Fractional Time-Delay Estimation for Linear Dynamic Systems: Practical Examples

Peter C. Young[a], Fengwei Chen[b]

[a]*Lancaster Environment Centre and Data Science Institute, Lancaster University, Lancaster LA1 4YQ, UK*
[b]*School of Automation, Chongqing University, Chongqing 400044, China*

Abstract

This report is associated with the paper *A Simple Robust Method of Fractional Time-Delay Estimation for Linear Dynamic Systems* that has been published in Automatica [2]. In particular, it describes in some detail three practical examples that had to be excluded from the paper because of size restrictions.

Keywords: continuous-time model, instrumental variable, identification, fractional time delay

1. Introduction

This report is concerned with estimating the parameters of single and multi-input, continuous-time systems from sampled input-output data, where any input time delays may be a fraction of the sampling interval. The report is associated directly with the paper *A Simple Robust Method of Fractional Time-Delay Estimation for Linear Dynamic Systems* that has been published in Automatica [2]. This paper discusses the *Fractional Time Delay* (FTD) algorithm in detail and includes both theoretical analysis and discussion of pre-filtering that may be required at times when the input signals to the system include sharp changes, as in the case of repeated steps or PRBS signals. The present report describes three practical examples that were originally part of the paper but were removed because of size limitations. The two-stage estimation procedure used in these examples is available as the rivcbjfd routine in the CAPTAIN toolbox for Matlab. It has been developed and used for some time and appears both robust, and widely applicable, as illustrated by the practical examples in this report. In particular, the routine avoids the potential difficulties that may be encountered when standard gradient-based optimization is used for solving this optimization problem directly (see discussion in the paper and [**?**])

2. Real Data Examples

This section describes two examples based on the analysis of real data. The first example is concerned with the data from an industrial gas furnace, first used in the well known book by Box and Jenkins [1]. The second example considers the estimated fractional delay relationship between effective rainfall and flow in a river catchment, where the fractional delay enhances the forecasting ability of the model. The final example is concerned with the modeling of unemployment percentage in the USA based on macro-economic data and is a follow-up of previous, related modeling study reported in an earlier paper [10].

Email addresses: p.young@lancaster.ac.uk (Peter C. Young), fengwei.chen@cqu.edu.cn (Fengwei Chen)

2.1. The Box–Jenkins Gas Furnace Data Revisited

This section describes an example based on the analysis of the real data shown in Fig. 1. These are a selected subset of the data that were first used in the well known book by Box and Jenkins [1] to show how their maximum likelihood approach to time series analysis could be used to identify the *discrete-time* (DT) *transfer function* (TF) between two time series, in this case the gas rate input $u(t_k)$ and the percentage CO_2 output $y(t_k)$ of an industrial gas furnace, at an original sampling interval $T = 0.15$ minutes.

The 75 sample subset of the data shown in Fig. 1 are obtained from the first 150 samples of the original data, using only every other sampled value of $u(t_k)$ and $y(t_k)$. The data have been selected in this manner for two reasons. First, previous analysis of the full data set [6] has shown that there is clear evidence that the dynamic behaviour is changing over the latter part of the data, causing variation in the parameters in an identified first-order DT model. Second, the data are sub-sampled to a sampling interval $T = 0.3$ minutes, so that the sampled data present a more difficult identification problem and the improvements obtained by the introduction of a FTD are revealed more clearly.

In this case, the rivcbjid routine in the CAPTAIN Toolbox identifies the following first-order, CT, linear model between $u(t)$ and the noise free output $x(t)$ from the integer sampled data $u(t_k)$ and $y(t_k)$:

$$\begin{cases} x(t) = \mathrm{e}^{-\tau p} \dfrac{b_0}{p + a_1} u(t) \\ y(t_k) = x(t_k) + \xi(t_k) \end{cases} \tag{1}$$

where the identified integer time delay is 2 sampling intervals, i.e. $\hat{\tau} = 0.6$ minutes; and $\xi(t_k)$ is DT noise modeled as an AR(4) autoregressive process. The estimated parameters are as follows, with the estimated standard errors shown in parentheses:

$$\hat{a}_1 = 6.784(\pm 0.524); \ \hat{b}_0 = -20.61(\pm 1.529) \tag{2}$$

so that the estimated steady state gain and time constant are $G = -3.038$ and $T_c = 0.147$ minutes, respectively. Using the same first-order model structure as in equation (1), the rivcbjfd routine in CAPTAIN, using the 'spline' interp1 option and an accuracy of 0.01 (100 sub-samples specified over each sampling interval),

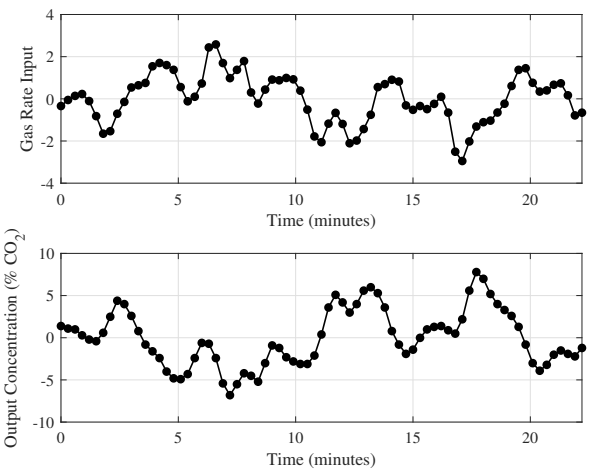

Figure 1: A sub-sampled subset of the original industrial gas furnace data [1] (first 75 samples with $T = 0.3$ minutes.)

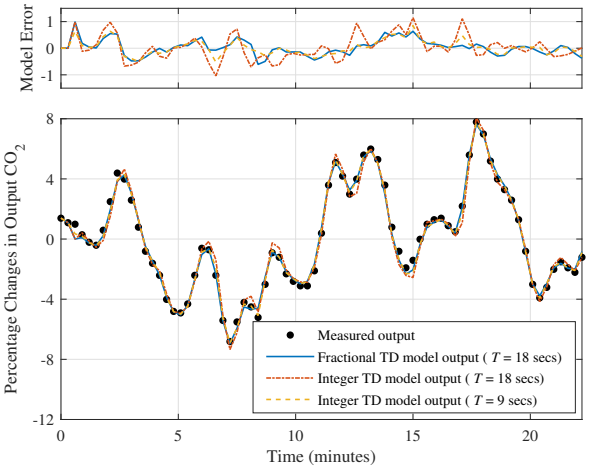

Figure 2: Comparison of the deterministic output responses of the integer and fractional time-delay models.

identifies a FTD model with an estimate of $\hat{\tau} = 0.453$ minutes. The estimated parameters of this model are as follows:

$$\hat{a}_1 = 3.262(\pm 0.106);\ \hat{b}_0 = -10.58(\pm 0.320) \qquad (3)$$

so that the steady state gain and time constant are now estimated as $G = -3.243$ and $T_c = 0.307$ minutes, respectively.

Both models explain the data quite well, with $R_T^2 = 0.982$ for the integer model and $R_T^2 = 0.992$ for the fractional model. The associated responses are shown in Fig. 2, where they are compared with the response of the rivcbj identified model with an integer time delay of 0.45 minutes, but based on the *original data* with a sampling interval of 0.15 minutes (i.e. the time delay is 3 sampling intervals). The estimates of the parameters in this latter model are as follows:

$$\hat{a}_1 = 3.206(\pm 0.071);\ \hat{b}_0 = -10.31(\pm 0.212). \qquad (4)$$

It is interesting to see how similar the parameters and the associated

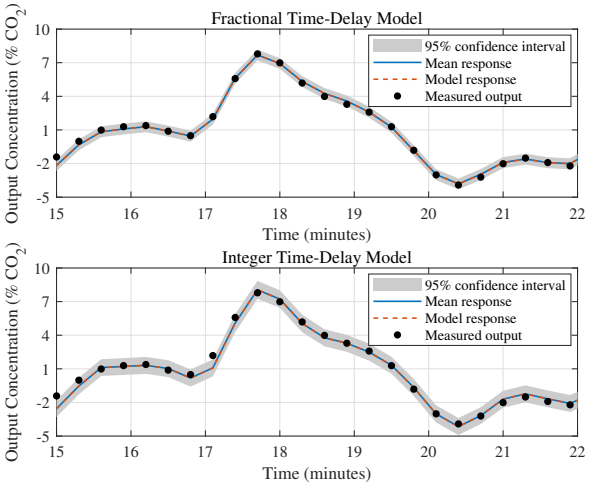

Figure 3: Comparison of the stochastic output responses of the integer and fractional time-delay models as obtained using Monte Carlo Simulation analysis.

model response of this model are to the FTD model, even though this latter model is based on only half the data. In fact, the fractional delay model, with $\hat{\tau} = 0.453$ minutes is effectively showing that the 3 sampling interval *integer* time-delay model is appropriate if the original, more rapidly sampled, data are utilized.

For comparison with the above models, the tfrivc routine in the CONTSID Toolbox identifies a quite similar FTD model, where the estimate of $\hat{\tau} = 0.454$ minutes and the estimated parameters are as follows:

$$\hat{a}_1 = 3.161(\pm 0.189);\ \hat{b}_0 = -10.28(\pm 0.557) \qquad (5)$$

so that the steady state gain and time constant are now estimated as $G = -3.253$ and $T_c = 0.316$ minutes, respectively.

Finally, the stochastic properties of the models obtained using MC simulation analysis, based on 10,000 realizations, confirm the above results. For clarity, Fig. 3 compares the latter section the stochastic output response of the integer and fractional delay models with the measured data where, as expected, we see that the 95% confidence bounds are visibly smaller in the fractional delay model case. The 95 percentile bounds on the estimated gain and time constant for all the identified models are shown in Table 1. Here, the two final columns show the 95 percentile bounds as a percentage of the associated median estimates. These reveal the improved confidence resulting from the application of the FTD estimation in these difficult sampling conditions and show that rivcbjfd estimation is advantageous in this regard.

2.2. *Rainfall-Flow Modeling and Forecasting*

Rainfall-flow models of rivers are important in flow and flood forecasting, as well as river catchment management and control studies. This example is based on the analysis of 'effective' rainfall and flow data for two very different catchments, one in Australia and one in the UK.

Table 1: Comparison of MC simulation results for steady state gain G and time constant T_c of the various integer and FTD models.

Algorithm (T)	Type (delay)	5% G	Median G	95% G	5% T_c	Median T_c	95% T_c	G %	T_c %
rivcbj (0.3 m)	integer (0.6 m)	−3.125	−3.038	−2.957	0.1309	0.1475	0.1693	5.52%	26.05%
rivcbjfd (0.3 m)	fractional (0.453 m)	−3.302	−3.242	−3.185	0.2911	0.3066	0.3238	3.62%	10.67%
tfrivc(0.3 m)	fractional (0.454 m)	−3.322	−3.253	−3.194	0.2882	0.3164	0.3510	3.95%	19.86%
rivcbj (0.15 m)	integer (0.45 m)	−3.386	−3.271	−3.162	0.3194	0.3410	0.3669	6.85%	13.93%

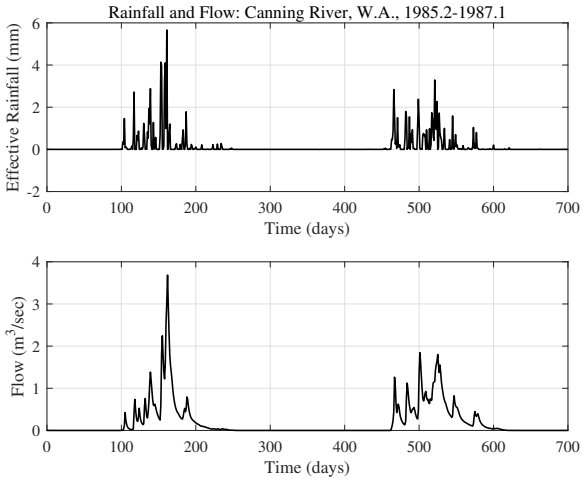

Figure 4: Daily rainfall and flow data in the Coweeta Catchment W.A. 1985-87.

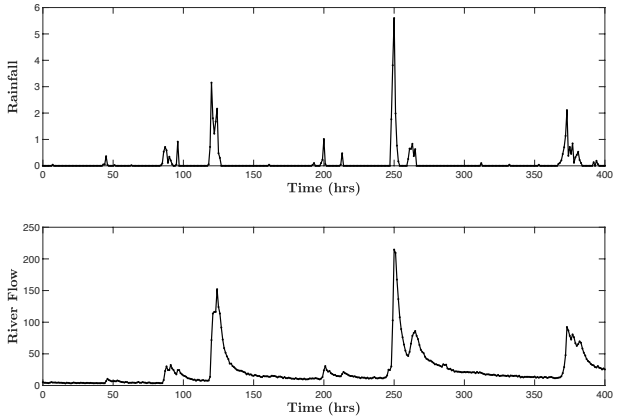

Figure 5: A short section of effective rainfall-flow data for a typical river catchment in the UK.

2.2.1. The Canning River, Western Australia

plotted in Fig. 4, which were measured daily over a two year period from 1985 to 1987 in a stretch of the ephemeral[1] Canning River located in the south west of Western Australia [7].

Model structure identification using the rivcbjid routine in the CAPTAIN Toolbox identifies the following 2nd-order, integer delay model between the effective rainfall and flow:

$$\begin{cases} x(t) = \mathrm{e}^{-1.0p}\dfrac{0.0713p^2 + 0.3100p + 0.0315}{p^2 + 0.4240p + 0.0221}u(t) \\ y(t_k) = x(t_k) + \xi(t_k) \end{cases} \tag{6}$$

where $\xi(t_k)$ is DT noise modeled as an AR(8) autoregressive process. The flow is explained reasonably well by this model, with a coefficient of determination $R_T^2 = 0.940$.

Using this same 2nd-order model structure, the rivcbjfd routine yields the following FTD model:

$$\begin{cases} x(t) = \mathrm{e}^{-0.73p}\dfrac{0.0864p^2 + 0.4044p + 0.0359}{p^2 + 0.4737p + 0.02453}u(t) \\ y(t_k) = x(t_k) + \xi(t_k) \end{cases} \tag{7}$$

with a 0.73 day (17.52 hours) time delay. The full estimation results are shown in Table 2, where we see that this fractional delay model improves the explanation of the river flow with $R_T^2 = 0.953$ compared with 0.940 for its integer time-delay equivalent. Moreover, as required in this application area, both the model structure and its estimated parameters make good physical sense (see [7]).

[1] i.e. the river flow ceases over the hot summer season

In particular, when the model is decomposed into a parallel connection of first-order systems, they have time constants of 2.4 and 16.9 days, together with a direct feed-through effect; and the associated 'partition percentages' of 54%, 40%, and 6%, respectively, define the relative surface, groundwater and instantaneous effects of the rainfall on the river flow.

The fractional delay model also performs better in a forecasting role, with the comparative R^2 values in the final column of 0.986 and 0.981, where R^2 is the standard one-step-ahead prediction-error coefficient of determination. This results in the 24 hour ahead forecasts being superior on 524 out of 700 days, or 75.0% of the time, with the mean absolute percentage error reduced from 13% to 11%. Note that there is no difference between the R_T^2 values when using 'linear' and 'pchip' interpolation, even though the estimated fractional time delays are a little different (0.73 and 0.65). However, the model using this 0.73 day time delay has a marginally higher R^2, suggesting slightly better 24 hour ahead forecasting performance.

The tfrivc routine in the CONTSID Toolbox can only be applied to TF models where the numerator order is less than that of the denominator. The results obtained under this constraint are shown in the bottom two sections of Table 1, which indicate that the rivcbjfd estimated models perform marginally better than the tfrivc estimated model in terms for R_T^2, and that the latter model has a smaller estimated fractional time delay (10.6 compared with 15.4 hours). However, because of its structure, the tfrivc estimated model is less acceptable in terms of its physical interpretation.

2.2.2. A Typical River in the UK

The effective rainfall-flow data in this case are plotted in Fig.5 The analysis for these data is not so detailed and is focussed on showing how the best results are obtained when the rivcbjfd tool is applied together with another, newer routine tool in the Captain

Table 2: Comparison of the integer and fractional time-delay estimation results obtained with the rivcbj, rivcbjfd and tfrivc routines for the effective rainfall-flow data.

Algorithm (Model)	Interpolation	a_1	a_2	b_0	b_1	b_2	τ	R_T^2	R^2
rivcbj ([2 3 1])	integer	0.4240	0.0221	0.0713	0.0310	0.0315	1	0.940	0.981
Standard Error		±0.0494	±0.0076	±0.0058	±0.0159	±0.0108			
rivcbjfd ([2 3 τ])	linear	0.4737	0.0245	0.0864	0.4044	0.0359	0.73	0.953	0.986
Standard Error		±0.0494	±0.0076	±0.0058	±0.0159	±0.0108			
rivcbjfd ([2 3 τ])	pchip	0.4708	0.0242	0.0733	0.4033	0.0354	0.65	0.953	0.985
Standard Error		±0.0478	±0.0075	±0.0059	±0.0150	±0.0106			
rivcbj ([2 2 1])	integer	0.6510	0.0336	0.4244	0.0479	—	1	0.939	0.975
Standard Error		±0.0583	±0.0079	±0.0159	±0.0108				
rivcbjfd ([2 2 τ])	linear	0.6007	0.0343	0.4668	0.0503	—	0.64	0.953	0.983
Standard Error		±0.0500	±0.0085	±0.0152	±0.01203				
rivcbjfd ([2 2 τ])	pchip	0.5723	0.0317	0.4498	0.0464	—	0.60	0.953	0.984
Standard Error		±0.0475	±0.0079	±0.0141	±0.0112				
tfrivc ([2 2 τ])	—	0.4084	0.0223	0.3242	0.0319	—	0.44	0.952	0.984
Standard Error		±0.0350	±0.0044	±0.0116	±0.0062				

Toolbox, rivcbjdd, which identifies and estimates the model in its full form, allowing for the transfer functions associated with each input to have different dynamic characteristics. The second input in this case is a unitary variable (all values equal to 1.0) that allows for the estimation of any initial conditions or related effects [see section A.1.1.3 of the tutorial appendix in 9].

First, let us consider the model in its simplest single input form with a fractional time delay on the effective rainfall input. This yields the following model:

$$\begin{aligned} x(t) &= \frac{25.26s^2 + 19.76s + 0.5502}{s^2 + 0.2174s + 0.0022}u(t - 0.46) \\ y(k) &= x(k) + \xi(k) \end{aligned} \tag{8}$$

where $\xi(k)$ is a third order, heteroscedastic AR(3) noise process. The parameter estimates with estimated standard errors in parentheses are as follows:

$$\begin{aligned} &0.2174(0.009);\ 0.0022(0.00019); \\ &25.26(0.773);\ 19.76(0.532);\ 0.550(0.041) \end{aligned} \tag{9}$$

The model explains the data quite well with $R_T^2 = 0.9789$ and the model response is compared with the measured flow in the top panel of Fig.6 Application of the Matlab residue routine to this model shows that the time constants T_i and steady state gains G_i, associated with each component (i=1,3) are as follows::

$$\begin{aligned} &\hat{T}_1 = 4.84;\ \hat{G}_1 = 60.72 \\ &\hat{T}_2 = 92.10;\ \hat{G}_2 = 159.4 \\ &\hat{G}_3 = 25.27(\text{instantaneous effect}) \end{aligned} \tag{10}$$

showing that the partition percentages between the rapid responses (instantantaneous plus quick) and the slow component are 35.05% and 64.95%, respectively.

If now the estimation of the initial conditions and their effects are introduced, the model is estimated as:

$$\begin{aligned} x_1(t) &= \frac{25.69s^2 + 19.34s + 0.482}{s^2 + 0.2087s + 0.00176}u(t - 0.47) \\ x_2(t) &= \frac{3.564s^2 + 1.12s - 0.0054}{s^2 + 0.2087s + 0.00176}u(t) \\ y(k) &= x(k) + \xi(k) \end{aligned} \tag{11}$$

where $\xi(k)$ is a heteroscedastic, second order AR(2) noise process. The parameter estimates with estimated standard errors in parentheses are as follows:

$$\begin{aligned} &0.2087(0.008);\ 0.0018(0.00019); \\ &25.69(0.677);\ 19.34(0.532);\ 0.482(0.036) \\ &3.56(2.27);\ 1.12(0.163);\ -0.0054(0.002) \end{aligned} \tag{12}$$

The model explains the data a little better than the single input model, with $R_T^2 = 0.9813$, and the model response is compared with the measured flow in the middle panel of Fig.6

Finally, the fully unconstrained model is estimated as:

$$\begin{aligned} x_1(t) &= \frac{-14.36s^2 + 14.83s + 0.866}{s^2 + 0.2768s + 0.005914}u(t - 0.47) \\ x_2(t) &= \frac{5.164s^2 - 0.0233s + 0.00029}{s^2 + 0.00167s + 3.225 * 10^{-5}}u(t) \\ y(k) &= x(k) + \xi(k) \end{aligned} \tag{13}$$

where $\xi(k)$ is a third order, heteroscedastic AR(5) noise process. The parameter estimates with estimated standard errors in parentheses are as follows:

$$\begin{aligned} &0.2768(0.012);\ 0.0059(0.00054); \\ &-14.36(1.128);\ 14.83(0.309);\ 0.866(0.067) \\ &0.00167(0.0024);\ 3.225 * 10^{-5}(4.424 * 10^{-6}); \\ &5.164(0.863);\ -0.0233(0.0046);\ 0.00029(4.936 * 10^{-6}) \end{aligned} \tag{14}$$

The model explains the data a little better than the other two models, with $R_T^2 = 0.9839$ and the model response is compared with the measured flow in the bottom panel of Fig.6.The main advantage of this model is that the second, nominally initial condition (IC),

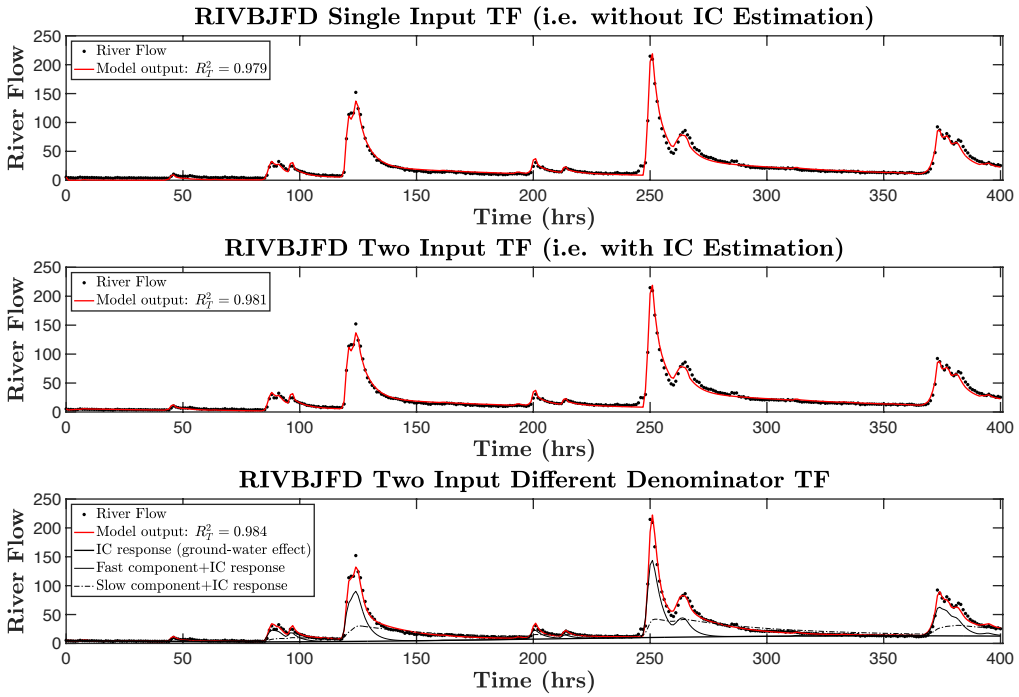

Figure 6: Response of the three models considered in this section: the top panel is the single input model with a fractional input time delay of 0.46 hours; the middle panel introduces a second unitary input to allow for the estimation of initial condition effects; and the lower panel is the full model of this same kind, allowing each TF component to have a different denominator polynomial.

component of the response $x_2(t)$ now explains a very slow component that is probably best interpreted as the effects of groundwater entering through the bed of the river channel. Also shown on the plot are the fast and slow components of the main channel response, added to the very slow IC component.

2.3. A Multi-Input Model for Unemployment in the USA

This example follows from a previous paper [10] that showed how *Relative Government Investment*[2] (RGI) and *Relative Private Investment* (RPI) affected percentage unemployment in the USA between 1948 and 1998. Here, the relative measures of the two inputs were identified using *State-Dependent Parameter* (SDP) estimation [8], which showed that it was necessary to divide both of the measured quarterly investment series by the *Gross National Product* (GNP). In the present example, the integer time delays that were identified as zero in the previous modeling are allowed to be fractional. Although this only improves the explanation of the unemployment series to a small extent, it demonstrates how the rivcbjfd routine is able to handle multiple inputs[3] and provide accurate estimates of small fractional time delays.

The data used in this example are shown in figure 7. The objective is to model these data with a continuous-time model, rather than the discrete-time model used in the previous study, and allow for the possibility of fractional time delays on each input. However, it is clear from the top panel in this plot that the unemployment rate has a large initial condition of 10.9% that must be estimated, together with the initial condition on its derivative, if a reasonable

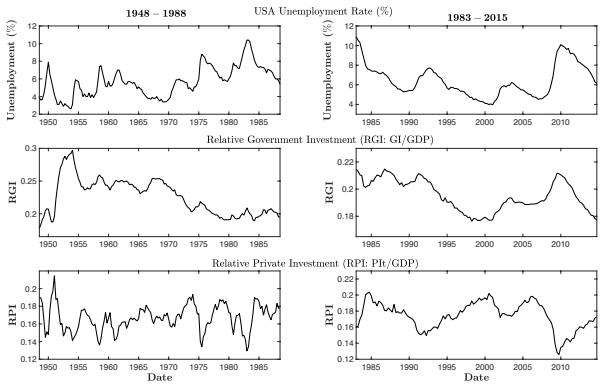

Figure 7: USA macro-economic data 1983-2013: unemployment rate (top panel) and relative measures of government expenditure and private investment below.

model is to be well identified[4]. The simplest approach to solving this problem is to refer to Laplace transform theory and introduce an additional input of unity over the whole observational interval, as outlined briefly in the appendix. Introducing this additional input, the model identified using the rivcbjid routine in CAPTAIN has the following form:

[2] As measured by government expenditure.

[3] Note that the rivcbjfd routine in CAPTAIN is not yet fully evaluated for multiple inputs, so must be used with caution in this regard.

[4] alternative strategies, such as removal of mean values or standardization of the data were tried but did not work as well as this

$$x(t) = \frac{b_{10}p + b_{11}}{s + a_{11}}u_1(t - \tau_1) + \frac{b_{20}p + b_{21}}{s + a_{11}}u_2(t - \tau_2) + \frac{b_{30}p + b_{31}}{s + a_{11}}u_3(t) \quad (15)$$

$$y(t_k) = x(t_k) + \xi(k)$$

where $u_1(t)$ is the RGI input; $u_2(t)$ is the RPI input; and $u_3(t) = 1.0 \; \forall \; t$ is the unity input introduced to handle the initial conditions. Note that there is no time delay on this input because this is associated with the estimation of the initial conditions at $t = 0$ (see above). In this initial model the two pure time delays τ_1 and τ_2 are identified as zero by the rivcbjid routine, where time delays are constrained to integer values.

Based on the initial model (15), the model is re-estimated, allowing for fractional time delays on the two main input series, using the rivcbjfd routine in CAPTAIN (as pointed out above the time delay on the unity input $u_3(t)$ is constrained to zero by not including it in the number of inputs to be considered for a FTD in rivcbjfd). The estimation results are presented in Table 3 where, in order to conserve space, only the results for the main two transfer functions associated with the two investment inputs are presented. As can be seen, the fractional time delay model explains the unemployment data marginally better than the integer delay model but only one significant FTD, with a small duration of 0.99 months, is identified on the RGI input and the FTD on the RPI input is zero). Figure 8 shows the optimization plot in this case, as obtained from rivcbjfd modeling analysis, where we see

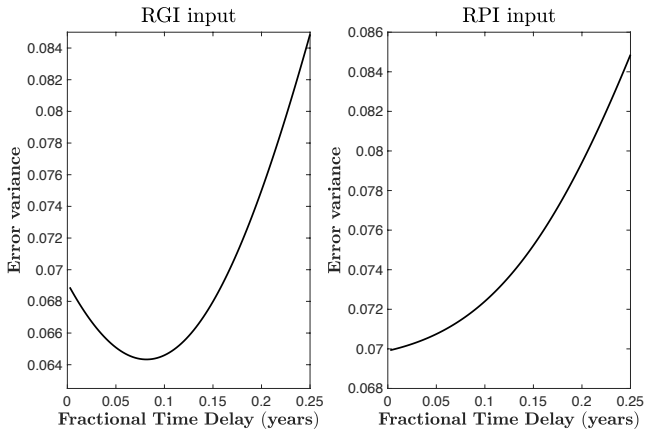

Figure 8: Optimization plots obtained from rivcbjfd modeling analysis.

that there is a clear minimum at 0.0825 years (0.99 months) in the left-hand panel for the RGI input, while there is no minimum for the RPI input in the right-hand panel, showing that the best delay in this case is zero. These results illustrate how the rivcbjfd routine has worked well and discovered the small delay on one input in this relatively difficult multi-input example, with a large initial condition on the output measurement. The table also shows that 'linear' interpolation produces marginally better results than 'pchip' interpolation in this example.

Finally, note that the rivcbjfd routine used in the above analysis was applied with the SRIVC option in (output error minimization) rather than full RIVC (prediction error minimization) as this produced superior results in this case. The noise model was then estimated subsequently based on the output error of the model and the AIC identified an AR(5) model. This two-step procedure has advantages in certain circumstances, such as this, where the theoretical assumptions for optimal maximum likelihood estimation using prediction error minimization are not completely satisfied. These require that the residual error series is a zero mean, serially uncorrelated sequence of random variables with a normal distribution. While statistical tests show that the error series satisfies the first two requirements, it is clearly does not have a normal distribution, as shown in figure 9.

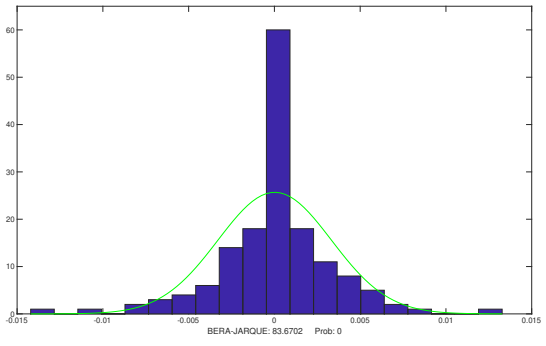

Figure 9: Histogram of residual series $e(k)$ showing its non-normal distribution.

The above situation can be encountered in practical applications where the time series, as in this example, are not obtained by planned experimentation but by observing the system's response to naturally occurring inputs during its normal operation. This is discussed in the epilogue *Good, Bad or Optimal?* of the first author's 2011 book [8], where this two step estimation procedure is referred to as the SRIVC-ARMA algorithm. However, one consequence of this is that the estimated covariance matrix is no longer a totally reliable measure of the parametric uncertainty and the standard error estimates must be considered with this in mind. In this case, they appear to make reasonable sense when used in Monte Carlo Simulation (MCS) analysis, as outlined below.

Finally, the validation ('ex-post forecasting') results are shown in figure 10 (next page). These are the result of Monte Carlo simulation analysis, where the FTD model estimated from the first 80 samples (marked by the vertical black dash-dot line) is used to forecast ahead to the end of the data. It is clear that the model, using only the measured RGI and RPI series as inputs and with no reference to the measured unemployment percentage over this period, continues to explain the data very well up to the vertical blue dash-dot line, which marks the start of the 2008 'Great Recession'. Not surprisingly after 2008, the explanation of the data is not quite so good, although quite sill acceptable. However, it becomes worse after 2015 (not shown) because the macro-economic system has changed by this date with the dominance of the 'Free Market'. This led Robert Reich, in his notable 2015 book 'Saving Capitalism' [**?**], to conclude that, 'the rich are becoming ever-more rich, the poor are getting poorer all the time'. This is clear from the recursive parameter estimates, (not shown), that have converged by 2008 but then start to change after this, with the changes larger still after 2015.

Table 3: Comparison of the integer and fractional time-delay estimation results obtained with the rivcbj, rivcbjfd routines for the macro-economic data.

Algorithm (Model)	Interpolation	a_{11}	b_{10}	b_{11}	b_{20}	b_{21}	τ (months)	R_T^2	R^2
rivcbj ([1 2 2 2])	integer	0.215	141.16	-5.798	-9.921	-44.11	[0 0]	0.9765	0.9929
Standard Error		±0.0198	±7.983	±1.786	±4.525	±1.783			
rivcbjfd ([1 2 2 2 τ_1 τ_2 τ_3])	linear	0.216	139.27	-6.176	-10.786	-43.07	[0.99 0.03]	0.9800	0.9948
Standard Error		±0.0183	±7.315	±1.639	±4.15	±1.610			
rivcbjfd ([1 2 2 2 τ_1 τ_2 τ_3])	pchip	0.216	138.27	-6.271	11.026	-43.047	[0.99 0.03]	0.9793	0.9942
Standard Error		±0.0187	±7.393	±1.663	±4.214	±1.636			

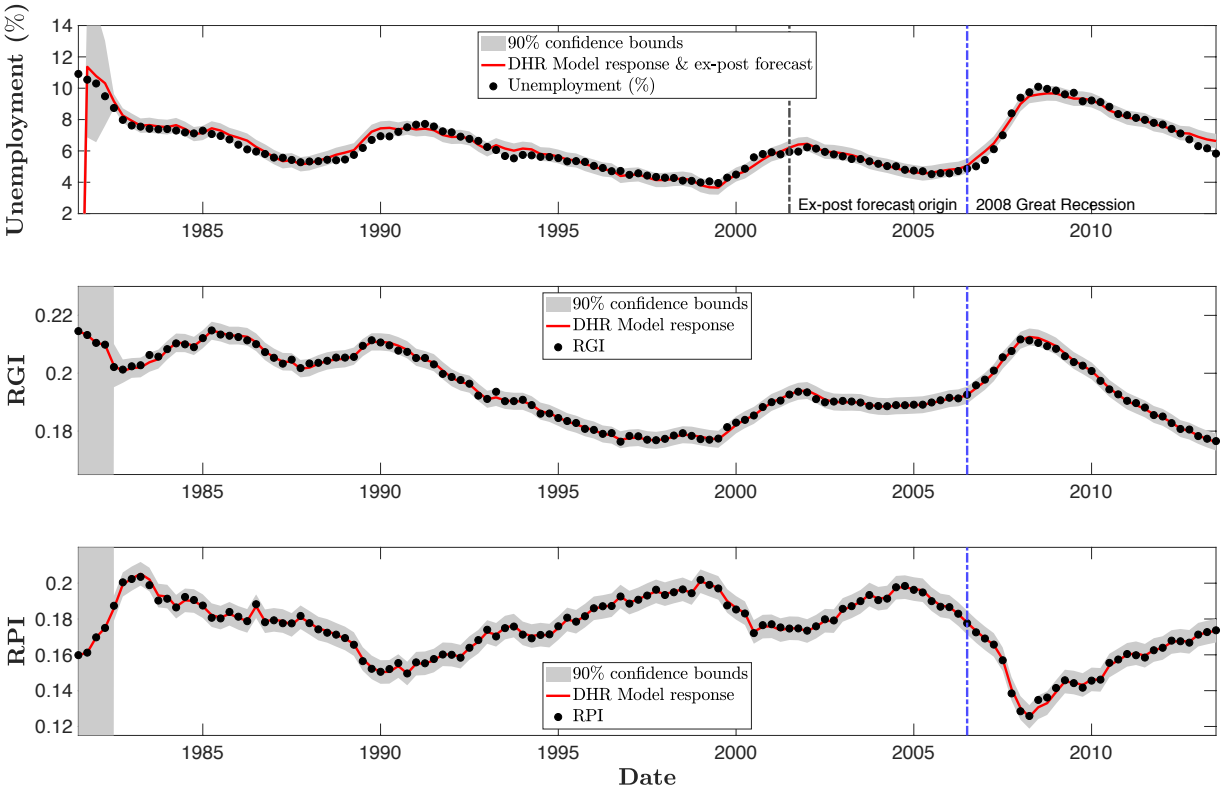

Figure 10: Validation responses using model estimated on data up to 2003 (i.e. ex-post forecasts from 2003 to 2015).

3. Conclusions

This report is primarily intended to provide additional material for the associated Automatica paper [2] concentrating on the provision of three practical examples: a re-examination of the Box-Jenkins gas furnace data using fractional delay model; showing how the introduction of a fractional delay affects the modeling and forecasting of flow in a river; and finally, the estimation and validation of a multi-input model for unemployment in the USA. These examples demonstrate that the proposed method of fractional delay estimation described in the paper, and implemented in the CAPTAIN Toolbox routine rivcbjfd, works well on these real data, reflecting the results obtained with other practical applications of this method over the past year.

Appendix: Estimating Initial Condition Effects in RIVC Estimation

A simple way[5] to illustrate how initial conditions can be accounted for in RIVC estimation is to consider the following first order differential equation model:

$$\frac{dx(t)}{dt} + a_1 x(t) = b_0 \frac{du(t)}{dt} + b_1 u(t) \tag{16}$$

The Laplace transform of this model (see any good control system text: e.g. [4]) allows for the incorporation of the initial conditions and yields the following equation *in the Laplace domain*:

$$sx(s) + x(0) + a_1 x(s) = b_0 su(s) + u(0) + b_1 u(s) \tag{17}$$

where 's' is the *Laplace operator*, while $x(0)$ and $u(0)$ are the initial conditions on $x(t)$ and $u(t)$, respectively. Consequently, reverting to the *differential operator* form of the equation, (where now $p = \frac{d}{dt}$), the associated CTF model is as follows:

$$x(t) = \frac{b_0 p + b_1}{p + a_1} u(t) + \frac{u(0) - x(0)}{p + a_1} 1.0 \tag{18}$$

As a result, it is straightforward to incorporate an allowance for initial condition effects into the RIVC algorithm by adding an additional constant input set to unity over the whole observational interval. The initial conditions are then estimated as the parameter(s) of the second TF in (18), which is associated with this additional input (here, a single parameter defined by $u(0) - x(0)$). The output of this second TF adds a response to the output that accounts for the effects of the initial conditions and so ensures a better explanation of the data over the time it takes for the effects of the initial conditions to decay.

References

[1] Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis Forecasting and Control*. Holden-Day: San Francisco.

[2] Chen, F. and Young, P. C. (2022). A simple robust method of fractional time-delay estimation for linear dynamic systems. *Automatica*, 137:110117.

[3] Chen, F., Zhuan, X., Garnier, H., and Gilson., M. (2018). Issues in separable identification of continuous-time models with time-delay. *Automatica*, 94:258–273.

[4] Kuo, B. C. and Golnaraghi, F. (2002). *Automatic Control Systems*. John Wiley & Sons, Inc., New York, NY, USA.

[5] Martinez, C. J. and Wise, W. R. (2003). Analysis of constructed treatment wetland hydraulics with the transient storage model OTIS. *Ecological Engineering*, 20(3):211 – 222.

[6] Young, P. C. (2002). Identification of time varying systems. In Unbehauen, H., editor, *Encyclopedia of Life Support Systems (EOLSS)*, volume 6.43: Control Systems, Robotics and Automation. EOLSS Publishers: Oxford.

[7] Young, P. C. (2008). Real-time flow forecasting. In Wheater, H., Sarooshian, S., and Sharma, K. D., editors, *Hydrological modelling in arid and semi-arid areas*, pages 113–138, Cambridge, UK. Cambridge University Press.

[8] Young, P. C. (2011). *Recursive Estimation and Time-Series Analysis: An Introduction for the Student and Practitioner*. Springer-Verlag, Berlin.

[9] Young, P. C. (2026). *The Art and Craft of Data-Based Mechanistic Modelling, Forecasting and Control*. Springer Nature.

[10] Young, P. C. and Pedregal, D. J. (1999). Macro-economic relativity: government spending, private capital investment and unemployment in the USA 1948-1998. *Structural Change and Economic Dynamics*, 10:359–380.

[5] Note that the analysis in this section is not a complete application of the Laplace transform. Normally, inverse Laplace transformation is incorporated after equation (17), above, to yield the *solution* of the differential equation in the time domain, whereas here we are only interested in the CTF form.