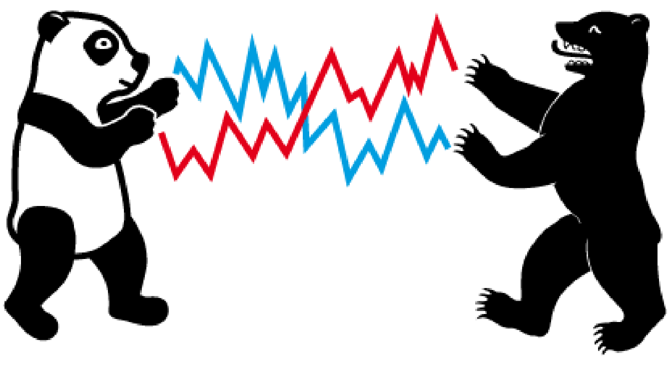
IRTG 1792 Discussion Paper 2018-020

A Regime Shift Model with Nonparametric Switching Mechanism

Haiqiang Chen *
Yingxing Li *
Ming Lin *
Yanli Zhu *[2]



* Xiamen University, People'S Republic of China
*[2] Hohai University, People'S Republic of China

This research was supported by the Deutsche Forschungsgemeinschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

A Regime Shift Model with Nonparametric Switching Mechanism

Haiqiang Chen[a], Yingxing Li[a], Ming Lin[a] and Yanli Zhu[b]

[a] Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian 361005, China.

[b] Business School, Hohai University, Nanjing, Jiangsu 211100, China.

May 16, 2018

Abstract

In this paper, we propose a new class of regime shift models with flexible switching mechanism that relies on a nonparametric probability function of the observed threshold variables. The proposed models generally embrace traditional threshold models with contaminated threshold variables or heterogeneous threshold values, thus gaining more power in handling complicated data structure. We solve the identification issue by imposing either global shape restriction or boundary condition on the nonparametric probability function. We utilize the natural connection between penalized splines and hierarchical Bayes to conduct smoothing. By adopting different priors, our procedure could work well for estimations of smooth curve as well as discontinuous curves with occasionally structural breaks. Bayesian tests for the existence of threshold effects are also conducted based on the posterior samples from Markov chain Monte Carlo (M-CMC) methods. Both simulation studies and an empirical application in predicting the U.S. stock market returns demonstrate the validity of our methods.

Key Words: Threshold Model, Nonparametric, Markov Chain Monte Carlo, Bayesian Inference, Spline.

JEL Classification:

*The authors acknowledge the supports from NSFC with grant numbers #71571152, #71571154, #11101341 and the Fundamental Research Funds for the Central Universities with grant number 2015B13114. E-mail addresses: hqchen2009@gmail.com (H. Chen), amyli999@hotmail.com (Y. Li), ming.lin50@gmail.com (M. Lin) and zhuyanli_921@126.com (Y. Zhu).

1 Introduction

Since the seminal work of Tong and Lim (1980) and Tong (1983), threshold models have been widely applied to analyze various kinds of economic data.[1] With the use of segmentation, it maintains the merit of simplicity as linear models do, but also captures various types of nonlinear phenomena, such as business cycles, jumps and asymmetry. Given their success in both theoretical and empirical studies, threshold models have attracted a remarkable amount of research attention and being extended in different ways, including the smooth transition threshold model (STAR) by Chan and Tong (1986), the double-threshold ARCH model by Li and Li (1996), the multiple-regimes threshold model by Tsay (1998). The estimation and asymptotics for threshold models as well as their variants have been well established by a sequence of studies, including Chan (1993), Hansen (1997), Tsay (1998), Seo and Linton (2007) and Gonzalo and Pitarakis (2012). A nice summary about the development of threshold models in last 30 years can be found in Hansen (2011) and Tong (2011).

Classical threshold models investigated in the aforementioned studies define regimes in a deterministic way, relying on an observed threshold variable relative to a fixed threshold value. Such a mechanism has its economic interpretation, but might not always be satisfied in empirical applications. For instance, it excludes the possibility of the existence of status mixtures. Moreover, it might be a difficult task to find such an appropriate threshold variable, or the threshold variable may be subjected to measurement errors. In addition, the homogeneity assumption on the threshold value may be too strong. All of the above challenges require more flexible regime switch mechanism. Quandt (1972) considers the pure random switching model which requires no observed threshold variable. Hamilton (1989) generalized the switching mechanism and proposed the Markov switching model, which assumes the latent regime state follows a Markov process. Such a model becomes a popular tool to analyze business cycles and regime shifts existing in various type of asset prices. However, the merit of requiring no threshold variable might also lead to its drawback, as the results are hard to interpret due to the untractable regime status.

In this paper, we propose a new class of regime shift models. Instead of modeling how

[1]For example, Durlauf and Johnson (1995) and Hansen (2000) use threshold regressions to study the multiple stages in economic growth models. Potter (1995), Tsay (1998) and Chen et al. (2012) apply threshold autoregressive models (TAR) to analyze the GDP growth rates, unemployment rates, inflation rates and stock returns.

the unobserved regime status transits, we focus on studying how the observed threshold variable and some latent factors, such as measurement errors or individual preference heterogeneity, affect the status of the regime simultaneously. Our model could be viewed as a partially-hidden switching model that combines both threshold models and random switching models. We adopt a nonparametric probability function to help classify the unobserved regime status. A closely related work is Wu and Chen (2007), but they adopt the parametric logistic link to model the relationship between the regime status and the threshold variable. Though our model might be more involved, the nonparametric setting enables us to enjoy more flexibility in model specification, and it also frees us from conducting various transformations to determine the appropriate form of the threshold variable. It could be shown that many popular models, including the random switching models and the classical threshold models, are our special cases. Moreover, we could handle the complicated situations when measurement errors or latent heterogeneity or time varying effects present. Hence the main advantage of our approach is obvious, as it provides a unified way to handle various complicated data structures and allows the data to reveal which might be the true nature of the switching mechanism.[2]

For the sake of identification, we propose a novel approach, which impose restrictions on the nonparametric probability function rather than on the regression coefficients associated with different regimes. Though it may be a conventional way to assume, for example, the first regression coefficient in status I is less than that in status II (Kim and Nelson, 1999; Wu and Chen, 2007; Henkel et al., 2011), such an approach might perform poorly if indeed the unknown regression coefficients are close to each other. Moreover, it implicitly assumes the existence of multiple regimes and thus hinders inference on the existence of threshold effects. In contrast, we avoid such a caveat by imposing constraints on the conditional probability of falling into status I given the observed threshold variable. We discuss various conditions for model identification, and demonstrate how to incorporate the constraints for estimation. In particular, two different types of constraints are considered. One is to assume some boundary condition, while the other is to assume a monotonic pattern. The former is less restrictive, while the latter explores the global feature of the underlying probability function and may

[2]Another related work is Cai et al. (2015), who extend STAR models by allowing the state variables to enter into the transition function in a nonparametric way. Their model can be regarded as a time-varying weighted average of two linear models, and the estimation can be obtained by least squares methods as no latent status variable is involved.

be more efficient if the monotonic condition holds. Estimating a nonparametric function with shape restrictions has also been studied by many authors, for instance, Ramsay (1998) and Ait-Sahalia and Duarte (2003). To our best knowledge, we may be the first to consider this in the regime shift model.

We employ a Bayesian approach as well as MCMC techniques for estimation. We model the nonparametric probability function via some transformation of the spline function, and utilize the natural connection between penalized splines and hierarchical Bayes to conduct smoothing. Moreover, we extend the choices of prior for the spline coefficients from the Gaussian distribution to some heavy-tailed distribution, so the approach could work for estimating smooth curve as well as discontinuous curve with occasionally jumps. It turns out our Bayesian approach is in the spirit of conditional penalized likelihood aiming at smoothing or variable selection. In the latter case, we also recommend choosing different penalty parameter such that the estimate could be locally adaptive. Based on the sample posterior, we also consider the nonstandard test on the existence of threshold effects. Simulations show the validity of our method. We also apply our method to analyze the U.S. stock market return. The results indicate that our approach could bring better out-of-sample predictability compared to various benchmark models.

The remainder of the paper is organized as follows. Section 2 introduces the model setup and inference procedure. Section 3 presents Monte Carlo simulations. Section 4 provides the empirical application, while Section 5 concludes.

2 Modeling and Inference

2.1 The Model

Consider the following two-state regime switching model

$$y_t = x_t'\beta_{I_t} + \sigma_{I_t}\varepsilon_t,\ t = 1, \cdots, T \tag{1}$$

where x_t's are exogenous covariates, ε_t's are i.i.d noise terms, with mean zero and variance 1. Each of the state variable I_t takes value either 1 or 2, *i.e.*, the model switches between two regimes. Parameters β_{I_t}'s and σ_{I_t}'s are the regression coefficients and the standard deviation, respectively. When I_t's are independent Bernoulli random variables, the model (1) is the random switching model proposed by Quandt (1972). If I_t is defined as $1 + \mathbf{1}(z_t < \gamma)$ such that the regime status is 1 when the threshold variable z_t is no less than a fixed value γ, the model (1) becomes the classical threshold model.

In this paper, we propose to model the regime switching function by the following nonparametric probability function, defined by

$$P(I_t = 1|z_t) = g(z_t) \quad \text{and} \quad P(I_t = 2|z_t) = 1 - g(z_t),$$

the model (1) becomes

$$y_t = \begin{cases} x_t'\beta_1 + \sigma_1\varepsilon_t, & \text{with probability } g(z_t); \\ x_t'\beta_2 + \sigma_2\varepsilon_t, & \text{with probability } 1 - g(z_t), \end{cases} \tag{2}$$

where z_t is an observed threshold variable. Obviously, the random switching model and the threshold model are two special cases with $g(z_t)$ as a constant or a step function, respectively. With various forms of $g(z_t)$, the model (2) could embrace many different models with complex data structure. The following two examples are some typical cases.

Example 1: Threshold model with heterogenous threshold values.

In some applications of threshold models, the threshold values could vary across individuals or over time. For example, in marketing science, an active research area is to study the asymmetric price effects to consumers' behavior, see Helson (1964), Kahneman and Tversky (1979), Chang et al. (1999) and Bell and Lattin (2000). Terui and Dahana (2006) consider a choice model in which consumers switch their utility function among different forms according to the relationship between the retail price and the reference price, an unknown price threshold. They consider incorporating a random component on the price threshold value in order to capture the heterogeneity of consumers' preference. In general, one may set up the following threshold model with heterogenous threshold values.

$$y_t = \begin{cases} x_t'\beta_1 + \sigma_1\varepsilon_t, & \text{if } z_t \geq \gamma_t; \\ x_t'\beta_2 + \sigma_2\varepsilon_t, & \text{if } z_t < \gamma_t, \end{cases}$$

where $\gamma_t = \gamma + u_t$ and u_t is independent of z_t. Suppose u_t has the cumulative distribution function $F_u(\cdot)$, then the regime switching function is

$$g(z_t) = P(z_t \geq \gamma + u_t \mid z_t) = P(u_t \leq z_t - \gamma \mid z_t) = F_u(z_t - \gamma).$$

Example 2: Threshold model with measurement errors.

In empirical studies, it may be quite common to have measurement error, thus resulting in loss of information or misleading conclusions, see Madansky (1959), Amemiya (1985, 1990) and Lewbel (1997). One example is about the q theory. In the literature. the marginal q (the ratio of a firm's market value to the replacement value) might be treated a good predictor for firm investment. However, empirical evidence suggested the financially constrained firms may violate this relation, see Fazzari et al. (1988), Barnett and Sakellaris (1998), Hu and Schiantarelli (1998) and Hansen (1999). However, Erickson and Whited (2000) argue that these findings could be artifacts due to the measurement errors in marginal q. In general, to handle threshold models with contaminated threshold variable, we propose the following model:

$$y_t = \begin{cases} x_t'\beta_1 + \sigma_1\varepsilon_t, & \text{if } z_t^* \geq \gamma; \\ x_t'\beta_2 + \sigma_2\varepsilon_t, & \text{if } z_t^* < \gamma, \end{cases}$$

$$z_t = z_t^* + v_t$$

where z_t^* is the true threshold variable, but could not be directly observed. Instead, we have its contaminant version z_t subjected to measurement error v_t that is independent of z_t^*. One could further extend this model to incorporate measurement errors in the regressors x_t's. Let $f_{z^*}(\cdot)$ and $f_v(\cdot)$ be the density functions of z_t^* and v_t, respectively. Then the regime probability function $g(z_t)$ could be expressed as

$$g(z_t) = P(z_t^* > \gamma | z_t) = \frac{\int_\gamma^\infty f_v(z_t - z_t^*) f_{z^*}(z_t^*) dz_t^*}{\int_{-\infty}^\infty f_v(z_t - z_t^*) f_{z^*}(z_t^*) dz_t^*}.$$

From the above two examples, our nonparametric random switching model is very flexible in handling various complicated data structure. When it is not clear what is the true switching mechanism, a parametric set up might be exposed to model misspecification risk, and model estimation bias. Therefore, we prefer to model the probability function $g(z)$ nonparametrically. We shall soon describe our estimation procedures, but we need to discuss the regime identification issues first.

2.2 Model Identification

We now discuss the identification conditions required for the proposed model (2). As an illustration, consider the following two threshold models:

$$\text{Model I:} \quad y_t = \begin{cases} x_t'\beta_1 + \sigma_1\varepsilon_t, & \text{with probability } g(z_t); \\ x_t'\beta_2 + \sigma_2\varepsilon_t, & \text{with probability } 1 - g(z_t), \end{cases}$$

and

$$\text{Model II:}\quad y_t = \begin{cases} x_t'\widetilde{\beta}_1 + \widetilde{\sigma}_1\varepsilon_t, & \text{with probability } \widetilde{g}(z_t); \\ x_t'\widetilde{\beta}_2 + \widetilde{\sigma}_2\varepsilon_t, & \text{with probability } 1 - \widetilde{g}(z_t). \end{cases}$$

When $\widetilde{\beta}_I = \beta_{3-I}$, $\widetilde{\sigma}_I = \sigma_{3-I}$ for $I = 1, 2$ and $\widetilde{g}(z_t) = 1 - g(z_t)$, Model I and Model II are identical. In other words, one may not able to distinguish the two regimes. For the sake of identifiability, some previous studies suggest to impose constraints on the regression coefficients in each regime. For example, Kim and Nelson (1999) and Wu and Chen (2007) require that $\beta_{1,1} < \beta_{2,1}$ to avoid ambiguity of regime classification.

However, in practice, one may not know in advance whether $\beta_{1,1}$ and $\beta_{2,1}$ are different, or in the worst scenario, the two regimes may share the same regression coefficients with different variance. Moreover, such an approach implicitly assume the existence of two regimes. If in fact all data are generated from the same status, it may still be tempted to yield different estimates of $\beta_{1,1}$ and $\beta_{2,1}$, thus misleadingly imply the existence of threshold effects. With respect to all the concerns mentioned above, we propose to set constraints on the nonparametric function $g(z)$ instead and introduce the following definitions.

Definition 1*: A probability function $g(z)$ is said to be informative about the regime status if there exists a value $\bar{z}$ such that $g(\bar{z}) \neq 0.5$, or equivalently, $P(I_t = 1 \mid z_t = \bar{z}) \neq P(I_t = 2 \mid z_t = \bar{z})$.*

Remark 1: An informative probability function is the most general condition we need to distinguish two regimes. When $P(I_t = 1 \mid z_t = \bar{z}) \neq P(I_t = 2 \mid z_t = \bar{z})$, one could define the two regimes such that $P(I_t = 1 \mid z_t = \bar{z}) > 0.5$. When the regime probability function is not informative, *i.e.*, $g(z) \equiv 0.5$ for all z, the model is unidentifiable.

Though the existence of an informative regime probability function might be helpful for model identification, it may not be very practical as one may not know in advance where this particular value $\bar{z}$ might be. Alternatively, we discuss the following two conditions that yield an informative regime probability function.

C1*: $g(z)$ satisfies the following boundary condition: $g(z_L) \neq 0.5$ or $g(z_U) \neq 0.5$, where z_L and z_U are the lower and upper limit of the support of z_t.*

C2*: $g(z)$ satisfies the following shape restriction: $g(z)$ is a monotonic function.*

Obviously, C2 implies C1, and C1 implies the regime probability function is informative. Hence either condition is sufficient for the sake of model identifiability. Without prior knowledge, we tend to believe that it might be more likely to find that $g(z) \neq 0.5$ when z approaches extreme values. This motivates us to consider C1 as a fairly mild condition. Condition C2 is more restrictive than C1, but it could also be satisfied in various models. The following proposition shows that $g(z) = P(I_t = 1 \mid z_t = z)$ is monotonically increasing in the aforementioned two motivating examples.

Proposition 1*: In example 1, the regime probability function $g(z)$ is monotonically increasing. In example 2, $g(z)$ is monotonically increasing if the measurement error follows a normal distribution.*

The proof is presented in the Appendix.

Remark 2: In many practical settings, shape restrictions may result from theoretical properties and stylized facts. For example, utility function, cost functions and profit functions in economics are well known to be monotonic increasing, with convex or concave shapes. In medicine sciences, the dose response curve in the phase I clinical trial has an increasing trend. In survival analysis or reliability analysis, the hazard rate and the failure rate might be modeled as monotonic curves as well. In our empirical application, the regime statuses could be interpreted as the bear and bull markets, and the threshold variable is set to be the growth rate of industry production. If the industry production is in a good shape, the market may be more likely to be a bull market. Hence we assume the regime probability function $g(z)$ to be monotonically increasing. When the shape restriction assumption is reliable, our estimate of $g(z)$ may be improved since it explores the global curve feature rather than the local boundary condition.

Now we discuss how to model the probability function $g(z)$ nonparametrically under Condition C1 or C2. Note that $g(z)$ also needs to satisfy $0 \leq g(z) \leq 1$. Hence we propose to treat $g(z)$ as a transformation of a nonparametric curve $f(z)$, i.e. $g(z) = H(f(z))$ for some pre-specified function $H(\cdot)$, and convert the problem into modeling $f(z)$ nonparametrically. The choices of H are quite flexible, and most invertible functions with range between [0,1] might be considered. In particular, we recommend two choices of H as follows.

$$C1': \ g(z) = H(f(z)) = \frac{\exp(f(z))}{\exp(f(z)) + 1},$$

or

$$C2': \ g(z) = H(f(z)) = \frac{\int_a^z \exp(f(u))du}{\int_a^z \exp(f(u))du + 1}. \tag{3}$$

For either choice, we adopt the p-th truncated power polynomial splines (TPPS) to model $f(z)$, *i.e.*,

$$f(z) = \alpha_0 + \sum_{i=1}^{p} \alpha_i z^i + \sum_{i=1}^{K} b_i (z - \tau_k)_+^p, \tag{4}$$

where $\tau_1, \cdots, \tau_K$ are K selected knots on the support of the threshold variable, and $x_+^i = \{\max(x, 0)\}^i$.

We first compare these two different choices and leave the discussions on how to choosing K and τ_j's in the next subsection. Choice C1$'$ could be used to model $g(z)$ under condition C1. When using C1$'$, we suggest first applying a monotonic transform to the threshold variable so that $z_t \geq 0$ and $P(I_t = 1 \mid z_t = 0) \neq 0.5$ after transform. We then define the first regime such that $g(0) = P(I_t = 1|z_t = 0) > 0.5$, or equivalently, requiring $\alpha_0 > 0$ in equation (4). The choice of the degree p reflects one's assumption on the smoothness of $f(z)$, or equivalently, $g(z)$. If $g(z) \in W^m$, *i.e.*, the total variation of its m-th derivative $\int \{g^{(m)}(z)\}^2 dz < \infty$, we could choose $p = m - 1$. In particular, if g is a discontinuous curve with finite structural breaks, we could let $p = 0$.

In contrast, choice C2$'$ could be used to model $g(z)$ as a monotonically increasing function to satisfy condition C2. For convention, we can define the first regime such that $g(z) = P(I_t = 1|z_t = z)$ is monotonically increasing. Similar to C1$'$, the choice of the degree p reflects the smoothness of g. If $g(z) \in W^m$, $m \geq 2$, we could choose $p = m - 2$. Though higher degree p could be used, choosing $p = 0$ or $p = 1$ will lead to more efficient computation as the integration in equation (3) has a closed form. The lower bound a of the integration should be less than z_L, the lower limit of the support of z_t. If z_t has an unbounded support, one should choose $p = 1$ and restrict $\alpha_1 > 0$ to guarantee the existence of the integration.

In the next subsection, we shall discuss how to choose K and τ_j's via penalized method. Our main task is to estimate the parameters $\beta_1, \beta_2, \sigma_1, \sigma_2$ and the nonparametric regime probability function $g(z)$. We adopt a unified estimating procedure that could work for both choices of H under C1$'$ and C2$'$.

2.3 Priors and Estimation

Recall that we express $f(z)$ via the p-th degree TPPS as in equation (4), where $\alpha_0, \cdots, \alpha_p$, $b_1, \cdots, b_K$ are unknown parameters and τ_k's are K selected knots. When the number of knots K increases as the sample size does, the spline representation is flexible enough to be treated as a nonparametric approach. As pointed out by Li and Ruppert (2008), the choice of K and τ_k's might not be important as long as K exists some lower bound. However, to prevent from overfitting which tends to interpolate the data, one need to control the roughness of the estimate. A simple idea is to incorporate a penalty term related to total variation of the highest order derivative $f^{(p+1)}(z)$.

Let $f_\varepsilon(\cdot)$ be the density function of ε_t in model (2), the likelihood of $\{y_t, x_t, z_t, t = 1, \cdots, T\}$ could be written as

$$L = \prod_{t=1}^{T} \left\{ f_\varepsilon \left(\frac{y_t - x_t\beta_1}{\sigma_1} \right) g(z_t) + f_\varepsilon \left(\frac{y_t - x_t\beta_2}{\sigma_2} \right) [1 - g(z_t)] \right\}.$$

Suppose $g(z_t) = H(f(z_t))$ for some pre-specified H, and $f(z)$ is the p-th degree TPPS defined in (4). It could be shown that $\int [f^{(p+1)}(z)]^2 dz = \sum_{i=1}^{K} b_k^2$ (Claeskens et al., 2009). In the spirit of penalized likelihood, we may propose to conduct smoothing by minimizing the following criterion:

$$-2\sum_{t=1}^{T} \log \left\{ \left\{ f_\varepsilon \left(\frac{y_t - x_t\beta_1}{\sigma_1} \right) g(z_t) + f_\varepsilon \left(\frac{y_t - x_t\beta_2}{\sigma_2} \right) [1 - g(z_t)] \right\} \right\} + \lambda \sum_{i=1}^{K} b_k^2, \qquad (5)$$

subjected to $\alpha_j > 0$, where $j = 0$ or 1, in correspondence to assuming condition C1 or C2 for model identification, respectively. Obviously, the nonparametric estimate $\hat{f}(z)$ is highly affected by the choice of λ. In practice, it is always a major issue to discuss how to determine an appropriate amount of penalty. Though some data drive methods like cross-validation or generalized cross-validation could be employed, an alternative solution might be to treat b_k's as Gaussian random coefficients $N(0, \sigma_b^2)$ and use likelihood principle to choose the penalty parameter. Note that log of the joint distribution defined by our model (2) with Gaussian priors on b_k's has exactly the same form as (5) if we define $\lambda = 1/\sigma_b^2$.

In traditional nonparametric regression, Ruppert et al. (2003) has adopted the above idea to determine the penalty term and conduct smoothing. However, their approach might not be applied in our case, as the computation is more involved when the joint distribution does not belong to the exponential family. Given the difficulty in solving the optimization problem (5), we decide to turn into the Bayesian approach. The use of the Bayesian approach has several advantages. First, it provides an elegant solution to handle parameter constraints and nonparametric smoothing simultaneously. In particular, the constrains $\alpha_j > 0$ could be easily incorporated if we let the support of the prior $P(\alpha_j)$ to be nonnegative, while smoothing could be conducted as we include an extra layer to model the prior distribution of σ_b^2 and let the posterior to tell us what amount of penalty might be appropriate. Second, we also gain more advantages in handling hypothesis testing as shown in the next subsection. Before we elaborate this, let us describe our estimating procedure in details.

In the model (2), the joint posterior distribution of the parameters $\beta_1, \beta_2, \sigma_1, \sigma_2$, $\alpha_0, \cdots, \alpha_p$, $b_1, \cdots, b_K$, σ_b^2 and the latent status variables $I_1, \cdots, I_T$ can be written as

$$\begin{aligned} & P(\beta_1, \beta_2, \sigma_1^2, \sigma_2^2, \alpha_{1:p}, b_{1:K}, \sigma_b^2, I_{1:T} \mid y_{1:T}, x_{1:T}, z_{1:T}) \\ \propto \; & P(\beta_1)P(\beta_2)P(\sigma_1^2)P(\sigma_2^2)\prod_{i=1}^{p} P(\alpha_i)P(\sigma_b^2)\prod_{k=1}^{K} P(b_k \mid \sigma_b^2) \\ & \times \prod_{t=1}^{T} g(z_t)^{2-I_t}[1-g(z_t)]^{I_t-1} f_\varepsilon\left(\frac{y_t - x_t\beta_{I_t}}{\sigma_{I_t}}\right). \end{aligned} \tag{6}$$

Without strong prior information, we use flat conjugate priors when possible. Assuming $\varepsilon_t \sim N(0,1)$, we assign the normal and inverse gamma priors for the regression coefficients and the variance for the error terms, respectively, *i.e.*, $\beta_j \sim N(\mu_\beta, \Sigma_\beta)$, $\sigma_j^2 \sim IG(A_\sigma, B_\sigma)$. For non-restricted spline coefficients α_i's, we may assume that $\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)$. For α_i that needs to satisfy the constraint $\alpha_i > 0$, we may assume the gamma distribution for convention, *i.e.*, $\alpha_i \sim G(A_\alpha, B_\alpha)$. For the rest of the spline coefficients, we first consider the typical hierarchical representation such that $\sigma_b^2 \sim IG(A_b, B_b)$ and $b_k \mid \sigma_b^2 \sim N(0, \sigma_b^2)$.

We adopt the Metropolis-within-Gibbs algorithm to draw samples $\{\beta_1^{(s)}, \beta_2^{(s)}, (\sigma_1^2)^{(s)}, (\sigma_2^2)^{(s)}$, $\alpha_{1:p}^{(s)}, b_{1:K}^{(s)}, (\sigma_b^2)^{(s)}, I_{1:T}^{(s)}, s = 1, \cdots, S\}$ from the joint posterior distribution (6). The posterior mean of the model parameters can be estimated by the averages of the MCMC samples. To implement the sampling algorithm, we summarize the full conditional distributions of each component of the parameters and the latent statuses as follows.

$$\begin{aligned} & P(\beta_j \mid rest) \sim N\left(\overline{\mu}_{\beta,j}, \overline{\Sigma}_{\beta,j}\right), \\ & P(\sigma_j^2 \mid rest) \sim IG(A_\sigma + \sum_{t:I_t=j} 1/2, B_\sigma + \sum_{t:I_t=j} (y_t - x_t\beta_j)^2/2), \\ & P(\alpha_{0:p} \mid rest) \propto \prod_{i=0}^{p} P(\alpha_i) \prod_{t=1}^{T} \left\{c_{t1}\, g(z_t) + c_{t2}\,[1-g(z_t)]\right\}, \\ & P(b_{1:K} \mid rest) \propto \prod_{i=1}^{K} P(b_i) \prod_{t=1}^{T} \left\{c_{t1}\, g(z_t) + c_{t2}\,[1-g(z_t)]\right\}, \\ & P(\sigma_b^2 \mid rest) \sim IG(A_b + K/2, B_b + \sum_k b_k^2/2), \\ & P(I_t = j \mid rest) = [c_{t1}\, g(z_t)]^{2-j}\{c_{t2}\,[1-g(z_t)]\}^{j-1} / \left\{c_{t1}\, g(z_t) + c_{t2}\,[1-g(z_t)]\right\}, \end{aligned} \tag{7}$$

where $\overline{\mu}_{\beta,j} = \overline{\Sigma}_{\beta,j}\left(\sigma_j^{-2}\sum_{t:I_t=j} x_t y_t' + \Sigma_\beta^{-1}\mu_\beta\right)$, $\overline{\Sigma}_{\beta,j} = \left(\sigma_j^{-2}\sum_{t:I_t=j} x_t x_t' + \Sigma_\beta^{-1}\mu_\beta\right)^{-1}$ and $c_{tj} = f_\varepsilon\left(\frac{y_t - x_t\beta_1}{\sigma_1}\right)$ for $j = 1, 2$. When simulating $\alpha_{0:p}$ and $b_{1:K}$, we have integrated out the regime status $I_{1:T}$ to facilitate fast convergence rate. The conditional distributions of $\alpha_{0:p}$ and $b_{1:K}$ might be involved, but we can draw samples via the Metropolis-Hastings method. See Gilks et al. (1995) and Liu (2001) for a detailed description of MCMC sampling.

Though it may be natural to treat b_k's as Gaussian random variables, we would also like to discuss other possible choices of the prior. The Gaussian prior might result in L^2 penalty that plays the key role of smoothing. However, we think it may be kind of restrictive as $g(z)$ might not be smooth. For example, in the classical threshold model, $g(z) = \mathbf{1}(z \geq \gamma)$, with a discontinuous point at the threshold value γ. When the smoothness assumption is violated, we may instead consider sparseness assumption on the spline coefficients b_k's, as we only have occasional jumps or structural break. Therefore, we recommend to choose Choice C1$'$ with $p = 0$, and turns the problem into conducting variable selection on nonzero b_k's. If we follow the idea of Park and Casella (2008), we may consider the Laplacian prior for b_k, *i.e.*, $p(b_k|\sigma_b^2) \sim \frac{1}{\sigma_{b,k}^2}\exp\{-(|b_k|/\sigma_b^2)\}$. This is equivalent to the Bayesian LASSO procedure, which includes the L^1 penalty in the log-likelihood. However, we do think the assumption that all b_k's share the same homogenous parameter σ_b^2 might be restrictive. Think of the classical threshold model. The curves are extremely smooth except at the break point. Therefore, an adaptive setting with each b_k's associated with various $\sigma_{b,k}^2$ might be more reasonable. Hence we use the following hierarchical representation:

$$p(b_k \mid \sigma_{b,k}^2) \sim \frac{1}{\sigma_{b,k}^2}\exp\{-(|b_k|/\sigma_{b,k}^2)\},\ \sigma_{b,k}^2 \sim IG(A_b, B_b).$$

The posterior distribution are similar as in (7), except that the conditional distribution of $\sigma_{b,k}^2$ becomes

$$P(\sigma_{b,k}^2 \mid rest) \sim IG(A_b + 1, B_b + |b_k|).$$

Note that the joint likelihood with different Laplacian prior of b_k's is equivalent to conducting an adaptive lasso procedure, which could achieve sparseness as it tries to minimizes

$$-2\sum_{t=1}^{T}\log\left\{\left\{f_\varepsilon\left(\frac{y_t - x_t\beta_1}{\sigma_1}\right)g(z_t) + f_\varepsilon\left(\frac{y_t - x_t\beta_2}{\sigma_2}\right)[1 - g(z_t)]\right\}\right\} + \sum_{i=1}^{K} 2|b_k|/\sigma_{b,k}^2.$$

2.4 Bayesian Inference

An obvious advantage of Bayesian approach lies in the fact that the posterior provides a powerful inferential machinery. With the help of MCMC, the credible interval for any parameter could be constructed for inference. Specifically, the level $100(1-\alpha)\%$ highest posterior density (HPD) interval of a particular parameter θ can be constructed as the minimum length interval that includes $100(1-\alpha)\%$ of the MCMC samples $\{\theta^{(s)}, s=1,\cdots,S\}$. We can also conduct significance test on each component $\beta_{j,i}$ of the regression coefficients β_j, $j=1,2$. Chen et al. (2010) defined the posterior odds for testing $H_0: \beta_{j,i}=0$ as

$$r(\beta_{j,i}) = \frac{\sum_{s=1}^{S} \mathbf{1}(\beta_{j,i}^{(s)} > 0)}{\sum_{s=1}^{S} \mathbf{1}(\beta_{j,i}^{(s)} \le 0)}.$$

The null hypothesis can be rejected at the α-level if $r(\beta_{j,i}) > (1-\alpha)/\alpha$ or $r(\beta_{j,i}) < \alpha/(1-\alpha)$.

Among different hypothesis questions, one may be particularly interested in checking whether the threshold effect really exists. The hypothesis is then constructed as

$$H_0: \beta_1 = \beta_2 \text{ and } \log(\sigma_1^2) = \log(\sigma_2^2) \quad \text{versus} \quad H_1: \beta_1 \neq \beta_2 \text{ or } \log(\sigma_1^2) \neq \log(\sigma_2^2). \tag{8}$$

Even for classical threshold model, this is a nonstandard test as the threshold value γ is a nuisance parameter which is not identified under the null (Davies, 1977). The frequentists' approach to handle the so called Davies' problem relies on taking supremum over the unidentified parameter, thus leading to a complicated null distribution associated with the test statistic. In Bayesian analysis, this kind of tests are often conducted based on the posterior odds of the hypotheses (Berger and Yang, 1994), which is defined as

$$\frac{P(H_0 \mid data)}{P(H_A \mid data)} = \frac{P(H_0)}{P(H_1)} \times \frac{P(data \mid H_0)}{P(data \mid H_1)},$$

where $P(H_0)/P(H_1)$ is the prior odds of the hypotheses, and $P(data \mid H_0)$, $P(data \mid H_1)$ are the marginal likelihoods.

....

For our nonparametric regime shift model, we propose to use the $100(1-\alpha)\%$ highest posterior density (HPD) interval based on the MCMC sample of the regression parameters to draw the inference. In particular, the testing procedure is designed as below.

1. Define $\Delta^{(s)} = \begin{pmatrix} \beta_1^{(s)} - \beta_2^{(s)} \\ \log[\sigma_1^2]^{(s)} - \log[\sigma_2^2]^{(s)} \end{pmatrix}$, $s=1,\cdots,S$.

2. Calculate

$$D^{(s)} = (\Delta^{(s)} - \mu)'\Sigma^{-1}(\Delta^{(s)} - \mu),$$

where μ and Σ are sample mean and sample variance of $\Delta^{(s)}$, $s = 1, \cdots, S$, respectively. Define

$$D^0 = (0 - \mu)'\Sigma^{-1}(0 - \mu),$$

where 0 is the origin.

3. Reject H_0 if D^0 is greater than the $1 - \alpha$ quantile of $D^{(s)}$, $s = 1, \cdots, S$.

Remark: Our inferential procedure is valid as our model identification condition is imposed on $g(z_t)$ only. The marginal posterior of the regression coefficient β_j's and σ_j's could then be used for inference on the existence of the threshold variables.

3 Conclusions

Though threshold models have been proposed for more than three decades, the growth of its literature is still ongoing. The regime switching mechanism of the classical threshold model is completely controlled by the observable threshold variable relative to the threshold value. Such a deterministic setup could endorse some nice properties, but may be too restrictive in real applications. In this paper, we propose a new type of regime switching models with nonparametric regime probability functions, which could be treated as a combination of classical threshold models and random switching models. The use of a nonparametric link function enable us to avoid model miss-specification risk, thus allowing us to better discover the sample feature and handle various complicated situations.

We solve the identification problem by imposing constraints on the regime probability function and adopt the penalized spline approach for nonparametric estimation. To circumvent the difficulty caused by high dimensional optimization as well as parameter constraint, we employ the Bayesian approach with MCMC sampling techniques. We consider various choices of the priors on the spline coefficients, that enable us to conduct handle automatic smoothing or variable selection. We also design a Bayesian test on the threshold effects based on the MCMC samples. We demonstrate the validity of our methods via simulations and an empirical example.

We conclude our paper by pointing out some potential topics for future research. First, it is possible to extend our prior and augmenting the prior of the scale parameter σ_b with a point mass at zero. This will help us to formally design a nonparametric test on the link function with parametric null versus general alternative. Second, it may be a very interesting direction to extend our model to allow multiple regimes or include more threshold variables in the regime probability function.

Appendix

Proof of Proposition 1: In example 1, the regime probability function $g(z) = F_u(z-\gamma)$, it is always a monotonically increasing function.

In example 2, suppose $v_t \sim N(0, \sigma_v^2)$, we have

$$\begin{aligned} g(z) &= \frac{\int_\gamma^\infty f_{z^*}(z^*) f_v(z-z^*) dz_t^*}{\int_{-\infty}^\infty f_{z^*}(z^*) f_v(z-z^*) dz^*} \\ &= \frac{\int_\gamma^\infty f_{z^*}(z^*)\, exp\{-\frac{(z-z^*)^2}{2\sigma_v^2}\} dz^*}{\int_{-\infty}^\gamma f_{z^*}(z^*)\, exp\{-\frac{(z-z^*)^2}{2\sigma_v^2}\} dz^* + \int_\gamma^\infty f_{z^*}(z^*)\, exp\{-\frac{(z-z^*)^2}{2\sigma_v^2}\} dz^*} \\ &= \frac{1}{\frac{\int_{-\infty}^\gamma f_{z^*}(z^*)\, exp\{-\frac{(z-z^*)^2}{2\sigma_v^2}\} dz^*}{\int_\gamma^\infty f_{z^*}(z^*)\, exp\{-\frac{(z-z^*)^2}{2\sigma_v^2}\} dz^*} + 1}. \end{aligned}$$

The first term in the denominator can be written as

$$\begin{aligned} \frac{\int_{-\infty}^\gamma f_{z^*}(z^*)\, exp\{-\frac{(z-z^*)^2}{2\sigma_v^2}\} dz^*}{\int_\gamma^\infty f_{z^*}(z^*)\, exp\{-\frac{(z-z^*)^2}{2\sigma_v^2}\} dz^*} &= \frac{\int_{-\infty}^\gamma f_{z^*}(z^*)\, exp\{-\frac{(z-\gamma+\gamma-z^*)^2}{2\sigma_v^2}\} dz^*}{\int_\gamma^\infty f_{z^*}(z^*)\, exp\{-\frac{(z-\gamma+\gamma-z^*)^2}{2\sigma_v^2}\} dz^*} \\ &= \frac{\int_{-\infty}^\gamma f_{z^*}(z^*)\, exp\{\frac{(2z-\gamma-z^*)(z^*-\gamma)}{2\sigma_v^2}\} dz^*}{\int_\gamma^\infty f_{z^*}(z^*)\, exp\{\frac{(2z-\gamma-z^*)(z^*-\gamma)}{2\sigma_v^2}\} dz^*}. \end{aligned}$$

Because $\int_{-\infty}^\gamma f_{z^*}(z^*)\, exp\{\frac{(2z-\gamma-z^*)(z^*-\gamma)}{2\sigma_v^2}\} dz^*$ is monotonically decreasing as z increases and $\int_\gamma^\infty f_{z^*}(z^*)\, exp\{\frac{(2z-\gamma-z^*)(z^*-\gamma)}{2\sigma_v^2}\} dz^*$ is monotonically increasing, so $g(z)$ is a monotonically increasing function. This completes the proof. $\square$

References

Ait-Sahalia, Y. and Duarte, J. (2003), "Nonparametric option pricing under shape restrictions," *Journal of Econometrics*, 116, 9–47.

Amemiya, Y. (1985), "Instrument variable estimator for the nonlinear errors in variables model," *Journal of Econometrics*, 28, 273–289.

— (1990), "Two stage instrumental variable estimators for the nonlinear errors in variables model," *Journal of Econometrics*, 44, 311–332.

Barnett, S. and Sakellaris, P. (1998), "Nonlinear response of firm investment to Q: Testing a model of convex and non-convex adjustment costs," *Journal of Monetary Economics*, 42, 261–288.

Bell, D. and Lattin, J. (2000), "Looking for loss aversion in scanner data: The confounding effect of price response heterogeneity," *Marketing Science*, 19, 185–200.

Berger, J. and Yang, R. (1994), "Noninformative priors and Bayesian testing for the AR(1) model," *Econometric Theory*, 10, 461–482.

Cai, N., Cai, Z., Fang, Y., and Xu, Q. (2015), "Forecasting major Asian exchange rates using a new semiparametric STAR model," *Empirical Economics*, 48, 407–426.

Chan, K. (1993), "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model," *The Annals of Statistics*, 21, 520–533.

Chan, K. and Tong, H. (1986), "On estimating thresholds in autoregressive models," *Journal of Time Series Analysis*, 7, 179–190.

Chang, K., Siddarth, S., and Weinberg, C. (1999), "The impact of heterogeneity in purchase timing and price responsiveness on estimates of sticker shock effects," *Marketing Science*, 18, 178–192.

Chen, H., Chong, T., and Bai, J. (2012), "Theory and applications of TAR model with two threshold variables," *Econometric Reviews*, 31, 142–170.

Chen, R., Guo, R., and Lin, M. (2010), "Self-selectivity in firm's decision to withdraw IPO: Bayesian inference for hazard models of bankruptcy with feedback," *Journal of the American Statistical Association*, 492, 1297–1309.

Claeskens, G., T Krivobokova, T., and Opsomer, J. D. (2009), "Asymptotic properties of penalized spline estimators," *Biometrika*, 96, 529–544.

Davies, R. (1977), "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 64, 247–254.

Durlauf, S. and Johnson, P. (1995), "Multiple regimes and cross-country growth behaviour," *Journal of Applied Econometrics*, 10, 365–384.

Erickson, T. and Whited, T. (2000), "Measurement error and the relationship between investment and q," *Journal of Political Economy*, 108, 1027–1057.

Fazzari, S., Hubbard, R., and Petersen, B. (1988), "Financing constraints and corporate investment," *Brookings Papers On Economic Activity*, 1, 141–195.

Gilks, W., Richardson, S., and Spiegelhalter, D. (1995), *Markov chain Monte Carlo in Practice*, Chapman & Hall.

Gonzalo, J. and Pitarakis, J. (2012), "Regime-specific predictability in predictive regressions," *Journal of Business & Economic Statistics*, 30(2), 229–241.

Hamilton, J. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57, 357–384.

Hansen, B. (1997), "Inference in TAR models," *Studies in Nonlinear Dynamics and Econometrics*, 2(1), 1–14.

— (1999), "Testing for linearity," *Journal of Economic Surveys*, 5, 551–576.

— (2000), "Sample splitting and threshold estimation," *Econometrica*, 68(3), 575–603.

— (2011), "Threshold autoregression in economics," *Statistics and Its Interface*, 4, 123–127.

Helson, H. (1964), *Adaptation-level theory*, New York: Harper.

Henkel, S., Martin, J., and Nadari, F. (2011), "Time-varying short-horizon predictability," *Journal of Financial Economics*, 99, 560–580.

Hu, X. and Schiantarelli, F. (1998), "Investment and capital market imperfections: A switching regression approach using U.S. firm panel data," *Review of Econometrics and Statistics*, 80, 466–479.

Kahneman, D. and Tversky, A. (1979), "Prospect theory: An analysis of decision under risk," *Econometrica*, 47, 263–292.

Kim, C. and Nelson, C. R. (1999), *State space models with regime switching : Classical and Gibbs-sampling approaches with applications*, MIT Press.

Lewbel, A. (1997), "Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D," *Econometrica*, 65(5), 1201–1213.

Li, C. and Li, W. (1996), "On a double-threshold autoregressive heteroscedastic time series model," *Journal of Applied Econometrics*, 11, 253–274.

Li, Y. and Ruppert, D. (2008), "On the asymptotics of penalized splines," *Biometrika*, 95, 415–436.

Liu, J. S. (2001), *Monte Carlo Strategies in Scientific Computing*, Springer, New York.

Madansky, A. (1959), "The fitting of straight lines with both variables are subject to error," *Journal of the American Statistical Association*, 54, 173–205.

Park, T. and Casella, G. (2008), "The Bayesian Lasso," *Journal of the American Statistical Association*, 103, 681–686.

Potter, S. (1995), "A nonlinear approach to U.S. GNP," *Journal of Applied Econometrics*, 2, 109–125.

Quandt, R. (1972), "A new approach to estimating switching regressions," *Journal of the American Statistical Association*, 67, 306–310.

Ramsay, J. (1998), "Estimating smooth monotone functions," *Journal of the Royal Statistical Society, Series B*, 60, 365–375.

Ruppert, D., Wand, M., and Carroll, R. (2003), *Semiparametric regression*, New York: Cambridge University Press.

Seo, M. and Linton, O. (2007), "A smoothed least squares estimator for threshold regression models," *Journal of Econometrics*, 141, 704–735.

Terui, N. and Dahana, W. (2006), "Estimating heterogeneous price thresholds," *Marketing Science*, 25(4), 384–391.

Tong, H. (1983), *Threshold models in nonlinear time series analysis*, New York: Springer Lecture Notes in Statistics.

— (2011), "Threshold models in time series analysis – 30 years on," *Statistics and Its Interface*, 4, 107–118.

Tong, H. and Lim, K. (1980), "Threshold autoregression, limit cycles and cyclic data," *Journal of the Royal Statistical Society*, Series B, 42, 245–292.

Tsay, R. (1998), "Testing and modeling multivariate threshold models," *Journal of the American Statistical Association*, 93, 1188–1202.

Wu, S. and Chen, R. (2007), "Threshold variable determination and threshold variable driven switching autoregressive models," *Statistica Sinica*, 17, 241–264.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

001 "Data Driven Value-at-Risk Forecasting using a SVR-GARCH-KDE Hybrid" by Marius Lux, Wolfgang Karl Härdle and Stefan Lessmann, January 2018.
002 "Nonparametric Variable Selection and Its Application to Additive Models" by Zheng-Hui Feng, Lu Lin, Ruo-Qing Zhu asnd Li-Xing Zhu, January 2018.
003 "Systemic Risk in Global Volatility Spillover Networks: Evidence from Option-implied Volatility Indices " by Zihui Yang and Yinggang Zhou, January 2018.
004 "Pricing Cryptocurrency options: the case of CRIX and Bitcoin" by Cathy YH Chen, Wolfgang Karl Härdle, Ai Jun Hou and Weining Wang, January 2018.
005 "Testing for bubbles in cryptocurrencies with time-varying volatility" by Christian M. Hafner, January 2018.
006 "A Note on Cryptocurrencies and Currency Competition" by Anna Almosova, January 2018.
007 "Knowing me, knowing you: inventor mobility and the formation of technology-oriented alliances" by Stefan Wagner and Martin C. Goossen, February 2018.
008 "A Monetary Model of Blockchain" by Anna Almosova, February 2018.
009 "Deregulated day-ahead electricity markets in Southeast Europe: Price forecasting and comparative structural analysis" by Antanina Hryshchuk, Stefan Lessmann, February 2018.
010 "How Sensitive are Tail-related Risk Measures in a Contamination Neighbourhood?" by Wolfgang Karl Härdle, Chengxiu Ling, February 2018.
011 "How to Measure a Performance of a Collaborative Research Centre" by Alona Zharova, Janine Tellinger-Rice, Wolfgang Karl Härdle, February 2018.
012 "Targeting customers for profit: An ensemble learning framework to support marketing decision making" by Stefan Lessmann, Kristof Coussement, Koen W. De Bock, Johannes Haupt, February 2018.
013 "Improving Crime Count Forecasts Using Twitter and Taxi Data" by Lara Vomfell, Wolfgang Karl Härdle, Stefan Lessmann, February 2018.
014 "Price Discovery on Bitcoin Markets" by Paolo Pagnottoni, Dirk G. Baur, Thomas Dimpfl, March 2018.
015 "Bitcoin is not the New Gold - A Comparison of Volatility, Correlation, and Portfolio Performance" by Tony Klein, Hien Pham Thu, Thomas Walther, March 2018.
016 "Time-varying Limit Order Book Networks" by Wolfgang Karl Härdle, Shi Chen, Chong Liang, Melanie Schienle, April 2018.
017 "Regularization Approach for NetworkModeling of German EnergyMarket" by Shi Chen, Wolfgang Karl Härdle, Brenda López Cabrera, May 2018.
018 "Adaptive Nonparametric Clustering" by Kirill Efimov, Larisa Adamyan, Vladimir Spokoiny, May 2018.
019 "Lasso, knockoff and Gaussian covariates: a comparison" by Laurie Davies, May 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

020 "A Regime Shift Model with Nonparametric Switching Mechanism" by Haiqiang Chen, Yingxing Li, Ming Lin and Yanli Zhu, May 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.