Estimating persistence for irregularly spaced historical data

Philip Hans Franses

Econometric Institute

Erasmus School of Economics

EI2020-03

Abstract

This paper introduces to the literature on Economic History a measure of persistence which is particularly useful if the data are irregularly spaced. An illustration to 10 historical unevenly spaced data series for Holland of 1738 to 1779 showed the merits of the methodology.

Key words: Irregularly spaced time series; Economic history; First order autoregression; Persistence

JEL code: C32; N01

This version: September 2019

Correspondence: Econometric Institute, Erasmus School of Economics, POB 1738 , NL-3000 DR Rotterdam, the Netherlands, franses@ese.eur.nl, +31104081273

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. Ideally, the constructed data are equally spaced, like per year of per ten years, as then basic time series analytical tools can be used to study the properties of the data. In this paper now the focus is on the analysis of *unequally* spaced data, which can occur in historical research.

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, $t = 1,2,\dots,T$, that is,

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

This model is called a first order autoregression, with acronym AR(1). The ε_t is a series of shocks (or news) that drives the data over time, and these shocks have mean 0 and common variance σ_ε^2, and over time these shocks are uncorrelated. In words, future shock or news cannot be predicted from past shocks or news. The α 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 Franses, van Dijk and Opschoor (2014, Chapter 3) for details.

The persistence of shocks to y_t is reflected by (functions of) the parameter α. 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

$$y_1 = \alpha y_0 + \varepsilon_1$$

where y_0 is some starting value[1]. The second observation would be

$$y_2 = \alpha y_1 + \varepsilon_2 = \alpha^2 y_0 + \varepsilon_2 + \alpha\varepsilon_1$$

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

$$y_t = \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2\varepsilon_{t-2} + \alpha^3\varepsilon_{t-3} + \cdots + \alpha^j\varepsilon_{t-j}$$

This expression shows that the immediate impact of a shock ε_t is equal to 1. The impact of a shock one period ago (which is ε_{t-1}) is α and the impact of a shock j periods ago is α^j. The total effect of a shock is thus

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

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 α 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

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

[1] In practice this starting value is usually taken as the first available observation, and then the estimation sample runs from $t = 2,3,4 \ldots, T$.

where log denotes the natural logarithm. When $\alpha = 0.5$, the $\tau_{0.95} = 4.322$, and when $\alpha = 0.9$, the $\tau_{0.95} = 28.433$. These persistence measures are informative about how many years (or periods) shocks last.

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 Figure 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. The issue is now how we can construct persistence measures, that is, functions of α like above, when the data follow a first order autoregression for such irregularly spaced data.

The paper proceeds as follows. The next section describes such 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 10 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.

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

$$y_{t_i} = \alpha_i y_{t_{i-1}} + \varepsilon_{t_i}$$

with

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

For easy of analysis, it is assumed here that ε_{t_i} is a white noise uncorrelated process with mean 0 and common variance[2]. Robinson (1977) defines τ as a measure of memory.

When we define

$$\alpha = \exp\left(-\frac{1}{\tau}\right)$$

the general AR(1) model can be written as

$$y_{t_i} = \alpha^{t_i - t_{i-1}} y_{t_{i-1}} + \varepsilon_{t_i}$$

When the data would be regularly spaced, then $t_i - t_{i-1} = 1$ and this model collapses into

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

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 the model reads as

$$y_t = \alpha^2 y_{t-2} + \varepsilon_t$$

---

[2] In Robinson (1977) it is assumed that the variance of the error process is

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

so that the variance of y_{t_i} is equal to 1. Here there is no need to make this assumption.

Estimation

Given a sample $\{t_i, y_{t_i}\}$, one can use Nonlinear Least Squares to estimate α (and hence τ). 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 Figure 2.

Table 3 presents the key variables relevant for estimation. The first column gives the demeaned and detrended irregularly spaced time series, that is x_{t_i}, where this variable follows from the OLS regression

$$y_{t_i} = \mu + \delta t + x_{t_i}$$

where $t = 1,2,3,\dots,T$ with $T = 42$ here. The demeaned and detrended data are in Figure 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 Nonlinear Least Squares to

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

and obtain an estimate of α and an associated standard error.

Illustration

Let us see how this works out for the 10 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 parameter is nog significant at the 5% level, and these are Sugar refinery and Army and Navy, but we do use the residuals of the auxiliary regressions in the subsequent analysis.

Table 5 reports on the estimated α parameters. The estimates range from 0.278 (Total size GDP of Holland) to 0.955 (Notaries). Comparing the estimated parameters with their associated standard errors, we see that 0 is included in the 95% confidence interval for International Trade, International Shipping, Domestic production, trade and shipping, Shipbuilding, and Total size GDP of Holland.

Table 6 presents the estimated persistence of shocks (news), measured by τ and the 95% duration interval $\tau_{0.95}$. Clearly, persistence is largest for Sugar refinery and Notaries. The parameter for Notaries 0.955 (Table 5) is very close to 1, so one might even claim that shocks to the Notaries sector in the observed period were permanent.

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 10 historical series for Holland of 1738 to 1779 showed the merits of the methodology.

Further applications should emphasize the practical relevance of the method. Also, an extension to an autoregressive process of higher order could be 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, van Dijk and Opschoor (2014).

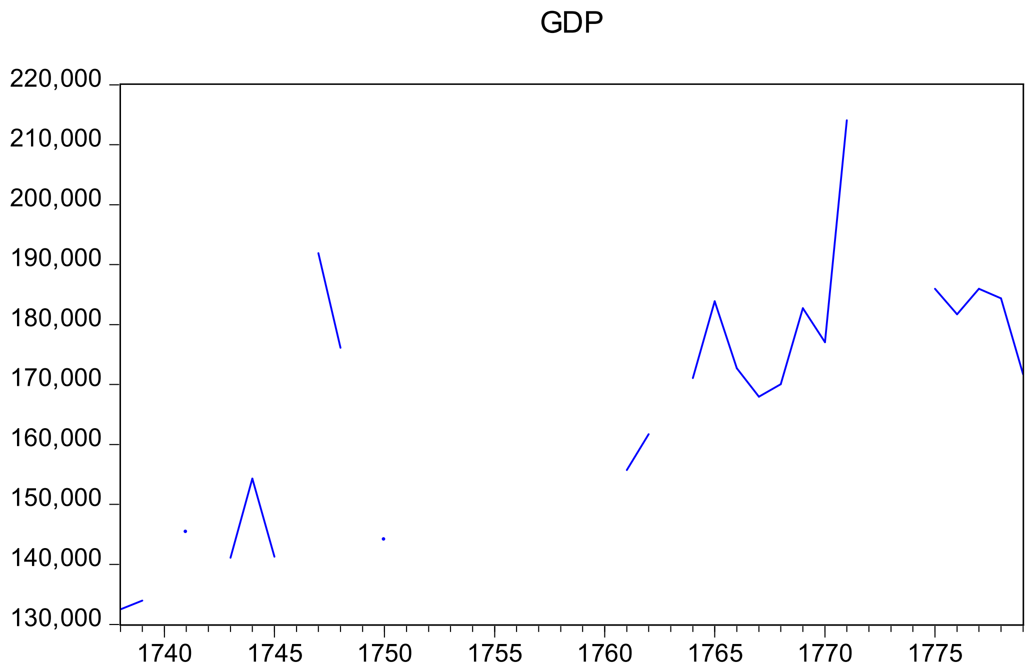

Figure 1: Total size GDP of Holland, 1738-1779

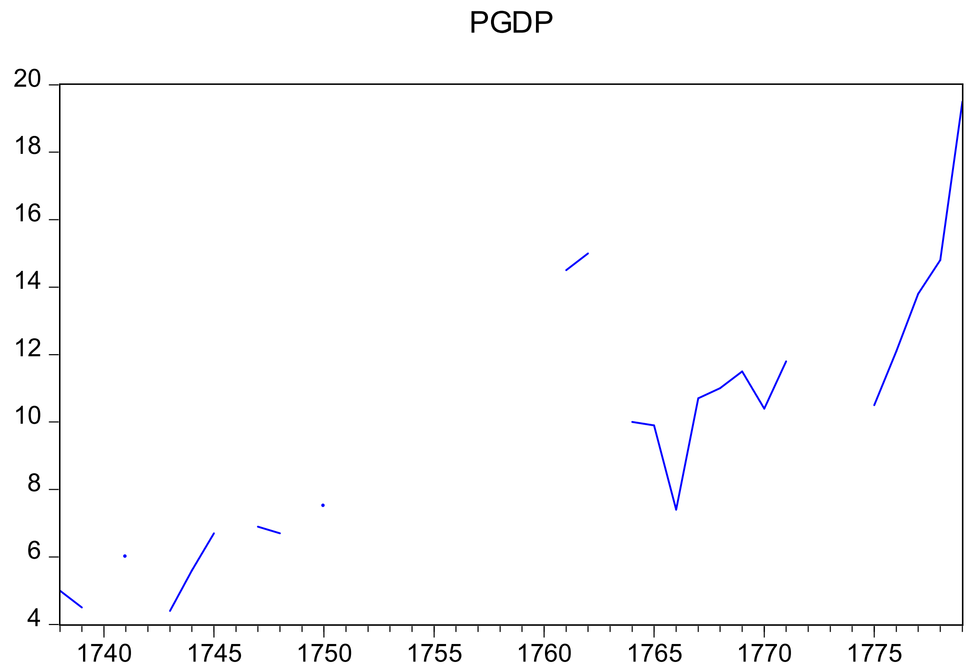

Figure 2: Weight of slave-based activities in GDP Holland, 1738-1779

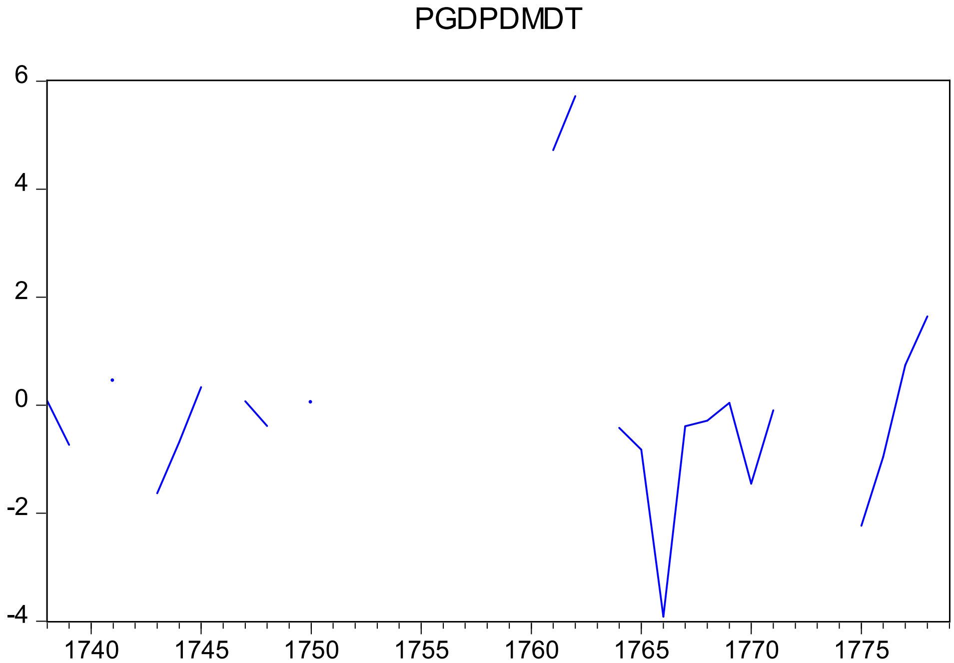

Figure 3: Weight of slave-based activities in GDP Holland, demeaned and detrended (DMDT), 1738-1779

Table 1: The variables

The variables	Acronym
International trade	IT
International shipping	IS
Domestic production, trade and shipping	DP
Shipbuilding	SB
Sugar refinery	SR
Notaries	NO
Army and Navy	AN
Total slave-based value added	VA
Total size GDP of Holland	GDP
Weight of slave-based activities in GDP Holland	PGDP

Source:

Brandon, P. and U. Bosma (2019), Calculating the weight of slave-based activities in the GDP of Holland and the Dutch Republic – Underlying methods, data and assumptions, *The Low Countries Journal of Social and Economic History*, 16 (2), 5-45, doi: 10.18352/tseg.1082

There is one other variable in the dataset, called Banking, but for this variable the sample is too small.

Table 2: The data

	IT	IS	DP	SB	SR	NO	AN	VA	GDP	PGDP
1738	3065	836	722	309	1208	220	274	6634	132494	5
1739	2807	771	661	273	959	220	278	5969	133983	4.5
1740	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1741	4281	1192	1008	352	1281	222	327	8663	145374	6
1742	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1743	2936	826	691	271	748	222	445	6139	141094	4.4
1744	4318	1187	1016	331	1022	222	530	8626	154306	5.6
1745	4705	1309	1108	616	938	223	610	9509	141286	6.7
1746	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1747	6723	1875	1583	1071	990	223	780	13245	191910	6.9
1748	5578	1562	1313	679	1239	226	1187	11784	176145	6.7
1749	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1750	5042	1314	1187	465	2017	225	542	10793	144076	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	12644	3549	2976	1231	1474	221	352	22548	155733	14.5
1762	13501	3793	3178	1720	1336	221	344	24193	161720	15
1763	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1764	9131	2401	2149	996	1550	221	324	17152	171071	10
1765	9824	2544	2313	1111	1384	220	309	18264	183898	9.9
1766	6707	1880	1579	714	1151	222	306	12720	172727	7.4
1767	10290	2714	2422	897	907	221	299	18022	167985	10.7
1768	10538	2826	2481	1202	890	224	328	18711	170075	11
1769	11909	3169	2804	1268	1005	222	319	20947	182748	11.5
1770	10620	2710	2500	975	682	222	334	18340	177069	10.4
1771	14558	3972	3427	1605	996	221	343	25332	214067	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	11144	2904	2623	1256	961	226	334	19448	185987	10.5
1776	13078	3239	3079	1203	822	226	363	22009	181702	12.1
1777	15174	3768	3572	1569	893	224	406	25626	185981	13.8
1778	16173	4239	3807	1837	621	246	407	27330	184359	14.8
1779	20060	5578	4722	1878	692	250	373	33554	171710	19.5

Table 3: Numerical example. PGDPDMDT means Weight of slave-based activities in GDP Holland, after demeaning (DM) and detrending (DT). DIFT is $t_i - t_{i-1}$

	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

Table 4: Regression on intercept and trend (with estimated standard errors in parentheses) using the regression $y_{t_i} = \mu + \delta t + x_{t_i}$

Variable	$\hat{\mu}$		$\hat{\delta}$	
International trade	2190	(839)	310	(31.3)
International shipping	656	(252)	80.0	(9.39)
Domestic production, trade and shipping	516	(197)	73.0	(7.36)
Shipbuilding	268	(111)	31.4	(4.12)
Sugar refinery	1250	(125)	-7.64	(4.66)
Notaries	219	(2.75)	0.24	(0.103)
Army and Navy	535	(78.3)	-4.93	(2.92)
Total slave-based value added	5654	(1378)	486	(51.4)
Total size GDP of Holland	142517	(5762)	1094	(215)
Weight of slave-based activities in GDP Holland	4.38	(0.878)	0.236	(0.033)

Table 5: Estimate of persistence (with estimated standard errors in parentheses) using NLS to the regression model $x_{t_i} = \alpha^{t_i - t_{i-1}} x_{t_{i-1}} + u_{t_i}$

Variable	$\hat{\alpha}$	
International trade	0.416	(0.242)
International shipping	0.437	(0.248)
Domestic production, trade and shipping	0.416	(0.242)
Shipbuilding	0.348	(0.237)
Sugar refinery	0.907	(0.072)
Notaries	0.955	(0.170)
Army and Navy	0.675	(0.120)
Total slave-based value added	0.404	(0.240)
Total size GDP of Holland	0.278	(0.281)
Weight of slave-based activities in GDP Holland	0.536	(0.214)

Table 6: Measures of persistence, measured in years

Variable	$\tau_{0.95}$	τ
International trade	1.14	3.42
International shipping	1.21	3.62
Domestic production, trade and shipping	1.14	3.42
Shipbuilding	0.947	2.84
Sugar refinery	10.2	30.1
Notaries	21.7	65.1
Army and Navy	2.54	7.62
Total slave-based value added	1.10	3.31
Total size GDP of Holland	0.78	2.34
Weight of slave-based activities in GDP Holland	1.60	4.80

References

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

Robinson, P.M. (1977), Estimation of a time series model from unequally spaced data, *Stochastic Processes and their Applications,* 6, 9-24.

Schulz, M. and M. Mudelsee (2002), REDFIT: estimating red-noise spectra directly from unevenly spaced paleoclimatic time series, *Computers & Geosciences*, 28, 421-426.

Van Zanden, J.L. and B. van Leeuwen (2012), Persistent but not consistent. The growth of national income in Holland, 1347-1807, *Explorations in Economic History*, 49, 119-130.