2 + \epsilon_y^2 - 2 \cdot \epsilon_x \cdot \epsilon_y)}{Var(z)} \\
&= \frac{E(\epsilon_x^2)}{Var(x)} \cdot \frac{Var(x)}{Var(z)} + \frac{E(\epsilon_y^2)}{Var(y)} \cdot \frac{Var(y)}{Var(z)} - 2 \cdot \frac{E(\epsilon_x \cdot \epsilon_y)}{Cov(x, y)} \cdot \frac{Cov(x, y)}{Var(z)} \\
&= \varphi(x) \cdot \frac{\sigma_x^2}{\sigma_z^2} + \varphi(y) \cdot \frac{\sigma_y^2}{\sigma_z^2} - 2 \cdot \frac{E(\epsilon_x \cdot \epsilon_y)}{\sigma_z^2 - \sigma_x^2 - \sigma_y^2} \cdot \frac{\sigma_z^2 - \sigma_x^2 - \sigma_y^2}{\sigma_z^2} \tag{16}
\end{aligned}$$

Aside from a change of sign on the cross-product terms, much is as before. To illustrate the practical application of this change, consider two simple examples, using the U.K. Current Account and U.S. Productivity Growth.

#### *U.K. Current Account*

Table 2 shows the results for revisions to the quarterly UK Current Account figures from 1996Q1 to 2015Q2.<sup>25</sup> The first line shows us that while the current account credits and debits both have NS ratios ( $\phi$ ) around 2%, that for the current account balance (which is simply their difference) is almost 15%. The next line shows that this is partly due to the fact that each of these components is more than 25 $\times$  as volatile as their sum. The final

Table 2: Data Revision in the UK Current Account: 1996Q1 to 2015Q2

Series	Credits	Debits	Cross-Moments	Balance
NS Ratios ( $\phi_i$ )	0.0245	0.0161	2.601	0.149
Variance Ratios ( $\sigma_i/\sigma_z$ )	26.45	34.15	0.202	1
$(\sigma_i/\sigma_z) \cdot \phi_i$	0.648	+ 0.550	$-2\times$ 0.524	= 0.149

**Notes:**

The figures displayed under “Cross-Moments” are the relevant cross moments of Credits and Debits, calculated as  $\sigma_{x,y}^2/\sigma_z^2$  for the Variance Ratio and  $E(\epsilon_x \cdot \epsilon_y)/\sigma_{x,y}^2$  for the NS Ratio.

line of the table allows us to verify the decomposition in (16) by taking into account the remaining terms in  $E(\epsilon_x \cdot \epsilon_y)$  and  $\sigma_z^2 - \sigma_x^2 - \sigma_y^2$  which are shown in the column marked “Cross”. This shows that the NS ratio for the current account balance would be much higher still (over 100%!) were it not for the positive correlation of revisions to the credits and debits.

*U.S. Productivity Growth*

Table 3: Data Revisions in US Productivity Growth: 1999Q2 to 2018Q2

Series	Output	Hours	Cross-Moments	OPH
NS Ratios	0.392	0.080	1.054	0.553
Variance Ratios	1.401	1.080	0.034	1.000
$(\sigma_i/\sigma_z) \cdot \phi_i$	0.550	+ 0.086	$-2\times$ 0.036	= 0.565

**Notes:**

The figures displayed under “Cross-Moments” are the relevant cross moments of Output and Hours, calculated as  $\sigma_{x,y}^2/\sigma_z^2$  for the Variance Ratio and  $E(\epsilon_x \cdot \epsilon_y)/\sigma_{x,y}^2$  for the NS Ratio. Growth is measured as the quarterly change in logs.

The minor discrepancy in the decomposition of the NS ratio for OPH is due to small inconsistencies between the initially published figures for OPH on the one hand and those for the Output and Hours series on the other.

Table 3 shows the results for revisions to growth in US productivity (measured as real output per hour worked – OPH) from 1999Q2 to 2018Q2.<sup>26</sup> Here one sees an example that combines growth rates with the difference between two series (the logs of real output and hours worked.) In the first line one again sees that the NS ratio for the difference of the two series (OPH) is larger than that for either of its components. In this case,

the NS ratios for all three series that are substantially larger than those in Table 2, with that for productivity growth exceeding 50%.<sup>27</sup> The decomposition shows that this largely reflects the importance of revisions in the real output series, which accounts for  $0.550/0.565 = 97\%$  of the NS ratio for productivity growth. In addition the cyclical comovement of output and hours implies that productivity is slightly less volatile than either of its components as shown by the variance ratios in the second line. This in turn magnifies the relative importance of the revisions in real output for measured productivity.

## Adjustment and Measurement Error

There are many situations in which macroeconomists are interested in series which abstract from some source of regularly observed variations. By far the most common among these are seasonally-adjusted series, which seek to remove regular annual fluctuations. Also important in some applications are quantities (such as potential output, or structural fiscal deficits) which abstract from the influence of business cycles. While the seasonal cycle in US GDP is of roughly the same magnitude as the business cycle, the problems of seasonal adjustment and related estimation errors and revisions have long occupied statisticians, while macroeconomists have largely focused on the analogous issues for business cycles. Both processes, however, contribute to the overall measurement error of cyclically-adjusted series. Their estimation error is therefore often layered on top of the types of estimation error mentioned in the previous section.

Most seasonal and cyclical adjustment methods rely on the application of linear filters, so that the adjusted series may be expressed as moving averages of past, present and future values of the unadjusted series. The linear state-space models used to derive many of these filters often provide a coherent framework for evaluating the degree of measurement error which such estimation adds to the adjusted series.

## *Seasonal Adjustment*

Seasonal adjustment is a challenging problem. Many countries find it sufficiently difficult that they prefer not to release seasonally-adjusted data.<sup>28</sup> The most widely-used seasonal adjustment approaches include the X12-ARIMA method developed at the US Census Bureau (see Findley et al. (1998)) and the TRAMO-SEATS model-based approach due to Gómez and Maravall (2000), which have both been incorporated in the Census Bureau's X13ARIMA-SEATS seasonal adjustment programs.<sup>29</sup> All grapple with the fundamental problem that the seasonal cycle is not a constant, but that it instead evolves over time. This means that as new observations become available, the modeling approach must decide how much of the observed change in the series is due to changes in the seasonally-adjusted value, and how much is due to changes in the seasonal cycle itself. Wright (2013) provides an insightful examination of this problem in the context of US Nonfarm Payrolls Employment and the sharp drop due to the 2008 recession. Here, another series with important seasonal cycles is considered; dividends on US stocks.<sup>30</sup>

### *Seasonally-Adjusted Dividends*

Figure 2 shows both the raw (in black) and seasonally-adjusted (SA, in red) monthly values for the log of the dividend series on the CRSP Value-Weighted index.<sup>31</sup> It shows very large monthly variations in dividends, reflecting the fact that most listed stocks pay quarterly dividends, but that the month in which they are paid varies from firm to firm.

To understand the importance of changes in the seasonal pattern, consider what happens if one simply regresses the monthly dividend growth rate on a constant and eleven monthly dummies. If the seasonal pattern were constant, the residuals from this regression should provide a seasonally-adjusted dividend growth rate. Figure 3 compares the autocorrelations of the raw data (solid line) to that of the regression residuals (dotted line.) It shows that almost all of the strong quarterly pattern of peaks and troughs in the autocorrelations of the raw series are preserved in the regression residuals, implying that the monthly dummies failed to capture much of the seasonal cycle in dividends.

One can better appreciate how the seasonal cycle evolves over time by applying the

## Original and Adjusted Series

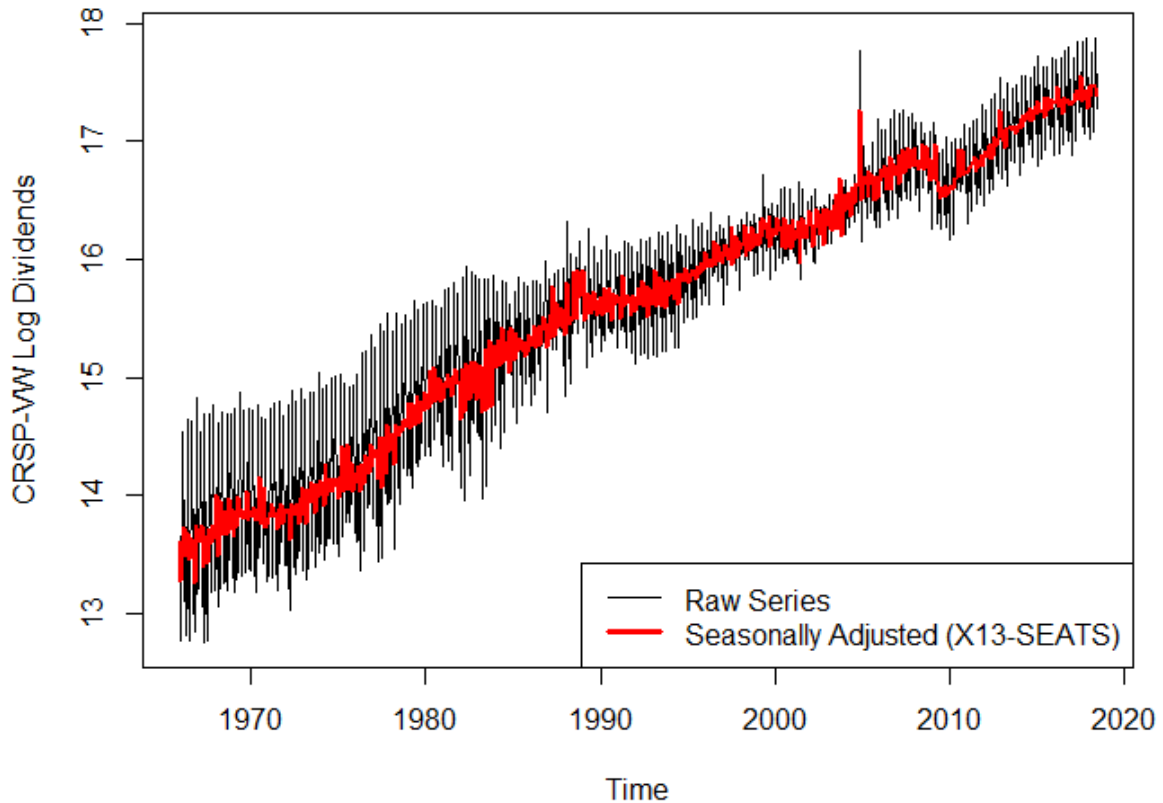


Figure 2: Monthly Stock Dividends

US Census Bureau’s X13 seasonal adjustment techniques to the dividend series.<sup>32</sup> Figure 4 shows how the estimated seasonal factors evolved over time by comparing the annual estimates of each month’s factor in a series of 12 plots. It is evident that dividends in the middle month of each quarter are consistently higher than those in the other months and that this is the dominant source of the seasonal variation. However, it is also clear the relative size of the “mid-quarter month” factor appears to have changed substantially over time, growing rapidly early in the sample, then falling to substantially less than its initial value by the mid-point of the sample, and then gradually recovering somewhat. Since the 12 seasonal factors are constrained to sum to zero, it is no surprise to see that these movements are largely mirrored by movements in the opposite direction in most of the other seasonal factors. As far measurement error is concerned, this volatility in the seasonal factors implies that simply capturing the *average* seasonal effect will typically

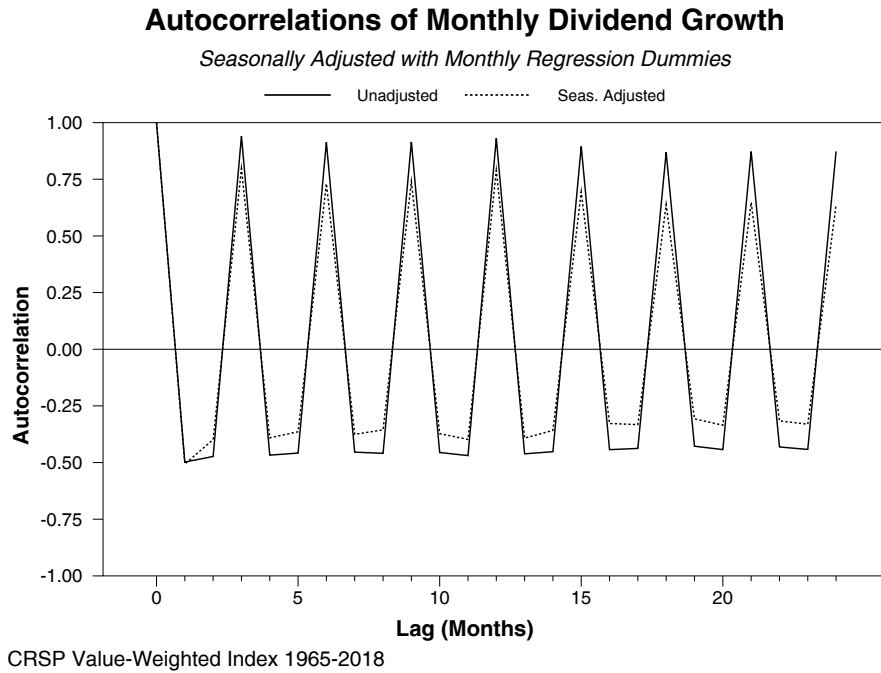


Figure 3: Seasonal Adjustment with Monthly Dummies

be insufficient, as was seen in Figure 3.

But how well do sophisticated methods capture these time-varying seasonals? One suggested diagnostic is to study how the estimated factors change as new observations become available. Revisions in the estimated factors are then interpreted as evidence of error in the initial estimates.<sup>33</sup> Figure 5 compares those initial estimates over the 2008 to 2011 period to those from the full data sample, showing that the revisions appear to be trivial. However, this ignores the fact that seasonal variations are much larger than the variations in the seasonally-adjusted estimates shown in Figure 2. Figure 6 shows the revisions in the latter during the same period, which appear to be more economically significant. In particular, it shows that the sharp fall in dividends in early 2009 was initially underestimated as it was (incorrectly) partially attributed to a change in the seasonal factor; only in retrospect did the change in the seasonally-adjusted dividends become clear. The same also occurs in reverse as dividends recover in early 2010.

These kinds of estimation errors will occur in most settings where there is a need to adjust the observed data to correct for a seasonal or other cyclical pattern that may evolve

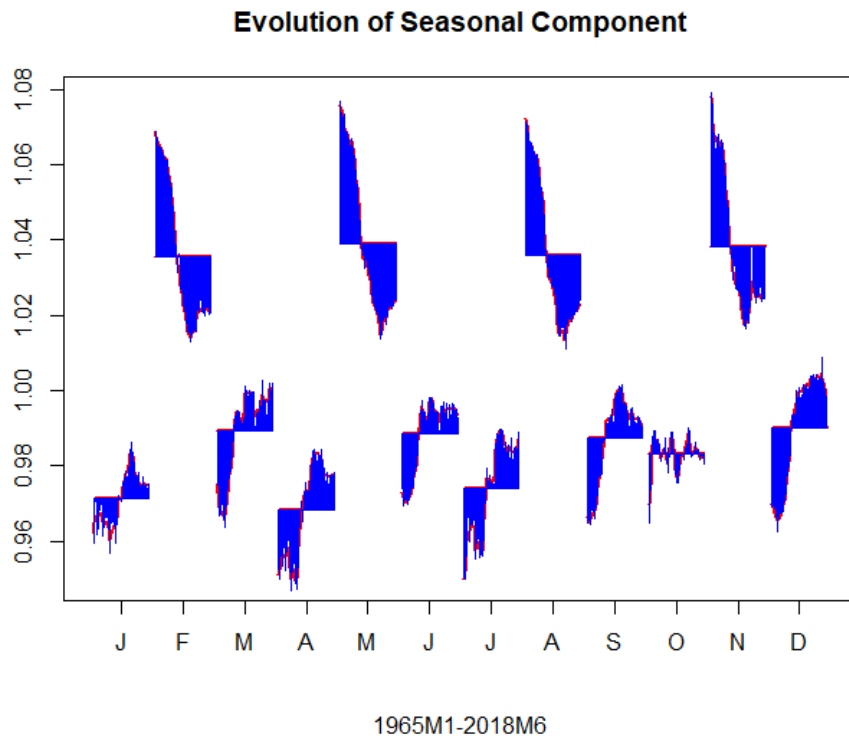


Figure 4: Evolution of Seasonal Factors

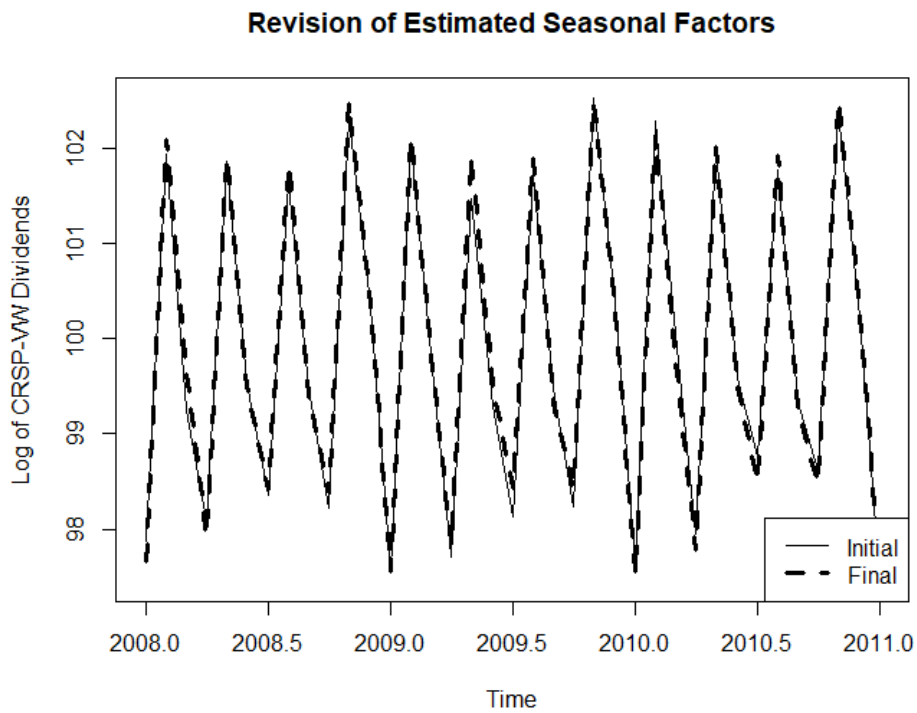


Figure 5: Revision in Seasonal Factors



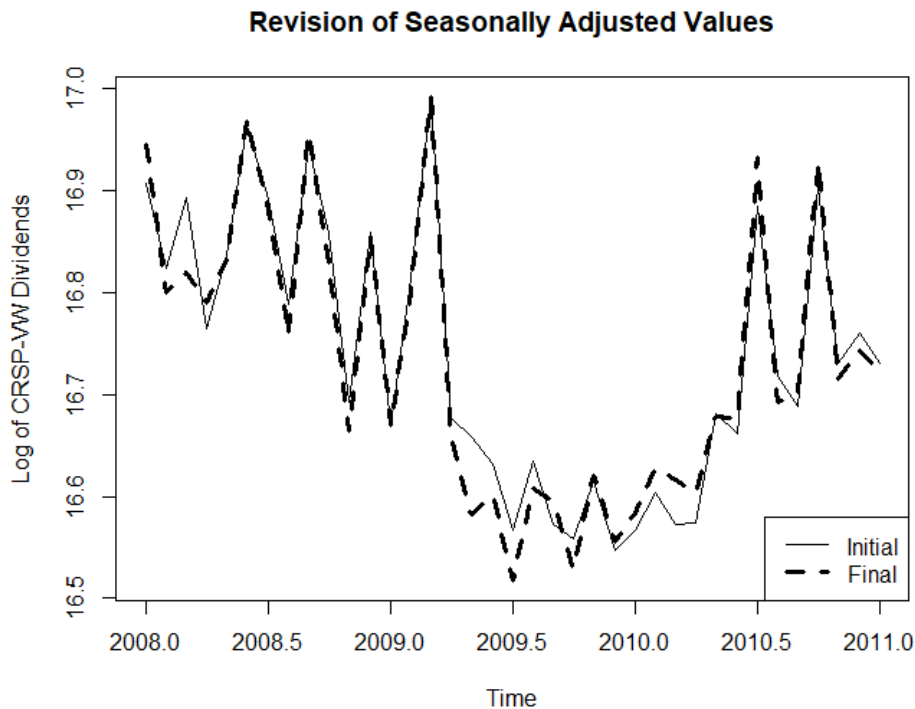


Figure 6: Revision in Seasonally-Adjusted Values

over time. In the wake of unusually large changes, such methods will tend to confuse the cyclical and cyclically-adjusted changes, something that will only be corrected with hindsight. This same phenomena occurs with business cycles, which is the next topic.

### *Trends and Cycles*

Just as macroeconomists find it useful to abstract from seasonal fluctuations, they also often prefer to abstract from business cycle fluctuations. For example, fiscal policy may be described in terms of a “structural” deficit, productivity growth trends may be measured in terms of “utilisation-adjusted” total factor productivity, GDP growth prospects may be compared to “potential output” and unemployment and interest rates may be compared to estimates of their respective “natural” rates. In every case, macroeconomists attempt to estimate the degree to which the observed data on GDP or productivity, etc. is influenced by temporary business cycle factors. With such estimation comes another potential source of estimation error.

Some of the methods used for cyclical adjustment are conceptually similar to those

used in seasonal adjustment in that they rely on removing specific dynamics from the data series (seasonal frequencies in the case of seasonal adjustment and business-cycle frequencies in the case of cyclical adjustment) using time-series techniques such as smoothing, band-pass filtering or structural state-space modeling. However, business cycle estimation also makes more use of multivariate modeling techniques (including structural vector autoregressions, cointegrating relationships and factor models) which use additional information in order to provide estimates of the business cycle that are more closely tied to economic theory.

It would be an understatement to say that there is a lack of consensus about which of these many different approaches is “best.” For that reason, most national statistical agencies refrain from publishing such series. However, among macroeconomists and policy institutions (such as central banks) a great variety of techniques are in use and results from many different methods are frequently consulted and compared. Because estimates often vary widely across methods (for example, see Orphanides and van Norden (2002)) this makes model uncertainty a much more important contributor to estimation error in the context of business cycle measurement than was seen above in the case of seasonal adjustment.<sup>34</sup>

To the extent that business cycle estimates are based on time-series techniques like those used in seasonal adjustment, studying their revision properties may give insight into the reliability of contemporary estimates of the cycle. Orphanides and van Norden (2002) examine estimates of US output gaps from several such methods and find that revisions are typically the same order of magnitude as the estimated gap.<sup>35</sup> This reflects a more serious problem than typically encountered in seasonal adjustment. Whereas most seasonally-adjusted series show a clear peak in their spectral density at the seasonal frequencies, few macroeconomic series show a comparable peak at business cycle frequencies.<sup>36</sup> This in turn makes separating the signal from the noise more difficult for business cycles than for seasonal fluctuations.

This difficulty with a purely time-series approach has increased the attractiveness of multivariate approaches to business cycle estimation. According to studies of their

revision properties, their success has been mixed, however. Numerous studies, such as Orphanides and van Norden (2002) have found that Phillips Curve relationships do little to improve the situation, reflecting the relatively weak link between output movements and inflation in recent decades in most major economies. Kurmann and Sims (2017) document large revisions in widely-used series of utilization-adjusted US TFP, but suggest an alternative adjustment method that may produce more precise estimates. Garratt, Mitchell and Vahey (2014) suggest that structural VAR methods tend to produce better results, a conclusion shared with Coibion, Gorodnichenko and Ulate (2017).

Several recent studies have examined cyclically-adjusted estimates used to guide macroeconomic policy decisions. These are informed by a variety of time-series and structural estimates as well as expert judgement. While in theory cyclically-adjusted quantities should not be systematically related to cyclical downturns, studies examining initial estimates of potential output in the wake of the 2008 recession and Global Financial Crisis find systematic overestimates of potential initially and subsequent large downward revisions. This includes Doern and Zuber (2017), who examine OECD estimates of potential output after recessions, Kuang and Mitra (2018), who look at the EC's estimates of potential output and structural balances, and Coibion, Gorodnichenko and Ulate (2017), who look at Federal Reserve Board and Congressional Budget Office estimates for the US, as well as IMF and OECD estimates for a number of other countries. Kuang and Mitra (2018) also find the reverse in the years following the crisis, with initial underestimates of potential output consistently being revised upwards.<sup>37</sup> All of these studies point to the need for awareness among macroeconomic policy advisors of the relative lack of precision of such estimates.

## Conclusions

Measurement error is seemingly omnipresent in macroeconomics. Some is caused by disagreement or confusion about what should be measured. Some is due to the statistical agency's challenge of producing timely, accurate and precise estimates with a fixed budget

and a diverse clientele. The economic importance of measurement error may be accentuated when the focus shifts from levels to growth rates, or from gross values to net values or ratios. Standard seasonal adjustment methods may give less reliable estimates in the aftermath of unusually large changes. Estimates which purport to correct for business cycle fluctuations are often substantially revised in the years following major recessions and recoveries.

For macroeconomists, increasingly the challenge is to (a) recognise the impact that measurement error may have on their analysis, and (b) where possible, minimize its effects. The examples mentioned above should help them recognise common situations in which measurement problems may be economically significant, and mechanisms which may increase their size and impact. That said, results for the US (and the UK) are far from universally representative; those working with data from other sources need to carefully check their quality. Analysis of data revisions may help identify some problems, but the absence of revisions is not a sufficient indication of reliability. In modeling measurement errors, it should be remembered that classical models of measurement errors fail to capture how published estimates may improve over time as more and better sources of information become available. A growing literature suggests methods for improving forecasts by using methods adapted to such measurement errors.<sup>38</sup> Finally, macroeconomic policy advice should be conditioned on a realistic assessment of the quality of information that decision makers have available.

## Further Reading

Stiglitz et al. (2009) and Jorgenson (2018) discuss efforts to measure aggregate economic welfare, while World Bank (2006), World Bank (2011) and Carey (2018) discuss attempts by statistical agencies to improve the system of National Accounts.

Croushore (2011) surveys research on data revision, as do the introductory literature surveys in Croushore and Stark (2001) and Jacobs and van Norden (2011).

Anderson and Kliesen (2006), Jacobs and van Norden (2016) and Kurmann and Sims (2017) look at the revision in various measures of productivity growth.

Garratt, Mitchell and Vahey (2014) examine the contribution of model uncertainty to business cycle estimation error.

The Clements and Galvao paper in this volume discuss the construction of forecasts in the presence of measurement error.

# Notes

<sup>1</sup>See Zwijnenburg (2015), Figures 7 and 8.

<sup>2</sup>For example, see Aghion et al. (2017) and the references therein.

<sup>3</sup>Accuracy refers to the closeness of an estimate to the true value. Precision refers to the closeness of multiple estimates to each other. Estimates which are widely scattered around the true value but whose average is correct may be said to be accurate but not precise. Estimates which are tightly clustered but consistently tend to deviate from the true value in a particular direction may be said to be precise but not accurate.

<sup>4</sup>For an greatly expanded discussion, see Stiglitz et al. (2009) or the excellent survey by Jorgenson (2018).

<sup>5</sup>See Carey (2018) for recent advances in their measurement.

<sup>6</sup>Bank of Ireland (2016), p. 7. The same source also describes the very large revisions to the Irish national accounts released that month.

<sup>7</sup>A somewhat related phenomena occurs in Norway, a country with a smaller population but substantial off-shore oil production. Rather than focus on measures of total GDP, the central bank focuses on “mainland” economic activity. For examples, see various issues of the Norges Bank’s Monetary Policy Report.

<sup>8</sup>The assumption is that  $E(v_t \cdot \epsilon_s) = 0$  for all  $t, s$  where  $\epsilon_s$  is the innovation in  $Z_s$  from its Wold representation. Note that  $v_t$  may be serially correlated and may be correlated across the  $n$  variables in an arbitrary fashion.

<sup>9</sup>For users of US data, the lack of seasonally-adjusted series for aggregate stock dividends is good example; see Pettenuzzo, Sabbatucci and Timmermann (2018) and the discussion below.

<sup>10</sup>See the OECD / Eurostat Guidelines on Revisions Policy and Analysis at <http://www.oecd.org/sdd/oecdeurostatguidelinesonrevisionspolicyandanalysis.htm>

<sup>11</sup>Bank of England fan charts from around this period showed that the Bank expected recent estimates of GDP growth to be underestimates. However, Robinson (2016) argues that the evidence of bias disappeared in the aftermath of the 2008 recession.

<sup>12</sup>The only case in which the two MSEs would be equal is the case in which the series was not revised.

<sup>13</sup>See the extensive survey by Croushore (2011), or the shorter surveys contained in Croushore and Stark (2001) or Jacobs and van Norden (2011).

<sup>14</sup>To see this, note that  $Cov(Z_t, v_t) = 0$  together with (1) implies that  $Cov(z_t, v_t) \neq 0$ . Once  $z_t$  is observed, this gives information on the unobserved value  $v_t$ . If values of  $v_t$  are independent across successive data releases, then  $Cov(z_t, v_t) > 0$ . This implies that unusually low values of  $z_t$  should be revised upwards, and vice versa.

<sup>15</sup>To see this, note that (1) implies  $Var(z_t) = Var(Z_t) + Var(v_t) + 2 \cdot Cov(Z_t, v_t)$ . When  $Cov(Z_t, v_t) = 0$ ,

this implies  $Var(z_t) < Var(Z_t)$  whenever  $Var(v_t) > 0$ .

<sup>16</sup>Slightly more precisely, if  $\tilde{z}_t$  is formed as the linear projection of  $Z_t$  onto the information set spanned by  $\{z_t, z_{t-1}, \dots\}$ , then the measurement error ( $\tilde{z}_t - Z_t$ ) must be orthogonal to  $\{z_t, z_{t-1}, \dots\}$ .

<sup>17</sup>To see this, note that one may rewrite (1) as  $Z_t = z_t - v_t$ , so  $Var(Z_t) = Var(z_t) + Var(v_t) - 2 \cdot Cov(z_t, v_t)$ . When  $Cov(z_t, v_t) = 0$ , this implies  $Var(Z_t) < Var(z_t)$  whenever  $Var(v_t) > 0$ .

<sup>18</sup>In addition, Now-Casting.com uses their methodology to produce reports for a variety of commercial clients.

<sup>19</sup>An referee noted that such arguments implicitly assume that the “correct” number of factors are extracted, and, in some cases, that measurement errors are uncorrelated across variables (i.e.  $H$  is diagonal.) The latter assumption is literally incredible.

<sup>20</sup>For clarity, the constants in the above equations have been suppressed and it is assumed that the system matrices  $A, T, R, H$  and  $Q$  are constant. Compare, for example, with equations (1) and (2) in Banbura et al. (2013).

<sup>21</sup>The dynamics for  $Y_t$  have been omitted for simplicity; see Jacobs and van Norden (2011) equations (7) and (8) and their section 3.2.2 for details.

<sup>22</sup>Jacobs and van Norden (2011) also show that in the more general case when  $Y_t$  is persistent, all these parameters are identified.

<sup>23</sup>The following derivation may be easily extended to the case where  $z$  is any linear combination of  $x$  and  $y$ . To see this, note that multiplying  $x$  by a factor of  $a$  will simply scale its covariance and that of its measurement error by  $a$  and their respective variances by  $a^2$ . Extensions to the case where  $z$  is a linear combination of  $n$  different variables is also straightforward; see the Appendix of Jacobs and van Norden (2016).

<sup>24</sup>For example, in the case where  $x$  follows a stationary AR(1) process with first-order autocorrelation  $\rho$ , the variance ratio will be proportional to  $1/(1 - \rho)$  and will be greater than 1 for all  $\rho \mid 0 < \rho < 1$ .

<sup>25</sup>Data are taken from the revision triangles available on the ONS website. Revisions are measured as the first estimate - the value reported 3 years later, which is the metric reported by the ONS.

<sup>26</sup>Figures are for output per hour worked in the non-farm business sector. Original vintage data are taken from the Federal Reserve Bank of St. Louis’ ALFRED for series OUTNFB, HOANBS and OPHNFB. Revisions are calculated using the first release for which all three series are available minus the value in the September 2018 vintage.

<sup>27</sup>See Anderson and Kliesen (2006), Jacobs and van Norden (2016) and Kurmann and Sims (2017) for additional analysis of revisions in this and related measures of US productivity growth.

<sup>28</sup>Moulton and Cowan (2016) notes several contemporary critiques of seasonal adjustment in US GDP and discusses how the Bureau of Economic Analysis (BEA) is trying to address them. See Wright (2018) for an early assessment of the BEA’s latest seasonal adjustment methods.

<sup>29</sup><http://www.census.gov/srd/www/x13as/>

<sup>30</sup>See also Pettenuzzo, Sabbatucci and Timmermann (2018), who develop an innovative high-frequency approach to seasonally-adjusting dividends.

<sup>31</sup>The CRSP Value-Weighted Index contains monthly dividend information reaching back to 1926. Due to limitations in the Census Bureau's X13 programs, the analysis below of seasonally-adjusted values will be restricted to the period starting in January 1965. Details of the seasonal adjustment are given below.

<sup>32</sup>The results below use a preliminary logarithmic transformation of the raw CRSP Dividend Index together with the SEATS approach, automatic model selection and no outliers. The purpose here is merely to illustrate how seasonal adjustment may introduce measurement errors; it is not meant to imply that this or any of the other X13 adjustment methods do an adequate job of seasonal adjustment for this series. See Pettenuzzo, Sabbatucci and Timmermann (2018) for a better approach.

<sup>33</sup>See Findley et al. (1998), section 2.2.2. Abeln, Jacobs and Ouwehand (2018) propose an alternative method of seasonal adjustment that avoids revision of the seasonal factors.

<sup>34</sup>See Garratt, Mitchell and Vahey (2014) for an analysis of the contribution of model uncertainty to business cycle estimation error.

<sup>35</sup>They report NS ratios typically in the range 0.5 to 1.0, although they found much smaller revisions using the univariate Beveridge-Nelson decomposition. Jönsson (2018) examines revisions from the new technique proposed by Hamilton (2018).

<sup>36</sup>See Cogley and Nason (1995).

<sup>37</sup>They further argue that excessive pessimism about potential output during this period led to excessive fiscal austerity, which in turn tended to reinforce the underestimation of potential output.

<sup>38</sup>For example, see the survey by Clements and Galvao in this volume.



## References

- Abeln, Barend, Jan P. A. M. Jacobs, and Pim Ouwehand.** 2018. “CAMPLET: Seasonal Adjustment Without Revisions.” *Journal of Business Cycle Research*.
- Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J Klenow, and Huiyu Li.** 2017. “Missing growth from creative destruction.” National Bureau of Economic Research.
- Anderson, R.G., and K.L. Kliesen.** 2006. “The 1990s acceleration in labor productivity: causes and measurement.” *Federal Reserve Bank of St. Louis Review*, 88(3): 181202.
- Banbura, Marta, Domenico Giannone, Michele Modugno, and Lucrezia Reichlin.** 2013. “Now-casting and the real-time data flow.” *Handbook of economic forecasting*, 2(Part A): 195–237.
- Bank of Ireland.** 2016. “Quarterly Bulletin.” Bank of Ireland 3.
- Bognanni, M., and J. Zito.** 2016. “New Normal or Real-Time Noise? Revisiting the Recent Data on Labor Productivity.” *Economic Commentary*, , (16).
- Carey, Glenn-Marie Lange Quentin Wodon Kevin.** 2018. *The Changing Wealth of Nations 2018: Building a Sustainable Future*. The World Bank.
- Cogley, Timothy, and James M Nason.** 1995. “Effects of the Hodrick-Prescott filter on trend and difference stationary time series Implications for business cycle research.” *Journal of Economic Dynamics and control*, 19(1-2): 253–278.
- Coibion, Olivier, Yuriy Gorodnichenko, and Mauricio Ulate.** 2017. “The cyclical sensitivity in estimates of potential output.” National Bureau of Economic Research.
- Croushore, Dean.** 2011. “Frontiers of real-time data analysis.” *Journal of economic literature*, 49(1): 72–100.
- Croushore, Dean, and Tom Stark.** 2001. “A real-time data set for macroeconomists.” *Journal of Econometrics*, 105: 111–130.

- Croushore, Dean, and Tom Stark.** 2003. “A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter?” *The Review of Economics and Statistics*, 85(3): 605–617.
- Cunningham, Alastair, Jana Eklund, Chris Jeffery, George Kapetanios, and Vincent Labhard.** 2012. “A State Space Approach to Extracting the Signal From Uncertain Data.” *Journal of Business & Economic Statistics*, 30(2): 173–180.
- Dovern, J., and C. Zuber.** 2017. “Recessions and Instable Estimates of Potential Output.” Department of Economics, Heidelberg University 639.
- Faust, Jon, John H Rogers, and Jonathan H Wright.** 2005. “News and Noise in G-7 GDP Announcements.” *Journal of Money, Credit and Banking*, 37(3): 403–419.
- Findley, David F., Brian C. Monsell, William R. Bell, Mark C. Otto, and Bor-Chung Chen.** 1998. “New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program.” *Journal of Business & Economic Statistics*, 16(2): 127–152.
- Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin.** 2000. “The Generalized Dynamic-Factor Model: Identification and Estimation.” *The Review of Economics and Statistics*, 82(4): 540–554.
- Garratt, Anthony, James Mitchell, and Shaun P Vahey.** 2014. “Measuring output gap nowcast uncertainty.” *International Journal of Forecasting*, 30(2): 268–279.
- Giannone, Domenico, Lucrezia Reichlin, and David Small.** 2008. “Nowcasting: The real-time informational content of macroeconomic data.” *Journal of Monetary Economics*, 55(4): 665–676.
- Gómez, Victor, and Agustín Maravall.** 2000. “Seasonal Adjustment and Signal Extraction Time Series.” *A Course in Time Series Analysis*, 202–246.
- Hamilton, James D.** 2018. “Why You Should Never Use the Hodrick-Prescott Filter.” *The Review of Economics and Statistics*, 0(0): null.

- Hamilton, Kirk, and Michael Clemens.** 1999. “Genuine savings rates in developing countries.” *The World Bank Economic Review*, 13(2): 333–356.
- Jacobs, Jan PAM, and Simon van Norden.** 2011. “Modeling data revisions: Measurement error and dynamics of true values.” *Journal of Econometrics*, 161(2): 101–109.
- Jacobs, Jan P.A.M., and Simon van Norden.** 2016. “Why are initial estimates of productivity growth so unreliable?” *Journal of Macroeconomics*, 47: 200–213.
- Jönsson, K.** 2018. “Real-time properties of a regression-based filter.” National Institute of Economic Research, Stockholm, Sweden.
- Jorgenson, Dale W.** 2018. “Production and Welfare: Progress in Economic Measurement.” *Journal of Economic Literature*, 56(3): 867–919.
- Kapetanios, George, and Tony Yates.** 2010. “Estimating time variation in measurement error from data revisions: an application to backcasting and forecasting in dynamic models.” *Journal of Applied Econometrics*, 25(5): 869–893.
- Kishor, N.K., and E.F. Koenig.** 2012. “VAR Estimation and Forecasting When Data Are Subject to Revision.” *Journal of Business & Economic Statistics*, 30: 181–190.
- Kozicki, Sharon, and Barak Hoffman.** 2004. “Rounding Error: A Distorting Influence on Index Data.” *Journal of Money, Credit and Banking*, 36: 319–338.
- Kuang, Pei, and Kaushik Mitra.** 2018. “Potential Output Pessimism and Austerity in the European Union.”
- Kurmann, André, and Eric Sims.** 2017. “Revisions in utilization-adjusted TFP and robust identification of news shocks.” National Bureau of Economic Research.
- Moulton, Brent R., and Benjamin D. Cowan.** 2016. “Residual Seasonality in GDP and GDI: Findings and Next Steps.” *Survey of Current Business*, 96(7): 1–6.
- Nordhaus, W.D., and J. Tobin.** 1972. “Is growth obsolete?” In *Economic Research: Retrospect and prospect, Volume 5, Economic growth*. 1–80. NBER.

- Orphanides, Athanasios, and Simon van Norden.** 2002. “The Unreliability of Output Gap Estimates in Real Time.” *Review of Economics and Statistics*, 84: 569–583.
- Pettenuzzo, Davide, Riccardo Sabbatucci, and Allan Timmermann.** 2018. “High-Frequency Cash Flow Dynamics.” UC San Diego Working Papers.
- Robinson, Lee.** 2016. “Rewriting history: Understanding revisions to UK GDP.” *Bank Underground*.
- Sargent, Thomas J.** 1989. “Two Models of Measurements and the Investment Accelerator.” *Journal of Political Economy*, 97(2): 251–287.
- Stiglitz, Joseph, Amartya Sen, Jean-Paul Fitoussi, et al.** 2009. “The measurement of economic performance and social progress revisited.” *Reflections and overview. Commission on the Measurement of Economic Performance and Social Progress, Paris*.
- Stock, James H, and Mark W Watson.** 2002. “Forecasting using principal components from a large number of predictors.” *Journal of the American statistical association*, 97(460): 1167–1179.
- World Bank.** 2006. *Where is the Wealth of Nations? Measuring Capital for the Twenty-First Century*. World Bank.
- World Bank.** 2011. *The Changing Wealth of Nations: Measuring Sustainable Development in the New Millennium*. World Bank.
- Wright, Jonathan H.** 2013. “Unseasonal seasonals?” *Brookings Papers on Economic Activity*, 2013(2): 65–126.
- Wright, Jonathan H.** 2018. “Seasonal Adjustment of NIPA data.” National Bureau of Economic Research.
- Youll, Robin.** 2008. “Dealing with potential bias in early estimates of GDP.” *Economic and Labour Market Review*, 2(3): 48–52.
- Zwijnenburg, Jorrit.** 2015. “Revisions of quarterly GDP in selected OECD countries.” *OECD Statistics Brief*, (22): 1–12.