Estimating persistence for irregularly spaced historical data

Philip Hans Franses

Accepted: 29 January 2021 / Published online: 17 February 2021
© The Author(s) 2021

Abstract
This paper introduces to the literature on Economic History a measure of persistence which is particularly useful when the data are irregularly spaced. An illustration to ten historical unevenly spaced data series for Holland of 1738 to 1779 shows the merits of the methodology. It is found that the weight of slave-based contribution in that period has grown with a deterministic trend pattern.

Keywords Irregularly spaced time series · Economic history · Slave trade · First order autoregression · Persistence

JEL Code C32 · N01

1 Introduction and motivation

One way to study economic history amounts to the construction and analysis of historical time series data, see for example van Zanden and van Leeuwen (2012) amongst many others. A particularly interesting period to study concerns the times of the Atlantic slave trade. One of the aspects of frequent examination concerns the contribution of slave trade to the size of an economy. Recent important studies are Eltis and Engerman (2000). Fatah-Black and van Rossum (2015) and Eltis et al. (2016). Another recent study is Brandon and Bosma (2019) who shows that 5 to 10% of Gross Domestic Product (GDP) in Holland around 1770 was based on slave trade, see Table 1.

An important feature to study concerns the trends in the data. Did the contribution to GDP of slave trade grow with a steady pace, like with a deterministic trend? Or, did that contribution jump to plateaus due to structural breaks, perhaps caused by technological developments? If it would be along a deterministic trend, then shocks to the data were not persistent. If the growth patterns followed sequences of structural breaks, then those shocks were persistent. Hence, it is of interest to study the persistence properties of the historical data.
Ideally, the constructed historical data are equally spaced, like per year or per ten years, as then basic time series analytical tools can be used to study the properties of the data. In the present paper the focus is on the analysis of unequally spaced data, which can also occur in historical research, as will be evident below.

### 2 Introductory remarks

An important property of time series data is, what is called, the persistence of shocks. Such persistence is perhaps best illustrated when we consider the following simple time series model for a variable $y_t$, which is observed for a sequence of $T$ years, $t = 1, 2, \ldots, T$, that is,

$$y_t = \alpha y_{t-1} + \epsilon_t$$

This model is called a first order autoregression, with acronym AR(1). The $\epsilon_t$ is a series of shocks (or news) that drives the data over time, and these shocks have mean 0 and common variance $\sigma^2$, and over time these shocks are uncorrelated. In other words, future shocks or news cannot be predicted from past shocks or news. The $\alpha$ is an unknown parameter that needs to be estimated from the data. Usually one relies on the ordinary least squares (OLS) method to estimate this parameter, see for example Franses et al. (2014, Chapter 3) for details.

In anAR(1) model,\(^1\) the persistence of shocks to $y_t$ is reflected by (functions of) the parameter $\alpha$. This is best understood by explicitly writing down all the observations on $y_t$ when the AR(1) is the model for these data. The first observation is then

$$y_1 = \alpha y_0 + \epsilon_1$$

---

\(^1\) If one were to consider an autoregression of higher order, then the measure of persistence is the sum of the autoregressive coefficients. One may also want to consider so-called fractionally integrated time series models, where the degree of differencing $d$ is a measure of persistence. Nonparametric methods to measure persistence also exist, like the number of times a time series crosses its mean value.
where $y_0$ is some known starting value, that can be equal to 0 or not. In practice this starting value is usually taken as the first available observation, and then the estimation sample runs from $t = 2, 3, 4 \ldots, T$. The second observation is

$$y_2 = \alpha y_1 + \epsilon_2 = \alpha^2 y_0 + \epsilon_2 + \alpha \epsilon_1$$

where the expression on the right-hand side now incorporates the expression for $y_1$. When this recursive inclusion of past observations is continued, we have for any $y_t$ observation that

$$y_t = \alpha^t y_0 + \epsilon_t + \alpha \epsilon_{t-1} + \alpha^2 \epsilon_{t-2} + \alpha^3 \epsilon_{t-3} + \ldots + \alpha^{t-1} \epsilon_1$$

This expression shows that the immediate impact of a shock $\epsilon_t$ is equal to 1. The impact of a shock one period ago (which is $\epsilon_{t-1}$) is $\alpha$ and the impact of a shock $j$ periods ago is $\alpha^j$. The total effect of a shock if $t \to \infty$ is thus

$$1 + \alpha + \alpha^2 + \alpha^3 + \ldots = \frac{1}{1 - \alpha}$$

when $|\alpha| < 1$. So, when $\alpha = 0.5$, the total effect of a shock is 2. When $\alpha = 0.9$, the total effect is 10. So, when $\alpha$ approaches 1, the impact gets larger. When $\alpha = 1$, the total effect is infinite. At the same time, when $\alpha = 1$, each shock in the past has the same permanent effect 1, as $1^j = 1$. In that case, shocks are said to have a permanent effect.

One may also be interested in, what is called, a duration interval. For example, a 95% duration interval is the time period $\tau_{0.95}$ within which 95% of the cumulative or total effect of a shock has occurred. It is defined by

$$\tau_{0.95} = \frac{\log(1 - 0.95)}{\log(\alpha)}$$

![GDP](image-url)
where log denotes the natural logarithm. When $\alpha = 0.5$, the $\tau_{0.95} = 4.32$, and when $\alpha = 0.9$, the $\tau_{0.95} = 28.4$. These persistence measures are informative about how many years (or periods) shocks last.

3 Motivation of this paper

In this paper the focus is on persistence measures in case the data do not involve a connected sequence of years but instead concern data with missing data at irregular intervals. Consider for example the data on Gross Domestic Product (GDP) in Holland for the sample 1738–1779 in Fig. 1. In principle the sample size is 42, but it is clear that various years with data are missing, and hence the sample effectively covers 24 years. Take for example the data in the final column of Table 2, which concern the Weights of slave-based activities in GDP Holland, for the sample 1738–1779. The data are in Fig. 2. The issue is now how we can construct persistence measures, that is, functions of $\alpha$ like above, when the data follow a first order autoregression for such irregularly spaced data.

The paper proceeds as follows. The next section presents a useful model for unevenly spaced data. It also deals with a step-by-step illustration of how to implement this method, which can be done using any statistical package. The empirical section implements this method for ten variables with irregularly spaced data, all of which appeared in a recent study of Brandon and Bosma (2019) on the economic impact of the Atlantic slave trade. The final section concludes.

4 Methodology

The starting point of our analysis is the representation of an AR(1) process given in Robinson (1977) (see also for example Schulz and Mudelsee, 2002). Suppose an AR(1) process is observed at times $t_i$ where $i = 1, 2, 3, \ldots, N$. A general expression for an AR(1) process with arbitrary time intervals is

$$y_{ti} = \alpha_i y_{t_{i-1}} + \epsilon_{ti}$$

with

$$\alpha_i = \exp\left(-\frac{t_i - t_{i-1}}{\tau}\right)$$

where $\tau$ is scaling the memory, see Robinson (1977). For easy of analysis, it is assumed here that $\epsilon_{ti}$ is a white noise uncorrelated process with mean 0 but with time-variation in the variance.$^2$ This means that in practice, one should correct for this heteroskedasticity by using the Newey West (1987) HAC estimator.

One may continue with (1) and (2), but it may be easier to define

$^2$ In Robinson (1977) it is assumed that the variance of the error process is.

$$\sigma^2 = 1 - \exp\left(-\frac{2(t_i - t_{i-1})}{\tau}\right)$$

so that the variance of $y_{ti}$ is equal to 1. Here there is no need to make this assumption.
Table 2 The data

| Year | IT | IS | DP | SB | SR | NO | AN | VA | GDP | PGDP |
|------|----|----|----|----|----|----|----|----|-----|------|
| 1738 | 3065 | 836 | 722 | 309 | 1208 | 220 | 274 | 6634 | 132,494 | 5 |
| 1739 | 2807 | 771 | 661 | 273 | 959 | 220 | 278 | 5969 | 133,983 | 4.5 |
| 1740 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1741 | 4281 | 1192 | 1008 | 352 | 1281 | 222 | 327 | 8663 | 145,374 | 6 |
| 1742 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1743 | 2936 | 826 | 691 | 271 | 748 | 222 | 445 | 6139 | 141,094 | 4.4 |
| 1744 | 4318 | 1187 | 1016 | 331 | 1022 | 222 | 530 | 8626 | 154,306 | 5.6 |
| 1745 | 4705 | 1309 | 1108 | 616 | 938 | 223 | 610 | 9509 | 141,286 | 6.7 |
| 1746 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1747 | 6723 | 1875 | 1583 | 1071 | 990 | 223 | 780 | 13,245 | 191,910 | 6.9 |
| 1748 | 5578 | 1562 | 1313 | 679 | 1239 | 226 | 1187 | 11,784 | 176,145 | 6.7 |
| 1749 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1750 | 5042 | 1314 | 1187 | 465 | 2017 | 225 | 542 | 10,793 | 144,076 | 7.5 |
| 1751 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1752 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1753 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1754 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1755 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1756 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1757 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1758 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1759 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1760 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1761 | 12,644 | 3549 | 2976 | 1231 | 1474 | 221 | 352 | 22,548 | 155,733 | 14.5 |
| 1762 | 13,501 | 3793 | 3178 | 1720 | 1336 | 221 | 344 | 24,193 | 161,720 | 15 |
| 1763 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1764 | 9131 | 2401 | 2149 | 996 | 1550 | 221 | 324 | 17,152 | 171,071 | 10 |
| 1765 | 9824 | 2544 | 2313 | 1111 | 1384 | 220 | 309 | 18,264 | 183,898 | 9.9 |
| 1766 | 6707 | 1880 | 1579 | 714 | 1151 | 222 | 306 | 12,720 | 172,727 | 7.4 |
| 1767 | 10,290 | 2714 | 2422 | 897 | 907 | 221 | 299 | 18,022 | 167,985 | 10.7 |
| 1768 | 10,538 | 2826 | 2481 | 1202 | 890 | 224 | 328 | 18,711 | 170,075 | 11 |
| 1769 | 11,909 | 3169 | 2804 | 1268 | 1005 | 222 | 319 | 20,947 | 182,748 | 11.5 |
| 1770 | 10,620 | 2710 | 2500 | 975 | 682 | 222 | 334 | 18,340 | 177,069 | 10.4 |
| 1771 | 14,558 | 3972 | 3427 | 1605 | 996 | 221 | 343 | 25,332 | 214,067 | 11.8 |
| 1772 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1773 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1774 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1775 | 11,144 | 2904 | 2623 | 1256 | 961 | 226 | 334 | 19,448 | 185,987 | 10.5 |
| 1776 | 13,078 | 3239 | 3079 | 1203 | 822 | 226 | 363 | 22,009 | 181,702 | 12.1 |
| 1777 | 15,174 | 3768 | 3572 | 1569 | 893 | 224 | 406 | 25,626 | 185,981 | 13.8 |
| 1778 | 16,173 | 4239 | 3807 | 1837 | 621 | 246 | 407 | 27,330 | 184,359 | 14.8 |
| 1779 | 20,060 | 5578 | 4722 | 1878 | 692 | 250 | 373 | 33,554 | 171,710 | 19.5 |
This makes that the general AR (1) model can be written as
\[ y_t = \alpha^{-1} y_{t-1} \epsilon_t \]

When the data would be regularly spaced, then \( t_i - t_{i-1} = 1 \) and this model collapses into
\[ y_t = \alpha y_{t-1} + \epsilon_t \]

which is the standard AR(1) model above. Or, suppose the data would be unequally spaced because of selective sampling each even observation, and all the odd observations would be called as missing, then \( t_i - t_{i-1} = 2 \), and then the model reads as
\[ y_t = \alpha^2 y_{t-2} + \epsilon_t \]

Before one proceeds with estimating the parameter in (3), one first needs to demean and detrend the data, see Robinson (1977).

5 Estimation

Given a sample \( \{ t_i, y_t \} \), one can use Nonlinear Least Squares (NLS) to estimate \( \alpha \) (and hence \( \tau \)). Table 3 provides the key variables relevant for estimation concerning the variable in Fig. 2. The first column gives the demeaned and detrended irregularly spaced time series, that is \( x_{i-t} \), where this variable follows from the OLS regression
\[ y_t = \mu + \delta t + x_{i-t} \]
Table 3 Numerical example. PGDPDMDT means Weight of slave-based activities in GDP Holland, after demeaning (DM) and detrending (DT). DIFT is $t_i - t_{i-1}$

| Year | PGDPDMDT | DIFT | PGDPDMDT(-DIFT) |
|------|----------|------|-----------------|
| 1738 | 0.075744 | 1    | NA              |
| 1739 | −0.736111| 1    | 0.075744        |
| 1740 | NA       | 1    | −0.736111       |
| 1741 | 0.446689 | 2    | −0.736111       |
| 1742 | NA       | 1    | 0.446689        |
| 1743 | −1.632230| 2    | 0.446689        |
| 1744 | −0.682778| 1    | −1.632230       |
| 1745 | 0.333192 | 1    | −0.682778       |
| 1746 | NA       | 1    | 0.333192        |
| 1747 | 0.072340 | 2    | 0.333192        |
| 1748 | −0.388786| 1    | 0.072340        |
| 1749 | NA       | 1    | −0.388786       |
| 1750 | 0.039440 | 2    | −0.388786       |
| 1751 | NA       | 1    | 0.039440        |
| 1752 | NA       | 2    | 0.039440        |
| 1753 | NA       | 3    | 0.039440        |
| 1754 | NA       | 4    | 0.039440        |
| 1755 | NA       | 5    | 0.039440        |
| 1756 | NA       | 6    | 0.039440        |
| 1757 | NA       | 7    | 0.039440        |
| 1758 | NA       | 8    | 0.039440        |
| 1759 | NA       | 9    | 0.039440        |
| 1760 | NA       | 10   | 0.039440        |
| 1761 | 4.721054 | 11   | 0.039440        |
| 1762 | 5.723825 | 1    | 4.721054        |
| 1763 | NA       | 1    | 5.723825        |
| 1764 | −0.422644| 2    | 5.723825        |
| 1765 | −0.824347| 1    | −0.422644       |
| 1766 | −3.920984| 1    | −0.824347       |
| 1767 | −0.391753| 1    | −3.920984       |
| 1768 | −0.289695| 1    | −0.391753       |
| 1769 | 0.040840 | 1    | −0.289695       |
| 1770 | −1.456449| 1    | 0.040840        |
| 1771 | −0.097761| 1    | −1.456449       |
| 1772 | NA       | 1    | −0.097761       |
| 1773 | NA       | 2    | −0.097761       |
| 1774 | NA       | 3    | −0.097761       |
| 1775 | −2.231562| 4    | −0.097761       |
| 1776 | −0.958341| 1    | −2.231562       |
| 1777 | 0.743064 | 1    | −0.958341       |
| 1778 | 1.644795 | 1    | 0.743064        |
| 1779 | NA       | 1    | 1.644795        |
where \( t = 1, 2, 3, \ldots, T \) with \( T = 42 \) here. The demeaned and detrended data are in Fig. 3. The next column in Table 3 contains the \( t_i - t_{i-1} \) with acronym DIFT. The last column of Table 3 reflects the new variable \( x_{t_i-1} \). With this new variable, one can apply NLS to

\[
x_{t_i} = \alpha^{t_i-t_{i-1}}x_{t_{i-1}} + u_{t_i}
\]

and obtain an estimate of \( \alpha \) and an associated HAC standard error.
6 Illustration

Let us see how this works out for the ten historical series in Table 2, which are taken from Brandon and Bosma (2019, Annex page XXX). Table 4 reports the estimation results for the auxiliary regression for demeaning and detrending. Two series do not seem to have a trend as the associated parameter is not significant at the 5% level, and these are Sugar refinery and Army and Navy. However, we do use the residuals of the auxiliary regressions in the subsequent analysis.

Table 5 reports on the estimated \( \alpha \) parameters. The estimates range from 0.278 (Total size GDP of Holland) to 0.907 (Sugar refinery). Comparing the estimated parameters with their associated HAC standard errors, we see that 0 is included in the 95% confidence interval only for Total size GDP of Holland. So, this variable fully follows a deterministic trend.

Table 6 presents the estimated persistence of shocks (news), measured the 95% duration interval \( \tau_{0.95} \) and by \( \tau \). Clearly, persistence is largest for Sugar refinery and Notaries. The parameter for Notaries is 0.862 (Table 5) is very close to 1, given its HAC standard error, so one might even claim that shocks to this sector in the observed period were permanent.

**Table 5** Estimate of persistence (with estimated HAC standard errors in parentheses, Newey and West, 1987) using NLS to the regression model \( x_t = \alpha t^{-\kappa} x_{t-1} + u_t \)

| Variable                                         | \( \hat{\alpha} \) | HAC standard error |
|--------------------------------------------------|--------------------|--------------------|
| International trade                              | 0.416 (0.165)      |
| International shipping                           | 0.437 (0.181)      |
| Domestic production, trade and shipping          | 0.416 (0.165)      |
| Shipbuilding                                     | 0.348 (0.171)      |
| Sugar refinery                                   | 0.907 (0.033)      |
| Notaries                                         | 0.862 (0.099)      |
| Army and Navy                                    | 0.675 (0.198)      |
| Total slave-based value added                    | 0.404 (0.167)      |
| Total size GDP of Holland                        | 0.278 (0.149)      |
| Weight of slave-based activities in GDP Holland  | 0.536 (0.152)      |

**Table 6** Measures of persistence, measured in years

| Variable                                         | \( \tau_{0.95} \) | \( \tau \) |
|--------------------------------------------------|-------------------|------------|
| International trade                              | 3.42              | 1.14       |
| International shipping                           | 3.62              | 1.21       |
| Domestic production, trade and shipping          | 3.42              | 1.14       |
| Shipbuilding                                     | 2.84              | 0.947      |
| Sugar refinery                                   | 30.7              | 10.2       |
| Notaries                                         | 20.2              | 6.73       |
| Army and Navy                                    | 7.62              | 2.54       |
| Total slave-based value added                    | 3.31              | 1.10       |
| Total size GDP of Holland                        | 2.34              | 0.781      |
| Weight of slave-based activities in GDP Holland  | 4.80              | 1.60       |
7 Conclusion

This paper has introduced to the literature on Economic History a measure of persistence which is particularly useful if the data are irregularly spaced. An illustration to ten historical series for the impact and contribution of slave trade in Holland of 1738–1779 showed the merits of the methodology.

When the question is addressed whether the contribution to GDP of slave trade has grown with a steady pace, like with a deterministic trend, or whether that contribution jumped to plateaus due to structural breaks, perhaps caused by technological developments, the following conclusion can be drawn. The persistence in the variables “Weight of slave-based activities in GDP Holland”, as measured by the parameters in an AR (1) regression, is equal to 0.536 with HAC standard error 0.214. This persistence is not equal to 1, meaning that there is no sign of occasional structural breaks with a long-lasting effect. Hence, in the considered period, the contribution to GDP has steadily grown with a deterministic pattern.

Further applications should emphasize the practical relevance of the method. Also, an extension to an autoregressive process of higher order could be relevant, in order to provide additional measures of persistence. An extension to fractionally integrated processes is also relevant. Finally, and this a further technical issue, that is, one may want to formally test if $\alpha = 1$. This amounts to a so-called test for a unit root, for which the asymptotic theory is different than standard, see for example Chapter 4 of Franses et al. (2014).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Brandon, P., Bosma, U.: Calculating the weight of slave-based activities in the GDP of Holland and the Dutch Republic–Underlying methods, data and assumptions. Low Ctries. J. Soc. Econ. Hist. 16(2), 5–45 (2019). https://doi.org/10.18352/tseg.1082

Eltis, D., Engerman, S.L.: The importance of slavery and the slave trade to industrializing Britain. J. Econ. Hist. 60(1), 123–144 (2000)

Eltis, D., Emmer, P.C., Lewis, F.D.: More than profits? The contribution of the slave trade to the Dutch economy: assessing Fatah-Black and Van Rossum. Slavery Abolit. 37(4), 724–735 (2016)

Fatah-Black, K., van Rossum, M.: Beyond profitability: The Dutch transatlantic slave trade and its economic impact. Slavery Abolit. 36(1), 63–83 (2015)

Franses, P.H., van Dijk, D.J., Opschoor, A.: Time Series Models for Business and Economic Forecasting. Cambridge University Press, Cambridge UK (2014)

Newey, W.K., West, K.D.: A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55(3), 703–708 (1987)

Robinson, P.M.: Estimation of a time series model from unequally spaced data. Stoch. Process. Appl. 6, 9–24 (1977)

Schulz, M., Mudelsee, M.: REDFIT: estimating red-noise spectra directly from unevenly spaced paleoclimatic time series. Comput. Geosci. 28, 421–426 (2002)
Van Zanden, J.L., van Leeuwen, B.: Persistent but not consistent. The growth of national income in Holland, 1347–1807. Explor. Econ. Hist. 49, 119–130 (2012)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.