



HAL
open science

Missing data in time series: A note on the equivalence of the dummy variable and the skipping approaches

Tommaso Proietti

► **To cite this version:**

Tommaso Proietti. Missing data in time series: A note on the equivalence of the dummy variable and the skipping approaches. *Statistics and Probability Letters*, 2009, 78 (3), pp.257. 10.1016/j.spl.2007.05.031 . hal-00594451

HAL Id: hal-00594451

<https://hal.science/hal-00594451>

Submitted on 20 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

Missing data in time series: A note on the equivalence of the dummy variable and the skipping approaches

Tommaso Proietti

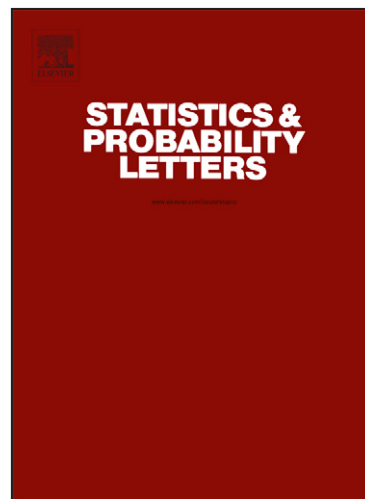
PII: S0167-7152(07)00222-2
DOI: doi:10.1016/j.spl.2007.05.031
Reference: STAPRO 4702

To appear in: *Statistics & Probability Letters*

Received date: 13 March 2006
Revised date: 22 February 2007
Accepted date: 23 May 2007

Cite this article as: Tommaso Proietti, Missing data in time series: A note on the equivalence of the dummy variable and the skipping approaches, *Statistics & Probability Letters* (2007), doi:10.1016/j.spl.2007.05.031

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/stapro

Missing data in time series: a note on the equivalence of the dummy variable and the skipping approaches

Tommaso Proietti

S.E.F. e ME.Q,

University of Rome, *Tor Vergata**

Abstract

This note shows the equivalence of the dummy variable approach and the skipping approach for the treatment of missing observations in state space models. The equivalence holds when the coefficient of the dummy variable is considered as a diffuse rather than a fixed effect. The equivalence concerns both likelihood inference and smoothed inferences.

Key Words: Kalman Filter, Smoothing, Influence, Cross-Validation.

*Address for correspondence: Dipartimento S.E.F. e ME.Q, Via Columbia 2, 00133 Rome (Italy). e-mail: tommaso.proietti@uniroma2.it

1 Introduction

A well known result is that estimating a missing observation by skipping the Kalman filter updating step is equivalent to introducing a dummy variable (additive outlier) in the measurement equation, filling the missing value arbitrarily. This result (in different frameworks) appears in a number of papers: Sargan and Drettakis (1974), Bruce and Martin (1989), Ljung (1993). A detailed discussion can be found in Fuller (1996, section 8.7). However, if the additive outlier is treated as a fixed effect, with zero covariance matrix, the likelihood is defined differently and a correction has to be computed in the second case, see Gómez et al. (1999). The correction factor is related to the determinantal term of the likelihood and depends in a simple fashion from quantities computed under the model for the complete observations, requiring a single run of the Kalman filter and smoothing filter.

To our knowledge, a proof the equivalence of the skipping approach and the dummy approach for the definition of the likelihood and for smoothing is not available. This note aims at bridging the gap, providing a simple proof that when the additive outlier is treated as diffuse, with arbitrarily large covariance matrix, the correction to the likelihood takes place automatically. This is convenient, as no extra programming effort is necessary once a programme handling diffuse initial conditions and regression effects has been implemented.

The equivalence is also carried forward to smoothed inferences, concerning the estimation of the states and the disturbances. The derivation of analytical expressions for the influence of an observation on these quantities, made in De Jong (1996), is greatly simplified in the dummy variable setup as they depend in a simple fashion on the output of the Kalman filter and smoother run on intervention variables.

The plan of the paper is the following: section 2 introduces the dummy variable approach for stationary state space models with no regression effects, under fixed and diffuse conditions, and derives the prediction error decomposition form of the like-

likelihood under the latter. In section 3 we present the alternative strategy of handling missing observations, known as the skipping approach, and prove that the likelihood for this model is equivalent to the dummy variable one. In section 4 the equivalence is extended to smoothed estimates of the states and the disturbances, and measures of influence of an observations are given, which depends in a simple way on the output of the Kalman filter and smoothing filter run on the intervention variable.

2 The Dummy Variable Approach

Let \mathbf{y}_t denote a vector stationary time series with N elements; the state space model is

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{G}_t \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, T, \quad (1)$$

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, T, \quad (2)$$

with $\boldsymbol{\alpha}_1 \sim \text{N}(\mathbf{a}_1, \sigma^2 \mathbf{P}_1)$, where \mathbf{a}_1 and $\sigma^2 \mathbf{P}_1$ denote the unconditional mean and covariance matrix of $\boldsymbol{\alpha}_t$, and $\boldsymbol{\varepsilon}_t \sim \text{NID}(\mathbf{0}, \sigma^2 \mathbf{I})$. The system matrices, $\mathbf{Z}_t, \mathbf{G}_t, \mathbf{T}_t, \mathbf{H}_t$, are functionally related to a vector of hyperparameters, $\boldsymbol{\theta}$.

The Kalman filter (KF) is a well-known recursive algorithm for computing the minimum mean square estimator of $\boldsymbol{\alpha}_t$ and its mean square error (MSE) matrix conditional on $\mathbf{Y}_{t-1} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{t-1}\}$. Defining

$$\mathbf{a}_t = \text{E}(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}), \quad \text{MSE}(\mathbf{a}_t) = \sigma^2 \mathbf{P}_t = \text{E}[(\boldsymbol{\alpha}_t - \mathbf{a}_t)(\boldsymbol{\alpha}_t - \mathbf{a}_t)' | \mathbf{Y}_{t-1}],$$

the filter consists of the following recursions:

$$\begin{aligned} \boldsymbol{\nu}_t &= \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t, & \mathbf{F}_t &= \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}'_t + \mathbf{G}_t \mathbf{G}'_t \\ q_t &= q_{t-1} + \boldsymbol{\nu}'_t \mathbf{F}_t^{-1} \boldsymbol{\nu}_t, & \mathbf{K}_t &= (\mathbf{T}_t \mathbf{P}_t \mathbf{Z}'_t + \mathbf{H}_t \mathbf{G}'_t) \mathbf{F}_t^{-1} \\ \mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t \boldsymbol{\nu}_t, & \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t \mathbf{L}'_t + \mathbf{H}_t \mathbf{J}'_t \end{aligned} \quad (3)$$

with $\mathbf{L}_t = \mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t$ and $\mathbf{J}_t = \mathbf{H}_t - \mathbf{K}_t \mathbf{G}_t$; $\boldsymbol{\nu}_t = \mathbf{y}_t - \text{E}(\mathbf{y}_t | \mathbf{Y}_{t-1})$ are the filter innovations, with MSE matrix $\sigma^2 \mathbf{F}_t$. The filter is started off with $\mathbf{a}_1 = \mathbf{0}$, $\mathbf{P}_1 = \mathbf{H}_0 \mathbf{H}'_0$ and

$q_0 = 0$. The log-likelihood for the model is, apart from a constant term,

$$\mathcal{L}(\mathbf{y}_1, \dots, \mathbf{y}_T; \boldsymbol{\theta}) = -\frac{1}{2} \left[NT \ln \sigma^2 + \sum_{t=1}^T \ln |\mathbf{F}_t| + \sigma^{-2} q_T \right] \quad (4)$$

where $q_T = \sum_{t=1}^T \boldsymbol{\nu}_t' \mathbf{F}_t^{-1} \boldsymbol{\nu}_t$.

Suppose that an intervention is included at $t = i$ so that the measurement equation becomes

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{I}_t(i) \boldsymbol{\delta} + \mathbf{G}_t \boldsymbol{\varepsilon}_t, \quad (5)$$

where $\mathbf{I}_t(i)$ is an indicator variable taking value 1 for $t = i$ and 0 elsewhere. For its statistical treatment, the KF (3) at $t = i$ is augmented by the following recursions:

$$\begin{aligned} \mathbf{V}_t^+ &= \mathbf{I}_t(i) \mathbf{I} - \mathbf{Z}_t \mathbf{A}_t^+ \\ \mathbf{A}_{t+1}^+ &= \mathbf{T}_t \mathbf{A}_t^+ + \mathbf{K}_t \mathbf{V}_t^+ = \mathbf{K}_i \mathbf{I}_t(i) + \mathbf{L}_t \mathbf{A}_t^+ \\ \mathbf{S}_t^+ &= \mathbf{S}_{t-1}^+ + \mathbf{V}_t^+ \mathbf{F}_t^{-1} \mathbf{V}_t^+, \\ \mathbf{s}_t^+ &= \mathbf{s}_{t-1}^+ + \mathbf{V}_t^+ \mathbf{F}_t^{-1} \boldsymbol{\nu}_t, \end{aligned} \quad (6)$$

for $t = i, \dots, T$ with starting conditions: $\mathbf{A}_i^+ = \mathbf{0}$, $\mathbf{S}_{i-1}^+ = \mathbf{0}$ and $\mathbf{s}_{i-1}^+ = \mathbf{0}$. This amounts to apply the KF to the intervention signature $\mathbf{I}_t(i) \mathbf{I}$.

When $\boldsymbol{\delta}$ is treated as a fixed effect, the log-likelihood can be written as (Rosenberg, 1973)

$$-\frac{1}{2} \left[NT \ln \sigma^2 + \sum_{t=1}^T \ln |\mathbf{F}_t| + \sigma^{-2} (q_T - 2 \mathbf{s}_T^+ \boldsymbol{\delta} + \boldsymbol{\delta}' \mathbf{S}_T^+ \boldsymbol{\delta}) \right].$$

The MLE of $\boldsymbol{\delta}$ is thus $\hat{\boldsymbol{\delta}} = \mathbf{S}_T^+{}^{-1} \mathbf{s}_T^+$ and the concentrated likelihood is

$$-\frac{1}{2} \left[NT \ln \sigma^2 + \sum_{t=1}^T \ln |\mathbf{F}_t| + \sigma^{-2} (q_T - \mathbf{s}_T^+ \mathbf{S}_T^+{}^{-1} \mathbf{s}_T^+) \right].$$

There is however a conceptual difficulty with the fixed effects model, as was clearly pointed out by Bell (1989, p. 408), in that "the use of an indicator variable lets the mean at a given point be anything while still assuming that the observation is normal with the same variance and covariances as other observations, whereas omitting observations makes no assumption at all about it".

In the sequel, δ is treated as a diffuse effect, that is $[\text{Cov}(\delta)]^{-1}$ converges to zero in the Euclidean norm (see De Jong, 1991), e.g. $\delta \sim N(\mathbf{0}, \kappa \mathbf{I})$, $\kappa \rightarrow \infty$; this is equivalent to making no assumption on the covariance of the i -th observation.

De Jong (1991) has shown that δ can be concentrated out of the likelihood function, so that $\hat{\delta} = \mathbf{S}_T^{+^{-1}} \mathbf{s}_T^+$ and $\text{MSE}(\hat{\delta}) = \sigma^2 \mathbf{S}_T^{+^{-1}}$. The diffuse log-likelihood function is

$$\begin{aligned} \mathcal{L}^{DV}(\mathbf{y}_1, \dots, \mathbf{y}_T; \theta) = & -\frac{1}{2}[N(T-1) \ln \sigma^2 + \sum_{t=1}^T \ln |\mathbf{F}_t| + \\ & \ln |\mathbf{S}_T^+| + \sigma^{-2}(q_T - \mathbf{s}_T^{+'} \mathbf{S}_T^{+^{-1}} \mathbf{s}_T^+)]. \end{aligned} \quad (7)$$

This function is the likelihood for a rank $T-1$ transformation of the observations, which makes the data invariant to δ .

The following theorem is a restatement of theorem 2 in De Jong and Penzer (1998).

Theorem 1. The estimate of δ and the diffuse LF can be written as

$$\hat{\delta} = \mathbf{M}_i^{-1} \mathbf{u}_i, \quad \text{MSE}(\hat{\delta}) = \sigma^2 \mathbf{M}_i^{-1}, \quad (8)$$

$$\mathcal{L}^{DV} = -\frac{1}{2}[N(T-1) \ln \sigma^2 + \sum_{i=1}^T \ln |\mathbf{F}_i| + \ln |\mathbf{M}_i| + \sigma^{-2}(q_T - \mathbf{u}_i' \mathbf{M}_i^{-1} \mathbf{u}_i)] \quad (9)$$

where \mathbf{u}_i and \mathbf{M}_i are the output at $t = i$ of the smoothing filter:

$$\begin{aligned} \mathbf{u}_t &= \mathbf{F}_t^{-1} \boldsymbol{\nu}_t - \mathbf{K}_t' \mathbf{r}_t & \mathbf{M}_t &= \mathbf{F}_t^{-1} + \mathbf{K}_t' \mathbf{N}_t \mathbf{K}_t \\ \mathbf{r}_{t-1} &= \mathbf{Z}_t' \mathbf{F}_t^{-1} \boldsymbol{\nu}_t + \mathbf{L}_t' \mathbf{r}_t & \mathbf{N}_{t-1} &= \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{Z}_t + \mathbf{L}_t' \mathbf{N}_t \mathbf{L}_t \end{aligned} \quad (10)$$

started with $\mathbf{r}_T = \mathbf{0}$ and $\mathbf{N}_T = \mathbf{0}$.

Proof. We begin by noting that $\mathbf{V}_i^+ = \mathbf{I}$ and $\mathbf{V}_t^+ = -\mathbf{Z}_t \mathbf{L}_{t,i+1} \mathbf{K}_i$ for $t = i+1, \dots, T$ with $\mathbf{L}_{t,i+1} = \mathbf{L}_{t-1} \cdots \mathbf{L}_{i+1}$ and $\mathbf{L}_{i+1,i+1} = \mathbf{I}$. Hence:

$$\begin{aligned} \mathbf{s}_T^+ &= \sum_{t=i}^T \mathbf{V}_t^{+'} \mathbf{F}_t^{-1} \boldsymbol{\nu}_t \\ &= \mathbf{F}_i^{-1} \boldsymbol{\nu}_i - \mathbf{K}_i' \sum_{t=i+1}^T \mathbf{L}_{t,i+1}' \mathbf{Z}_t' \mathbf{F}_t^{-1} \boldsymbol{\nu}_t \\ &= \mathbf{F}_i^{-1} \boldsymbol{\nu}_i - \mathbf{K}_i' \mathbf{r}_i \\ &= \mathbf{u}_i \end{aligned}$$

$$\begin{aligned}
\mathbf{S}_T^+ &= \sum_{t=i}^T \mathbf{V}_t^+ \mathbf{F}_t^{-1} \mathbf{V}_t^+ \\
&= \mathbf{F}_i^{-1} + \mathbf{K}'_i \left(\sum_{t=i+1}^T \mathbf{L}'_{t,i+1} \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{L}_{t,i+1} \right) \mathbf{K}_i \\
&= \mathbf{F}_i^{-1} + \mathbf{K}'_i \mathbf{N}_i \mathbf{K}_i \\
&= \mathbf{M}_i
\end{aligned}$$

Replacing into the expressions for $\hat{\delta}$ and (7) yields the result. ■

Using a different argument, De Jong and Penzer (1998) show that

$$\mathbf{y}_t - \mathbb{E}(\mathbf{y}_t | \mathbf{y}_1, \dots, \mathbf{y}_{t-1}, \mathbf{y}_{t+1}, \dots, \mathbf{y}_T) = \mathbf{M}_t^{-1} \mathbf{u}_t.$$

The next theorem provides an alternative expression for the likelihood function, based on the one-step-ahead prediction error decomposition. This will prove useful in the comparison with that arising from the skipping approach.

Theorem 2. For model (5), let $\hat{\mathbf{F}}_t = \text{MSE}(\boldsymbol{\nu}_t | \mathbf{Y}_{t-1})$, and $\hat{\boldsymbol{\nu}}_t = \mathbb{E}(\boldsymbol{\nu}_t | \mathbf{Y}_{t-1})$, where $\hat{\mathbf{F}}_t = \mathbf{F}_t + \mathbf{V}_t^+ \mathbf{S}_{t-1}^{+^{-1}} \mathbf{V}_t^+$ and $\hat{\boldsymbol{\nu}}_t = \boldsymbol{\nu}_t - \mathbf{V}_t^+ \mathbf{S}_{t-1}^{+^{-1}} \mathbf{s}_{t-1}$. Then,

$$\begin{aligned}
\mathcal{L}^{DV} &= -\frac{1}{2} \left[N(T-1) \ln \sigma^2 + \sum_{t=1}^{i-1} \ln |\mathbf{F}_t| + \sum_{t=i+1}^T \ln |\hat{\mathbf{F}}_t| + \right. \\
&\quad \left. \sigma^{-2} \left(q_{i-1} + \sum_{t=i+1}^T \hat{\boldsymbol{\nu}}_t' \hat{\mathbf{F}}_t^{-1} \hat{\boldsymbol{\nu}}_t \right) \right]. \tag{11}
\end{aligned}$$

Proof. To show that the determinantal part of the LF is as stated we provide the following recursion for $|\mathbf{S}_T^+|$:

$$\begin{aligned}
|\mathbf{S}_T^+| &= |\mathbf{S}_{T-1}^+ + \mathbf{V}_T^+ \mathbf{F}_T^{-1} \mathbf{V}_T^+| \\
&= |\mathbf{S}_{T-1}^+| |\mathbf{I} + \mathbf{S}_{T-1}^{+^{-1}} \mathbf{V}_T^+ \mathbf{F}_T^{-1} \mathbf{V}_T^+| \\
&= |\mathbf{S}_{T-1}^+| |\mathbf{I} + \mathbf{F}_T^{-1} \mathbf{V}_T^+ \mathbf{S}_{T-1}^{+^{-1}} \mathbf{V}_T^+| \\
&= |\mathbf{S}_{T-1}^+| |\mathbf{F}_T|^{-1} |\hat{\mathbf{F}}_T|
\end{aligned}$$

Iterating this result for $t = T-1, T-2, \dots, i+1$ and recalling that $\mathbf{S}_i^+ = \mathbf{F}_i^{-1}$ produces:

$$\ln |\mathbf{S}_T^+| = \sum_{t=i+1}^T \ln |\hat{\mathbf{F}}_t| - \sum_{t=i}^T \ln |\mathbf{F}_t|.$$

Moreover,

$$q_T - \mathbf{s}'_T \mathbf{S}_T^{+-1} \mathbf{s}_T^+ = q_{T-1} - \mathbf{s}'_{T-1} \mathbf{S}_{T-1}^{+-1} \mathbf{s}_{T-1}^+ + \hat{\boldsymbol{\nu}}'_T \hat{\mathbf{F}}_T^{-1} \hat{\boldsymbol{\nu}}_T$$

which, applied recursively, yields

$$q_T - \mathbf{s}'_T \mathbf{S}_T^{+-1} \mathbf{s}_T^+ = q_i - \mathbf{s}'_i \mathbf{S}_i^{+-1} \mathbf{s}_i^+ + \sum_{t=i+1}^T \hat{\boldsymbol{\nu}}'_t \hat{\mathbf{F}}_t^{-1} \hat{\boldsymbol{\nu}}_t.$$

Now, as $q_i = q_{i-1} + \boldsymbol{\nu}'_i \mathbf{F}_i^{-1} \boldsymbol{\nu}_i$ and $\mathbf{s}'_i \mathbf{S}_i^{+-1} \mathbf{s}_i^+ = \boldsymbol{\nu}'_i \mathbf{F}_i^{-1} \boldsymbol{\nu}_i$, result (11) follows directly. ■

3 The Skipping Approach

When the i -th observation is missing, the KF is forced to skip the updating step at time $t = i$, so that

$$\mathbf{a}_{i+1}^{(m)} = \mathbf{T}_i \mathbf{a}_i, \quad \mathbf{P}_{i+1}^{(m)} = \mathbf{T}_i \mathbf{P}_i \mathbf{T}'_i + \mathbf{H}_i \mathbf{H}'_i, \quad (12)$$

and $q_i^{(m)} = q_{i-1}$.

From time $i + 1$ on the KF (3) is run with $\boldsymbol{\nu}_t, \mathbf{a}_t, q_t, \mathbf{K}_t, \mathbf{P}_{t+1}, \mathbf{L}_t, \mathbf{J}_t$ replaced by $\boldsymbol{\nu}_t^{(m)}, \mathbf{a}_t^{(m)}, q_t^{(m)}, \mathbf{K}_t^{(m)}, \mathbf{P}_{t+1}^{(m)}, \mathbf{L}_t^{(m)}, \mathbf{J}_t^{(m)}$. See Harvey *et al.* (1998).

The log-likelihood function $\mathcal{L}^{(m)} = \mathcal{L}(\mathbf{y}_1, \dots, \mathbf{y}_{i-1}, \mathbf{y}_{i+1}, \dots, \mathbf{y}_T; \boldsymbol{\theta})$ is:

$$\begin{aligned} \mathcal{L}^{(m)} = & -\frac{1}{2} [N(T-1) \ln \sigma^2 + \sum_{t=1}^{i-1} \ln |\mathbf{F}_t| + \sum_{t=i+1}^T \ln |\mathbf{F}_t^{(m)}| + \\ & \sigma^{-2} \left(\sum_{t=1}^{i-1} \boldsymbol{\nu}'_t \mathbf{F}_t^{-1} \boldsymbol{\nu}_t + \sum_{t=i+1}^T \boldsymbol{\nu}'_t^{(m)} \mathbf{F}_t^{(m)-1} \boldsymbol{\nu}_t^{(m)} \right)]. \end{aligned} \quad (13)$$

Theorem 3. $\mathcal{L}^{(m)} = \mathcal{L}^{DV}$

Proof. The KF resulting from the skipping approach is related to the full sample KF (3) by the following equations:

$$\begin{aligned} \boldsymbol{\nu}_t^{(m)} &= \boldsymbol{\nu}_t - \mathbf{V}_t^+ \mathbf{S}_{t-1}^{+-1} \mathbf{s}_{t-1}^+, & \mathbf{F}_t^{(m)} &= \mathbf{F}_t + \mathbf{V}_t^+ \mathbf{S}_{t-1}^{+-1} \mathbf{V}_t^+, \\ \mathbf{K}_t^{(m)} &= \mathbf{K}_t - \mathbf{A}_{t+1}^+ \mathbf{S}_t^{+-1} \mathbf{V}_t^+ \mathbf{F}_t^{-1}, & & \\ \mathbf{a}_{t+1}^{(m)} &= \mathbf{a}_{t+1} - \mathbf{A}_{t+1}^+ \mathbf{S}_t^{+-1} \mathbf{s}_t^+, & \mathbf{P}_{t+1}^{(m)} &= \mathbf{P}_{t+1} + \mathbf{A}_{t+1}^+ \mathbf{S}_t^{+-1} \mathbf{A}_{t+1}^+, \end{aligned} \quad (14)$$

where $\mathbf{A}_t^+ = \mathbf{L}_{t,i+1}\mathbf{K}_i$. These relations hold for $t = i + 1$: from (12),

$$\mathbf{a}_{i+1}^{(m)} = \mathbf{a}_{i+1} - \mathbf{K}_i\boldsymbol{\nu}_i = \mathbf{a}_{i+1} - \mathbf{A}_{i+1}^+\mathbf{S}_i^{+,-1}\mathbf{s}_i^+$$

and

$$\mathbf{P}_{i+1}^{(m)} = \mathbf{P}_{i+1} + \mathbf{K}_i\mathbf{F}_i\mathbf{K}_i' = \mathbf{P}_{i+1} + \mathbf{A}_{i+1}^+\mathbf{S}_i^{+,-1}\mathbf{A}_{i+1}'.$$

Hence

$$\boldsymbol{\nu}_{i+1}^{(m)} = \mathbf{Z}_{i+1}\mathbf{a}_{i+1}^{(m)} = \boldsymbol{\nu}_{i+1} - \mathbf{Z}_{i+1}\mathbf{A}_{i+1}^+\mathbf{S}_i^{+,-1}\mathbf{s}_i^+ = \boldsymbol{\nu}_{i+1} - \mathbf{V}_{i+1}^+\mathbf{S}_i^{+,-1}\mathbf{s}_i^+$$

and

$$\mathbf{F}_{i+1}^{(m)} = \mathbf{Z}_{i+1}\mathbf{P}_{i+1}^{(m)}\mathbf{Z}_{i+1}' + \mathbf{G}_{i+1}\mathbf{G}_{i+1}' = \mathbf{F}_{i+1} + \mathbf{V}_{i+1}^+\mathbf{S}_i^{+,-1}\mathbf{V}_{i+1}'.$$

The formula for the gain matrix is obtained noticing that

$$\mathbf{F}_{i+1}^{(m)-1} = \mathbf{F}_{i+1}^{-1} - \mathbf{F}_{i+1}^{-1}\mathbf{V}_{i+1}^+\mathbf{S}_{i+1}^{+,-1}\mathbf{V}_{i+1}'\mathbf{F}_{i+1}^{-1}$$

whence

$$\begin{aligned} \mathbf{K}_{i+1}^{(m)} &= (\mathbf{T}_{i+1}\mathbf{P}_{i+1}^{(m)}\mathbf{Z}_{i+1}' + \mathbf{H}_{i+1}\mathbf{G}_{i+1}')\mathbf{F}_{i+1}^{(m)-1} \\ &= (\mathbf{K}_{i+1}\mathbf{F}_{i+1} - \mathbf{T}_{i+1}\mathbf{K}_i\mathbf{F}_i\mathbf{V}_{i+1}'^{(m)})\mathbf{F}_{i+1}^{(m)-1} \\ &= \mathbf{K}_{i+1} - \mathbf{A}_{i+2}^+\mathbf{S}_{i+1}^{+,-1}\mathbf{V}_{i+1}'\mathbf{F}_{i+1}^{-1} \end{aligned}$$

In conclusion, $\boldsymbol{\nu}_t^{(m)} = \hat{\boldsymbol{\nu}}_t$ and $\mathbf{F}_t^{(m)} = \hat{\mathbf{F}}_t$. ■

When δ is treated as a fixed effect, the correction that has to be applied to the determinantal part of the likelihood is $-.5 \ln |\mathbf{M}_i|$, which is available from a run of the smoothing filter (10).

4 Influence and deletion diagnostics

In this section we use the previous results in a different perspective. Assuming that the full sample is available we aim at computing measures of influence on smoothed

inferences. De Jong (1989) proved that the smoothed estimate of the state at t , $\tilde{\mathbf{a}}_t = \mathbb{E}(\boldsymbol{\alpha}_t | \mathbf{Y}_T)$, and its MSE matrix, $\sigma^2 \tilde{\mathbf{P}}_t = \mathbb{E}[(\boldsymbol{\alpha}_t - \tilde{\mathbf{a}}_t)(\boldsymbol{\alpha}_t - \tilde{\mathbf{a}}_t)' | \mathbf{Y}_T]$, are

$$\tilde{\mathbf{a}}_t = \mathbf{a}_t + \mathbf{P}_t \mathbf{r}_{t-1}, \quad \tilde{\mathbf{P}}_t = \mathbf{P}_t - \mathbf{P}_t \mathbf{N}_{t-1} \mathbf{P}_t,$$

where \mathbf{r}_{t-1} and \mathbf{N}_{t-1} are given in the second line of (10).

Now, let $\tilde{\mathbf{a}}_t^{(m)} = \mathbb{E}(\boldsymbol{\alpha}_t | \mathbf{Y}_T^{(i)})$, where $\mathbf{Y}_T^{(i)} = (\mathbf{y}_1, \dots, \mathbf{y}_{i-1}, \mathbf{y}_{i+1}, \dots, \mathbf{y}_T)$ is the information at T excluding \mathbf{y}_i . Moreover, let $\tilde{\mathbf{P}}_t^{(m)} = \mathbb{E}[(\boldsymbol{\alpha}_t - \tilde{\mathbf{a}}_t^{(i)})(\boldsymbol{\alpha}_t - \tilde{\mathbf{a}}_t^{(m)})' | \mathbf{Y}_T^{(m)}]$.

Theorem 4.

$$\begin{aligned} \tilde{\mathbf{a}}_t &= \tilde{\mathbf{a}}_t^{(m)} + (\mathbf{A}_t^+ + \mathbf{P}_t \mathbf{R}_{t-1}^+) \mathbf{M}_i^{-1} \mathbf{u}_i, \\ \tilde{\mathbf{P}}_t &= \tilde{\mathbf{P}}_t^{(m)} - (\mathbf{A}_t^+ + \mathbf{P}_t \mathbf{R}_{t-1}^+) \mathbf{M}_i^{-1} (\mathbf{A}_t^+ + \mathbf{P}_t \mathbf{R}_{t-1}^+)', \end{aligned}$$

where

$$\mathbf{R}_{t-1}^+ = \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{V}_t^+ + \mathbf{L}_t' \mathbf{R}_t^+, \quad (15)$$

with $\mathbf{A}_t^+ = \mathbf{0}$ for $t < i + 1$. Also, for $t < i$, $\mathbf{V}_t^+ = \mathbf{0}$ and $\mathbf{R}_{t-1}^+ = \mathbf{L}_t' \mathbf{R}_t^+ = \mathbf{L}_{i,t}' \mathbf{R}_{i-1}^+$.

Proof. The orthogonal set $\{\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_T\}$ is a linear transformation of the set

$$\{\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_{i-1}, \boldsymbol{\nu}_{i+1}^{(m)}, \dots, \boldsymbol{\nu}_T^{(m)}, \mathbf{u}_i\}.$$

This set is orthogonal too, since \mathbf{u}_i and $\boldsymbol{\nu}_t^{(m)}$ depend only on $\{\boldsymbol{\nu}_j, j \geq i\}$, and

$$\text{Cov}(\boldsymbol{\nu}_t^{(m)}, \mathbf{u}_i) = \text{Cov}(\boldsymbol{\nu}_t - \mathbf{V}_t^+ \mathbf{S}_{t-1}^{+-1} \mathbf{s}_{t-1}^+, \sum_i^T \mathbf{V}_t^+ \mathbf{F}_t^{-1} \boldsymbol{\nu}_t) = \mathbf{0}, \forall t > i.$$

Thus, applying a standard result in uncorrelated linear projection:

$$\begin{aligned} \tilde{\mathbf{a}}_t &= \tilde{\mathbf{a}}_t^{(m)} + \sigma^{-2} \text{Cov}(\boldsymbol{\alpha}_t - \mathbf{a}_t^{(m)}, \mathbf{u}_i) \mathbf{M}_i^{-1} \mathbf{u}_i \\ &= \tilde{\mathbf{a}}_t^{(m)} + \sigma^{-2} \text{Cov}(\boldsymbol{\alpha}_t - \mathbf{a}_t + \mathbf{A}_t^+ \mathbf{S}_{t-1}^{+-1} \mathbf{s}_{t-1}^+, \mathbf{u}_i) \mathbf{M}_i^{-1} \mathbf{u}_i \\ &= \tilde{\mathbf{a}}_t^{(m)} + \left(\mathbf{P}_t \sum_{j=i}^T \mathbf{L}_{t,j}' \mathbf{Z}_j' \mathbf{F}_j^{-1} \mathbf{V}_j^+ + \mathbf{A}_t^+ \right) \mathbf{M}_i^{-1} \mathbf{u}_i \end{aligned}$$

The formula for the MSE matrix is derived similarly from

$$\text{MSE}(\boldsymbol{\alpha}_t | \mathbf{Y}_T) = \text{MSE}(\boldsymbol{\alpha}_t | \mathbf{Y}_T^{(i)}) - \text{Cov}(\boldsymbol{\alpha}_t, \mathbf{u}_i) \mathbf{M}_i^{-1} \text{Cov}(\boldsymbol{\alpha}_t, \mathbf{u}_i)'. \quad \blacksquare$$

Hence,

$$\tilde{\mathbf{a}}_t - \tilde{\mathbf{a}}_t^{(m)} = (\mathbf{A}_t^+ + \mathbf{P}_t \mathbf{R}_{t-1}^+) \mathbf{M}_i^{-1} \mathbf{u}_i \quad (16)$$

provides the measure of influence of the i -th observation on the state estimate at time t . The quantities on the right hand side are readily available from the augmented Kalman filter for model (5); $\sigma^{-2} \text{Cov}(\boldsymbol{\alpha}_t, \mathbf{u}_i) = \mathbf{A}_t^+ + \mathbf{P}_t \mathbf{R}_{t-1}^+$ is the *leverage* of \mathbf{y}_i on $\tilde{\mathbf{a}}_t$ (De Jong, 1996).

We now show that

$$\mathbf{R}_{i-1}^+ = \mathbf{Z}'_i \mathbf{M}_i - \mathbf{T}'_i \mathbf{N}_i \mathbf{K}_i. \quad (17)$$

$$\begin{aligned} \mathbf{R}_{i-1}^+ &= \mathbf{Z}'_i \mathbf{F}_i^{-1} + \mathbf{L}'_i \mathbf{Z}'_{i+1} \mathbf{F}_{i+1}^{-1} \mathbf{V}_{i+1}^+ + \cdots + \mathbf{L}'_{T,i} \mathbf{Z}'_T \mathbf{F}_T^{-1} \mathbf{V}_T^+ \\ &= \mathbf{Z}'_i \mathbf{F}_i^{-1} - (\mathbf{L}'_i \mathbf{Z}'_{i+1} \mathbf{F}_{i+1}^{-1} \mathbf{Z}_{i+1} \mathbf{K}_i + \cdots + \mathbf{L}'_{T,i} \mathbf{Z}'_T \mathbf{F}_T^{-1} \mathbf{Z}_T \mathbf{L}_{T,i+1} \mathbf{K}_i) \\ &= \mathbf{Z}'_i \mathbf{F}_i^{-1} - \mathbf{L}'_i \mathbf{N}_i \mathbf{K}_i \\ &= \mathbf{Z}'_i \mathbf{F}_i^{-1} - \mathbf{T}'_i \mathbf{N}_i \mathbf{K}_i + \mathbf{Z}'_i \mathbf{K}'_i \mathbf{N}_i \mathbf{K}_i \\ &= \mathbf{Z}'_i \mathbf{M}_i - \mathbf{T}'_i \mathbf{N}_i \mathbf{K}_i \end{aligned}$$

which proves (17). Therefore, we recover the result derived in De Jong (1996), $\tilde{\mathbf{a}}_i - \tilde{\mathbf{a}}_i^{(m)} = \mathbf{P}_i (\mathbf{Z}'_i \mathbf{M}_i - \mathbf{T}'_i \mathbf{N}_i \mathbf{K}_i)$.

Let $\mathbf{e}_t = \text{E}(\boldsymbol{\varepsilon}_t | \mathbf{Y}_T)$ denote the smoothed disturbance. Koopman (1993) shows that

$$\mathbf{e}_t = \mathbf{G}_t \mathbf{F}_t^{-1} \boldsymbol{\nu}_t + \mathbf{J}'_t \mathbf{r}_t.$$

We are now interested in assessing the influence of the i -th observation on the smoothed estimate of the disturbance $\boldsymbol{\varepsilon}_t$.

For this purpose, we denote $\mathbf{e}_t^{(m)} = \text{E}(\boldsymbol{\varepsilon}_t | \mathbf{Y}_T^{(m)})$, and, taking the expectation of both sides of (1) alternatively with respect to \mathbf{Y}_T and $\mathbf{Y}_T^{(i)}$,

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \tilde{\mathbf{a}}_t + \mathbf{G}_t \mathbf{e}_t, \\ \mathbf{y}_t &= \mathbf{Z}_t \tilde{\mathbf{a}}_t^{(m)} + \mathbf{G}_t \mathbf{e}_t^{(m)}, \end{aligned}$$

we write

$$\begin{aligned} \mathbf{Z}_t(\tilde{\mathbf{a}}_t - \tilde{\mathbf{a}}_t^{(m)}) &= -\mathbf{G}_t(\mathbf{e}_t - \mathbf{e}_t^{(m)}), \quad t \neq i, \\ \mathbf{y}_i - \mathbb{E}(\mathbf{y}_i | \mathbf{Y}_T^{(i)}) &= \mathbf{Z}_i(\tilde{\mathbf{a}}_i - \tilde{\mathbf{a}}_i^{(m)}) + \mathbf{G}_i(\mathbf{e}_i - \mathbf{e}_i^{(m)}). \end{aligned} \quad (18)$$

Moreover, from the transition equation (2)

$$\tilde{\mathbf{a}}_{t+1} - \tilde{\mathbf{a}}_{t+1}^{(m)} = \mathbf{T}_t(\tilde{\mathbf{a}}_t - \tilde{\mathbf{a}}_t^{(m)}) + \mathbf{H}_t(\mathbf{e}_t - \mathbf{e}_t^{(m)}). \quad (19)$$

Theorem 5. Defining

$$\mathbf{E}_t^+ = \mathbf{G}_t' \mathbf{F}_t^{-1} \mathbf{V}_t^+ + \mathbf{J}_t' \mathbf{R}_t^+ \quad (20)$$

the scaled smoothed disturbances of the transition equation are given by

$$\mathbf{H}_t(\mathbf{e}_t - \mathbf{e}_t^{(m)}) = \mathbf{H}_t \mathbf{E}_t^+ \mathbf{M}_i^{-1} \mathbf{u}_i. \quad (21)$$

Moreover,

$$\mathbf{H}_i(\mathbf{e}_i - \mathbf{e}_i^{(m)}) = \mathbf{H}_i(\mathbf{G}_i' \mathbf{M}_i - \mathbf{H}_i' \mathbf{N}_i \mathbf{K}_i) \mathbf{M}_i^{-1} \mathbf{u}_i.$$

Proof.

$$\begin{aligned} \mathbf{H}_t(\mathbf{e}_t - \mathbf{e}_t^{(m)}) &= (\tilde{\mathbf{a}}_{t+1} - \tilde{\mathbf{a}}_{t+1}^{(m)}) - \mathbf{T}_t(\tilde{\mathbf{a}}_t - \tilde{\mathbf{a}}_t^{(m)}) \\ &= [(\mathbf{A}_{t+1}^+ + \mathbf{P}_{t+1} \mathbf{R}_t^+) - \mathbf{T}_t(\mathbf{A}_t^+ + \mathbf{P}_t \mathbf{R}_{t-1}^+)] \mathbf{M}_i^{-1} \mathbf{u}_i \\ &= [\mathbf{K}_t \mathbf{V}_t^+ + (\mathbf{T}_t \mathbf{P}_t \mathbf{L}_t + \mathbf{H}_t \mathbf{J}_t') \mathbf{R}_t^+ - \mathbf{T}_t \mathbf{P}_t (\mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{V}_t^+ + \mathbf{L}_t' \mathbf{R}_t^+)] \mathbf{M}_i^{-1} \mathbf{u}_i \\ &= [\mathbf{K}_t \mathbf{V}_t^+ + \mathbf{H}_t \mathbf{J}_t' \mathbf{R}_t^+ - \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{V}_t^+] \mathbf{M}_i^{-1} \mathbf{u}_i \\ &= \mathbf{H}_t (\mathbf{G}_t' \mathbf{F}_t^{-1} \mathbf{V}_t^+ + \mathbf{J}_t' \mathbf{R}_t^+) \mathbf{M}_i^{-1} \mathbf{u}_i. \end{aligned}$$

Note that, for $t < i$, $\mathbf{E}_t^+ = \mathbf{J}_t' \mathbf{R}_t^+$. Also, $\mathbf{H}_0(\mathbf{e}_0 - \mathbf{e}_0^{(m)}) = \mathbf{H}_0 \mathbf{H}_0' \mathbf{R}_0^+ \mathbf{M}_i^{-1} \mathbf{u}_i$.

In order to prove the last statement, we first note that

$$\begin{aligned} \mathbf{R}_i^+ &= \mathbf{Z}_{i+1}' \mathbf{F}_{i+1}^{-1} \mathbf{V}_{i+1}^+ + \mathbf{L}_{i+1} \mathbf{Z}_{i+2}' \mathbf{F}_{i+2}^{-1} \mathbf{V}_{i+2}^+ + \cdots + \\ &\quad \mathbf{L}_{T,i+1}' \mathbf{Z}_T' \mathbf{F}_T^{-1} \mathbf{V}_T^+ \\ &= -(\mathbf{Z}_{i+1}' \mathbf{F}_{i+1}^{-1} \mathbf{Z}_{i+1} \mathbf{K}_i + \mathbf{L}_{i+1} \mathbf{Z}_{i+2}' \mathbf{F}_{i+2}^{-1} \mathbf{Z}_{i+2} \mathbf{L}_{i+1} \mathbf{K}_i + \cdots + \\ &\quad \mathbf{L}_{T,i+1}' \mathbf{Z}_T' \mathbf{F}_T^{-1} \mathbf{Z}_T \mathbf{L}_{T,i+1} \mathbf{K}_i) \\ &= -\mathbf{N}_i' \mathbf{K}_i \end{aligned}$$

which gives

$$\begin{aligned}
\mathbf{E}_i^+ &= \mathbf{G}'_i \mathbf{F}_i^{-1} \mathbf{V}_i^+ + \mathbf{J}'_i \mathbf{R}_i^+ \\
&= \mathbf{G}'_i \mathbf{F}_i^{-1} + \mathbf{H}'_i \mathbf{R}_i^+ - \mathbf{G}'_i \mathbf{K}'_i \mathbf{R}_i^+ \\
&= \mathbf{G}'_i \mathbf{M}_i - \mathbf{H}'_i \mathbf{N}_i \mathbf{K}_i \quad \blacksquare
\end{aligned}$$

Theorem 6. For $t \neq i$ the scaled smoothed measurement disturbance is given as follows:

$$\mathbf{G}_t(\mathbf{e}_t - \mathbf{e}_t^{(m)}) = \mathbf{G}_t \mathbf{E}_t^+ \mathbf{M}_i^{-1} \mathbf{u}_i \quad (22)$$

Proof. From (18)

$$\begin{aligned}
\mathbf{G}_t(\mathbf{e}_t - \mathbf{e}_t^{(m)}) &= -\mathbf{Z}_t(\tilde{\mathbf{a}}_t - \tilde{\mathbf{a}}_t^{(m)}) \\
&= -\mathbf{Z}_t(\mathbf{A}_t^+ + \mathbf{P}_t \mathbf{R}_{t-1}^+) \mathbf{M}_i^{-1} \mathbf{u}_i \\
&= [-\mathbf{Z}_t \mathbf{A}_t^+ - \mathbf{Z}_t \mathbf{P}_t (\mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{V}_t^+ + \mathbf{L}'_t \mathbf{R}_t^+)] \mathbf{M}_i^{-1} \mathbf{u}_i \\
&= [\mathbf{V}_t^+ - (\mathbf{F}_t - \mathbf{G}_t \mathbf{G}'_t) \mathbf{F}_t^{-1} \mathbf{V}_t^+ - \mathbf{Z}_t \mathbf{P}_t \mathbf{L}'_t \mathbf{R}_t^+] \mathbf{M}_i^{-1} \mathbf{u}_i \\
&= \mathbf{G}_t (\mathbf{G}'_t \mathbf{F}_t^{-1} \mathbf{V}_t^+ + \mathbf{J}'_t \mathbf{R}_t^+) \mathbf{M}_i^{-1} \mathbf{u}_i \\
&= \mathbf{G}_t \mathbf{E}_t^+ \mathbf{M}_i^{-1} \mathbf{u}_i.
\end{aligned}$$

In the derivation we used the easily established relation: $\mathbf{Z}_t \mathbf{P}_t \mathbf{L}'_t = -\mathbf{G}_t \mathbf{J}'_t$. \blacksquare

Theorem 7. The influence of \mathbf{y}_i on the smoothed disturbance:

$$\mathbf{e}_t - \mathbf{e}_t^{(m)} = \mathbf{E}_t^+ \mathbf{M}_i^{-1} \mathbf{u}_i. \quad (23)$$

Proof. The proof is immediate, as the matrix $[\mathbf{G}'_t \mathbf{H}'_t]'$ has full column rank. \blacksquare

In conclusion, the computation of the influence for $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\alpha}_t$ via the forward recursion (19), depend on quantities readily available from a run of the smoothing filter on the dummy variable. \mathbf{E}_t^+ provides a measure of leverage of \mathbf{y}_i on \mathbf{e}_t .

5 Conclusions

The paper has showed the equivalence between the skipping approach and the dummy variable (additive outlier) approach for both likelihood and smoothed inferences, and use the latter for deriving suitable algorithms for computing deletion diagnostics. The extension to the class of nonstationary state space models is available from the author.

References

- Bell, W. R. (1989). Discussion of the paper by Bruce and Martin: Leave- k -out diagnostics for time series, *Journal of the Royal Statistical Society, Series B*, 51, 408-409.
- Bruce, A.G., and Martin, R.D. (1989). Leave- k -out diagnostics for time series, *Journal of the Royal Statistical Society, Series B*, 51, 363-424 (with discussion).
- Burman J.P. (1980). Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society, Series A* 143, 321-337.
- De Jong P. (1989). Smoothing and Interpolation with the State Space Model, *Journal of the American Statistical Association*, 84, 1085-1088.
- De Jong, P. (1991). The Diffuse Kalman Filter, *Annals of Statistics*, 19, 1073-1083.
- De Jong, P. (1996). Fixed Interval Smoothing, *Working paper*, London School of Economics.
- De Jong, P., and Penzer, J. (1998). Diagnosing Shocks in Time Series, *Journal of the American Statistical Association*, 93 796-806.
- Gómez V., Maravall A., and Peña, D. (1999). Missing Observations in ARIMA Models: Skipping Approach versus Additive Outlier Approach, *Journal of Econometrics*, 88, 341-363.

- Fuller W. A. (1996). *Introduction to Statistical Time Series*. Wiley Series in Probability and Statistics, New York, John Wiley and Sons.
- Ljung, G.M. (1993). On Outlier Detection in Time Series, *Journal of the Royal Statistical Society, Series B*, 55, 559-567.
- Koopman, S.J. (1993). Disturbance Smoother for State Space Models, *Biometrika*, 80, 117-126.
- Rosenberg, B. (1973). Random coefficient models: the analysis of a cross-section of time series by stochastically convergent parameter regression, *Annals of Economic and Social Measurement*, 2, 399-428.
- Sargan, J.D., and Drettakis, E.G. (1974). Missing Data in an Autoregressive Model, *International Economic Review*, 15, 39-58.