# PART IV

# Interpolating monthly Swiss GDP in a Kalman filter framework

**Abstract**

We estimate deseasonalized monthly series for Swiss gross domestic product (GDP) at constant prices of 1990 for the period 1980-1997. They are consistent with the quarterly figures estimated by the State Secretariat for Economic Affairs and obtained by including information contained in related series, in particular following the expenditure definition of GDP. We present a general approach using the Kalman filter technique nesting a great variety of interpolation setups. We evaluate competing models and provide a time series that can be used by other researchers.

**JEL codes**: E32, E37

**Keywords**: GDP, Interpolation, Kalman filter, National accounting, Switzerland.

---

## 1 Introduction

For economic studies using quarterly data, a low number of observations can cause serious flaws in the quality of quantitative analysis. In vector autoregressions (VAR) with relatively short time series

[0]This part is based on Cuche and Hess (1999a, 1999b) written with Martin K. Hess.

for example, many degrees of freedom are used up in the estimation, reducing drastically its power. Moreover, monthly frequency is sometimes implied by the assumptions of the model to be estimated, while only quarterly data is released[1].

Therefore, economists are sometimes forced to use variables that proxy GDP and that are available at a higher frequency. In many countries, a common proxy is industrial production (IP) which is often recorded at monthly frequency. In Switzerland, it is difficult to find such a monthly indicator for aggregate productive activity. The IP index is a series at a quarterly frequency, and other series like business surveys or filled orders can only be used as GDP proxies with some reservations. Hence, in cases where adequate proxies are not at hand, monthly estimates of GDP by interpolation provide a solution for this problem[2].

Whether to replace a proxy variable by an interpolated one or not depends on the available data series and on the empirical economic model considered. The evaluation of the trade-off between potential benefits and disadvantages of both approaches is beyond the scope of this paper and is omitted. The goal of this paper is to provide a monthly deseasonalized real GDP series for empirical research[3].

Chow and Lin (1971) were the first to present a coherent and easily applicable econometric approach that handles interpolation problems for stock and flow variables. Assuming a linear relation between the series of interest (series for which observations are missing, i.e. monthly GDP) and other data with more frequent recording (related series), they estimate a univariate regression equation. This multiple regression approach is flexible enough to take into account

[1] The official quarterly GDP figures for Switzerland are interpolated and published by the State Secretariat for Economic Affairs. Furthermore, an official annual GDP is calculated by the Federal Statistics Office producing the national income accounts. The quarterly estimates are then corrected and published again to match the official annual statistics.

[2] We define interpolation as a process of computing flow or stock series at a higher frequency than the original one. In this terminology, we do not distinguish between interpolation and distribution which is often done in studies with both, stock and flow variables. Here, we present models that exclusively serve for in-sample interpolations and not for out-of-sample predictions.

[3] In our view, deseasonalized time series are of greater interest as they are handy to use in economic models. To estimate a seasonalized series, the seasonality is separately estimated and then added to the deseasonalized series, as done for example for quarterly GDP estimates by the State Secretariat for Economic Affairs.

heteroscedasticity and low-order autocorrelation in the residuals.

More recent studies make use of the Kalman filter (Bernanke, Gertler and Watson (1997) and Harvey and Pierse (1984)). This dynamic framework is much more flexible, since it is capable of nesting more models than the Chow and Lin framework.

In this paper, the focus is directed on econometric details such as the issue of stationarity and cointegration in different Kalman filter configurations. Recent innovative techniques are analyzed theoretically and then evaluated empirically. An overview of estimated monthly GDP series produced by various model setups is provided. We evaluate different combinations of methods and related series with the aim to get the most appropriate monthly GDP. For this task, several selection criteria as well as a simulated interpolation from annual to quarterly data are used.

Before estimating the model, we evaluate competing related series. We identify the series containing the highest amount of information for the interpolation. The choice criteria for the related monthly series are based on the expenditure definition of GDP and on statistical properties of the comovement with GDP. However, the dearth of Swiss data at higher frequency limits severely the choice of these variables. Therefore, we consider other related series, for example foreign aggregate economic activity, as alternatives for interpolation. In fact, all related series that closely and robustly move together with quarterly GDP could be appropriate series helping to extract monthly GDP. With these related series available, it is then possible to estimate monthly GDP for Switzerland for 1980-1997 in different model setups[4].

The paper is organized as follows. It starts in section 2 with a short survey of the interpolation literature. In section 3, we briefly review the Kalman filter methodology and present the different interpolation models. In section 4, various related series are evaluated and described. We give an overlook of our results in section 5. We then evaluate the appropriateness of these interpolations. Section 6 concludes.

[4] We exclusively concentrate our investigation on the period 1980-1997 because these figures are compatible with the new national accounting system in Switzerland, the European System of Integrated Economic Accounts (ESA) 78. This standard was introduced in Switzerland in 1996, but the State Secretariat for Economic Affairs calculated quarterly GDP figures back to 1980. See Schwaller and Parnisari (1997) for a comprehensive survey.

## 2 Related Literature

As Lanning (1986) illustrates, economists facing missing data have basically two different ways to solve that problem. A first approach is to estimate the missing data simultaneously with the economist's model parameters, thereby considering the missing observations as any other parameter. The second way is a two-step approach where in a first step the missing data, which could be independent from the economist's model, are interpolated. In a second step, the new augmented series are used to estimate the economist's model. Lanning found that the simultaneous approach yields estimates of the economist's model parameters that have a greater variance, and thus are less reliable, than the model parameters estimated with complete data in the second stage. Based on these empirical findings, he suggests using the two-step approach. Related literature about this procedure can be subdivided into the following three classes[5].

First, the seminal approach for the use of the univariate multiple regression technique with related series was presented by Chow and Lin (1971, 1976) in a unified framework which allows treating the interpolation of stocks and flows variables. This approach was able to overcome the problems faced by Friedman (1962) who treated stocks and flows in different ways. Specifically, they could deal with the requirement that if an observed flow value is distributed among the corresponding subintervals, the higher frequency estimates must add up to the original flow variable. Until now, this univariate regression approach has been widely used for interpolation due to its easier implementation than the state-space approach. This argument seems to more than just outweigh the potential advantages of more sophisticated procedures like the Kalman filter. An annual GDP is for example interpolated for Mexico by De Alba (1990). Schmidt (1986) gives a survey of this method, interpolating personal income of US regions.

Second, still with related series, Denton (1971), Fernandez (1981), and Litterman (1983) proposed an approach that minimizes a weighted quadratic loss function on the difference between the series to be estimated and a linear combination of the observed related series.

[5] The signal extraction literature is very vast and difficult to objectively classify. Here, we only review the interpolation literature, without considering general approaches such as the problem of unobserved components in economic time series or the estimation of irregularly missing data.

This strategy is related to the Chow and Lin regression and allows for the use of data in first difference. An illustration with Portuguese data is given in Pinheiro and Coimbra (1993).

Third, Bernanke, Gertler and Watson (1997) have recently used a state-space model to interpolate real GDP in the US. Their approach is to first estimate monthly components of nominal GDP plus the GDP deflator and then to aggregate the individual estimates. The methodology they followed was suggested by Harvey and Pierse (1984) who provided a general framework - state-space formulations for stock and flow variables, for stationary and nonstationary series, and with or without related series - to estimate missing observations in economic time series. Solving such state-space models requires the use of the Kalman filter. A Kalman filter interpolation is done for Canadian GDP by Guay, Milbourne, Otto and Smith (1990).

Hereafter, we present a state-space framework introduced by Harvey and Pierse (1984). This general formulation allows us to rewrite all three classes of models as well as much simpler models that do not use related series.

# 3 Models

## 3.1 Kalman Filter

A useful method for extracting signals is to write down a model linking the unobserved and observed variables in a state-space representation according to Kalman (1960, 1963). The multivariate Kalman filter is an algorithm for sequentially updating a linear projection on the vector of interest. A general review is given here and a more detailed description in the appendix[6]. We present various configurations of the state-space system in the next section on interpolation models.

The state-space representation is given by a system of two vector equations. First, the state or transition equation describes the dynamics of the state vector $\boldsymbol{\xi}_t$ containing the unobserved variables we estimate. The second type of equation represents the observation or measurement equation linking the state vector to the vector con-

[6] Very detailed descriptions of the Kalman filter technique can be found in the Handbook of Econometrics by Hamilton (1994a) and in his textbook (Hamilton, 1994b). See Aoki and Havenner (1991), Gourieroux and Monfort (1997), Harvey (1989), and Lütkepohl (1993) for further useful contributions.

taining the observed variables $\mathbf{y}_t^+$. The equations of this system for $t = 1, \ldots, T$ where $T$ is the number of monthly observations are the following:

$$\boldsymbol{\xi}_{t+1} = \mathbf{F}_t \boldsymbol{\xi}_t + \mathbf{C}_t' \mathbf{x}_{t+1} + \mathbf{R}_t \mathbf{u}_{t+1}, \tag{1}$$
$$\mathbf{y}_t^+ = \mathbf{A}_t' \mathbf{x}_t^* + \mathbf{H}_t' \boldsymbol{\xi}_t + \mathbf{N}_t \mathbf{v}_t. \tag{2}$$

In addition to the unobserved and observed variables of interest, vector equations (1) and (2) contain the so-called related series ($\mathbf{x}_t$ and $\mathbf{x}_t^*$) as exogenous variables in each equation. Both equations have multinormally distributed error terms: $\begin{pmatrix} \mathbf{u}_t \\ \mathbf{v}_t \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} \end{pmatrix}\right)$. Premultiplied by matrices $\mathbf{R}_t$ and $\mathbf{N}_t$, these orthogonal disturbances transform into nonorthogonal residuals within each vector equation. The coefficient matrices $\mathbf{F}_t$, $\mathbf{C}_t'$, $\mathbf{R}_t$, $\mathbf{A}_t'$, $\mathbf{H}_t'$, $\mathbf{N}_t$, and the two variance-covariance matrices $\mathbf{Q}$ and $\mathbf{G}$ are estimated by maximizing the log-likelihood function of this system.

## 3.2 Interpolation Models

### 3.2.1 Overview

In this section, we adapt the general state-space representation (1) and (2) to our problem in different ways, specifically the inclusion of related series and assumed stochastic processes for monthly GDP[7]. The interpolation framework[8] for $t = 1, \ldots, T$ becomes:

$$\boldsymbol{\xi}_{t+1} = \mathbf{F} \boldsymbol{\xi}_t + \mathbf{C}' \mathbf{x}_{t+1} + \mathbf{R} \mathbf{u}_{t+1}, \tag{3}$$
$$y_t^+ = \mathbf{a}_t' \mathbf{x}_t^* + \mathbf{h}_t' \boldsymbol{\xi}_t. \tag{4}$$

The state vector equation (3) describes the vector dynamics of the unobserved variable, monthly GDP $y_t$, stacked in the state vector $\boldsymbol{\xi}_t = \begin{pmatrix} y_t & y_{t-1} & y_{t-2} \end{pmatrix}'$. The $[3 \times 1]$ dimension serves to take the three months within a quarter together in order to satisfy the

[7] The Kalman filter algorithm and the log-likelihood function are displayed in Appendix A.

[8] In all the models, quarterly GDP $y_t^+$ is given each month, $y_1^+ = 0$, $y_2^+ = 0$, $y_3^+ =$ first quarterly value, $y_4^+ = 0$, $y_5^+ = 0$, $y_6^+ =$ second quarterly value, etc. Note that we observe $\frac{T}{3}$ quarterly values for $T$ months to interpolate. Contrary to the convention when stacking quarterly observations to a column vector $\mathbf{y}^+$, we do not include zero observations resulting thus in a $\left[\frac{T}{3} \times 1\right]$ vector.

sum-up constraint[9]. The exact formulation of this state vector equation is difficult, because there is no prior knowledge about the true process driving monthly GDP. In order to shed light on this issue, we compare results of various competing setups in section 5. We assume time-invariant coefficients for the matrices $\mathbf{F}$, $\mathbf{C}'$, and $\mathbf{R}$.

Equation (4) relates the state vector to the observed quarterly GDP $y_t^+$. Following Harvey and Pierse (1984), this observation equation represents the constraint that the sum of three monthly observations within a quarter must equal the quarterly observed GDP. Hence, this equality constraint implies that the error term $\mathbf{N}_t\mathbf{v}_t$ disappears from the observation equation. The sum-up constraint is introduced by coefficient vector $\mathbf{a}_t'$ and $\mathbf{h}_t'$, depending on the models presented in the following section[10].

All the specifications of the state-space models described hereafter correspond to different assumptions depending on whether related series ($\mathbf{x}_t$ or $\mathbf{x}_t^*$) are used or not and on the characteristics of the data to interpolate (stochastic process and stationarity). The properties of the data such as the order of integration and the assumed stochastic driving process of monthly GDP influence the representation of the state equation. Possible related series influence the setup of the state vector equation and the observation equation to in turn affect the coefficients contained in $\mathbf{C}'$ and $\mathbf{a}_t'$. We add related series in order to evaluate their statistical relevance. The selected assumptions are also guided by simplicity and technical considerations about the construction of the Kalman filter.

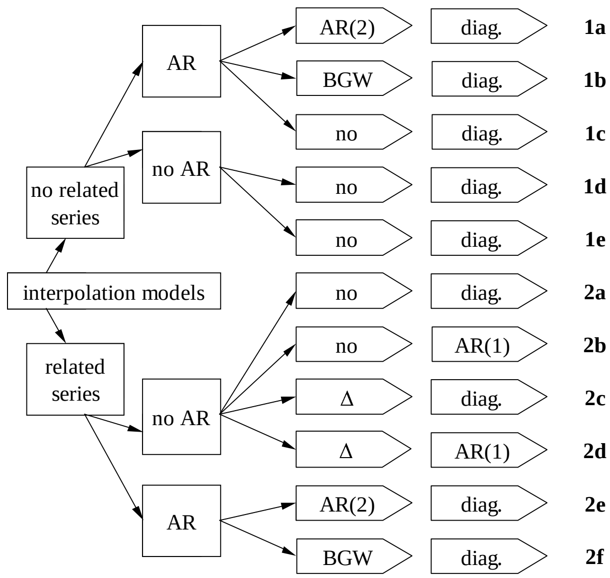
Hence, we focus on two broad classes of Kalman filter models summarized in figure 1. The first class of models is designed without related series. We assume that there is enough information in the autocovariance function of the quarterly series and in the assumed low-order autoregressive (AR) process of monthly GDP. Moreover, we combine this assumption with alternative ways to treat nonsta-

[9] We could imagine a $[12 \times 1]$ vector, implying an interpolation from annual to monthly data. This procedure would have the advantage to use probably more accurate annual data as a basis for the interpolation. On the other hand, it would give the Kalman filter too much freedom in the adjustment to match the annual observations.

[10] Not treated in this text is the introduction of the sum-up constraint by augmenting the state-space representation with a 'cumulator function' $y_t^c$ which accumulates monthly GDP observations in a given quarter: $y_t^c = \sum_{s=0}^{r} y_{t-s}$ where $r = 0$ for $t = 1, 4, 7, \ldots, T-2$, $r = 1$ for $t = 2, 5, 8, \ldots, T-1$, and $r = 2$ for $t = 3, 6, 9, \ldots, T$. See Harvey (1989) for more details and Kobler (1999) for an application to Swiss GDP.

tionary series (models **1a-c**). Contrasting with these AR models, there are two 'naive' models that neither follow an AR process nor include related series.

Figure 1: Overview of Interpolation Models




*Note*: First arrow column displays correction of nonstationarity. Second arrow column concerns the residuals form. Last column displays model numbers. AR(2) stands for an AR(1) process in first difference rewritten as an AR(2) in level; BGW means correction according to Bernanke, Gertler, and Watson (1997); Δ uses a first difference operator; in absence of correction done by the model, we mention 'no'; 'diag.' indicates no autocorrelation in the residuals; AR(1) stands for residuals following an AR(1) process.

However, it is not necessary to run the Kalman filter, because simple calculus produces the same results. For each quarter, model **1d** returns three equal monthly values, namely the third of the corresponding quarterly observation. This obviously indicates that the Kalman filter corrects the estimates at each quarter taking into account a possible trend in the data. Model **1e** gives for each quarter three monthly GDP following a quarterly linear trend centered around the monthly mean of the quarter[11]. We take model **1e** as our

[11] Model **1d** needs a constant term as explanatory variable in order to calculate the third of the quarterly observation. Model **1e** interpolates monthly observations linearly within a quarter, where we assume that we can split each quarter

benchmark because of its intuitive setup. The second class of models introduces related series in order to extract information for the interpolation of monthly GDP. Within this group, we distinguish between the assumptions that monthly GDP does not follow an AR process (models **2a-d**) and that it does (models **2e-f**). We further enrich this second class of models with different ways to treat nonstationarity and with different assumptions about monthly residuals.

In the next paragraphs, we show the various models **1a–c** and **2a-f** in detail.

### 3.2.2 Models without Related Series

**3.2.2.1 Model 1a** In our first model, we assume that the first difference of monthly GDP follows a stationary AR(1) process $\Delta y_t = \phi \Delta y_{t-1} + u_t$ in order to treat nonstationarity. $\Delta y_t$ is the first difference of monthly GDP, $\phi$ is a coefficient constrained to lie inside the unit circle, and $u_t$ is a iid error term with distribution $N\left(0, \sigma_u^2\right)$. In order to find a starting value for the GDP series, it is imperative to write this AR(1) as an AR(2) of the series in level[12]:

$$y_t = (1+\phi)\, y_{t-1} - \phi y_{t-2} + u_t. \tag{5}$$

This equation written in companion form yields the state equation (6) for $t = 1, \ldots, T$ where $\boldsymbol{\xi}_t = \begin{pmatrix} y_t & y_{t-1} & y_{t-2} \end{pmatrix}'$.

$$\begin{pmatrix} y_{t+1} \\ y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} 1+\phi & -\phi & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} u_{t+1} \\ u_t \\ u_{t-1} \end{pmatrix} \tag{6}$$

Note that this formulation sets $\mathbf{C}' = \mathbf{0}$ in equation (3). The observation equation incorporates the sum-up constraint leaving out the

(except the first one) into an initial value $y_{t-3}$ which is the last month of the previous quarter and a step $d_t$ for $t = 4, 5, \ldots, T$ according to the following equation: $(\varphi y_{t-3} + d_t) + (\varphi y_{t-3} + 2d_t) + (\varphi y_{t-3} + 3d_t) = y_t^+$. As quarterly GDP $y_t^+$, the step $d_t$ is given each month, $d_4 = 0$, $d_5 = 0$, and $d_6 =$ monthly step of second quarter, etc. $\varphi$ is a scalar that takes on 1 for $t = 6, 9, \ldots, T$ and 0 for $t = 4, 5, 7, \ldots, T-1$.

[12] This 'transformation' yields the same likelihood and the same estimator for $\phi$ as the original equation in first difference. However, this form has the characteristic to produce a system with explosive eigenvalues.

related series. This implies $\mathbf{a}_t' = \mathbf{0}$. $\mathbf{h}_t'$ takes on two different values depending on the respective month:

$$\begin{aligned} y_t^+ &= \mathbf{h}_t'\boldsymbol{\xi}_t, \\ \text{where } \mathbf{h}_t' &= \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}, \text{ for } t = 1, 2, 4, 5, 7, \ldots, T-1, \\ \text{where } \mathbf{h}_t' &= \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}, \text{ for } t = 3, 6, 9, \ldots, T. \end{aligned} \tag{7}$$

**3.2.2.2 Model 1b** Recently, an alternative interpolative method was suggested by Bernanke, Gertler and Watson (1997). It consists in using GDP integrated of order one (I(1)) with a cointegrated series $p_t$ such that we compute a new monthly stationary series[13] $y_t^s = \frac{y_t}{p_t}$. $p_t$ is a scaling variable such that $y_t^s$ is nontrending. This approach relies on a calculated multiplicative cointegration that holds at both, monthly and quarterly frequencies. For the dynamic specification of $y_t^s$, now forming the elements of state vector $\boldsymbol{\xi}_t$, we assume the AR(1) process $y_t^s = \phi y_{t-1}^s + u_t$. Compared to model **1a,** $y$ is replaced by $y^s$, and $\begin{pmatrix} \phi & 0 & 0 \end{pmatrix}$ becomes the first row in $\mathbf{F}$. The observation equation is also different because we have to 'neutralize' the division by the I(1) series $p_t$. This is done by constraining the measurement equation in such a way that quarterly values $y_t^+$ are restored in setting $\mathbf{a}_t' = \mathbf{0}$ and redefining vector $\mathbf{h}_t'$:

$$\begin{aligned} y_t^+ &= \mathbf{h}_t'\boldsymbol{\xi}_t, \\ \text{where } \mathbf{h}_t' &= \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}, \text{ for } t = 1, 2, 4, 5, 7, \ldots, T-1, \\ \text{where } \mathbf{h}_t' &= \begin{pmatrix} p_t & p_{t-1} & p_{t-2} \end{pmatrix}, \text{ for } t = 3, 6, 9, \ldots, T. \end{aligned} \tag{8}$$

**3.2.2.3 Model 1c** Finally, Bomhoff (1994) suggests using the series in level arguing that the Kalman filter does not require the user to make a definite decision about the need for differencing the data. The Kalman filter offers automatic processing capacity for a wide range of nonstationary time series[14]. Hence, we write down the law governing the process as if the series were stationary: $y_t = \phi y_{t-1} + u_t$. This model is similar to model **1a** but with a first row of matrix $\mathbf{F}$ defined as $\begin{pmatrix} \phi & 0 & 0 \end{pmatrix}$.

[13] The ratio $y_t^s = \frac{y_t}{p_t}$ is chosen as a general framework that avoids assuming a particular cointegrating vector which is impossible to estimate at the monthly level.

[14] The models without related series and without AR processes (**1d** and **1e**) have already suggested this feature.

### 3.2.3 Models with Related Series and without AR Structure

A main criticism of models **1a-c** is that they extract signals only from the presumed stochastic process of the original series without adding new information. We could speak about 'fool-yourself' models[15] to generate monthly GDP. Therefore, it seems that we would be better off enriching the model with additional information. For this purpose, we now include explanatory series that are related to the series to interpolate.

In all the models, we may introduce the related series either in the measurement equation (4) for the generalized least squares (GLS) estimator (models **2a-d**), or in the state equation (3) for the Kalman filter algorithm (models **2e-f**).

**3.2.3.1 Model 2a-b** Chow and Lin (1971, 1976) show how related series can be used to interpolate lower frequency data in order to get higher frequency data with a GLS estimator. They assume that monthly GDP $y_t$ is described by a simple regression of $y_t$ on $l$ related series $x_t$, in matrix notation $\mathbf{y}_{GLS} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}_{GLS}$, where the variance-covariance of the error term is $\mathbf{V} = E\left(\mathbf{u}_{GLS}\mathbf{u}_{GLS}'\right)$. They also assume the same relationship at the quarterly level: $\mathbf{y}^+ = \mathbf{X}^+\boldsymbol{\beta} + \mathbf{u}^+$, where $\mathbf{X}^+$ is a matrix with quarterly averages of three months of related series, and $\mathbf{V}^+ = E\left(\mathbf{u}^+\mathbf{u}^{+\prime}\right)$ a variance-covariance matrix. $\mathbf{V}^+$ is thus a function of $\mathbf{V}$[16].

The Kalman filter configuration of the Chow and Lin setup, defining vector $\boldsymbol{\xi}_t$ as suggested by Harvey and Pierse (1984), is:

$$\boldsymbol{\xi}_{t+1} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \boldsymbol{\xi}_t + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{u}_{t+1}, \tag{9}$$

$$y_t^+ = \mathbf{a}_t'\mathbf{x}_t^* + \mathbf{h}_t'\boldsymbol{\xi}_t, \tag{10}$$

where $\boldsymbol{\xi}_t = \begin{pmatrix} y_t - \mathbf{x}_t'\mathbf{c} \\ y_{t-1} - \mathbf{x}_{t-1}'\mathbf{c} \\ y_{t-2} - \mathbf{x}_{t-2}'\mathbf{c} \end{pmatrix}$ with $\mathbf{c} = \frac{1}{3}\boldsymbol{\beta}$, $\mathbf{x}_t^* = \sum_{j=t-2}^{t} \mathbf{x}_j$, $\mathbf{h}_t' = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$ and $\mathbf{a}_t' = \mathbf{0}$ for $t = 1, 2, 4, 5, 7, \ldots, T-1$, and $\mathbf{h}_t' =$

[15] Thanks to Mark Watson for bringing up this expression.

[16] In order to use identical coefficients $\boldsymbol{\beta}$ in the monthly and quarterly regressions, Chow and Lin have to use a $[T \times 1]$ vector $\mathbf{y}_{GLS}$ that contains figures three times larger than the monthly estimates of the Kalman filter setup.

$\begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$ and $\mathbf{a}_t' = \mathbf{c}'$ for $t = 3, 6, 9, \ldots, T$. As the related series and their coefficients are contained in the state vector, we set $\mathbf{C}' = \mathbf{0}$ in equation (3) and re-introduce them as $\mathbf{a}_t'\mathbf{x}_t^*$ in the observation equation.

Chow and Lin (1971, 1976) directly calculate a best linear unbiased estimator for the monthly series from the trace minimization of the variance-covariance matrix $V[\hat{\mathbf{y}}_{GLS}]$. Applying a GLS method, they avoid numerical optimization problems contained in the Kalman filter procedure[17,18]. The estimates of monthly GDP are

$$\hat{\mathbf{y}}_{GLS} = \mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\Lambda}(\mathbf{V})\hat{\mathbf{u}}^+ \tag{11}$$

where $\hat{\mathbf{y}}_{GLS}$ denotes the vector of the monthly estimates in matrix notation. This special fitted value has two parts: a traditional fitted value $\mathbf{X}\hat{\boldsymbol{\beta}}$ with the influence of related series and an interpolation-corrected residual term $\boldsymbol{\Lambda}(\mathbf{V})\hat{\mathbf{u}}^+$.

$\hat{\boldsymbol{\beta}}$ is a GLS estimator of the regression between quarterly GDP data $\mathbf{y}^+$ and their 'quarterly' related series $\mathbf{X}^+$:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{+'}\mathbf{V}^{+^{-1}}\mathbf{X}^+\right)^{-1}\mathbf{X}^{+'}\mathbf{V}^{+^{-1}}\mathbf{y}^+. \tag{12}$$

The weighting matrix in this regression is the inverse of the variance-covariance matrix $\mathbf{V}^+$ of the quarterly residuals $\mathbf{u}^+$. Hence, the assumptions about $\mathbf{V}$ directly influence the distribution of $\hat{\boldsymbol{\beta}}$ and the $\left[T \times \frac{T}{3}\right]$ matrix $\boldsymbol{\Lambda}$ for the dissemination of the quarterly residuals over the monthly estimated GDP. These quarterly residuals are crucial for interpolation, because the traditional fitted monthly values

[17] Appendix B describes the Chow and Lin regression, and Appendix C shows that the Kalman filter and the Chow and Lin regression yield the same estimates by maximum likelihood.

[18] An alternative state-space form producing the same results is the following:

$$\begin{aligned}
\begin{pmatrix} y_{t+1} \\ y_t \\ y_{t-1} \end{pmatrix} &= \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} \mathbf{c}' \\ \mathbf{0}' \\ \mathbf{0}' \end{pmatrix} \mathbf{x}_{t+1} \\
&\quad + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} u_{t+1} \\ u_t \\ u_{t-1} \end{pmatrix}, \qquad (9') \\
y_t^+ &= \mathbf{h}_t'\boldsymbol{\xi}_t, \qquad (10')
\end{aligned}$$

where $\mathbf{h}_t' = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$, for $t = 1, 2, 4, 5, 7, \ldots, T-1$, and where $\mathbf{h}_t' = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$, for $t = 3, 6, 9, \ldots, T$.

$\mathbf{X}\hat{\boldsymbol{\beta}}$ do not sum up to quarterly observations. Therefore, the residuals $\mathbf{u}^+$ must be 'redistributed' to the monthly GDP values according to the weighting matrix $\boldsymbol{\Lambda}$ to correct this shortcoming.

Models **2a** and **2b** differ in $\mathbf{V}$. In model **2a,** we assume that the variance-covariance of the monthly residuals $\mathbf{V}$ is a simple diagonal matrix $\sigma^2_{u_{GLS}}\mathbf{I}_T$. It implies that $\mathbf{V}^+$, the variance-covariance matrix of quarterly residuals is equal to $\frac{\sigma^2_{u_{GLS}}}{3}\mathbf{I}_{\frac{T}{3}}$, and $\boldsymbol{\Lambda}$ is equal to $\mathbf{I}_{\frac{T}{3}} \otimes \mathbf{i}_3$.

However, the diagonal variance-covariance matrix $\mathbf{V}$ is rarely supported by the data. A way to improve this setting is to allow for serial correlation in the error term. Hence, we assume for the model **2b** that the error term follows an AR(1), $u_{GLS_t} = \varrho u_{GLS_{t-1}} + \varepsilon_t$ where $\varepsilon_t$ is a white noise, yielding a variance-covariance matrix:

$$\mathbf{V} = \begin{pmatrix} 1 & \varrho & \varrho^2 & \dots & \varrho^{T-1} \\ \varrho & 1 & \varrho & & \\ \varrho^2 & \varrho & 1 & & \\ \vdots & & & \ddots & \vdots \\ \varrho^{T-1} & & & \dots & 1 \end{pmatrix} \frac{\sigma^2_\varepsilon}{1-\varrho^2}. \tag{13}$$

This specification introduces a different $\hat{\boldsymbol{\beta}}$ and a new redistribution matrix $\boldsymbol{\Lambda}$ depending now on $\varrho$. The more $\varrho$ tends to zero, the more the $\boldsymbol{\Lambda}$ matrix converges to $\mathbf{I}_{\frac{T}{3}} \otimes \mathbf{i}_3$. Hence, if the serial autocorrelation is significant, redistribution is less 'rigid' than in model **2a**, and the quarterly residuals are not only spread out over their corresponding months but also influence monthly GDP of surrounding quarters in a 'smoother' way.

**3.2.3.2 Model 2c-d** A variation of models **2a-b**, as suggested by Denton (1971), Fernandez (1981), and Litterman (1983) is to use first differences of time series in the regression instead of levels in order to account for nonstationarity, specifically not treated in the Chow and Lin framework. In model **2c,** they assume that the variance-covariance of the error term $\mathbf{V}$ is $\sigma^2_{u_{GLS}}\mathbf{I}_T$ and in model **2d**, that the error terms follow an AR(1) with a variance-covariance matrix $\mathbf{V}$ equal to matrix (13). They suggest estimating $\hat{\boldsymbol{\beta}}$ with a weighting matrix equal to the inverse of the quarterly equivalent of

$(\mathbf{D}'\mathbf{V}'\mathbf{D})^{-1}$ where

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 & \dots \\ -1 & 1 & 0 & \\ 0 & -1 & 1 & \\ \vdots & & & \ddots \end{pmatrix}. \tag{14}$$

Matrix $\mathbf{D}$ is a first difference operator. From this follows that matrix $\mathbf{\Lambda}$ implies for both models a new redistribution of the quarterly residuals, where a weighted moving average of quarterly residuals is given to each monthly estimated GDP.

### 3.2.4 Models with Related Series and AR Structure

**3.2.4.1 Model 2e** In addition to the introduction of related series explained in the previous class of models, we assume here in addition that monthly GDP is characterized by an AR structure. The nonstationarity correction is similar as in model **1a**. After the inclusion of related series to model **1a**, the state equation becomes equation (6) plus the term $\mathbf{C}'\mathbf{x}_{t+1}$ where $\mathbf{x}_{t+1}$ includes $l$ related series.

**3.2.4.2 Model 2f** This model is similar to model **1b** with added related series. As in model **2e,** matrix $\mathbf{C}'$ characterizes their impact on monthly GDP.

# 4 Data

## 4.1 Signal Extraction from Related Series

A key factor in the present interpolation problem is the signal extraction from related series. Besides the assumption about the dynamics of GDP, related series data represent the main information source for interpolation. These data must fulfill two requirements.

First, they need to be correlated with the series to interpolate. The higher the systematic comovements with GDP are, the stronger is the signal that can be exploited to fill the gaps. If however there is only a modest information content in the related series, this comes at the cost of noise introduced in the interpolated series. The choice of the related series is therefore crucial in order to successfully estimate a series at higher frequency.

Second, the related series need to be available in the desired higher frequency of the interpolated GDP. The fact that there are not many macroeconomic series available at monthly frequency imposes a strong restriction in Switzerland. This leads us to use also foreign variables that are correlated with the desired related series.

These two points require a thorough investigation for the task of choosing the correct related series. Amemiya (1980) suggests a joint strategy based on economic-theoretic considerations and on statistical evidence. Economic intuition often indicates which related series to choose and what functional form they should have. Moreover, it is convenient to have a single statistical measure to choose related series that produce the 'best' result. These two aspects, intuitive approach and choice metrics, should be viewed as forming a single choice package rather than being in competition with each other. They allow making a final choice of the series which we use in our models. We present both elements of this selection process in detail in the following section.

## 4.2 Choice of Related Series

### 4.2.1 Economic Intuition

The most natural way to approach the series selection problem is to split up GDP into its expenditure components, private consumption ($C$), private domestic investments ($I$), government expenses ($G$), and net exports ($X - M$):

$$Y = C + I + G + X - M. \tag{15}$$

With the exception of exports and imports, none of these series is available at the higher frequency. Therefore, it is necessary to identify related series that proxy for the desired components.

An alternative to breaking GDP into its expenditure components is to benefit from the characteristics of Switzerland as a small open economy and the important comovement between domestic and foreign business cycles. Taking into consideration monthly foreign economic indicators allows us to choose the related series from a broader data set as Switzerland's closest trade partners have traditionally large statistical databases.

### 4.2.2 Statistical Evaluation

The case discussed here is the search for individual proxy variables in economic models[19,20]. Suppose, we identify a set of related data series $\mathbf{X}$ out of which variable $x_k$ is unobservable. Furthermore, the variable $y$ that we want to interpolate depends linearly on $\mathbf{X}$.

$$y_t = \alpha_0 + \alpha_1 x_{1,t} + \alpha_2 x_{2,t} + \ldots + \alpha_k x_{k,t} + \ldots + u_t \qquad (16)$$

The goal is to choose the best observable proxy for $x_k$. In cases like this, an informal method often applied is replacing $x_k$ with the variable $z_k$ which yields the highest $R^2$ of all possible candidates in equation (16). Leamer (1983) shows that if the proxy variables are assumed to depend linearly on $x_k$ and the error terms are assumed to be normally iid, the best proxy is the one that produces the highest $R^2$. In the univariate regression $z_{i,t} = \delta_i x_{k,t} + \varepsilon_{i,t}$, the particular proxy $z_i$ which yields the smallest variance $\sigma^2_{\varepsilon_i}$ could be defined as the best one. Leamer (1983) uses a likelihood ratio test to show the unambiguous negative relationship between the variance of the error term and $R^2$.

Another popular method which can be applied to a wider range of competing models than the one $R^2$ criterion above is the method of penalized likelihood. The best known examples in this class of criteria are the Akaike (1974) Information Criterion (AIC) and the Schwarz (1978) Information Criterion (SIC). In this class of criteria, a term that acts to punish additional coefficients is added to the negative of the likelihood function. Smaller values of the criteria are preferable.

## 4.3 Data Description

For a long time, Switzerland has stayed far behind other European countries in the development of economic statistical data. In 1996, as part of a reform program, national accounting was adapted to the European System of Integrated Economic Accounts (ESA) 78[21].

[19] From the limited flexibility in applying economic theories due to data availability restrictions follows that the statistical evaluation must take a more important place than it usually would according to Amemiya (1980).

[20] At this stage of the text, we describe the data selection within one group of related series as described in section 4.2.1.

[21] Up to and including 1996, Swiss GDP was recorded following the OECD 58 standard. According to the Federal Statistics Office, it is planned to adopt the ESA 95 standard within a few years.

Thereafter, GDP was calculated differently. The Federal Statistics Office dated the series back to 1980 such that there is now a data sample of more than 18 years or 73 quarterly observations. The figures to be interpolated are deflated and deseasonalized[22].

The related series[23] in the national accounting approach have been identified as retail sales $x^{rs}$ to proxy for private consumption and as non-utilized construction loans to proxy for investment $x^{nl}$. These monthly available proxies have been selected based on the criteria described in the previous section. Furthermore, we include exports $x^X$ and imports $x^M$. All the series are entered in levels[24]. Government expenditure was dropped in the national accounting approach due to its low covariance with the business cycle. This would have introduced too much noise and moreover, there is no sensible proxy for it at monthly frequency.

As foreign series, we use a composite index[25] of IP $x^{comip}$, British IP $x^{ukip}$, and German IP $x^{brdip}$. IP are the foreign monthly available series that move closest with the Swiss business cycle of all the related foreign series considered (results not reported).

Prior to estimation, we have excluded several potential series based on economic arguments or on the statistical evaluation of the previous section. French IP, Italian IP, survey data by the KOF[26], labor market figures, exchange rates, and commodity prices were statistically eliminated. We have neither included variables that have proved to have predictive power for GDP such as the term spread because of unrealistic assumptions on the lead-lag relationship that would have been necessary. Table 1 and figure 2 give an overlook over the related series used in this paper.

[22] Deseasonalization was executed using the X12/ARIMA method of the US Bureau of Census.

[23] All the series, with the exception of real GDP given by the State Secretariat for Economic Affairs, are provided by Datastream.

[24] The models transform the level vectors into the desired form.

[25] IP of 5 countries (major trade partners of Switzerland) are weighted according to the share of Swiss exports to the respective countries in 1996.

[26] Institute for Business Cycle Research of the Swiss Federal Institute of Technology, Zurich.

Table 1: Data Description

Descriptive Statistics

| | $\mu$ | $\sigma$ | AR(1) | JB | ADF |
|---|---|---|---|---|---|
| $gdp$ | 1.33 | 2.95 | 0.25 | 0.05 | -4.38* |
| $x^{rs}$ | 3.20 | 46.29 | -0.65* | 77.88* | -11.04* |
| $x^{nl}$ | -0.97 | 21.39 | 0.25* | 199.18* | -2.98* |
| $x^{X}$ | 4.09 | 49.03 | -0.57* | 53.17* | -7.91* |
| $x^{M}$ | 4.29 | 51.11 | -0.61* | 85.52* | -8.06* |
| $x^{brdip}$ | 1.40 | 21.73 | -0.43* | 1370.58* | -6.26* |
| $x^{ukip}$ | 1.31 | 13.17 | -0.22* | 6.71* | -5.30* |
| $x^{comip}$ | 1.52 | 12.16 | -0.32* | 68.53* | -5.63* |

Dynamic Correlations with $gdp$

| | $-3$ | $-2$ | $-1$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| $x^{rs}$ | 0.01 | 0.12 | -0.01 | 0.09 | 0.13 | 0.12 | 0.02 |
| $x^{nl}$ | 0.30 | 0.37 | 0.30 | 0.23 | 0.23 | 0.14 | 0.16 |
| $x^{X}$ | 0.10 | 0.15 | 0.18 | 0.26 | 0.25 | -0.02 | -0.07 |
| $x^{M}$ | 0.16 | 0.17 | 0.20 | 0.09 | 0.26 | -0.05 | 0.03 |
| $x^{brdip}$ | 0.03 | 0.12 | 0.34 | 0.25 | 0.35 | 0.21 | 0.14 |
| $x^{ukip}$ | 0.03 | 0.09 | 0.17 | 0.05 | -0.01 | -0.09 | -0.17 |
| $x^{comip}$ | 0.05 | 0.18 | 0.41 | 0.27 | 0.30 | 0.19 | 0.09 |

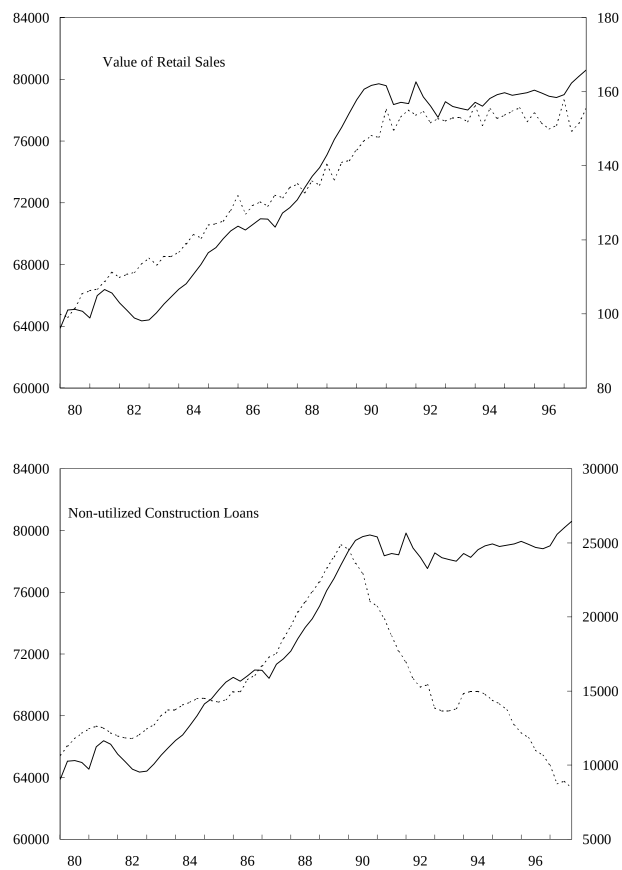
*Note*: Annualized statistical figures are calculated for quarterly growth rates of GDP and for monthly growth rates for all other variables. $gdp$ = Gross domestic product; $x^{rs}$ = Value of retail sales; $x^{nl}$ = Non-utilized construction loans; $x^{X}$ = Exports volume; $x^{M}$ = Imports volume; $x^{brdip}$ = IP in Germany; $x^{ukip}$ = IP in UK; $x^{comip}$ = Composite index of IP. All variables except $x^{comip}$ are seasonally adjusted. $\mu$ = Mean; $\sigma$ = Standard deviation; AR(1) = First-order autoregressive coefficient; JB = Jarque-Bera test; ADF = Augmented Dickey-Fuller test. Null hypotheses: i) first-order AR coefficient test, $H_0$: AR-coefficient = 0; ii) JB test, $H_0$: normal distribution; iii) ADF test, $H_0$: unit root. Rejection of the null hypothesis at the 1% significance level (*) and at the 5% significance level (**). Dynamic correlations with $gdp$ are cross-correlations of lags and leads (between -3 and 3) of quarterly growth rate of related series with quarterly GDP growth rate. Source: Datastream and State Secretariat for Economic Affairs.

During the 18 years of observations, the state of the Swiss economy can be roughly divided into two parts. Figure 2 clearly shows the phases of economic growth and prosperity in the eighties and of stagnation in the nineties. During its recession, Switzerland exhibited the lowest real GDP growth of all European countries[27].

[27] We decided not to take into account this structural break in the estimation of the Kalman parameters. It would mean to combine our interpolation models with time-varying parameters generally dealt within the Kalman filter framework

Table 1 reports basic summary statistics of the quarterly and monthly series used for interpolation. Following the integration results from figure 2 and from augmented Dickey-Fuller (ADF) tests for all the variables (not reported), we find that all the series in levels are nonstationary. Hence, we report the results for growth rates. The ADF tests and the AR(1) regressions concerning the growth rates confirm that the level of the series is not stationary. The different values of the contemporary cross-correlations also confirm the requirement of the comovements of the related series with quarterly GDP.

Figure 2: GDP and Related Series




or with the use of dummy variables. However, as shown with our calculus models, the Kalman filter considers the changing trend over time in computing the monthly estimates.

Figure 2 *Continued*

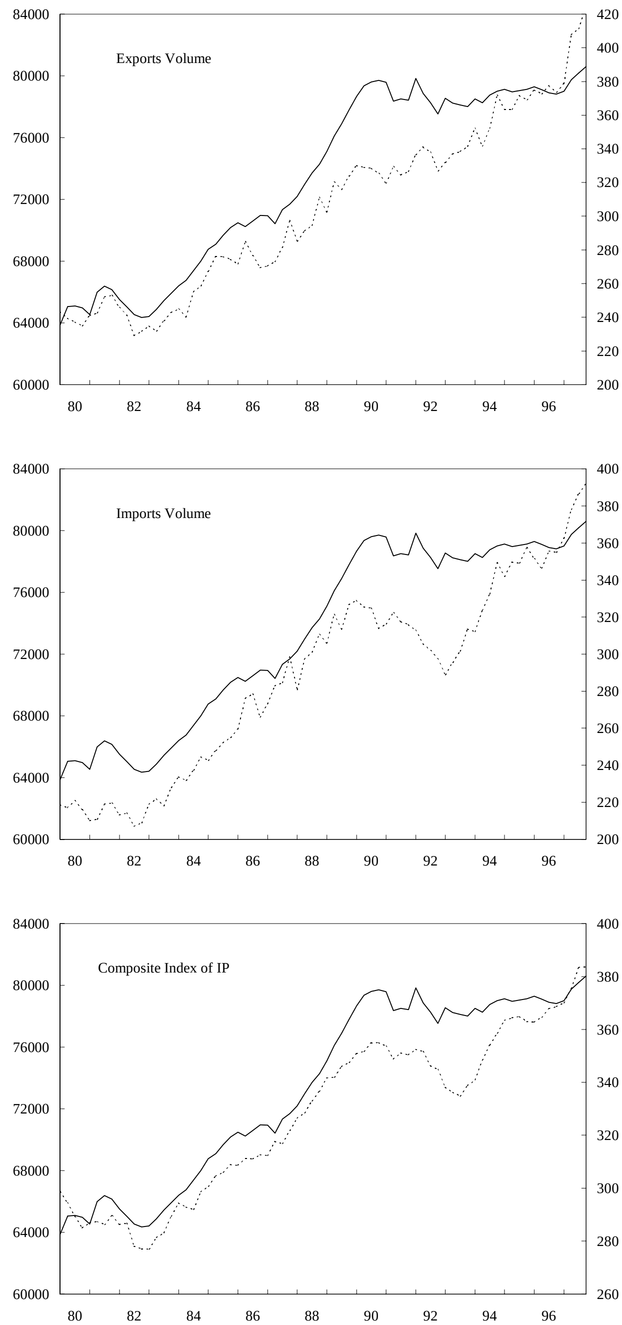
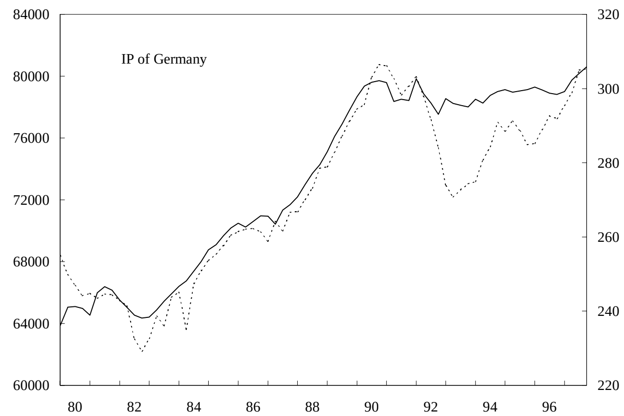
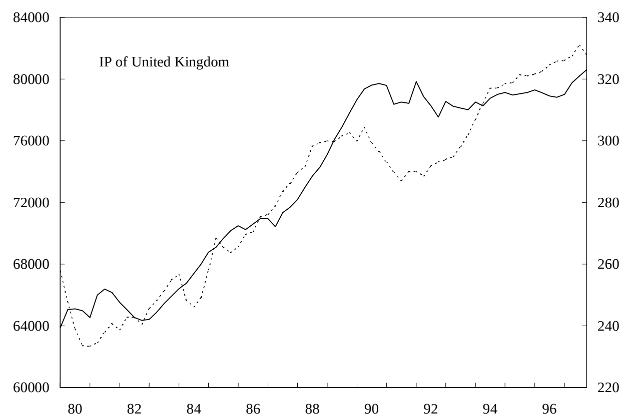


Figure 2 *Continued*




*Note*: GDP uses solid line in mio CHF; related series use dashed line. Non-utilized construction loans in mio CHF, other related series as index points. Source: Datastream and State Secretariat for Economic Affairs.

Finally, these cross-correlations also show why we only consider contemporary relationships between the related series and the quarterly GDP. It is difficult to find robust leads and lags - the so-called stylized facts of the business cycles literature - between GDP and our proxy variables.

We also perform a Johansen (1991) test in order to check for cointegration that is needed for the evaluation of the applicability of the Bernanke, Gertler and Watson (1997) approach. It is natural to assume that $x^{rs}$ is moving along with GDP. We therefore test the quarterly proxy for cointegration. The test results reject the hypothesis of no cointegration at the 1% significance level. These results are displayed in table 2.

Table 2: Cointegration Test of *gdp* and $x^{rs}$

| 80:01-97:12 | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $H_0$ | $H_a$ | LR | B | $H_0$ | $H_a$ | LR |
| A1 | 0 | 2 | 25.11* | B1 | 0 | 1 | 23.38* |
| A2 | 1 | 2 | 1.73 | B2 | 1 | 2 | 1.73 |

*Note*: Cointegration tests are performed with quarterly data. $gdp$ = Gross domestic product; $x^{rs}$ = Value of retail sales. $H_0$ = null hypothesis; $H_a$ = alternative hypothesis; for each hypothesis, given figure is number of cointegration relations; LR = likelihood ratio statistic. Tests are run assuming linear trend in data and an intercept in the cointegrating equation and in the vector autoregression. Two lags are included. Test A, null hypothesis of $h$ cointegrating relations against the alternative of no cointegration. LR is the weighted sum of the $(3-h)$-smallest eigenvalues. Test B, null hypothesis of $h$ cointegrating relations against the alternative of $h+1$ relations. LR is the weighted $h^{th}$ largest eigenvalue. Rejection of the null hypothesis at the 1% significance level (*) and at the 5% significance level (**).

We do not report the tests of other potentially cointegrated variables with an economic interpretation. All the tests reveal that only the quarterly GDP and $x^{rs}$ are cointegrated. Hence, we use $x^{rs}$ either as a related series or as the detrending series $p_t$ in the Bernanke, Gertler and Watson (1997) framework[28]. Since we cannot directly test the necessary multiplicative cointegration, an ADF test on the quarterly equivalent of $y^s = \frac{y}{x^{rs}}$ reveals that this ratio is stationary at the 1% significance level.

# 5 Results

## 5.1 Overview

The interpolation results are displayed in table 3. It contains for each model statistical information about the estimated series for the

[28]To prevent the detrending series from introducing excessive volatility in the system, we take only the low frequency part of $x^{rs}$ after Hodrick-Prescott filtering. The main objective of detrending GDP can still be maintained.

period 1981-1997[29,30], the related series, the information criterion, the log-likelihood, and key indicators for the annualized growth rate of the monthly interpolated GDP. Two mean-squared errors (MSE) for the evaluation of the models are given. The first one is between the level of the interpolated benchmark (model **1e**) and the interpolated series of each model, respectively. The second one is the MSE between the observed quarterly GDP and a simulated quarterly interpolated GDP from annual data within the model in question in order to compare how the interpolation model would have performed at a frequency where models can be selected unambiguously based on an available data set.

Table 3: Interpolation Results

| Model | 1, 1a | 2, 1c | 3, 2a | 4, 2a |
|---|---|---|---|---|
| Series | - | - | $x^{comip}$ | $x^{nl}$ |
| AIC | 8.05 | 10.08 | 14.66 | 17.31 |
| log L | -562.56 | -573.41 | -636.59 | -733.33 |
| $\mu$ | 1.32 | 1.33 | 1.20 | 1.25 |
| $\sigma$ | 4.30 | 5.11 | 9.93 | 5.22 |
| AR(1) | 0.21* | 0.06 | -0.37* | -0.02 |
| JB | 308.06* | 15.59* | 12.09** | 192.13* |
| ADF | -5.47* | -5.96* | -5.79* | -5.30* |
| 1e | 57.51 | 64.39 | 136.60 | 84.40 |
| AQ | 443.69 | 379.60 | 454.82 | 456.27 |

| Model | 5, 2a | 6, 2b | 7, 2b | 8, 2b |
|---|---|---|---|---|
| Series | $x^{rs,nl,X,M}$ | $x^{comip}$ | $x^{nl}$ | $x^{rs,nl,X,M}$ |
| AIC | 13.06 | 15.89 | 17.32 | 13.34 |
| log L | -575.38 | -681.65 | -733.97 | -585.70 |
| $\mu$ | 1.27 | 1.29 | 1.30 | 1.41 |
| $\sigma$ | 14.13 | 6.42 | 3.49 | 12.21 |
| AR(1) | -0.53* | -0.14** | 0.73* | -0.53* |
| JB | 2.77 | 17.93* | 16.83* | 0.52 |
| ADF | -6.40* | -5.53* | -4.20* | -6.85* |
| 1e | 179.80 | 94.25 | 65.16 | 153.63 |
| AQ | 498.24 | 360.40 | 311.72 | 470.60 |

[29] Due to initial oscillations using the Kalman filter, we discard twelve months of observations, which otherwise would have influenced the results.

[30] Note that we interpolate GDP with information that is available ex post. This ensures that monthly values sum to the quarterly observations. In certain empirical studies, the monthly business cycle indicator should represent the information set of decision makers in the respective period. In Switzerland, quarterly GDP is published about 10 weeks after the reference quarter. In this case, we recommend to use a series not influenced by the sum-up constraint. The presented methods generate such a series as a by-product (not reported).

Table 3 *Continued*

| Model | 9, 2c | 10, 2c | 11, 2c | 12, 2d |
|---|---|---|---|---|
| Series | $x^{comip}$ | $x^{nl}$ | $x^{rs,nl,X,M}$ | $x^{comip}$ |
| AIC | 15.63 | 14.42 | 14.38 | 15.27 |
| log L | -677.97 | -627.11 | -623.52 | -664.89 |
| $\mu$ | 1.30 | 1.30 | 1.34 | 1.27 |
| $\sigma$ | 3.65 | 3.51 | 4.66 | 6.80 |
| AR(1) | 0.63* | 0.72* | 0.14** | -0.23* |
| JB | 6.58** | 15.93* | 9.62* | 13.45* |
| ADF | -4.30* | -4.17* | -5.57* | -4.91* |
| 1e | 66.29 | 65.47 | 75.33 | 104.63 |
| AQ | 373.99 | 319.38 | 281.87 | 346.69 |

| Model | 13, 2d | 14, 2d | 15, 2e | 16, 2e |
|---|---|---|---|---|
| Series | $x^{nl}$ | $x^{rs,nl,X,M}$ | $x^{nl}$ | $x^{rs,nl,X,M}$ |
| AIC | 14.30 | 14.09 | 8.05 | 8.04 |
| log L | -622.86 | -614.83 | -562.55 | -562.55 |
| $\mu$ | 1.28 | 1.35 | 1.28 | 1.29 |
| $\sigma$ | 4.69 | 7.96 | 4.30 | 4.58 |
| AR(1) | 0.10 | -0.36* | 0.22* | 0.11 |
| JB | 332.63* | 15.08* | 306.50* | 156.83* |
| ADF | -4.92* | -5.63* | -5.45* | -5.56* |
| 1e | 81.83 | 108.53 | 57.29 | 60.84 |
| AQ | 388.83 | 376.60 | 384.66 | 404.59 |

*Note*: $x^{rs}$ = Value of retail sales; $x^{nl}$ = Non-utilized construction loans; $x^X$ = Exports volume; $x^M$ = Imports volume; $x^{comip}$ = Composite index of IP. All estimations include a constant, models 2c and 2d transform time trend into constant. Descriptive statistics are for annualized growth rates of the interpolated GDP for 1981-1997. log L = Value of log-likelihood function; $\mu$ = Mean; $\sigma$ = Standard deviation; AR(1) = First-order autoregressive coefficient; JB = Jarque-Bera test; ADF = Augmented Dickey-Fuller test. Null hypotheses: i) first-order AR coefficient test, $H_0$: AR-coefficient = 0; ii) JB test, $H_0$: normal distribution; iii) ADF test, $H_0$: unit root. Rejection of the null hypothesis at the 1% significance level (*) and at the 5% significance level (**). 1e = Root-MSE with 1e for 1981-1997; AQ = Root-MSE Annual→Quarterly for 1982-1996.

Note, that table 3 is constructed in order to evaluate the models with respect to two basic directions. First, it is important to know whether the inclusion of related series (class **2**) performs better than the ‘fool-yourself’ class **1**. Second, we investigate the appropriate treatment of nonstationarity and analyze whether applying recent techniques perform better than traditional ones.

## 5.2 Evaluation of Related Series

It is desirable to have an economic model underlying the interpolation instead of a purely econometric and mechanical procedure. Econometrically, the conclusion whether to include related series or not is however ambiguous. AIC and likelihood ratio tests show that introducing related series does not always enhance the performance of the interpolation as it involves costs of additional noise in the interpolated series. All the models generating too much volatility relative to the annualized standard deviation of the quarterly GDP estimates are not displayed in table 3 as they are economically not meaningful[31].

Related series could possibly break the regular pattern within a quarter, produced by all interpolation procedures without related series[32]. However, as shown in figure 3, series **16** for example is not able to break the pattern. The figure shows monthly estimates and the published quarterly GDP[33]. The cyclical pattern within the quarter, illustrated on the bottom, is an average difference for the three months within the quarter between series **16** and the benchmark **1e** for growth and decline periods, respectively. The deviations are significant for the first and the last observation within the quarter and lead us to reject the model for economic reasons. Moreover, we find that, in all series of type **2e**, the inclusion of related series, relative to model **1a**, even exacerbates the pattern.

The suggestion of Bomhoff (1994) that the Kalman filter accounts for nonstationarity cannot be generalized for interpolation with AR structure in the state equation. Explosive eigenvalues, responsible for the pattern, are introduced in models **1a** and **2e** by construction.

[31] We restrict ourselves to models that produce series with an annualized standard deviation lower than five times the variability of the growth rate of the official quarterly GDP estimates (15%). Comparisons between monthly and quarterly values of industrial production growth in various countries show that the annualized values of monthly standard deviation are two to five times higher than quarterly ones which serve as a reference.

[32] The pattern is systematically convex or concave if the model has an AR structure, depending on growth state of the economy. Monthly GDP estimates produced by model **1e** are linear and model **1d** produces monthly estimates which equal one third of the corresponding quarterly GDP.

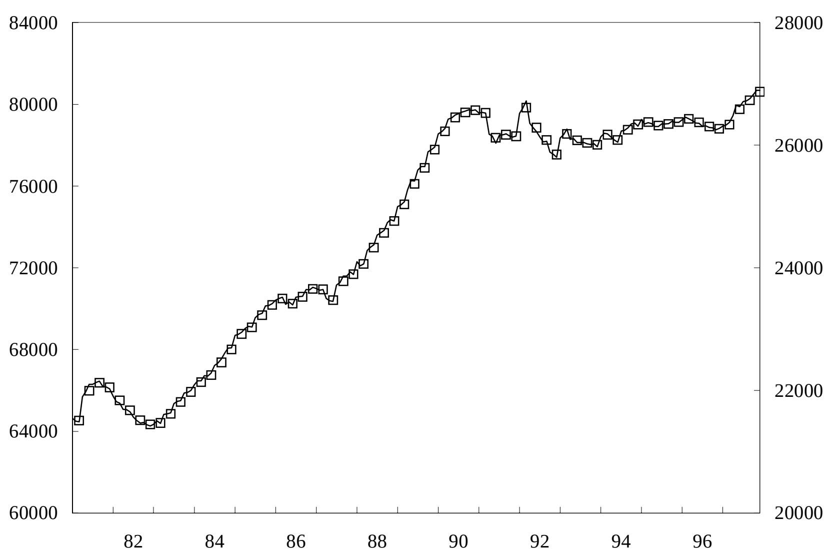
[33] The three months of each quarter sum to the value of this quarter. However, the line of the monthly interpolated GDP does not exactly pass through the points of quarterly GDP as the former is scaled by a factor of one third. The fact that the squares are not exactly on the line cannot be interpreted as a quality indicator.

In model **1c**, the pattern is implied by estimating a $\phi$ close to one. We eliminate the pattern by removing its source, the AR structure, and we use models **2a-d** which assume no AR process.

Regarding the two sets of related series, one observes that in general, related series based on the open economy assumption introduce less volatility in the generated growth rates than the national accounting variables. For the related series $x^{comip}$ and $x^{nl}$ this relation is reversed[34]. However, including additional variables in the national accounting approach increases the volatility considerably. To further investigate the characteristics of the most appropriate related series, note that within each model the log-likelihood values show that the national accounting approach is preferable even if not always significantly.

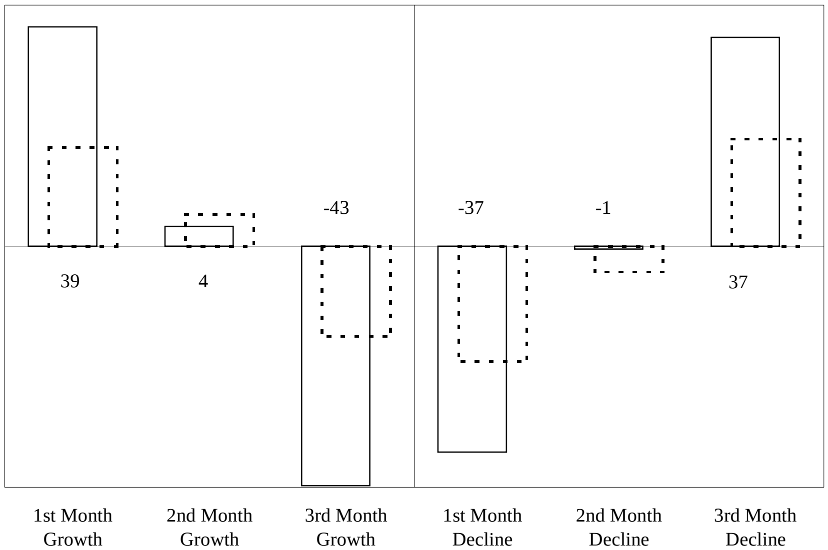
Another evaluation criterion is the MSE of a model series with respect to the benchmark **1e**. The results indicate that in general adding related series increases the MSE reflecting an increase in volatility as the models deviate more from the smooth benchmark.

Figure 3: Interpolated Series with Pattern




[34]Of all the series that could not be distinguished by statistical evaluation in section 4.2.2, $x^{comip}$ is found to be the most useful related series of the open economy approach. Results using $x^{brdip}$ and $x^{ukip}$ are therefore not reported in table 3.

Figure 3 *Continued*




*Note*: Interpolated monthly GDP (right-hand scale) is represented by the solid line; squares represent the published quarterly GDP estimates (left-hand scale). Series and patterns in mio CHF. Solid rectangle = Pattern; Dashed rectangle = t-distributed critical value.

As this criterion is a rather soft one and as there are models with the contrary effect, it does not seem suitable for model evaluation. Moreover, our benchmark is mainly founded on practical reasons and hence, cannot be regarded as an objective measure for model evaluation.

## 5.3 Evaluation of Techniques

The comparison between different interpolation setups and the question whether modern setups perform better than traditional ones are closely linked to the treatment of stationarity. First of all, within the regression-based methods the correction for nonstationarity proposed by Denton and Fernandez (DF, model **2c**) produces results that are qualitatively only slightly better than the classic Chow and Lin method using level series (CL, model **2a**).

The effect of modeling AR(1) error terms in the CL-model (model **2b**) and in the DF-model (model **2d**) is not clear. In the CL-models, the likelihood falls while for the DF-models it increases, when AR(1) error terms are considered. The standard deviation of the generated series rises in the CL-models and behaves irregularly in the DF-setups.

Models constructed with Bernanke, Gertler and Watson (1997) are clearly worse than the ones reported in table 3, both in terms of cyclical regularity and volatility. This procedure neglects the fact that the Kalman filter already corrects the nonstationarity of the data.

Generally, regression-based models yield good estimators while Kalman filter routines sometimes struggle with the numerical optimization. In case of model equivalence, we recommend for practical reasons the use of analytical solutions. However, due to its flexibility, the Kalman filter is able to model a much richer set of assumptions about the properties of monthly GDP, while the GLS approach utterly fails to model any stochastic behavior. This makes the Kalman filter an unavoidable tool when analyzing competing interpolation models.

### 5.4 A Monthly GDP Estimate

Based on this mixed evidence concerning the two directions, we recommend the series **5** for further research. Due to the absence of AR structure, it does not display a regular pattern and the series exhibits moderate volatility as shown in figure 4 and in table 4.

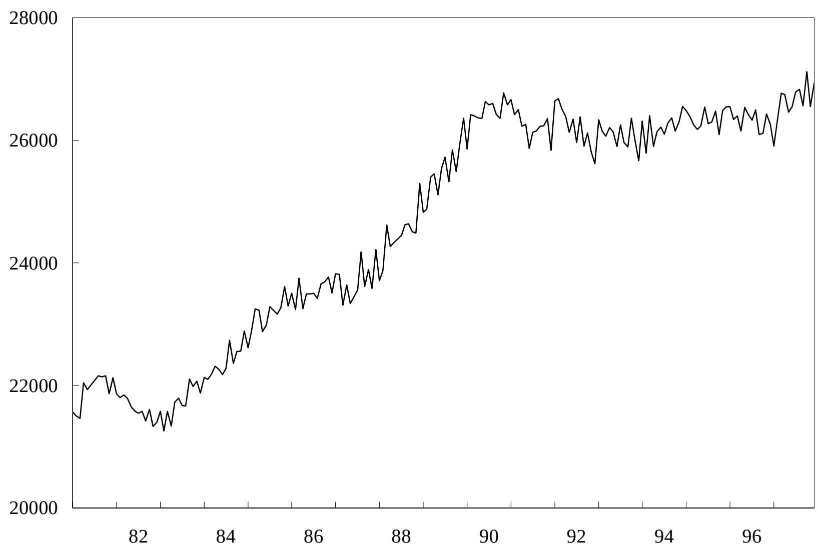
Can this extensive selection procedure be confirmed by first interpolating annual to quarterly data and then comparing the resulting quarterly series with the official GDP estimates? If yes, then we would have a very handy tool for the evaluation of competing interpolation models. Of course, the underlying assumption that the best annual interpolation model is also the best quarterly one is strong, but if this criterion does well, it could be used as suggestive evidence in similar problems. Moreover, there is no reason to think that the frequency change has a fundamental impact on the performance of the models[35].

Surprisingly, the results show that it is not always the case that models with highest likelihood are the best interpolating models at

[35] Another way to apply this proposal would be to select the model with the best AIC for the interpolation from annual data and to see whether the same model also produces the best AIC for the interpolation of monthly data from quarterly data.

the lower frequency. Within the GLS-based class just model **2c** confirms our expectations. For all models with a pattern, applying this method makes no sense. Therefore, we conclude that this approach may be used as an indicator only but certainly not as a selection criterion.

Figure 4: Monthly GDP 81:01-97:12




*Note*: Monthly GDP in mio CHF.

# 6 Conclusion

We describe a setup that nests a wide range of interpolation models in the literature and we apply it to Swiss GDP. The goal of this paper is to evaluate alternative interpolation models and then to produce a monthly deseasonalized real GDP available for researchers and practitioners.

With respect to the nonstationarity and to the usefulness of related series, it is difficult to a priori present a clear-cut answer how these issues can be best treated. Our results show that the ways to consider the nonstationarity problem and to solve it with the second-order AR structure (models **1a** and **2e**) or with the detrending method (models **1b** and **2f**) are not suitable for Swiss data.

These two methods impose an econometric burden on the produced data that cannot be carried further for an economic interpretation. The nonstationarity correction made by the filter itself seems to be sufficient.

Table 4: Interpolated GDP

| | GDP | | GDP | | GDP | | GDP |
|---|---|---|---|---|---|---|---|
| 81:01 | 21563 | 83:04 | 21337 | 85:07 | 23286 | 87:10 | 23888 |
| 81:02 | 21503 | 83:05 | 21730 | 85:08 | 23230 | 87:11 | 23580 |
| 81:03 | 21461 | 83:06 | 21793 | 85:09 | 23165 | 87:12 | 24213 |
| | *64527* | | *64860* | | *69681* | | *71682* |
| 81:04 | 22043 | 83:07 | 21670 | 85:10 | 23265 | 88:01 | 23706 |
| 81:05 | 21932 | 83:08 | 21666 | 85:11 | 23615 | 88:02 | 23872 |
| 81:06 | 22003 | 83:09 | 22104 | 85:12 | 23297 | 88:03 | 24615 |
| | *65978* | | *65440* | | *70177* | | *72193* |
| 81:07 | 22082 | 83:10 | 21986 | 86:01 | 23504 | 88:04 | 24270 |
| 81:08 | 22157 | 83:11 | 22066 | 86:02 | 23235 | 88:05 | 24336 |
| 81:09 | 22141 | 83:12 | 21876 | 86:03 | 23751 | 88:06 | 24389 |
| | *66380* | | *65928* | | *70490* | | *72995* |
| 81:10 | 22158 | 84:01 | 22130 | 86:04 | 23255 | 88:07 | 24449 |
| 81:11 | 21864 | 84:02 | 22096 | 86:05 | 23496 | 88:08 | 24622 |
| 81:12 | 22124 | 84:03 | 22178 | 86:06 | 23495 | 88:09 | 24639 |
| | *66146* | | *66404* | | *70246* | | *73710* |
| 82:01 | 21867 | 84:04 | 22314 | 86:07 | 23501 | 88:10 | 24505 |
| 82:02 | 21802 | 84:05 | 22267 | 86:08 | 23419 | 88:11 | 24484 |
| 82:03 | 21843 | 84:06 | 22175 | 86:09 | 23661 | 88:12 | 25295 |
| | *65512* | | *66756* | | *70581* | | *74284* |
| 82:04 | 21793 | 84:07 | 22275 | 86:10 | 23688 | 89:01 | 24823 |
| 82:05 | 21650 | 84:08 | 22734 | 86:11 | 23770 | 89:02 | 24877 |
| 82:06 | 21582 | 84:09 | 22359 | 86:12 | 23508 | 89:03 | 25400 |
| | *65025* | | *67368* | | *70966* | | *75100* |
| 82:07 | 21547 | 84:10 | 22553 | 87:01 | 23824 | 89:04 | 25453 |
| 82:08 | 21577 | 84:11 | 22558 | 87:02 | 23811 | 89:05 | 25112 |
| 82:09 | 21417 | 84:12 | 22887 | 87:03 | 23310 | 89:06 | 25541 |
| | *64541* | | *67998* | | *70945* | | *76106* |
| 82:10 | 21606 | 85:01 | 22616 | 87:04 | 23639 | 89:07 | 25723 |
| 82:11 | 21333 | 85:02 | 22898 | 87:05 | 23334 | 89:08 | 25329 |
| 82:12 | 21400 | 85:03 | 23248 | 87:06 | 23441 | 89:09 | 25842 |
| | *64339* | | *68762* | | *70414* | | *76894* |
| 83:01 | 21575 | 85:04 | 23229 | 87:07 | 23553 | 89:10 | 25490 |
| 83:02 | 21259 | 85:05 | 22877 | 87:08 | 24176 | 89:11 | 25936 |
| 83:03 | 21575 | 85:06 | 22985 | 87:09 | 23613 | 89:12 | 26363 |
| | *64409* | | *69091* | | *71342* | | *77789* |

Table 4 *Continued*

| | GDP | | GDP | | GDP | | GDP |
|---|---|---|---|---|---|---|---|
| 90:01 | 25861 | 92:01 | 26642 | 94:01 | 26317 | 96:01 | 26551 |
| 90:02 | 26420 | 92:02 | 26679 | 94:02 | 25790 | 96:02 | 26342 |
| 90:03 | 26398 | 92:03 | 26515 | 94:03 | 26406 | 96:03 | 26399 |
| | *78679* | | *79836* | | *78513* | | *79292* |
| 90:04 | 26368 | 92:04 | 26386 | 94:04 | 25901 | 96:04 | 26154 |
| 90:05 | 26359 | 92:05 | 26130 | 94:05 | 26142 | 96:05 | 26539 |
| 90:06 | 26634 | 92:06 | 26344 | 94:06 | 26216 | 96:06 | 26242 |
| | *79361* | | *78860* | | *78259* | | *79115* |
| 90:07 | 26580 | 92:07 | 25961 | 94:07 | 26102 | 96:07 | 26327 |
| 90:08 | 26603 | 92:08 | 26383 | 94:08 | 26281 | 96:08 | 26494 |
| 90:09 | 26424 | 92:09 | 25909 | 94:09 | 26368 | 96:09 | 26095 |
| | *79607* | | *78253* | | *78751* | | *78916* |
| 90:10 | 26361 | 92:10 | 26120 | 94:10 | 26155 | 96:10 | 26116 |
| 90:11 | 26774 | 92:11 | 25801 | 94:11 | 26303 | 96:11 | 26430 |
| 90:12 | 26582 | 92:12 | 25621 | 94:12 | 26557 | 96:12 | 26265 |
| | *79717* | | *77542* | | *79015* | | *78811* |
| 91:01 | 26663 | 93:01 | 26336 | 95:01 | 26486 | 97:01 | 25906 |
| 91:02 | 26418 | 93:02 | 26146 | 95:02 | 26386 | 97:02 | 26328 |
| 91:03 | 26499 | 93:03 | 26070 | 95:03 | 26254 | 97:03 | 26771 |
| | *79580* | | *78552* | | *79126* | | *79005* |
| 91:04 | 26233 | 93:04 | 26208 | 95:04 | 26179 | 97:04 | 26748 |
| 91:05 | 26261 | 93:05 | 26137 | 95:05 | 26238 | 97:05 | 26459 |
| 91:06 | 25869 | 93:06 | 25900 | 95:06 | 26544 | 97:06 | 26549 |
| | *78363* | | *78245* | | *78961* | | *79756* |
| 91:07 | 26128 | 93:07 | 26251 | 95:07 | 26271 | 97:07 | 26792 |
| 91:08 | 26151 | 93:08 | 25966 | 95:08 | 26296 | 97:08 | 26834 |
| 91:09 | 26234 | 93:09 | 25983 | 95:09 | 26475 | 97:09 | 26566 |
| | *78513* | | *78110* | | *79042* | | *80192* |
| 91:10 | 26237 | 93:10 | 26364 | 95:10 | 26098 | 97:10 | 27119 |
| 91:11 | 26355 | 93:11 | 25982 | 95:11 | 26486 | 97:11 | 26556 |
| 91:12 | 25840 | 93:12 | 25667 | 95:12 | 26547 | 97:12 | 26936 |
| | *78432* | | *78013* | | *79131* | | *80611* |

*Note*: Quarterly values in italics. All values in mio CHF.

Our results further show that in particular cases, related series can be very useful. In this case, economic interpretation, backed on our two economic models, is based on the comparison of the volatilities between the growth rates of the quarterly values and the computed monthly series, and some subsidiary indicators. However, including related series does not systematically improve the results of the base case.

Finally, the data does not seem to unambiguously confirm the expected long–run hypothesis between the interpolation at a monthly and at a quarterly level. A more rigorous econometric analysis would be needed to know if this comparison transgresses short-run considerations.

For the interpolation of Swiss GDP we suggest using an approach with the four related series retail sales, non-utilized construction loans, exports, and imports, the former two of which are proxying for consumption and investment which are not monthly recorded.

## References

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control* AC-19: 716–723.

Amemiya, T. (1980). Selection of regressors, *International Economic Review* 21(2): 331–354.

Aoki, M. and Havenner, A. (1991). State space modeling of multiple time series, *Econometric Review* 10(1): 1–59.

Bernanke, B. S., Gertler, M. and Watson, M. W. (1997). Systemic monetary policy and the effects of oil price shocks, *Brookings Papers on Economic Activity* 28(1): 91–157.

Bomhoff, E. J. (1994). *Financial Forecasting for Business and Economics*, Dryden.

Chow, G. C. and Lin, A.-L. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series, *Review of Economics and Statistics* 53(4): 372–375.

Chow, G. C. and Lin, A.-L. (1976). Best linear unbiased estimation of missing observations in an economic time series, *Journal of the American Statistical Association* 71(355): 719–721.

Cuche, N. A. and Hess, M. K. (1999a). Eine monatliche Datenreihe für das schweizerische Bruttoinlandprodukt, Konjunkturindikator und Hilfsgrösse für die Forschung, *Die Volkswirtschaft, Magazin für Wirtschaftspolitik* 72(9): 32–33.

Cuche, N. A. and Hess, M. K. (1999b). Estimating monthly GDP in a general Kalman filter framework: Evidence from Switzerland, *Working paper 9902*, Study Center Gerzensee.

De Alba, E. (1990). Estimación del PIB trimestral para México, *Estudios Económicos* 5(2): 359–370.

Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual totals, *Journal of the American Statistical Association* 66(333): 99–102.

Fernandez, R. B. (1981). A methodological note on the estimation of time series, *Review of Economics and Statistics* 63(3): 471–476.

Friedman, M. (1962). The interpolation of time series by related series, *Journal of the American Statistical Association* 57(300): 729–757.

Gourieroux, C. and Monfort, A. (1997). *Time Series and Dynamic Models*, Cambridge University.

Guay, R., Milbourne, R. D., Otto, G. and Smith, G. W. (1990). Estimation du PIB mensuel canadien: 1962 à 1985, *L'Actualité économique, Revue d'analyse économique* 66(1): 14–30.

Hamilton, J. D. (1994a). State-space models, *Handbook of Econometrics IV ed. by R. F. Engle and D. L. McFadden*, Elsevier.

Hamilton, J. D. (1994b). *Time Series Analysis*, Princeton University.

Harvey, A. C. (1989). *Forecasting, Structural Series Models and the Kalman Filter*, Cambridge University.

Harvey, A. C. and Pierse, R. G. (1984). Estimating missing observations in economic time series, *Journal of the American Statistical Association* 79(385): 125–131.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* 59(6): 1551–80.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, *Journal of Basic Engineering, Transactions of the ASME Series D* 82: 35–45.

Kalman, R. E. (1963). New methods in Wiener filtering theory, *Proceedings of the First Symposium of Engineering Applications of Random Function Theory and Probability ed. by J. L. Bogdanoff and F. Kozin*, Wiley.

Kobler, A. E. (1999). Sources and Dynamics of Macroeconomic Fluctuations in Switzerland: Evidence from a Structural Vector Autoregressive Approach. Ph.D. Dissertation, Zurich.

Lanning, S. G. (1986). Missing observations: A simultaneous approach versus interpolation by related series, *Journal of Economic and Social Measurement* 14(1): 155–163.

Leamer, E. E. (1983). Model choice and specification analysis, *Handbook of Econometrics I ed. by Z. Griliches and M. D. Intriligator*, North-Holland.

Litterman, R. B. (1983). A random walk, Markov model for the distribution of time series, *Journal of Business and Economic Statistics* 1(2): 169–173.

Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*, 2nd edn, Springer.

Pinheiro, M. and Coimbra, C. (1993). Distribution and extrapolation of time series by related series using logarithms and smoothing penalties, *Economia* 12(3): 359–374.

Schmidt, J. R. (1986). A general framework for interpolation, distribution, and extrapolation of a time series by related series, *Regional Econometric Modeling ed. by M. R. Perryman and J. R. Schmidt*, Kluwer.

Schwaller, A. and Parnisari, B. (1997). Die Quartalsschätzungen des Bruttoinlandproduktes auf Grundlage der revidierten Volkswirtschaftlichen Gesamtrechnung (ESVG 78), *Mitteilungsblatt für Konjunkturfragen, Federal Office for Economic Development and Labour* (1): 3–24.

Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics* 6: 147–164.

# Appendix

## Appendix A Kalman Filter Algorithm and Log-Likelihood Function

We show the iteration steps of the Kalman filter. We also give the log-likelihood function of our system. All our interpolation models are based on equations (3) and (4), $\boldsymbol{\xi}_{t+1} = \mathbf{F}\boldsymbol{\xi}_t + \mathbf{C}'\mathbf{x}_{t+1} + \mathbf{R}\mathbf{u}_{t+1}$ and $y_t^+ = \mathbf{a}_t'\mathbf{x}_t^* + \mathbf{h}_t'\boldsymbol{\xi}_t$. The Kalman filter iteration, update, prediction, and MSE steps at time $t$ are given by the following loop. At time $t$, we assume that $y_0^+, \ldots, y_t^+$ are known. The related series $\mathbf{x}$ and $\mathbf{x}^*$ are known up to $t+1$. The predictions done at time $t-1$ for $t$ are known: $\hat{\boldsymbol{\xi}}_{t|t-1}, \hat{y}^+_{t|t-1}$. The corresponding MSE are also known: $\mathbf{P}_{t|t-1} = \text{MSE}\left(\hat{\boldsymbol{\xi}}_{t|t-1}\right)$ and $\text{MSE}\left(\hat{y}^+_{t|t-1}\right)$.

Update Step

$$\hat{\boldsymbol{\xi}}_{t|t} = \hat{\boldsymbol{\xi}}_{t|t-1} + \underbrace{\mathbf{P}_{t|t-1}\mathbf{h}_t\left(\text{MSE}\left(\hat{y}^+_{t|t-1}\right)\right)^{-1}}_{Gain}\left(y_t^+ - \hat{y}^+_{t|t-1}\right)$$
$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{h}_t\left(\text{MSE}\left(\hat{y}^+_{t|t-1}\right)\right)^{-1}\mathbf{h}_t'\mathbf{P}_{t|t-1}$$

Prediction Step

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t} + \mathbf{C}'\mathbf{x}_{t+1}$$
$$\hat{y}^+_{t+1|t} = \mathbf{a}'_{t+1}\mathbf{x}^*_{t+1} + \mathbf{h}'_{t+1}\hat{\boldsymbol{\xi}}_{t+1|t}$$

MSE Step

$$\text{MSE}\left(\hat{\boldsymbol{\xi}}_{t+1|t}\right) = \mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}' + \mathbf{R}\mathbf{Q}\mathbf{R}'$$
$$\text{MSE}\left(\hat{y}^+_{t+1|t}\right) = \mathbf{h}'_{t+1}\mathbf{P}_{t+1|t}\mathbf{h}_{t+1}$$

Log-Likelihood Function

Each observation is normally distributed:

$$y_t^+ \mid \left(y_0^+, \ldots, y_{t-1}^+, \mathbf{x}_1, \ldots, \mathbf{x}_t, \mathbf{x}_1^*, \ldots, \mathbf{x}_t^*\right) \sim N\left(\mathbf{a}_t'\mathbf{x}_t^* + \mathbf{h}_t'\hat{\boldsymbol{\xi}}_{t|t-1}, \mathbf{h}_t'\mathbf{P}_{t|t-1}\mathbf{h}_t\right).$$

The log-likelihood function for the whole sample is

$$\sum_{t=1}^{\frac{T}{3}} \ln f\left(y_t^+\right) = -\frac{T}{6}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{\frac{T}{3}} \ln\left|\mathbf{h}_t'\mathbf{P}_{t|t-1}\mathbf{h}_t\right| - \frac{1}{2}\sum_{t=1}^{\frac{T}{3}} \frac{\left(y_t^+ - \mathbf{a}_t'\mathbf{x}_t^* - \mathbf{h}_t'\hat{\boldsymbol{\xi}}_{t|t-1}\right)^2}{\mathbf{h}_t'\mathbf{P}_{t|t-1}\mathbf{h}_t},$$

$$\sum_{t=1}^{\frac{T}{3}} \ln f\left(y_t^{+}\right) = -\frac{T}{6}\ln\left(2\pi\right) - \frac{1}{2}\sum_{t=1}^{\frac{T}{3}} \ln\left|\mathbf{h}_t'\left(\mathbf{F}\mathbf{P}_{t-1|t-1}\mathbf{F}' + \mathbf{R}\mathbf{Q}\mathbf{R}'\right)\mathbf{h}_t\right|$$

$$-\frac{1}{2}\sum_{t=1}^{\frac{T}{3}} \frac{\left(y_t^{+} - \mathbf{a}_t'\mathbf{x}_t^{*} - \mathbf{h}_t'\left(\mathbf{F}\hat{\boldsymbol{\xi}}_{t-1|t-1} + \mathbf{C}'\mathbf{x}_t\right)\right)^2}{\mathbf{h}_t'\left(\mathbf{F}\mathbf{P}_{t-1|t-1}\mathbf{F}' + \mathbf{R}\mathbf{Q}\mathbf{R}'\right)\mathbf{h}_t}.$$

## Appendix B Chow and Lin Regression

We show hereafter the Chow and Lin regression model[36]. Chow and Lin assume a true model for the monthly GDP explained by $l$ related series given in matrix notation for the whole sample of $T$ observations: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ where $\mathbf{V} = E(\mathbf{uu}')$ is the variance-covariance matrix of the error terms. With help of a $\left[\frac{T}{3} \times T\right]$ matrix $\mathbf{C}_D = \frac{1}{3}\left(\mathbf{I}_{\frac{T}{3}} \otimes \mathbf{i}_3'\right)$, they transform this model to match the quarterly observed GDP. The quarterly vector can thus be expressed:

$$\mathbf{y}^+ = \mathbf{C}_D\mathbf{y} = \mathbf{C}_D\mathbf{X}\boldsymbol{\beta} + \mathbf{C}_D\mathbf{u} = \mathbf{X}^+\boldsymbol{\beta} + \mathbf{u}^+$$

where $E\left(\mathbf{u}^+\mathbf{u}^{+\prime}\right) = \mathbf{V}^+ = \mathbf{C}_D\mathbf{V}\mathbf{C}_D'$ and where $\mathbf{X}^+$ is a $\left[\frac{T}{3} \times l\right]$ matrix with quarterly averages of related series. Chow and Lin look then for a $\left[T \times \frac{T}{3}\right]$ matrix $\mathbf{A}$ that fills the gap between quarterly and estimated monthly data such that

$$\hat{\mathbf{y}} = \mathbf{A}\mathbf{y}^+.$$

In this search they impose an unbiased estimated monthly series $\hat{\mathbf{y}}$:

$$E(\hat{\mathbf{y}} - \mathbf{y}) = E\left(\mathbf{A}\left(\mathbf{X}^+\boldsymbol{\beta} + \mathbf{u}^+\right) - \mathbf{X}\boldsymbol{\beta} - \mathbf{u}\right) = E\left(\left(\mathbf{A}\mathbf{X}^+ - \mathbf{X}\right)\boldsymbol{\beta}\right) = \mathbf{0}$$

implying $\mathbf{AX}^+ - \mathbf{X} = \mathbf{0}$ and giving then an expression for the difference between the true monthly series and the estimated one: $\hat{\mathbf{y}} - \mathbf{y} = \mathbf{Au}^+ - \mathbf{u}$. To find the optimal matrix $\mathbf{A}$, they minimize, under the constraint of unbiasedness, the trace of the variance-covariance matrix $V[\hat{\mathbf{y}}]$, so minimizing the sum of all the variances corresponding to each 'observation'. $V[\hat{\mathbf{y}}]$ is the following equation.

$$\begin{aligned} E(\hat{\mathbf{y}} - \mathbf{y})^2 &= \mathbf{A}E\left(\mathbf{u}^+\mathbf{u}^{+\prime}\right)\mathbf{A}' - \mathbf{A}E\left(\mathbf{u}^+\mathbf{u}'\right) - E\left(\mathbf{u}\mathbf{u}^{+\prime}\right)\mathbf{A}' + E\left(\mathbf{uu}'\right) \\ &= \mathbf{AV}^+\mathbf{A}' - \mathbf{AV}^{+"} - \mathbf{V}^{"+}\mathbf{A}' + \mathbf{V} \end{aligned}$$

They minimize with respect to $\mathbf{A}$ the following Lagrange function with help of a $[l \times T]$ Lagrange multiplier $\mathbf{M}'$:

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}tr\left(\mathbf{AV}^+\mathbf{A}' - \mathbf{AV}^{+"} - \mathbf{V}^{"+}\mathbf{A}' + \mathbf{V}\right) - tr\left(\mathbf{M}'\left(\mathbf{AX}^+ - \mathbf{X}\right)\right), \\ &= \frac{1}{2}tr\left(\mathbf{AV}^+\mathbf{A}'\right) - tr\left(\mathbf{AV}^{+"}\right) + \frac{1}{2}tr(\mathbf{V}) - tr\left(\mathbf{X}^+\mathbf{M}'\mathbf{A}\right) + tr\left(\mathbf{M}'\mathbf{X}\right) \end{aligned}$$

yielding $\mathbf{A}$:

$$\begin{aligned} \mathbf{A} &= \mathbf{X}\left(\mathbf{X}^{+\prime}\mathbf{V}^{+^{-1}}\mathbf{X}^+\right)^{-1}\mathbf{X}^{+\prime}\mathbf{V}^{+^{-1}} \\ &\quad + \mathbf{V}^{"+}\mathbf{V}^{+^{-1}}\left(\mathbf{I} - \mathbf{X}^+\left(\mathbf{X}^{+\prime}\mathbf{V}^{+^{-1}}\mathbf{X}^+\right)^{-1}\mathbf{X}^{+\prime}\mathbf{V}^{+^{-1}}\right). \end{aligned}$$

$\hat{\mathbf{y}}$ is then given by the following fitted values.

$$\hat{\mathbf{y}} = \mathbf{A}\mathbf{y}^+$$

[36] We do not write GLS subscripts.

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{X}\overbrace{\left(\mathbf{X}^{+\prime}\mathbf{V}^{+^{-1}}\mathbf{X}^{+}\right)^{-1}\mathbf{X}^{+\prime}\mathbf{V}^{+^{-1}}\mathbf{y}^{+}}^{\hat{\boldsymbol{\beta}}} \\ &\quad +\mathbf{V}^{"+}\mathbf{V}^{+^{-1}}\underbrace{\left(\mathbf{I}_{\frac{T}{3}}-\mathbf{X}^{+}\left(\mathbf{X}^{+\prime}\mathbf{V}^{+^{-1}}\mathbf{X}^{+}\right)^{-1}\mathbf{X}^{+\prime}\mathbf{V}^{+^{-1}}\right)\mathbf{y}^{+}}_{\hat{\mathbf{u}}^{+}} \\ &= \mathbf{X}\hat{\boldsymbol{\beta}}+\left(\mathbf{V}^{"+}\mathbf{V}^{+^{-1}}\right)\hat{\mathbf{u}}^{+}=\mathbf{X}\hat{\boldsymbol{\beta}}+\boldsymbol{\Lambda}\hat{\mathbf{u}}^{+}\end{aligned}$$

The monthly series is computed by the third of all the elements of vector $\hat{\mathbf{y}}$.

## Appendix C Comparison of Log-Likelihood Functions

In a slightly different notation, the Kalman filter presented in section 'Models with Related Series and without AR Structure' yields the same likelihood as Chow and Lin. This shows that the results are the same for both methods. We assume the structural equation

$$y_t = \mathbf{x}_t'\mathbf{c} + u_t, \text{ for } t = 1, \ldots, T, \text{ where } u_t \text{ is iid and } E(u_t^2) = \sigma_u^2.$$

We define state vector

$$\boldsymbol{\xi}_t = \begin{pmatrix} y_t - \mathbf{x}_t'\mathbf{c} \\ y_{t-1} - \mathbf{x}_{t-1}'\mathbf{c} \\ y_{t-2} - \mathbf{x}_{t-2}'\mathbf{c} \end{pmatrix},$$

state equation

$$\boldsymbol{\xi}_{t+1} = \mathbf{I}_3 \begin{pmatrix} u_{t+1} \\ u_t \\ u_{t-1} \end{pmatrix},$$

and measurement equation

$$y_t^+ = \mathbf{a}_t'\mathbf{x}_t^* + \mathbf{h}_t'\boldsymbol{\xi}_t,$$

where $\mathbf{x}_t^* = \sum_{j=t-2}^{t} \mathbf{x}_j$, $\mathbf{h}_t' = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$ and $\mathbf{a}_t' = \mathbf{0}$ for $t = 1, 2, 4, 5, 7, \ldots, T-1$, and $\mathbf{h}_t' = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$ and $\mathbf{a}_t' = \mathbf{c}'$ for $t = 3, 6, 9, \ldots, T$. We further assume

$$\hat{\boldsymbol{\xi}}_{t|t-1} = \begin{pmatrix} \hat{y}_{t|t-1} - \mathbf{x}_t'\mathbf{c} \\ \hat{y}_{t-1|t-1} - \mathbf{x}_{t-1}'\mathbf{c} \\ \hat{y}_{t-2|t-1} - \mathbf{x}_{t-2}'\mathbf{c} \end{pmatrix} \text{ and } \mathbf{P}_{t|t-1} = \begin{pmatrix} \sigma_u^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix}.$$

Finally, the log-likelihood function for this Kalman filter is

$$\begin{aligned} \sum_{t=1}^{\frac{T}{3}} \ln f\left(y_t^+\right) &= -\frac{T}{6}\ln(2\pi) - \frac{T}{6}\ln\left(3\sigma_u^2\right) \\ &\quad -\frac{1}{6\sigma_u^2}\sum_{\tau=1}^{\frac{T}{3}}\left(y_{3\tau}^+ - \mathbf{a}_{3\tau}'\mathbf{x}_{3\tau}^* - \mathbf{h}_{3\tau}'\hat{\boldsymbol{\xi}}_{3\tau|3\tau-1}\right)^2, \\ &= -\frac{T}{6}\ln(2\pi) - \frac{T}{6}\ln\left(3\sigma_u^2\right) \\ &\quad -\frac{1}{6\sigma_u^2}\sum_{\tau=1}^{\frac{T}{3}}\left(\ y_{3\tau}^+ - \mathbf{c}'\left(\mathbf{x}_{3\tau} + \mathbf{x}_{3\tau-1} + \mathbf{x}_{3\tau-2}\right)\ \right)^2. \end{aligned}$$

Chow and Lin assume the following regression $\mathbf{y}^+ = \mathbf{X}^+\boldsymbol{\beta} + \mathbf{u}^+$ or for each observation $y_t^+ = \boldsymbol{\beta}'\mathbf{x}_t^+ + u^+$ meaning that quarterly observations are $N\left(\boldsymbol{\beta}'\mathbf{x}_t^+, \sigma_{u^+}^2\right)$. The log-likelihood function for this Chow and Lin regression is

$$\sum_{t=1}^{\frac{T}{3}} \ln f\left(y_t^+\right) = -\frac{T}{6}\ln(2\pi) - \frac{T}{6}\ln\left(\sigma_{u^+}^2\right) - \frac{1}{2}\left(\sigma_{u^+}^2\right)^{-1}\sum_{\tau=1}^{\frac{T}{3}}\left(y_{3\tau}^+ - \boldsymbol{\beta}'\mathbf{x}_{3\tau}^+\right)^2.$$

This equation is equivalent to the Kalman filter log-likelihood assuming $\boldsymbol{\beta}' = 3\mathbf{c}'$ and $\mathbf{x}_t^+ = \frac{1}{3}\mathbf{x}_t^*$. Finally, the variance of quarterly observations $\sigma_{u^+}^2$ equals three times the variance of monthly observations $\sigma_u^2$.