# APPROACHES IN USING EXPECTATION-MAXIMIZATION ALGORITHM FOR MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS OF A CONSTRAINED STATE SPACE MODEL WITH AN EXTERNAL INPUT SERIES

Chengliang Huang[1], Xiao-Ping Zhang[1]and Fang Wang[2]

[1]Department of Electrical and Computer Engineering, Ryerson University
Toronto, On, Canada
c28huang@ryerson.ca, xzhang@ee.ryerson.ca
[2]School of Business and Economics, Wilfrid Laurier University
Waterloo, On, Canada
fwang@wlu.ca

***ABSTRACT***

*EM algorithm is popular in maximum likelihood estimation of parameters for state-space models. However, extant approaches for the realization of EM algorithm are still not able to fulfill the task of identification systems, which have external inputs and constrained parameters. In this paper, we propose new approaches for both initial guessing and MLE of the parameters of a constrained state-space model with an external input. Using weighted least square for the initial guess and the partial differentiation of the joint log-likelihood function for the EM algorithm, we estimate the parameters and compare the estimated values with the "actual" values, which are set to generate simulation data. Moreover, asymptotic variances of the estimated parameters are calculated when the sample size is large, while statistics of the estimated parameters are obtained through bootstrapping when the sample size issmall. The results demonstrate that the estimated values are close to the "actual" values.Consequently, our approaches are promising and can applied in future research.*

***KEYWORDS***

*State-Space Model, Maximum Likelihood Estimation, Expectation Maximization Algorithm, Kalman filtering and smoothing, Asymptotic variances, Bootstrapping*

## 1. INTRODUCTION

One of the advantages of the state-space models for linear dynamic systems is their ability to fit more parsimonious structures with fewer parameters to describe a multivariate time series. As a result, the application of state-space models is not limited to engineering practice.

For example, in certain scenarios, the dynamics of the brand equity of a firm can be represented by a state-space model in the form of Kalman filter [1],

David C. Wyld et al. (Eds) : DBDM, CICS, CSIP, AI&FL, SCOM, CSE, CCNET-2016
pp. 57–66, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.60506

$$\begin{cases} x_t = \alpha x_{t-1} + \gamma u_t + e_t \\ y_t = h x_t + w_t \end{cases}, \qquad (1)$$

where $x_t$ is the invisible brand equity, $u_t$ is the investment, the external input, $y_t$ is the brand performance, the output, at a certain step $t$. The transition coefficient, $\alpha(|\alpha| < 1)$, and the input coefficient, $\gamma$, are parameters of interest. The observation coefficient, $h$, is constrained to be a constant, i.e., $h$=1. The observation noise, $w_t$, the process noise,$e_t$, and the initial state,$x_0$, at each step, $\{x_0, e_1, ..., e_t, w_1 ... w_t\}$, are all assumed to be mutually independent, where $w_t \sim N(0, \sigma_w^2)$,$e_t \sim N(0, \sigma_e^2)$, and $x_0 \sim N(\mu_0, \sigma_0^2)$.

As exemplified in Equation (1) where $h$=1, a constrained state-space modemeans that some parameters, some elements in the parameter matrix of a Kalman filter, are fixed or shared thus not all of them have to be estimated. Conversely, if the model is unconstrained, any parameters or any elements of each parameter matrix will be estimated. Moreover, the model has an time-variant external input series, $u_t$. By the way, parameter estimation of state space models with external inputs can be seen as supervised problems while that of state space models without external inputs can be treated as unsupervised problems.

Maximum likelihood estimation (MLE) [2] is used to obtain the time-invariant parameters of the Kalman filter from input, $u_t$, and the output, $y_t$. Because of the existence of hidden variables, $x_t$, in the formulated likelihood function, expectation-maximization (EM) algorithm [3], an iterative method, is used to complete MLE of parameters in state-space models.

The approaches for the implementation of EM algorithm in the MLE of unconstrained state-space model are delivered in [4]-[7]. The approaches for certain constrained state-space models are covered in [7]-[10]. Since those constraints are quite specific, the approaches proposed are lack of generalicity. On the other hand, approaches for state-space models with external inputs are usually not provided except in [7]-[9]. However, the state space mode in [7]-[9] is not typical because the inputs are constants or do not affect the hidden variable. Especially in [5], the external input is actually a constant, which is a parameter to be estimated. Therefore, as we know from extant literature, no innovative methods are explored to estimate the unknown parameters of such constrained state-space modelsin the recent years.

Our research is to find appropriate methods for the supervised problem: using EM algorithm for the MLE of those constrained state-space models who have external inputs. In this paper, we use the model represented by Equation (1) as an example. Our target is to estimate the parameters,$\alpha$, $\gamma$, $\sigma_e$, $\sigma_w$, $\mu_0$ and $\sigma_0$. Our work is carried out mainly in two phases: (i) initial guessing using our innovative approach, and (ii) implementation of MLE using EM algorithm with approaches different from those in extant literature. In addition, if the sample size is large, we use the asymptotic variances of the estimated parameters to check the accuracy of the estimation. If the sample size is small, we introduced bootstrapping to examine the distribution of the estimated parameters.

## 2. INITIAL GUESSING

The initial guess is performed through two steps. Firstly, the system parameter $\alpha$ and the variances $\sigma_e$, $\sigma_w$ and $\sigma_0$ are guessed using autocovariance of the observations. Then the initial state mean $\mu_0$ and input parameter $\gamma$ are guessed using weighted linear square.

### 2.1.GuessingSystem Parameter and the Variances

Denote $\eta_t = \sum_{i=0}^{\infty} \alpha^i e_{t-i}$, from (1) it can be proved that

$$x_t = \frac{\gamma u_t}{1-\alpha} + \eta_t \ . \tag{2}$$

For $h = 0,1,2 \dots \dots$, we can obtain the covariance and autocovariance of $\eta_t$ as

$$\gamma_\eta(h) = \frac{\sigma_e^2 \alpha^h}{1-\alpha^2}, \tag{3}$$

while it is obvious that

$$\sigma_0^2 = \frac{\sigma_e^2}{1-\alpha^2}. \tag{4}$$

which means that we don't have to estimate $\sigma_e$ and $\sigma_0$ separately, as was performed in extant literature.

Moreover, we have the variance of $y_t$

$$\gamma_y(0) = v_0 + v_w, \tag{5}$$

and the covariance of $y_t$ when $h = 1, 2, 3 \cdots$,

$$\gamma_y(h) = \gamma_\eta(h). \tag{6}$$

Hence we can obtain the guessed initial values for system parameter,$\alpha$, and standard deviations, $\sigma_e$ and $\sigma_w$, from

$$\begin{cases} \alpha = \frac{\gamma_y(2)}{\gamma_y(1)} \\ \sigma_e^2 = \frac{(1-\alpha^2)\gamma_y(1)}{\alpha} \\ \sigma_w^2 = \gamma_y(0) - \frac{\sigma_e^2}{1-\alpha^2} \end{cases} . \tag{7}$$

## 2.2. Guessing Initial State Mean and Input Parameter

Denoting

$$z_t = \sum_{i=1}^{t} \alpha^{t-i} u_i, \tag{8}$$

and

$$\zeta_t = \sum_{i=1}^{t} \alpha^{t-i} e_i + w_t, \tag{9}$$

we have,

$$y_t = \alpha^t \mu_0 + \gamma z_t + \zeta_t, \tag{10}$$

where $\zeta_t \sim N(0, \sigma_\zeta^2)$ and

$$\sigma_\zeta^2 = \frac{1-\alpha^{2t}}{1-\alpha^2} \sigma_e^2 + \sigma_w^2 \tag{11}$$

In order to estimate $\mu_0$ and $\gamma$, we perform linear regression between $y_t$, as dependent variable, and $\alpha^t$ and $z_t$, as independent variables, using *T* samples of $u_t$ and $y_t$. Since $\zeta_t$ is heteroscadestical, we apply weighted least square (WLS). WLS finds its optimum when the weighted sum, *S*, of squared residuals is minimized where

$$S = \sum_{t=1}^{T} \frac{(y_t - \alpha^t \mu_0 - \gamma z_t)^2}{\frac{1-\alpha^{2t}}{1-\alpha^2}\sigma_e^2 + \sigma_w^2}. \tag{12}$$

Denote that $v_t = (1-\alpha^{2t})\sigma_e^2 + (1-\alpha^2)\sigma_w^2$, we solve the gradient equation (regarding $\mu_0$ and $\gamma$ respectively) for the sum of squares

$$\begin{cases} \sum_{t=1}^{T} \frac{\alpha^{t}(y_t - \alpha^{t}\mu_0 - \gamma z_t)(1-\alpha^2)}{v_t} = 0 \\ \sum_{t=1}^{T} \frac{z_t(y_t - \alpha^{t}\mu_0 - \gamma z_t)(1-\alpha^2)}{v_t} = 0 \end{cases}. \tag{13}$$

Therefore, we will have initial guess about as below:

$$\begin{cases} \gamma = \frac{\sum_{t=1}^{T}\frac{\alpha^{2t}}{v_t}\sum_{t=1}^{T}\frac{z_t y_t}{v_t} - \sum_{t=1}^{T}\frac{\alpha^{t} y_t}{v_t}\sum_{t=1}^{T}\frac{\alpha^{t} z_t}{v_t}}{\sum_{t=1}^{T}\frac{\alpha^{2t}}{v_t}\sum_{t=1}^{T}\frac{z_t^2}{v_t} - \sum_{t=1}^{T}\frac{\alpha^{t} z_t}{v_t}\sum_{t=1}^{T}\frac{\alpha^{t} z_t}{v_t}} \\ \mu_0 = \frac{\sum_{t=1}^{T}\frac{\alpha^{t} y_t}{v_t}\sum_{t=1}^{T}\frac{z_t^2}{v_t} - \sum_{t=1}^{T}\frac{\alpha^{t} z_t}{v_t}\sum_{t=1}^{T}\frac{y_t z_t}{v_t}}{\sum_{t=1}^{T}\frac{\alpha^{2t}}{v_t}\sum_{t=1}^{T}\frac{z_t^2}{v_t} - \sum_{t=1}^{T}\frac{\alpha^{t} z_t}{v_t}\sum_{t=1}^{T}\frac{\alpha^{t} z_t}{v_t}} \end{cases}. \tag{14}$$

## 3. ESTIMATION USING EM ALGORITHM

The conditional density for the states and outputs are,

$$P(x_t|x_{t-1}) = \frac{1}{\sigma_e\sqrt{2\pi}} exp\left[-\frac{(x_t - \alpha x_{t-1} - \gamma u_t)^2}{2\sigma_e^2}\right], \tag{15}$$

$$P(y_t|x_t) = \frac{1}{\sigma_w\sqrt{2\pi}} exp\left[-\frac{(y_t - x_t)^2}{2\sigma_w^2}\right]. \tag{16}$$

Assuming a Gaussian initial state density

$$P(x_0) = \frac{1}{\sigma_0\sqrt{2\pi}} exp\left[-\frac{(x_0 - \mu_0)^2}{2\sigma_0^2}\right], \tag{17}$$

By the Markov property implicit in this model, we calculate the joint probability, not the partial probability used by Shumway (2011), regarding all T samples of $x_t$ and $y_t$, denoted as $\{x\}$ and $\{y\}$ respectively:

$$P(\{x\},\{y\}) = P(x_0)\prod_{t=1}^{T} P(x_t|x_{t-1})\prod_{t=1}^{T} P(y_t|y_{t-1}) \tag{18}$$

We denote the joint log probability as

$$\Omega = \log P(\{x\},\{y\}). \tag{19}$$

According to Equation (4), we only need to estimate the parameter set, $\Psi = \{\alpha,\ \gamma,\ \sigma_e,\ \sigma_w,\ \mu_0\}$, through maximizing the objective function:

$$\Omega(\alpha,\ \gamma,\ \sigma_e,\ \sigma_w,\ \mu_0) = -\frac{(1-\alpha^2)(x_0-\mu_0)^2}{2\sigma_e^2} - \frac{1}{2}\log\frac{\sigma_e^2}{1-\alpha^2} - \sum_{t=1}^{T}\frac{(x_t - \alpha x_{t-1} - \gamma u_t)^2}{2\sigma_e^2} - \frac{1}{2}T\log\sigma_e^2 - \sum_{t=1}^{T}\frac{(y_t - x_t)^2}{2\sigma_w^2} - \frac{1}{2}T\log\sigma_w^2 - \frac{2T+1}{2}\log(2\pi). \tag{20}$$

### 3.1. EM Algorithm

Since the objective function expressed by Equation(20) depends on the unobserved data series,$x_t (t = 1, 2, \dots T)$, we consider applying the EM algorithm conditionally with respect to the observed output series $y_1,\ y_2,\ \dots,\ y_T$. The objective function above has an input series. Accordingly, the input coefficient has to be estimated. Consequently, our approaches are unlike the approaches[6] used inthe implementation ofEM algorithm for linear dynamic systems.

The EM algorithm mainly has two steps: the E-STEP and the M-Step. During the E-Step, the parameters are assumed known, the hidden states and their variance are estimated over all the samples, and then the likelihood function constructed from joint probability are calculated.

During the M step, we have to find the parameter set $\widehat{\Psi}(k) = \{\alpha(k), \gamma(k), \sigma_e(k), \sigma_w(k), \mu_0(k)\}$ for the $k$th counts of the recursions by maximizing the conditional expectation, or the above objective function.

The overall procedure for the implementation of EM algorithm is as below:

(i) Initialize the procedure by selecting the guessed values as starting values for the parameters.
On iteration k, (k=1,2,……)

(ii) Compute the log-likelihood (optional),

(iii) Use the parameters to obtain the smoothed values of the hidden states and their correlations, for t= 1,2, …...,T.

(iv) Use the smoothed values to calculate the updated parameters.

(v) Repeat Steps (ii) – (iv) to convergence.

We mainly perform two sub-steps in the E-step of EM algorithm: Kalman filtering and Kalman smoothing.

## 3.2. Kalman Filtering and Smoothing

Assuming thatwe already know the parameter set $\{\alpha, \gamma, \sigma_e, \sigma_w, \mu_0, \sigma_0\}(x_0 \sim N(\mu_0, \sigma_0^2))$, and the observations $y_t$ and $u_t$, we have the estimation of the hidden state, as well as the variances estimated based on the observations for the period 1 to $t$.

$$x_{t|t-1} = \alpha x_{t-1|t-1} + \gamma u_t, \tag{21a}$$

$$v_{t|t-1} = \alpha^2 v_{t-1|t-1} + \sigma_e^2, \tag{21b}$$

$$\widetilde{y}_t = y_t - x_{t|t-1}, \tag{21c}$$

$$s_t = v_{t|t-1} + v_w, \tag{21d}$$

$$k_t = v_{t|t-1} s_t^{-1}, \tag{21e}$$

$$x_t^t = x_{t|t-1} + k_t \widetilde{y}_t, \tag{21f}$$

$$v_t^t = (1 - k_t) v_{t|t-1}, \tag{21g}$$

where $x_{0|0} = \mu_0$ and $v_{0|0} = \sigma_0^2$.

According to [6],to compute$E[x_t|\{y, u\}] \equiv x_t^T$ and the correlation matrices $p_t \equiv v_t^T + x_t^T (x_t^T)'$ one performs a set of backward recursion using

$$j_{t-1} = \alpha \frac{v_{t-1|t-1}}{v_{t|t-1}}, \tag{22}$$

$$x_{t-1}^T = x_{t-1|t-1} + j_{t-1}\left(x_t^T - \alpha x_{t-1|t-1} - \gamma u_t\right), \tag{23}$$

$$v_{t-1}^T = v_{t-1|t-1} + j_{t-1}\left(v_t^T - v_{t|t-1}\right)j_{t-1}', \tag{24}$$

where $x_T^T = x_{T|T}$ and $v_T^T = v_{T|T}$. We also have $p_{t,t-1} \equiv V_{t,t-1}^T + x_t^T(x_{t-1}^T)'$, where $v_{t,t-1}^T$ can be obtained through the backward recursions

$$v_{t-1,t-2}^T = v_{t-1|t-1}j_{t-2}' + j_{t-1}\left(v_{t,t-1}^T - \alpha v_{t-1|t-1}\right)j_{t-2}', \tag{25}$$

which is initialized using $v_{T,T-1}^T = \alpha(1-k_T)v_{T-1}^{T-1}$.

Note that the state estimate, $x_t^T$**,** differs from the one computed in a Kalman filter in that it is the smoothed estimator of $x_t$ based on all of the observed data (and input data), i.e. it depends on past and future observations; the Kalman filter estimates$E[x_t|\{y\}_1^t]$ is the usual Kalman filter estimator based on the obsearved data up to the current time instant $\boldsymbol{t}$**.**

## 3.3. Expected Log-Likelihood Formulation

After we have got the expected values for $x_0$and$x_t$ as $x_0^T \equiv\ E[x_0|\{y,u\}]$ and $x_t^T \equiv E[x_t|\{y,u\}]$ respectively, we can calculate the expectation of the log-likelihood

$$E(\Omega) = E[\log P(\{x\},\{y\})]. \tag{26}$$

Denote

$$\begin{cases} P_{t-1}^T = \sum_{t=1}^T E[x_{t-1}x_{t-1}'|\{y\}] \\ P_t^T = \sum_{t=1}^T E[x_t x_t'|\{y\}] \\ P_{t,t-1}^T = \sum_{t=1}^T E[x_t x_{t-1}'|\{y\}] \end{cases}, \begin{cases} M_{t-1}^T = \sum_{t=1}^T E(x_{t-1}u_t) \\ M_t^T = \sum_{t=1}^T E(x_t u_t) \\ U_t^T = \sum_{t=1}^T u_t^2 \end{cases}, \text{ and } \begin{cases} W_{t-1}^T = \sum_{t=1}^T E(x_{t-1}y_t) \\ W_t^T = \sum_{t=1}^T E(x_t y_t) \\ Y_t^T = \sum_{t=1}^T y_t^2 \end{cases},$$

we have

$$E(\Omega) = -\frac{1}{2}E(\Omega_0) - \frac{1}{2}E(\Omega_1) - \frac{1}{2}E(\Omega_2) - \frac{2T+1}{2}\log(2\pi) \tag{27}$$

where

$$E(\Omega_0) = \frac{1-\alpha^2}{\sigma_e^2}\left[v_0^T + (x_0^T - \mu_0)^2\right] + \log\frac{\sigma_e^2}{1-\alpha^2} \tag{27a}$$

$$E(\Omega_1) = \sigma_e^{-2}\left(P_t^T + \alpha^2 P_{t-1}^T + \gamma^2 U_t^T - 2\alpha P_{t,t-1}^T - 2\gamma M_t^T + 2\alpha\gamma M_{t-1}^T\right) + T\log\sigma_e^2 \tag{27b}$$

$$E(\Omega_2) = \sigma_w^{-2}(Y_t^T + P_t^T - 2W_t^T) + T\log\sigma_w^2 \tag{27c}$$

## 3.4. Estimation of the Parameters

We use the first order condition on partial derivatives of $E(\Omega)$ to individual parameters to obtain the gradient and then the values of individual parameters. This method is not the multivariate regression approach [6]. The parameters are chosen when the objective function is maximized, i.e., the gradients are all zero. The estimates of $\alpha$, $\gamma$, $\sigma_e^2$, $\sigma_w^2$ and $\mu_0$are from below five equations:

$$\alpha(1-\alpha^2)v_0^T - \alpha\sigma_e^2 - \alpha(1-\alpha^2)P_{t-1}^T + (1-\alpha^2)P_{t,t-1}^T - \gamma(1-\alpha^2)M_{t-1}^T = 0 \tag{28a}$$

$$\gamma U_t^T - M_t^T + \alpha M_{t-1}^T = 0 \tag{28b}$$

$$(1-\alpha^2)v_0^T - (1+T)\sigma_e^2 + P_t^T + \alpha^2 P_{t-1}^T + \gamma^2 U_t^T - 2\alpha P_{t,t-1}^T - 2\gamma M_t^T + 2\alpha\gamma M_{t-1}^T = 0 \tag{28c}$$

$$\sigma_w^2 = \frac{1}{T}(Y_t^T + P_t^T - 2W_t^T) \tag{28d}$$

$$\mu_0 = x_0^T \tag{28e}$$

Moreover, we use the second orders of the derivatives of $E(\Omega)$ to calculate the second derivatives and then the information matrix. Most of the second order derivatives of $E(\Omega)$ are zero except those listed below:

$$\frac{\partial^2 E(\Omega)}{\partial \alpha^2} = \frac{v_0^T + \left(x_0^T - \mu_0\right)^2 - P_{t-1}^T}{\sigma_e^2} - \frac{1+\alpha^2}{(1-\alpha^2)^2}, \tag{29}$$

$$\frac{\partial^2 E(\Omega)}{\partial \gamma \partial \alpha} = \frac{\partial^2 E(\Omega)}{\partial \alpha \partial \gamma} = -\frac{M_{t-1}^T}{\sigma_e^2}, \tag{30}$$

$$\frac{\partial^2 E(\Omega)}{\partial \sigma_e \partial \alpha} = \frac{\partial^2 E(\Omega)}{\partial \alpha \partial \sigma_e} = \frac{2\left(\alpha P_{t-1}^T - P_{t,t-1}^T + \gamma M_{t-1}^T\right) - 2\alpha\left[v_0^T + \left(x_0^T - \mu_0\right)^2\right]}{\sigma_e^3}, \tag{31}$$

$$\frac{\partial^2 E(\Omega)}{\partial \gamma^2} = -\frac{U_t^T}{\sigma_e^2} \tag{32}$$

$$\frac{\partial^2 E(\Omega)}{\partial \sigma_e \partial \gamma} = \frac{\partial^2 E(\Omega)}{\partial \gamma \partial \sigma_e} = \frac{2\left(\gamma U_t^T - M_t^T + \alpha M_{t-1}^T\right)}{\sigma_e^3} \tag{33}$$

$$\frac{\partial^2 E(\Omega)}{\partial \sigma_e^2} = \frac{T+1}{\sigma_e^2} - \frac{3\left(1-\alpha^2\right)\left[v_0^T + \left(x_0^T - \mu_0\right)^2\right]}{\sigma_e^4} - \frac{3\left(P_t^T + \alpha^2 P_{t-1}^T + \gamma^2 U_t^T - 2\alpha P_{t,t-1}^T - 2\gamma M_t^T + 2\alpha\gamma M_{t-1}^T\right)}{\sigma_e^4} \tag{34}$$

$$\frac{\partial^2 E(\Omega)}{\partial \sigma_w^2} = \frac{T}{\sigma_w^2} - \frac{3\left(Y_t^T + P_t^T - 2W_t^T\right)}{\sigma_w^4} \tag{35}$$

$$\frac{\partial^2 E(\Omega)}{\partial \mu_0^2} = -\frac{1-\alpha^2}{\sigma_e^2} \tag{36}$$

According to Cramer-Rao Theorem, the MLE is an efficient estimate. When the sample size is large enough, the asymptotic variances of the estimates can be considered as the metric of the accuracy of the estimation. The asymptotic variances are calculated using the inverse of the information matrix, which is the inverse of the negative of the expected value of the Hessian matrix. The vector of the asymptotic variances of the estimates is

$$\begin{bmatrix} \dfrac{(T+1)(\alpha^2-1)^2 U\sigma_e^2}{(T+1)(1-\alpha^2)^2(\mathrm{P}_0 U - U\sigma_0^2 - \mathrm{M}_0^2) + (1+T-\alpha^2+T\alpha^2)U\sigma_e^2} \\ \dfrac{[(1+T)(1-\alpha^2)^2(\mathrm{P}_0 - \sigma_0^2) + (1+T-\alpha^2+T\alpha^2)\sigma_e^2]\sigma_e^2}{(T+1)(1-\alpha^2)^2(\mathrm{P}_0 U - U\sigma_0^2 - \mathrm{M}_0^2) + (1+T-\alpha^2+T\alpha^2)U\sigma_e^2} \\ \dfrac{[(1-\alpha^2)^2(\mathrm{P}_0 U - \mathrm{M}_0^2 - U\sigma_0^2) + (1+\alpha^2)U\sigma_e^2]\sigma_e^2}{(T+1)(1-\alpha^2)^2(\mathrm{P}_0 U - U\sigma_0^2 - \mathrm{M}_0^2) + (1+T-\alpha^2+T\alpha^2)U\sigma_e^2} \\ \dfrac{\sigma_w^2}{2T} \\ \sigma_0^2 \end{bmatrix}$$

If the sample size is small, we introduce boot-strapping procedure where the estimates are obtained from likelihood constructed from re-sampled standardized innovation, $\tilde{y}_t$, in Equation (21c). Moreover, the mean squared errors (MSE) of the state variables which is estimated from Equations (21a-g) using estimated parameters are also estimated.

## 4. SIMULATION AND RESULTS

The output data is generated through presetting the input series and the values of the parameters. We implement the initial guessing, EM iteration, and finally obtain the parameter estimates. This makes it easier to evaluate our work by comparing the actual values with the estimated values, or by checking the standard deviation of the estimates.

### 4.1. Data Generation

We generate data from the state-space model described as Equation (1). We assume $\alpha = 0.8$, $\gamma = 1.5$, and $\mu_0 = 0$. Moreover, the process noises, $e_t$, and observation noise, $w_t$, are generated independently where $e_t \sim N(0, 1.1^2)$ and $w_t \sim N(0, 0.9^2)$. We assume that the input, $u_t$, is a slow changing periodical square wave signal whose period is 10 time unit. The standard deviation of initial state, $\sigma_0$, is not needed during the data generation but can be calculated according to Equation (2). The expected log-likelihood can be calculated using Equation (27). Both are treated as "actual" values to be compared with guessed values and estimated values.

We performed our simulation using two different sample sizes: the large size of 1000 and the small size of 50. When the sample size is small, we applied bootstrapping method to estimate the accuracy of the estimate.

### 4.2. Results

We provide the results of the simulation with small sample size of 50 in Table 1, and the results of the simulation with large sample size of 1000 in Table 2.

Table 1. The parameters estimate with small sample size

| Parameters | Actual | Guessed | Estimated | Std. Dev. |
|---|---|---|---|---|
| $\alpha$ | 0.8 | 0.879 | 0.801 | 0.023 |
| $\gamma$ | 1.5 | 1.075 | 1.438 | 0.104 |
| $\sigma_e$ | 1.1 | 1.400 | 0.708 | 0.158 |
| $\sigma_w$ | 0.9 | 0.922 | 0.862 | 0.193 |
| $\mu_0$ | 0 | 0.992 | 2.143 | 1.231 |
| $\sigma_0$ | 1.83 | 3.811 | 1.183 | 0.252 |

In Table 1and Table 2, we displayed the actual values, the guessed values, the estimated values and the standard deviations of the estimated values for transition coefficient, $\alpha$, input coefficient, $\gamma$, standard deviation of process errors, $\sigma_e$, standard deviation of observation errors, $\sigma_w$, mean of initial state, $\mu_0$, and standard deviation of initial state, $\sigma_0$. In general, the guessed value is near the actual value while the estimated value is much more close to the actual value than the guessed ones for the parameters of the most interest: $\alpha$ and $\gamma$. The standard deviations in Table 1 are from bootstrapped distribution while the standard deviations in Table 2 are from the asymptotic variances.

Table 2: The parameters estimated with large sample size

| Parameters | Actual | Guessed | Estimated | Std. Dev. |
|---|---|---|---|---|
| $\alpha$ | 0.8 | 0.879 | 0.800 | 0.0001 |
| $\gamma$ | 1.5 | 1.276 | 1.424 | 0.0012 |
| $\sigma_e$ | 1.1 | 2.024 | 1.033 | 0.0011 |
| $\sigma_w$ | 0.9 | 0.533 | 0.971 | 0.0005 |
| $\mu_0$ | 0 | 2.459 | 1.667 | 2.9614 |
| $\sigma_0$ | 1.83 | 4.250 | 1.721 | NA |

In general, the guessed value is near the ‘actual” value while the estimated value is much closer to the “actual” value than the guessed ones, especially for the parameters of interest: $\alpha$ and $\gamma$. It is worth noting that the deviations of the estimated values are larger than the asymptotic ones due to the imperfectly generated data.

## 5. CONCLUSIONS

The research is to validate our innovative approaches in the application of EM algorithm in the MLE of a constrained dynamic linear system with external input. There are two main contributions in this research. Firstly, we realized that $\sigma_0$ and $\sigma_e$ has the relationship expressed by Equation (4) thus we don’t have to estimate both of them during the implementation of EM algorithm. Accordingly, the likelihood function in Equation (27) is not similar with those researchers who ignored the relationship. Secondly, in initial guessing of the value of input coefficient and the mean of initial state, we introduce weighted least square for the guessing of input coefficient, $\gamma$, and the mean of the initial state, $\mu_0$.

It is obvious that more techniques have to be discovered for the initial guessing, and the estimating based on the guessed initial guessing of, the parameter values, especially for the implementation of the M-step of the EM algorithm. The approaches we proposed can be a new start point for the future research on the estimation of dynamic systems with higher dimensions of external inputs, hidden states and observations.

## REFERENCES


[1] Kalman,R. E.,(1960) "A new approach to linear filtering and prediction problems",Journal of Basic Engineering, 82 (1), pp. 35–45.

[2] James D. Hamilton, Time Series Analysis, Princeton University Press, Princeton, New Jersey, 1996.

[3] Dempster,A.P., Laird, N.M. and Rubin, D.B. (1977) "Maximum likelihood from incomplete data via the EM algorithm",Journal of the Royal Statistical Society, Series B,Vol. 39, No.1, pp. 1–38.

[4] Ghahramani,Z. and Hinton,G. E. (1996) “Parameter estimation for linear dynamic systems”,Technical reportCRC-TR-96-2, Department of Computer Science, University of Toronto.

[5] Deng, L. and Shen, X. (1997) “Maximum likelihood in statistical estimation of dynamic systems: decomposition algorithm and simulation results”,Signal Processing, Vol. 57, No. 1.

[6] Shumway,Robert H. and Stoffer,David S. (2011) Time Series Analysis and Its Applications: With R Examples, New York, NY, Springer, New York.

[7] Holmes, Elizabeth Eli. (2013)“Derivation of an EM algorithm for constrained and unconstrained multivariate autoregressive state-space (MARSS) models”,Technical Report, Northwest Fisheries Center, NOAA Fisheries.

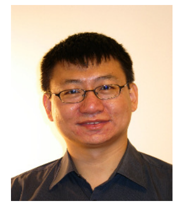
[8] Wu,Lilian Shiao-Yen, Pai,Jeffrey S. and Hosking,J. R. M.(1996) “An algorithm for estimating parameters of state-space models”,Statistics & Probability Letters, Vol. 28, pp. 99-106.

[9] Zuur, A. F.,Fryer, R. J.,Jolliffe,I. T.,Dekker, R. and Beukema, J. J. (2003) “Estimating common trends in multivariate time series using dynamic factor analysis”,Environmetrics, Vol. 14, No. 7, pp. 665-685.

[10] Dewar,M. Scerri, K. and Kadirkamanathan, V. (2009) “Data-driven spatio-temporal modeling using the integro-difference equation”,IEEE Transactions on Signal Processing,Vol. 57, No. 1, pp.83-91.

## AUTHORS

**Chengliang Huang** is currently a Ph. D. candidate at Ryerson University. He had been a senior engineer and project manager at Huawei and Intel for four years respectively. He graduated as MASc in Electrical and Computer Engineering, Ryerson University in 2009, MBA in Economic Management, Beijing Jiaotong University and B. Sci. in Electronic Engineering, Fudan University in 1990.He has published papers on peer-reviewed academic journals, including IEEE Sensor Journal and Wireless Personal Communications. His research interest includes wireless communications, random signal processing, and system identification, time series analysis, marketing data mining/analysis, and marketing productivity.



**Xiao-Ping Zhang** (M'97, SM'02) received the B.S. and Ph.D. degrees from Tsinghua University, in 1992 and 1996, respectively, all in electronic engineering. He holds an MBA in Finance, Economics and Entrepreneurship with Honors from Booth School of Business, the University of Chicago. Since Fall 2000, he has been with the Department of Electrical and Computer Engineering, Ryerson University, where he is now Professor, Director of Communication and Signal Processing Applications Laboratory (CASPAL) and Program Director of Graduate Studies. His research interests include multimedia communications and signal processing, multimedia retrieval and video content analysis, sensor networks and electronic systems, computational intelligence, and applications in bioinformatics, finance, and marketing. He is a frequent consultant for biotech companies and investment firms. He is cofounder and CEO for EidoSearch, Inc., which offers a content-based search and analysis engine for financial data. Dr. Zhang is a registered Professional Engineer in Ontario, Canada, a Senior Member of IEEE and a member of Beta Gamma Sigma Honor Society. He is the publicity co-chair for ICME'06 and program co-chair for ICIC'05. He is currently an Associate Editor for IEEE Transactions on Signal Processing, IEEE Signal Processing Letters and Journal of Multimedia**.**

**Dr. Wang's** research focuses on brand equity assessment, marketing data mining/analysis, long-term marketing productivity, and marketing strategy. She also works on e-commerce and information systems related topics. Her research on brand equity assessment is funded by the Social Sciences and Humanities Research Council of Canada (SSHRC). She is a key team member for a NSERC (Natural Sciences and Engineering Research Council of Canada) supported project on financial market information analysis. Her work has appeared in many academic journals, including Journal of Marketing, Journal of the Academy of Marketing Science, Information & Management, Journal of Consumer Marketing, among others.