Markov regime switching in mean and in fractional integration parameter.

Harun Özkan[†] Thanasis Stengos[‡] Ege Yazgan[§¶]

April 2016

To be appeared in Communications in Statistics-Simulation and Computation

Abstract

We propose a specific general Markov-regime switching estimation both in the long memory parameter d and the mean of a time series. Following Tsay and Härdle (2009) we employ Viterbi algorithm, which combines the Viterbi procedures, in two state Markov-switching parameter estimation. It is well-known that existence of mean break and long memory in time series can be easily confused with each other in most cases. Thus, we aim at observing the deviation and interaction of mean and d estimates for different cases.

A Monte Carlo experiment reveals that the finite sample performance of the proposed algorithm for a simple mixture model of Markov-switching mean and d changes with respect to the fractional integrating parameters and the mean values for the two regimes.

JEL Classification: C12, C22

Keywords: Markov regime switching, fractional integration, long memory time series.

[†]Istanbul Bilgi University, Kurtulus Deresi Caddesi, Dolapdere, Beyoglu, Istanbul, 34440, Turkey. Email: harun.ozkan@bilgi.edu.tr

[‡]Department of Economics, University of Guelph, 50 Stone Road East Guelph, Ontario, Canada. Email: tstengos@uoguelph.ca

[§]Istanbul Bilgi University, Kurtulus Deresi Caddesi, Dolapdere, Beyoglu, Istanbul, 34440, Turkey. Email: ege.yazgan@bilgi.edu.tr

[¶]We want to thank an anonymous referee for his/her very helpful comments on an earlier version of the paper. The second author wants to acknowledge finanacial support for NSERC of Canada.

1. Introduction

The use of fractional or long memory methods has been extensive in econometrics as they have been found to be quite effective in describing the behavior of many macroeconomic and financial data (Lobato and Velasco (2000) Ding et al. (1993)). In particular, recent findings suggest that long memory phenomenon is observed in LIBOR (Cajueiro and Tabak, 2005b, 2007a), interest rate (Cajueiro and Tabak, 2007b), trading volume (Lobato and Velasco, 2000; Lux and Kaizoji, 2007), and volatility of returns (Lobato and Velasco, 2000; Cajueiro and Tabak, 2005c; Granger and Ding, 1994) -albeit controversial findings in terms of long memory behavior are present for the stock returns(Cajueiro and Tabak, 2005a; Limam, 2003; Willinger et al., 1999).

Similarly, since the seminal work of Hamilton (1989) the Markov–switching model has been a popular vehicle to analyze economic phenomena that are likely to obey regime changes. Recently, Tsay and Härdle (2009) (hereafter TH) has combined the two approaches in a unified framework by introducing a Markov-switching-ARFIMA (MS-ARFIMA) process which extends the hidden Markov model with a latent state variable, allowing for the different regimes to have different degrees of long memory. Recent papers in the literature have been concerned with changes in the persistence of a univariate time series, considering primarily a shift from a unit root process $[I(1)]$ to a stationary process $[I(0)]$ or vice versa at some unknown date over the sample under consideration. In that strand of the literature the analysis centers on the properties of estimators (and tests) in these extreme cases, see Perron (2006) for a survey of testing procedures and Chong (2001) and Kejriwal and Perron (2010) for some recent results on the properties of the break estimators. These models however deal with the extreme dichotomy of $[I(1)]$ versus $[I(0)]$ and do not allow for long memory and fractional integration.

Models that allow for different long memory regimes have been used in the literature but the regime switching is forced by an observable state variable as opposed to the latent nature of the state variable in the MS–ARFIMA model, see Haldrup and Nielsen (2006). The main motivation behind the TH MS–ARFIMA model has been the observation by Diebold and Inoue (2001) that a mixture model of latent Markov-switching mean can generate long memory dependence. In other words, structural change and long memory may be easily confused in estimation. Hence, the main emphasis of the TH approach has been to disentangle the impact of long memory dependence on the estimates of the latent regime parameters in the MS–ARFIMA framework. It is worth noting that in this context the direct application of the EM algorithm used by Hamilton (1989) and Hamilton (1990) is not applicable due to the non-Markovian nature of the model. One of the main contributions of TH is the use of the Viterbi algorithm to estimate the MS–ARFIMA model. This algorithm is capable of tackling the hidden Markov process observed in a general ARFIMA framework, something that is not generally possible with the EM algorithm. The algorithm discussed above is not the only choice practically available in estimation of the parameters of Markov regime switching model. Other widely used algorithms are, for example, the forward-backward algorithm, the Baum-Welch algorithm and the BCJR algorithm. The forward-backward algorithm is an inference algorithm for hidden Markov models which computes the posterior marginals of all hidden state variables given a sequence of observations. The Baum-Welch algorithm is a special forward-backward algorithm and also relies on EM algorithm. The BCJR algorithm, also called Maximum posteriori probability

(MAP) decoder, named after Bahl et al. (2006), relies on maximization of posteriori probabilities. These algorithm has several modified and tailored versions*.

However, the TH analysis did not consider regime switching in the long memory parameter but only in the mean both in their simulations and their empirical application. Yet long memory parameter regime switching may have similar contamination effects on the estimation of the mean parameters as it would be the case in the opposite case considered by TH. This possibility in fact was indicated by Diebold and Inoue (2001) concern mentioned above. In this note we explicitly consider the case where the long memory parameter may be subject to regime switching, while the mean can be unchanging or regime switching itself. By allowing the direct impact of long memory regime switching on the mean parameters (whether in regime switching mode or not) would allow us to asses the possible contamination and impact that such long memory structural break may have on the mean estimates. We conduct a Monte Carlo simulation that considers contamination to be going both ways between mean and long memory parameter breaks. Our results suggest that in addition to the findings by TH that only considered breaks in the mean parameter, breaks in the long memory parameter can have similar effects on the (in sample) fitting ability of the model irrespective of the presence of breaks in the mean parameter, confirming the contamination concern raised by Diebold and Inoue (2001).

The rest of the paper is organized as follows. In the next section we present the model that we analyze as well as the Viterbi algorithm we use following TH. We then proceed to present our simulations that analyze the possible contamination that could run from long memory parameter breaks to mean parameter breaks and vice-versa. Finally we conclude.

2. Viterbi maximum likelihood EM algorithm

Consider the fractionally integrated process y_t defined as

$$(1-L)^d y_t = \mu + \varepsilon_t \tag{1}$$

where ε_t is white noise, L is the lag operator, d is the fractional integration parameter and μ is real valued drift.

Let $y_1 = 0$ and $\hat{y} = \phi_{t1} y_{t-1} + \ldots + \phi_{t1} y_1$ be the one-step predictors of the process y_t. The coefficients has the following recursive structure:

$$\begin{aligned} \phi_{tt} &= \left[\gamma(t) - \sum_{i=1}^{t-1} \gamma(t-i)\right] \nu_{t-1} \\ \phi_{tj} &= \phi_{tj} - \phi_{tt}\phi_{t-1t-j}, \qquad j = 1, \ldots, T-1 \\ \nu_{tj} &= \nu_{t-1}(1-\phi_{tt}^2), \qquad t = 1, \ldots, T-1 \end{aligned} \tag{2}$$

where $\gamma(t)$ is the autocovariance function of order t, $\nu_0 = \gamma_0$. In a fractionally integrated processes

*Examining the performance of these other algorithms is beyond the scope of the present note and is left for future research.

such as specified in Equation 1, we have

$$\gamma(t) = \frac{\Gamma(1-2d)\Gamma(d-t)}{\Gamma(d)\Gamma(1-d)\Gamma(1-d-t)}$$

where $\Gamma(x)$ is the gamma function.

Defining the prediction error $e_t = y_t - \hat{y}_t$, then $e_t = Ly$ where L is

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -\phi_{11} & 1 & 0 & \dots & 0 \\ -\phi_{22} & -\phi_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi_{T-1T-1} & -\phi_{T-1T-2} & -\phi_{T-1T-3} & \dots & 1 \end{bmatrix} \quad (3)$$

Setting $\Gamma_\theta = LDL'$, where D is diagonal matrix with $diag(\nu_0, \dots, \nu_{T-1})$, we have $\det \Gamma_\theta = \prod_{j=1}^{n} \nu_{j-1}$. Consequently, $Y'\Gamma_\Theta^{-1}Y = e'^{-1}e$, where $Y = (y_1, \dots, y_T)$ and $e_t = y_t - \hat{y}_t$. Then, the log-likelihood function may be written as

$$\mathcal{L}(\theta) = -\frac{1}{2}\sum_{t=1}^{T} \log \nu_{t-1} - \frac{1}{2}\sum_{t=1}^{T} \frac{e_t^2}{\nu_{t-1}} \quad (4)$$

for the model specified in Equation 1.

Now, we consider a 2−state homogeneous Markov chain S_t taking values 1 or 2. Let $S_{t t=1}^{T}$ be the latent sample path of the Markov chain. At each time point t, S_t can assume only an integer value of 1 or 2, and its transition probability matrix is

$$\mathcal{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \quad (5)$$

where $p_{ij} = \mathbb{P}(S_t = j \mid S_{t-1} = i)$ and $p_{i1} + p_{i2} = 1$ for $i = 1, 2$.

We specify the corresponding regime switching fractionally integrated process as

$$(1-L)^{d^{s_t}} y_t = \mu^{s_t} + \varepsilon_t^{s_t} \quad (6)$$

with the unobserved state vector $S = (s_t)_{1 \le t \le T}$.

In this study, we are going to employ the Viterbi algorithm to estimate the unobserved state vector.S. The Viterbi algorithm is forward decoding procedure widely used in signal processing problems with Hidden Markov Models (HMM) specification. The main idea behind this algorithm is to "decode" the sequence of states s_t in the Markov chain iteratively starting from time 1 to T. Given the previous state, s_{t-1}, and the unconditional likelihood function for the observations up to t, $(y_s)_{0 \le s \le t}$ and a parameter vectors ξ_1 and ξ_2, it enables us to choose the most probable state at t, i.e., s_t.

In particular, specifiying the unconditional likelihood functions for both states at t as

$$l_t^j = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(y_t - \hat{y}_t^{(j)})^2}{2\sigma_j^2}} \tag{7}$$

where $\hat{y}_t^{(j)} = \hat{y}_t^{(1)}(\xi_j,\ (y_u)_{1\leq u\leq t})$ and $\xi_j = (\mu_j, d_j, \sigma_j,),\ j = 1, 2$ and it can be calculated by $Ly + y$. Given the likelihood function,, the parameter vectors ξ^j, and the transition probabilities P, by Viterbi specification we can iteratively estimate the path of the states s_t as follows:

$$\mathcal{S}_t = \arg\max_{s_t \in \{1,2\}} (l_t^{s_t} P(s_t | \mathcal{S}_{t-1})) \tag{8}$$

given s_0 and the initial probabilities $\pi_j = \mathbb{P}(s_0 = j)$, and $\hat{y_0}$, where $\mathcal{S}_\sqcup$ is the Viterbi decoded state at t.

Then, the estimation of parameter vectors ξ_j and the transition probabilities p_{ij} is obtained by maximixzing the following log-likelihood function:

$$\mathcal{L}(\xi_1, \xi_2, y) = -\frac{1}{2}\sum_{t=1}^{T} \log \sigma_j - \frac{1}{2}\sum_{t=1}^{T} \frac{e_t^2}{\sigma_j} + \sum_{t=1}^{T} \mathbb{P}(\mathcal{S}_t | \mathcal{S}_{t-1}) \tag{9}$$

The computational complexity of this maximization is $\mathcal{O}(n^2)$. In comparison, the EM algorithm of Hamilton (1989), which is exhaustively used in estimation of Markov regime switching models, cannot be employed in this case (when d is allowed to be state depended) for one main reason: the number of possible state paths the EM algorithm is going to account for is n^T (since $n = 2$ in our case, it is 2^T, hence, implying a computational complexity of $O(2^T)$. Thus, EM is computationally much more demanding in comparison to Viterbi algorithm, which is $\mathcal{O}(n^2)$ in computational complexity. Furthermore, Tsay (2008) argues, that the model we examine above "... cannot be written in a state-space form due to the presence of a fractional differencing parameter, implying that we cannot apply the EM algorithm considered in Hamilton (1990) ... because the non-Markovian nature of the model prevents us from using the results in (4.2) of Hamilton (1990)" (Tsay, 2008, page. 3).

TH made a similar argument that the EM algorithm proposed by Hamilton (1989) is unstable for estimating the Markov regime switching d. We confirmed that by conducting a small scale computation exercise using the EM algorithm†.

In the context of the present model, even though the log-likelihood function given by Equation 9, is well-defined, the properties of the estimators of the long-memory parameter and the mean level under regime switching are not derived. We conjecture that an analytical derivation of the

†However, , we made a naïve experiment to estimate d_{s_t} by Hamilton (1989) method, for a series generated by the process $(1-L)^{d^{s_t}}\ y_t = \mu^{s_t} + \varepsilon_t$ with the true values of $d_1 = 0, d_2 = 0.5, \mu_1 = 0, \mu_2 = 0$ and $\varepsilon_t \sim iidN(0, 0.25)$. We obtained $\widehat{d_1} = 0.142, \widehat{d_2} = 0.411, \widehat{\mu}_2 = -0.12, \widehat{\mu}_2 = 0.42$, and $\sigma_{\varepsilon_t}^2 = 0.32$. From a computational perspective, it also took much longer to compute the estimators than the Viterbi algorithm did; with a 2.7 GHz i7 single thread processor it takes 1.8 times longer time than that of the Viterbi algorithm computation.

asymptotic distribution of these estimators might entail an analytical framework such as Hansen (2000).

We now proceed with the application of the Viterbi algorithm in a Monte Carlo simulation where we will study breaks in both long memory and mean parameters.

3. The Monte Carlo experiment

The Monte Carlo experiments that we conduct in this note is to consider the simplest possible framework of analysis an MS-ARFIMA $(0, d, 0)$ to do the analysis as we are interested in isolating the effects of spillovers between the regime shifts of the two simple parameters in the model, μ and d. We carry out the Monte Carlo experiment for the Cartesian product space $\mu_1 \times \mu_2 \times d_1 \times d_2$ of true μ_1, μ_2, d_1, d_2 vales such that $\mu_i \in \{0, 0.5, 1.0\}$, $d_i \in \{-0.50, 0.00, 0.25, 0.50, 0.75, 1.00\}$ for $i = 1, 2$.

First, we generated the series y_t

$$(1-L)^{d^{s_t}}\, y_t = \mu^{s_t} + \varepsilon_t$$

where $s_t = 1$ if $t \leq T/2$ and 2 otherwise; ε_t are *iid* and distributed $N(0, 0.25)$; T is 256. d_1, d_2, μ_1, and μ_2 are selected such that $d_i \in \{-0.50, 0.00, 0.25, 0.50, 0.75, 1.00\}$, $\mu_i \in \{0.0, 0.5, 1.0\}$ for $i = 1, 2$.

Then, we obtain the estimates of $\widehat{d_1}, \widehat{d_2}, \widehat{\mu}_2, \widehat{\mu}_2$, and $\widehat{\sigma}_{\varepsilon t}$ values by the procedure described above. We conduct 1000 replications for the (d_1, d_2, μ_1, μ_2) quadruple.

The results are presented in Tables 1 to 4 and Figure 1. Table 1 simply presents the average parameter estimates over the total number of replications, Table 2 presents the root mean square errors for the parameter estimates, Table 3 the mean absolute biases and finally Table 4 the in sample root mean squared errors of the model fit. The results of Table 4 are best seen in Figure 1 and it becomes clear that whether there is a break in the mean (different values of μ_1, μ_2) or not the model fit seems to be unaffected (as measured by the root mean squared error). In other words we observe that structural breaks in the long memory parameter produce similar in sample estimation patterns irrespective of the presence of breaks in the mean parameter. Hence, these results confirm the original concern raised by Diebold and Inoue (2001), something that was not possible to be seen in the work of TH who only considered in his simulation study breaks in the mean but not in the long memory parameter d. The upshot of our analysis is that one has to be careful in claiming the occurrence of structural breaks in such models as the presence of such breaks may be difficult to disentangle. For that purpose one would need to develop more appropriate joint testing procedures that would be able to discern the nature of possible such breaks. The results also confirm that as d, the long memory parameter gets close to unity, the estimates of the mean level parameter μ tend to become ininconsistentonsistent as expected in the presence of unit roots.

4. Conclusion

In this note we consider a general Markov-regime switching estimation both in the long memory parameter d and the mean of a time series. Following Tsay and Härdle (2009) we employ the Viterbi algorithm in a two state Markov-switching parameter estimation. Since it has been suggested in the literature that mean and long memory parameter breaks in time series can be easily confused with each other Diebold and Inoue (2001), our aim is to produce evidence for this assertion by means of a Monte Carlo simulation that considers both breaks in both the mean and the long memory parameters. Our results confirm the above assertion, since we observe that structural breaks in the long memory parameter produce similar in sample estimation patterns irrespective of the presence of breaks in the mean parameter. Our results are complementary to TH who only considers breaks in the mean parameter in his simulation study. Our results call for the development for joint tests that would be able to discern the nature of possible such breaks. However, given the results of our analysis there are many outstanding issues that we leave for future research. First, a more comprehensive Monte Carlo simulation study would need to allow for a more general MS-ARFIMA (p, d, q)structure. Furthermore, one would need to investigate the asymptotic properties of the estimators of μ and $d$we obtain under regime switching. Even though the likelihood function is well defined, in the context of unknown breaks, we conjecture that one would need to use a framework that is used in threshold models with unknown thresholds, see Hansen (2000). An investigation of such a problem is beyond the scope of this paper and is left for future research.

References

Bahl, L., J. Cocke, F. Jelinek, and J. Raviv (2006, September). Optimal decoding of linear codes for minimizing symbol error rate (corresp.). *IEEE Trans. Inf. Theor. 20*(2), 284–287.

Cajueiro, D. O. and B. M. Tabak (2005a). Possible causes of long-range dependence in the brazilian stock market. *Physica A: Statistical Mechanics and its Applications 345*(3), 635–645.

Cajueiro, D. O. and B. M. Tabak (2005b). Ranking efficiency for emerging equity markets ii. *Chaos, Solitons & Fractals 23*(2), 671–675.

Cajueiro, D. O. and B. M. Tabak (2005c). Testing for time-varying long-range dependence in volatility for emerging markets. *Physica A: Statistical Mechanics and its Applications 346*(3), 577–588.

Cajueiro, D. O. and B. M. Tabak (2007a). Long-range dependence and multifractality in the term structure of libor interest rates. *Physica A: Statistical Mechanics and its Applications 373*, 603–614.

Cajueiro, D. O. and B. M. Tabak (2007b). Time-varying long-range dependence in us interest rates. *Chaos, Solitons & Fractals 34*(2), 360–367.

Chong, T. T.-L. (2001). Structural change in ar (1) models. *Econometric Theory 17*(01), 87–155.

Diebold, F. X. and A. Inoue (2001). Long memory and regime switching. *Journal of econometrics 105*(1), 131–159.

Ding, Z., C. W. Granger, and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance 1*(1), 83–106.

Granger, C. W. and Z. Ding (1994). Stylized facts on the temporal and distributional properties of daily data from speculative markets. *UCSD Department of Economics Discussion Paper*, 94–19.

Haldrup, N. and M. O. Nielsen (2006). A regime switching long memory model for electricity prices. *Journal of econometrics 135*(1), 349–376.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357–384.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of econometrics 45*(1), 39–70.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica 68*(3), 575–603.

Kejriwal, M. and P. Perron (2010). Testing for multiple structural changes in cointegrated regression models. *Journal of Business & Economic Statistics 28*(4), 503–522.

Limam, I. (2003). Is long memory a property of thin stock markets? international evidence using arab countries. *Review of Middle East Economics and Finance 1*(3), 251–266.

Lobato, I. N. and C. Velasco (2000). Long memory in stock-market trading volume. *Journal of Business & Economic Statistics 18*(4), 410–427.

Lux, T. and T. Kaizoji (2007). Forecasting volatility and volume in the tokyo stock market: Long memory, fractality and regime switching. *Journal of Economic Dynamics and Control 31*(6), 1808–1843.

Perron, P. (2006). Dealing with structural breaks. *Palgrave handbook of econometrics 1*, 278–352.

Tsay, W.-J. (2008). Analysing inflation by the arfima model with markov-switching fractional differencing parameter.

Tsay, W.-J. and W. K. Härdle (2009). A generalized arfima process with markov-switching fractional differencing parameter. *Journal of Statistical Computation and Simulation 79*(5), 731–745.

Willinger, W., M. S. Taqqu, and V. Teverovsky (1999). Stock market prices and long-range dependence. *Finance and stochastics 3*(1), 1–13.

Tables and Figures

Table 1: True parameters and estimated parameters by the Monte Carlo study.

d	$(\mu_1,\mu_2)=(0,0)$				$(\mu_1,\mu_2)=(0,0.5)$				$(\mu_1,\mu_2)=(0,1)$			
	d_1	d_2	μ_1	μ_2	d_1	d_2	μ_1	μ_2	d_1	d_2	μ_1	μ_2
(-0.5,-0.5)	0.080	-0.344	-0.083	0.011	-0.505	-0.335	-0.042	0.563	-0.472	-0.363	0.032	0.918
(-0.5, 0)	-0.566	-0.156	-0.065	0.149	-0.407	-0.161	0.099	0.532	-0.172	0.068	0.015	0.905
(-0.5, 0.25)	-0.518	-0.235	-0.095	0.158	-0.455	-0.161	0.099	0.545	-0.472	0.078	0.015	0.905
(-0.5, 0.5)	-0.580	0.530	-0.135	0.003	-0.627	0.445	-0.014	0.524	-0.595	0.388	0.013	0.896
(-0.5, 0.75)	-0.510	0.730	-0.091	0.003	-0.627	0.725	-0.014	0.558	-0.595	0.788	0.010	0.913
(-0.5, 1)	-0.415	0.906	0.047	0.172	-0.338	1.080	0.161	0.717	-0.466	0.844	-0.017	1.036
(0.0, 0.0)	-0.096	-0.017	-0.240	-0.024	-0.256	0.141	-0.079	0.562	-0.109	0.095	-0.057	0.880
(0.0, 0.25)	-0.106	-0.252	-0.014	0.008	0.144	0.254	-0.079	0.562	0.110	0.245	-0.055	0.895
(0.0, 0.5)	0.028	0.425	-0.082	-0.099	0.015	0.496	-0.111	0.594	0.108	0.553	-0.004	0.927
(0.0, 0.75)	0.017	0.785	-0.082	-0.099	0.115	0.725	-0.111	0.524	0.095	0.714	-0.074	0.927
(0.0, 1.0)	0.058	1.008	0.127	-0.071	0.282	0.963	-0.175	0.573	0.041	0.971	-0.094	1.387
(0.5, 0.5)	0.524	0.518	-0.136	0.056	0.542	0.536	-0.074	0.529	0.448	0.549	0.089	1.207
(0.5, 0.75)	0.556	0.768	-0.106	0.046	0.552	0.752	-0.074	0.519	0.476	0.746	0.029	1.207
(0.5, 1)	0.465	0.934	-0.140	0.213	0.528	1.036	-0.071	0.528	0.513	1.132	-0.013	1.155
(.25,1)	0.235	0.964	-0.119	0.115	0.215	1.014	-0.073	0.582	0.234	1.175	-0.053	1.154
(.75,1)	0.765	0.957	-0.136	0.103	0.765	1.217	-0.071	0.608	0.713	1.107	-0.031	1.168
(1, 1)	1.149	1.188	-0.015	-0.232	0.826	1.004	0.082	0.681	0.950	1.198	-0.139	0.996

Table 2: Root mean squared errors for the parameters computed as $\sqrt{N^{(-1)}\sum_i^N(\widehat{x_i}-x)^2}$ where x may be d_1, d_2, μ_1, μ_2

d	$(\mu_1,\mu_2)=(0,0)$				$(\mu_1,\mu_2)=(0,0.5)$				$(\mu_1,\mu_2)=(0,1)$			
	d_1	d_2	μ_1	μ_2	d_1	d_2	μ_1	μ_2	d_1	d_2	μ_1	μ_2
(-0.5,-0.5)	0.580	0.156	0.083	0.011	0.005	0.165	0.042	0.063	0.028	0.137	0.032	0.082
(-0.5, 0)	0.066	0.156	0.065	0.149	0.093	0.161	0.099	0.032	0.328	0.068	0.015	0.095
(-0.5, 0.25)	0.018	0.485	0.095	0.158	0.045	0.411	0.099	0.045	0.028	0.172	0.015	0.095
(-0.5, 0.5)	0.080	0.030	0.135	0.003	0.127	0.055	0.014	0.024	0.095	0.112	0.013	0.104
(-0.5, 0.75)	0.010	0.020	0.091	0.003	0.127	0.025	0.014	0.058	0.095	0.038	0.010	0.087
(-0.5, 1)	0.085	0.094	0.047	0.172	0.162	0.080	0.161	0.217	0.034	0.156	0.017	0.036
(0.0, 0.0)	0.096	0.017	0.240	0.024	0.256	0.141	0.079	0.062	0.109	0.095	0.057	0.120
(0.0, 0.25)	0.106	0.502	0.014	0.008	0.144	0.004	0.079	0.062	0.110	0.005	0.055	0.105
(0.0, 0.5)	0.028	0.075	0.082	0.099	0.015	0.004	0.111	0.094	0.108	0.053	0.004	0.073
(0.0, 0.75)	0.017	0.035	0.082	0.099	0.115	0.025	0.111	0.024	0.095	0.036	0.074	0.073
(0.0, 1.0)	0.058	0.008	0.127	0.071	0.282	0.037	0.175	0.073	0.041	0.029	0.094	0.387
(0.5, 0.5)	0.024	0.018	0.136	0.056	0.042	0.036	0.074	0.029	0.052	0.049	0.089	0.207
(0.5, 0.75)	0.056	0.018	0.106	0.046	0.052	0.002	0.074	0.019	0.024	0.004	0.029	0.207
(0.5, 1)	0.035	0.066	0.140	0.213	0.028	0.036	0.071	0.028	0.013	0.132	0.013	0.155
(.25,1)	0.015	0.036	0.119	0.115	0.035	0.014	0.073	0.082	0.016	0.175	0.053	0.154
(.75,1)	0.015	0.043	0.136	0.103	0.015	0.217	0.071	0.108	0.037	0.107	0.031	0.168
(1, 1)	0.149	0.188	0.015	0.232	0.174	0.004	0.082	0.181	0.050	0.198	0.139	0.004

Table 3: Absolute estimation errors computed as $N^{(-1)} \sum_i^N |\widehat{x}_i - x|$ where x is d_1, d_2, μ_1, μ_2.

d	$(\mu_1, \mu_2) = (0, 0)$				$(\mu_1, \mu_2) = (0, 0.5)$				$(\mu_1, \mu_2) = (0, 1)$			
	d_1	d_2	μ_1	μ_2	d_1	d_2	μ_1	μ_2	d_1	d_2	μ_1	μ_2
(-0.5,-0.5)	0.580	0.158	0.130	0.101	0.025	0.166	0.107	0.119	0.035	0.139	0.104	0.132
(-0.5, 0)	0.070	0.157	0.119	0.179	0.096	0.163	0.141	0.107	0.328	0.072	0.097	0.138
(-0.5, 0.25)	0.029	0.485	0.139	0.188	0.053	0.412	0.140	0.112	0.035	0.173	0.103	0.139
(-0.5, 0.5)	0.084	0.036	0.170	0.101	0.129	0.058	0.100	0.102	0.098	0.116	0.098	0.145
(-0.5, 0.75)	0.025	0.032	0.134	0.102	0.129	0.034	0.099	0.114	0.098	0.045	0.100	0.133
(-0.5, 1)	0.088	0.096	0.111	0.199	0.163	0.083	0.188	0.239	0.040	0.158	0.102	0.107
(0.0, 0.0)	0.101	0.032	0.260	0.104	0.257	0.143	0.129	0.115	0.111	0.098	0.116	0.157
(0.0, 0.25)	0.108	0.502	0.101	0.100	0.146	0.028	0.128	0.114	0.113	0.022	0.115	0.146
(0.0, 0.5)	0.037	0.081	0.127	0.139	0.030	0.021	0.150	0.135	0.110	0.057	0.098	0.125
(0.0, 0.75)	0.031	0.041	0.130	0.140	0.117	0.045	0.151	0.105	0.098	0.042	0.124	0.123
(0.0, 1.0)	0.062	0.025	0.161	0.125	0.283	0.050	0.202	0.124	0.047	0.039	0.136	0.400
(0.5, 0.5)	0.035	0.029	0.168	0.114	0.048	0.043	0.127	0.107	0.059	0.053	0.133	0.229
(0.5, 0.75)	0.062	0.028	0.147	0.112	0.057	0.023	0.126	0.104	0.035	0.021	0.106	0.229
(0.5, 1)	0.042	0.070	0.172	0.232	0.038	0.047	0.122	0.104	0.030	0.135	0.097	0.185
(.25,1)	0.027	0.043	0.156	0.153	0.041	0.026	0.124	0.130	0.025	0.177	0.111	0.184
(.75,1)	0.029	0.050	0.168	0.141	0.026	0.218	0.125	0.148	0.043	0.110	0.103	0.197
(1, 1)	0.151	0.190	0.097	0.252	0.175	0.021	0.124	0.208	0.053	0.199	0.171	0.097

Table 4: In-sample mean squared errors

Cases	(d_1, d_2)	$(\mu_1, \mu_2) = (0, 0)$	$(\mu_1, \mu_2) = (0, 0.5)$	$(\mu_1, \mu_2) = (0, 1)$
1	(-0.5,-0.5)	0.2582	0.3142	0.3573
2	(-0.5, 0)	0.1879	0.1247	0.1611
3	(-0.5, 0.25)	0.3402	0.2915	0.3296
4	(-0.5, 0.5)	0.3960	0.3810	0.2860
5	(-0.5, 0.75)	0.4560	0.4230	0.4360
6	(-0.5, 1)	0.5930	0.5634	0.5417
7	(0.0, 0.0)	0.0909	0.1302	0.1297
8	(0.0, 0.25)	0.1083	0.1473	0.1104
9	(0.0, 0.5)	0.1020	0.1278	0.1160
10	(0.0, 0.75)	0.3099	0.2571	0.2152
11	(0.0, 1.0)	0.5420	0.5930	0.6020
12	(0.25, 0.25)	0.1064	0.1000	0.1219
13	(0.25, 0.5)	0.1483	0.2318	0.1186
14	(0.25, 0.75)	0.3227	0.1665	0.2251
15	(0.25, 1.0)	0.5081	0.5770	0.6555
16	(0.5,0.5)	0.4953	0.4434	0.4674
17	(0.5, 0.75)	0.6573	0.7166	0.7883
18	(0.5, 1)	0.8193	0.8158	0.8699
19	(0.75, 0.75)	0.7105	0.6416	0.7095
20	(0.75, 1.0)	0.8044	0.7900	0.8426
1	(1, 1)	0.9674	0.9410	0.9839

Figure 1: Average in-sample RMSE of 1000 replications for different (d_1, d_2, μ_1, μ_2) quadruple cases.