Aggregate Output Measurements

A Common Trend Approach*

Martín Almuzara† Gabriele Fiorentini‡ Enrique Sentana§

This version: January 2023

[Link to the latest version]

Abstract

We analyze a model for N different measurements of a persistent latent time series when measurement errors are mean-reverting, which implies a common trend among measurements. We study the consequences of overdifferencing, finding potentially large biases in maximum likelihood estimators of the dynamics parameters and reductions in the precision of smoothed estimates of the latent variable, especially for multiperiod objects such as quinquennial growth rates. We also develop an R^2 measure of common trend observability that determines the severity of misspecification. Finally, we apply our framework to US quarterly data on GDE and GDI, obtaining an improved aggregate output measure.

Keywords: Cointegration, GDE, GDI, Overdifferencing, Signal Extraction.

JEL Classification: C32, E01.

*We are grateful to Dante Amengual, Borağan Aruoba, Peter Boswijk, Frank Diebold, Ulrich Müller, Tommaso Proietti, Richard Smith, and Mark Watson for useful discussions, as well as audiences at Princeton, ESCoE (London 2018), ESEM (Köln 2018), Computational and Financial Econometrics (Pisa 2018) and SAEe (Madrid 2018). Yoosoon Chang and two anonymous referees have also provided very useful feedback. Of course, the usual caveat applies. The first and third authors acknowledge financial support from the Santander Research chair at CEMFI, while the second one is grateful to MIUR through the PRIN project "High-dimensional time series for structural macroeconomic analysis in times of pandemic". The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

†Federal Reserve Bank of New York: martin.almuzara@ny.frb.org

‡Universitá di Firenze and RCEA: gabriele.fiorentini@unifi.it

§CEMFI and CEPR: sentana@cemfi.es

1 Introduction

Aggregate measurements, particularly those of output, are a key input to research economists and policy makers. Assessing the state of business cycles, making predictions of future economic activity, and detecting long-run trends in national income are some of their most popular uses. These measurements are typically regarded as noisy estimates of the quantities of interest, but accounting for the role of measurement error in applications is a difficult task. An important exception arises when more than one measurement of the same quantity is available. This makes it possible to combine the different measurements to produce a better estimate, ideally assigning higher weights to more precise ones.

In the US, the Bureau of Economic Analysis (BEA) reports both the expenditure-based Gross Domestic Expenditure (GDE) measure of output and its income-based Gross Domestic Income (GDI) counterpart. If the sources and methods of the statistical office were perfect, then the two would be identical. In practice, however, they differ (see Landefeld, Seskin, and Fraumeni (2008) for a review). The frequent, and at times noticeable, discrepancy between them (officially known as *statistical discrepancy*) has been recently the subject of active debate in academic and policy circles,[1] and various proposals for improved measures of economic activity have been discussed (see, e.g. Nalewaik (2010), Nalewaik (2011), Greenaway-McGrevy (2011), and Aruoba, Diebold, Nalewaik, Schorfheide, and Song (2016)).[2] The *GDPplus* measure of Aruoba et al. (2016), for example, is currently released on a monthly schedule by the Federal Reserve Bank of Philadelphia.

In this paper, we propose improved output measures under the assumption that alternative measurements in levels do not systematically diverge from each other over the long run. While economic activity, like several other macro aggregates, arguably displays a strong stochastic trend, one would expect statistical discrepancies to mean-revert. In that case, measurements in levels would share a common trend. Somewhat surprisingly, though, the standard practice is to rely on models that do not impose this common trend, working instead with the growth rates of measurements. To cite a few references, Smith et al. (1998), Nalewaik (2010), Nalewaik (2011), Greenaway-McGrevy (2011) and Aruoba et al. (2016) all apply signal extraction techniques to a model of the first differences of log GDP. So does the literature on GDP data revisions, e.g., Aruoba (2008), Jacobs and van Norden (2011) and **?**.

[1]See Grimm (2007) for a detailed methodological insight.

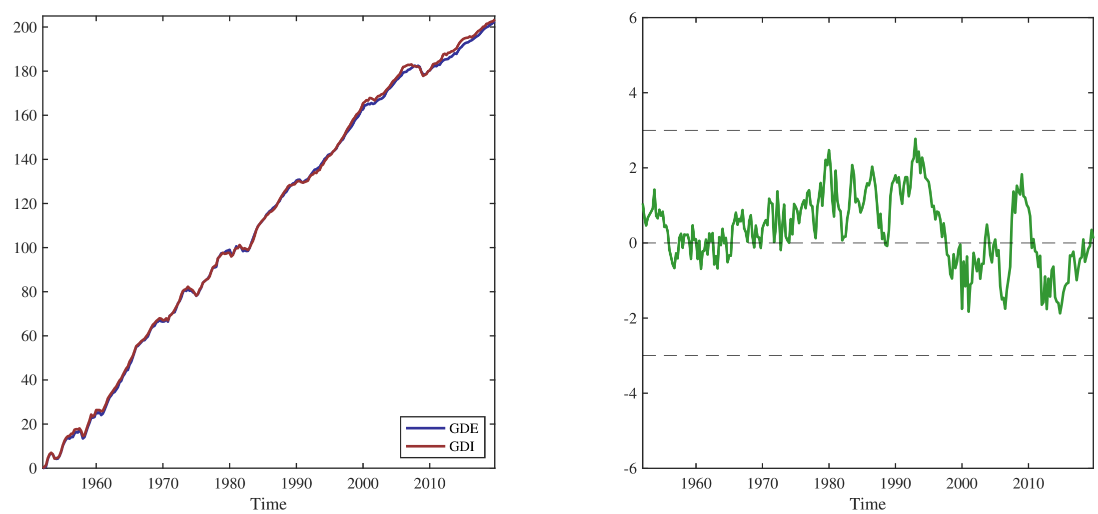
[2]Stone, Champernowne, and Meade (1942) is the first known reference to the signal-extraction framework of our paper. Early literature is surveyed in Weale (1992). See also Smith, Weale, and Satchell (1998).

In this respect, our main goal is to explore the implications of neglecting a common trend in levels for both parameter estimators and smoothed estimates of latent variables. Specifically, we follow Smith et al. (1998) in analyzing a model in which N different measurements y_t of an unobserved quantity x_t are available, so that

$$y_t = x_t 1_{N\times 1} + v_t,$$

with v_t denoting measurement errors in levels and $1_{N\times 1}$ a vector of N ones. In contrast to the literature, though, we model x_t as I(1) – i.e., Δx_t is stationary and strictly invertible – but v_t as I(0). The discrepancies between measurements $y_{it} - y_{jt} = v_{it} - v_{jt}$ are thus cointegrating relationships, reflecting that mean reversion keeps alternative measurements from diverging. As a result, the measurement errors in first differences, Δv_t, are overdifferenced.

Figure 1 shows US data counterparts to y_t and $y_{it} - y_{jt}$:



(a) Output measurements in levels

(b) Statistical discrepancy

FIGURE 1. GDE and GDI. We use November 2020's release of BEA national accounts estimates spanning 1952Q1-2019Q4; (a) 100×logs of GDE and GDI subtracting their 1952Q1 values, i.e., percentage (log) growth in GDE and GDI accumulated since 1952Q1; (b) Differences between 100×log of GDE and 100×log of GDI.

The parameters that describe the dynamics of x_t are typically of interest in themselves, as they inform important dimensions of business cycles and enter signal-extraction calculations. For that reason, we begin by studying the effects of ignoring cointegration among the elements of y_t on estimation procedures. We focus on Gaussian maximum likelihood estimators (MLE) in a simple parametric setup in which the model for x_t is correctly specified but that of v_t is not because of the neglected common trend. Our main finding is that even if x_t and v_t are stochastically independent, estimators of the autocorrelation parameters of x_t will be affected by misspecification in the dynamics of v_t, displaying potentially large biases and increased asymptotic variances. At the same time, we show that if the statistical model assumes Gaussian autoregressive dynamics for both Δx_t and Δv_t, then the estimators of their unconditional means and variances will be asymptotically unaffected. Consequently, the impact of misspecification will be confined to the autocorrelation structure of Δx_t.

Moreover, we prove that the extent to which inferences will be impaired is governed by (i) the severity of overdifferencing in measurement errors, and (ii) the overall signal-to-noise ratio. The more severely overdifferenced the elements of Δv_t are (i.e., the further away from unit root processes those measurement errors are), the stronger the dynamic misspecification resulting from the omitted common trend will be. In addition, a low degree of signal observability, which we quantify by means of an R^2 measure of the relative contribution of x_t and v_t to the variation in observables, amplifies the role of incorrect modeling assumptions on v_t. In the limiting case of $R^2 = 1$, x_t is observable and misspecification in v_t inconsequential.[3] Our results therefore complement those in Chang, Miller, and Park (2009), who derive the asymptotic distribution of the Gaussian MLE in a dynamic factor model with a single common trend. While Chang et al. (2009) study the case of unknown loadings under correct specification, we focus on the case of known loadings (equal to $1_{N\times 1}$) but subject to the dynamic misspecification induced by overdifferencing.[4]

Prediction, filtering and smoothing of x_t given data on y_t – signal extraction, for short – constitute the other main focus of our paper. Given that the uncertainty of signal extraction calculations does not vanish in large T samples, unlike that of parameter estimators, we study their behavior at the pseudo-true parameter values, i.e., at the probability limits of ML estimators. Thus, we leverage on our estimation results to establish the suboptimality as a signal extraction technique of the Kalman-filter-based methods that neglect the common trend.

We find that the effect of ignoring the common trend is substantially different when signal extraction targets a short-run object and a long-run one. In particular, confidence sets for a long-run object such as an average of Δx_t over a relatively large time span are highly sensitive to even modest amounts of overdifferencing in Δv_t. This result is important because long-run

[3] In unreported simulation experiments, we explore the possibility that biases in parameter estimators may be reduced by means of a flexible model of the serial dependence structure of measurement errors in first differences. Specifically, we model Δv_t as a set of independent univariate AR(p) models with p large. Our analysis suggests that bias reduction is thus possible, but at the expense of significant precision loss. Large-p, large-T double-asymptotics in this context appear to be an interesting (but challenging) avenue for future research.

[4] Another difference with Chang et al. (2009) is that their baseline analysis assumes a random walk common trend while ours assumes it to be ARI(1,1).

objects are relevant to empirical questions about slowly evolving trends in macro variables. One example originates in the recurrent debate about growth deceleration in industrialized economies (e.g., Gordon (2016)). Another instance is the secular stagnation hypothesis, which implies a downward trend in interest rates (e.g., Hansen (1939) and Summers (2015)). Similarly, the apparent secular decline in labor shares (e.g., Kaldor (1957), Blanchard (1997) and Karabarbounis and Neiman (2014)) provides another case in point.

On the empirical side, we fit our proposed common trend model to US data on GDE and GDI. Through standard Kalman smoothing calculations, we obtain an improved measure of economic activity, which we compare to other existing measures in the literature. We then use our improved measure to assess the robustness of a variety of empirical facts on economic activity, involving both short- and long-run objects. Our main findings are the following: (1) point estimates of the serial correlation structure of economic activity appear robust to common trend assumptions, (2) the same seems to be true of point estimates of the quarterly average rate of growth in GDP, but (3) our common trend model gives rise to lower signal extraction uncertainty about economic activity than its competitors. Our third finding is conceptually important because point estimates of latent variables cannot be justified by an appeal to consistency – uncertainty about latent variables remains high regardless of the sample size, implying that such estimates must be accompanied by a measure of their precision. This is particularly important from an empirical point of view because the "putative" precision of estimates of economic activity which do not impose a common trend is so low that no sharp conclusion can be drawn about trends in growth from them. In contrast, our common-trend model provides noticeably more precise inference about such long-run objects.[5]

Of course, whether or not there is a common trend is an empirical question in its own right. The evidence that the statistical discrepancy between US GDE and GDI, although persistent, is mean-reverting is suggestive but not conclusive.[6] Yet, the fact that, absent a common trend, the probability of observing large deviations between different measurements tends to one, lends strong support to our framework in the context of aggregate measurement problems.

The rest of the paper is organized as follows. In section 2 we present the basic setup. Section 3 discusses the properties of maximum likelihood estimators while section 4 is devoted to filtering. We report the results of our empirical analysis in section 5. Finally, section 6 concludes. Additional results are relegated to appendix A and the supplemental material.

[5]Interestingly, Chang et al. (2009) show that there is cointegration between the true series x_t and the smoothed estimates that exploit the correctly specified model, which reinforces the case for imposing the common trend in order to obtain a filter capable of closely tracking the level of the signal.

[6]This is most probably related to the low power attributed to cointegration tests.

Notation. We use $\omega_{t_0:t_1}$ to denote the sequence $\{\omega_t\}_{t=t_0}^{t_1}$. If ω_t is a $d_1 \times d_2$ array for all t, and if it raises no confusion, we also use $\omega_{t_0:t_1}$ to denote the $d_1(t_1 - t_0 + 1) \times d_2$ array obtained by vertical concatenation of the terms of $\{\omega_t\}_{t=t_0}^{t_1}$. Analogously, $\psi_{1:N}$ denotes the column vector $(\psi_1, \ldots, \psi_N)'$. We write $\mathbb{E}_T[\omega_t] = T^{-1}\sum_{t=1}^{T}\omega_t$ for the sample average of $\omega_{1:T}$, $\mathbb{E}[\omega_t]$ for its population counterpart, "$\xrightarrow{p}$" for convergence in probability and "$\Longrightarrow$" for weak convergence.

2 Model

In our setup, the statistical office collects N (log) measurements y_t of an unobserved scalar (log) quantity x_t. Let v_t be the vector of (multiplicative) measurement errors so that, in first differences,

$$\Delta y_t = \Delta x_t 1_{N\times 1} + \Delta v_t, \quad t = 1, \ldots, T. \tag{1}$$

For a sample $\Delta y_{1:T}$, the data generating process is given by the probability distribution $\mathbb{P}$.

Assumption 1. $\mathbb{P}$ *satisfies the following:*

(A) The time series $\Delta x_{0:T}, v_{1,0:T}, \ldots, v_{N,0:T}$ are cross-sectionally independent;

(B) Δx_t is a Gaussian AR(1) process: For some values $\mu_0, \rho_0 \in (-1, 1), \sigma_0 > 0$,

$$\begin{aligned}\Delta x_0 &\sim N(\mu_0, \sigma_0^2),\\ \Delta x_t|\Delta x_{0:(t-1)} &\sim N\Big(\mu_0 + \rho_0(\Delta x_{t-1} - \mu_0), (1-\rho_0^2)\sigma_0^2\Big), \quad t = 1, \ldots, T;\end{aligned}$$

(C) v_{it} is a Gaussian AR(1) process: For some values $\rho_i \in (-1, 1], \sigma_i > 0$,

$$\begin{aligned}v_{i0} &\sim N\left(0, \frac{(1+\rho_i)}{2}\sigma_i^2\right),\\ v_{it}|v_{i,0:(t-1)} &\sim N\left(\rho_i v_{i,t-1}, \frac{(1+\rho_i)}{2}\sigma_i^2\right), \quad t = 1, \ldots, T, \ i = 1, \ldots, N.\end{aligned}$$

Assumptions (A) and (B) are made in essentially every paper in the literature (e.g., Smith et al. (1998), Greenaway-McGrevy (2011), Aruoba et al. (2016), and Almuzara, Amengual, and Sentana (2019)). Independence between Δx_t and measurement errors rules out cyclical patterns in the statistical discrepancy. Although potentially of substantive interest, introducing dependence between Δx_t and v_t or across the v_{it}'s complicates identification of the spectra of latent variables. Similarly, AR(1) dynamics for Δx_t is generally agreed to be a reasonable benchmark for economic activity data. Normality is unnecessary for most of our analysis, but since our focus is on the modeling of measurement errors and the role of dynamic misspecification, we adopt it for ease of exposition.

According to Assumption (B), we can regard $\Delta x_{0:T}$ as a segment from a strictly stationary process $\Delta x_{-\infty:\infty}$,

$$\Delta x_t = (1-\rho_0)\mu_0 + \rho_0 \Delta x_{t-1} + \sqrt{1-\rho_0^2}\sigma_0 \varepsilon_{0t},$$

with $\varepsilon_{0t} \overset{iid}{\sim} N(0,1)$. Our parameterization of the process for the signal ensures that $\mathbb{E}\left[\Delta x_t\right] = \mu_0$ and $\mathrm{Var}\left(\Delta x_t\right) = \sigma_0^2$, which do not depend on ρ_0, thereby separating these unconditional moments from the parameters governing the dynamics of Δx_t. Thus, we can summarize the serial dependence structure of the growth rate by its spectral density

$$f_0(\lambda) = \sigma_0^2 \frac{(1-\rho_0^2)}{(1-\rho_0 e^{i\lambda})(1-\rho_0 e^{-i\lambda})} = \sigma_0^2 \left(\sum_{\ell=-\infty}^{\infty} \rho_0^{|\ell|} e^{i\ell\lambda} \right).$$

Assumption (C) implies Δv_{it} is overdifferenced, the severity of overdifferencing increasing as ρ_i moves away from unity. In fact, Δv_{it} is a strictly noninvertible ARMA(1,1) process, except in the limiting case $\rho_i = 1$, when Δv_{it} becomes white noise. We can view $\Delta v_{i,0:T}$ as a segment from a strictly stationary process $\Delta v_{i,-\infty:\infty}$,

$$\Delta v_{it} = \rho_i \Delta v_{i,t-1} + \sqrt{\frac{(1+\rho_i)}{2}}\sigma_i \Delta\varepsilon_{it},$$

with $\varepsilon_{it} \overset{iid}{\sim} N(0,1)$. We have $\mathbb{E}\left[\Delta v_{it}\right] = 0$ and $\mathrm{Var}\left(\Delta v_{it}\right) = \sigma_i^2$, and the spectral density of Δv_{it} is

$$f_i(\lambda) = \sigma_i^2 \frac{(1+\rho_i)(1-e^{i\lambda})(1-e^{-i\lambda})}{2(1-\rho_i e^{i\lambda})(1-\rho_i e^{-i\lambda})},$$

which vanishes at frequency $\lambda = 0$ if $\rho_i \neq 1$ – an unequivocal symptom of overdifferencing.

When $\rho_i \neq 1$ for all i, the spectral density matrix of Δy_t at $\lambda = 0$ is $f_0(0) 1_{N\times N}$. Therefore, it is singular with finite positive diagonal, implying the cointegration (of rank $N-1$) of y_t. Thus, y_t is driven by a single common trend, x_t, while the statistical discrepancies $d_{ij,t} = y_{it} - y_{jt}$ are cointegrating relationships.[7]

[7] There are $N(N-1)/2$ statistical discrepancies but only $N-1$ of them are linearly independent. For example, all the discrepancies with respect to a fixed measurement j form a basis of the cointegration space. An error-correction

Henceforth, we assume the econometrician formulates a statistical model $\mathcal{P} = \left\{\mathbb{P}_\theta : \theta \in \Theta\right\}$ where $\theta = (\vartheta', \psi'_{1:N})'$ with $\vartheta = (\mu, \rho, \sigma)'$ and $\Theta = \Theta_x \times \Theta_v$, $\Theta_x \subset \mathbb{R} \times (-1, 1) \times \mathbb{R}_{>0}$ and $\Theta_v \subset \mathbb{R}^N_{>0}$. The distribution $\mathbb{P}_\theta$ is such that

(a) The time series $\Delta x_{0:T}, \Delta v_{1,0:T}, \ldots, \Delta v_{N,0:T}$ are cross-sectionally independent;

(b) $\Delta x_0 \sim N(\mu, \sigma^2)$ and $\Delta x_t | \Delta x_{0:(t-1)} \sim N\Big(\mu + \rho(\Delta x_{t-1} - \mu), (1 - \rho^2)\sigma^2\Big)$, $t = 1, \ldots, T$;

(c) $\Delta v_{it} \overset{iid}{\sim} N(0, \psi_i^2)$, $i = 1, \ldots, N$.

From (a) and (b) it follows that the econometrician has correctly specified the model for $\Delta x_{0:T}$ conditional on $\vartheta_0 = (\mu_0, \rho_0, \sigma_0)' \in \Theta_x$, an assumption we maintain in what follows. Similarly, $\sigma_{1:N} \in \Theta_v$. In contrast, the model for the observed data $\Delta y_{1:T}$ is misspecified unless $\rho_i = 1$ for all i. In effect, (c) captures the idea that the econometrician neglects the common trend in y_t caused by the mean reversion of measurement errors because she assumes that $v_t = \sum_{\tau=1}^t \Delta v_\tau + v_0$ is a set of N independent random walks.

To ease the comparisons, the statistical model is also parameterized so that $\mathbb{E}_\theta \left[\Delta x_t\right] = \mu$ and $\mathrm{Var}_\theta \left(\Delta x_t\right) = \sigma^2$, where the subscript θ indicates moments of the assumed distribution, so that the implied spectral density of Δx_t becomes

$$f_\vartheta(\lambda) = \sigma^2 \frac{(1 - \rho^2)}{(1 - \rho e^{i\lambda})(1 - \rho e^{-i\lambda})},$$

which coincides with f_0 at $\vartheta = \vartheta_0$. For measurement errors, $\mathbb{E}_\theta \left[\Delta v_{it}\right] = 0$ and $\mathrm{Var}_\theta \left(\Delta v_{it}\right) = \psi_i^2$. Importantly, the assumed spectral density matrix of Δy_t at $\lambda = 0$ is $f_\vartheta(0) 1_{N\times N} + \mathrm{diag}(\psi^2_{1:N})$, which is nonsingular.

Identification. A statistical model that makes use of assumption (a) attains nonparametric identification of the spectra of latent variables. Given a spectral density matrix $f_{\Delta y}$ for the observables, equation (1) and assumption (a) deliver

$$f_{\Delta y}(\lambda) = f_{\Delta x}(\lambda) 1_{N\times N} + \mathrm{diag}\left[f_{\Delta v}(\lambda)\right],$$

where $f_{\Delta x}$ is the spectral density of Δx_t, $f_{\Delta v}$ is the N-dimensional vector of spectral densities of $\Delta v_{1t}, \ldots, \Delta v_{Nt}$ and $1_{N\times N}$ is a square matrix with N^2 ones. Therefore, the ij-th entry of $f_{\Delta y}(\lambda)$, for any $i \neq j$, equals $f_{\Delta x}(\lambda)$, which subtracted from the diagonal of $f_{\Delta y}(\lambda)$ yields $f_{\Delta v}(\lambda)$. In

representation can be derived along the lines of Chang et al. (2009).

fact, assumption (a) imposes overidentifying restrictions on $f_{\Delta y}$ for $N > 2$, as it implies that the off-diagonal elements of $f_{\Delta y}$ must be equal. Consequently, the joint probability distribution of the time series $\{\Delta x_t, \Delta v_t\}$ is identified under Gaussianity, provided one adds some restrictions on the unconditional means of the latent variables, which are necessary because there are $N+1$ unconditional means but we only observe N measurements. Assumption 1, for example, imposes that the expectation of all measurement errors are zero, which is enough to identify $\mu_0 = \mathbb{E}[\Delta x_t]$ for any $N \geq 1$.

2.1 Observability of the signal: a key parameter

Measures of the relative contributions of signal and noises to variation in observables are often important for understanding the quality of estimation and filtering in unobserved components models. To develop such a measure, we use the idea of minimal sufficient statistic for dynamic factor models in Fiorentini and Sentana (2019). With L the lag operator and $F_i(.)$ the autocovariance generating function of Δv_{it},[8] the Generalized Least Squares (GLS) estimator of Δx_t based on the past, present and future of Δy_t is

$$\Delta y_t^* = \frac{\sum_{i=1}^N F_i^{-1}(L)\Delta y_{it}}{\sum_{i=1}^N F_i^{-1}(L)} = \Delta x_t + \frac{\sum_{i=1}^N F_i^{-1}(L)\Delta v_{it}}{\sum_{i=1}^N F_i^{-1}(L)}.$$

Fiorentini and Sentana (2019) show that Δy_t^*, a one-dimensional linear filter applied to Δy_t, contains all relevant information about Δx_t in Δy_t, in the sense that the application of the Kalman filter to Δy_t^* delivers the same predictions for Δx_t as the Kalman filter applied to Δy_t. We denote the resulting error by Δv_t^*, whose spectral density is given by $f_*(\lambda) = \left(\sum_{i=1}^N f_i^{-1}(\lambda)\right)^{-1}$.

Fiorentini and Sentana (2019) also derive the frequency-domain analogue to Δy_t^*, namely

$$\sum_{t=-\infty}^{\infty} \Delta y_t^* e^{i\lambda t} = \Delta y^*(\lambda) = \Delta x(\lambda) + \frac{\sum_{i=1}^N f_i^{-1}(\lambda)\Delta v_i(\lambda)}{\sum_{i=1}^N f_i^{-1}(\lambda)} = \Delta x(\lambda) + \Delta v^*(\lambda).$$

In this context, we take

$$R^2(\lambda) = \frac{f_0(\lambda)}{f_0(\lambda) + f_*(\lambda)},$$

i.e. the fraction of the variance of $\Delta y^*(\lambda)$ explained by $\Delta x(\lambda)$, as an indicator of the degree of observability of the signal at frequency λ. This measure is useful to gauge the frequencies at

[8] That is $F_j(e^{i\lambda}) = f_j(\lambda)$.

which the effect of misspecification in measurement errors is more severe for inferences about Δx_t. In addition, we can obtain an overall measure of observability of the signal by simply replacing spectral densities by their integrals over $[0, 2\pi]$, which yields

$$R^2 = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_*^2},$$

since $\sigma_0^2 = \int_0^{2\pi} f_0(\lambda)\, d\lambda$ and $\sigma_*^2 = \int_0^{2\pi} f_*(\lambda)\, d\lambda$. Interestingly, this is the usual measure of observability (or signal strength) in the error-in-variable model

$$\frac{\sum_{i=1}^N \sigma_i^{-2} \Delta y_{it}}{\sum_{i=1}^N \sigma_i^{-2}} = \Delta x_t + \frac{\sum_{i=1}^N \sigma_i^{-2} \Delta v_{it}}{\sum_{i=1}^N \sigma_i^{-2}},$$

and coincides with $R^2(\lambda)$ at each frequency λ when Δv_{it} is white noise for each i. Thus, a "more observable" signal is indicated by $R^2(\lambda)$ and R^2 closer to unity.[9] In particular, when one of the measurement error variances is zero, $R^2(\lambda) = 1$ and $R^2 = 1$.

3 Estimation

Let $\hat{\theta}$ be the Gaussian MLE of θ. We give asymptotics for $\hat{\theta}$ in the following setup:

Assumption 2. *As $T \to \infty$, the parameters $\mu_0, \rho_0, \sigma_0, \rho_{1:N}, \sigma_{1:N}$ are held constant.*

Remark. An alternative local embedding in which parameters drift in a $1/\sqrt{T}$-neighborhood of a fixed value can be used with little change as long as the autoregressive roots $\rho_{0:N}$ are bounded away from unity. To keep the exposition focused, though, we do not allow for local-to-unity asymptotics for the persistence of measurement errors. A setup in which $\rho_i = 1 - \varrho_i/T$ with ϱ_i held fixed would capture a situation in which the researcher is uncertain about imposing cointegration because the probability that a unit-root test on the differences $y_{it} - y_{jt}$ rejects the null remains bounded between 0 and 1 as $T \to \infty$ (see, e.g., Cavanagh (1985), Chan and Wei (1987) and Phillips (1987)). Still, our analysis suggests that the difference between unit roots and near unit roots is very relevant for constructing inference for long-run objects (see supplemental appendix SM.C) but not so for estimation.

[9]As an alternative, one could use the signal-noise ratios $q(\lambda) = f_0(\lambda)/f_*(\lambda)$ and $q = \sigma_0^2/\sigma_*^2$. Nevertheless, R^2-type measures are easier to interpret because they are bounded between 0 and 1.

Our main estimation result, whose proof appears in appendix A, is as follows:

Theorem 1. *Let $\{\tilde{\mu}, \tilde{\sigma}, \tilde{\psi}_{1:N}\}$ be the maximum likelihood estimator from the static model (i.e., the model that assumes (a), (b) with $\rho = 0$, and (c)). Similarly, let $\{\hat{\mu}, \hat{\rho}, \hat{\sigma}, \hat{\psi}_{1:N}\}$ be the maximum likelihood estimator from the dynamic model $\mathcal{P}$. Then, under assumptions 1 and 2,*

$$\sqrt{T}\begin{pmatrix} \hat{\mu} - \tilde{\mu} \\ \hat{\sigma} - \tilde{\sigma} \\ \hat{\psi}_{1:N} - \tilde{\psi}_{1:N} \end{pmatrix} = o_p(1)\,.$$

Further, for some B and V,

$$\sqrt{T}(\hat{\rho} - (\rho_0 + B)) \implies N(0, V).$$

Therefore, $\hat{\mu} \xrightarrow{p} \mu_0$, $\hat{\sigma} \xrightarrow{p} \sigma_0$ and $\hat{\psi}_i \xrightarrow{p} \sigma_i$ for all i. The estimators of the unconditional mean and variance parameters of the latent variables obtained from the static and dynamic models are asymptotically normal at the usual rate,[10] and, perhaps more surprisingly, they have the same asymptotic covariance matrix. The consequences of neglecting the common trend are, thus, confined to the autocorrelation structure of Δx_t. A univariate example in supplemental appendix SM.A provides further intuition.

Theorem 1 has many implications. First, one can estimate the model parameters without loss of asymptotic precision in two steps: maximizing the static model log-likelihood for $\{\mu, \sigma, \psi_{1:N}\}$ first, and then the dynamic log-likelihood for ρ after plugging in $\{\tilde{\mu}, \tilde{\sigma}, \tilde{\psi}_{1:N}\}$.[11] Second, the unconditional R^2 measure of signal observability is consistently estimated even if the model is misspecified, unlike its frequency-domain counterpart. Third, the estimator of ρ_0 will typically be inconsistent, and at least when normality holds, it will display higher asymptotic variance than the estimator from the model that correctly imposes the common trend in levels.

We can implicitly characterize the inconsistency term B by means of the spectral condition

$$\int_0^{2\pi} \cos(\lambda)\left(\frac{f_{\bar{\mathfrak{y}}}(\lambda)}{f_{\bar{\mathfrak{y}}}(\lambda) + \sigma_*^2}\right)^2 \left(f_{\bar{\mathfrak{y}}}(\lambda) - f_0(\lambda) + \sigma_*^2 - \tilde{f}(\lambda)\right) d\lambda = 0, \tag{2}$$

[10] If the loadings on the common trend x_t, which we assume are $1_{N\times 1}$, had to be estimated instead, the results in Chang et al. (2009) imply that a linear combination of them would converge to a nonstandard distribution at the rate T along a direction determined by the cointegration space.

[11] In fact, our proof suggests that the asymptotic equivalence between static and dynamic MLEs would survive in the presence of forms of dynamic misspecification other than the one we consider in this paper, and for more general dynamic models when the latent variables follow autoregressive processes but not when they have moving average components.

where $\bar{\vartheta} = (\mu_0, \rho_0 + B, \sigma_0)$, σ_*^2 is defined in section 2, and $\tilde{f}(\lambda) = \sum_{i=1}^N \sigma_i^{-4} f_i(\lambda)/[\sum_{i=1}^N \sigma_i^{-2}]^2$ is the spectrum of $\sum_{i=1}^N \sigma_i^{-2}\Delta v_{it}/\sum_{i=1}^N \sigma_i^{-2}$, i.e., the true error in the GLS minimal sufficient statistic for Δx_t computed under the misspecified model.

In Appendix A we derive a time-domain counterpart to (2), namely:

$$\text{(3)} \quad \mathrm{Cov}\Big(\mathbb{E}_{\bar{\theta}}\left[\Delta x_{t-1}\middle|\Delta y_{-\infty:\infty}\right], \mathbb{E}_{\bar{\theta}}\left[\Delta x_t\middle|\Delta y_{-\infty:\infty}\right]\Big) = \mathrm{Cov}_{\bar{\theta}}\Big(\mathbb{E}_{\bar{\theta}}\left[\Delta x_{t-1}\middle|\Delta y_{-\infty:\infty}\right], \mathbb{E}_{\bar{\theta}}\left[\Delta x_t\middle|\Delta y_{-\infty:\infty}\right]\Big).$$

This means that B adjusts to match two types of covariances between the smoothed values of Δx_{t-1} and Δx_t obtained with the misspecified model: the covariance taken under the data generating process $\mathbb{P}$ and the covariance computed using the misspecified model $\mathbb{P}_{\bar{\theta}}$ evaluated at the pseudo-true value $\bar{\theta}$.

When either $\rho_i = 1$ for all i or $\sigma_i = 0$ for at least one i, we have that $\tilde{f}(\lambda) = \sigma_*^2$ for all λ. As a consequence, one can set $f_{\bar{\vartheta}} = f_0$, which implies a consistent estimator of ρ_0 with $B = 0$. By continuity, the inconsistency term B will be small when the extent of misspecification is small (ρ_i's all close to unity) or when the observability of the signal is high (R^2 close to unity). In contrast, noticeable biases may arise when one moves away from those limiting cases, as we illustrate in the next section.

AR(p) dynamics. Our approach can be easily extended to a model in which Δx_t is an AR(p) process with unconditional mean μ_0, unconditional variance σ_0^2, and AR coefficients $\varrho_{0,1:p}$ and the estimated model correctly specifies the dynamics of Δx_t. Under assumptions 1-(A) and (C), the asymptotic equivalence between $\hat{\mu}, \hat{\sigma}, \hat{\psi}_{1:N}$ and $\tilde{\mu}, \tilde{\sigma}, \tilde{\psi}_{1:N}$ of theorem 1 remains, and so does asymptotic normality of $\sqrt{T}(\hat{\varrho}_{1:p} - \bar{\varrho}_{1:p})$ where $\hat{\varrho}_{1:p}$ is the MLE of $\varrho_{1:p}$ and $\bar{\varrho}_{1:p}$ is the pseudo-true value. These are characterized by a set of spectral conditions analogous to (2), i.e.,

$$\int_0^{2\pi} \cos(\ell\lambda)\left(\frac{f_{\bar{\vartheta}}(\lambda)}{f_{\bar{\vartheta}}(\lambda) + \sigma_*^2}\right)^2 \left(f_{\bar{\vartheta}}(\lambda) - f_0(\lambda) + \sigma_*^2 - \tilde{f}(\lambda)\right) d\lambda = 0, \quad \ell = 1, \ldots, p,$$

and a set of time-domain conditions analogous to (3),

$$\mathrm{Cov}\Big(\mathbb{E}_{\bar{\theta}}\left[\Delta x_{t-\ell}\middle|\Delta y_{-\infty:\infty}\right], \mathbb{E}_{\bar{\theta}}\left[\Delta x_t\middle|\Delta y_{-\infty:\infty}\right]\Big) = \mathrm{Cov}_{\bar{\theta}}\Big(\mathbb{E}_{\bar{\theta}}\left[\Delta x_{t-\ell}\middle|\Delta y_{-\infty:\infty}\right], \mathbb{E}_{\bar{\theta}}\left[\Delta x_t\middle|\Delta y_{-\infty:\infty}\right]\Big)$$

for $\ell = 1, \ldots, p$. Some numerical experiments (available upon request) suggest that in this case the roots $\bar{\phi}_{1:p}$ that satisfy $\prod_{\ell=1}^p (1 - \bar{\phi}_\ell z) = (1 - \bar{\varrho}_1 z - \cdots - \bar{\varrho}_p z^p)$ are subject to downward bias relative to the true roots $\phi_{0,1:p}$ that satisfy $\prod_{\ell=1}^p (1 - \phi_{0,\ell} z) = (1 - \varrho_{0,1} z - \cdots - \varrho_{0,p} z^p)$.

3.1 Numerical and simulation evidence

We complement our foregoing discussion of estimation with some insights from numerical and simulation calculations. To begin with, we compute expression (2) by numerical quadrature to obtain the inconsistency in the estimation of ρ_0 as a function of the observability of the signal and the severity of overdifferencing. We set $\mu_0 = 3$, $\rho_0 = 0.5$ and $\sigma_0 = 3.25$.[12] We also take $N = 2$ and let R^2 (with $\sigma_1 = \sigma_2$) and $\rho_1 = \rho_2$ vary over the interval $(0, 1)$.

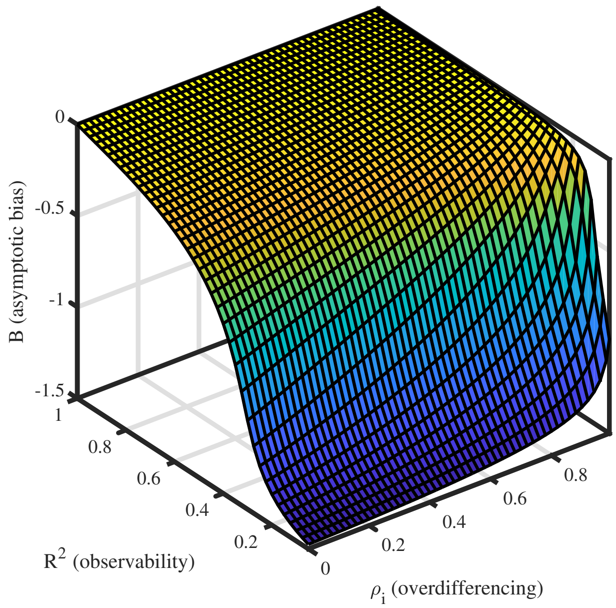


FIGURE 2. Numerical computation of asymptotic bias B in the estimation of ρ_0 for different extents of overdifferencing $\rho_1 = \rho_2$ and signal observability R^2. The true value is $\rho_0 = 0.5$. The integral in (2) is approximated by numerical quadrature with a fine grid on the interval $[0, 2\pi]$.

We display the results for this exchangeable design in figure 2. They clearly confirm our intuition about the roles of ρ_i and R^2 in determining B, with the inconsistency growing quickly as R^2 decreases below 0.5 even for moderate amounts of overdifferencing. Importantly, we always find that $B \leq 0$ under the form of misspecification we analyze in this paper. The rationale is as follows. Equation (2) shows that $\rho_0 + B$ is set to match a weighted average of the difference between $f_{\bar{\vartheta}}$ and $f_0 + \sigma_*^2 - \tilde{f}$, which is depressed at lower frequencies compared to the true spectrum f_0 by the effect of overdifferencing. To see this, note that since f_ϑ is an AR(1) spectrum, lower values of f_ϑ at low frequencies with σ fixed at σ_0 require decreasing ρ. In

[12] Since they represent an affine transformation of the data, the parameters μ_0 and σ_0 (given R^2) are irrelevant for both B and the finite-sample behavior of the ML estimators. Nevertheless, we choose the values of these parameters to match estimates from US quarterly data on economic activity for the period 1952Q1-2019Q4, so that our simulated data resembles the actual dataset in our empirical application. Other sample periods usually lead to different estimates of μ_0 and σ_0 but leave ρ_0 and measures of overdifferencing and signal observability practically unchanged.

particular, at frequency $\lambda = 0$, we have $f_\vartheta(0) = (1+\rho)/(1-\rho)$ which decreases with ρ and, more generally, $\frac{\partial f_\vartheta(\lambda)}{\partial \rho} = -2 f_\vartheta(\lambda)\left[\frac{\rho}{1-\rho^2} + \frac{\rho - \cos(\lambda)}{1+\rho^2 - 2\rho\cos(\lambda)}\right]$, which is negative for low values of λ. Hence, $\text{plim}_{T\to\infty}\, \hat{\rho} = \rho_0 + B < \rho_0$.[13]

We next present simulation evidence on the finite-sample properties of the following three estimators of θ: (i) maximum likelihood for the model in first differences (i.e., $\hat{\theta}$), (ii) the two-step procedure suggested by theorem 1, and (iii) maximum likelihood for the model in levels. The results are summarized in tables 1, 2 and 3. They show that the approximation in theorem 1 works very well in realistic sample sizes and setups. The correlation between $\hat{\theta}$ and the two-step estimator is virtually one, as one would expect from their asymptotic equivalence, and the inconsistency in $\hat{\rho}$ is close to the values for B obtained from equation (2). Not surprisingly, the model in levels outperforms its competitors, although not by much for unconditional moments. The results for a second symmetric design in which $\rho_1 = \rho_2 = 0$ and for an asymmetric design, which we present in appendix SM.B, display the same patterns.

Remark. The behavior of B as ρ_i approaches unity for fixed R^2 can be obtained from figure 2. Our calculations suggest that $\sqrt{T}|B| = o(1)$ when $\rho_i = 1 - \varrho_i/T$ for all i, and that $\sqrt{T}|B| = O(1)$ would require the alternative embedding $\rho_i = 1 - \varrho_i/\sqrt{T}$ instead. Such an embedding would allow us to pretest the existence of a bias in the estimation of ρ_0. Although we do not formally prove these statements, they convey a sense of the relevance of estimation biases in applications. Note that if Δx_t were observable, the standard error of $\hat{\rho}$ for a sample of seventy years of quarterly data ($T = 280$) would be roughly $\sqrt{(1-\rho_0^2)/T} \approx 0.05$. If, for example, the data were generated from the common-trend model with parameters $(\rho_i, R^2) = (0.35, 0.85)$, $(\rho_i, R^2) = (0.92, 0.50)$ or $(\rho_i, R^2) = (0.98, 0.30)$, then the estimation of the model in differences would yield a bias of size comparable to the standard error. These values seem plausible for a large number of applications. In fact, when R^2 is 0.5 or below, values of ρ_i which are only slightly below unity can already cause severe downward bias in the estimation of ρ_0.

[13] Another way to look at the fact that $B \le 0$ is that over-differencing makes f_i be close to zero for frequencies λ which are near zero (f_i is real and continuous), and so

$$\tilde{f}(\lambda) = \frac{\sum_{i=1}^N \sigma_i^{-4} f_i(\lambda)}{\left[\sum_{i=1}^N \sigma_i^{-2}\right]^2} < \frac{\sum_{i=1}^N \sigma_i^{-4}}{\left[\sum_{i=1}^N \sigma_i^{-2}\right]^2} = \sigma_*^2$$

for small λ. As a result, $f_0(\lambda) + \left[\tilde{f}(\lambda) - \sigma_*^2\right] < f_0(\lambda)$ for small λ. A pseudo-true value $\rho_0 + B$ lower than ρ_0 is needed to match $f_\vartheta(\cdot)$ with $f_0(\cdot) + \left[\tilde{f}(\cdot) - \sigma_*^2\right]$. We thank a referee for providing this alternative argument.

TABLE 1. Monte Carlo simulation for $\rho_1 = \rho_2 = 0.85$ and $R^2 = 0.30$.

		True	Differences	Two-step	Levels
μ_0	mean	3	3.001	3.001	3.002
	stderr		0.344	0.344	0.345
	corr			1	0.998
ρ_0	mean	0.5	0.289	0.292	0.481
	stderr		0.204	0.195	0.178
	corr			0.942	0.451
σ_0	mean	3.25	3.233	3.194	3.204
	stderr		0.558	0.588	0.598
	corr			0.96	0.825
ρ_i	mean	0.85			0.831
	stderr				0.072
σ_i	mean	7.021	6.995	7.01	6.996
	stderr		0.385	0.391	0.384

NOTES. Number of samples is $n_{\text{MC}} = 2,000$, sample size is $T = 280$, and parameter values are given under column "True". Rows "mean" and "stderr" show mean and standard deviation across simulations of each estimator; "corr" shows the correlation with MLE in differences of the other two estimators. The bias in $\hat{\rho}$ is to be compared with the theoretical inconsistency $B \approx -0.23$ computed from equation (2) as indicated in the text.

4 Signal extraction

In general, neglecting the common trend should negatively impact filtered and smoothed estimates of the latent variables. We can identify two channels through which this happens: one important for short-run calculations, and the other for long-run calculations. We begin with the short-run channel, i.e., the downward bias in $\hat{\rho}$.

Consider the filtered estimate of Δx_t,

$$\Delta \hat{x}_t = \mathbb{E}_{\hat{\theta}}\left[\Delta x_t \middle| \Delta y_{1:T}\right].$$

As is well-known, the filtering error $\Delta \hat{x}_t - \Delta x_t$ is $O_p(1)$ for large T. This is in contrast to the estimation error $\hat{\theta} - \bar{\theta}$, with $\bar{\theta} = (\mu_0, \rho_0 + B, \sigma_0, \sigma_{1:N}')'$, which is $o_p(1)$. Therefore, we can obtain a good approximation to the behavior of $\Delta \hat{x}_t - \Delta x_t$ if we simply abstract from estimation uncertainty and focus on filtered estimates at pseudo-true values,

$$\Delta \bar{x}_t = \mathbb{E}_{\bar{\theta}}\left[\Delta x_t \middle| \Delta y_{1:T}\right] = \mu_0 + \sum_{\tau=1}^{T} \bar{\phi}_{\tau,T}(\Delta y_\tau - \mu_0 1_{N\times 1}),$$

where the conditional expectation is affine because of the normality assumptions in (b)-(c).

On the other hand, the ideal filter from a mean-square error perspective is the conditional

TABLE 2. Monte Carlo simulation for $\rho_1 = \rho_2 = 0.85$ and $R^2 = 0.50$.

		True	Differences	Two-step	Levels
μ_0	mean	3	3.002	3.002	3.002
	stderr		0.341	0.341	0.341
	corr			1	0.999
ρ_0	mean	0.5	0.41	0.408	0.488
	stderr		0.111	0.11	0.105
	corr			0.986	0.876
σ_0	mean	3.25	3.229	3.22	3.221
	stderr		0.322	0.321	0.326
	corr			0.979	0.954
ρ_i	mean	0.85			0.822
	stderr				0.089
σ_i	mean	4.596	4.583	4.589	4.581
	stderr		0.266	0.272	0.268

NOTES. Number of samples is $n_{\text{MC}} = 2,000$, sample size is $T = 280$, and parameter values are given under column "True". Rows "mean" and "stderr" show mean and standard deviation across simulations of each estimator; "corr" shows the correlation with MLE in differences of the other two estimators. The bias in $\hat{\rho}$ is to be compared with the theoretical inconsistency $B \approx -0.08$ computed from equation (2) as indicated in the text.

TABLE 3. Monte Carlo simulation for $\rho_1 = \rho_2 = 0.85$ and $R^2 = 0.85$.

		True	Differences	Two-step	Levels
μ_0	mean	3	3.001	3.002	3.001
	stderr		0.342	0.343	0.343
	corr			0.999	0.999
ρ_0	mean	0.5	0.478	0.475	0.49
	stderr		0.062	0.062	0.065
	corr			0.998	0.934
σ_0	mean	3.25	3.231	3.23	3.238
	stderr		0.196	0.196	0.205
	corr			0.999	0.934
ρ_i	mean	0.85			0.781
	stderr				0.24
σ_i	mean	1.931	1.925	1.926	1.914
	stderr		0.148	0.165	0.252

NOTES. Number of samples is $n_{\text{MC}} = 2,000$, sample size is $T = 280$, and parameter values are given under column "True". Rows "mean" and "stderr" show mean and standard deviation across simulations of each estimator; "corr" shows the correlation with MLE in differences of the other two estimators. The bias in $\hat{\rho}$ is to be compared with the theoretical inconsistency $B \approx -0.01$ computed from equation (2) as indicated in the text.

mean under the correctly specified model $\mathbb{P}$,[14]

$$\Delta x_t^* = \mathbb{E}\left[\Delta x_t \middle| \Delta y_{1:T}\right] = \mu_0 + \sum_{\tau=1}^{T} \phi_{\tau,T}^* (\Delta y_\tau - \mu_0 1_{N\times 1}).$$

The discrepancy between the weights $\bar{\phi}_{1:T,T}$ and $\phi_{1:T,T}^*$ is of interest because we can decompose $\Delta\bar{x}_t - \Delta x_t$ into two orthogonal components: (i) the optimal filtering error $\Delta x_t^* - \Delta x_t$, whose variance cannot be reduced any further in the class of measurable functions of $\Delta y_{1:T}$ with bounded second moments, and (ii) the difference between the optimal and suboptimal filters $\Delta\bar{x}_t - \Delta x_t^*$.

To illustrate the consequences for signal extraction of neglecting the common trend in levels, figure 3 provides a comparison of the weights for our baseline calibration when overdifferencing is not so severe ($\rho_1 = \rho_2 = 0.85$) and the degree of observability varies from low ($R^2 = 0.30$) to high ($R^2 = 0.85$). We do so for the two leading signal extraction exercises encountered in practice: the computation of $\Delta\bar{x}_t$ and Δx_t^* for values of t in the middle of the sample, and for $t = T$ (i.e., "nowcasting").

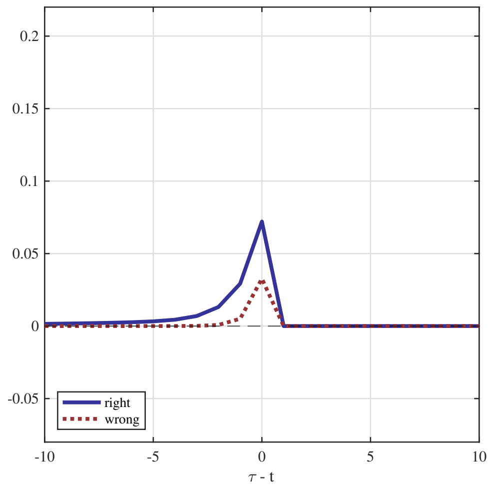
In both cases, it is clear that the filters from misspecified models tend to assign lower weights to nearby observations relative to what is optimal, the difference being larger the lower R^2 is. For the most part, this is explained by the fact that the suboptimal filters assume the signal to be less persistent than it actually is, as B is negative and grows in absolute value as R^2 decreases. Intuitively, the negative value of B resulting from neglecting the common trend leads the econometrician to underestimate the information content of current data.

Naturally, when overdifferencing is more severe, so is its impact on signal extraction. To support this claim, appendix SM.B shows an analogous weight comparison in a design with $\rho_1 = \rho_2 = 0$. For a given R^2, more severe overdifferencing means a larger downward bias in the estimation of the persistence of the signal, which implies even more depressed weights for informative nearby observations.

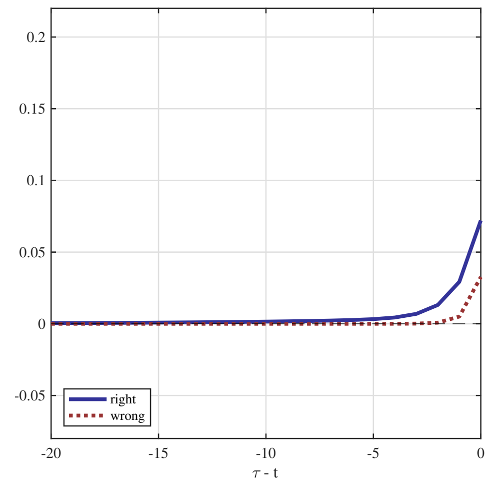
4.1 Long-run objects

By a long-run object we mean a weighted average $X = \sum_{t=1}^{T} \omega_t \Delta x_t$, where the weights $\omega_{1:T}$ satisfy $\|\omega_{1:T}\| = \sqrt{\sum_{t=1}^{T} \omega_t^2} = O\left(1/\sqrt{T}\right)$. They are suitable to quantify trends in aggregate quantities and therefore regularly show up in empirical studies of growth. Compared to smoothed estimates of short-run objects, neglecting the common trend in the level measurements affects

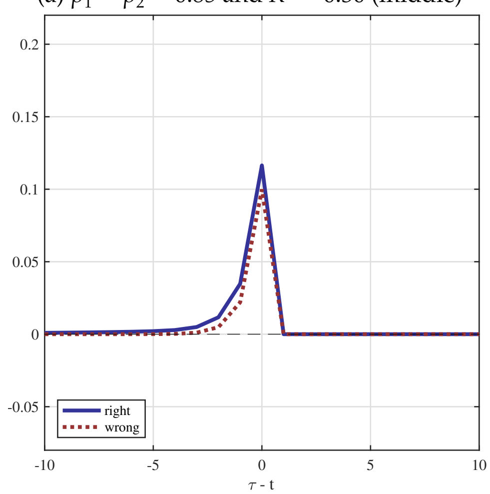
[14] The data in levels enable the use of y_0 in $\mathbb{E}\left[\Delta x_t \middle| \Delta y_{1:T}, y_0\right]$, which dominates $\Delta x_t^* = \mathbb{E}\left[\Delta x_t \middle| \Delta y_{1:T}\right]$ unless $\rho_0 = 0$.



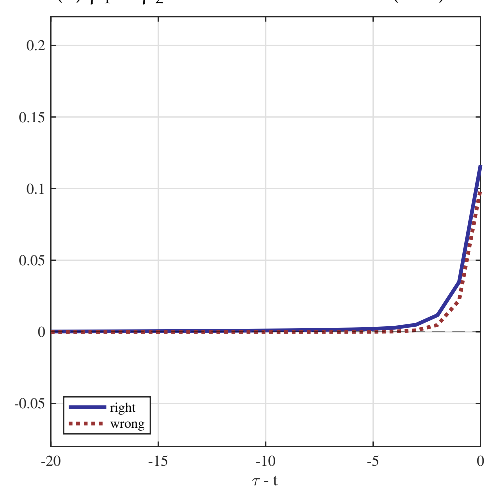
(a) $\rho_1 = \rho_2 = 0.85$ and $R^2 = 0.30$ (middle)



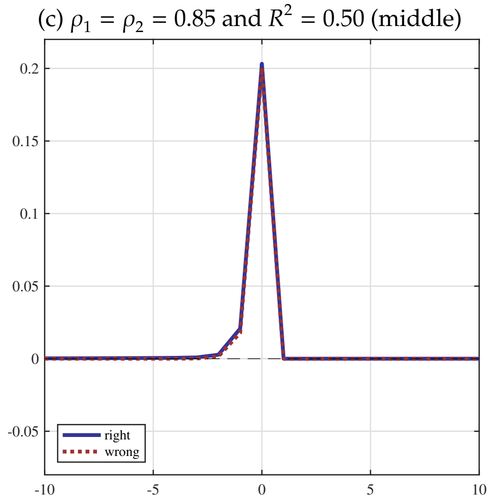
(b) $\rho_1 = \rho_2 = 0.85$ and $R^2 = 0.30$ (end)



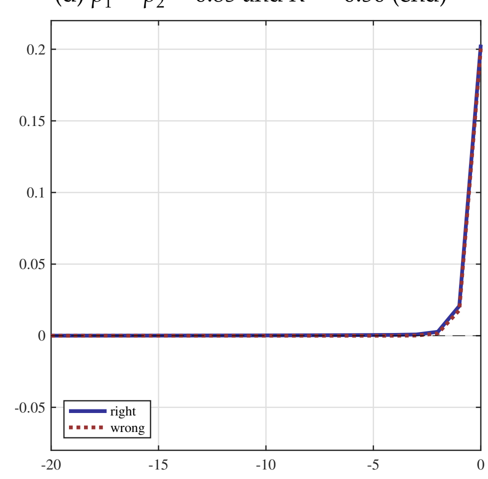
(c) $\rho_1 = \rho_2 = 0.85$ and $R^2 = 0.50$ (middle)



(d) $\rho_1 = \rho_2 = 0.85$ and $R^2 = 0.50$ (end)



(e) $\rho_1 = \rho_2 = 0.85$ and $R^2 = 0.85$ (middle)



(f) $\rho_1 = \rho_2 = 0.85$ and $R^2 = 0.85$ (end)

FIGURE 3. Weights of Kalman smoother. Horizontal axis is $\tau - t$; vertical axis is first entry of $\bar{\phi}_{\tau,T}$ (red) and $\phi^*_{\tau,T}$ (blue). Panels (a), (c) and (e) display weights for $t \approx T/2$ (middle), and panels (b), (d) and (f) for $t = T$ (end). The filters are computed using $\mu_0 = 3$, $\rho_0 = 0.50$, $\sigma_0 = 3.25$. Wrong filter uses $\rho_0 + B$ as AR root with B computed from (2).

inferences about long-run objects though a different channel, namely, by inflating measures of their uncertainty, such as standard errors or confidence intervals. This can be appreciated in the comparison offered in figure 6 in the empirical analysis . We provide a theoretical discussion of this phenomenon in appendix SM.C.

4.2 Simulation evidence (continued)

We compare the finite-sample behavior of the filters discussed above using the same simulation designs as in subsection 3.1 (see tables 1-3). In each simulated sample, we first obtain maximum likelihood estimates of both the misspecified and correctly specified models, and then we compute the corresponding smoothed estimates of Δx_t for $t \approx T/2$ and $t = T$. We present the results for those designs that set $\rho_1 = \rho_2 = 0.85$ in tables 4 and 5. Appendix SM.B reports additional results setting $\rho_1 = \rho_2 = 0$ and $\rho_1 = 0, \rho_2 = 0.95$.

As a general rule, $T = 280$ seems large enough for $\Delta\bar{x}_t$ to provide a good approximation to $\Delta\hat{x}_t$. The same is true for Δx_t^* and the filter from the correctly specified model evaluated at the maximum likelihood estimates, which we call $\Delta\hat{x}_t^*$. The main differences in precision appear between $\Delta\bar{x}_t$ and Δx_t^* rather than between the filters evaluated at the ML estimates and their limiting values.

The effect of neglecting the common trend when measurement errors are highly persistent seems modest in our simulations, with an increase of at most 7% in root MSE relative to the optimal filter in low-R^2 designs. However, more severe overdifferencing combined with a low R^2 leads to a substantial reduction in the precision of filters, as appendix SM.B illustrates. Therefore, researchers should be particularly concerned about their modeling assumptions on measurement error when the R^2 measure we propose in the paper is 0.5 or less, something we already saw in the estimation results.

It is interesting to note that the nowcasting estimate $\Delta\hat{x}_T$ is less affected by misspecification than the smoothed estimate for an observation in the middle of the sample, as one would expect given that the sample is relatively less informative (and therefore receives smaller weights) when filtering Δx_T.

5 An improved aggregate output measure

In this section, we apply our framework to the US quarterly GDE and GDI data displayed in figure 1 with the objective of constructing a new improved measure of economic activity.

TABLE 4. Monte Carlo simulation for $\rho_1 = \rho_2 = 0.85$ and $t \approx T/2$.

		$\Delta\hat{x}_t$	$\Delta\bar{x}_t$	$\Delta\hat{x}^*_t$	Δx^*_t
$R^2 = 0.30$	RMSE	3.15	3.18	3.13	3.12
	increase	0.04	0.04	0.02	
$R^2 = 0.50$	RMSE	3.05	3.05	3.04	3.03
	increase	0.02	0.01	0.02	
$R^2 = 0.85$	RMSE	2.88	2.88	2.87	2.88
	increase	0.01	0	0.01	

NOTES. Number of samples is $n_{\text{MC}} = 2,000$ and sample size is $T = 280$. Columns "$\Delta\hat{x}_t$" and "$\Delta\bar{x}_t$" refer to the wrong filter at the ML estimates and pseudo true values, respectively. Columns "$\Delta\hat{x}^*_t$" and "Δx^*_t" refer to the right filter at the ML estimates and true values, respectively. Root MSE and increase in MSE as a fraction of the MSE of Δx^*_t are indicated for each filter and R^2.

TABLE 5. Monte Carlo simulation for $\rho_1 = \rho_2 = 0.85$ and $t = T$.

		$\Delta\hat{x}_t$	$\Delta\bar{x}_t$	$\Delta\hat{x}^*_t$	Δx^*_t
$R^2 = 0.30$	RMSE	3.12	3.14	3.1	3.08
	increase	0.04	0.04	0.02	
$R^2 = 0.50$	RMSE	3.02	3.01	3.01	3
	increase	0.02	0.01	0.02	
$R^2 = 0.85$	RMSE	2.87	2.87	2.87	2.87
	increase	0.01	0	0.01	

NOTES. Number of samples is $n_{\text{MC}} = 2,000$ and sample size is $T = 280$. Columns "$\Delta\hat{x}_t$" and "$\Delta\bar{x}_t$" refer to the wrong filter at the ML estimates and pseudo true values, respectively. Columns "$\Delta\hat{x}^*_t$" and "Δx^*_t" refer to the right filter at the ML estimates and true values, respectively. Root MSE and increase in MSE as a fraction of the MSE of Δx^*_t are indicated for each filter and R^2.

Specifically, we use the November 2020's release of BEA national accounts estimates for the period 1952Q1-2019Q4 and define $y_{1t} = 400 \times \ln(\text{GDE}_t)$ and $y_{2t} = 400 \times \ln(\text{GDI}_t)$ so that their first differences Δy_{1t} and Δy_{2t} indicate the annualized (geometric) growth rates. The statistical discrepancy, which we compute as $d_{12,t} = (y_{1t} - y_{2t})/4 = 100 \times \ln(\text{GDE}_t/\text{GDI}_t)$, is then roughly the percentage by which GDE exceeds GDI in levels.[15] Remarkably, the levels of GDE and GDI have remained within 3% of each other for about 70 years, lending strong support to our claim that the two measurements are cointegrated.

Table 6 reports maximum likelihood estimates for both the parameters of the common trend model $\mathbb{P}$ and the statistical model $\mathcal{P}$ we discussed in section 3. As expected from Theorem 1, there is no significant disagreement between different estimates of the unconditional moment

[15] We begin the sample in 1952Q1 to coincide with the Treasury-Fed accord. As is well known, this accord established in its modern terms the separation between monetary and fiscal policies, inaugurating a period of more stable behavior of economic aggregates in comparison to the immediate aftermath of World War II. In turn, we end our sample at 2019Q4 to avoid the use of the yet provisional (and highly variable) data from 2020. Thus, all the data in our sample has been subject to at least one annual revision by the BEA.

TABLE 6. Estimates of model parameters for US data.

		Differences	Two-step	Levels
μ_0	estimate	2.994	2.997	2.989
	stderr	(0.338)	(0.204)	(0.338)
ρ_0	estimate	0.488	0.485	0.499
	stderr	(0.057)	(0.048)	(0.057)
σ_0	estimate	3.237	3.227	3.223
	stderr	(0.186)	(0.151)	(0.184)
ρ_1	estimate			-0.097
	stderr			(0.272)
σ_1	estimate	1.49	1.387	1.314
	stderr	(0.115)	(0.149)	(0.130)
ρ_2	estimate			0.941
	stderr			(0.021)
σ_2	estimate	1.113	1.239	1.338
	stderr	(0.137)	(0.162)	(0.113)

NOTES. The sample period is 1952Q1-2019Q4 ($T = 271$). Rows "estimate" and "stderr" show point estimate and standard error for each estimator. A subindex 1 in ρ_i and σ_i refers to GDE while a subindex 2 refers to GDI. The point estimate for signal observability R^2 is 0.922 and a 95% confidence interval for R^2 is $[0.901, 0.943]$.

parameters $\{\mu_0, \sigma_0, \sigma_1, \sigma_2\}$. The estimates of our R^2 measure of common trend observability are high at about 0.92, with a small confidence interval around them. Estimates for ρ_0, in turn, are all near 0.5, with a seemingly small downward bias in the estimators from the models that neglect the common trend. These patterns are in line with the theoretical and simulation analysis in section 3.[16] The estimates of the autoregressive coefficient ρ_2 implies that the time series of GDI's measurement error in levels, v_{2t}, seems stationary but highly persistent. In contrast, we cannot reject that the GDE's measurement error in levels, v_{1t}, is white noise. This difference in the persistence of measurement errors may be the result of the fact that GDE and GDI are computed from different sources and with different methods, therefore relying on inputs which themselves differ in their dynamics. Next, we discuss some of the implications of these results.

Our first consideration is about the serial dependence of the statistical discrepancy, $d_{12,t}$. Figure 4, in particular, shows that our assumption of AR(1) measurement errors in levels does a good job at replicating the autocorrelations of this variable. Although the statistical discrepancy is highly persistent because the GDI's measurement error in level dominates it, it is also evident that the serial dependence steadily declines, being already fairly low after 12 quarters. We also note that, if anything, the autocorrelations of the statistical discrepancy tend to decrease faster in the data than in our model, although the difference is small relative to the sampling

[16] When we restrict our sample to the one used by Aruoba et al. (2016), we obtain estimates of $\{\mu_0, \rho_0, \sigma_0, \sigma_1, \sigma_2\}$ comparable to theirs. For the subsample 1960Q1-2011Q4 that they use, the variance of the signal is slightly lower, and so is the R^2 measure of common trend observability.

uncertainty.

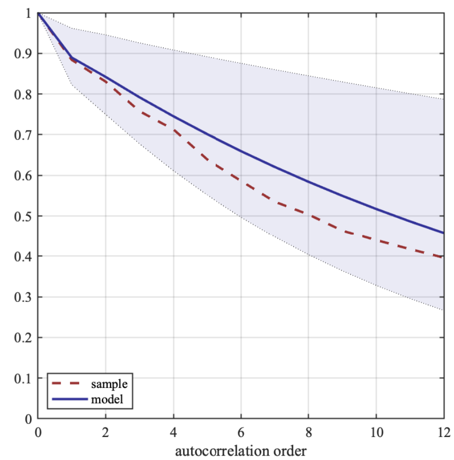
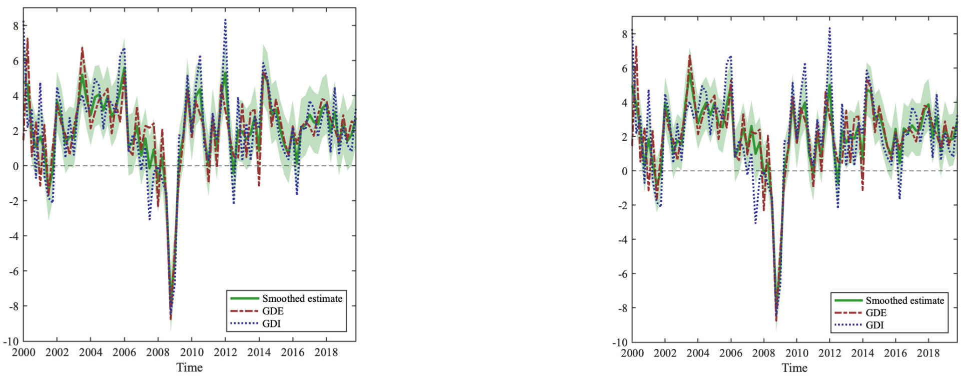


FIGURE 4. Autocorrelations of the statistical discrepancy. The solid blue line contains the autocorrelations implied by model $\mathbb{P}$ at point estimates. Shaded area is a pointwise 95% confidence interval for each lag.

A related observation is that the model in first differences leads to an implausibly high probability of long-run divergence between GDE and GDI in levels. To get a sense for it, $d_{12,T}|d_{12,0} \sim N(d_{12,0}, 0.9 \times T)$ with the estimates of the dynamically misspecified model at hand. A quick calculation indicates that the probability that today we would observe a divergence between GDE and GDI higher than 3% is 0.99 if the two aggregate output measurements were not cointegrated.

The second consideration refers to the impact of neglecting the common trend in levels on inferences about parameters and latent variables. Because Δx_t is highly observable, our theoretical results in sections 3 and 4 lead us to expect no significant divergence between the models in differences and in levels with regards to maximum likelihood estimates and smoothed estimates of what we have called short-run objects. We have already confirmed the similarity of the estimates in table 6. In turn, figure 5 confirms our results for the smoothed estimates of Δx_t, as one can hardly distinguish one model from the other in terms of the conditional mean and variance of Δx_t given $\Delta y_{1:T}$.[17] Still, pointwise confidence intervals are shorter for the model that imposes the common trend: their average length is 3.5% (in annualized growth) for the model in differences against 3% for the model in levels.

[17]In fact, the smoothed series obtained by assuming that GDE and GDI are not cointegrated lies within the credible sets obtained under the assumption that the model that imposes cointegration is correct. Therefore, the difference between the two smoothed series, which is largest in the period preceding the fall of Lehman Brothers but does not show any business cycle variation, is not significant.

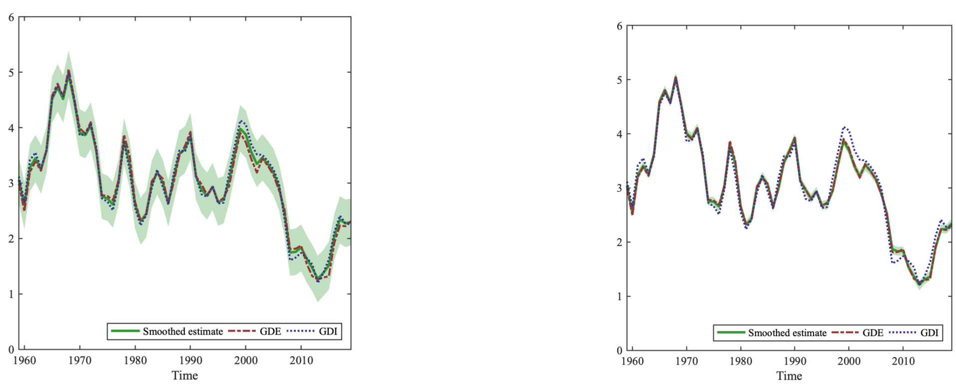


(a) Model in differences

(b) Model in levels

FIGURE 5. Smoothed estimates of Δx_t (short-run object). The solid green line represents the smoothed estimates and the shaded area represents 95% confidence intervals (pointwise for each t).

However, the fact that GDE's measurement error is essentially white noise does affect inferences about long-run objects. Figure 6 illustrates this feature with the 8-year moving averages of Δx_t. We take overlapping 8-year intervals for the purposes of averaging out the typical business-cycle periodicity.[18] As expected from our results in appendix SM.C, there is substantially less uncertainty for the model that exploits the common trend. Specifically, the average length of the confidence intervals is 1% for the model in differences and 0.2% for the model in levels. This reduced uncertainty is particularly important for assessing changes in aggregate trends, as such changes are typically small.



(a) Model in differences

(b) Model in levels

FIGURE 6. Smoothed estimates of $h^{-1} \sum_{\ell=1}^{h} \Delta x_{t-\ell+1}$ ($h = 32$, long-run object). Solid green line represents the smoothed estimates and shaded area represents 99% confidence intervals (pointwise for each t).

[18]In this respect, we follow Müller and Watson (2008) but similar patterns arise when we use 5- or 10-year intervals instead.

Finally, note that our trend estimates track far more closely 8-year moving averages of GDE growth than those of GDI, which is again explained by the low persistence of v_{1t}. Therefore, we may conclude that empirical patterns about economic activity previously obtained from low-frequency averages of GDE are robust to the presence of measurement error in view of the small degree of filter uncertainty implied by our model.

6 Conclusion

From a practical point of view, the first lesson we can extract from our study regarding aggregate output measurement is that the need to account for a common trend in levels hinges critically on how important measurement errors are in driving observable variation. For quarterly or annual data, measurement errors might be small; for monthly and, particularly, for high-frequency data the opposite should be expected. Although no direct measure of economic activity exists at the monthly frequency, nowcasting exercises at high frequencies typically contain larger amounts of measurement errors as they feed on noisier, more preliminary, input data. The nature of what is being measured matters too as different economic concepts have different associated degrees of noisiness. For example, it is not the same to look at the quarterly growth rate of GDP than to look at its quinquennial counterpart. As a practical prescription, we recommend the estimation of the R^2 measure of trend observability we develop in the paper. We also strongly urge researchers to always impose a common trend, especially when the R^2 turns out low – an R^2 below 0.5 should be cause of concern.

Moreover, our econometric analysis yields several insights which are of theoretical interest. First, we prove that estimators of unconditional first and second moments under the misspecified model are asymptotically equivalent to static model estimators. Second, we show that the form of misspecification studied in this paper causes a downward bias in the estimated persistence of the signal. And third, we highlight that the misspecified model will tend to overstate uncertainty of smoothed estimates of latent variables, and dramatically so for long-run objects. Although we derive these results in a simplified parametric model, our methods of analysis allow easy extension to more general setups. In particular, our analysis may be adapted to dynamic factor models with nontrivial cross-sectional dimensions – models in which the usual data preprocessing may likely lead to overdifferencing.

On the empirical side, we construct a new improved measure of US aggregate output from GDE and GDI data. Unlike existing signal-extraction measures, ours allows GDE and

GDI's measurement errors in levels to mean-revert, a property that fits well with the data. Still, given that signal observability is high in this application, our estimates of the parameters of the dynamics of output growth are not affected much by ignoring the stationarity of the measurement errors. Nevertheless, our common trend approach delivers noticeable reductions in the implied uncertainty of smoothed estimates of true output growth. Specifically, measured in terms of root mean square errors, the reductions are around 15% for short-run objects and 80% for long-run ones. One important practical issue that we have neglected in this paper is the regular updating of the GDE and GDI measures by the BEA. In Almuzara, Amengual, Fiorentini, and Sentana (2022), we are currently exploring this relevant research avenue within the common trends framework of this paper.

Appendix A Proof of theorem 1 and derivation of equations (2) and (3)

Although for the sake of brevity, we do not discuss frequency-domain ML estimation (see, e.g., Fiorentini, Galesi, and Sentana (2018)) or Bayesian estimation (e.g., Durbin and Koopman (2012)), before presenting the proof, it is useful to describe the way estimates are produced. We can obtain numerically equivalent Gaussian MLEs of θ, $\hat{\theta}$, by means of two algorithms. The first one exploits the Kalman filter to recursively compute the one-period ahead conditional means and variances of observables appearing in the log-likelihood function. The second is the EM algorithm, which, for some initial $\hat{\theta}^{(0)}$, updates parameter estimates by iterating over

$$\hat{\theta}^{(s)} = \underset{\theta\in\Theta}{\operatorname{argmax}}\, \mathbb{E}_{\hat{\theta}^{(s-1)}}\left[\sum_{t=1}^{T} \ln p_{\vartheta}\left(\Delta x_t|\Delta x_{1:(t-1)}\right) + \sum_{t=1}^{T}\sum_{i=1}^{N} \ln p_{\psi_i}\left(\Delta v_{it}\right)\middle| y_{1:T}\right],$$

$$\ln p_{\vartheta}\left(\Delta x_t|\Delta x_{1:(t-1)}\right) = -\frac{1}{2}\left[\ln\left(2\pi(1-\rho^2)\sigma^2\right) + \frac{(\Delta x_t - (1-\rho)\mu - \rho\Delta x_{t-1})^2}{(1-\rho^2)\sigma^2}\right],$$

$$\ln p_{\psi_i}\left(\Delta v_{it}\right) = -\frac{1}{2}\left[\ln(2\pi\psi_i^2) + \frac{\Delta v_{it}^2}{\psi_i^2}\right].$$

The EM algorithm alternates between smoothing the so-called complete-data likelihood using the current value $\hat{\theta}^{(s-1)}$ for expectation calculations (the E-step), and maximizing the resulting smoothed function to yield a new value $\hat{\theta}^{(s)}$ (M-step). See Dempster, Laird, and Rubin (1977), Ruud (1991), and Watson and Engle (1983). If the algorithm converges, we have $\hat{\theta} = \lim_{s\to\infty}\hat{\theta}^{(s)}$. This EM algorithm is particularly relevant because our proof relies heavily on a generalization of Louis (1982) score formula, which we call EM principle, formalized in Almuzara et al. (2019, th. 1). Consider the functions

$$
\begin{aligned}
g_\mu(\theta) &= \sqrt{T}\left(\frac{\mathbb{E}_T\left[\Delta x_t - \rho\Delta x_{t-1}\right]}{1-\rho} - \mu\right), \\
g_\rho(\theta) &= \sqrt{T}\mathbb{E}_T\left[(\Delta x_{t-1}-\mu)(\Delta x_t - \mu) - \rho(\Delta x_{t-1}-\mu)^2\right], \\
g_\sigma(\theta) &= \sqrt{T}\left(\frac{\mathbb{E}_T\left[((\Delta x_t - \mu) - \rho(\Delta x_{t-1}-\mu))^2\right]}{1-\rho^2} - \sigma^2\right), \\
g_{\psi_i}(\theta) &= \sqrt{T}\left(\mathbb{E}_T\left[\Delta v_{it}^2\right] - \psi_i^2\right), \quad i = 1,\dots,N.
\end{aligned}
\tag{A.1}
$$

These are proportional to the scaled average scores of the complete-data log-likelihood for the misspecified model. Maximum likelihood estimates $\hat{\theta}$ are characterized by the first-order necessary conditions

$$
\mathbb{E}_{\hat{\theta}}\left[g_\mu(\hat{\theta})\middle|\Delta y_{1:T}\right] = \mathbb{E}_{\hat{\theta}}\left[g_\rho(\hat{\theta})\middle|\Delta y_{1:T}\right] = \mathbb{E}_{\hat{\theta}}\left[g_\sigma(\hat{\theta})\middle|\Delta y_{1:T}\right] = \mathbb{E}_{\hat{\theta}}\left[g_{\psi_i}(\hat{\theta})\middle|\Delta y_{1:T}\right] = 0.
$$

Define the auxiliary functions

$$
\begin{aligned}
\tilde{g}_\mu(\theta) &= \sqrt{T}\left(\mathbb{E}_T\left[\Delta x_t\right] - \mu\right), \\
\tilde{g}_\sigma(\theta) &= \sqrt{T}\left(\mathbb{E}_T\left[(\Delta x_t - \mu)^2\right] - \sigma^2\right),
\end{aligned}
$$

and note that the maximum likelihood estimates for the static model $\tilde{\theta}$ (i.e., the restricted maximum likelihood estimates subject to $\rho = 0$) satisfy

$$
\mathbb{E}_{\tilde{\theta}}\left[\tilde{g}_\mu(\tilde{\theta})\middle|\Delta y_{1:T}\right] = \mathbb{E}_{\tilde{\theta}}\left[\tilde{g}_\sigma(\tilde{\theta})\middle|\Delta y_{1:T}\right] = \mathbb{E}_{\tilde{\theta}}\left[g_{\psi_i}(\tilde{\theta})\middle|\Delta y_{1:T}\right] = 0.
\tag{A.2}
$$

The first lemma will allow us to replace g_μ and g_σ by the much simpler $\tilde{g}_\mu$ and $\tilde{g}_\sigma$.

Lemma 1. *Let $\hat{\theta}$ be the maximum likelihood estimator for the misspecified model. Under assumptions 1 and 2,*

$$
\mathbb{E}_{\hat{\theta}}\left[g_\mu(\hat{\theta})\middle|\Delta y_{1:T}\right] = \mathbb{E}_{\hat{\theta}}\left[\tilde{g}_\mu(\hat{\theta})\middle|\Delta y_{1:T}\right] + o_p(1) \text{ and } \mathbb{E}_{\hat{\theta}}\left[g_\sigma(\hat{\theta})\middle|\Delta y_{1:T}\right] = \mathbb{E}_{\hat{\theta}}\left[\tilde{g}_\sigma(\hat{\theta})\middle|\Delta y_{1:T}\right] + o_p(1).
$$

Proof. For any $\theta \in \Theta$,

$$
g_\mu(\theta) - \tilde{g}_\mu(\theta) = \frac{\rho(\Delta x_T - \Delta x_0)}{(1-\rho)\sqrt{T}}.
$$

One can then show that $\mathbb{E}_\theta\left[\Delta x_0 \middle| \Delta y_{1:T}\right] = O_p(1)$ and $\mathbb{E}_\theta\left[\Delta x_T \middle| \Delta y_{1:T}\right] = O_p(1)$, which leads to

$$\mathbb{E}_\theta\left[g_\mu(\theta) - \tilde{g}_\mu(\theta) \middle| \Delta y_{1:T}\right] = O_p\left(1/\sqrt{T}\right).$$

In particular, this implies that $\mathbb{E}_{\hat{\theta}}\left[g_\mu(\hat{\theta}) - \tilde{g}_\mu(\hat{\theta}) \middle| \Delta y_{1:T}\right] = o_p(1)$.

Turning to the score function with respect to σ, for any $\theta \in \Theta$,

$$g_\sigma(\theta) + \frac{2\rho}{1-\rho^2} g_\rho(\theta) - \tilde{g}_\sigma(\theta) = \frac{\rho^2((\Delta x_T - \mu)^2 - (\Delta x_0 - \mu)^2)}{(1-\rho^2)\sqrt{T}}.$$

Since $\mathbb{E}_\theta\left[(\Delta x_0)^2 \middle| \Delta y_{1:T}\right] = O_p(1)$ and $\mathbb{E}_\theta\left[(\Delta x_T)^2 \middle| \Delta y_{1:T}\right] = O_p(1)$,

$$\mathbb{E}_\theta\left[g_\sigma(\theta) + \frac{2\rho}{1-\rho^2} g_\rho(\theta) - \tilde{g}_\sigma(\theta) \middle| \Delta y_{1:T}\right] = O_p\left(1/\sqrt{T}\right).$$

And since $\mathbb{E}_{\hat{\theta}}\left[g_\rho(\hat{\theta}) \middle| \Delta y_{1:T}\right] = 0$, we finally get $\mathbb{E}_{\hat{\theta}}\left[g_\sigma(\hat{\theta}) - \tilde{g}_\sigma(\hat{\theta}) \middle| \Delta y_{1:T}\right] = o_p(1)$. ∎

Remark. A subtlety in the previous proof is that order-in-probability statements refer to $\mathbb{P}$ while expectations refer to $\mathbb{P}_\theta$. For the sake of brevity, we omit the proof that this difference is inconsequential.

Let $\theta_{\rho=0}$ be the parameter vector θ in which we set $\rho = 0$. In an abuse of notation, we will also occasionally identify $\theta_{\rho=0}$ with the subvector that excludes the $\rho = 0$ component – note that, trivially, $\tilde{\theta}$ and $\tilde{\theta}_{\rho=0}$ represent the same parameter value. For the proof of lemma 2, the following remark will be useful:

Remark. Sample spaces of $\Delta y_{1:T}$, $\Delta x_{0:T}$, and $\Delta v_{1:T}$ are $\mathcal{Y} = \mathbb{R}^{NT}$, $\mathcal{X} = \mathbb{R}^{T+1}$, and $\mathcal{V} = \mathbb{R}^{NT}$. Probability distributions $\mathbb{P}$ and $\mathbb{P}_\theta$, $\theta \in \Theta$, may be taken to be measures on the Borel sets of $\mathcal{X} \times \mathcal{V}$ with probability statements about $\Delta y_{1:T}$ interpreted by means of the inverse image of the mapping in (1). The measure $\mathbb{P}$ and the model $\mathcal{P}$ are then dominated by Lebesgue measure λ on the Borel sets of $\mathcal{X} \times \mathcal{V}$. Consequently, densities exist by the Radon-Nikodym theorem. In contrast, conditional distributions of $\Delta x_{0:T}$ and $\Delta v_{1:T}$ given $\Delta y_{1:T}$ implied by $\mathbb{P}$ and $\mathcal{P}$ are dominated by the σ-finite measure $\lambda_{\Delta y_{1:T}}$ on the Borel sets of the hyperplane defined by (1) for fixed $\Delta y_{1:T}$, rather than by $\mathcal{P}$. Therefore, conditional densities exist with respect to $\lambda_{\Delta y_{1:T}}$ in that case too.

Lemma 2. *Under assumptions 1 and 2,*

$$\sqrt{T}\left(\mathbb{E}_{\hat{\theta}}\left[\mathbb{E}_T\left[\Delta x_t\right]\middle|\Delta y_{1:T}\right]-\mathbb{E}_{\hat{\theta}_{\rho=0}}\left[\mathbb{E}_T\left[\Delta x_t\right]\middle|\Delta y_{1:T}\right]\right)=o_p(1)\,,$$
$$\sqrt{T}\left(\mathbb{E}_{\hat{\theta}}\left[\mathbb{E}_T\left[\Delta x_t^2\right]\middle|\Delta y_{1:T}\right]-\mathbb{E}_{\hat{\theta}_{\rho=0}}\left[\mathbb{E}_T\left[\Delta x_t^2\right]\middle|\Delta y_{1:T}\right]\right)=o_p(1)\,,$$
$$\sqrt{T}\left(\mathbb{E}_{\hat{\theta}}\left[\mathbb{E}_T\left[\Delta v_{it}^2\right]\middle|\Delta y_{1:T}\right]-\mathbb{E}_{\hat{\theta}_{\rho=0}}\left[\mathbb{E}_T\left[\Delta v_{it}^2\right]\middle|\Delta y_{1:T}\right]\right)=o_p(1)\,,\quad i=1,\ldots,N.$$

Proof. Let x denote the latent variables $\{\Delta x_{0:T}, v_{1:N,1:T}\}$ and y the observables $\Delta y_{1:T}$. Under the restriction $\rho = 0$, the model for x has density

$$p_\eta(x) = b\exp\left[T\cdot(\eta' S(x) - a(\eta))\right]$$

with respect to measure λ. Similarly, the density of x given y is an exponential family with density

$$p_\eta(x|y) = b\exp\left[T\cdot(\eta' S(x) - a(\eta|y))\right]$$

with respect to measure λ_y. Measures λ and λ_y are defined in the remark above, b is a constant, $\eta = \eta(\mu, \sigma, \psi_{1:N})$ is a function of the original parameters, $a(\cdot)$ and $a(\cdot|y)$ are functions of η, and the sufficient statistics are $S(x) = \mathbb{E}_T\left[(\Delta x_t, \Delta x_t^2, \Delta v_{1t}^2, \ldots, \Delta v_{Nt}^2)'\right]$.

Define $\hat{S} = \mathbb{E}_{\hat{\theta}}\left[S(x)\right]$ and note that if x is such that $S(x) = \hat{S}$, then the densities $p_\eta(x)$ and $p_\eta(x|y)$ are maximized at $\hat{\eta} = \eta(\hat{\mu}, \hat{\sigma}, \hat{\psi}_{1:N})$.

In addition,

$$\hat{S} = \frac{\partial a(\hat{\eta})}{\partial \eta} = \frac{\partial a(\hat{\eta}|y)}{\partial \eta} = \mathbb{E}_{\hat{\theta}_{\rho=0},}\left[S(x)\middle|y\right]$$

where this follows from well-known properties of exponential families (Jørgensen and Labouriau, 2012, Th. 1.17 and 1.18). Since $\mathbb{E}_{\hat{\theta}}\left[S(x)\middle|y\right] = \hat{S} + o_p\left(1/\sqrt{T}\right)$ by virtue of lemma 1, the current lemma immediately follows. ∎

The rest of the argument for the asymptotic equivalence between $\hat{\theta}_{\rho=0}$ and $\tilde{\theta}_{\rho=0}$ is standard. Specifically, if G collects the static model estimating equations (A.2) (suitably scaled by T), and $\bar{\theta}_{\rho=0}$ denotes the (common) probability limit of the two estimators, a Taylor expansion gives

$$o_p(1) = G(\hat{\theta}_{\rho=0}) - G(\tilde{\theta}_{\rho=0}) = \left[H(\bar{\theta}_{\rho=0}) + o_p(1)\right]\times\sqrt{T}(\hat{\theta}_{\rho=0} - \tilde{\theta}_{\rho=0}),$$

where $H(\bar{\theta}_{\rho=0})$ is a fixed nonsingular matrix, which in turn implies $\sqrt{T}(\hat{\theta}_{\rho=0} - \tilde{\theta}_{\rho=0}) = o_p(1)$.

We now turn to characterizing the pseudo-true value for the estimator $\hat{\rho}$. From (A.1) and (a small variation of) lemma 2, we obtain

$$\mathbb{E}_T\left[\mathbb{E}_{\hat{\theta}}\left[(\Delta x_{t-1} - \hat{\mu})(\Delta x_t - \hat{\mu})\middle|\Delta y_{1:T}\right]\right] - \hat{\rho}\hat{\sigma}^2 = o_p(1)\,.$$

Let $\bar{\rho} = \rho_0 + B$ be the probability limit of $\hat{\rho}$. Our discussion of equivalence between dynamic-model and static-model maximum likelihood estimators has already established that $\hat{\mu} \xrightarrow{p} \mu_0$, $\hat{\sigma} \xrightarrow{p} \sigma_0$ and $\hat{\psi}_i \xrightarrow{p} \sigma_i$. Thus, $\hat{\rho}\hat{\sigma}^2 \xrightarrow{p} \bar{\rho}\sigma_0^2$. Also let $\bar{\theta}$ be the probability limit of $\hat{\theta}$.

For any $\theta \in \Theta$, we have

$$\begin{aligned}
\mathbb{E}_\theta\left[(\Delta x_{t-1} - \mu)(\Delta x_t - \mu)\middle|\Delta y_{1:T}\right] &= \mathrm{Cov}_\theta(\Delta x_{t-1}, \Delta x_t|\Delta y_{1:T}) \\
&\quad + (\mathbb{E}_\theta\left[\Delta x_{t-1}|\Delta y_{1:T}\right] - \mu)(\mathbb{E}_\theta\left[\Delta x_{t-1}|\Delta y_{1:T}\right] - \mu) \\
&= \mathbb{E}_\theta\left[\mathrm{Cov}_\theta(\Delta x_{t-1}, \Delta x_t|\Delta y_{1:T})\right] \\
&\quad + (\mathbb{E}_\theta\left[\Delta x_{t-1}|\Delta y_{1:T}\right] - \mu)(\mathbb{E}_\theta\left[\Delta x_{t-1}|\Delta y_{1:T}\right] - \mu) \\
&= \mathrm{Cov}_\theta(\Delta x_{t-1}, \Delta x_t) - \mathrm{Cov}_\theta\Big(\mathbb{E}_\theta\left[\Delta x_{t-1}|\Delta y_{1:T}\right], \mathbb{E}_\theta\left[\Delta x_t|\Delta y_{1:T}\right]\Big) \\
&\quad + (\mathbb{E}_\theta\left[\Delta x_{t-1}|\Delta y_{1:T}\right] - \mu)(\mathbb{E}_\theta\left[\Delta x_{t-1}|\Delta y_{1:T}\right] - \mu)
\end{aligned}$$

The second line follows from properties of the normal distribution, while the third follows from a well-known identity for covariances. Now, $\mathrm{Cov}_\theta(\Delta x_{t-1}, \Delta x_t) = \rho\sigma^2$ which leads to

$$\begin{aligned}
o_p(1) &= \mathbb{E}_T\left[\mathbb{E}_{\hat{\theta}}\left[(\Delta x_{t-1} - \hat{\mu})(\Delta x_t - \hat{\mu})\middle|\Delta y_{1:T}\right]\right] - \hat{\rho}\hat{\sigma}^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}(\mathbb{E}_{\hat{\theta}}\left[\Delta x_{t-1}|\Delta y_{1:T}\right] - \hat{\mu})(\mathbb{E}_{\hat{\theta}}\left[\Delta x_{t-1}|\Delta y_{1:T}\right] - \hat{\mu}) \\
&\quad - \frac{1}{T}\sum_{t=1}^{T}\mathrm{Cov}_{\hat{\theta}}\Big(\mathbb{E}_{\hat{\theta}}\left[\Delta x_{t-1}|\Delta y_{1:T}\right], \mathbb{E}_{\hat{\theta}}\left[\Delta x_t|\Delta y_{1:T}\right]\Big)
\end{aligned}$$

Let $T \to \infty$,

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}(\mathbb{E}_{\hat{\theta}}\left[\Delta x_{t-1}|\Delta y_{1:T}\right] - \hat{\mu})(\mathbb{E}_{\hat{\theta}}\left[\Delta x_{t-1}|\Delta y_{1:T}\right] - \hat{\mu}) &\xrightarrow{p} \mathrm{Cov}\Big(\mathbb{E}_{\bar{\theta}}\left[\Delta x_{t-1}|\Delta y_{-\infty:\infty}\right], \mathbb{E}_{\bar{\theta}}\left[\Delta x_t|\Delta y_{-\infty:\infty}\right]\Big), \\
\frac{1}{T}\sum_{t=1}^{T}\mathrm{Cov}_{\hat{\theta}}\Big(\mathbb{E}_{\hat{\theta}}\left[\Delta x_{t-1}|\Delta y_{1:T}\right], \mathbb{E}_{\hat{\theta}}\left[\Delta x_t|\Delta y_{1:T}\right]\Big) &\xrightarrow{p} \mathrm{Cov}_{\bar{\theta}}\Big(\mathbb{E}_{\bar{\theta}}\left[\Delta x_{t-1}|\Delta y_{-\infty:\infty}\right], \mathbb{E}_{\bar{\theta}}\left[\Delta x_t|\Delta y_{-\infty:\infty}\right]\Big).
\end{aligned}$$

We take limits by replacing sample averages by expectations, $\hat{\theta}$ by $\bar{\theta}$, and smoothing with respect to $\Delta y_{1:T}$ by smoothing with respect to $\Delta y_{-\infty:\infty}$.Smoothing with respect to $\Delta y_{1:\infty}$ and $\Delta y_{-\infty:\infty}$ give the same result (the proof of which we omit), but the second turns out to be more convenient.

Both limits are covariances between the smoothed values of Δx_{t-1} and Δx_t obtained using the misspecified model, but in the first case the covariance is taken under the data generating process $\mathbb{P}$ while in the second the covariance is computed using the misspecified model $\mathbb{P}_{\bar{\theta}}$ evaluated at the pseudo-true value $\bar{\theta}$. In summary, $\bar{\rho}$ is characterized by the equation

$$\mathrm{Cov}\Big(\mathbb{E}_{\bar{\theta}}\left[\Delta x_{t-1}\middle|\Delta y_{-\infty:\infty}\right], \mathbb{E}_{\bar{\theta}}\left[\Delta x_t\middle|\Delta y_{-\infty:\infty}\right]\Big) = \mathrm{Cov}_{\bar{\theta}}\Big(\mathbb{E}_{\bar{\theta}}\left[\Delta x_{t-1}\middle|\Delta y_{-\infty:\infty}\right], \mathbb{E}_{\bar{\theta}}\left[\Delta x_t\middle|\Delta y_{-\infty:\infty}\right]\Big).$$

In order to establish the spectral condition (2) we note that the Fourier transform of $\mathbb{E}_{\bar{\theta}}\left[\Delta x_t\middle|\Delta y_{-\infty:\infty}\right]$ is the Wiener-Kolmogorov filter,

$$\Delta x_\infty(\lambda) = \frac{f_{\bar{\vartheta}}(\lambda)}{f_{\bar{\vartheta}}(\lambda) + \sigma_*^2} \frac{\sum_{i=1}^N \sigma_i^{-2} \Delta y_i(\lambda)}{\sum_{i=1}^N \sigma_i^{-2}}$$

where $\Delta y_i(\lambda)$ is the Fourier transform of the time series $\Delta y_{i,-\infty:\infty}$. The filter Δx_∞ has spectrum

$$\left(\frac{f_{\bar{\vartheta}}(\lambda)}{f_{\bar{\vartheta}}(\lambda) + \sigma_*^2}\right)^2 (f_0(\lambda) + \tilde{f}(\lambda))$$

under the data generating process $\mathbb{P}$ and spectrum

$$\left(\frac{f_{\bar{\vartheta}}(\lambda)}{f_{\bar{\vartheta}}(\lambda) + \sigma_*^2}\right)^2 (f_{\bar{\vartheta}}(\lambda) + \sigma_*^2)$$

under the misspecified model $\mathbb{P}_{\bar{\theta}}$. Hence, by Fourier inversion,

$$\begin{aligned}
\mathrm{Cov}\Big(\mathbb{E}_{\bar{\theta}}\left[\Delta x_{t-1}\middle|\Delta y_{-\infty:\infty}\right], \mathbb{E}_{\bar{\theta}}\left[\Delta x_t\middle|\Delta y_{-\infty:\infty}\right]\Big) &= \int_0^{2\pi} \cos(\lambda)\left(\frac{f_{\bar{\vartheta}}(\lambda)}{f_{\bar{\vartheta}}(\lambda) + \sigma_*^2}\right)^2 (f_0(\lambda) + \tilde{f}(\lambda))\, d\lambda, \\
\mathrm{Cov}_{\bar{\theta}}\Big(\mathbb{E}_{\bar{\theta}}\left[\Delta x_{t-1}\middle|\Delta y_{-\infty:\infty}\right], \mathbb{E}_{\bar{\theta}}\left[\Delta x_t\middle|\Delta y_{-\infty:\infty}\right]\Big) &= \int_0^{2\pi} \cos(\lambda)\left(\frac{f_{\bar{\vartheta}}(\lambda)}{f_{\bar{\vartheta}}(\lambda) + \sigma_*^2}\right)^2 (f_{\bar{\vartheta}}(\lambda) + \sigma_*^2)\, d\lambda,
\end{aligned}$$

whence equation (2) follows.

References

Almuzara, M., D. Amengual, G. Fiorentini, and E. Sentana (2022): "GDP solera: The ideal vintage mix," Cemfi working paper.

Almuzara, M., D. Amengual, and E. Sentana (2019): "Normality tests for latent variables," *Quantitative Economics*, 10, 981–1017.

Aruoba, S. B. (2008): "Data revisions are not well behaved," *Journal of Money, Credit and Banking*, 40, 319–340.

Aruoba, S. B., F. X. Diebold, J. Nalewaik, F. Schorfheide, and D. Song (2016): "Improving GDP measurement: A measurement-error perspective," *Journal of Econometrics*, 191, 384–397.

Blanchard, O. J. (1997): "The medium run," *Brookings Papers on Economic Activity*, 28, 89–158.

Cavanagh, C. L. (1985): "Roots local to unity," Harvard university manuscript.

Chan, N. H. and C. Z. Wei (1987): "Asymptotic inference for nearly nonstationary AR(1) processes," *Annals of Statistics*, 15, 1050–1063.

Chang, Y., J. I. Miller, and J. Y. Park (2009): "Extracting a common trend: Theory with some applications," *Journal of Econometrics*, 150, 231–247.

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

Durbin, J. and S. J. Koopman (2012): *Time Series Analysis by State Space Methods*, Oxford University Press, second ed.

Fiorentini, G., A. Galesi, and E. Sentana (2018): "A spectral EM algorithm for dynamic factor models," *Journal of Econometrics*, 205, 249–279.

Fiorentini, G. and E. Sentana (2019): "Dynamic specification tests for dynamic factor models," *Journal of Applied Econometrics*, 34, 325–346.

Gordon, R. J. (2016): "Perspectives on the rise and fall of American growth," *American Economic Review: Papers and Proceedings*, 106, 72–76.

Greenaway-McGrevy, R. (2011): "Is GDP or GDI a better measure of output? A statistical approach," Working Paper 2011–08, Bureau of Economic Analysis.

Grimm, B. T. (2007): "The statistical discrepancy," Working Paper 2007-01, Bureau of Economic Analysis.

Hansen, A. (1939): "Economic progress and declining population growth," *American Economic Review*, 29, 1–15.

Jacobs, J. P. A. M. and S. van Norden (2011): "Modeling data revisions: Measurement error and dynamics of "true" values," *Journal of Econometrics*, 161, 101–109.

Jørgensen, B. and R. Labouriau (2012): *Exponential Families and Theoretical Inference*.

Kaldor, N. (1957): "A Model of Economic Growth," *Economic Journal*, 67, 591–624.

Karabarbounis, L. and B. Neiman (2014): "The global decline of the labor share," *Quarterly Journal of Economics*, 129, 61–103.

Landefeld, J. S., E. P. Seskin, and B. M. Fraumeni (2008): "Taking the pulse of the economy: Measuring GDP," *Journal of Economic Perspectives*, 22, 193–216.

Louis, T. A. (1982): "Finding the observed information matrix when using the EM algorithm," *Journal of the Royal Statistical Society, Series B*, 44, 226–233.

Müller, U. K. and M. W. Watson (2008): "Testing models of low-frequency variability," *Econometrica*, 76, 979–1016.

Nalewaik, J. (2010): "The income- and expenditure-side measures of output growth," *Brookings Papers on Economic Activity*, 1, 71–106.

——— (2011): "The income- and expenditure-side measures of output growth – An update through 2011Q2," *Brookings Papers on Economic Activity*, 2, 385–402.

Phillips, P. C. B. (1987): "Towards a unified asymptotic theory for autoregression," *Biometrika*, 74, 535–547.

Ruud, P. (1991): "Extensions of estimation methods using the EM algorithm," *Journal of Econometrics*, 49, 305–341.

Smith, R. J., M. R. Weale, and S. E. Satchell (1998): "Measurement error with accounting constraints: Point and interval estimation for latent data with an application to U.K. gross domestic product," *Review of Economic Studies*, 65, 109–134.

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

Summers, L. H. (2015): "Demand side secular stagnation," *American Economic Review: Papers and Proceedings*, 105, 60–65.

Watson, M. W. and R. F. Engle (1983): "Alternative algorithms for estimation of dynamic MIMIC models, factor, and time varying coefficient regression models," *Journal of Econometrics*, 23, 385–400.

Weale, M. (1992): "Estimation of data measured with error and subject to linear restrictions," *Journal of Applied Econometrics*, 7, 167–174.