Construction of Annual Data Sets for the Econometric Analysis of Long-Run Money Demand

Bachelor Thesis

Department of Economics

University of Klagenfurt

Markus Köfler

Supervised by

Univ.-Prof. Dipl.-Ing. Dr. Martin Wagner

June 13, 2023

Contents

Abstract

Collecting annual economic time series data presents several challenges, including converting monthly or quarterly series to annual observations, changing base years, adjusting levels, linking series together, and dealing with missing values. This thesis primarily provides explanations to the methodologies applied to the gathered data, though, also serves as a guide for researchers and practitioners seeking to overcome similar obstacles when compiling economic data sets. First, the distinct characteristics of economic variables collected will be examined in a conceptual manner and related to the proper methods to temporally aggregate series from high to low frequency. Also, simple splicing and linking procedures for constructing long-run series will be addressed. Finally, a univariate imputation algorithm, which has been deployed to fill out gaps in the data, will be proposed. A thorough documentation of the data and the Python program code are provided in the appendix for replication purposes.

Abbreviations

ABS	Australian Bureau of Statistics
ANN	Artificial Neural Networks
ARIMA	Auto-Regressive Integrated Moving Average
BEA	Bureau of Economic Analysis
BLS	Bureau of Labor Statistics
BOC	Bank of Canada
BOE	Bank of England
BOJ	Bank of Japan
BOK	Bank of Korea
CPI	Consumer Price Index
ECB	European Central Bank
ECOS	Economic Statistics System
ESRI	Economic and Social Research Institute
EWMA	Exponential Weighted Moving Average
GDP	Gross Domestic Product
HCPI	Harmonized Consumer Price Index
IFS	International Financial Statistics
IMF	International Monetary Fund
LSTM	Long-Short Term Memory
LWMA	Linear Weighted Moving Average
MSE	Mean Squared Error

OECD	Organisation for Economic Co-operation and Development
ONS	Office for National Statistics
PCE	Private Consumption Expenditure
RBA	Reserve Bank of Australia
RMSE	Root Mean Squared Error
SECO	State Secretariat for Economic Affairs
SMA	Simple Moving Average
SNB	Swiss National Bank

List of Figures

1 Introduction

For the research project *Instability and Nonlinearity of Long-Run Money Demand: Econometric Theory and Empirical Analysis* lead by my supervisor*, a data set consisting of a wide range of countries covering extensive periods must be constructed to perform empirical analysis. The economic variables of interest are the Gross Domestic Product (GDP) and a form of consumption expenditure from the national accounts category both in nominal and real terms, monetary aggregates (at least one narrow and one broad aggregate), one series for each short and long term interest rates (in most cases, 3-month treasury bill rates and 10-year government bond yields), Consumer Price Index (CPI) and lastly, the unemployment rate.‡
A large data set fulfilling most of the criteria has been made publicly available by Benati et al. (2019) and can be retrieved from the Federal Reserve Bank of Minneapolis' database (link to spreadsheet) complemented by an appendix[1] including the documentation. This data set has been taken as a starting point. Since Benati et al. (2019) collected data only until 2014 to 2017, each series needed to be continued up to present. Relevant variables not yet contained in this data set needed to be added. Within the scope of this thesis, I contributed to the data collection by gathering data for eight economies, whereby a total of 124 time series were involved. Most time series have been obtained from primary sources as in national statistics agencies or central banks; in a few exceptions, from large organizations like International Monetary Fund (IMF) or the World Bank, if not otherwise possible.

Identifying the data on the web, however, was just a small fraction of the entire work. Many series were provided in monthly or quarterly periodicity, and therefore had to be aggregated to annual frequency. There is no one-fits-all method to achieve this, as it is crucial to account for the nature and context of the variable when choosing an appropriate aggregation method. Furthermore, updates of reference periods in real series for national accounts and price indices, methodological revisions as well as discontinued series entail

* At this stage, I would like to thank Univ.-Prof. Dipl.-Ing. Dr. Martin Wagner for giving me the chance to participate in the project

‡ Special thanks go to Univ.-Ass. Sebastian Veldhuis, BSc. MSc. for the many helpful discussions and support throughout the data collection process.

[1] Benati et al., 2019

appropriate adjustments in order to be linked. Finally, a hand full of series had missing values for which no actual records could be found. To overcome this issue, alternative approaches are available for calculating the interrupted period.

The aim of the following chapters is to clarify and justify the methodologies carried out in the data collection process by providing both theoretical background in conjunction to instructions for practical implementation.

2 The Concept of Stocks and Flows in Economics

We encounter stocks and flows every day, let that be a glass of water that is filled up, the salary earned, the fuel level left in the tank of our cars and many other situations, the list goes on. Despite it being trivial to grasp in real life situations, the concept of stocks and flows in scientific domains may sometimes be not as intuitive anymore, and people struggle to differentiate between a stock and a flow. Having a solid understanding of stocks and flows is tremendously helpful when working with economic data, as different treatments depending on the nature of the variable are necessary.

The MIT researcher Jay W. Forrester describes the relationship between stocks and flows in his famous book *Industrial Dynamics* (1961) as follows:

> The levels are the accumulations within the system. They are inventories, goods in transit, bank balances, factory space, and number of employees. Levels are the present values of those variables that have resulted from the accumulated difference between inflows and outflows.[2]

In essence, stocks, levels or state variables[3] describe quantities measured at a snapshot in time[4], reflecting the amount of inflows and outflows of a quantity of a particular commodity, material or resource that have accumulated over time. Flows are responsible for the (absolute) change in stock and are thus often referred to as rates of flow, or just rates. To be more precise, *positive flows* (inflows) alter the value of a stock, whereas *negative flows* (outflows) deplete the stock.[5] Therefore, stocks and flows are inter-dependent as they imply each other.[6] One key difference to distinguish between the two is that stocks are expressed only in quantities of a specific unit, unlike flows, which are stated in quantities of a certain unit per time.

The relationship between flow and stock variables can be illustrated with the example of wealth. Wealth is a stock variable that is affected by several flow variable, such as net income, that is, the sum of income minus (consumption) expenditure. If an individual earns a salary of $\$50,000$ per year, this is an inflow of wealth which temporarily alters

[2] Forrester, 1980
[3] Sterman, 2000
[4] Martin, 1997
[5] Zuhu and Ratha, 1997
[6] Wagner, 2006

the level of wealth, whereas expenditure on groceries (immediate consumption of consumables) of $\$4,000$ per year would be an outflow, eventually depleting wealth, given that other influencing factors like, among many others, inflation and interest are not taken into consideration [7]. Over time, if individuals save parts of their income i.e., spending less money on consumption than they earned, their stock of wealth will increment. In our example, the wealth of the individual will rise by $46000 after one year. In this way, flow variables impact stock variables. The acquisition of assets such as company shares, real estate or cryptocurrencies are in this sense flows, however, the stock of wealth remains the same (not considering price fluctuations of these assets). Such flows do not raise or lower the wealth (outflows of from bank account and inflows to asset balance which cancel out); they only change its composition.

Some scientific disciplines like physics, or biology work with continuous time in certain situations. Though, in empirical economics time is discrete. This is not a debate about whether time in general is continuous or discrete, but about the fact that economic activity is not observable at any moment. Economic activity generally happens continuously, meaning that there is no time at which the economy is put on pause, however, the only feasible way to record economic activity is in a sequential manner.[8]

---

[7] In economic jargon, this would be a *ceteris paribus* statement.

[8] Wikipedia, 2023

3 A Mathematical View on Economic Variables

When looking at economic data on a simple line chart or in a spread sheet application, rarely someone ponders how these numbers have been collected in the first place. Though, when working with economic time series, it is beneficial to have a good understanding of what the data at hand is representing. In this section, I briefly want to outline the key characteristics of the economic variables I have been working with throughout the data gathering phase. I use high school math concepts to explain the variables, which will be of relevance for later topics such as temporal aggregation and splicing.

3.1 Stock and Flow Variables

Using basic calculus, the rate of change in continuous time can be determined by taking the point derivative (w.r.t. time). Knowing the function for a stock $S(t)$ which changes as a result of a flow over time, the underlying flow can be obtained as follows:

$$F(t \to t + \Delta t) = S'(t) = \lim_{\Delta t \to 0} \frac{S(t + \Delta t) - S(t)}{\Delta t}$$

A graphical representation of this relationship between a stock and the corresponding flow is depicted in Figure 1:

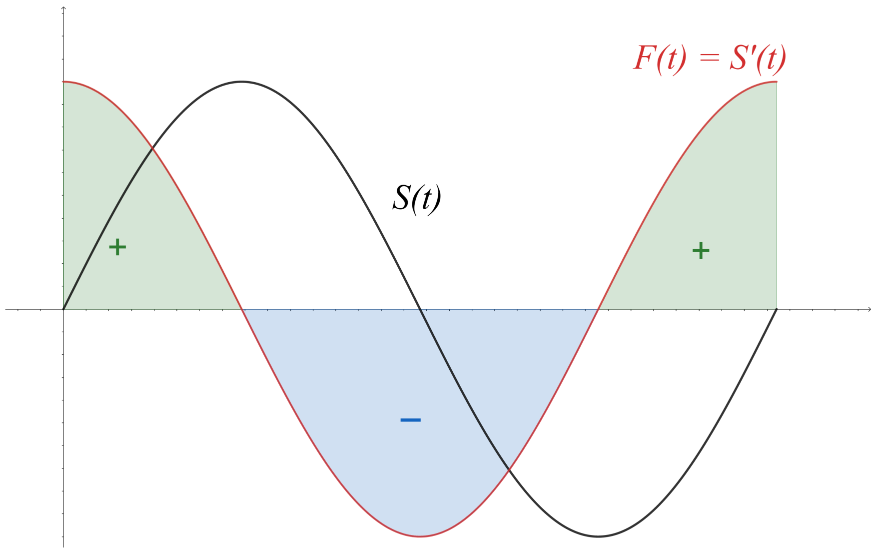

Figure 1: Relationship between stocks and slows in continuous time

As long as the flow stays positive, the stock grows, when the flow becomes negative, the stock depletes. By reversing the derivative, i.e., integrating the flow, the value of a stock at a future time point can be determined:

$$S(t+\Delta t) = S(t) + \int_{t}^{t+\Delta t} F(t)dt$$

This implies that a stock at $t + \Delta t$ is the result of an initial stock at t plus the net flow that has occurred between t and $t + \Delta t$. Rearranging the above equation gives the net flow over the interval from t to $t + \Delta t$, emerging from the absolute difference in stock, recorded at the endpoints of the interval:

$$\int_{t}^{t+\Delta t} F(t)dt = S(t+\Delta t) - S(t)$$

In economics, the time steps are discrete, so Δt is not infinitely small. Figure 2 visualizes the stock-flow relationship, whereby the stock is recorded in discrete sequences. The underlying flow can still occur continuously:

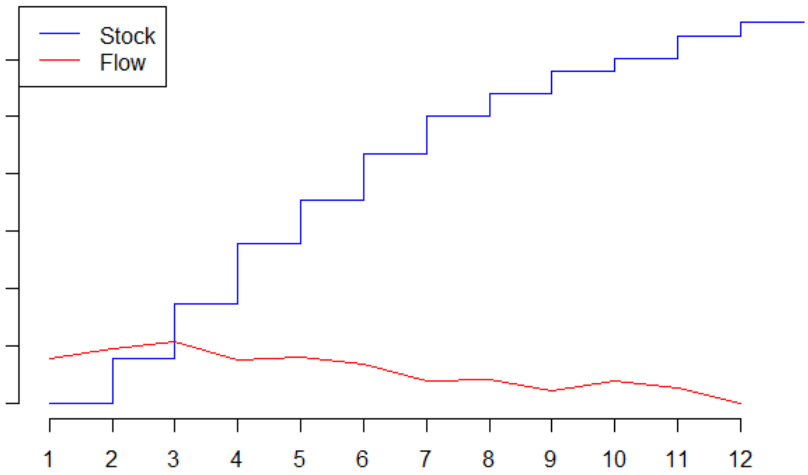

Figure 2: Stock variable recorded in discrete time

Regarding flow variables in discrete time, not the actual rate of flow (point derivative) is measured, but the total flow over an interval. Series of flow variables are for this reason conventionally labelled as integrated flows. But how can someone distinguish between a stock variable and a flow variable in economics? At this stage, we take a step back from a pure mathematical perspective and engage the context of the variable. As already mentioned, while flow variables measure the accumulated or integrated flows over a period, a stock variable measures the level at a specific point in time. The accumulation of a flow variable at the beginning of the time period subject to measurement always starts from 0, whereas the accumulation of a flow which changes a stock is added to the previous stock value recorded. With that being said, if a flow $F(t \to t + \Delta t)$ is 0 for the entire interval, then the flow variable that measures the accumulated flow $\int_t^{t+\Delta t} F(t)dt$ would be 0 too. A stock variable on the other hand, which only changes due to the accumulation of the flow, will just remain the same, so $S(t + \Delta t) = S(t)$, but it will still have a value.

For example, the GDP is a flow variable because it is a representation for the value of finalized goods and services produced by the economy during a specified period, usually one year. Accordingly, the flows from production are initialized at 0 at the beginning of each new period; hence, flow variables only develop a value over time. The same principle applies to national accounts in general. Income, consumption expenditure or investment are all economic flow variables. Monetary aggregates, on the other hand, are stock variables as they express the currency or other financial assets circulating in the economy at the moment of recording. The stock changes due to currency issued or destroyed by the central bank, and other transactions of monetary liabilities within the economy governed by money supply and demand.[9]

Another characteristic to point out is that integrated flows of national accounts are solely measured on a positive scale (in the first quadrant) and are weakly monotonic increasing, implying that $\int_0^t F(t)dt \leq \int_0^{t+\Delta t} F(t)dt$ inside the interval being measured. The reason for this is that no outflows can occur such as negative investment or negative consumption, thus, the lowest rate of flow is 0. For monetary aggregates, outflows can certainly happen, leading to a depletion in stock. Still, the lowest possible value economic stock variables can take on is 0 as there is no such thing as, for instance, negative currency in circulation.

3.2 Ratios and Relative Rates

Ratios are simple fractions of two quantities sharing the same unit.[10] Ratios do not express a change over time, and therefore do not posses a time dimensions. Popular occurrences of ratios in economics are elasticities, exchange rates or the unemployment rate. The latter two can cause confusion due to their terminologies. In the case of the unemployment rate, despite being called a rate, it is in this sense a ratio because it represents the number of the unemployed in relation to the labor force, two stationary observations in time, and not the rate at which the unemployment level changes with time.

[9] European Central Bank, 2023
[10] Young, 2023

Relative rates on the other hand are fractions of two quantities of different units, most frequently in the form of quantities or units per time.[11] Relative rates with a time component may also be called *temporal rates* of change. They express, how a variable changes over time in relation to the initial value, which gives them a time dimension. The word *per*, again, is a decent indicator when distinguishing between ratios and relative rates. Examples of economic variables which measure relative changes of quantities would be any kind of interest rates or growth rates. Although relative rates are a form of describing a flow over time, I want to point out that relative rates must not be confused with flow variables. This is important when it comes to temporal aggregation, as flow variables and relative rates are treated differently, as explained in 4.4. Data both for ratios and relative rates are conventionally expressed in percentages.

3.3 Index Numbers

Index numbers are statistical measures which were originally introduced for summarizing changes in a set of related variables over time. US economist Irving Fisher (1922) argues that combining these variables into a single index number allows for a better grasp of the overall trend in the economy. Without index numbers, measuring changes in the wages, prices or production would be a tedious task. In terms of measuring the change in price level over time, someone might have to look at the prices of individual items, as for some items, the price might rise by a high percentage, other items might get only slightly more expensive or even cheaper. Introducing a price index like the CPI makes it much easier, as it summarizes the *weighted average of prices* and sets it into relation of a defined benchmark period, the *base year*.[12] Popular index-types are the *Paasche* index, using the current year's quantities as weights, the *Laspeyres* index, using the base year's quantities as weights, and *Fisher* index, being a combination of Paasche and Laspeyres.[13]

[11] Wang, 2018
[12] Fisher, 1922
[13] Diewert, 2021

4 Temporal Aggregation

The task for this bachelor's thesis was to gather annual data sets of macroeconomic variables from several countries and Central Banks. Sometimes, no annual but only higher frequency time series were provided. In this sense, the question came up, how to properly convert from high frequency time series (monthly, quarterly) to annual values. In this section, I conceptually deal with the temporal aggregation of time series data for different economic variables.

First, to define the term temporal aggregation, it is the technique of summarizing or grouping time series data over a time period to create a coarser representation of the data.[14] A plethora of aggregation functions and methods, among which summation, averaging, minimum or maximum and first or last are most frequently implemented. Despite the broad selection of aggregation functions or concepts, it needs to be carefully considered what aggregation methods are appropriate for which type of variable. In this section and upcoming subsections, I use the constants H to denote the number of sub-annual observations of the high frequency time series, and L to denote the number of sub-annual observations of the low frequency time series, so L suggests the number of observations persisting after the aggregation process. The constant K is attained from the division of H by L. Note that H, L and K are integers.

Also, I want to impose two restrictions:

1. $H \wedge L \wedge K \in \mathbb{N}$
2. it holds that $L \leq \frac{H}{2}$

The restrictions imply that H is a multiple of L and that the floor division of H and L leading to K yields no remainder. This is to exclude possible edge cases in which the instructions would not work, for example, if observations H are of irregular frequency. Those three constants will be used in the upcoming notations, where I aim to introduce universally applying methods to temporally aggregate variables of distinct kinds. The left-hand side thereby represents the resulting low frequency series, the right-hand side the aggregation of the high frequency series.

[14] Gamper, Böhlen, and Jensen, 2009

4.1 Stock Variables

4.1.1 Point-in-Time Sampling

The most straight forward approach to temporally aggregate higher frequency time series of a stock variable to low frequency is point-in-time sampling[15]. Point-in-time sampling in this regard is appropriate because stock variables are measured at a specific moment in time representing a certain quantity or level at that moment. It is a good practice to consistently select the same time point(s) for each year, commonly either the first or last observation of the high frequency time series. The vast majority of public financial authorities, such as the European Central Bank (ECB), publish data for stock measures from the last day of the reference period.[16] For annual time series, this would be the stock value of the last calendar day, i.e., December 31.

If the original time series is given with H number of observations in year t, to temporally aggregate to L number of observations, every Kth element is selected, whereby $K = \frac{H}{L}$.

$$\{x_{t,l}\}_{\substack{t=1,2,\ldots,T \\ l=1,2,\ldots,L}} = \{x_{t,Kl}\}_{\substack{t=1,2,\ldots,T \\ l=1,2,\ldots,L}}$$

To demonstrate the above formulation in an exemplary way, I use the monetary aggregate M1 which is given in monthly frequency for the years 1980 to 2022, so $H = 12$ since a year has 12 months. To aggregate to quarterly values, so $L = 4$ since a year consists of four quarters, with $K = \frac{12}{4} = 3$, we can write:

$$\{\mathrm{M1}_{t,l}\}_{\substack{t=1980,1981,\ldots,2022 \\ l=1,2,\ldots,4}} = \{\mathrm{M1}_{t,3*l}\}_{\substack{t=1980,1981,\ldots,2022 \\ l=1,2,\ldots,4}}$$

Thereby, the values of March, June, September, and December are selected to establish a low frequency series consisting of Q1, Q2, Q3 and Q4.

[15] Marcellino, 1999

[16] European Central Bank, 2023

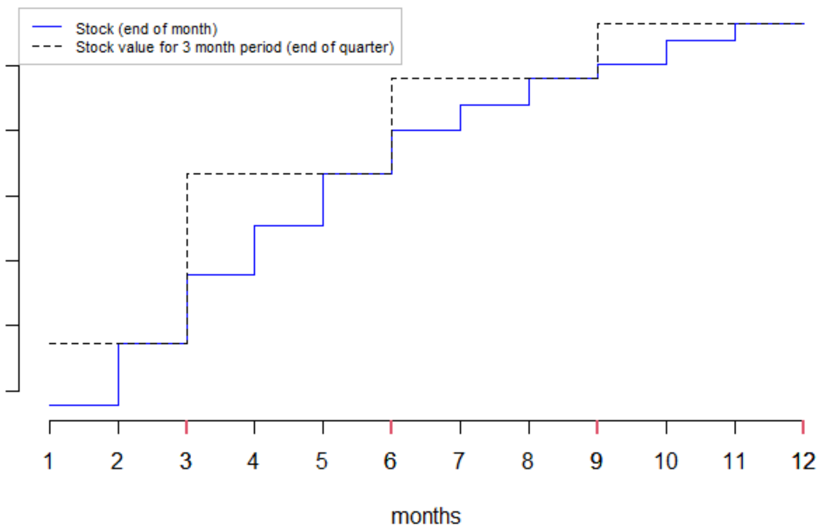

Figure 3: Temporal aggregation from monthly to quarterly frequency

The process to temporally aggregate the M1 from monthly frequency to a single annual observation, set $H = 12$ and $L = 1$, implying that $K = 12$:

$$\{\mathrm{M1}_t\}_{t=1980,1981,\ldots,2022} = \{\mathrm{M1}_{t,12}\}_{t=1980,1981,\ldots,2022}$$

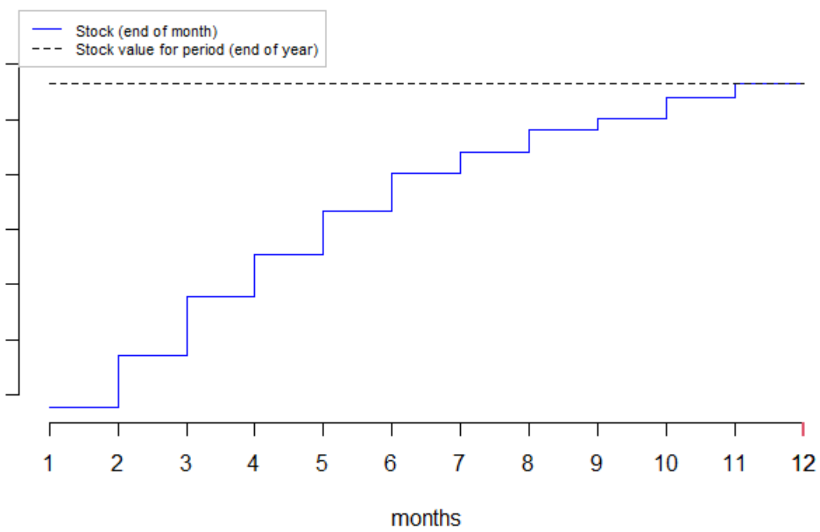

Figure 4: Temporal aggregation from monthly to annual frequency

4.1.2 Averaging

Besides the end-of-period approach, another conventional way for publishing series of stock variables is in the form of period averages. Such series for monetary aggregates are then labelled as average amounts outstanding. For example, the M1 may be published in quarterly frequency in units of quarterly average stock. Temporal aggregation of a high frequency series of a stock variable by computing the arithmetic mean of K consecutive stock observations within a year t can be achieved as follows:

$$\{x_{t,l}\}_{\substack{t=1,2,\ldots,T\\ l=1,2,\ldots,L}} = \left\{\frac{1}{K}\sum_{k=1}^{K} x_{t,k+(l-1)*K}\right\}_{\substack{t=1,2,\ldots,T\\ l=1,2,\ldots,L}}$$

Considering the same scenario of a monthly (high frequency) time series of the M1 with $H = 12$ as proposed before in section 4.1.1, to convert to quarterly M1, so $L = 4$ and $K = 3$, the average of the three months can be applied for temporal aggregation of the M1:

$$\{\mathrm{M1}_{t,l}\}_{\substack{t=1980,1981,\ldots,2022\\ l=1,2,\ldots,4}} = \left\{\frac{1}{3}\sum_{k=1}^{3} \mathrm{M1}_{t,k+(l-1)*3}\right\}_{\substack{t=1980,1981,\ldots,2022\\ l=1,2,\ldots,4}}$$

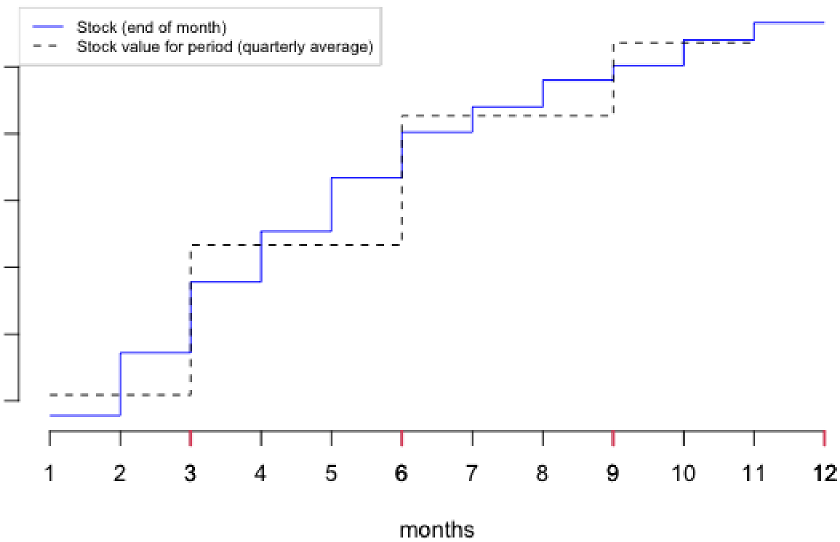

Figure 5: Temporal aggregation from monthly to quarterly frequency

Similarly, converting from monthly to annual values with $H = 12$, $L = 1$ and $K = 12$ we can write:

$$\{\mathrm{M1}_{t,l}\}_{t=1980,1981,\ldots,2022} = \left\{\frac{1}{12}\sum_{k=1}^{12} \mathrm{M1}_{t,k}\right\}_{t=1980,1981,\ldots,2022}$$

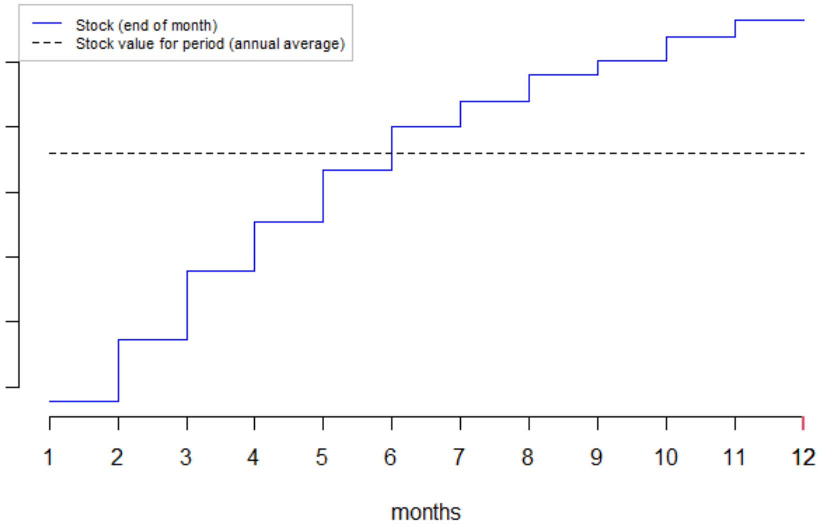

Figure 6: Temporal aggregation from monthly to annual frequency

The resulting low frequency time-series in the case of using the arithmetic mean method is then interpreted as the average stock for the aggregated period.

4.2 Flow Variables

4.2.1 Summation

As discussed in section 3, flow variables in economics such as national accounts measure the integrated flows of economic activity; in terms of the GDP, the total value of goods and services produced in an economy over a period. The properties of integrals allow for adding up integrated flows of sub-periods to cover a broader interval. Figure 7 below depicts the growing integrated flow over a year as a result of summing up the four quarterly integrated flows (note that the horizontal axis is given in months). The annual aggregate of quarterly integrated flows (that is, $\int_0^{12} F(t)dt$) is indicated by dashed borders:

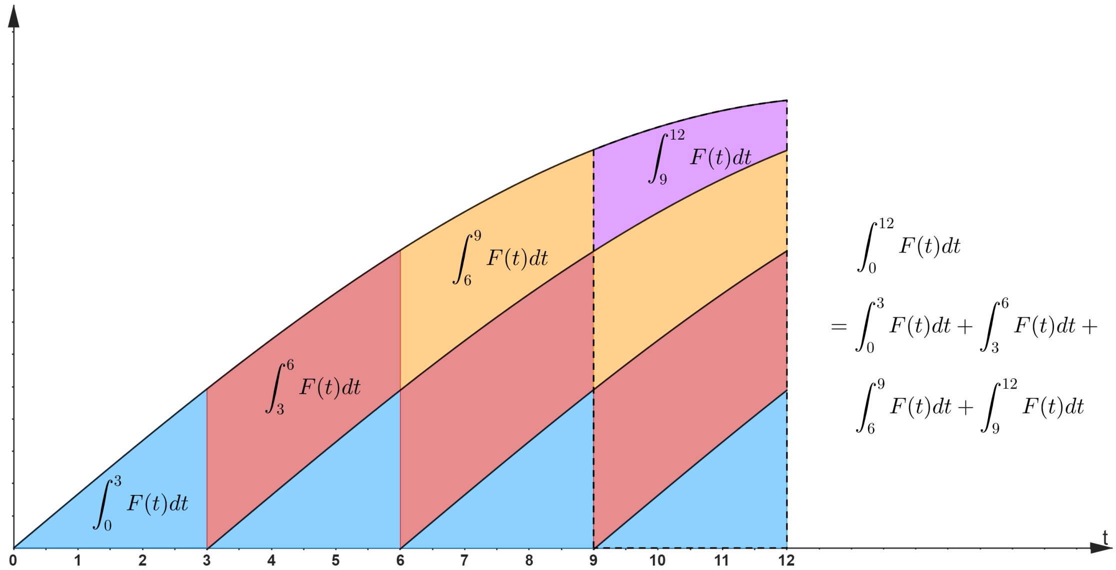

Figure 7: Aggregation of integrated flows

Considering a flow variable given in high frequency with H number of observations for year t, to aggregate to low frequency L in year t, $K = \frac{H}{L}$ consecutive observations from the high frequency time series are summed up:

$$\{x_{t,l}\}_{\substack{t=1,2,\ldots,T \\ l=1,2,\ldots,L}} = \left\{ \sum_{k=1}^{K} x_{t,k+(l-1)*K} \right\}_{\substack{t=1,2,\ldots,T \\ l=1,2,\ldots,L}}$$

For instance, a monthly time series of the real GDP denoted as Y with $H = 12$ (12 months) can be temporally aggregated to quarterly frequency $L = 4$ (four quarters in a year) by summing up $K = \frac{H}{L} = 3$ consecutive months:

$$\{Y_{t,l}\}_{\substack{t=1980,1981,\ldots,2022 \\ l=1,2,\ldots,4}} = \left\{ \sum_{k=1}^{3} Y_{t,k+(l-1)*3} \right\}_{\substack{t=1980,1981,\ldots,2022 \\ l=1,2,\ldots,4}}$$

Even more effortless is the temporal aggregation process from monthly to annual frequency, which is to add up the values of all 12 months:

$$\{Y_t\}_{t=1980,1981,\ldots,2022} = \left\{ \sum_{k=1}^{12} Y_{t,k} \right\}_{t=1980,1981,\ldots,2022}$$

4.2.2 Averaging

Statistical agencies of some countries follow the approach of publishing high frequency time series of national accounts at annualized levels by multiplying each sub-annual observation by the number of sub-annual observations H. Consequently, monthly or quarterly values are artificially altered by factors of 12 or 4 respectively, potentially causing confusion as the total flows cannot be summed up anymore to obtain true annual aggregates.[17] For instance, the Bureau of Economic Analysis (BEA) annualizes their national accounts data with the argument of better comparability across series of differing periodicity.[18] A simple way to deal with this issue is to take the arithmetic mean of sub-annual observations such that the factor H cancels out.

Applying the arithmetic mean on a series of sub-annual observations $Hx_1, Hx_2, \ldots, Hx_H$ of a national account variable, whereby each observation has been multiplied by H to mimic annualized levels, cancels out the factor:

$$\frac{Hx_1 + Hx_2 + \cdots + Hx_H}{H} = x_1 + x_2 + \cdots + x_H$$

Recall that annual values are obtained from a high frequency time series by setting $L = 1$, implying that $K = H$, a generalized formula can be denoted as:

$$\{x_{t,l}\}_{\substack{t=1,2,\ldots,T \\ l=1,2,\ldots,L}} = \left\{ \frac{1}{K} \sum_{k=1}^{K} K x_{t,k+(l-1)*K} \right\}_{\substack{t=1,2,\ldots,T \\ l=1,2,\ldots,L}}$$

4.3 Ratios

Since ratios do not suggest a change over time, summation is not a meaningful aggregation method. Point-in-time sampling or taking the arithmetic average (same as for stock variables, which is described in section 4.1) would be appropriate to apply. Still, it is advised to appraise the context of the variable. Looking at a monthly series for the unemployment rate, selecting the June or December value for representing an entire year can be misleading, as a high seasonal volatility (standard deviation) is involved for most countries. Still, if done so, then the observation of the same month should be consistently chosen for representing the respective year. Rather applying the arithmetic mean would be a better measure to summarize the labor market situation over a period of one year. Although not explicitly stated, I empirically investigated that annual data for the unemployment rate from most publishers of labor statistics, such as Eurostat or the ECB, is identical to the arithmetic average of monthly observations.

[17] International Monetary Fund, 2017
[18] Fox, 2022

4.4 Relative Rates

4.4.1 Compounding

Unlike the aggregation process for flow variables in section 4.2 (which express absolute changes), relative rates of change are aggregated by multiplication instead of summation. The notation is slightly tweaked by replacing the partial summation by the product notation to achieve a compounding effect.

Considering a high frequency time series of rates (in decimal format), again, with H numbers of observations for year t, it can be temporally aggregated by sequential multiplication of $K = \frac{H}{L}$ consecutive observations to obtain a low frequency time series with L number of observations in year t:

$$\{r_{t,l}\}_{\substack{t=1,2,\dots,T\\ l=1,2,\dots,L}} = \left\{ \left[\prod_{k=1}^{K} 1 + r_{t,k+(l-1)*K} \right] - 1 \right\}_{\substack{t=1,2,\dots,T\\ l=1,2,\dots,L}}$$

A rate that stays constant throughout all H observations in t years ($r_{t,h} = \bar{r}$) allows for simpler notation:

$$\{\bar{r}_{t,l}\}_{\substack{t=1,2,\dots,T\\ l=1,2,\dots,L}} = \left\{ (1+\bar{r})^K - 1 \right\}_{t=1,2,\dots,T}$$

Assuming the interest rate for a financial asset is given in monthly frequency also on a monthly basis (indicating that it compounds for longer holding periods than one month), to convert to quarterly frequency we can therefore set $H = 12$, $L = 4$ and $K = \frac{12}{4} = 3$:

$$\{r_{t,l}\}_{\substack{t=1,2,\dots,T\\ l=1,2,\dots,4}} = \left\{ \left[\prod_{k=1}^{3} 1 + r_{t,k+(l-1)*K} \right] - 1 \right\}_{\substack{t=1,2,\dots,T\\ l=1,2,\dots,4}}$$

For aggregating from monthly to annual figures, set $H = 12$, $L = 1$ and $K = \frac{12}{1} = 12$:

$$\{r_{t,l}\}_{t=1,2,\dots,T} = \left\{ \left[\prod_{k=1}^{12} 1 + r_{t,k} \right] - 1 \right\}_{t=1,2,\dots,T}$$

From quarters to annual frequency, set $H = 4$ and $L = 1$, so $K = 4$:

$$\{r_{t}\}_{t=1,2,\dots,T} = \left\{ \left[\prod_{k=1}^{4} 1 + r_{t,k} \right] - 1 \right\}_{t=1,2,\dots,T}$$

4.4.2 Geometric Mean

Especially interest rates on financial assets such as long-term government bond yields or 3-month treasury bill rates are typically expressed as percentage per annum (p.a.), regardless of their maturity. A compounding hereby is not applicable as the interest rates are already provided on an annual basis. Nevertheless, time series for interest rates may be reported in quarterly, monthly, sometimes even daily frequency. The go-to method to temporally aggregate the interest rates would be to compute the average. Not the arithmetic mean, but the geometric mean would be the mathematically correct method for relative rates[19] because it takes into account the compounding effect of relative rates over time. The temporal aggregation process using the geometric mean would be to multiply K consecutive sub-annual observations and taking the Kth root of the product, and finally, subtract 1 to obtain the average rate in a decimal representation:

$$\{r_{t,l}\}_{\substack{t=1,2,\dots,T\\ l=1,2,\dots,L}} = \left\{ \left[\prod_{k=1}^{K} 1 + r_{t,k(l-1)*K} \right]^{\frac{1}{K}} - 1 \right\}_{\substack{t=1,2,\dots,T\\ l=1,2,\dots,L}}$$

4.4.3 Arithmetic Mean

In many incidents, the arithmetic mean is applied to temporally aggregate interest rates, despite the geometric mean being the mathematically more suitable approach. Reformulating the geometric mean of interest r_t by taking logs (natural logarithm) shows that the smaller the interest rates (though, $r_t \geq 0$), the better the arithmetic mean approximates the geometric mean. Considering the product rule of the natural logarithm which states that $\ln(xy) = \ln(x) + \ln(y)$, it can be shown that:

$$\ln\left(\left[\prod_{t=1}^{T} 1 + r_t \right]^{\frac{1}{T}} \right) = \frac{1}{T} \sum_{t=1}^{T} \ln(1 + r_t),$$

and for small, non-negative interest rates $\ln(1 + r_t) \approx r_t$, so the geometric mean can be approximated by the arithmetic mean:

$$\frac{1}{T} \sum_{t=1}^{T} \ln(1 + r_t) + 1 \approx \frac{1}{T} \sum_{t=1}^{T} 1 + r_t$$

Because of their properties it can be shown that arithmetic mean $\geq$ geometric mean. If interest rates are rather high (above 10%), the arithmetic mean notably overestimates the interest rates.

[19] Wild, 2021

4.5 Indices

The go-to method to temporally aggregate high frequency indices to low frequency is taking the arithmetic mean for the sake of simplicity. Both Eurostat and Office for National Statistics (ONS) use a chain-linked Laspeyres-type index for establishing the Harmonized Consumer Price Index (HCPI). Disaggregate index series in monthly frequency are constructed in a way that the annual arithmetic mean equals 100 in the base year.[20,21] Also, annual figures of the CPI for the US published by the Bureau of Labor Statistics (BLS) are identical to the arithmetic mean of monthly CPI observations. Note that the aggregated value for the base year might not always amount to 100. This may happen if not an entire year has been chosen as a base period, but an observation of a higher frequency series, which is the case for Switzerland's CPI, whereby July 2015 has been chose and the base period.

[20] Office for National Statistics (ONS), 2019

[21] Eurostat, n.d.

5 Splicing and Linking

There is a strong demand among econometricians, and economic analysts for data covering prolonged time spans while ensuring consistency. It is quite rare to find time series that meet these criteria. In case, multiple shorter time series covering different periods are available, they can be combined to cover the desired long horizon. Assuming an ideal situation in which the series to be linked have both been collected under the same definition and methodology, they can be concatenated (end-to-beginning) without further modifications. An expired time series covering a period from 0 to $T + j$, and the new series of the same variable ranging from T to $T + n$ ($j < n$, and if $j > 0$, then there is an overlap between the two series, for which the values are identical), then the two series can simply be concatenated at any point within the overlapping interval between T and $T + j$.[22] However, economic data is regularly subject to 1) methodological improvements, 2) change of conceptual definition and composition, and 3) corrections of errors, ultimately causing breaks or inconsistencies.[23] The related challenge arising from this is that time series cannot be simply joined, but must be transformed somehow in order to be linked. Such a procedure is called splicing. Quoting Irving's (2021) words:

> Splicing in this context refers to the combining or joining of more than one method to form a complete time series. Several splicing techniques are available if it is not possible to use the same method or data source in all years . . . [This involves] techniques that can be used to combine methods to minimi[z]e the potential inconsistencies in the time series. Each technique can be appropriate in certain situations, as determined by considerations such as data availability and the nature of the methodological modification.[24]

Conducting a full revision is arguably the most accurate approach, but simultaneously the most impractical one as well. In a nutshell, the costs in terms of effort and time for conducting such a sophisticated reconstruction adhering to the present principles established by statistical offices is for most independent researchers not feasible. Although not as accurate, simple linking and splicing techniques such as rebasing, retropolation and

[22] Organisation for Economic Co-operation and Development (OECD), 2023
[23] Di Fonzo, 2003
[24] Irving, Nakane, and Villarin, 2021

interpolation, may be used, just to name a few.[25,26]

For the upcoming sections, I propose two conditions which must be fulfilled:

1. time series to be joined are in annual frequency
2. the old series X_t^{old} reaches $T + j$, and the new series X_t^{new} starts at T and reaches $T + n$, with $0 \leq j < n$, so there is a minimum overlap of one observation in T

If condition 2 is violated, say, X_t^{old} only reaches $T - 3$, imputation techniques can be applied to approximate missing observations, as covered later in section 6. Whenever splicing techniques are brought into play, it is usually recommended to adjust the older discontinued series to match the new series for the reason that it is assumed that the new series has been constructed with the latest methodology, ultimately making the data more accurate. For monetary aggregates or national accounts, the old series will be adjusted to the level of the most recent series, while for indices or other series that involve some kind of reference period, the reference period of the latest available series will be adopted.

5.1 Rebasing of Index Numbers

There is a high likelihood when working with time series of index numbers to face different base years. Either as good practice, or because of regulation, the base year should regularly be updated in a specified time window, typically every five years. This is due to practical reasons, namely, to improve international comparability and achieve a more accurate reflection of the economy of the present time.[27] To construct a longer, cohesive series of an index by linking an old series to the new series, whereby their base years differ, it is a good practice to rebase the old index to match the new index' base year, for the reasons stated previously. To rebase an old index, say with base year 2015, to the new base year 2020, the entire index series is divided by the index value of the desired new base period, that is, the 2020 value. Thereby, the growth rates (the relative change from one year to a subsequent year) are preserved.[28]

Let i_t^{b1} be the old time series of an index with $b1$ as the base year, and i_t^{b2} the time series of the new index with $b2$ as base year, then the old index can be rebased by dividing the

[25] Prados de la Escosura, 2017
[26] Tzavidis,
[27] Eurostat, 2023
[28] Fox, 2022

whole index by the index number in $b2$, which is held constant:

$$\{i_t^{b2}\}_{t=1,2,\ldots,T} \simeq \left\{\frac{i_t^{b1}}{i_{b2}^{b1}} * 100\right\}_{t=1,2,\ldots,T}$$

Once the old index is rebased to the new base year, and there are no major deviations between the old and the new series, both time series can simply be concatenated. Alongside updating the base year, some statistical institutions make use of this opportunity to introduce revisions to the index series.[29] A proponent of this practice would be the BLS revising the CPI index frequently, whereas Statistics Canada refrains from doing so. One critical part here to point out is that while rebasing does not impact growth rates, the revisions do. As a consequence, some discrepancies towards more recent years might be observable (hence, the $\simeq$ sign).

5.1.1 Chain-Type Index

Real figures of national accounts are computed by using an index which is used for adjusting for inflation. Such indices are derived from aggregates of specific price and quantity combinations for goods and services. As the name suggests, the fixed base year method uses the price level of the fixed base year and thereby establishes fixed weights, which soon become outdated over time as the economy advances. By way of illustration, a series of real GDP with a fixed reference period of 1990 does not capture the gain in relative importance of computer and software products over the following decades. The majority of developed countries have gradually introduced chain linking during the last decade of the 20th century to address this issue. These chained volume statistics are computed by using the price and weight structures of the preceding year.[30] So, the base year is updated annually, and the values obtained from year to year are linked. The reference period of chained volume series thereby expresses that the underlying chain-type index is equal to 100 in this year, where the value of the real value is equal to the nominal value. The chain-linked series of a national account of the US with reference period t is then stated as *Chained* t US dollars.[31] A key advantage of chaining is that it preserves the year-on-year growth rates, regardless of the reference period.[32,33]

[29] Eurostat, 2023
[30] Economic and Social Commission for Asia and the Pacific (ESCAP), 2019
[31] Fox, 2022
[32] Dippelsman, Josyula, and Métreau, 2016
[33] Eurostat, 2022

Figure 8 displays how updating the base year does not fundamentally influence the growth rates of the real GDP series. The small deviations in (B) can be explained by revisions imposed by the BEA, as they updated the reference period from 2009 to 2012.

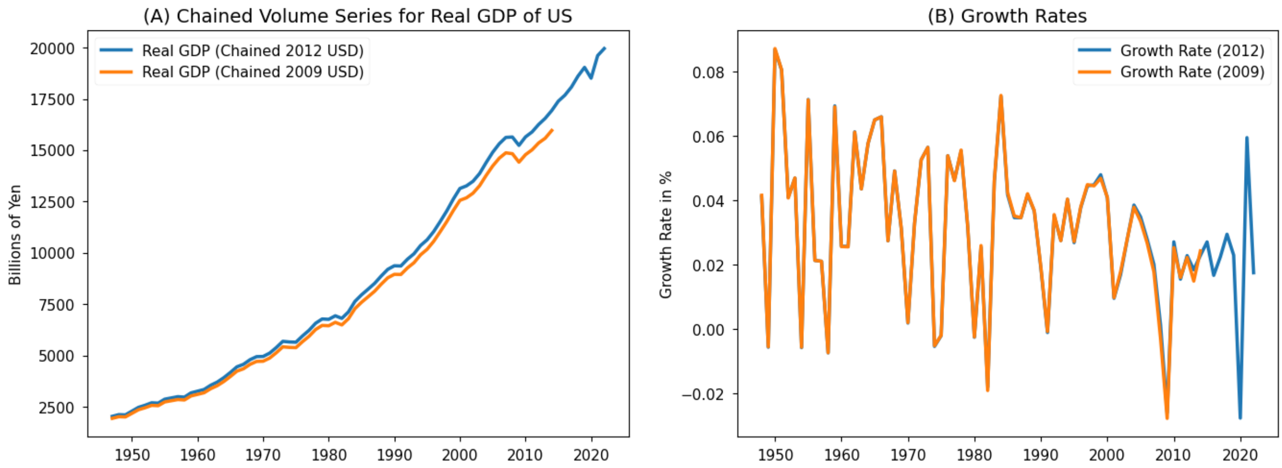

Figure 8: Chained volume real GDP with different reference years

5.2 Retropolation with Growth Rates

The prevalent technique for linking time series in scholarly works is a form of backward extrapolation of the most recently published (new) time series by using the growth rates of preceding (old) time series. No fixed agreed-upon terminology can be found in the literature for this technique, henceforth, I will adopt the proposed expression *retropolation* as suggested by, among others, de la Fuente (2009). By doing so, the older series is adjusted in a way to smoothly transition into the new series at its starting point. Thereby, the levels of the old series are modified while the growth rates are preserved.[34] This method comes with the major advantage of being straightforward to execute. Furthermore, retropolation can be applied for splicing both stock and flow variables since growth rates measure the relative change in value from one observation to its timely successor. Even if two series are scaled differently or have different reference periods, the growth rates should reflect a similar trend in most cases. Also, the broader the overlap span of the series, the better.

The simple growth rate in year t of a time series can be denoted as follows:

$$r_t = \frac{x_t - x_{t-1}}{x_{t-1}} = \frac{x_t}{x_{t-1}} - 1$$

[34] de la Fuente, 2009

Alternatively, the simple growth factor (defined as the factor by which x_{t-1} must be multiplied to obtain x_t)[35] can be used to create a shortcut, as the deduction of 1 and later adding 1 again is redundant:

$$g_t = \frac{x_t}{x_{t-1}} = 1 + r_t$$

Given an older time series X_t^{old}, available for the years $t = 0, 1, 2, \ldots, T + j$, and the most recent time series X_t^{new} for the years $t = T, T + 1, T + 2, \ldots, T + n$, and given the growth rates of both series show a high similarity, then the new series can be retropolated by the growth rate of the old time series r_t^{old} (in decimal format) for $t = 1, 2, \ldots, T$ from the linking point T back to 0. Initially, we start with the first recorded observation of the new series X_T^{new}, and divide it by the growth factor of the old series at time T. The result will be the input for the next operation to achieve a recursive retropolation performed back to $t = 0$. I additionally introduce $m = 0, 1, 2, \ldots, T - 1$, and as m increases, a further backwards calculation can be achieved:

$$\left\{\hat{X}_{T-m-1}\right\} = \left\{\frac{X}{1 + r_{T-m}^{old}}\right\}; \text{ with } X = \begin{cases} X_T^{new}, & \text{if } m = 0 \\ \hat{X}_{T-m}, & \text{otherwise} \end{cases}$$

The resulting retropolated series $\hat{X}_t$ is then linked to the new series X_t^n.

As a splendid example to illustrate the retropolation method in practice, I showcase the application on the M3 of Japan. Consider Figure 9, a recent series (MD02'MAM1NEM3M3MO from 2003 to 2022) accompanied by five discontinued series (MD02'MAMS1EN10 from 1971 to 1999, MD02'MAMS1EN10NE and MD02'MAMS1ENM3 from 1996 to 1999, MD02'MAMS3EN10NE and MD02'MAMS3ENM3 from 1998 to 2008) are provided by the Bank of Japan (BOJ) (A), all referring to end of period stock (that is, the December value of the corresponding year). For all series, the growth rates (indicated by g_ before the series code) have been computed before analyzing their potential for linking (B). In (C), the growth rates of selected series have been concatenated (g_MD02'MAMS1EN10 from 1972 to 1998, g_MD02'MAMS3ENM3 from 1999 to 2003, and g_MD02'MAM1NEM3M3MO from 2004 to 2022), which can be justified as they together form a cohesive trend. Finally, the merged series of growth rate has been used to retropolate the most recent series back to 1971 (D). It does not matter whether the retropolation is initiated from the first observation of the most recent series in 2003 using the growth rates of the discontinued

[35] Hüpen, 2002

series up to 2003, or retropolating from the 2022 value with the entire linked-together series of growth rates - the values of the most series will be the same.

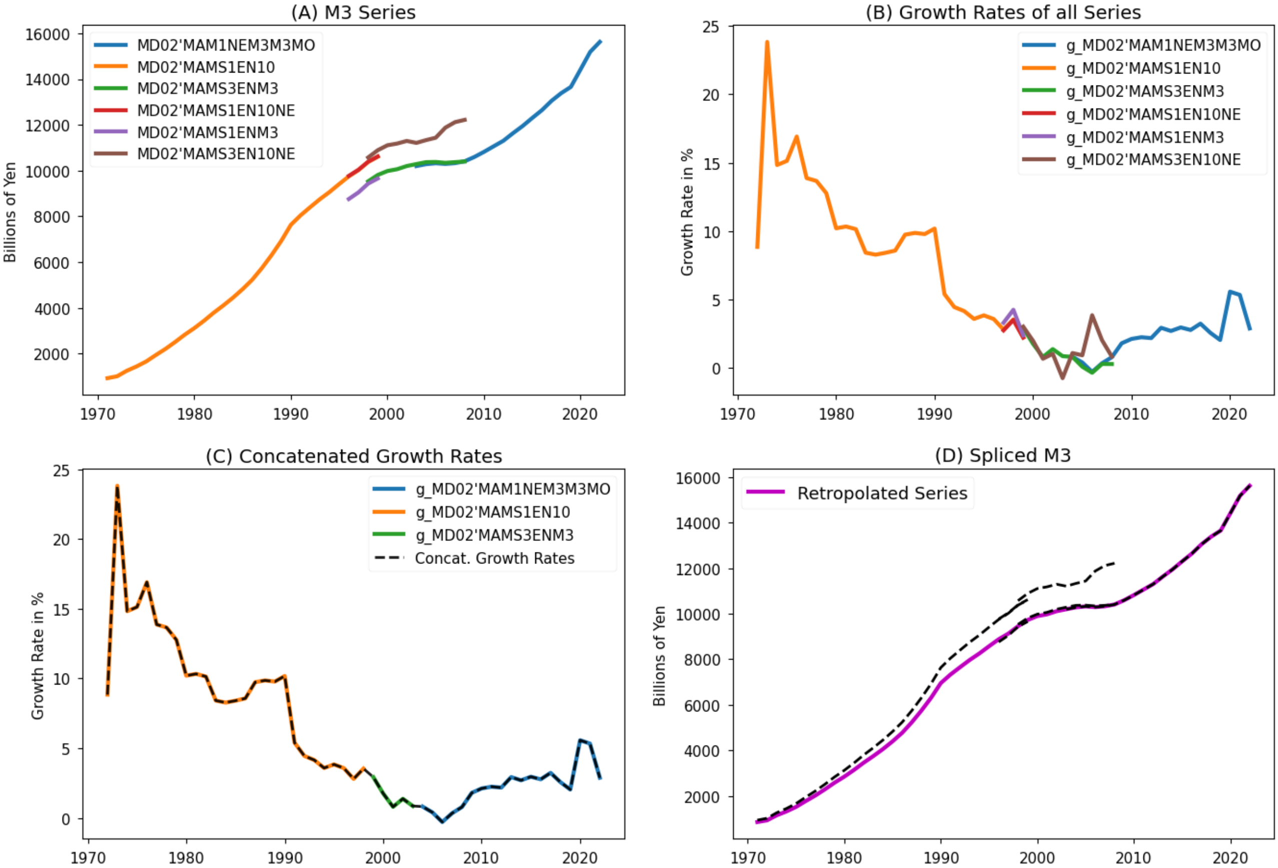

Figure 9: Retropolation procedure applied to M3 of Japan

Again, a crucial requirement for this technique to be appropriate to apply is that the growth rates of the two series to be linked show at least similarity to some extent, ideally, they are identical. If significant deviations from one another in overlapping periods occur, retropolation should not be used.[36] Due to this limitation, the series for real final consumption expenditure of households, released by BOJ, could not be spliced by retropolation, in spite of chain-linking, which supposedly preserves the year-on-year growth rates. The discrepancies in the series (A) as well as growth rates (B) are depicted in Figure 10.

[36] Irving, Nakane, and Villarin, 2021

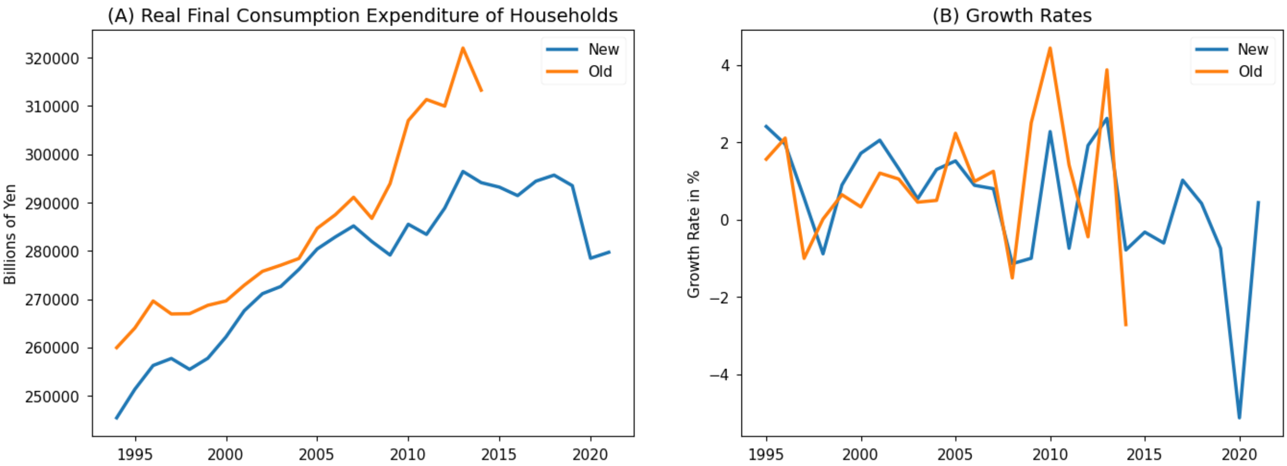

Figure 10: Inconsistencies in series for real final consumption expenditure of Japan

The new series for real final consumption expenditure of households has been added to the data collection. The end users themselves may decide whether to perform splicing or not.

6 Imputation of Missing Data

Directly related to linking or splicing time series is the question, how to cope with missing values. A statistical procedure named imputation can be deployed to fill out gaps in the data. A broad selection of univariate and multivariate imputation methods is available. Univariate methods purely rely on a single series, typically available data point of the same series subject to imputation, whereas multivariate methods also consider related variables to derive approximations for the missing data. Univariate techniques are definitely more heavily used, as the implementability of multivariate approaches depends on the availability of related data, which is seldomly the case.
Simple univariate imputation involve mean, median, mode, last or random imputation, advanced methods are spline interpolation, time series models such as Auto-Regressive Integrated Moving Average (ARIMA), deep learning models like feed-forward Artificial Neural Networks (ANN) and Long-Short Term Memory (LSTM), and even clustering algorithms such as K-Means or K-Nearest Neighbours.[37] In order to test for accuracy, simulation trials are conducted by artificially cutting out chunks of the series and then comparing the imputed values to the true, hidden values. The usual statistical error metrics for time series prediction tasks are R^2, Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Data imputation is oftentimes viewed from a critical standpoint for assorted reasons that I will address later. A general recommendation is to impute gaps only if they are small, since for longer intervals, the accuracy will decline severely.[38,39]

6.1 Linear Weighted Moving Average

Several versions of moving average algorithms exist, among the most popular ones, the Simple Moving Average (SMA), the Linear Weighted Moving Average (LWMA), and the Exponential Weighted Moving Average (EWMA). While SMA has no weighting involved (each input is equally important), LWMA assigns arithmetically linearly and EWMA exponentially decreasing weights the further the distance between the predictor and the target observations.[40] The distributed weights thereby always add up to 1. These moving

[37] Flores, Tito, and Centty, 2019
[38] Wijesekara and Liyanage, 2020
[39] Demirhan and Renwick, 2018
[40] Demirhan and Renwick, 2018

average algorithms are typically used in finance and technical trading applications, to smooth data by mitigating short run volatility. Nevertheless, researchers identified them as suitable imputation models. In contrast to sophisticated machine learning models, these moving average algorithms are fairly comprehensible and interpretable. For filling out missing observations in the series, I deployed the LWMA algorithm. A key advantage of LWMA over SMA is that as the window size (m values before plus n after the target observation) increases, the impact of observations further away is reduced. In contrast, the closer predictor observations are to the target point, the higher the effect on the output, making LWMA more responsive to recent fluctuations than SMA.
According to the literature, setting $m = n$ seems to be best practice. The total window size should also be appropriately accustomed to the size of the gap k. Applicable open-sourced software packages only allow to specify either the total window size or the number of observations considered in each direction, because they utilize vector operations. Though, the LWMA algorithm I propose has the ability to pass in asymmetric parameter arguments for m and n. A vectorized version would run undoubtedly more efficiently in terms of computing power, but at the expense of flexibility. Nevertheless, computing efficiency in the context of time series imputation is not a factor to worry about. Also, instead of assigning weights of $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}...$, weights are assigned as integers in decreasing order, and then divided by their sum, so that they add up to 1. For establishing a mathematical representation of the algorithm, missing values (marked as NA or NaN, standing for *Not Available* and *Not a Number*, respectively) have been swapped out by zeros.
Given that x_t is missing,

$$x_t = \frac{\sum_{i=1}^{m}(m-i+1) \times x_{t-i} + \sum_{j=1}^{n}(n-j+1) \times x_{t+j}}{\sum_{i=1}^{m}(m-i+1) \times L_i + \sum_{j=1}^{n}(n-j+1) \times L_j}$$

The terms $m - i - 1$ and $n - j + 1$ are the linearly decreasing weights of the observations within the window. L_i and L_j in the formula are numerical representatives of logicals, also known as Boolean or dummy variables, which are 0 if the corresponding input value is also 0 (indicating that it is missing) in order to eliminate the excess weights from the denominator:

$$L_i, L_j = \begin{cases} 1, & \text{if } x_{t-i}, x_{t+j} > 0 \\ 0, & \text{otherwise} \end{cases}$$

The imputation problem at hand can be described as follows: The gap in the time series is of size k, and the linear weighted moving average is computed of m observations before the target observation and n observations successive to the target observation. For each iteration, the window moves forward to adjacent observations. The value obtained for t in the first iteration will then become an input for subsequent iterations, also receiving the highest weight of m for computing the next observation at $t+1$, the second highest weight of $m-1$ for computing the value at $t+2$ and so on. Accordingly, the higher m and n are set, the more previous and future observations are considered, given they are not missing, but again, the highest weights are assigned to the neighbour observations, if available. An illustration of the algorithm is depicted in Figure 11 below:

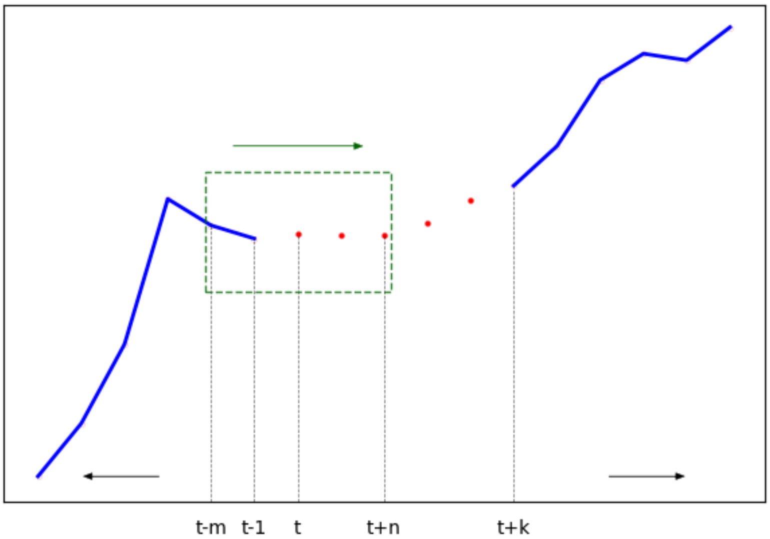

Figure 11: Imputation by moving average algorithm

The parameters m and n are both set to 2 (so the window is said to be symmetric), the moving window (green box) is therefore comprised of $m+n=4$ observations. For each iteration, the window moves forward in time by 1. Be aware that the value computed for t will be used in the next iteration with weight 2 for obtaining $t+1$.

Intuitively speaking, the gaps in the data I attempted to impute using LWMA have most likely been induced by significant historical events, such as WWII. The graphical results are shown in Figure 12. For the nominal GDP of Switzerland in (A), which has only $k=3$ missing observations for the years 1911-1913, LWMA works quite well in establishing a connecting trend. As depicted in (B), a clear spread in the level in the series for M1 of Switzerland is observable. It is unclear, whether a change in definition occurred. According to one source, in the intra-war period from 1914 to 1918, the Swiss government fluted the money market by issuing extreme amounts of currency, which lead to a high inflation rate.[41] This might explain the drastic rise in M1. Though, the imputed values are associated with high uncertainty in terms of correctness. Regarding the series for nominal GDP in (C) and M1 in (D) of Japan, the window size, that is $m+n$, have been increased to 12 and 8 respectively to avoid kinks and constant slopes by covering wider intervals. The imputed series are included in the final data set, whereby the imputed observations are highlighted.

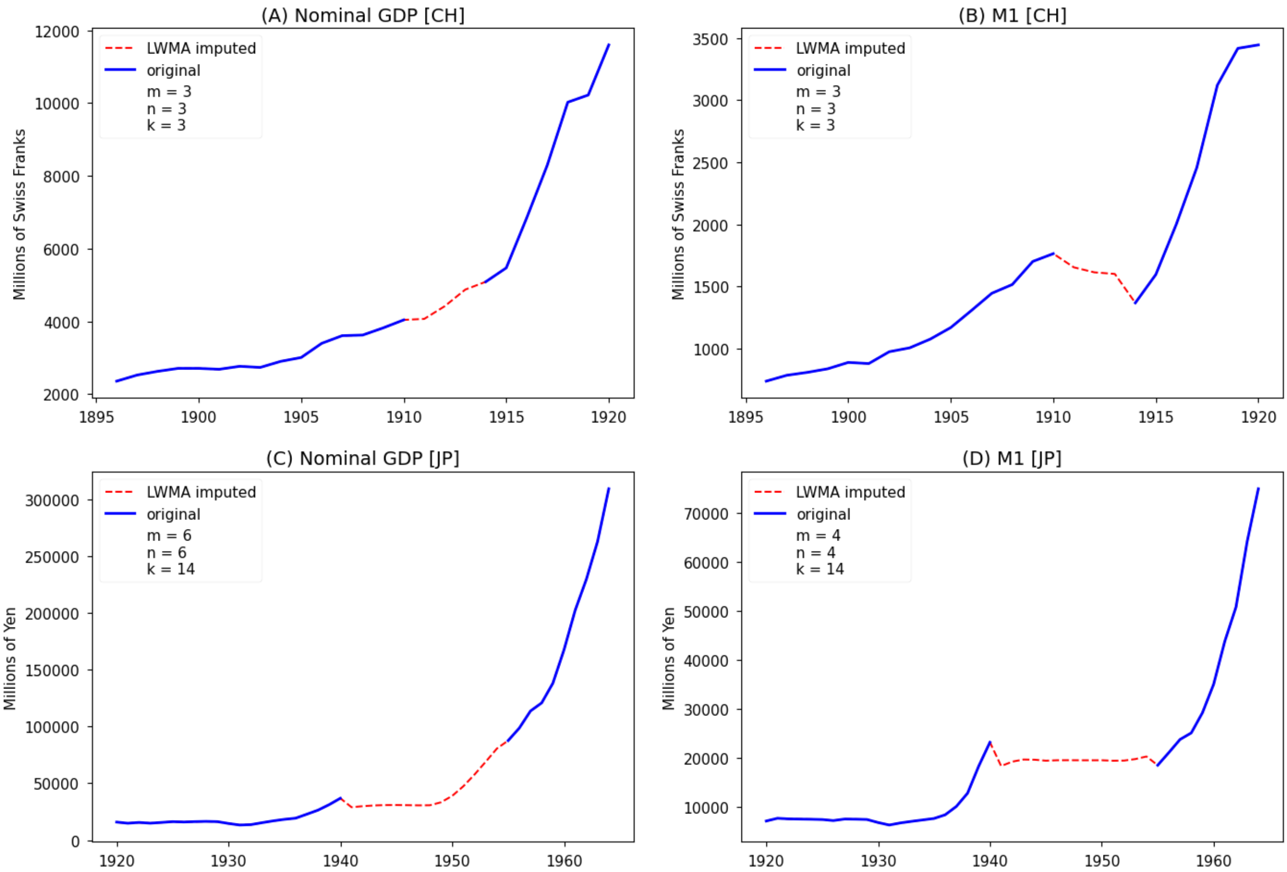

Figure 12: Time series imputed by LWMA

In Figure 12, it is clearly visible, that the large gaps of size $k = 14$ in the Japanese series are problematic for the LWMA algorithm. It is generally not advised to impute such large gaps by any univariate method if the volatility can be expected to be extraordinarily high[42], which is especially the case in times of crisis. Multivariate imputation may produce

[41] Diem and Maissen, n.d.

[42] Kline, 2022

much more accurate approximations using related variables, if available.

An experiment demonstrates the limitations of missing value imputation. A time series for the real expenditure of households available online at Bank of Korea (BOK) has been taken, whereby the values for the periods from 1997 to 1999 and from 2013 to 2015 have been discarded ($k = 3$), and the LWMA algorithm with different window sizes has been applied on iteratively. The accuracy has been assessed by the lowest RMSE of all possible symmetric window sizes ranging from $n = m = 1$ up to $n = m = 8$. This metric is broadly used to quantify and compare prediction performances across different models by measuring the average deviation of the prediction from the actual value.[43] The formula for the RMSE is given as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(\hat{x}_t - x_t)^2}{T}},$$

with $\hat{x}_t - x_t$ being the residual from the fitted to the true value.
The best accuracy has been achieved with $n = m = 2$ for 1997 to 1999, and with $n = m = 7$ for 2013 to 2015. The RMSE score is more than 12 times larger in the 1997 to 1999 period, where the upward trend breaks. Conversely, in the period 2013 to 2015, the error of the imputed values is minimal. Figure 13 shows the results of the experiment for symmetric window sizes of 4, 8 and 14, including the lowest RMSE score achieved in the bottom right corners:

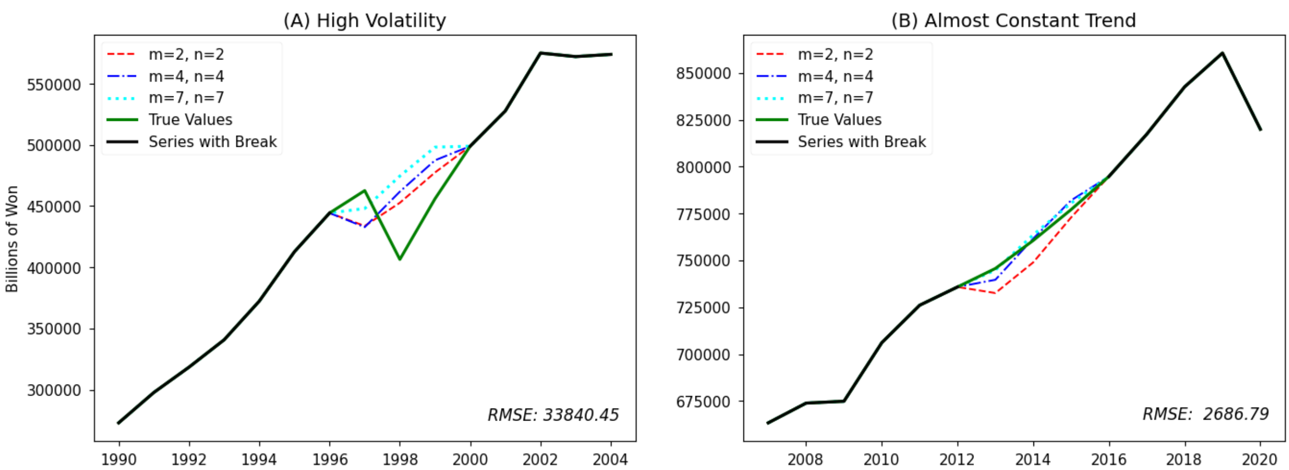

Figure 13: Performance evaluation of LWMA

As this experimental trial shows, imputed values shall not be considered as being accurate

[43] Statista, n.d.

by any means, no matter which method or technique has been applied. The LWMA does reasonably well in filling up gaps aligned with the ongoing trend, but fails in outlier periods. Based on this, the final decision is up to the analyst of the data, whether to consider imputing gaps or not. For this reason, imputed values for the four series shown in Figure 12 have been highlighted in the data collection, and may be discarded if necessary.

A summary of the main arguments for and against an imputation with LWMA is given in the table below:

Pros	**Cons**
easy to implement and interpret	high uncertainty for large gaps
preservation of long-term trend	susceptible to trend outliers

7 Conclusion

Working with large amounts of economic data can undoubtedly be challenging. A good grasp of the variables and the context attached to them is critical when applying any modifications to the data. There is no definite right or wrong when it comes to choosing the correct methodology; yet, looking at the task to accomplish conceptually enables someone to distinguish between appropriate and inappropriate practices. Knowing that monetary aggregates are stock observations recorded at a time point and national accounts are measured over a period, it would neither make sense to sum up monthly observations of the M1 to obtain an annual observation, nor would it make sense to select Q4 of a quarterly series of GDP to represent an entire year.

The temporal aggregation methods actually applied in practice are summarized in the table below:

Type	**Variable**	**Aggregation**
Stocks	Monetary Aggregates	point-in-time sampling, arithmetic mean
Flows	National Accounts	summation, arithmetic mean [a]
Ratios	Unemployment Rate	arithmetic mean
Interest Rates	Government Bond Yield, 3-Month Maturity Yield	arithmetic mean, geometric mean [b]
Price Indices	CPI, GDP Deflator	arithmetic mean

[a] If integrated flows were provided in annualized levels.

[b] Geometric mean has been preferred over arithmetic mean for series with high interest rates if it was unclear what Benati et al. did to obtain annual values.

When trying to construct long-run time series by assembling multiple shorter series, splicing and linking procedures are brought into play. A simplistic, nonetheless powerful splicing method is the use of growth rates. Regardless of comparing indices of different base years, flow or stock variables, as long as growth rates of the two or more series to be joined yield strong similarities, the series can be merged. Other, more sophisticated methods may lead to better accuracy and reliability, though, at the expense of time and effort. Given this trade-off, most independent researchers rather adhere to simple methodologies.

Depending on the final purpose of the data, missing values can be replaced by actual numbers. Various techniques and algorithms are available, all of them come with the downside of potentially misrepresenting the ground truth severely. Especially for wider gaps in the data, accuracy can be expected to decline. The LWMA algorithm used to fill up gaps in the data collection performs well in establishing a connecting trend. For longer gaps and trend breaks, the imputed values shall be treated with caution.

References

Benati, Luca et al. (2019). *Online Appendix For: International Evidence On Long-Run Money Demand*. In: 588. DOI: https://doi.org/10.21034/sr.588.

de la Fuente, Angel (Dec. 2009). *A Mixed Splicing Procedure for Economic Time Series*. In: *Barcelona Economics Working Paper Series* CESifo Working Paper No. 2876. DOI: http://dx.doi.org/10.2139/ssrn.1532638.

Demirhan, Haydar and Zoe Renwick (2018). *Missing Value Imputation for Short to Mid-Term Horizontal Solar Irradiance Data*. In: *Applied Energy* 225, pp. 998–1012.

Di Fonzo, Tommaso (2003). *Constrained Retropolation of High-Frequency Data Using Related Series: A Simple Dynamic Model Approach*. In: *Statistical Methods & Applications* 12.1, pp. 109–119. DOI: https://doi.org/10.1007/BF02511587.

Diem, Aubrey and Thomas Maissen (n.d.). *Switzerland - World War I and Economic Crisis*. Accessed on May 6, 2023. URL: https://www.britannica.com/place/Switzerland/Recent-developments.

Diewert, Erwin (2021). *CPI Theory Chapter 2 Basic Index Number Theory*. Accessed on May 22, 2023. URL: https://www.imf.org/~/media/Files/Data/CPI/companion-publication/chapter-1-basic-index-number-theory.ashx?la=en.

Dippelsman, Robert, Venkat Josyula, and Eric Métreau (2016). *Fixed Base Year vs. Chain Linking in National Accounts: Experience of Sub-Saharan African Countries*. 16/133. URL: https://www.imf.org/external/pubs/ft/wp/2016/wp16133.pdf.

Economic and Social Commission for Asia and the Pacific (ESCAP) (Apr. 2019). *Review of Country Practices on Rebasing and Linking National Accounts Series*. URL: https://unstats.un.org/unsd/DA-SEA-Asia/Documents%5C%20-%5C%20Lao%5C%20WS/Lao%5C%20WS-Country%5C%20practices%5C%20on%5C%20rebasing%5C%20and%5C%20linking%5C%20NA.pdf.

European Central Bank (2023). *Monetary Aggregates and Counterparts*. Accessed on March 26, 2023. URL: https://www.ecb.europa.eu/stats/money_credit_banking/monetary_aggregates/html/index.en.html.

Eurostat (July 2022). *Short-Term Business Statistics and (Annual) Chain Linking*. Accessed on April 26, 2023. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Short-term_business_statistics_and_(annual)_chain_linking.

Eurostat (2023). *Statistics Explained, Your Guide to European Statistics.* Accessed on April 26, 2023. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Main_Page.

– (n.d.). *Procedure for the Aggregation of Indices.* Accessed on May 25, 2023. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Harmonised_indices_of_consumer_prices_(HICP)_-_methodology#Aggregation_procedure.

Fisher, Irving (1922). *The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability.* Reviewed by Bruce D. Mudgett in The ANNALS of the American Academy of Political and Social Science (1923). Houghton Mifflin Company. DOI: https://doi.org/10.1177/000271622310700151.

Flores, Anibal, Hugo Tito, and Deymor Centty (2019). *Model for Time Series Imputation based on Average of Historical Vectors, Fitting and Smoothing.* In: *International Journal of Advanced Computer Science and Applications* 10.10, pp. 346–347.

Forrester, Jay W. (1980). *Industrial Dynamics.* 10. print. Cambridge, Mass. [u.a.] : MIT Press, pp. 63–72.

Fox, Douglas R. (2022). *NIPA Handbook: Concepts and Methods of the U.S. National Income and Product Accounts.* Accessed on May 5, 2023. U.S. Bureau of Economic Analysis. Chap. 4. URL: https://www.bea.gov/resources/methodologies/nipa-handbook.

Gamper, Johann, Michael Böhlen, and Christian S. Jensen (2009). *Temporal Aggregation.* In: *Encyclopedia of Database Systems.* Ed. by LING LIU and M. TAMER ÖZSU. Boston, MA: Springer US, pp. 2924–2929. ISBN: 978-0-387-39940-9. DOI: https://doi.org/10.1007/978-0-387-39940-9_386.

Hüpen, Rolf (2002). *Zur Berechnung von Wachstumsraten in Diskreter Zeit.* URL: https://www.ruhr-uni-bochum.de/wista/download/Beilagen/Wachstumsraten.pdf.

International Monetary Fund (2017). *Quarterly National Accounts Manual 2017.* International Monetary Fund, pp. 159–162. URL: https://www.imf.org/external/pubs/ft/qna/pdf/2017/chapter7.pdf.

Irving, William, Hideaki Nakane, and Jose Ramon T. Villarin (2021). *Time Series Consistency.* In: *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse

Gas Inventories. IPCC. Chap. 5.3. URL: https://www.ipcc-nggip.iges.or.jp/public/2019rf/vol1.html.

Kline, Adrienne (2022). *Implementation and Limitations of Imputation Methods*. Accessed on May 6, 2023. URL: https://towardsdatascience.com/implementation-and-limitations-of-imputation-methods-b6576bf31a6c.

Marcellino, Massimiliano (1999). *Some Consequences of Temporal Aggregation in Empirical Analysis*. In: *Journal of Business Economic Statistics* 17.1. Accessed on March 7, 2023, pp. 129–136. ISSN: 07350015. URL: http://www.jstor.org/stable/1392244.

Martin, Leslie A. (1997). *Beginner Modeling Exercises*. In: *MIT System Dynamics in Education*. URL: https://ocw.mit.edu/courses/15-988-system-dynamics-self-study-fall-1998-spring-1999/c14686381688f6677e9b8bbbce13249f_modeling.pdf.

Office for National Statistics (ONS) (2019). *Consumer Price Indices Technical Manual*. Accessed on May 25, 2023. URL: https://www.ons.gov.uk/economy/inflationandpriceindices/methodologies/consumerpriceinflationtechnicalmanual.

Organisation for Economic Co-operation and Development (OECD) (2023). *Methodology - Linking Time Series, OECD*. Accessed on April 26, 2023. URL: https://www.oecd.org/sdd/leading-indicators/methodology-linking-time-series-oecd.htm.

Prados de la Escosura, Leandro (Sept. 2017). *Spanish Economic Growth, 1850–2015*. Palgrave Macmillan Cham, pp. 169–174. DOI: https://doi.org/10.1007/978-3-319-58042-5.

Statista (n.d.). *Root Mean Square Error (RMSE)*. Accessed on May 8, 2023. URL: https://de.statista.com/statistik/lexikon/definition/303/root_mean_square_error/.

Sterman, John (Jan. 2000). *Business Dynamics, System Thinking and Modeling for a Complex World*. In: *International Institute for Educational Planning (IIEP) UNESCO* 19, pp. 10–19. URL: https://www.researchgate.net/profile/John-Sterman/publication/44827001_Business_Dynamics_System_Thinking_and_Modeling_for_a_Complex_World/links/54359e480cf2643ab9867cb0/Business-Dynamics-System-Thinking-and-Modeling-for-a-Complex-World.pdf.

Tzavidis, Nikolaos (n.d.). *Aspects of Estimation Procedures at Eurostat with Some Emphasis in the Over Space Harmonization*. Accessed on April 28, 2023. (*phdthesis*). Chap. 3

Backward Calculation Techniques, pp. 53–54. URL: http://www2.stat-athens.aueb.gr/~jpan/diatrives/Tzavidis/chapter3.pdf.

Wagner, Reinhard (2006). *Stock-Flow-Thinking und Bathtub Dynamics: Eine Theorie von Bestands- und Flussgrößen.* (*PhD dissertation*). Universität Klagenfurt. URL: https://www.fraktalwelt.de/systeme/rw_diss_stock_flow.pdf.

Wang, Meizhong (2018). *Key Concepts of Intermediate Level Math Textbook.* BCcampus, pp. 143–148. URL: https://collection.bccampus.ca/textbooks/key-concepts-of-intermediate-level-math-bccampus-204.

Wijesekara, W. M. L. K. N. and Liwan Liyanage (2020). *Comparison of Imputation Methods for Missing Values in Air Pollution Data: Case Study on Sydney Air Quality Index.* In: *Advances in Information and Communication.* Ed. by Kohei Arai, Supriya Kapoor, and Rahul Bhatia. Springer International Publishing, pp. 257–269. ISBN: 978-3-030-39442-4.

Wikipedia, contributors (2023). *Discrete Time and Continuous Time — Wikipedia, The Free Encyclopedia.* Accessed on May 20, 2023. URL: https://en.wikipedia.org/wiki/Discrete_time_and_continuous_time&oldid=1140807041.

Wild, Wolfgang (2021). *Finanzmathematik für Nicht-Mathematiker.* Linde Verlag, p. 108.

Young, Ian (2023). *Mathematics for Public and Occupational Health Professional.* Ryerson University, pp. 33–38. URL: https://pressbooks.library.torontomu.ca/ohsmath/.

Zuhu, Helen and Manas Ratha (1997). *Graphical Integration of Linear Flows.* In: *MIT System Dynamics in Education.*

A Data Documentation

Hereby follows a comprehensive description of the time series and methodologies used to construct the data collection. The tables include the name of the new variable, date of last access (in the form of D|M|Y), the unit, period covered and frequency, methodologies applied, and other supplementary details. The series have been added to the data set provided by Benati et al. Shorter descriptions are included in the column header of each series in the data collection.

A.1 Australia

A.1.1 National Accounts

All series stem from *Table 36. Expenditure on Gross Domestic Product (GDP), Chain volume measures and Current prices, Annual* from Australian Bureau of Statistics (ABS) (link). Values are reflecting a June-to-June interval, not December to December. Therefore, the 2022 observations reflect the period June 2021 to June 2022. Regarding nominal national accounts, ABS states that "estimates are valued at the prices of the period to which the observation relates. For example, estimates for this financial year are valued using this financial year's prices. This contrasts to chain volume measures where the prices used in valuation refer to the prices of the previous year". Due to revisions, nominal figures in recent years slightly deviate from those from Benati et al. (see link).

Variable	GROSS DOMESTIC PRODUCT - Original - DERIVED - A2302467A
Source	ABS
Last Accessed	2.3.2022
Description	Unit: Millions Australian Dollars Structure: Time-series from Sep 1959 to Dec 2022 Frequency: quarterly
Methodology	To obtain annual values, the four quarters have been temporally aggregated by summation.
Notes	Since the Series for nominal GDP (column B in Benati et al. data set https://gpih.ucdavis.edu) uses the GDP from ABS from 2001 onwards, it is appropriate to continue the series with the data with a nominal GDP series of ABS

Variable	Gross domestic product: Current prices ;- original - DERIVED - A2304617J
Source	ABS
Last Accessed	2.3.2023
Description	Unit: Millions Australian dollars Structure: Time-series from Jun 1960 to Jun 2022 Frequency: annual
Methodology	The series has been retropolated by the growth rates of the series for nominal GDP provided by Benati et al. as follows: *Nominal GDP* from 1959 to 1950, and *GDP mln* from 1949 to 1870

Variable	Gross domestic product: Chain volume measures - original - DERIVED - A2304755F
Last Accessed	2.3.2023
Description	Unit: Millions of Chained 2012 USD (ref. period: 2020-2021) Structure: Time-series from 1929 to 2021 Frequency: yearly

Variable	Households ; Final consumption expenditure: Current prices - original - DERIVED - A2304591W
Last Accessed	2.3.2023
Description	Unit: Millions Australian dollars Structure: Time-series from Jun 1960 to Jun 2022 Frequency: annual

Variable	Households ; Final consumption expenditure: Chain volume measures - original - DERIVED - A2304724R
Last Accessed	2.3.2023
Description	Unit: Millions of Chained Australian dollars (ref. period: 2020-2021) Structure: Time-series from Jun 1960 to Jun 2022 Frequency: annual
Methodology	Rebasing by growth rates. Growth rates show a high similarity; thus, series can be linked.

A.1.2 Price Index

A series for the CPI index of Australia is available in the table *Consumer Price Inflation - G1* at the Reserve Bank of Australia (RBA) (link).

Variable	Consumer price index; All groups - Original - Index: 2011/12=100 - GCPIAG
Description	Unit: Index (Base period: 2011/12=100) Structure: Time-series from Jun (Q2) 1922 to Dec (Q4) 2022 Frequency: quarterly
Methodology	Temporal aggregation from quarters to years using the arithmetic mean for each year's available observations. 1922 value has been calculated as follows: (Q2+Q3+Q4)/3.

A.1.3 Unemployment Rate

Series for the unemployment rate of Australia have been retrieved from IMF (link) and ABS (link). The forecasted figures for 2023 to 2027 have been discarded from IMF's series.

Variable	Unemployment rate
Last Accessed	3.3.2023
Description	IMF: Unit: percentage Structure: Time-series from 1980 to 2027 Frequency: annual ABS: Unit: percentage rate (seasonally adjusted) Structure: Time-series from Jan 2013 to Jan 2023 Frequency: monthly
Methodology	Temporal aggregation of monthly series provided by ABS to annual frequency using arithmetic mean of all months within a year. A comparison of aggregated ABS' and IMF's series for has been conducted for validation purposes. Both series are identical. For this reason, only the longer series from IMF has been added to the data collection.

A.1.4 Interest Rates

Series for long and short-term interest rates have been obtained from the RBA. Another series for long-term interest rates of government bonds has been taken from Fred St. Louis Fed, since the 10-year government bond series from RBA is only available for the recent decade.

Variable	Bank Accepted Bills/Negotiable Certificates of Deposit-3 months; monthly average - original - FIRMMBAB90
Last Accessed	2.3.2023
Description	Unit: percentage rate Structure: Time-series from Jun 1969 to Feb 2023 Frequency: monthly
Methodology	Temporal aggregation from monthly to annual values using arithmetic mean, as Benati et al. stated in their appendix.

Variable	NSW Treasury Corporation 10 year bond - Yields on New South Wales Treasury Corporation bonds, 10 years maturity - FCMYGBNT10 Table F2.1
Last Accessed	3.3.2023
Description	Unit: percentage rate Structure: Time-series from Jun 2013 to Feb 2023 Frequency: monthly
Methodology	Temporal aggregation from monthly to annual values using arithmetic mean.

Variable	Interest Rates, Government Securities, Government Bonds for Australia, Percent per Annum, Monthly, Not Seasonally Adjusted - INTGSBAUM193N
Last Accessed	3.3.2023
Description	Unit: percentage rate Structure: Time-series from Jul 1969 to Nov 2019 Frequency: monthly
Methodology	Temporal aggregation to annual frequency using the arithmetic mean for the months available. Note that the aggregated value for 1969 is the average of Jul to Dec, the 2019 value is the average from Jan to Nov.
Limitations	Besides the missing month for 1969 and 2019, time series cannot be completed to 2022. Values are not similar enough to justify linking with the series provided by RBA.

A.1.5 Monetary Aggregates

All monetary aggregates have been retrieved from the RBA (link). All series have first been converted from billions to millions of Australian dollars (multiplication by 1000), then temporally aggregated using the arithmetic mean to match the approach ofy Benati et al.

Variable	Monetary Base - DMAMMB
Last Accessed	2.3.2023
Description	Unit: Billions of Australian Dollars (in data collection: millions) Structure: Time-series from Jan 1975 to Jan 2023 Frequency: monthly
Limitations	1975 value is the arithmetic average of Feb to Dec, as the Jan observation was not available.

Variable	M1 - DMAM1N
Last Accessed	2.3.2023
Description	Unit: Billions of Australian Dollars (in data collection: millions) Structure: Time-series from Feb 1975 to Jan 2023 Frequency: monthly
Methodology	Linked to M1 series of Benati et al. in 1975 as the series are identical between 1975 and 2002.
Limitations	1975 value is the arithmetic average of Feb to Dec, as the Jan observation was not available. Up to 2002, the new series and the old series from Benati et al. are identical. Since 2002, the new M1 grows above the old series, most likely due to a change in definition. No information could be investigated.

Variable	M3 - DMAM3N
Last Accessed	2.3.2023
Description	Unit: Billions of Australian Dollars (in data collection: millions) Structure: Time-series from Jul 1959 to Jan 2023 Frequency: monthly
Methodology	Linked to M3 series of Benati et al. in 1959.
Limitations	1959 value is the arithmetic average of Jul to Dec, as the Jan to Jun observations were not provided.

Variable	Broad Money (M3 plus) - DMABMN
Last Accessed	2.3.2023
Description	Unit: Billions of Australian Dollars (in data collection: millions) Structure: Time-series from Aug 1976 to Jan 2023 Frequency: monthly
Limitations	1975 value is the arithmetic average of Aug to Dec, as the Jan to Jul observations were not provided.

A.2 Canada

Multiple versions of series for nominal and real national account variables are available at Statistics Canada's homepage: A series for the Nominal GDP has been taken from IMF and converted from US dollars to millions of Canadian dollars through the use of exchange rates provided by the Bank of Canada (BOC). A series for GDP at constant prices has been accessed via the CANSIM-API (R-script to access data). Remaining national account series have been taken from Statistics Canada's data portal. Retroplolation and linking has been performed with old series form Benati et al. to obtain long-run series.

A.2.1 National Accounts

Variable	Nominal GDP
Source	IMF
Last Accessed	5.3.2023
Description	Unit: Thousands of billions of US dollars Structure: Time-series from 1980 to 2022 Frequency: annual
Methodology	Firstly, thousands of billion USD have been converted to billions of USD (multiplication by 1000). Secondly, the conversion from USD to CAD has been done by using exchange rates from BOC (series *Legacy Annual Average Rates IEXA0101* for 1997-2016, and *Annual average exchange rates FXAUSDCAD* for 2017-2022). The converted series (in CAD) has been linked to the series of Nominal GDP provided by Benati et al. in 1997 onwards.

Variable	Gross domestic product, income-based, quarterly (x 1,000,000) at market prices
Source	Statistics Canada: Table: 36-10-0103-01 (formerly CANSIM 380-0063)
Last Accessed	6.4.2023
Description	Unit: Billions of Canadian dollars, seasonally adj, annual rate Structure: Time-series from Q1 1961 to Q4 2022 Frequency: quarterly
Methodology	Temporal aggregation from quarters to annual observations using arithmetic mean, since series is given in annualized levels.

Variable	Historical (real-time) releases of gross domestic product (GDP) at basic prices, by industry, monthly (x 1,000,000)
Source	Statistics Canada: Table: 36-10-0491-01 (formerly CANSIM 379-8031)
Last Accessed	8.5.2023
Description	Unit: Millions of Canadian dollars (2012 prices), seasonally adj. annual rates Structure: Time-series from Jan 1997 to Feb 2023 Frequency: monthly
Methodology	Arithmetic mean to temporally aggregate from months to years, as values are in annual rates. Retropolation has been performed from 1997 to 1870 using growth rates from Benati et al.'s *Real GDP* series. Growth rates match well in overlap period.

Variable	Quarterly expenditure-based, gross domestic product, Canada, in chained (2012) and current dollars
Source	Statistics Canada: Table: 36-10-0104-01 (formerly CANSIM 380-0064)
Last Accessed	4.3.2023
Description	Unit: Millions of 2012 Chained Canadian dollars, seasonally adj. annual rates Structure: Time-series from Q1 1961 to Q4 2022 Frequency: quarterly
Methodology	Arithmetic mean to temporally aggregate from quarters to years, as values are in annual rates. Retropolation performed from 1960 to 1870 using growth rates from Benati et al.'s *Real GDP* series. Growth rates match well in overlap period.

Variables	Household final consumption expenditure and Final consumption expenditure in:Gross domestic product, expenditure-based, provincial and territorial, annual (x 1,000,000)
Source	Statistics Canada: Table: 36-10-0222-01 (formerly CANSIM 384-0038)
Last Accessed	1.6.2023
Description	Both Series: Unit: Millions of Canadian dollars at current prices Structure: Time-series from 1981 to 2021 Frequency: annual
Methodology	Household final consumption expenditure linked to *Personal expenditure on consumer goods and services* from Benati et al. Linking has been chosen over retropolation because Benati et al. also linked a similar series from 1981 to v96730339 (from 1926 to 1980) and there are no major differences in levels.

Variables	Household final consumption expenditure and Final consumption expenditure in: Gross domestic product, expenditure-based, provincial and territorial, annual (x 1,000,000)
Source	Statistics Canada: Table: 36-10-0222-01 (formerly CANSIM 384-0038)
Last Accessed	1.6.2023
Description	Both Series: Unit: Millions of Chained 2012 Canadian dollars, seasonally adj. annual rates Structure: Time-series from 1981 to 2021 Frequency: annual
Methodology	Retropolation performed for Households final consumption expenditure from 1980 to 1926 using growth rates from Benati et al.'s *Personal expenditure on consumer goods and services at market prices* series. Growth rates match well in overlap period.

A.2.2 Price Index

Variable	Consumer Price Index, annual average, not seasonally adjusted
Source	Statistics Canada: Table 18-10-0005-01
Last Accessed	4.3.2023
Description	Unit: Index 2002=100 Structure: Time-series from 1914 to 2022 Frequency: annual
Methodology	Linked to rebased *CPI* series from Benati et al. Rebasing from 1992=99.975 to 2002=100 yields the exact same index numbers as the new CPI index, therefore, linking can be justified.

A.2.3 Unemployment Rate

Benati et al. did not include a series for the unemployment rate in their data set; hence, the variable has been taken from IMF. For unknown reasons, the 1975 observation is missing in IMF's series. The LWMA algorithm has been deployed to impute the missing observation.

Variable	Labor Markets, Unemployment Rate, Percent - 2023 - M03
Source	IMF
Last Accessed	4.3.2023
Description	Unit: Percentage Structure: Time-series from 1969 to 2021 Frequency: annual
Methodology	Missing vlaue for 1975 has been imputed by LWMA, by setting $m = n = 2$. The imputed value is highlighted in the data collection.
Limitations	Missing 1975 value. Has already been double-checked, whether the 1975 has accidentally not been selected.

A.2.4 Interest Rates

Both 3-month treasury bill rate and 10-year government bond yields are available at the BOC. They have been linked to the series provided by Benati et al.

Variable	Treasury bill auction - 3 month - Average yields - V80691303
Last Accessed	5.3.2023
Description	Unit: Percentage, average yields Structure: Time-series from Jan 4 2000 to Feb 28 2022 Frequency: daily (irregular frequency)
Methodology	Temporal aggregation using the arithmetic average of all daily values. Linked to *3-month Treasury bill rate* of Benati et al.

Variable	Over 10 year Government of Canada marketable bonds - Average yield - CDN.AVG.OVER.10.AVG
Last Accessed	5.3.2023
Description	Unit: Percentage, average yield Structure: Time-series from Jan 2 2001 to Mar 2 2023 Frequency: daily (irregular frequency)
Methodology	Temporal aggregation using the arithmetic average of all daily values. Linked to *Over 10-year government marketable bonds* of Benati et al.

A.2.5 Monetary Aggregates

All series of monetary aggregates have been taken from BOC. All series come in monthly frequency, and have been temporally aggregated using the arithmetic mean to match the approach of Benati et al.

Variables	Currency outside banks (Unadjusted) - V37173 M2 (gross) (Unadjusted) - V41552786 M2+ (gross) (Unadjusted) - V41552788 M2++ (gross) (Unadjusted) - V41552801 M3 (gross) (Unadjusted) - V41552785
Last Accessed	5.3.2023
Link	link
Description	All series: Unit: Millions of Canadian dollars (Month-end) Frequency: monthly Currency outside banks: Structure: Time-series from Jan 1946 to Dec 2022 M2, M2+ & M2++: Structure: Time-series from Jan 1968 to Dec 2022 M3: Structure: Time-series from Jan 1970 to Dec 2022
Methodology	All series have been temporally aggregated by the arithmetic mean. Only the series for currency outside banks has then been linked to *Currency outside banks at month-end* from Benati et al.

Variable	M1B (gross) (currency outside banks, chartered bank chequable deposits, less inter-bank chequable deposits) - v41552795 - Table: 10-10-0116-01 (formerly CANSIM 176-0025)
Last Accessed	5.3.2023
Link	link
Description	Unit: Millions of Canadian dollars (Month-end) Structure: Time-series from Jan 1968 to Dec 2022 Frequency: annual
Methodology	Temporally aggregation to annual values using the arithmetic mean.

A.3 Euro Area

Quarterly data for the Euro Area provided by Benati et al. has been constructed from the Area Wide Model (AWM) database (link), which covers macroeconomic variables back to 1970, and series from the ECB's *Statistical Data Warehouse*. To obtain annual values, the provided data has been temporally aggregated before it has been continued up to 2022 by most recent series of the ECB data warehouse (link). An exception applies to the series for short-term interest rate, whereby no suitable series could be identified. The data for the short-term interest rate series for most recent years has been obtained from the Organisation for Economic Co-operation and Development (OECD). Hereby, it is worth mentioning that the ECB simply picks the last observation of high-frequency series (point-in-time) of interest rates to establish annual frequent series, rather than applying a kind of mean. Although proposed differently in this thesis, the approach of the ECB to select the end of period observation has been adopted to ensure consistency.

The quarterly series in the data set from Benati et al. have been temporally aggregated as follows:

Variable	**Aggregation**
National Accounts	summation
HICP	arithmetic mean
Unemployment Rate	arithmetic mean
Interest rates	end of period
M1, M2 & M3	end of period

A.3.1 National Accounts

Variable	Gross domestic product at market prices MNA.Q.Y.I8.W2.S1.S1.B.B1GQ._Z._Z._Z.EUR.V.N
Last Accessed	28.5.2023
Description	Unit: Millions of Euro Structure: Time-series from Q1 1995 to Q4 2022 Frequency: quarterly
Methodology	Temporal aggregation to annual frequency by summing up quarterly observations. Linked to temporally aggregated *Nominal GDP* of Benati et.al.

Variable	Gross domestic product at market prices (Chain linked volume, 2015=100) MNA.Q.Y.I8.W2.S1.S1.B.B1GQ._Z._Z._Z.EUR.LR.N
Last Accessed	2.6.2023
Description	Unit: Millions of Chained 2015 Euros Structure: Time-series from 1995 to 2022 Frequency: annual
Methodology	Growth rates of temporally aggregated *GDP (Real)* from Benati et al. have been used to retropolate the new GDP at market prices from 1995 to 1970.

Variable	Private final consumption at current prices MNA.Q.N.I8.W0.S1M.S1.D.P31._Z._Z._T.EUR.V.N
Last Accessed	2.6.2023
Description	Unit: Millions of Euro Structure: Time-series from 1995 to 2022 Frequency: annual

Variable	Private final consumption (Chain linked volume, 2015=100) MNA.Q.N.I8.W0.S1M.S1.D.P31._Z._Z._T.EUR.LR.N
Last Accessed	2.6.2023
Description	Unit: Millions of Chained 2015 Euro Structure: Time-series from 1995 to 2022 Frequency: annual
Methodology	Retropolation by growth rates of temporally aggregated *Private consumption* (originally from AWM under abbreviation PCR) from Benati et al. for 1995 to 1970.

A.3.2 Price Index

Since Benati et al. did not include a series for a price index in their data set, the HCPI has been taken from the AWS database, temporally aggregated by using the arithmetic mean, rebased to 2015, and then linked to the new HICP index series of the ECB.

Variables	HICP - Overall index - ICP.M.U2.N.000000.4.INX [ECB] HICP [AWS]
Last Accessed	2.6.2023
Description	HICP [AWS]: Unit: Index 1996=100 Structure: Time-series from Q1 1970 to Q4 2017 Frequency: quarterly HICP [ECB]: Unit: Index 2015=100 Structure: Time-series from 1996 to 2022 Frequency: annual
Methodology	HICP [AWS] has been temporally aggregated by the arithmetic average such that 1996=100 before rebasing it to 2015=100. Then, HICP [AWS] has been linked to HICP [ECB].

A.3.3 Unemployment Rate

Variable	Unemployment Rate IESS.Q.I8.S.UNEHRT.TOTAL0.15_74.T
Last Accessed	2.6.2023
Description	Unit: Percentage Structure: Time-series from 1999 to 2022 Frequency: annual
Methodology	Linked to aggregated temporally *Unemployment Rate* from Benati et al.

A.3.4 Interest Rates

A long-run series for 10-year government bond yield ranging from 1970 to 2022 is available at the ECB's *Statistical Data Warehouse*. A suitable series for short-term interest rate could not be found; hence, it has been taken from the OECD. Because the AWS database offers a quarterly series for the short-term rate from 1970 to 2017, the series from OECD has been used to continue the series from AWS from 2018 to 2022.

Variable	10-year Government Benchmark bond yield Euro area FM.M.U2.EUR.4F.BB.U2$_1$0$Y.YLD$
Last Accessed	2.6.2023
Description	Unit: Percentage Structure: Time-series from 1970 to 2022 Frequency: annual
Notes	The *Long-Term Interest Rate (Nominal in percent)* from Benati et al. has been temporally aggregated using the end-of-period approach. The annual series of 10-year government bond yield is almost identical to the aggregated series of Benati et al.

Variables	Short-term interest rates [OECD] STN [AWS]
Last Accessed	2.6.2023
Description	Both series: Unit: Percentage STN [AWS]: Structure: Time-series from Q1 1970 to Q4 2017 Frequency: quarterly Short-term interest rates [OECD]: Structure: Time-series from 1994 to 2022 Frequency: annual
Methodology	The STN series from AWS has been temporally aggregated selecting the last quarter of each year. The series from OECD has been linked to the aggregated STN series.
Notes	Both series are very similar, hence, a linking can be justified.

A.3.5 Monetary Aggregates

Monetary aggregates M1, M2 and M3 are available at the ECB's *Statistical Data Warehouse*. All series have been temporally aggregated by selecting the year-end stock, that is, the Q4 observation.

Variables	M1 - BSI.M.U2.Y.V.M10.X.1.U2.2300.Z01.E M2 - BSI.M.U2.Y.V.M20.X.1.U2.2300.Z01.E M3 - BSI.M.U2.Y.V.M30.X.1.U2.2300.Z01.E
Last Accessed	28.5.2023
Description	All series: Unit: Millions of Euro Structure: Time-series from Q1 1995 to Q4 2022 Frequency: quarterly
Methodology	Following the methodology of the ECB, the series for monetary aggregates have been temporally aggregated by selecting Q4 of each year to establish an annual time series. All three series have been linked to aggregated *M1*, *M2* and *M3* of Benati et al.

A.4 Japan

A.4.1 National Accounts

All series for both nominal and real national accounts stem from Economic and Social Research Institute (ESRI) from table *1. Real Gross Domestic Product (Expenditure approach: at current prices and chain-linked)* IV. *Main Time Series* (link).

Variables	5. Gross domestic product (expenditure approach) and (1) Final consumption expenditure of households
Last Accessed	08.04.2023
Description	Both series: Unit: Billions of Yen (current prices) Structure: Time-series from 1994 to 2021 (Calendar Year) Frequency: annual
Methodology	Retropolation performed for new nominal GDP from 1996 with growth rates of *Nominal GDP* from Benati et al. Retropolation has been used rather than linking because there are notable differences in the new series and the old series from the original data set. The growth rates match well. Since the old series for the nominal GDP had no records for the period from 1941 to 1954, the series has been imputed by LWMA (parameter arguments: $m = n = 6$). The new series for final consumption expenditure of households could not be spliced, as neither the levels, nor the show reasonable similarity.
Limitations	The growth rates used for retropolation from 1955 backwards may therefore not be regarded as reliable, and as a result, the retropolated values before 1955 shall be used carefully, considering the lengthy gap of 14 missing observations.

Variables	5. Gross domestic product (expenditure approach) and (1) Final consumption expenditure of household
Last Accessed	8.4.2023
Description	Both series: Unit: Billions of Chained 2015 Yen Structure: Time-series from 1994 to 2021 Frequency: annual
Methodology	The new series for real GDP has been retropolated with growth rates of old series for *Real GDP* from Benati et al. Growth rates are not perfectly identical, but overall, show an acceptable degree of similarity.

A.4.2 Price Index

Variable	2020-Base Consumer Price Index
Source	Statistics of Japan e-Stat
Last Accessed	08.04.2023
Link	link
Description	Unit: Index 2020=100 Structure: Time-series from 1970 to 2022 Frequency: annual

A.4.3 Unemployment Rate

Variable	Unemployment Rate
Source	Statistics Bureau of Japan
Last Accessed	08.04.2023
Link	link
Description	Unit: Percentage (annual average figures - Results of whole Japan) Structure: Time-series from 1953 to 2022 Frequency: annual
Limitations	Reliability concerns of the Statistics Bureau of Japan are stated in the spreadsheet as follows: "Reliability of data before 1973. Several revisions and sampling approaches reconsidered over time [...] Estimation of 2011 value due to loss of data caused by the great earthquake of Japan in that year."

A.4.4 Inerest Rates

Series for the overnight rate, 3-month call rate and basic discount rate are available at the BOJ's data portal (link). Series for 3-month money market rate as well as a series for 10-year government bond yield from OECD have also been added to the data collection.

Variables	Call Rate, Uncollateralized Overnight, Average (Daily) - FM01'STRDCLUCON Call Rate, Uncollateralized 3 Months/Average - FM02'STRACLUC3M The Basic Discount Rate and Basic Loan Rate - IR01'MADR1M
Source	BOJ
Last Accessed	11.4.2023
Description	All Series: Unit: Percentage rate (average) Frequency: annual Overnight Call rate: Structure: Time-series from 1998 to 2022 3-Month Call rate: Structure: Time-series from 1989 to 2022 Basic Discount & Loan Rate: Structure: Time-series from 1882 to 2022
Limitations	Value for 1882 of the basic discount and loan rate is the average of October, November and December values only. Earlier observations are not available.

Variables	3-Month Money Market Rate and 10-year Government Bonds yields
Source	OECD
Last Accessed	11.4.2023
Description	Both Series: Unit: Percentage rate Frequency: annual 3-Month Money Market Rate: Structure: Time-series from 2002 to 2022 10-year Government Bonds yields: Structure: Time-series from 1989 to 2022

A.4.5 Monetary Aggregates

All series for monetary aggregates have been obtained from the BOJ's data portal (link). Long-run series for M1, M2 and M3 have been created via splicing of several shorter series. Precise description of the splicing and linking procedures are included in the headings. Also, the old series from Benati et al. has been imputed by the LWMA algorithm to retropolate the spliced new M1 using the growth rates of the old series (similar to nominal GDP in A.4.1).

Variable	Monetary Base/Average Amounts Outstanding - MD01'MABS1AN11
Last Accessed	8.4.2023
Description	Unit: 100 millions of Yen (average) Structure: Time-series from 1960 to 2022 Frequency: annual
Methodology	Linked to *Monetary Base* of Benati et al.

Variables	_M1/Amounts Outstanding at End of Period/Money Stock - MD02'MAM1NEM3M1MO, MD02'MAMS1EN01 and MD02'MAMS3EN01
Last Accessed	8.4.2023
Description	All Series: Unit: Millions of Yen Frequency: monthly MD02'MAMS1EN01: Structure: Time-series from Jan 1955 to Mar 1999 MD02'MAMS3EN01: Structure: Time-series from Apr 1998 to Mar 2008 MD02'MAM1NEM3M1MO: Structure: Time-series from Apr 2003 to Dec 2022
Methodology	Splicing has been done on a disaggregated monthly frequency. After splicing, the series has been temporally aggregated using the arithmetic mean. The splicing has been done as follows: (1) Computation of growth rates (2) A long series of growth rates has been established as follows: 1. From Jan 1955 to Mar 1998: MD02'MAMS1EN01 - (discontinued)_M1/Amounts Outstanding at End of Period/(Reference) Money Stock (Based on excluding Foreign Banks in Japan, etc., through March 1999) 2. From Apr 1998 to Mar 2008: MD02'MAMS3EN01 - (discontinued)_M1/Amounts Outstanding at End of Period /(Reference) Money Stock (from April 1998 to March 2008) 3. From Apr 2008 to Dec 2022: MD02'MAM1NEM3M1MO - _M1/Amounts Outstanding at End of Period/Money Stock (3) Retropolation (still on monthly frequency) starting from the 2022 value of MD02'MAM1NEM3M1MO using the linked growth rate series. (4) Finally, the M1 stock was temporally aggregated by taking monthly averages (5) The missing values in the old M1 series from Benati et al. has been LWMA-imputed (parameter arguments: $n = m = 4$) (6) The growth rates of the imputed old M1 series has been used to further retropolated from 1954 to 1885 (like nominal GDP)

Variable	M2/Average Amounts Outstanding/Money Stock - MD02'MAM1NAM2M2MO
Last Accessed	8.4.2023
Description	Unit: Millions of Yen Structure: Time-series from 2003 to 2022 Frequency: monthly
Methodology	MD02'MAM1NAM2M2MO has been linked to old M2 of Benati et al.

Variables	M3/Amounts Outstanding at End of Period/Money Stock - MD02'MAM1NEM3M3MO, (discontinued)_M3+CDs-Money Trusts/Amounts Outstanding at End of Period /(Reference) Money Stock (from April 1998 to March 2008) - MD02'MAMS3ENM3 and (discontinued)M3+CDs (old)/Amounts Outstanding at End of Period/(Reference) Money Stock (Based on excluding Foreign Banks in Japan, etc., through March 1999) - MD02'MAMS1EN10
Last Accessed	8.4.2023
Description	All Series: Unit: Millions of Yen Frequency: monthly MD02'MAMS1EN10: Structure: Time-series from 1971 to 1999 MD02'MAMS3ENM3: Structure: Time-series from 1998 to 2008 MD02'MAM1NEM3M3MO: Structure: Time-series from 2003 to 2022
Methodology	The splicing has been done as follows: (1) Computation of growth rates (2) A long series of growth rates has been established as follows: 1. From 1972 to 1998: MD02'MAMS1EN10 2. From 1999 to 2003: MD02'MAMS3ENM3 3. From 2004 to 2022: MD02'MAM1NEM3M3MO (3) Retropolation starting from the 2022 value of MD02'MAM1NEM3M3MO using the linked growth rate series

A.5 South Korea

Most series are available at Economic Statistics System (ECOS), which is the data system managed by the BOK (link). The unemployment rate and series for short and long term interest rates have been taken from IMF.

A.5.1 National Accounts

All series for national accounts have been taken from ECOS. No linking or splicing was necessary, as no series from Benati et al. ranges further back in the past than the new series.

Variables	Gross domestic product (current prices, won) - 10101, Gross national income (current prices, won) - 10102 and GDP deflator 2015=100 - 90103 in: 2.1.1.1. Main Annual Indicators (reference year 2015)
Last Accessed	12.4.2023
Description	All Series: Structure: Time-series from 1953 to 2022 Frequency: annual GDP and GNI: Unit: Billions of Won GDP deflator: Unit: Index 2015=100

Variables	Gross domestic product at market prices (GDP) - 1400 and Gross national income (GNI) - 1800 in: 2.1.2.1.2. GDP and GNI by Economic Activities (seasonally adjusted, chained 2015 year prices, quarterly)
Last Accessed	12.4.2023
Description	Both Series: Unit: Billions of Chained 2015 Won Structure: Time-series from Q1 1960 to Q4 2022 Frequency: quarterly
Methodology	Both real GDP and real GNI have been temporally aggregated by summation of the four quarters.

Variables	Final consumption - 10101, Household consumption - 1010111 and Private consumption - 1010110 in: 2.1.2.2.3. Expenditures on GDP (not seasonally adjusted, current prices, quarterly & annual)
Last Accessed	13.4.2023
Description	All Series: Unit: Billions of Won Structure: Time-series from 1953 to 2022 Frequency: annual

Variables	Final consumption - 10101, Household consumption - 1010111 and Private consumption - 1010110 in: 2.1.2.2.4. Expenditures on GDP (not seasonally adjusted, chained 2015 year prices, quarterly & annual)
Last Accessed	12.4.2023
Description	All Series: Unit: Billions of Chained 2015 Won Structure: Time-series from 1953 to 2022 Frequency: annual

A.5.2 Price Index

Variable	CPI index (Code: 0, Wgt: 1000) in: 4.2.1. Consumer Price indices
Last Accessed	13.4.2023
Description	Unit: Index 2020=100 Structure: Time-series from 1965 to 2022 Frequency: annual

A.5.3 Unemployment Rate

A series for the unemployment rate is available at ECOS. Because this series only covers the recent 23 years, another series for the unemployment rate of Korea has been taken from IMF's *World Economic Outlook* to cover a longer period.

Variables	Unemployment Rate [ECOS] - I61BC in: 8.6.2. Summary of Economically Active Pop. and Unemployment Rate [IMF]
Last Accessed	13.4.2023
Description	Both Series: Unit: Percentage Frequency: annual Unemployment Rate [IMF]: Structure: Time-series from 1980 to 2022 Unemployment Rate [ECOS]: Structure: Time-series from 2000 to 2022
Methodology	Both series have been added to the data collection as they are almost identical (difference at most $\pm 0.1\%$) for available observations.
Limitations	A monthly series for the unemployment rate has also been accessed from ECOS. This monthly series starts in Jun 1999, and the arithmetic average of Jun to Dec values has been used to derive the annual value for 1999. The 1999 observation has been added and highlighted in the data collection as it does not represent the entire year average, but only of the second half of 1999.

A.5.4 Interest Rates

No suitable series for short and long-run interest rates have been identified from the BOK. As alternatives, three series for interest rates have been retrieved from the IMF's financial data portal International Financial Statistics (IFS) (link) as Benati et al. also used data from IFS.

Variables	Discount Rate - FID_PA, Money Market Rate - FIMM_PA and Government Bonds - FIGB_PA
Last Accessed	13.4.2023
Description	All Series: Unit: Percentage Frequency: annual Discount Rate: Structure: Time-series from 1964 to 2022 Money Market Rate: Structure: Time-series from 1976 to 2022 Government Bonds: Structure: Time-series from 1963 to 2022
Methodology	The new discount rate series has been linked to *Discount Rate* of Benati et al., which originally stems from IFS too.

A.5.5 Monetary Aggregates

Series for Monetary Base, M1 and M2 are available at ECOS. Benati et al. included a series for Monetary Base and M1 in the form of average stock, and a series for the M2 is given in end-of-period terms. The most plausible explanation for this is that the average series starts in 1986, while the end of series covers a longer time span as it starts in 1960. Following their approach, the new series for Monetary Base and M1 are expressed in annual average stock, for the M2, both an end-of-period and an average stock series has been added to the data collection.

Variables	1.1.1.1.2. Components of Monetary Base (Average) - ABA1 1.1.2.1.2. M1 By Type (Average) - BBLA00 1.1.3.1.2. M2 By Type (Average) - BBHA00 1.1.3.1.4. M2 By Type (End of) - BBGA00
Last Accessed	13.4.2023
Description	All Series: Unit: Billions of Won Frequency: annual Monetary Base: Structure: Time-series from 1971 to 2022 M1: Structure: Time-series from 1970 to 2022 M2 (Average): Structure: Time-series from 1986 to 2022 M2 (End of): Structure: Time-series from 1960 to 2022

A.6 Switzerland

A.6.1 National Accounts

Except from a reconstructed long-run series (from the Federal Statistical Office (BFS)), all series for national accounts have been taken from State Secretariat for Economic Affairs (SECO) (link). Benati et al. constructed an extended series for the nominal GDP, however, with a three-year gap from 1911 to 1913. The missing values have been imputed by the LWMA algorithm and used for splicing.

Variable	GDP at current prices, millions of Swiss francs Gross domestic product, long time series FSO nr: je-e-04.02.01.08
Source	Federal Statistical Office (BFS)
Last Accessed	7.4.2023
Link	link
Description	Unit: Millions of Swiss Francs Structure: Time-series from 19248 to 2021 Frequency: annual
Limitations	According to the Federal Statistical Office, this series should be used with caution. They state: "The data prior to 1995 has been obtained by retropolation using evolution rates, which have been estimated on the basis on old accounting systems (OECD 1952, ESA 1979, ESA 1995). [...] Therefore, this long time series cannot ensure conceptual and methodological consistency for the entire period."

Variable	Gross domestic product (B.1*b from table nom_y) in: ESA 2010, Annual aggregates of Gross Domestic Product, expenditure approach (SFSO, SECO), sport event adjusted data
Last Accessed	7.4.2023
Description	Unit: Millions of Swiss Francs (at current prices) Structure: Time-series from 1980 to 2022 Frequency: annual
Methodology	Retropolated new nominal GDP by growth rates of LWMA-imputed *Nominal GDP* from Benati et al. The parameter arguments for the algorithm have been set to $m = n = 3$. Imputed values are highlighted in the data collection.

Variable	Gross Domestic Product (B.1*b from table real_y) in: ESA 2010, Annual aggregates of Gross domestic Product, expenditure approach (SFSO, SECO), sport event adjusted data
Last Accessed	7.4.2023
Description	Unit: Millions of Chained 2015 Swiss Francs Structure: Time-series from 1980 to 2022 Frequency: annual
Methodology	Retropolation by growth rates of *Real GDP* from Benati et al. back to 1892.

Variables	Households and NPISH and Final consumption expenditure
Last Accessed	7.4.2023
Description	Both Series: Unit: Millions of Swiss Francs (at current prices) Structure: Time-series from 1980 to 2022 Frequency: annual
Methodology	The series Households and NPISH has been retropolated to 1948 by growth rates of *Nominal Consumption* from Benati et al. The series Final consumption expenditure has been added to the data collection.

Variables	Households and NPISH and Final consumption expenditure
Last Accessed	7.4.2023
Description	Both Series: Unit: Millions of Chained 2015 Swiss Francs Structure: Time-series from 1980 to 2022 Frequency: annual
Methodology	The series Households and NPISH has been retropolated to 1892 by growth rates of *Real Consumption* from Benati et al. The series Final consumption expenditure has been added to the data collection.

A.6.2 Price Index

A series for the CPI of Switzerland is available at the Swiss National Bank (SNB). Note that the aggregated index does not amount to 100 in 2015 because the reference observation is the month July.

Variable	Consumer prices - total National index - July 2015 = 100
Last Accessed	6.4.2023
Description	Unit: Index July 2015=100 Structure: Time-series from Jan 1922 to Mar 2023 Frequency: monthly
Methodology	Temporal aggregation to annual frequency by using the arithmetic mean of the months within a year. The growth rates (CPI inflation rate) of the new aggregated series is identical to the growth rates of the old series from Benati et al. for available observations.

A.6.3 Unemployment Rate

A series for the jobless rate is available at the SNB. For validation purposes, a series for Switzerland's unemployment rate from the IMF has been taken.

Variables	Total Jobless Rate [SNB] and Unemployment Rate [IMF]
Last Accessed	6.4.2023
Description	Both Series: Unit: Percentage Jobless Rate [SNB]: Structure: Time-series from Jan 1948 to Mar 2023 Frequency: monthly Unemployment Rate [IMF]: Structure: Time-series from 1980 to 2022 Frequency: annual
Methodology	To temporally aggregate the Jobless Rate from monthly to annual figures, the arithmetic mean of monthly values has been used. A comparison of SNB's (aggregated) and IMF's series show that the available observations are identical.

A.6.4 Interest Rates

A variety of short and long-term interest rates are available at the SNB. Series for yields on government bonds with maturities of 1, 2, 5, 10, 20 and 30 years have been included in the data collection.

Variables	Schweiz - SARON - 1 Tag [SARON], Schweiz - 1-Tages-Geld (Tomorrow next) - 1 Tag [1TGT] and Schweiz - Eidg. Geldmarktbuchforderungen - 3 Monate [EG3M]
Last Accessed	6.4.2023
Link	link
Description	All Series: Unit: Percentage (end of month rate) Frequency: monthly SARON: Structure: Time-series from Jan 1999 to Mar 2023 1TGT: Structure: Time-series from Jan 1972 to Mar 2023 EG3M: Structure: Time-series from Jan 1992 to Mar 2023
Methodology	All three series have been temporally aggregated by the arithmetic mean.

Variables	Spot interest rates on Swiss confederation bond issues for selected maturities 1J, 2J, 5J, 10J, 20J and 30J
Last Accessed	6.4.2023
Link	link
Description	All Series: Unit: Percentage (end of month rate) Frequency: monthly 1J, 2J, 5J, 10J and 20J: Structure: Time-series from Jan 1988 to Mar 2023 30J: Structure: Time-series from Jan 1997 to Mar 2023
Methodology	All six series have been temporally aggregated by the arithmetic mean.

A.6.5 Monetary Aggregates

Monthly series for Monetary Base, M1, M2 and M3 are available at SNB's data portal. Similar to *Nominal GDP*, Benati et al. provide a long-run series of the M1, whereby the values for the period 1911-1913 are missing. These values have been imputed by the LWMA algorithm.

Variable	Monetary Base
Last Accessed	6.4.2023
Link	link
Description	Unit: Billions of Swiss Francs Structure: Time-series from Jan 1950 to Feb 2023 Frequency: monthly
Methodology	Monthly series has been converted to millions of Swiss Francs (x1000) prior to temporal aggregation by using the arithmetic mean. The series has then been linked to *M0* from Benati et al.

Variables	Currency in circulation - B,B Monetary aggregate M1 - B,GM1 Monetary aggregate M2 - B,GM2 Monetary aggregate M3 - B,GM3
Last Accessed	6.4.2023
Link	link
Description	All Series: Unit: Millions of Swiss Francs Structure: Time-series from Dec 1984 to Feb 2023 Frequency: monthly
Methodology	All four series have been temporally aggregated using the arithmetic mean, which is the same method Benati et al. implemented. The new M1 and M3 have been linked to the LWMA-imputed *M1* (parameter arguments set to $m = n = 3$) and *M3* from Benati et al., respectively.

A.7 United Kingdom

A thorough data set containing a large number of variables named *Three Centuries of Macroeconomic Data* has been provided by the Bank of England (BOE) (link to three-centuries_v2.3.xlsx). This data set has been used to construct updated long-run series.

A.7.1 National Accounts

All series for national accounts are available at ONS. Variables from the *Three Centuries of Macroeconomic Data* have been used for splicing (description for spliced long-run series are included in the header of the data collection). A series of growth rates for the GDP deflator is also available at ONS.

Variable	YBHA Gross Domestic Product at market prices - Current prices - Seasonally adjusted £m
Last Accessed	24.2.2023
Link	link
Description	Unit: Millions of Pounds Sterling Structure: Time-series from 1948 to 2022 Frequency: annual
Methodology	Retropolation has been done by using the growth rates of *Nominal UK GDP* from Benati et al. A retropolated series of YBHA back until 1700 is also included in the data collection.

Variables	ABMI Gross Domestic Product: chained volume measures - Seasonally adjusted £m and A2.Real GDP(A) in: Three Centuries of Macroeconomic Data
Last Accessed	24.2.2023
Link	link
Description	Both Series: Frequency: annual A2.Real GDP(A): Unit: Millions of Chained 2013 Pounds Sterling Structure: Time-series from 1700 to 2015 ABMI: Unit: Millions of Chained 2019 Pounds Sterling Structure: Time-series from 1948 to 2022
Methodology	Retropolation of ABMI by growth rates of A2.Real GDP(A).

Variables	GDP Deflator [ONS] - Year on Year growth - seasonally adjusted - Series ID: MNF3 and GDP deflator at market prices [threecenturies_v2.3] - 2013=100 in: A1. Headline series
Last Accessed	25.2.2023
Link	link to GDP Deflator [ONS]
Description	Both Series: Frequency: annual GDP deflator [threecenturies_v2.3]: Unit: Index 2013=100 Structure: Time-series from 1700 to 2016 GDP Deflator [ONS]: Unit: Percent Structure: Time-series from 1949 to 2022
Methodology	Deflator index from *Three Centuries of Macroeconomic Data* has been continued by the deflator growth rates provided by ONS. The growth rates of the series from threecent_v2.3 are very similar to the year-on-year change series from ONS.

Variable	ABPB / NAT0 - Final Consumption Expenditure in: Table 0a: UK national and domestic household final consumption expenditure, current prices, annual data, non-seasonally adjusted, £ millions
Last Accessed	13.4.2023
Link	link
Description	Unit: Millions of Pounds Sterling Structure: Time-series from 1997 to 2021 Frequency: annual
Methodology	New series has been linked to *Nominal UK consumption at market prices* from Benati et al.

Variable	ABJR - Household final consumption expenditure - National concept Chain Volume Measure - seasonally adjusted - £m
Last Accessed	13.4.2023
Link	link
Description	Unit: Millions of Chained 2019 Pounds Sterling Structure: Time-series from 1948 to 2022 Frequency: annual
Methodology	New series has been retropolated by growth rates of *Real consumption* from Benati et al.

A.7.2 Price Index

A series for the CPIH and a long-run series beginning in 1800 for the RPI are provided by ONS.

Variable	Consumer Prices Index including owner occupiers' housing costs (CPIH) - Overall Index - K02000001
Last Accessed	14.5.2023
Link	link
Description	Unit: Index Jul 2015=100 Structure: Time-series from Jan 1968 to Feb 2023 Frequency: monthly
Methodology	Temporal aggregation by using the arithmetic mean.

Variable	RPI - Retail Price Index/Composite Price Index - CDKO
Last Accessed	27.2.2023
Link	link
Description	Unit: Index Jan 1974=100 Structure: Time-series from 1800 to 2022 Frequency: annual
Methodology	A series of the RPI with base Jan 1974=100 and base 2015=100 have been added to the data collection.
Limitations	Reliability before 1947 cannot be assured. ONS states: "[...] The table does not fall within the scope of ONS due to the limitations of some of the primary sources, particularly pre-1947, used to construct the index."

A.7.3 Unemployment Rate

Variable	Claimant Count - Seasonally Adjusted Percentage - BCJE
Source	ONS
Last Accessed	27.2.2023
Link	link
Description	Unit: Percentage Structure: Time-series from 1971 to 2022 Frequency: annual
Methodology	Linked to *Unemploymen Rate* from Benati et al.

A.7.4 Interest Rates

An annual series for short-term interest rates has been retrieved from OECD and a monthly series for long-term interest rates from the BOE.

Variable	Short-term interest rate
Last Accessed	1.3.2023
Link	link
Description	Unit: Percentage Structure: Time-series from 1986 to 2022 Frequency: annual

Variable	Yield from British Government Securities 10 year Nominal Zero Coupon - IUMAMNZC (monthly average)
Last Accessed	28.2.2023
Link	link
Description	Unit: Percentage Structure: Time-series from Jan 1982 to Jan 2023 Frequency: monthly
Methodology	The series has been temporally aggregated by the geometric mean. The interest rates have first been converted from percentages to decimal numbers in the form of $1 + r_t$, and after aggregation back to percentages. Then, the aggregated series has been linked to *10 year/medium-term government bond* from Benati et al.

A.7.5 Monetary Aggregates

Series for Monetary Base, M1, M2, and M4 are available at the BOE's database (link) in both seasonally adjusted and non-seasonally adjusted terms. The M4 is only available in seasonally adjusted terms. A series for the discontinued M0 of England until 2016 is available at Fred St. Louis Fed. Variables from the *Three Centuries of Macroeconomic Data* have been used for linking to obtain updated long-run series.

Variable	M0 - MBM0UKM Monetary Base M0 in the United Kingdom, Millions of British Pounds, Annual, Not Seasonally Adjusted
Last Accessed	24.2.2023
Link	link
Description	Unit: Millions of British Pounds Sterling (year-end stock) Structure: Time-series from 1833 to 2016 Frequency: annual
Notes	Series originally stems from the data set *A Millennium of Macroeconomic Data for the UK*. The M0 is similar to Monetary Base from BOE until 2005.

Variables	Monetary Base - Notes and Coins in Circulation - LPMAVAA Monthly average amount outstanding of total sterling notes and coin in circulation, excluding backing assets for commercial banknote issue in Scotland and Northern Ireland total (in sterling millions), not seasonally adjusted in: Table A1.1.1 (Money and lending) M0 - A1.Headline series in: Three Centuries of Macroeconomic Data
Last Accessed	13.4.2023
Link	link for Monetary Base
Description	Both Series: Unit: Millions of Pounds Sterling M0: Structure: Time-series from 1790 to 2015 Frequency: annual Monetary Base: Structure: Time-series from Jun 1969 to Mar 2023 Frequency: monthly
Methodology	The series has been temporally aggregated by selecting the Dec observation. The aggregated series has then been linked to *M0* from *Three Centuries of Macroeconomic Data* from the BOE.
Notes	Seasonally adjusted series are also included in the data collection.

Variables	M1 - LPMVWYE M2 - LPMVWYH M3 - LPMVWXL Monthly amounts outstanding of monetary financial institutions' sterling and all foreign currency M1,M2,M3 (UK estimate of EMU aggregate) liabilities to private and public sectors (in sterling millions), not seasonally adjusted in: Table A2.3
Last Accessed	25.2.2023
Link	link
Description	All Series: Unit: Millions of Pounds Sterling Frequency: monthly M1 and M2: Structure: Time-series from Sep 1986 to Dec 2022 M3: Structure: Time-series Jan from 1987 to Dec 2022
Methodology	All series have been temporally aggregated by selecting the Dec observation.
Notes	Seasonally adjusted series are also included in the data collection.

Variables	M4 - LPMAUYN Monthly amounts outstanding of M4 (monetary financial institutions' sterling M4 liabilities to private sector) (in sterling millions) seasonally adjusted M4 Broad Money - A1.Headline series in: Three Centuries of Macroeconomic Data
Last Accessed	14.5.2023
Link	link to M4 from BOE
Description	Structure: Time-series from Jun 1982 to Dec 2022 Frequency: monthly Both Series: Unit: Millions of Pounds Sterling M4 Broad Money: Structure: Time-series from 1844 to 2015 Frequency: annual M4: Structure: Time-series from Jun 1969 to Mar 2023 Frequency: monthly
Methodology	The series has been temporally aggregated by selecting the Dec observation. The aggregated series has then been linked to *M4 Broad Money* from *Three Centuries of Macroeconomic Data* from the BOE.

A.8 United States

Most series provided by established statistical agencies for US macroeconomic data have been spliced or linked using the long-run series from Benati et al.

A.8.1 National Accounts

Nominal and real GDP have been taken from the BEA (link), nominal and real Private Consumption Expenditure (PCE) from Fred St. Louis Fed (link).

Variable	GDP - Gross Domestic Product, Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate
Last Accessed	8.3.2023
Description	Unit: Billions of USD Structure: Time-series from Q1 1947 to Q4 2022 Frequency: quarterly
Methodology	Temporal aggregation to yearly observations by using the arithmetic mean, since quarterly figures are expressed as annualized levels. GDP has been linked to *Nominal GDP* from Benati et al., as values in the overlapping period are identical.

Variable	GDPC1 - Real Gross Domestic Product, Billions of Chained 2012 Dollars, Quarterly, Seasonally Adjusted Annual Rate
Last Accessed	8.3.2023
Description	Unit: Billions of Chained 2012 USD Structure: Time-series from Q1 1947 to Q4 2022 Frequency: quarterly
Methodology	Temporal aggregation to yearly observations by using the arithmetic mean, since quarterly figures are expressed as annualized levels. Retropolation has been applied to construct a long-run series for the real GDP by using the growth rates of *Real GDP* from Benati et al. (Chained 2009 USD).

Variable	PCE - Personal Consumption Expenditures, Billions of Dollars, Monthly, Seasonally Adjusted Annual Rate
Last Accessed	12.12.2023
Description	Unit: Billions of USD Structure: Time-series from Jan 1959 to Dec 2022 Frequency: monthly
Methodology	Temporal aggregation to yearly observations by using the arithmetic mean, since monthly figures are expressed as annualized levels. The new series fro nominal PCE has been linked to *Nominal Personal consumption expenditure* from Benati et al.

Variables	PCECCA and PCEC96 - Real Personal Consumption Expenditures, Billions of Chained 2012 Dollars, Annual, Not Seasonally Adjusted
Last Accessed	14.5.2023
Description	Both Series: Unit: Billions of Chained 2012 USD Frequency: annual PCECCA: Structure: Time-series from 1929 to 2021 PCEC96: Structure: Time-series from 2002 to 2022
Methodology	The linked series added to the data collection is comprised as follows: from 1929 its PCECCA to 2001, from 2002 to 2022, its PCEC96. A linking of the two series can be justified because their values in the entire overlapping period only differ at the second decimal place. The linked series has been further retropolated in the past by using the growth rates of *Real PCE* from Benati et al.

A.8.2 Price Index

A series for the CPI of the US covering over a century is available at the BLS (link).

Variable	CPI for All Urban Consumers (CPI-U) - CUUR0000SA0
Last Accessed	14.5.2023
Description	Unit: Index 1982-84=100 Structure: Time-series from 1913 to 2023 Frequency: annual

A.8.3 Unemployment Rate

No series for the unemployment rate could be identified, thus, a series for the unemployment rate has been constructed by using the series for the number of unemployed and the series for number of people in the labor force accessed via Python-API (Python-Notebook to access data) provided by BLS.

Variables	Unemployment Level - LNS13000000 Civilian Labor Force - LNS11000000
Last Accessed	8.12.2022
Description	Both Series: Unit: Number of people in thousands (Seasonally Adjusted) Structure: Time-series from 1948 to 2022 Frequency: monthly
Methodology	Temporal aggregation of both series has been done by using the arithmetic mean of monthly observations. The unemployment rate has been derived from the two series as follows: $$\text{Unemplyoment Rate}\,(\%) = \frac{\text{Unemployment Level}}{\text{Civilian Labor Force}} \times 100$$ Despite the detour, the computed values are identical to the available observations of the series provided by Benati et al.

A.8.4 Interest Rates

Variable	3-Month Treasury Bill Secondary Market Rate - RIFSGFSM03NA
Source	Fred St. Louis Fed
Last Accessed	8.3.2023
Link	link
Description	Unit: Percentage Structure: Time-series from 1954 to 2022 Frequency: annual
Methodology	Linked to *US Treasury Bill* from Benati et al. as values are identical in the overlap period.

Variable	Moody's Seasoned Baa Corporate Bond Yield, Percent, Monthly, Not Seasonally Adjusted
Source	Fred St. Louis Fed
Last Accessed	9.3.2022
Link	link
Description	Unit: Percentage Structure: Time-series from Jan 1919 to Feb 2023 Frequency: monthly
Methodology	Temporal aggregation by taking average of monthly observations. Linked to *Yield on corporate bonds* from Benati et al. as values are identical in the overlap period.
Notes	According to Fred St. Louis Fed: "These instruments are based on bonds with maturities 20 years and above."

A.8.5 Monetary Aggregates

Variables	M1 Money Stock (Seasonally Adjusted) - H6/H6_M1/M1.M and M2 Money Stock (Seasonally Adjusted) - H6/H6_M2/M2.M in: H.6 Money Stock Measures
Source	Board of Governors the Federal Reserve System
Last Accessed	8.12.2022
Link	link
Description	Both Series: Unit: Billions of USD Structure: Time-series from 1959 to 2022 Frequency: monthly
Methodology	Both Series have been temporally aggregated by using the arithmetic mean to derive the average annual stock. Furthermore, both new series have been linked to *M1* and *M2* of Benati et al., respectively.

Variables	M3 for the United States - MABMM301USM189S [stock] and M3 for the United States - MABMM301USA657S [growth rates]
Source	Fred St. Louis Fed
Last Accessed	25.2.2023
Link	link
Description	M3 [stock]: Unit: Billions of USD Structure: Time-series from Jan 1960 to Jan 2023 Frequency: monthly M3 [growth rates]: Unit: Percentage Structure: Time-series from 1960 to 2022 Frequency: monthly
Methodology	The monthly M3 [stock] series has been temporally aggregated by using the arithmetic mean to derive the average annual stock. The 1960 value of the series M3 [growth rates] has been used to calculate the 1959 value for the M3: $$M3_{1959} = \frac{M3_{1960}}{1 + r_{1960}/100}$$

B Program Code

B.1 Retropolation

```
# function takes in the first observation of the new series at T
# and a list of growth rates (in percent) for the period of T-m
def retropolation(endval:float, grates:list):
  l = []
  for x in reversed(grates):
    l.append(endval)
    endval = endval/(x/100+1)
  return l[::-1]
```

B.2 Linear Weighted Moving Average

```
import numpy as np
import pandas as pd

def lwma(series, m=2, n=2):

  # check if series argument is an instance of pd.Series
  # copy to avoid modifying the input time series
  if type(series) != type(pd.Series):
    copy = pd.Series(series).copy()

  else:
    copy = series.copy()

  # generating the indices of the time series
  all = pd.Series(range(len(copy)))

  # replacing NaN values by 0
  copy.fillna(0, inplace=True)

  # boolean mask whether value is missing or not
  zero = float(0)
  zero_mask = copy.eq(zero)

  # getting the indices of the missing values using the mask
  zero_idx = all.index[zero_mask]

  # generating the weights
```

```
# highest weights for obs close to the target obs
# weights decrease linearly
w_m = pd.Series(range(1, m+1)[::-1])
w_n = pd.Series(range(1, n+1)[::-1])

# indices for how many obs before and after
# are taken into consideration
mback = pd.Series(range(1, m+1))
nforw = pd.Series(range(1, n+1))

# select obs in the current window
for t in zero_idx:
  back_m = copy.iloc[t - mback]
  forw_n = copy.iloc[t + nforw]

  # m values back in time times their weights
  bwd = []
  for i in range(m):
    bwd.append(back_m.iloc[i] * w_m.iloc[i])

  # n values forward in time times their weights
  fwd = []
  for j in range(n):
    fwd.append(forw_n.iloc[j] * w_n.iloc[j])

  # boolean, whether values backward or forward are 0 or not
  # if 0 -> they are missing in the data (NaN values)
  L_i = [1 if i > 0 else 0 for i in bwd]
  L_j = [1 if j > 0 else 0 for j in fwd]

  # denominator: comprised of the sum of weights, minus the weights
  # that are not considered if values of fwd or bwd are 0
  denom = 0
  for i in range(m):
    denom += w_m.iloc[i] * L_i[i]
  for j in range(n):
    denom += w_n.iloc[j] * L_j[j]

  missing_val = 0
  for i in range(m):
    missing_val += bwd[i] * L_i[i]
  for j in range(n):
```

```
            missing_val += fwd[j] * L_j[j]

        missing_val /= denom

        copy.iloc[t] = missing_val

    return copy
```