Interpolation and Shock Persistence of Prewar U.S. Macroeconomic Time Series: A Reconsideration

Dezhbakhsh, Hashem and Levy, Daniel

Emory University, Bar-Ilan University, Emory University, ICEA, ISET at TSU, and RCEA

17 March 2022

Forthcoming
Economics Letters

Interpolation and shock persistence of prewar U.S. macroeconomic time series: A reconsideration*

Hashem Dezhbakhsh [a], Daniel Levy [a,b,c,d, e] **

[a] *Department of Economics, Emory University, Atlanta, GA 30322, USA*
[b] *Department of Economics, Bar-Ilan University, Ramat-Gan 5290002, Israel*
[c] *Rimini Centre for Economic Analysis, Via Patara, 3 – 47921, Rimini (RN), Italy*
[d] *International Centre for Economic Analysis, Wilfrid Laurier University, Waterloo, Ontario, Canada*
[e] *International School of Economics, Tbilisi State University, 0108 Tbilisi, Georgia*

Last Revision: February 9, 2022

Abstract: The U.S. prewar output series exhibit smaller shock-persistence than postwar-series. Some studies suggest this may be due to linear interpolation used to generate missing prewar data. Monte Carlo simulations that support this view generate large standard-errors, making such inference imprecise. We assess analytically the effect of linear interpolation on a nonstationary process. We find that interpolation indeed reduces shock-persistence, but the interpolated series can still exhibit greater shock-persistence than a pure random walk. Moreover, linear interpolation makes the series periodically nonstationary, with parameters of the data generating process and the length of the interpolation time-segments affecting shock-persistence in conflicting ways.

JEL Classification: C01, C02, E01, E30, N10

Keywords: Linear Interpolation, Random Walk, Shock-Persistence, Nonstationary series, Periodic nonstationarity, Stationary series, Prewar US Time Series

* We thank an anonymous reviewer for comments and the editor Eric Young for advice. Louis Johnston was very helpful in clarifying our inquiries about the use of linear interpolation in constructing historical time series. We are responsible for any remaining errors.

** Corresponding author at: Department of Economics, Bar-Ilan University, Ramat-Gan 5290002, Israel.
E-mail address: Daniel.Levy@biu.ac.il (D. Levy)

1. Introduction

Most prewar U.S. data including output, CPI, etc. were observed only in benchmark years, several years apart. The missing observations were usually reconstructed by linear interpolation of the benchmark observations, sometimes padded with a serially correlated term (See, e.g., Friedman and Schwartz 1982, Romer 1989, Balke and Gordon 1989, Johnston and Williamson 2018). Interpolation was also used to construct many historical series for the UK (Measuring Worth 2022).[1]

Studies find that US prewar output data is less shock-persistent than the U.S. postwar output data.[2] US prewar-data also has different shock-persistence properties than other countries' prewar data (Cogley 1990).

Stock and Watson (1986) suggest that this difference could be due to linear interpolation of missing values for the prewar data. Although, as Romer (1989) notes, much of our knowledge of the macroeconomy during the prewar periods is based on these interpolated data, the effect of interpolation on shock-persistence has been rarely examined.[3] An exception is Jaeger (1990) who assesses the effect of linear interpolation using Monte Carlo simulations for a random-walk with MA(1) errors. He generates a 50-observation time series, "original series," and a corresponding "interpolated series," where the 10^{th}, 20^{th}, 30^{th}, 40^{th} and 50^{th} observations are generated by the above data generating process (DGP), and the rest are linearly interpolated. He then calculates the variance-ratio measure of shock-persistence for the "original" and "interpolated series" over 500-replications, where variance-ratio is defined as the ratio of the variance of the series' s-period growth (long variance) to the variance of the series' 1-period growth (short variance), and finds that linear interpolation indeed reduces shock-persistence.

While Jaeger's experiments are well-designed, his estimates are very imprecise. For example, 2-SE confidence-bounds around his variance-ratio estimates fall within (0, 2+) or even (0, 3+), making it impossible to infer whether or not the ratio is less than 1 (stationary) or larger than 1 (non-stationary). Thus, one cannot infer from these results that interpolation necessarily reduces shock-persistence.

We examine analytically the effect of interpolation on the shock-persistence of a non-stationary series. The advantage of analytical approach is that it can identify interpolation effects that are distinct

[1] The use of interpolation is not limited to the prewar data. For example, Levy and Chen (1994) employ the method of linear interpolation to construct quarterly values of the US postwar capital stock and capital stock depreciation series using the series' annual observations. Levy et al. (2020) apply linear interpolation to weekly retail scanner price data during the 1989–1997 period, to determine the values of missing observations.

[2] Examples include DeLong and Summers (1988), Campbell and Mankiw (1987), Stock and Watson (1986), Cecchetti and Lam (1994), Murray and Nelson (2000), and Charles and Darné (2010), among others.

[3] Interpolation, however, has consequences for other issues as well, including dating of business cycles (Charles, et al. 2014), classification of the 19th century inflation (Kaufmann 2020), transmission of monetary policy in the EU (Ehrmann 2000), data periodicity (Franses 2013), and chaotic dynamics (Orlando and Zimatore 2018).

and not diluted by sampling-variation, which is the hallmark of simulations. We assume the same DGP as Jaeger (1990)—a random-walk with MA(1) errors, for parsimony and comparability. We first derive the variance-ratio of the original, non-interpolated series, and then develop a linear interpolation model for the series with a general benchmark cycle, replacing all non-benchmark data points with their linearly interpolated values and a moving average padding. We then derive the variance-ratio for the interpolated series and compare it to the variance ratio of the original series to ascertain the effect of interpolation on shock-persistence.

Our results provide analytical support for Jaeger's (1990) Monte Carlo findings. Additionally, we uncover a few other interpolation-caused effects. Although interpolation reduces shock-persistence, interpolated series may still exhibit high shock-persistence with variance-ratios greater than 1. Also, linear interpolation makes the series periodically non-stationary, with parameters of the DGP and the length of the interpolated time segments affecting shock-persistence in conflicting ways.

We proceed as follows. In section 2, we present the DGP and derive its shock-persistence measure analytically. In section 3, we develop the interpolation model for this DGP and analytically derive the corresponding shock-persistence measures. In section 4, we compare the results for non-interpolated and interpolated series. Section 5 concludes.

2. Shock-persistence of a random-walk with MA(1) errors

Following Jaeger (1990), we assume a random-walk with MA(1) errors and no-drift:

$$Y_T = Y_{T-1} + U_T, \quad \text{and} \quad U_T = \varepsilon_T - \theta\varepsilon_{T-1}$$

where $\varepsilon_T \sim \text{iid}\left(0, \sigma_\varepsilon^2\right)$ is white noise, and $|\theta| < 1$ for invertibility.

As Jaeger (1990) notes, this setup captures the main features of the interpolation procedure, as described by Romer (1986, 1989). The model is also parsimonious: it is simple, yet it can capture the dynamics of many macroeconomic time series, and thus it and similar models are frequently employed in macroeconomic time series analysis (e.g., Nelson and Plosser 1982, Campbell and Mankiw 1987, Cochrane 1988, Cogley 1990, etc.).

To measure shock-persistence, we follow Jaeger (1990) and others (Cochrane 1988, Cogley 1990, Leung 1992, Levy and Dezhbakhsh 2003, etc.) by using variance-ratio measure which is defined as $V_Y = \sigma_{s,Y}^2 / s\sigma_{1,Y}^2$, where $\sigma_{s,Y}^2 = \text{var}\left(Y_T - Y_{T-s}\right)$ is the variance of the series' s-period growth—"long-variance," and $\sigma_{1,Y}^2 = \text{var}\left(Y_T - Y_{T-1}\right)$ is the variance of the series' 1-period growth—"short-variance." The short variance for the above process is

$$
\begin{aligned}
\sigma_{1,Y}^2 &= \operatorname{var}\left(Y_T - Y_{T-1}\right) \\
&= \operatorname{var}\left(U_T\right) \\
&= \gamma_0
\end{aligned}
$$

where $\gamma_0 = \left(1+\theta^2\right)\sigma_\varepsilon^2$.

The long variance for this process is given by

$$
\begin{aligned}
\sigma_{s,Y}^2 &= \operatorname{var}\left(Y_T - Y_{T-s}\right) \\
&= \operatorname{var}\left(\sum_{j=0}^{s-1} U_{T-j}\right) \\
&= \sum_{j=0}^{s-1} \operatorname{var}\left(U_{T-j}\right) + 2\sum_{j=0}^{s-2}\sum_{i=j}^{s-2} \operatorname{cov}\left(U_{T-j}, U_{T-i-1}\right) \\
&= s\gamma_0 + 2(s-1)\gamma_1 + 2(s-2)\gamma_2 + 2(s-3)\gamma_3 + \cdots + 2\gamma_{s-1} \\
&= s\gamma_0 + 2(s-1)\gamma_1 \\
&= s\left(1+\theta^2\right)\sigma_\varepsilon^2 - 2(s-1)\theta\sigma_\varepsilon^2
\end{aligned}
$$

where γ_0 is given above, $\gamma_1 = -\theta\sigma_\varepsilon^2$, and $\gamma_j = 0$ for $j \geq 2$.

The variance-ratio for the original, non-interpolated random walk with MA(1) series, is therefore

$$
\begin{aligned}
V_Y &= \frac{\sigma_{s,Y}^2}{s\sigma_{1,Y}^2} \\
&= \frac{s\left(1+\theta^2\right)\sigma_\varepsilon^2 - 2(s-1)\theta\sigma_\varepsilon^2}{s\left(1+\theta^2\right)\sigma_\varepsilon^2} \\
&= 1 + \frac{2(1-s)\theta}{s\left(1+\theta^2\right)} \qquad (1)
\end{aligned}
$$

3. Shock-persistence of interpolated random-walk with MA(1) errors

To model interpolated series, we divide the original time series Y_T into segments of equal length *s*, drop all but one observation within each segment, and reconstruct the "missing" observations by linearly interpolating the remaining observations. To facilitate the conversion, rewrite Y_T as $y_{t,i}$, where $t = 0,1,2,...$ and $i = 1,2,...,s$, where $s \geq 2$. Then, $Y_1 = y_{0,1}$, $Y_2 = y_{0,2}$, …, $Y_s = y_{0,s}$, $Y_{s+1} = y_{1,1}$, …, $Y_{st+i} = y_{t,i}$.

In this notation, each period *t* contains *s* sub-periods. For example, *t* and *i* could denote years and quarters with $s = 4$ (e.g., if quarterly observations are obtained by interpolating annual observations), or decades and years within decades, with $s = 10$ (e.g. if annual observations are obtained by interpolating decennial benchmark observations). We thus rewrite the series as

$$y_{t,i} = y_{t,i-1} + u_{t,i} \quad \text{with} \quad u_{t,i} = \varepsilon_{t,i} - \theta\varepsilon_{t,i-1}, \tag{2}$$

and

$$y_{t,s} = y_{t,s-1} + u_{t,s} \quad \text{with} \quad u_{t,s} = \varepsilon_{t,s} - \theta\varepsilon_{t,s-1} \tag{3}$$

for $i = 1, 2, \ldots, s$, where $\varepsilon_{t,i} \sim \text{iid}\left(0, \sigma_\varepsilon^2\right)$ corresponds to ε_T the same way that $y_{t,i}$ corresponds to Y_T.

Now suppose that in each sub-period only one data-point is observable: the benchmark observation pertaining to the end of the period $y_{t,s}$. The remaining $s-1$ "missing" observations are generated using a segmented linear interpolation. In addition, to account for deviations from trend, the series is padded by adding a moving average component. Here we follow Jaeger (1990) by adding the MA(1) part of the original series $\varepsilon_{t,i} - \theta\varepsilon_{t,i-1}$ for padding. Thus,

$$x_{t,i} = \frac{i}{s} y_{t,s} + \frac{s-i}{s} y_{t-1,s} + \left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1}\right)\psi\left(i \neq s\right) \tag{4}$$

where $x_{t,i}$ is the interpolated series, and $\psi\left(g\right)$ is an indicator function which equals 1 if *g* is true and 0 otherwise, to ensure that MA padding is applied only to the interpolated observations.[4]

Lagging (4) one-period, we obtain

$$x_{t,i-1} = \frac{i-1}{s} y_{t,s} + \frac{s-i+1}{s} y_{t-1,s} + \left(\varepsilon_{t,i-1} - \theta\varepsilon_{t,i-2}\right)\psi\left(i \neq 1\right). \tag{5}$$

One-period difference of the interpolated series equals

$$\begin{aligned} x_{t,i} - x_{t,i-1} &= \frac{1}{s}\left(y_{t,s} - y_{t-1,s}\right) + \left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1}\right)\psi\left(i \neq s\right) - \left(\varepsilon_{t,i-1} - \theta\varepsilon_{t,i-2}\right)\psi\left(i \neq 1\right) \\ &= \frac{1}{s}\Big[\left(\varepsilon_{t,1} - \theta\varepsilon_{t-1,s}\right) + \left(\varepsilon_{t,2} - \theta\varepsilon_{t,1}\right) + \left(\varepsilon_{t,3} - \theta\varepsilon_{t,2}\right) + \cdots + \left(\varepsilon_{t,s-1} - \theta\varepsilon_{t,s-2}\right) + \left(\varepsilon_{t,s} - \theta\varepsilon_{t,s-1}\right)\Big] \\ &\quad + \Big[\left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1}\right)\psi\left(i \neq s\right)\Big] - \Big[\left(\varepsilon_{t,i-1} - \theta\varepsilon_{t,i-2}\right)\psi\left(i \neq 1\right)\Big] \end{aligned}$$

Therefore, the short-variance of the interpolated series $\sigma_{1,x}^2$, as shown in the Appendix, equals:

$$\begin{aligned} \sigma_{1,x}^2 &= \text{var}\left(x_{t,i} - x_{t,i-1}\right) \\ &= \left(\frac{\sigma_\varepsilon^2}{s^2}\right)\Big[s\left(1+\theta^2\right)\left(2s-1\right) + 2\theta\left(s^2 - 3s + 1\right)\Big] \end{aligned} \tag{6}$$

To compute the long-variance of the interpolated series $\sigma_{s,x}^2$, we start with *s*-period lag

[4] In benchmark periods, the original and the interpolated series coincide by construction. i.e., in (4), $x_{t,i} = y_{t,i}$ when $i = s$.

$$y_{t-1,s} = y_{t-2,s} + \sum_{j=1}^{s} u_{t-1,j} \, . \tag{7}$$

Then, the interpolated series equals

$$\begin{aligned} x_{t,i} &= \frac{i}{s} y_{t,s} + \frac{s-i}{s} y_{t-1,s} + \left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1}\right)\psi\left(i \neq s\right) \\ &= \frac{i}{s}\left(y_{t-1,s} + \sum_{j=1}^{s} u_{t,j} \right) + \frac{s-i}{s} y_{t-1,s} + \left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1}\right)\psi\left(i \neq s\right) \\ &= y_{t-1,s} + \frac{i}{s}\sum_{j=1}^{s} u_{t,j} + \left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1}\right)\psi\left(i \neq s\right) \end{aligned}$$

Using (7), $x_{t,i}$ equals

$$x_{t,i} = y_{t-2,s} + \sum_{j=1}^{s} u_{t-1,j} + \frac{i}{s}\sum_{j=1}^{s} u_{t,j} + \left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1}\right)\psi\left(i \neq s\right) \tag{8}$$

Lagging (8), we obtain

$$x_{t-1,i} = y_{t-2,s} + \frac{i}{s}\sum_{j=1}^{s} u_{t-1,j} + \left(\varepsilon_{t-1,i} - \theta\varepsilon_{t-1,i-1}\right)\psi\left(i \neq s\right) \tag{9}$$

Using (8) and (9), the long difference of the interpolated series equals

$$\begin{aligned} x_{t,i} - x_{t-1,i} &= \left(\frac{s-i}{s}\right)\sum_{j=1}^{s} u_{t-1,j} + \left(\frac{i}{s}\right)\sum_{j=1}^{s} u_{t,j} + \left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1} - \varepsilon_{t-1,i} + \theta\varepsilon_{t-1,i-1}\right)\psi\left(i \neq s\right) \\ &= \left(\frac{s-i}{s}\right)\Big[\left(\varepsilon_{t-1,1} - \theta\varepsilon_{t-2,s}\right) + \left(\varepsilon_{t-1,2} - \theta\varepsilon_{t-1,1}\right) + \left(\varepsilon_{t-1,3} - \theta\varepsilon_{t-1,2}\right) + \cdots \\ &\quad + \left(\varepsilon_{t-1,s-1} - \theta\varepsilon_{t-1,s-2}\right) + \left(\varepsilon_{t-1,s} - \theta\varepsilon_{t-1,s-1}\right)\Big] \\ &\quad + \left(\frac{i}{s}\right)\Big[\left(\varepsilon_{t,1} - \theta\varepsilon_{t-1,s}\right) + \left(\varepsilon_{t,2} - \theta\varepsilon_{t,1}\right) + \left(\varepsilon_{t,3} - \theta\varepsilon_{t,2}\right) + \cdots \\ &\quad + \left(\varepsilon_{t,s-1} - \theta\varepsilon_{t,s-2}\right) + \left(\varepsilon_{t,s} - \theta\varepsilon_{t,s-1}\right)\Big] \\ &\quad + \left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1} - \varepsilon_{t-1,i} + \theta\varepsilon_{t-1,i-1}\right)\psi\left(i \neq s\right) \end{aligned}$$

Therefore, the long variance of the interpolated series $\sigma_{s,x}^2 = \mathrm{var}\left(x_{t,i} - x_{t-1,i}\right)$, as shown in the Appendix, equals

$$\begin{aligned} \sigma_{s,x}^2 &= \left(\frac{\sigma_\varepsilon^2}{s^2}\right)\left[\left(\frac{1+\theta^2}{3}\right)\left[\left(2s^3 + 6s^2 - 8s + 3\right) + 3\left(2s-4\right) + 3(s+1)\right]\right] \\ &\quad - \left(\frac{\theta}{3}\right)\left(\frac{\sigma_\varepsilon^2}{s^2}\right)\left[\left(4s^3 - 3s^2 + 2s - 3\right) + 6\left(2s-4\right) + 3(s+1)\right] \end{aligned}$$

$$= \left(\frac{\sigma_\varepsilon^2}{s^2}\right)\left(\frac{1+\theta^2}{3}\right)\left(2s^3+6s^2+s-6\right)-\left(\frac{\sigma_\varepsilon^2}{s^2}\right)\left(\frac{\theta}{3}\right)\left(4s^3-3s^2+17s-24\right) \tag{10}$$

Thus, the variance-ratio for the interpolated series, using (6) and (10) and $k = s$, is given by

$$V_x = \frac{\sigma_{s,x}^2}{s\sigma_{1,x}^2}$$

$$= \frac{\left(\frac{\sigma_\varepsilon^2}{s^2}\right)\left(\frac{1+\theta^2}{3}\right)\left(2s^3+6s^2+s-6\right)-\left(\frac{\sigma_\varepsilon^2}{s^2}\right)\left(\frac{\theta}{3}\right)\left(4s^3-3s^2+17s-24\right)}{s\left(\frac{\sigma_\varepsilon^2}{s^2}\right)\left[s\left(1+\theta^2\right)\left(2s-1\right)+2\theta\left(s^2-3s+1\right)\right]}$$

$$= \frac{\left(1+\theta^2\right)\left(2s^3+6s^2+s-6\right)-\theta\left(4s^3-3s^2+17s-24\right)}{3s^2\left(1+\theta^2\right)\left(2s-1\right)+6s\theta\left(s^2-3s+1\right)} \tag{11}$$

4. Shock-persistence comparisons for original and interpolated series

A variance-ratio $V_Y = \sigma_{s,Y}^2 / s\sigma_{1,Y}^2$ smaller than 1, suggests that the long-run component of the series as measured by $\sigma_{s,Y}^2$, is stable relative to year-to-year changes as measured by $\sigma_{1,Y}^2$.

Shock-persistence for the original series is given by equation (1), where $s\left(1+\theta^2\right) > 0$ because $s \geq 2$ and $-1 < \theta < 1$. The size of V_Y therefore depends on the sign of θ: if $\theta < 0$, then $V_Y > 1$, and if $\theta > 0$, then $V_Y < 1$. For a random-walk with white noise errors ($\theta = 0$) and no-drift, the variance ratio equals 1 as equation (1) shows. See Table 1 and Figure 1.

Shock-persistence measure for the interpolated series, given by equation (11), depends on the MA parameter θ and on the interpolation segment length *s* which in this case also represents the time-length of the long-difference. To identify the impact of interpolation, we examine how shock-persistence varies with the two parameters θ and *s*, by numerically evaluating the variance-ratio of both the original and the interpolated series for various values of the parameters.

Table 1 exhibits the results. Several observations are noteworthy. First, variance-ratio of the original random-walk exhibits significant variation with respect to θ and *s*, varying from 0.03 (no shock-persistence), to 1.97 (significant shock-persistence). For all values of *s*, variance-ratio declines as θ increases. The decline is steeper for larger values of *s*. This directional consistency, however, is not reciprocal. For negative values of θ, variance-ratio increases with *s*, but for positive values of θ, variance-ratio decreases with *s*. For $\theta = 0$, it is 1 for all values of *s*.

Second, the interpolated series has a smaller variance-ratio than the original series for each parameter value, confirming that interpolation reduces shock-persistence. This finding provides analytical support for Jaeger's (1990) simulation estimates which, as noted, had little inferential value due to large standard errors of these estimates. Jaeger's simulations results reported in Table 1 (last column), and their comparable counterparts from our theoretical work reported in column 4, highlights this point.

Third, while interpolation reduces the shock-persistence of a nonstationary series, the interpolated series may still exhibit significant shock-persistence. Thus, a low shock-persistence cannot be automatically attributed to interpolation. Indeed, Leung (1992) finds that the persistence differences between US and UK, which Jaeger uses to corroborate his findings, hold only for the particular UK output series that Jaeger used (NNP series). Alternative UK series (GDP at market price or factor cost) are similar in terms of shock-persistence to the interpolated prewar US output series. Thus, a low shock-persistence is not a sole artifact of interpolation.

Figure 1 displays the pattern of variance-ratio of the original series as data generating parameter θ and interpolation segment s change. As the figure shows, the variance-ratio may increase or decrease beyond the reference point of 1 as θ varies, if the series is a pure random-walk. The changes are more pronounced at higher values of s but more variable with respect to s at lower values of s. We also observe a prominent symmetry in the behavior of the variance-ratio for positive and negative values of θ, as s increases.

Figure 2 displays the pattern of variance-ratio of the interpolated series as we vary the DGP parameter θ and the interpolation segment s. Here we find that when $\theta < 0$, there are smaller changes in variance-ratio away from the reference point of 1 than when $\theta > 0$. Moreover, for positive values of θ, variance-ratio drops sharply with s at low values of s, then it remains steady as s increases. There is no symmetry in the behavior the variance ratio of the interpolated series. This is a further indication of the nonuniformity in the effect of linear interpolation on nonstationary series. These patterns reinforce our conclusion that interpolation affects data in complicated ways that go beyond simple rules.

Figure 3 shows the difference between the variance ratios of the original and interpolated series $V_y - V_x$ as we vary the parameters θ and s. The difference is higher for negative values of θ than for positive values of θ. Further, the effect of increase in s is relatively sharp for lower values of s.

Fourth, as we show in the Appendix, the variance-ratio for the interpolated series exhibits periodic nonstationarity as the underlying moments are conditional on observation index i.[5] Using iterated expectation, we removed this conditionality, and thus eliminated the dependency on individual i's.

[5] Dezhbakhsh and Levy (1994) derive the variance, the covariance, and the autocorrelation functions of linearly interpolated trend-stationary series, and find that they all vary with i, which they term "periodic variation."

However, the presence of s in the shock-persistence measures we derive highlights this point.

In closing, we note a subtle statistical point. The variations and changes in the variance ratio parameter due to change in model parameters or interpolation length that we alluded to in the above discussions refer to the actual parameters derived from the DGP or its interpolated form. The sampling variation and uncertainty that comes with any inference about variance ratio parameter would depend on sample size and method of estimation and must not be confounded with the above theoretical results that are derived by changing the underlying model parameters.

5. Concluding Remarks

Stock and Watson (1986) suggest that linear interpolation of prewar U.S. macroeconomic series is the likely cause of the shock-persistence difference between these series and most European prewar or similar U.S. postwar series. Jaeger's (1990) simulation results is a rare confirmation of this assertion, but the large variations in his simulation-based persistence measures detracts from its inference value. We derive analytically the impact of interpolation on shock-persistence in Jaeger's model, and also go further to uncover a few additional interpolation-caused effects.

Our findings are as follows. First, linear interpolation reduces the shock-persistence of a random-walk with MA(1) errors, confirming Stock and Watson's (1986) conjecture and Jaeger's (1990) simulation results. Second, however, the interpolated random-walk series, may still exhibit significant shock persistence, with variance-ratio attaining values greater than 1, suggesting that interpolation is not synonymous with low shock persistence. Third, linear interpolation introduces periodic non-stationarity in a series. Fourth, the effect of linear interpolation on shock-persistence depends on the parameters of the underlying DGP as well as on the interpolation segment length.

Overall, our results suggest that using a simple rule to describe the effect of interpolation on shock-persistence would be an overreach even in a simple model such as Jaeger's (1990). That is because the determinants of this effect, which include the DGP parameters, the length of the interpolation segment, and the periodic nonstationarity that interpolation introduces, interact in complex ways. An important implication of this finding is that alternative causes of the difference in the persistence estimates for the U.S. prewar and postwar data merits further investigation.

We suspect that the results we report here are specific to the particular model we study and thus any statements about their generalizability should be made with caution. It is, therefore, important that future work examines the implications of linear interpolation for more general classes of models, and for both stationary and non-stationary series.

References

Balke, N., Gordon, R., 1989. The estimation of prewar gross national product: methodology and new evidence. ***Journal of Political Economy*** 97 (1), 38–92.

Campbell, J., Mankiw, N., 1987. Are output fluctuations transitory? ***Quarterly Journal of Economics*** 102, 857–880.

Cecchetti, S., Lam, P., 1994. Variance-ratio tests: Small-sample properties with an application to international output data. ***Journal of Business and Economic Statistics*** 12 (2), 177–186.

Charles, A., Darné, O, 2010. A note on the uncertain trend in US real GNP. Working Paper, HAL-00547737.

Charles, A., Darné, O., Diebolt, C., 2014. A revision of the US business-cycles chronology, 1790–1928. ***Economics Bulletin*** 34 (1), 234–244.

Cochrane J., 1988. How big is the random walk in GNP? ***Journal of Political Economy*** 96(5), 893–920.

Cogley, T, 1990. International evidence on the size of the random walk in output. ***Journal of Political Economy*** 98 (3), 501–518.

DeLong, B., Summers, L., 1988. How does macroeconomic policy affect output? ***Brookings Papers on Economic Activity*** 2, 433–480.

Dezhbakhsh, H., Levy, D., 1994. Periodic properties of interpolated time series. ***Economics Letters*** 44, 221–228.

Ehrmann, M., 2000. Comparing monetary policy transmission across European countries. ***Weltwirtschaftliches Archiv*** 136 (1), 58–83.

Franses, P., 2013. Data revisions and periodic properties of macroeconomic data. ***Economics Letters*** 120, 139–141.

Friedman, M., Schwartz, A., 1982. Monetary Trends in the US and the UK, NBER.

Jaeger, A., 1990. Shock persistence and the measurement of prewar output series. ***Economics Letters*** 34, 333–337.

Johnston, L., Williamson, S., 2018. The annual real and nominal GDP for the United States, 1790–1928," www.measuringworth.com/datasets/usgdp12/sourcegdp.php (accessed: Sept. 13, 2018).

Kaufmann, D., 2020. Is deflation costly after all? The perils of erroneous historical classifications. ***Journal of Applied Econometrics*** 35(5), 614–628.

Leung, S., 1992. Changes in the behavior of output in the United Kingdom, 1856–1990. ***Economics Letters*** 40, 435-444.

Levy, D., Chen, H., 1994. Estimates of the aggregate quarterly capital stock series for the post-war U.S. economy. ***Review of Income and Wealth*** 40, 317–349.

Levy, D., Dezhbakhsh, H., 2003. International evidence on output fluctuation and shock persistence. ***Journal of Monetary Economics*** 50, 1499–1530.

Levy, D., Snir, A., Chen, H., Gotler, A., 2020. Not all price endings are created equal: Price points and asymmetric price rigidity. ***Journal of Monetary Economics*** 110, 33–49.

Measuring Worth, 2022. Notes on three centuries of UK GDP and seven centuries of English GDP. https://measuringworth.com/datasets/ukgdp/#, accessed February 8, 2022.

Murray, C., Nelson, C., 2000. The uncertain trend in U.S. GDP. Journal of Monetary Economics 46(1), 79–95.

Nelson, C., Plosser, C., 1982. Trends and random walks in macroeconomic time series. ***Journal of***

Monetary Economics 10, 139–162.

Orlando, G., Zimatore, G., 2018. Recurrence quantification analysis of business cycles. ***Chaos, Solitons and Fractals*** 110, 82–94.

Romer, C., 1986. Is the Stabilization of the Postwar Economy a Figment of the Data? ***American Economic Review*** 76(3), 314–334.

Romer, C., 1989. The prewar business cycle reconsidered: new estimates of gross national product, 1869–1908. ***Journal of Political Economy*** 97, l–37.

Stock, J., Watson, M., 1986. Does GNP have a unit root? ***Economic Letters*** 22, 147–151.

Table 1

Values of the variance-ratio for the original (non-interpolated) and interpolated series, for different values of θ and s.

		$s=5$	$s=10$	$s=15$	$s=20$	$s=25$	$s=30$	Jaeger $s=10$
$\theta=-0.99$	Original	1.79	1.90	1.93	1.95	1.96	1.97	
	Interpolated	1.26	1.27	1.28	1.29	1.30	1.30	
$\theta=-0.75$	Original	1.77	1.86	1.87	1.91	1.92	1.93	
	Interpolated	1.22	1.22	1.22	1.22	1.23	1.23	
$\theta=-0.50$	Original	1.64	1.72	1.74	1.76	1.77	1.77	1.69
	Interpolated	1.09	1.04	1.02	1.01	1.01	1.01	0.98
$\theta=-0.25$	Original	1.38	1.42	1.44	1.45	1.45	1.45	1.37
	Interpolated	0.86	0.75	0.71	0.69	0.68	0.67	0.65
$\theta=0.00$	Original	1.00	1.00	1.00	1.00	1.00	1.00	0.93
	Interpolated	0.59	0.46	0.41	0.39	0.38	0.37	0.38
$\theta=0.25$	Original	0.62	0.58	0.56	0.55	0.55	0.55	0.50
	Interpolated	0.38	0.25	0.22	0.19	0.19	0.18	0.16
$\theta=0.50$	Original	0.36	0.28	0.25	0.24	0.23	0.23	
	Interpolated	0.25	0.14	0.11	0.09	0.08	0.08	
$\theta=0.75$	Original	0.23	0.14	0.10	0.09	0.08	0.07	
	Interpolated	0.20	0.10	0.07	0.05	0.04	0.04	
$\theta=0.99$	Original	0.20	0.10	0.07	0.05	0.04	0.03	
	Interpolated	0.19	0.09	0.06	0.04	0.03	0.03	

Notes

1. The table reports the variance-ratio measures of shock-persistence of the original and interpolated series for the values of θ between –0.99 and 0.99, and for the values of s between 5 and 30.
2. The DGP is random walk with MA(1) errors with parameter θ and the length of the long-difference for the variance-ratio s.
3. The figures in the table were computed using equations (1) and (11), for the original and for the interpolated series, respectively.
4. The source of the figures presented in the last column is Jaeger (1990), p. 336, Table 2, column labeled $k=10$.
5. The shaded cells indicate comparable figures.

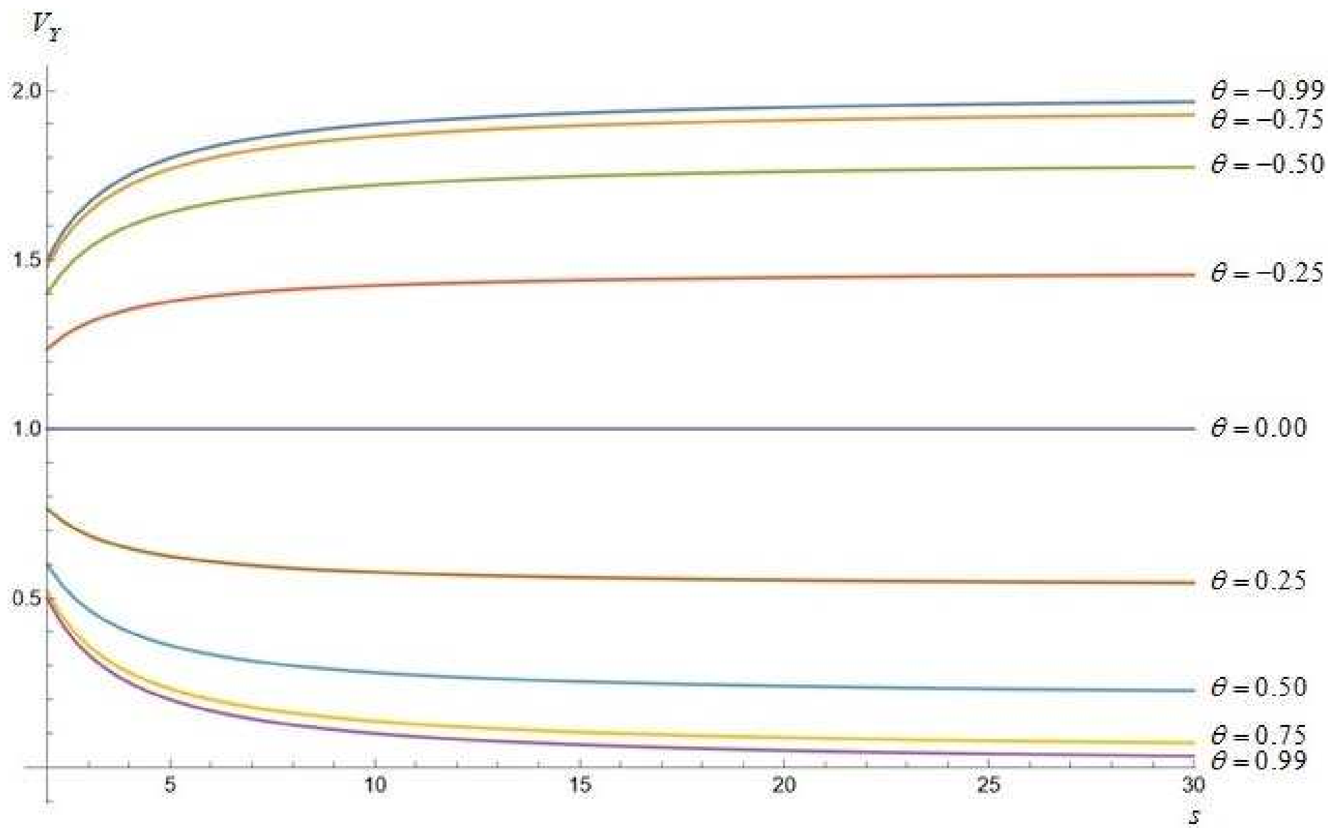

Fig. 1. Effects of changing the values of θ and s on the variance-ratio of the original series

Notes:

1. The figure shows the values of the variance-ratio of the original series V_y given in equation (1), for values of θ between -0.99 and 0.99, and for values of s between 3 and 30.
2. The DGP is random walk with MA(1) errors with parameter θ and the length of the long-difference for the variance-ratio s.

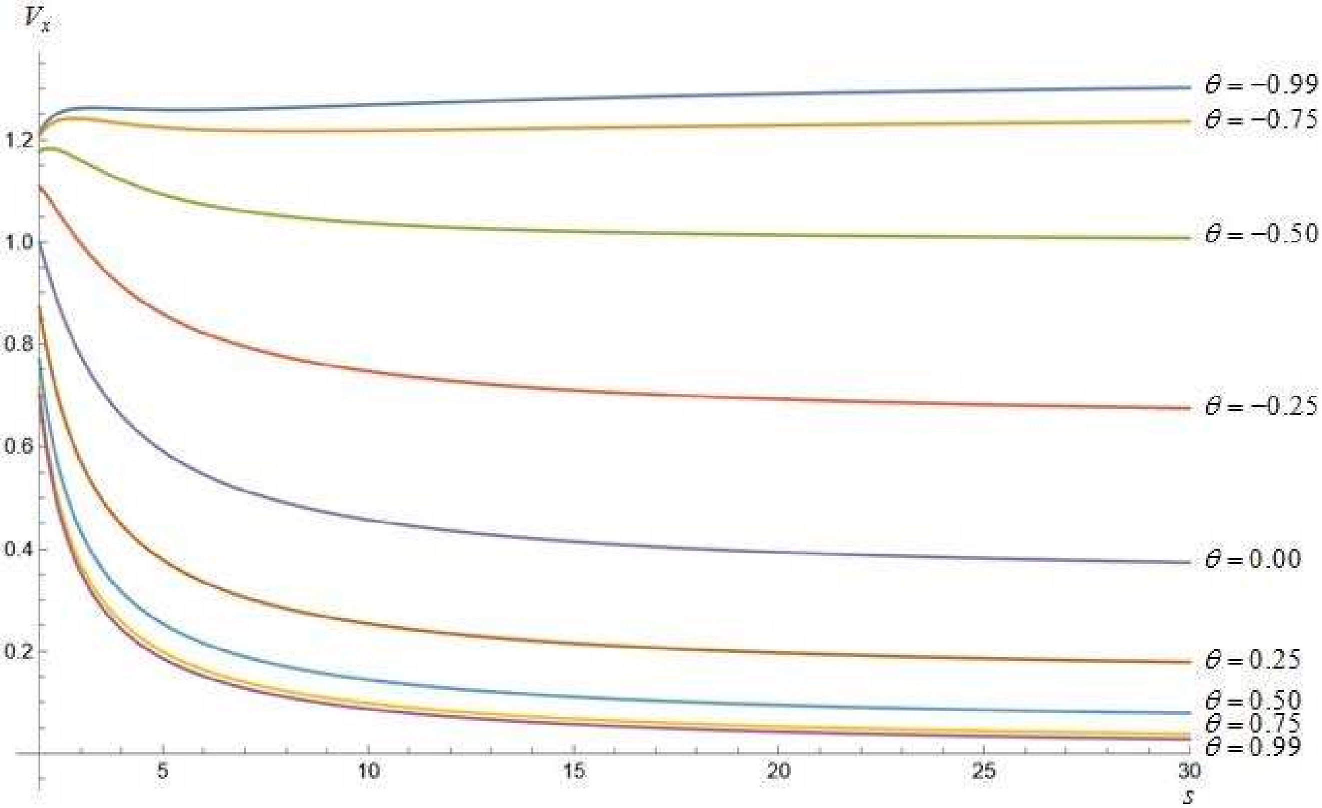

Fig. 2. Effects of changing the values of θ and s on the variance-ratio of the interpolated series

Notes:
The figure shows the values of the variance-ratio of the interpolated series V_x given in equation (11). See the notes underneath Figure 1 for more details.

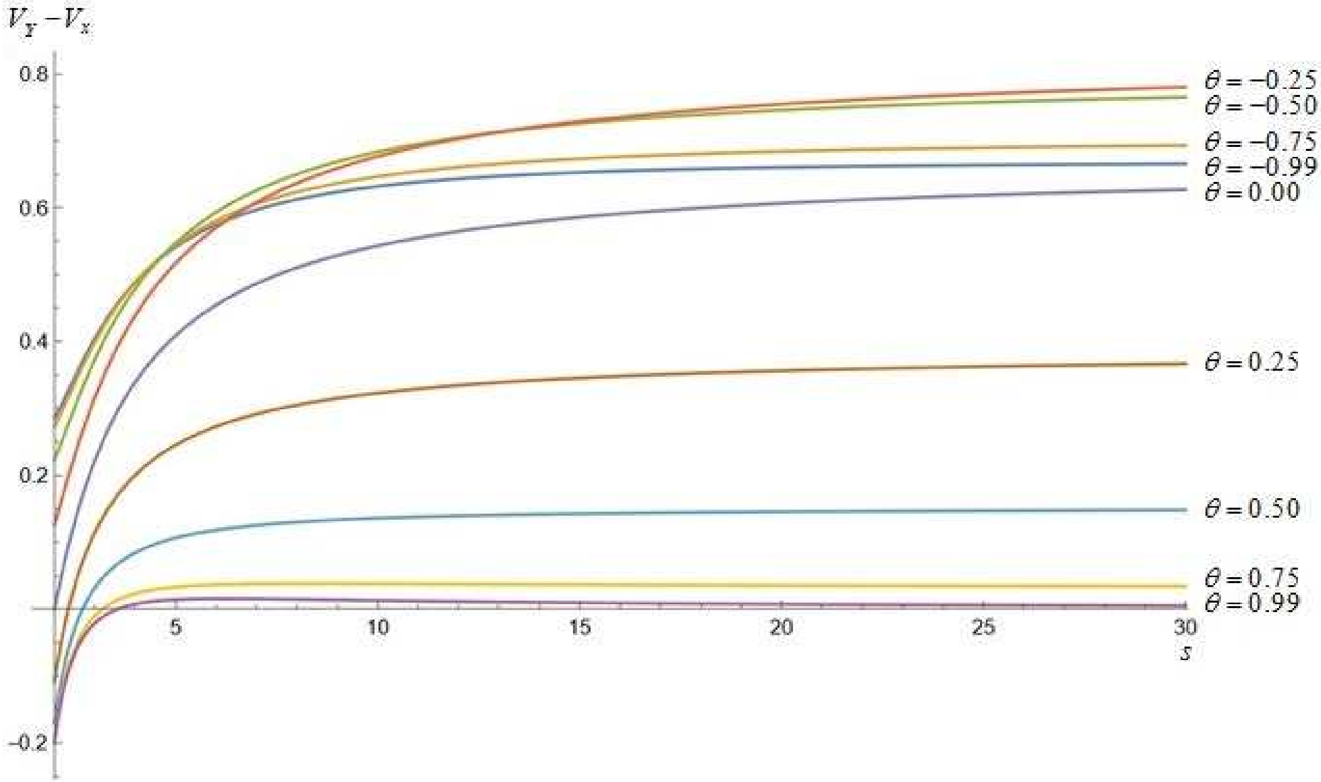

Fig. 3. The difference between the values of the variance ratio of the original and interpolated series, for different values of θ and s

Notes:
The figure shows the difference between the values of the variance-ratio of the original and interpolated series $V_y - V_x$. See the notes underneath Figure 1 for more details.

Online Supplementary Appendix

(Not for Publication)

Interpolation and Shock Persistence of Prewar U.S. Macroeconomic Time Series: A Reconsideration

Hashem Dezhbakhsh
Department of Economics, Emory University
Atlanta, GA 30322, USA
econhd@emory.edu

Daniel Levy
Department of Economics, Bar-Ilan University
Ramat-Gan 5290002, ISRAEL,
Department of Economics, Emory University
Atlanta, GA 30322, USA,
Rimini Center for Economic Analysis, ITALY, and
International Centre for Economic Analysis, CANADA
ISET at TSU, Tbilisi, GEORGIA
Daniel.Levy@biu.ac.il

February 9, 2022

To obtain the variance ratio for the interpolated series (section 3 in the paper), we need to find $\sigma_{1,x}^2 = \operatorname{var}\left(x_{t,i} - x_{t,i-1}\right)$, the variance of the series' 1-period growth or the "short variance," and $\sigma_{s,x}^2 = \operatorname{var}\left(x_{t,i} - x_{t-1,i}\right)$, the variance of the series' s-period growth or the "long variance."

<u>Short Variance</u>

The short variance, i.e., the variance of the "short difference" can be obtained by taking the variance of the short difference in the interpolated series, as shown below.

$$
\begin{aligned}
\sigma_{1,x}^2 &= \operatorname{var}\left(x_{t,i} - x_{t,i-1}\right) \\
&= \operatorname{var}\left[(1/s)\left(y_{t,s} - y_{t-1,s}\right) + \left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1}\right)\psi\left(i \neq s\right) - \left(\varepsilon_{t,i-1} - \theta\varepsilon_{t,i-2}\right)\psi\left(i \neq 1\right)\right] \\
&= \operatorname{var}\left\{(1/s)\left[\left(\varepsilon_{t,1} - \theta\varepsilon_{t-1,s}\right) + \left(\varepsilon_{t,2} - \theta\varepsilon_{t,1}\right) + \left(\varepsilon_{t,3} - \theta\varepsilon_{t,2}\right) + \cdots + \left(\varepsilon_{t,s-1} - \theta\varepsilon_{t,s-2}\right) + \left(\varepsilon_{t,s} - \theta\varepsilon_{t,s-1}\right)\right]\right\} \\
&\quad + \operatorname{var}\left[\left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1}\right)\psi\left(i \neq s\right)\right] \\
&\quad + \operatorname{var}\left[\left(\varepsilon_{t,i-1} - \theta\varepsilon_{t,i-2}\right)\psi\left(i \neq 1\right)\right] \\
&\quad + 2\operatorname{cov}\left\{\left[(1/s)\left[\left(\varepsilon_{t,1} - \theta\varepsilon_{t-1,s}\right) + \left(\varepsilon_{t,2} - \theta\varepsilon_{t,1}\right) + \left(\varepsilon_{t,3} - \theta\varepsilon_{t,2}\right) + \cdots + \left(\varepsilon_{t,s-1} - \theta\varepsilon_{t,s-2}\right) + \left(\varepsilon_{t,s} - \theta\varepsilon_{t,s-1}\right)\right]\right],\right. \\
&\quad \left.\left[\left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1}\right)\psi\left(i \neq s\right)\right]\right\} \\
&\quad - 2\operatorname{cov}\left\{\left[(1/s)\left[\left(\varepsilon_{t,1} - \theta\varepsilon_{t-1,s}\right) + \left(\varepsilon_{t,2} - \theta\varepsilon_{t,1}\right) + \left(\varepsilon_{t,3} - \theta\varepsilon_{t,2}\right) + \cdots + \left(\varepsilon_{t,s-1} - \theta\varepsilon_{t,s-2}\right) + \left(\varepsilon_{t,s} - \theta\varepsilon_{t,s-1}\right)\right]\right],\right. \\
&\quad \left.\left[\left(\varepsilon_{t,i-1} - \theta\varepsilon_{t,i-2}\right)\psi\left(i \neq 1\right)\right]\right\} \\
&\quad + 2\operatorname{cov}\left\{\left[\left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1}\right)\psi\left(i \neq s\right)\right], \left[\left(\varepsilon_{t,i-1} - \theta\varepsilon_{t,i-2}\right)\psi\left(i \neq 1\right)\right]\right\}
\end{aligned}
$$

Applying the definitions of variances and covariances, we obtain

$$
\begin{aligned}
\sigma_{1,x}^2 &= (1/s)^2\left\{\operatorname{var}\left(\varepsilon_{t,1} - \theta\varepsilon_{t-1,s}\right) + \cdots + \operatorname{var}\left(\varepsilon_{t,s} - \theta\varepsilon_{t,s-1}\right) + 2E\left[\left(\varepsilon_{t,1} - \theta\varepsilon_{t-1,s}\right)\left(\varepsilon_{t,2} - \theta\varepsilon_{t,1}\right)\right] + \cdots\right. \\
&\quad \left. + 2E\left[\left(\varepsilon_{t,s-1} - \theta\varepsilon_{t,s-2}\right)\left(\varepsilon_{t,s} - \theta\varepsilon_{t,s-1}\right)\right]\right\} \\
&\quad + \operatorname{var}\left[\left(\varepsilon_{t,i} - \theta\varepsilon_{t,i-1}\right)\psi\left(i \neq s\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&+\operatorname{var}\left[\left(\varepsilon_{t,i-1}-\theta\varepsilon_{t,i-2}\right)\psi\left(i\neq 1\right)\right] \\
&+2\operatorname{cov}\left\{\left[\left(1/s\right)\left[\left(\varepsilon_{t,1}-\theta\varepsilon_{t-1,s}\right)+\left(\varepsilon_{t,2}-\theta\varepsilon_{t,1}\right)+\left(\varepsilon_{t,3}-\theta\varepsilon_{t,2}\right)+\cdots+\left(\varepsilon_{t,s-1}-\theta\varepsilon_{t,s-2}\right)+\left(\varepsilon_{t,s}-\theta\varepsilon_{t,s-1}\right)\right]\right],\right. \\
&\left.\left[\left(\varepsilon_{t,i}-\theta\varepsilon_{t,i-1}\right)\psi\left(i\neq s\right)\right]\right\} \\
&-2\operatorname{cov}\left\{\left[\left(1/s\right)\left[\left(\varepsilon_{t,1}-\theta\varepsilon_{t-1,s}\right)+\left(\varepsilon_{t,2}-\theta\varepsilon_{t,1}\right)+\left(\varepsilon_{t,3}-\theta\varepsilon_{t,2}\right)+\cdots+\left(\varepsilon_{t,s-1}-\theta\varepsilon_{t,s-2}\right)+\left(\varepsilon_{t,s}-\theta\varepsilon_{t,s-1}\right)\right]\right],\right. \\
&\left.\left[\left(\varepsilon_{t,i-1}-\theta\varepsilon_{t,i-2}\right)\psi\left(i\neq 1\right)\right]\right\} \\
&+2\operatorname{cov}\left\{\left[\left(\varepsilon_{t,i}-\theta\varepsilon_{t,i-1}\right)\psi\left(i\neq s\right)\right],\left[\left(\varepsilon_{t,i-1}-\theta\varepsilon_{t,i-2}\right)\psi\left(i\neq 1\right)\right]\right\}
\end{aligned}
$$

which can be simplified

$$
\begin{aligned}
\sigma_{1,x}^{2} &= \left(1/s\right)^{2}\left[s\left(\sigma_{\varepsilon}^{2}+\theta^{2}\sigma_{\varepsilon}^{2}\right)+\left(s-1\right)\left(-2\theta\sigma_{\varepsilon}^{2}\right)\right]+\left[\left(\sigma_{\varepsilon}^{2}+\theta^{2}\sigma_{\varepsilon}^{2}\right)\psi\left(i\neq s\right)\right] \\
&+\left[\left(\sigma_{\varepsilon}^{2}+\theta^{2}\sigma_{\varepsilon}^{2}\right)\psi\left(i\neq 1\right)\right]+\left[\left(2\theta\sigma_{\varepsilon}^{2}\right)\psi\left(i\neq 1, i\neq s\right)\right] \\
&+\left[\left(\frac{2\sigma_{\varepsilon}^{2}}{s}-\frac{2\theta\sigma_{\varepsilon}^{2}}{s}-\frac{2\theta\sigma_{\varepsilon}^{2}}{s}+\frac{2\theta^{2}\sigma_{\varepsilon}^{2}}{s}\right)\psi\left(i\neq 1, i\neq s\right)\right] \\
&+\left[\left(\frac{2\sigma_{\varepsilon}^{2}}{s}-\frac{2\theta\sigma_{\varepsilon}^{2}}{s}+\frac{2\theta^{2}\sigma_{\varepsilon}^{2}}{s}\right)\psi\left(i=1\right)\right] \\
&+\left[\left(-\frac{2\sigma_{\varepsilon}^{2}}{s}+\frac{2\theta\sigma_{\varepsilon}^{2}}{s}-\frac{2\theta^{2}\sigma_{\varepsilon}^{2}}{s}\right)\psi\left(i=2\right)\right] \\
&+\left[\left(-\frac{2\sigma_{\varepsilon}^{2}}{s}+\frac{2\theta\sigma_{\varepsilon}^{2}}{s}+\frac{2\theta\sigma_{\varepsilon}^{2}}{s}-\frac{2\theta^{2}\sigma_{\varepsilon}^{2}}{s}\right)\psi\left(s\geq i>2\right)\right]
\end{aligned}
$$

Collecting terms, we obtain:

$$
\begin{aligned}
\sigma_{1,x}^{2}\left(i\right) &= \left(\frac{\sigma_{\varepsilon}^{2}}{s^{2}}\right)\left[s\left(1+\theta^{2}\right)-2\left(s-1\right)\theta\right]+\left[\sigma_{\varepsilon}^{2}\left(1+\theta^{2}\right)\psi\left(i\neq s\right)\right] \\
&+\left[\sigma_{\varepsilon}^{2}\left(1+\theta^{2}\right)\psi\left(i\neq 1\right)\right]+\left[\sigma_{\varepsilon}^{2}\left(2\theta\right)\psi\left(i\neq 1, i\neq s\right)\right] \\
&+\left(\frac{\sigma_{\varepsilon}^{2}}{s}\right)\left[\left(2-4\theta+2\theta^{2}\right)\psi\left(i\neq 1, i\neq s\right)\right]+\left(\frac{\sigma_{\varepsilon}^{2}}{s}\right)\left[\left(2-2\theta+2\theta^{2}\right)\psi\left(i=1\right)\right] \\
&+\left(\frac{\sigma_{\varepsilon}^{2}}{s}\right)\left[\left(-2+2\theta-2\theta^{2}\right)\psi\left(i=2\right)\right]+\left(\frac{\sigma_{\varepsilon}^{2}}{s}\right)\left[\left(-2+4\theta-2\theta^{2}\right)\psi\left(s\geq i>2\right)\right]
\end{aligned}
$$

Note that the variance depends on i which is the index within the interpolation segment. We use the notation $\sigma_{1,x}^{2}(i)$ to emphasize the dependence of the short variance of the interpolated series on i. This interpolation-caused dependency is referred to as *periodic nonstationarity*.[1] To remove this conditionality, we integrate i out of this equation using the fact that i follows a uniform distribution with support [1, 2, …s].

Using the summation rules,

$$\frac{1}{s}\sum_{i=1}^{s}\psi(i\neq 1)=\frac{1}{s}\sum_{i=1}^{s}\psi(i\neq s)=\frac{s-1}{s}$$

$$\frac{1}{s}\sum_{i=1}^{s}\psi(i=1)=\frac{1}{s}\sum_{i=1}^{s}\psi(i=2)=\frac{1}{s}$$

$$\frac{1}{s}\sum_{i=1}^{s}\psi(i\neq 1,i\neq s)=\frac{1}{s}\sum_{i=1}^{s}\psi(s\geq i>2)=\frac{s-2}{s}$$

we obtain

$$\begin{aligned}\sigma_{1,x}^{2}&=\left(\frac{\sigma_{\varepsilon}^{2}}{s^{2}}\right)\left[s\left(1+\theta^{2}\right)-2(s-1)\theta\right]\\&+\sigma_{\varepsilon}^{2}\left[\left(1+\theta^{2}\right)\left(\frac{s-1}{s}\right)+\left(1+\theta^{2}\right)\left(\frac{s-1}{s}\right)+2\theta\left(\frac{s-2}{s}\right)\right]\\&+\left(\frac{\sigma_{\varepsilon}^{2}}{s}\right)\left[\left(2-4\theta+2\theta^{2}\right)\left(\frac{s-2}{s}\right)+\left(2-2\theta+2\theta^{2}\right)\left(\frac{1}{s}\right)\right.\\&\left.+\left(-2+2\theta-2\theta^{2}\right)\left(\frac{1}{s}\right)+\left(-2+4\theta-2\theta^{2}\right)\left(\frac{s-2}{s}\right)\right]\end{aligned}$$

which is the unconditional short variance of the interpolated series. Simplifying the equation, after collecting terms, we obtain

[1] Dezhbakhsh and Levy (1994) derive the variance, the covariance, and the autocorrelation functions of linearly interpolated trend-stationary series, and find that they all vary with i, which they term "periodic variation."

$$\sigma_{1,x}^2 = \left(\frac{\sigma_\varepsilon^2}{s^2}\right)\left[s\left(1+\theta^2\right)\left(2s-1\right)+2\theta\left(s^2-3s+1\right)\right]$$

This is the expression for the short variance that is given in equation (6) in the paper.

<u>Long Variance</u>

The long variance, i.e., the variance of the "long difference" can be obtained by taking the variance of the long difference in the interpolated series, as shown below.

$$\begin{aligned}
\sigma_{s,x}^2 &= \operatorname{var}\left(x_{t,i}-x_{t-1,i}\right)\\
&= \operatorname{var}\left[\left(\frac{s-i}{s}\right)\sum_{j=1}^{s}u_{t-1,j}+\left(\frac{i}{s}\right)\sum_{j=1}^{s}u_{t,j}+\left(\varepsilon_{t,i}-\theta\varepsilon_{t,i-1}-\varepsilon_{t-1,i}+\theta\varepsilon_{t-1,i-1}\right)\psi\left(i\neq s\right)\right]\\
&=\left(\frac{s-i}{s}\right)^2\operatorname{var}\left[\left(\varepsilon_{t-1,1}-\theta\varepsilon_{t-2,s}\right)+\left(\varepsilon_{t-1,2}-\theta\varepsilon_{t-1,1}\right)+\cdots+\left(\varepsilon_{t-1,s}-\theta\varepsilon_{t-1,s-1}\right)\right]\\
&+\left(\frac{i}{s}\right)^2\operatorname{var}\left[\left(\varepsilon_{t,1}-\theta\varepsilon_{t-1,s}\right)+\left(\varepsilon_{t,2}-\theta\varepsilon_{t,1}\right)+\cdots+\left(\varepsilon_{t,s}-\theta\varepsilon_{t,s-1}\right)\right]\\
&+\operatorname{var}\left[\left(\varepsilon_{t,i}-\theta\varepsilon_{t,i-1}-\varepsilon_{t-1,i}+\theta\varepsilon_{t-1,i-1}\right)\psi\left(i\neq s\right)\right]\\
&+2\operatorname{cov}\left\{\left(\frac{s-i}{s}\right)\left[\left(\varepsilon_{t-1,1}-\theta\varepsilon_{t-2,s}\right)+\left(\varepsilon_{t-1,2}-\theta\varepsilon_{t-1,1}\right)+\cdots+\left(\varepsilon_{t-1,s}-\theta\varepsilon_{t-1,s-1}\right)\right],\right.\\
&\left.\left(\frac{i}{s}\right)\left[\left(\varepsilon_{t,1}-\theta\varepsilon_{t-1,s}\right)+\left(\varepsilon_{t,2}-\theta\varepsilon_{t,1}\right)+\cdots+\left(\varepsilon_{t,s}-\theta\varepsilon_{t,s-1}\right)\right]\right\}\\
&+2\operatorname{cov}\left\{\left[\left(s-i\right)/s\right]\left[\left(\varepsilon_{t-1,1}-\theta\varepsilon_{t-2,s}\right)+\left(\varepsilon_{t-1,2}-\theta\varepsilon_{t-1,1}\right)+\cdots+\left(\varepsilon_{t-1,s}-\theta\varepsilon_{t-1,s-1}\right)\right],\right.\\
&\left.\left[\left(\varepsilon_{t,i}-\theta\varepsilon_{t,i-1}-\varepsilon_{t-1,i}+\theta\varepsilon_{t-1,i-1}\right)\psi\left(i\neq s\right)\right]\right\}\\
&+2\operatorname{cov}\left\{\left(i/s\right)\left[\left(\varepsilon_{t,1}-\theta\varepsilon_{t-1,s}\right)+\left(\varepsilon_{t,2}-\theta\varepsilon_{t,1}\right)+\cdots+\left(\varepsilon_{t,s}-\theta\varepsilon_{t,s-1}\right)\right],\right.\\
&\left.\left[\left(\varepsilon_{t,i}-\theta\varepsilon_{t,i-1}-\varepsilon_{t-1,i}+\theta\varepsilon_{t-1,i-1}\right)\psi\left(i\neq s\right)\right]\right\}
\end{aligned}$$

which yields

$$
\begin{aligned}
\sigma_{s,x}^2 =& \left(\frac{s-i}{s}\right)^2 \left[s\left(\sigma_\varepsilon^2+\theta^2\sigma_\varepsilon^2\right)+\left(s-1\right)\left(-2\theta\sigma_\varepsilon^2\right)\right] \\
&+\left(\frac{i}{s}\right)^2 \left[s\left(\sigma_\varepsilon^2+\theta^2\sigma_\varepsilon^2\right)+\left(s-1\right)\left(-2\theta\sigma_\varepsilon^2\right)\right] \\
&+\left[\left(\sigma_\varepsilon^2+\theta^2\sigma_\varepsilon^2+\sigma_\varepsilon^2+\theta^2\sigma_\varepsilon^2\right)\psi\left(i\neq s\right)\right] \\
&+\left(\frac{s-i}{s}\right)\left(\frac{i}{s}\right)\left(-2\theta\sigma_\varepsilon^2\right) \\
&+\left(\frac{s-i}{s}\right)\left[2\left(-\theta\sigma_\varepsilon^2-\sigma_\varepsilon^2+\theta\sigma_\varepsilon^2-\theta^2\sigma_\varepsilon^2\right)\psi\left(i=1\right)+2\left(-\sigma_\varepsilon^2+2\theta\sigma_\varepsilon^2-\theta^2\sigma_\varepsilon^2\right)\psi\left(i\neq 1, i\neq s\right)\right] \\
&+\left(\frac{i}{s}\right)\left[2\left(\sigma_\varepsilon^2-\theta\sigma_\varepsilon^2+\theta^2\sigma_\varepsilon^2\right)\psi\left(i=1\right)+2\left(\sigma_\varepsilon^2-2\theta\sigma_\varepsilon^2+\theta^2\sigma_\varepsilon^2\right)\psi\left(i\neq 1, i\neq s\right)\right]
\end{aligned}
$$

which after further simplification becomes

$$
\begin{aligned}
\sigma_{s,x}^2(i) =& \sigma_\varepsilon^2\left[\left(\frac{s-i}{s}\right)^2+\left(\frac{i}{s}\right)^2\right]\left[s\left(1+\theta^2\right)-2\left(s-1\right)\theta\right] \\
&+\sigma_\varepsilon^2\left(2+2\theta^2\right)\left[\psi\left(i\neq s\right)\right]-2\sigma_\varepsilon^2\theta i\left(\frac{s-i}{s^2}\right) \\
&+\sigma_\varepsilon^2\left(\frac{s-i}{s}\right)\left[\left(-2-2\theta^2\right)\psi\left(i=1\right)+\left(-2+4\theta-2\theta^2\right)\psi\left(i\neq 1, i\neq s\right)\right] \\
&+\sigma_\varepsilon^2\left(\frac{i}{s}\right)\left[\left(2-2\theta+2\theta^2\right)\psi\left(i=1\right)+\left(2-4\theta+2\theta^2\right)\psi\left(i\neq 1, i\neq s\right)\right]
\end{aligned}
$$

where we use the notation $\sigma_{s,x}^2(i)$ to emphasize that the long variance of the interpolated series also depends on *i*. Following the same steps as above to remove the conditionality of the long variance on *i*, and using the summation for the indicator functions as we did above in this Appendix, we obtain

$$
\begin{aligned}
\sigma_{s,x}^2 =& \sigma_\varepsilon^2\left(\frac{s^2-2s\overline{i}+2\overline{i}^2}{s^2}\right)\left[s\left(1+\theta^2\right)-2\left(s-1\right)\theta\right] \\
&+\sigma_\varepsilon^2\left(2+2\theta^2\right)\left[\left(\frac{1}{s}\right)\sum_{i=1}^{s}\psi\left(i\neq s\right)\right]-2\sigma_\varepsilon^2\theta\left(\frac{s\overline{i}-\overline{i}^2}{s^2}\right) \\
&+\sigma_\varepsilon^2\left(\frac{s-\overline{i}}{s}\right)\left\{\left(-2-2\theta^2\right)\left[\left(\frac{1}{s}\right)\sum_{i=1}^{s}\psi\left(i=1\right)\right]+\left(-2+4\theta-2\theta^2\right)\left[\left(\frac{1}{s}\right)\sum_{i=1}^{s}\psi\left(i\neq 1, i\neq s\right)\right]\right\} \\
&+\sigma_\varepsilon^2\left(\frac{\overline{i}}{s}\right)\left\{\left(2-2\theta+2\theta^2\right)\left[\left(\frac{1}{s}\right)\sum_{i=1}^{s}\psi\left(i=1\right)\right]+\left(2-4\theta+2\theta^2\right)\left[\left(\frac{1}{s}\right)\sum_{i=1}^{s}\psi\left(i\neq 1, i\neq s\right)\right]\right\}
\end{aligned}
$$

where

$$\bar{i} = \left(\frac{1}{s}\right)\sum_{i=1}^{s} i = \frac{s+1}{2} \qquad \text{and} \qquad \overline{i}^{\,2} = \left(\frac{1}{s}\right)\sum_{i=1}^{s} i^2 = \frac{(s+1)(2s+1)}{6}$$

After the substitution, we have

$$\begin{aligned}
\sigma_{s,x}^2 &= \sigma_\varepsilon^2 \left\{ \frac{s^2 - 2\left[s(s+1)/2\right] + 2\left[(s+1)(2s+1)/6\right]}{s^2} \right\} \left[s\left(1+\theta^2\right) - 2(s-1)\theta \right] \\
&\quad + \sigma_\varepsilon^2 \left(2+2\theta^2\right)\left(\frac{s-1}{s}\right) - 2\sigma_\varepsilon^2 \theta \left\{ \frac{s\left[(s+1)/2\right] - \left[(s+1)(2s+1)/6\right]}{s^2} \right\} \\
&\quad + \sigma_\varepsilon^2 \left\{ \frac{s - \left[(s+1)/2\right]}{s} \right\} \left\{ \left(-2-2\theta^2\right)\left(\frac{1}{s}\right) + \left(-2+4\theta-2\theta^2\right)\left(\frac{s-2}{s}\right) \right\} \\
&\quad + \sigma_\varepsilon^2 \left[\frac{(s+1)/2}{s} \right] \left\{ \left(2-2\theta+2\theta^2\right)\left(\frac{1}{s}\right) + \left(2-4\theta+2\theta^2\right)\left(\frac{s-2}{s}\right) \right\}
\end{aligned}$$

which, after simplification yields

$$\sigma_{s,x}^2 = \left(\frac{\sigma_\varepsilon^2}{s^2}\right)\left(\frac{1+\theta^2}{3}\right)\left(2s^3+6s^2+s-6\right) - \left(\frac{\sigma_\varepsilon^2}{s^2}\right)\left(\frac{\theta}{3}\right)\left(4s^3-3s^2+17s-24\right)$$

This is the expression for the long variance that is given in equation (10) in the paper.

Mathematica commands for plotting Figures 1, 2, and 3

Figure 1

$$\text{Plot}[\{(\frac{s(1+(-0.99)^2)+\left(2(1-s)(-0.99)\right)}{s(1+(-0.99)^2)}),$$

$$\left(\frac{s(1+(-0.75)^2)+\left(2(1-s)(-0.75)\right)}{s(1+(-0.75)^2)}\right),$$

$$\left(\frac{s(1+(-0.5)^2)+\left(2(1-s)(-0.5)\right)}{s(1+(-0.5)^2)}\right),$$

$$\left(\frac{s(1+(-0.25)^2)+\left(2(1-s)(-0.25)\right)}{s(1+(-0.25)^2)}\right),$$

$$\left(\frac{s(1+(0)^2)+\left(2(1-s)(0)\right)}{s(1+(0)^2)}\right),$$

$$\left(\frac{s(1+(0.25)^2)+\left(2(1-s)(0.25)\right)}{s(1+(0.25)^2)}\right),$$

$$\left(\frac{s(1+(0.5)^2)+\left(2(1-s)(0.5)\right)}{s(1+(0.5)^2)}\right),$$

$$\left(\frac{s(1+(0.75)^2)+\left(2(1-s)(0.75)\right)}{s(1+(0.75)^2)}\right),$$

$$(\frac{s(1+(0.99)^2)+\left(2(1-s)(0.99)\right)}{s(1+(0.99)^2)})\},$$

$\{s, 2{,}30\}$,

PlotLabels → {"Teta = −0.99", "Teta = −0.75", "Teta = −0.50", "Teta = −0.25", "Teta = 0.00", "Teta = 0.25", "Teta = 0.50", "Teta = 0.75", "Teta = 0.99"}, PlotTheme−> "Monochrome"]

Figure 2

Plot[{$\frac{(1+(-0.99)^2)(2s^3+6s^2+s-6)-(-0.99)(4s^3-3s^2+17s-24)}{3s^2(1+(-0.99)^2)(2s-1)+6s(-0.99)(s^2-3s+1)}$,

$\frac{(1+(-0.75)^2)(2s^3+6s^2+s-6)-(-0.75)(4s^3-3s^2+17s-24)}{3s^2(1+(-0.75)^2)(2s-1)+6s(-0.75)(s^2-3s+1)}$,

$\frac{(1+(-0.5)^2)(2s^3+6s^2+s-6)-(-0.5)(4s^3-3s^2+17s-24)}{3s^2(1+(-0.5)^2)(2s-1)+6s(-0.5)(s^2-3s+1)}$,

$\frac{(1+(-0.25)^2)(2s^3+6s^2+s-6)-(-0.25)(4s^3-3s^2+17s-24)}{3s^2(1+(-0.25)^2)(2s-1)+6s(-0.25)(s^2-3s+1)}$,

$\frac{(1+(0)^2)(2s^3+6s^2+s-6)-(0)(4s^3-3s^2+17s-24)}{3s^2(1+(0)^2)(2s-1)+6s(0)(s^2-3s+1)}$,

$\frac{(1+(0.25)^2)(2s^3+6s^2+s-6)-(0.25)(4s^3-3s^2+17s-24)}{3s^2(1+(0.25)^2)(2s-1)+6s(0.25)(s^2-3s+1)}$,

$\frac{(1+(0.5)^2)(2s^3+6s^2+s-6)-(0.5)(4s^3-3s^2+17s-24)}{3s^2(1+(0.5)^2)(2s-1)+6s(0.5)(s^2-3s+1)}$,

$\frac{(1+(0.75)^2)(2s^3+6s^2+s-6)-(0.75)(4s^3-3s^2+17s-24)}{3s^2(1+(0.75)^2)(2s-1)+6s(0.75)(s^2-3s+1)}$,

$\frac{(1+(0.99)^2)(2s^3+6s^2+s-6)-(0.99)(4s^3-3s^2+17s-24)}{3s^2(1+(0.99)^2)(2s-1)+6s(0.99)(s^2-3s+1)}$},

{s, 2,30},

PlotLabels → {"Teta = −0.99", "Teta = −0.75", "Teta = −0.50", "Teta = −0.25", "Teta = 0.00", "Teta = 0.25", "Teta = 0.50", "Teta = 0.75", "Teta = 0.99"}, PlotTheme−> "Monochrome"]

Figure 3

$$\text{Plot}[\{(\frac{s(1+(-0.99)^2)+\left(2(1-s)(-0.99)\right)}{s(1+(-0.99)^2)} - \frac{(1+(-0.99)^2)(2s^3+6s^2+s-6)-(-0.99)(4s^3-3s^2+17s-24)}{3s^2(1+(-0.99)^2)(2s-1)+6s(-0.99)(s^2-3s+1)}),$$

$$\left(\frac{s(1+(-0.75)^2)+\left(2(1-s)(-0.75)\right)}{s(1+(-0.75)^2)} - \frac{(1+(-0.75)^2)(2s^3+6s^2+s-6)-(-0.75)(4s^3-3s^2+17s-24)}{3s^2(1+(-0.75)^2)(2s-1)+6s(-0.75)(s^2-3s+1)}\right),$$

$$\left(\frac{s(1+(-0.5)^2)+\left(2(1-s)(-0.5)\right)}{s(1+(-0.5)^2)} - \frac{(1+(-0.5)^2)(2s^3+6s^2+s-6)-(-0.5)(4s^3-3s^2+17s-24)}{3s^2(1+(-0.5)^2)(2s-1)+6s(-0.5)(s^2-3s+1)}\right),$$

$$\left(\frac{s(1+(-0.25)^2)+\left(2(1-s)(-0.25)\right)}{s(1+(-0.25)^2)} - \frac{(1+(-0.25)^2)(2s^3+6s^2+s-6)-(-0.25)(4s^3-3s^2+17s-24)}{3s^2(1+(-0.25)^2)(2s-1)+6s(-0.25)(s^2-3s+1)}\right),$$

$$\left(\frac{s(1+(0)^2)+\left(2(1-s)(0)\right)}{s(1+(0)^2)} - \frac{(1+(0)^2)(2s^3+6s^2+s-6)-(0)(4s^3-3s^2+17s-24)}{3s^2(1+(0)^2)(2s-1)+6s(0)(s^2-3s+1)}\right),$$

$$\left(\frac{s(1+(0.25)^2)+\left(2(1-s)(0.25)\right)}{s(1+(0.25)^2)} - \frac{(1+(0.25)^2)(2s^3+6s^2+s-6)-(0.25)(4s^3-3s^2+17s-24)}{3s^2(1+(0.25)^2)(2s-1)+6s(0.25)(s^2-3s+1)}\right),$$

$$\left(\frac{s(1+(0.5)^2)+\left(2(1-s)(0.5)\right)}{s(1+(0.5)^2)} - \frac{(1+(0.5)^2)(2s^3+6s^2+s-6)-(0.5)(4s^3-3s^2+17s-24)}{3s^2(1+(0.5)^2)(2s-1)+6s(0.5)(s^2-3s+1)}\right),$$

$$\left(\frac{s(1+(0.75)^2)+\left(2(1-s)(0.75)\right)}{s(1+(0.75)^2)} - \frac{(1+(0.75)^2)(2s^3+6s^2+s-6)-(0.75)(4s^3-3s^2+17s-24)}{3s^2(1+(0.75)^2)(2s-1)+6s(0.75)(s^2-3s+1)}\right),$$

$$(\frac{s(1+(0.99)^2)+\left(2(1-s)(0.99)\right)}{s(1+(0.99)^2)} - \frac{(1+(0.99)^2)(2s^3+6s^2+s-6)-(0.99)(4s^3-3s^2+17s-24)}{3s^2(1+(0.99)^2)(2s-1)+6s(0.99)(s^2-3s+1)})\},$$

$\{s, 2{,}30\}$,

PlotLabels → {"Teta = −0.99", "Teta = −0.75", "Teta = −0.50", "Teta = −0.25", "Teta = 0.00", "Teta = 0.25", "Teta = 0.50", "Teta = 0.75", "Teta = 0.99"}, PlotTheme−> "Monochrome"]