CHANGE POINT ANALYSIS AND REGIME SWITCHING MODELS

PAUL WESKAMP AND MARKUS HÖCHSTÖTTER

Abstract. Change points and changing regimes can be observed in many statistical time series. In this paper we analyze the most influential approaches of modeling these occurrences. Along with the theory some of the most interesting applications are presented. We find out that use of terminology is not unique, leading to considerable differences between the models, as will be pointed out.

1. Introduction

In time series there are many examples where we can observe break, turning, or change points from where on the series seems to follow a different regime than before. Considering e.g. stock markets, we will most likely find the time series of returns to have a different structure after a crash than before. So does a crash involve a decrease in mean for one, but it will probably also have a deep impact on the variance.

This observation is highly interesting for the statistical modeler. It is trivial that a model, which takes the change point into account, will capture the subject more precisely than one which ignores it. Consequently, there have been many attempts to implement one or more change points and the corresponding different regimes into one model.

Especially in the last two decades statisticians have been concerned with this issue. There have been different terms to denote the respective events, like "break point" or "turning point". Nevertheless, the most common terms were "change point" for the point in time where the change takes place and "regime switch" for the occurrence of a different regime after the change point. These are the terms referred to in the sequel.

Within the scope of this paper we have identified the approaches, which were to our knowledge most influential and in our opinion most promising. In chapter 2 the corresponding models are introduced and chapter 3 shows a few of the already existing applications, which we know of. Chapter 4 summarizes, including parallels and differences between the different approaches.

Date: December 15, 2010.

Key words and phrases. Change Point, Regime Switching.

To display the ideas of the researchers as accurate as possible, their actual notation has been adopted in the model presentation. Thus, the variables used are introduced in the respective sections.

2. Analyses and Models

2.1. **A Diagnostic Approach to the Timing of Structural Breaks.** A very simple and intuitive approach to the finding of change points is illustrated by Isaac and Griffin [8] . The authors reject "ahistorism" in time series analysis in favor of a historically sensitive approach. They state that many quantitative analyses fail to realize their potential because sudden or gradual temporal changes of (historical-) structural relationships are generally ignored in quantitative explorations of (historical) processes.

Isaac and Griffin analyze dozens of analyses of labor organization and conflict to find out that there is not one that systematically assesses the temporal stability of time-series parameter estimates, regards historical periodization, and offers a theorized rationale for choosing the starting and ending dates in the analysis at the same time.

To demonstrate that these three features of analyses of historical series are crucial, Isaac and Griffin present the "backward" strategy, which illustrates the likelihood of sudden and dramatic turning points in historical relationships. They apply the strategy to the correlations and covariances between union growth and change in strike frequency, union disbandings and change in strike frequency (both for 1882–1980), and union foundings and change in strike frequency (for 1882–1976).

To start the procedure, Isaac and Griffin estimate the correlation and covariance over the years 1882–1980 (1882–1976 for foundings). Then they sequentially drop the first year of the series, so that the second association refers to 1883–1980, the third to 1884–1980, and so on until the last relationship is estimated over the years 1962–1980. So all relationships presented are for a period up to and including 1980 and become exclusively more recent, as the most distant year is deleted. The results of this analysis are shown as illustration in Figure 1.

This procedure they use assesses the temporal stability of time series estimates and establishes change points (Isaac and Griffin refer to "shift" or "break" or "turning points") where the relationships studied alters suddenly with the removal of a single year.

2.2. **Intervention Analysis Methods.** If the timing of a change point is known then Box and Jenkins provide a methodology to determine the magnitude of the change. If T is the location of the change point on the time axis (referred to as "intervention event"), their transfer function model [2] allows a mean shift in a response variable. Let Y_t be the time series that is subject to regime changes. The model

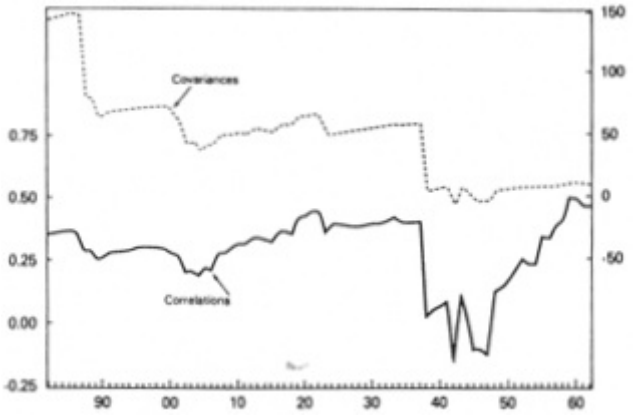


FIGURE 1. Correlations (Left-Hand) Axis and Covariances (Right-Hand) Axis between Union Growth and Change in Strike Frequency for Periods from 1882–1980 to 1962–1980 [8]
Note to Figure 1: Entries on the X-axis refer to the year beginning the historical period encapsulating the relationship graphed. Thus, "90" refers to the period 1890–1980, "50" to the period 1950–1980, etc.

considered then has the form

$$Y_t = \frac{\omega(B) B^b}{\delta(B)} \xi_t + N_t,$$

where the term $I_t = \delta^{-1}(B)\,\omega(B)\,B^b \xi_t$ represents the effects of the intervention event in terms of the deterministic input series ξ_t with lag operator $B = 1 - \nabla[3]$. N_t is the noise superimposed on the series Y_t. It is assumed that N_t follows an ARIMA(p, d, q) model $\phi(B) N_t = \theta(B) a_t$, with $\phi(B) = \varphi(B)(1-B)^d$. Box and Jenkins found two common types of deterministic input variables ξ_t to be useful to represent the impact of intervention events on a time series. Both of these are indicator variables taking only the values 0 and 1 to denote the nonoccurrence and occurrence of intervention. One type is a step function at time T, given by

$$S_t^{(T)} = \begin{cases} 0 & t < T \\ 1 & t \geq T \end{cases},$$

which would typically be used to represent the effects of an intervention that are expected to remain permanently after time T to some extent. The other type is a pulse function at T, given by

$$P_t^{(T)} = \begin{cases} 0 & t \neq T \\ 1 & t = T \end{cases},$$

which could represent the effects of an intervention that are temporary or transient and will die out after time T.

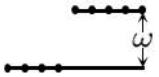
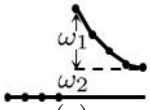
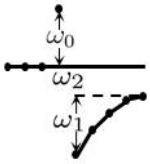
Input			
$S_t^{(T)}$	 Step	Pulse	$P_t^{(T)}$
Output			
$\frac{\omega(B)}{\delta(B)} S_t^{(T)}$			$\frac{\omega(B)}{\delta(B)} P_t^{(T)}$
$\omega B S_t^{(T)}$	 (a)	 (d)	$\frac{\omega_1 B}{1-\delta B} P_t^{(T)}$
$\frac{\omega B}{1-\delta B} S_t^{(T)}$	 (b)	 (e)	$\left\{ \frac{\omega_1 B}{1-\delta B} + \frac{\omega_2 B}{1-B} \right\} P_t^{(T)}$
$\frac{\omega B}{1-B} S_t^{(T)}$	 (c)	 (f)	$\left\{ \omega_0 + \frac{\omega_1 B}{1-\delta B} + \frac{\omega_2 B}{1-B} \right\} P_t^{(T)}$

TABLE 1. Responses to a step and a pulse input: (a), (b), (c) response to a step input for various simple transfer function models; (d), (e), (f) response to a pulse input for some models of interest.

Several different response patterns $I_t = \delta^{-1}(B)\,\omega(B)\,B^b \xi_t$ are possible through different choices of the transfer function. Table 1 shows the responses for various simple transfer functions with both step and pulse indicators as input.

In general, the parameter estimates and their standard errors for the intervention model

$$Y_t = \frac{\omega(B)\,B^b}{\delta(B)} \xi_t + \frac{\theta(B)}{\varphi(B)} a_t,$$

are obtained by the least squares method of estimation for transfer function-noise models [4].

2.3. **Testing for Structural Change in Conditional Models.** Bruce E. Hansen proposes a general test for structural change in linear regression models, which also allows for structural change in regressors [6]. His test does not presuppose knowledge of the timing of the change point.

In the standard linear regression model

$$y_{ni} = x'_{ni}\beta_{ni} + e_{ni},$$

where y_{ni} is real valued and x_{ni} is an m-vector, structural change in the conditional distribution arises through the coefficient β_{ni}. The structural change in β_{ni} takes the form

$$\beta_{ni} = \begin{cases} \beta & i < t_0 \\ \beta + \theta_n & i \geq t_0 \end{cases},$$

where θ_n takes the form

$$\theta_n = \frac{\delta\sigma}{\sqrt{n}}$$

with δ fixed. The error e_{ni} is a martingale difference: $\mathbb{E}\left(e_{ni}|\,\mathfrak{I}_{ni-1}\right) = 0$, where $\mathfrak{I}_{ni-1}$ is the sigma-field generated by current x_{ni} and lagged values of (x_{ni}, e_{ni}).

The parameter $t_0 \in [t_1, t_2]$ indicates the relative timing of the change point and θ_n indicates the magnitude of the shift. If there is a change point $H_0 : \theta_n = 0$ will be rejected against $H_1 : \theta_n \neq 0$ and the parameter δ indicates the degree of structural change.

Under H_0 the model reduces to $y_{ni} = x'_{ni}\beta_n + e_{ni}$, which does not depend on t_0. The ordinary least squares estimators (OLS) is denoted $\widehat{\beta}$, the residuals $\widehat{e}_i$, and the variance estimate $\widehat{\sigma}^2 = (n-m)^{-1}\sum_{i=1}^{n}\widehat{e}_i^2$.

Under the alternative $H_1 : \theta_n \neq 0$, the model can be written as

$$y_{ni} = x'_{ni}\beta_{ni} + x'_{ni}\theta_n I\left(i \geq t_0\right) + e_{ni},$$

where $I\left(\cdot\right)$ is the indicator function.

For any fixed t, the regression function above can be estimated by OLS, yielding estimates $(\widehat{\beta}_t, \widehat{\theta}_t)$, residuals $\widehat{e}_{it}$ and variance estimates $\widehat{\sigma}_t^2 = (n-2m)^{-1}\sum_{i=1}^{n}\widehat{e}_{it}^2$. The least squares estimate of the change point is given by $\widehat{t} = \operatorname{argmin}\,\widehat{\sigma}_t^2$. Set $\widetilde{\beta} = \widehat{\beta}_t$ and $\widetilde{e}_i = \widehat{e}_{i\widehat{t}}$.

The standard test for H_0 against H_1 for known t is the Wald statistic:

$$F_t = \frac{\left(n-m\right)\widehat{\sigma}^2 - \left(n-2m\right)\widehat{\sigma}_t^2}{\widehat{\sigma}_t^2},$$

which is equivalent to the likelihood ratio statistic when e_{ni} is iid $\mathrm{N}(0, \sigma^2)$. The likelihood ratio test for the true but unknown change point t_0 is equivalent to $\sup F_n = \sup{}_t F_t$, where the supremum is taken over $t \in (t_1, t_2)$.

Hansen shows that this theory holds for all asymptotically stationary processes but not for asymptotically non-stationary processes. In this

case, the asymptotic distributions of the test statistics are different than those originally tabulated.

Hansen constructs two forms of what he calls "Fixed Regressor Bootstraps", which treats the regressors x_{ni} as if they are fixed (exogenous) even when they contain lagged dependent variables. One is appropriate if the error e_{ni} is homoskedastic and the other is appropriate under heteroskedasticity.

For the homoskedastic bootstrap, let $y_{ni}(b) : i = 1, \ldots, n$ be a random sample from the $\mathrm{N}(0,1)$ distribution. Regress $y_{ni}(b)$ on x_{ni} to get residual variance $\widehat{\sigma}^2(b)$ and regress $y_{ni}(b)$ on x_{ni} and $x_{ni}I(i \leq t)$ to get residual variance $\widehat{\sigma}_t^2(b)$ and Wald sequence

$$F_t(b) = \frac{(n-m)\,\widehat{\sigma}^2(b) - (n-2m)\,\widehat{\sigma}_t^2(b)}{\widehat{\sigma}_t^2(b)}.$$

The bootstrap test statistic is $\sup F_n(b) = \sup_{t_1 \leq t \leq t_2} F(b)_t$. The bootstrap p-value is $p_n = 1 - G_n(\sup F_n)$, where $G_n(x) = P(\sup F_n(b) \leq x \,|\mathfrak{I}_n)$ is the conditional distribution function of $\sup F_n(b)$. The bootstrap test rejects H_0 when p_n is small.

For heteroskedastic errors one has to substitute $y_{ni}^h(b) = u_i(b)\widetilde{e}_i$, for $y_{ni}(b)$ in the test statistic, where $\{u_i(b) : i = 1, \ldots, n\}$ is an iid $\mathrm{N}(0,1)$ sample and $\widetilde{e}_i$ are the regression residuals defined above. The new test statistic $\sup F_n^n(b)$ has the bootstrap distribution $G_n^h(x) = P(\sup F_n^h(b) \leq x \,|\mathfrak{I}_n)$ and the bootstrap p-value $p_n^h = 1 - G_n^h(\sup F_n)$.

$G_n(\cdot)$ is unknown but may be calculated by simulation. The simulated bootstrap p-value $p_n(j)$ is calculated by counting the percentage of simulated bootstrap test statistics $\sup F_n(j), j = 1, \ldots, J$ ((conditionally) independent draws from the distribution $\sup F_n(b)$), which exceed the sample value $\sup F_n$. As $J \to \infty$, $p_n(j) \to p_n$ almost surely, so the error due to simulation can be made arbitrarily small. For the case of heteroskedastic errors, a simulation estimate may be constructed similarly.

2.4. **Change Point Analysis for Dependence Structures.** A more elaborate model of the detection of changes in the dependence structure is shown by Dias and Embrechts [5]. They work out a methodology to test for a change point in a copula and estimation of the time of change: Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random vectors in $\mathbb{R}^m$ with distribution functions $F(x; \theta_1, \eta_1), \ldots, F(x; \theta_n, \eta_n)$ respectively, where θ_i and η_i are parameters of the distribution functions such that $\theta_i \in \Theta^{(1)} \subseteq \mathbb{R}^d$ and $\eta_i \in \Theta^{(2)} \subseteq \mathbb{R}^p$ for $1 \leq i \leq n$.

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_n \text{ and } \eta_1 = \eta_2 = \ldots = \eta_n$$

$$H_A : \theta_1 = \ldots = \theta_k \neq \theta_{k+1} = \ldots = \theta_n \text{ and } \eta_1 = \eta_2 = \ldots = \eta_n$$

Here k^* is the location or time of the (assumed) single change-point. All the parameters $(\theta, \eta) \in \Theta^{(1)} \times \Theta^{(2)}$ are supposed to be unknown

under both hypotheses. If $k^* = k$ is known, the null hypothesis will be rejected for small values of the test statistic

$$\Delta_k = \frac{\sup_{(\theta,\eta)\in\Theta^{(1)}\times\Theta^{(2)}} \prod_{1\leq i\leq n} f\left(X_i;\theta,\eta\right)}{\sup_{(\theta,\theta',\eta)\in\Theta^{(1)}\times\Theta^{(1)}\times\Theta^{(2)}} \prod_{1\leq i\leq k} f\left(X_i;\theta,\eta\right)\prod_{k<i\leq n} f\left(X_i;\theta',\eta\right)}.$$

If

$$L_k(\theta,\eta) = \sum_{1\leq i\leq k} \log f(X_i;\theta,\eta)$$

and

$$L_k^*(\theta,\eta) = \sum_{k<i\leq n} \log f(X_i;\theta,\eta)$$

then the likelihood ratio equation can be written as

$$-2\log(\Delta k) = 2\left[L_k\left(\widehat{\theta}_k,\widehat{\eta}_k\right) + L_k^*\left(\theta_k^*,\widehat{\eta}_k\right) - L_k\left(\widehat{\theta}_n,\widehat{\eta}_n\right)\right].$$

When k is unknown, H_0 will be rejected for large values of

$$Z_n = \max_{1\leq k\leq n}\left(-2\log\left(\Delta_k\right)\right).$$

If we assume there is exactly one change-point, then the maximum likelihood estimator for the time of the change is given by

$$\widehat{k}_n = min\left\{1\leq k\leq n : Z_n = -2\log(\Delta_k)\right\}$$

In the case that there is no change, $\widehat{k}_n$ will take a value near the limits of the sample.

In order to construct a confidence interval for the time of the change, we need to know or to approximate the distribution of $\widehat{k}_n - k_0$, where k_0 is the true time of change. One approximation can be obtained using bootstrap methodology. One of the simplest approaches to construct a bootstrap confidence interval is the percentile method. Suppose that $x_1, x_2, \ldots, x_n$ is the sample from which we estimated the time of the change $\widehat{k}_n$, assuming that the data came from a population with distribution function in the copula family C_θ. Then we have the maximum likelihood estimates $\widehat{\theta}_b$ and $\widehat{\theta}_a$ of the parameter of the distribution before and after the time of the change, respectively. From this we replicate the original sample simulating N samples of size n from the fitted distribution

$$F_{\widehat{\theta}_b,\widehat{\theta}_a,\widehat{k}_n}(x_i) = C_{\widehat{\theta}_b}(x_i)1_{\left\{i\leq\widehat{k}_n\right\}} + C_{\widehat{\theta}_a}(x_i)1_{\left\{i>\widehat{k}_n\right\}}$$

and compute the estimated time of the change $\widehat{k}_{n,i}^*$ for each replicate sample, $i = 1,\ldots,N$. These replicates allow us to estimate the distribution function of $\widehat{k}_n^* - \widehat{k}_n$, where $\widehat{k}_n^*$ is the time of change estimate of a resample from a population with distribution function $F_{\widehat{\theta}_b,\widehat{\theta}_a,\widehat{k}_n}$. The bootstrap principle, for confidence intervals, consists on assuming that

we can approximate the quantiles of $\widehat{k}_n - k_0$ by the quantiles of $\widehat{k}_n^* - \widehat{k}_n$. In order to obtain the quantiles k_α^*, where

$$P\left(k_\alpha^* \leq k_0 \leq k_{1-\alpha}^*\right) \approx 1 - 2\alpha$$

from the bootstrap replicates $\widehat{k}_{n,i}^*(i = 1, \ldots, N)$ we note that, if $\widehat{k}_{n,1}^*, \ldots, \widehat{k}_{n,N}^*$ are independent and identically distributed with distribution function H, then

$$\mathbb{E}\left[\widehat{k}_{n,(j)}^*\right] = H^{-1}\frac{j}{N+1},$$

where $\widehat{k}_{n,(j)}^*$ denotes the jth ordered value. Using this result an estimate for $k_\alpha^* = H^{-1}(\alpha)$ is $\widehat{k}_{n,((N+1)\alpha)}^*$. Finally substituting these values in

$$P(k_\alpha^* \leq k_0 \leq k_{1-\alpha}^*) \approx 1 - 2\alpha$$

we obtain that the bootstrap confidence interval for the time of the change k_0 :

$$\left(\widehat{k}_{n,((N+1)\alpha)}^*, \widehat{k}_{n,((N+1)(1-\alpha))}^*\right).$$

2.5. **Estimating and Testing Linear Models with Multiple Structural Changes.** Bai and Perron develop a model, which allows multiple structural changes, i.e. change points, and propose test statistics for such a model [1].

Underlying is the following multiple linear regression with m breaks ($m + 1$ regimes):

$$y_t = x_t'\beta + z_t'\delta_j + u_t$$

($t = T_{j-1} + 1, \ldots, T_j$), for $j = 1, \ldots, m + 1$ and $T_0 = 0$ and $T_{m+1} = T$. In this model y_t is the observed independent variable, $x_t(p \times 1)$ and $z_t(q \times 1)$ are vectors of covariates, and β and $\delta_j(j = 1, \ldots, m + 1)$ are the corresponding vectors of coefficients; u_t is the disturbance. The break points $(T_1, \ldots, T_m)$ are explicitly unknown. The purpose is to estimate the unknown regression coefficients together with the break points when T observations on (y_t, x_t, z_t) are available. In this model β is not subject to shifts and is estimated using the entire sample. When $p = 0$ the model becomes a pure structural change model, where all the coefficients are subject to change.

The true value of a parameter is denoted as a 0 superscript. In particular $\delta^0 = (\delta_1^{0\prime}, \ldots, \delta_{m+1}^{0\prime})'$ and $(T_1^0, \ldots, T_m^0)$ are, respectively, the true values of the parameters δ and of the break points. Bai and Perron first estimate the unknown coefficients $(\beta^0, \delta_1^0, \ldots, \delta_{m+1}^0, T_1^0, \ldots, T_m^0)$, assuming $\delta_i^0 \neq \delta_{i+1}^0(1 \leq i \leq m)$. The estimation of the coefficients is not restricted such that the regression function is continuous at the turning points. Bai and Perron focus on discrete shifts.

The actual estimation is based on the least-squares principle. For each m-partition $(T_1, \ldots, T_m)$, denoted T_j, the associated least-squares estimates of β and δ_j are obtained by minimizing the sum of squared

residuals $\sum_{i=1}^{m+1}\sum_{t=T_{i-1}+1}^{T_i}[y_t - x_t'\beta - z_t'\delta_i]^2$. The resulting estimates are denoted $\widehat{\beta}(\{T_j\})$ and $\widehat{\delta}(\{T_j\})$ and substituted into the objective function. The estimated break points $(\widehat{T}_1, \ldots, \widehat{T}_m)$, denoting the sum of squared residuals as $S_t(T_1, \ldots, T_m)$, are such that

$$(\widehat{T}_1, \ldots, \widehat{T}_m) = \text{argmin}_{T_1,\ldots,T_m} S_t(T_1, \ldots, T_m),$$

where the minimization is taken over all partitions $(T_1, \ldots, T_m)$ such that $T_i - T_{i-1} \geq q$. Thus, the break point estimators are global minimizers of the objective function. The regression parameter estimates are the associated least-squares estimates at the estimated m-partition $\{\widehat{T}_j\}$, i.e. $\widehat{\beta} = \widehat{\beta}(\{\widehat{T}_j\})$ and $\widehat{\delta} = \widehat{\delta}(\{\widehat{T}_j\})$. Bai and Perron proof the consistency and convergence to the true value of the estimated break fractions under a set of technical assumptions for their model.

Bai and Perron propose three kinds of test statistics for their model. One tests for no break ($m = 0$) against some fixed number of breaks ($m = k$). Let $(T_1, \ldots, T_k)$ be a partition and define

$$F_T(T_1, \ldots, T_k; q) = \left(\frac{T - (k+1)q - p}{kq}\right) \frac{\widehat{\delta}' R'(R(\overline{Z}' M_X \overline{Z})^{-1} R')^{-1} R\widehat{\delta}}{SSR_k},$$

where $\overline{Z}$ is the matrix which diagonally partitions Z at the k-partition, i.e. $\overline{Z} = \text{diag}(Z_1, \ldots, Z_{k+1})$ with $Z_i = (z_{T_{i-1}+1}, \ldots, z_{T_i})'$. R is the conventional matrix such that $(R\delta)' = (\delta_1' - \delta_2', \ldots, \delta_k' - \delta_{k+1}')$ and $M_X = I - X(X'X)^{-1}X'$. SSR_k is the sum of squared residuals under the alternative hypothesis, which depends on $(T_1, \ldots, T_k)$. For the asymptotic analysis Bai and Perron impose some restrictions on the test statistic, in particular they restrict each break date to be asymptotically distinct and bounded from the boundaries of the sample. The asymptotic critical values are obtained via simulations.

The second test is called double maximum test and is designed to test for no structural break versus an unknown number of breaks given some upper bound M:

$$\text{Dmax}F_T(M, q, a_1, \ldots, a_M) = \max_{1 \leq m \leq M} a_m \sup_{(T_1,\ldots,T_m)} F_T(T_1, \ldots, T_m; q)$$

defined for some fixed weights $a_1, \ldots, a_M$. The weights may reflect the imposition of some priors on the likelihood of various numbers of breaks. An obvious candidate is to set all weights equal to unity. Since for any fixed q the critical values of the individual tests $\sup(T_1, \ldots, T_m) F_T(T_1, \ldots, T_m; q)$ decrease as m increases, the marginal p-values decrease with m and may lead to a test with low power if the number of breaks is large. The following statistic addresses this shortcoming:

$$\text{WDmax}F_T(M, q) = max_{1 \leq m \leq M} \frac{c(q, \alpha, 1)}{c(q, \alpha, m)} \times \sup_{(T_1,\ldots,T_m)} F_T(T_1, \ldots, T_m; q)$$

with $c(q, \alpha, m)$ as the asymptotic critical value of the test $\sup_{(T_1,\ldots,T_m)} F_T(T_1, \ldots, T_m; q)$ for a significance level α.

The third statistic tests for l versus $l+1$ breaks. If the overall minimal value of the sum of squared residuals (over all segments where an additional break is included) is sufficiently smaller than the sum of squared residuals from the l break model, it will be rejected in favor of a model with $l+1$ breaks. The The test is defined by

$$F_T(l+1|l) = \left\{ S_t\left(\widehat{T}_1, \ldots, \widehat{T}_l\right) - \min_{1\le i\le l+1} \inf_{\tau\in\Delta_{i,\eta}} S_t\left(\widehat{T}_1, \ldots, \widehat{T}_{i-1}, \tau, \widehat{T}_i, \ldots, \widehat{T}_l\right) \right\} / \sigma^2$$

where

$$\Delta_{i,\eta} = \left\{ \tau; \widehat{T}_{i-1} + \left(\widehat{T}_i - \widehat{T}_{i-1}\right)\eta \le \tau \le \widehat{T}_i + \left(\widehat{T}_i - \widehat{T}_{i-1}\right)\eta \right\}$$

and $\widehat{\sigma}^2$ is a consistent estimate of σ^2 under the null hypothesis. Finally Bai and Perron propose a sequential estimation of the break points, as well for a known number as for an unknown number. The sequential estimation yields consistent estimates of the break points, which are not guaranteed to be identical to those obtained by global optimization though, but allows using least squares operations which are only of relatively small complexity order.

2.6. **A Bayesian Change Point Model.** Western and Kleykamp describe a change point model that treats the change point in a time series as a parameter to be estimated and which yields statistical inferences about regression coefficients [12].

Western and Kleykamp propose to model the change of causal variables over time in a regression model with a dummy variable, which takes the value of zero up to the time point marking the end of the first regime, and the value one thereafter:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 I_t(\theta) + \beta_3 I_t(\theta) x_t + e_t,$$

where $I_t(\theta)$ is the dummy variable as a function of the change point θ, y_t the dependent variable, x_t the independent variable, and e_t the error term.

To find the change point θ Western and Kleykamp propose a simple method, which tries a range of values for θ and examines the goodness of fit of the model. They try their approach for the Isaac and Griffin example of the relationship between strikes and unionization. To do this they try their model for the breaks $\theta = 1915, 1916, \ldots, 1950$ and record the respective R^2 statistic for each year. Then they choose the value of θ with the maximum R^2.

Western and Kleykamp statistically justify their ad-hoc method by the fact that the time series of R^2 statistics is proportional to the profile log likelihood for θ, assuming that the error term is normally distributed. Thus the time point of maximum R^2 is the maximum

likelihood of θ. The shortcoming so far is though, that from a Bayesian perspective, inferences about the coefficients are optimistic because the model has not correctly accounted for uncertainty about θ.

The Bayesian change point model of Western and Kleykamp addresses this shortcoming considering the vector model

$$\widehat{y} = X_\theta \beta,$$

where y is a vector of observations on the dependent variable, the matrix X_θ includes all regressors and β is a vector of regression coefficients. X_θ is a function of θ because different change points will yield different regressors. The contribution of an observation y_t to the likelihood of θ is

$$L_t(\theta; y_t) \propto \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{(y_t - \widehat{y}_t)^2}{2\sigma^2}\right],$$

assuming that y conditionally follows a normal distribution, where the likelihood depends on θ through the regression function. Because the likelihood function for this model is discontinuous for discrete values of θ, it is difficult to obtain unconditional inferences about the regression coefficients β using standard likelihood methods.

Inferences can be obtained using a Bayesian approach, which specifies prior distributions of the parameters:

$$\begin{aligned} p(\tau) &= \text{Gamma}(n_0, s_0) & \tau = \sigma^-2 \\ p(\beta) &= N(\beta_0, V_0) & \\ p(\theta) &= (T-1)^{-1} & \tau = 1, \ldots, T-1 \\ p(y|X_\theta) &= N(\widehat{y}, \tau^-1), & \end{aligned}$$

where $\text{Gamma}(a, b)$ is a Gamma distribution with shape parameter a and expectation a/b. The prior probability distribution of θ is uniform and discrete for each time point. A noninformative prior for τ sets n_0 and s_0 to small numbers, e.g. $0,001$. A noninformative prior for β sets $\beta_0 = 0$ and V_0 to a diagonal matrix with large prior variances, say 100.

Estimation of the parameters can be approached with a Gibbs sampler, which simulates draws from the posterior distribution. To implement the Gibbs sampler, posterior distributions are specified for each parameter conditional on all the other parameters in the model. This ultimately yields draws from the unconditional posterior distribution.

The Gibbs sampler begins by setting initial values of the change point, the precision, and the regression coefficients. Then it iterates over the following steps:

(1) Draw the precision τ^* from $\text{Gamma}(n_0 + n/2, s_0 + SS^*/2)$, where the current value of the sums of squares $SS^* = \sum e_t^*$, and $e_t^* = y_t - x'_{\theta^* t}\beta^*$.
(2) Draw the vector of regression coefficients β^* from the multivariate normal distribution $N(\beta_1, V_1)$ with mean vector

$\beta_1 = V_1(V_0\beta_0 + \tau^* X'_{\theta^*} y)$, and covariance matrix $V_1 = (V_0^{-1}\beta_0 + \tau^* X'_{0^*} X_{0^*})^{-1}$.

(3) Draw the change point θ^* from the discrete distribution $p(\theta = t|y) = L^*(\theta = t; y)/\sum_t L^*(\theta = t; y)$, where the likelihood is evaluated at each time point $\theta = 1, \ldots, T-1$ using current values of the parameters β^* and τ^*.

The change point model involves a type of Bayesian model averaging, where inference about the coefficients proceeds by taking the sum of their conditional posteriors and weighting by the posterior probability of each model. Thus, Bayesian inference accounts for the uncertainty about the location of θ by integrating over, rather than conditioning on θ.

2.7. **Regime-Switching Models.** Hamilton proposes an autoregressive time series model, which specifies the fixed number of regimes a priori. A transition matrix describes the probability of moving from one regime to another for each time point in a series.

Hamilton considers a first-order autoregressive model of the form

$$y_t = c_{s_t} + \Phi y_{t-1} + \xi_t,$$

with $\xi_t = N(0, \sigma^2)$ and $t = 1, 2, \ldots, t_0, t_0 + 1, t_0 + 2, \ldots$, where t_0 is a known change point. I.e. s_t is a random variable that, as a result of institutional changes, assumes the value $s_t = 1$ for $t = 1, 2, \ldots, t_0$ and $s_t = 2$ for $t = t_0 + 1, t_0 + 2, \ldots$. The simplest specification of a probabilistic model of what caused the change from $s_t = 1$ to $s_t = 2$ is a two-state Markov chain with

$$\Pr(s_t = j|s_{t-1} = i, s_{t-2} = k, \ldots, y_{t-1}, y_{t-2}, \ldots) = \Pr(s_t = j|s_{t-1} = i) = p_{ij}.$$

Assuming one only infers s_t's operation through the observed behavior of y_t, the parameters necessary to fully describe the probability law governing y_t are then the variance of the Gaussian innovation σ^2, the autoregressive coefficient Φ, the two intercepts c_1 and c_2 , and the two state transition probabilities, p_{11} and p_{22}. Therefore, it is assumed that the probability of a change in regime depends on the past only through the value of the most recent regime.

The inference about the value of s_t will take the form of two probabilities $\xi_{jt} = \Pr(s_t = j|\Omega_t; \theta)$ for $j = 1, 2$, where these two probabilities sum to unity by construction. Here $\Omega_t = \{y_t, y_{t-1}, \ldots, y_1, y_0\}$ denotes the set of observations obtained as of date t, and θ is a vector of population parameters, which for the above example is $\theta = (\sigma, \Phi, c_1, c_2, p_{11}, p_{22})'$ and presumed to be known for now. The inference is performed iteratively for $t = 1, 2, \ldots, T$, with step t accepting as input the values

$$\xi_{i,t-1} = \Pr(s_{t-1} = i|\Omega_{t-1}; \theta)$$

for $i = 1, 2$ and producing as output ξ_{jt}. In order to perform this iteration one needs the densities under the two regimes,

$$\eta_{jt} = f(y_t|s_t = j, \Omega_{t-1}; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y_t - c_j - \Phi y_{t-1})^2}{2\sigma^2}\right],$$

for $j = 1, 2$. Given the input $\xi_{i,t-1}$ one can calculate the conditional density of the tth observation from

$$f(y_t|\Omega_{t-1}; \theta) = \sum_{i=1}^{2}\sum_{j=1}^{2} p_{ij}\xi_{i,t-1}\eta_{jt},$$

for $j = 1, 2$ and the output is

$$\xi_{jt} = \frac{\sum_{i=1}^{2} p_{ij}\xi_{i,t-1}\eta_{jt})}{f(y_t|\Omega_{t-1}; \theta)}.$$

The result of the executed iteration will be the sample conditional log likelihood of the observed data

$$\log f(y_1, y_1, \ldots, y_T|y_0; \theta) = \sum_{t=1}^{T} \log f(y_t|\Omega_{t-1}; \theta)$$

for the specified value of θ. An estimate of θ can then be obtained by maximizing $\log f(y_1, y_1, \ldots, y_T|y_0; \theta)$ by numerical optimization.

For the value of ξ_{i0}, which is necessary to start these iterations, there are several options to choose from. If the Markov chain is presumed to be ergodic, the unconditional probabilities

$$\xi_{i0} = \Pr(s_0 = i) = \frac{1 - p_{jj}}{2 - p_{ii} - p_{jj}}$$

can be used. Otherwise one can also simply set $\xi_{i0} = 1/2$ or estimate ξ_{i0} itself by maximum likelihood.

Hamilton also generalizes his model to a $(r \times 1)$ vector of observations y_t, whose density depends on N separate regimes. Let $\Omega_t = \{y_t, y_{t-1}, \ldots, y_1\}$ be the observations through date t, P be an $(N \times N)$ matrix whose row j, column i element is the transition probability p_{ij}, η_t be an $(N \times 1)$ vector whose jth element $f(y_t|s_t = j, \Omega_{t-1}; \theta)$ is the density in regime j, and $\widehat{\xi}_{t|t}$ be an $(N \times 1)$ vector whose jth element is $\Pr(s_t = j|\Omega_t; \theta)$. Then the conditional density of y_t and the conditional probability of s_t generalize to

$$\begin{aligned} f(y_t|\Omega_{t-1}; \theta) &= \mathbf{1}'(P\widehat{\xi}_{t-t|t-1} \odot \eta_t) \\ \widehat{\xi}_{t-t} &= \frac{\widehat{\xi}_{t-t|t-1} \odot \eta_t}{f(y_t|\Omega_{t-1}; \theta)} \end{aligned}$$

where $\mathbf{1}$ denotes an $(N \times 1)$ vector all of whose elements are unity and $\odot$ denotes element-by-element multiplication.

3. Applications

Change point or regime switching models are appropriate for applications in different areas. Isaac and Griffin developed their approach for use in social sciences, more precisely macrosociological processes and historical change. They show, using their method, for the relationship between unionization and strikes that magnitude and even sign of the relationship differ substantially from one period to the next and that the precise year at which one chooses to begin or end an historical analysis can have profound implications for the quantitative findings.

Dias and Embrechts apply their change point analysis for dependence structures in finance and insurance. In finance they demonstrate their model on a pricing example. They show that change point analysis in dependence structures is important because products priced correctly in a "normal" market environment may be severely mispriced in "extreme" market conditions. The estimated expected values in the simulated example vary considerably before and after the change and differ both from the estimated value of the homogenous model which ignors the change.

In insurance Dias and Embrechts apply their model to Danish fire data, which consists of $n = 512$ bivariate observations of losses to contents and losses to profits. They use their method to test for changes in the parameter of the dependence structure and find out one cannot reject the hypothesis, that there is no change, although a Kolmogorov-Smirnov goodness of fit test suggests it.

Western and Kleykamp analyze the real wage growth in 18 OECD countries for the period 1965–1992 [12]. They model the shift in wage-setting regimes for country i at time t as:

$$y_{it} = \alpha_0 + x'_{it}\alpha + I_t(\theta)\beta_0 + I_t(\theta)(x'_{it}\beta) + \epsilon_{it},$$

where covariates are collected in the vector x'_{it}, and the dummy variable $I_t(\theta)$ equals 1 for $t \geq \theta$ and 0 otherwise. The error ϵ_{it} is assumed to follow a normal distribution. The independent variables are unemployment, inflation, productivity growth, bargaining centralization, labor government, and union density.

Contrary to the original Analysis of Western and Healy, which assumed a change point after the first OPEC oil shock in 1973–1974, Western and Kleykamp find the best fitting model with $\theta = 1976$. In this model only productivity growth has a significant influence on wage growth before the regime change. Afterwards also the inflation becomes significant.

Western and Kleykamp also compare an ordinary least squares estimator for the independent variables to their Bayesian model but it turns out that there are few differences in the results. The evidence for the break point is so strong in the data that it makes little difference if one conditions on the assumption that $\theta = 1976$.

Van Norden and Schaller use an extension of Hamilton's Markov switching techniques to describe an analyze stock market returns [11]. To determine whether there is switching in stock market returns, they consider four specifications. In the first, for the null hypothesis of no switching, stock market returns R are drawn from a single Gaussian distribution with mean α_0 and variance σ_0:

$$R_t = \alpha_0 + \sigma_0 \epsilon_t$$

and ϵ_t is a standard Gaussian variable.

Van Norden and Schaller consider three alternative hypotheses. In the first, returns are drawn from two distributions with different means (α_0 and α_1):

$$R_t = \alpha_0(1 - S_t) + \alpha_1 S_t + \sigma_0 \epsilon_t$$

where S_t is a binary state variable which follows a first-order Markov chain:

$$\Pr(S_t = 0 | S_{t-1} = 0) = q$$

$$\Pr(S_t = 1 | S_{t-1} = 1) = p.$$

In the second alternative, returns are drawn from two distributions with the same mean but different variances (σ_0 and σ_1):

$$R_t = \alpha_0 + [\sigma_0(1 - S_t) + \sigma_1 S_t]\epsilon_t.$$

The third alternative hypothesis allows for different means and variances:

$$R_t = \alpha_0(1 - S_t) + \alpha_1 S_t + [\sigma_0(1 - S_t) + \sigma_1 S_t]\epsilon_t.$$

The distribution from which stock market returns are drawn is determined by the state variable S_t.

Van Norden and Schaller test their hypotheses on the CRSP value-weighted monthly stock market excess returns for the period January 1929 to December 1989. Their results imply very strong rejection of the null hypothesis of no switching. There is a high likelihood of switching in means and even higher likelihood of switching in variances or switching in both, means and variances.

The mean switching model shows enormous differences in the two states. While state 0 has a slightly positive mean excess return of 0.0082, state 1 hast a negative mean excess return of -0.1705. Then again the probability is 98.5% to remain in state 0 but only 26.4% to remain in state 1.

The variance switching model shows that the variance of excess returns is approximately three times as high in state 1 as in state 0. Both states are quite persistent with 99.1% probability to remain in state 0 and 94.1% in state 1.

The third model, which allows switching in both means and variances, shows very similar behavior to the variance switching only model concerning the variance and the transition probabilities. The mean of

state 0 is very similar to the one from the switching in means model but the mean of state 1 is not significantly different from 0.

Furthermore, van Norden and Schaller are consistent with previous studies that suggest that Markov switching models do a better job of capturing stock market returns than ARCH or GARCH specifications. Also they find little evidence of higher order Markov effects.

Van Norden and Schaller also study multivariate specifications. More specifically they test whether, after controlling for switching, there is evidence that stock market returns can be predicted using the log price-dividend ratio. To do so, they use the following specification:

$$R_t = \alpha_0 + \beta_0 d_{t-1} + [\sigma_0(1 - S_t) + \sigma_1 S_t]\epsilon_t,$$

where d_{t-1} is the log price-dividend ratio.

The empirical estimates of $\beta_0 = -0.0046$ are significantly different from 0 and therefore provide strong evidence of predictability. A high price-dividend ratio therefore predicts low excess returns. If one allows switching also in means, as possible in

$$R_t = \alpha_0(1-S_t)+\alpha_1 S_t+\beta_0 d_{t-1}(1-S_t)+\beta_1 d_{t-1}S_t+[\sigma_0(1-S_t)+\sigma_1 S_t]\epsilon_t,$$

Van Norden and Schaller find out that β is only significantly different from zero in state 0. Its value and significance in this state do not differ from the one in the first specification.

Finally van Norden and Schaller examine whether the transition probabilities vary over time. In particular, whether the transition probabilities are influenced by the price-dividend ratio. The motivation is that it would be highly desirable for investors if they were able to predict that regime 1 would occur in the following period, based on currently available information.

Van Norden and Schaller consider the following data generating process:

$$\begin{aligned}
R_t =& \alpha_0(1 - S_t) + \alpha_1 S_t + \beta_0 d_{t-1}(1 - S_t) \\
&+ \beta_1 d_{t-1} S_t + [\sigma_0(1 - S_t) + \sigma_1 S_t]\epsilon_t \\
\Pr(S_t = 0|S_{t-1} = 0) =& q(d_{t-1}) \\
\Pr(S_t = 1|S_{t-1} = 1) =& p(d_{t-1}) \\
q =& \Phi(\gamma_{qo} + \gamma_{qd} d_{t-1}) \\
p =& \Phi(\gamma_{po} + \gamma_{pd} d_{t-1})
\end{aligned}$$

By estimating this model, on the one hand, for β_0 and β_1 constraint to zero and on the other hand unconstrained, van Norden and Schaller test whether predictability is spurious. They find out, that the predictive ability of the price-dividend ratio for returns cannot be attributed to the assumption of constant transition probabilities. On the other hand they also realize, that allowing for time variation in the transition probabilities has little effect on the estimate of β_0. Nevertheless

Model	Type of Change
Isaac and Griffin	Covariance of time series
Box and Jenkins	Intercept term of an autoregression
Hansen	Coefficients of a linear regression
Dias and Embrechts	Parameters of a copula
Bai and Perron	Coefficients of a linear regression
Western and Kleykamp	Coefficients of a linear regression
Hamilton	Intercept term of an autoregression

TABLE 2. Type of Change Implemented in each Model

van Norden and Schaller find the "bad" state hard to predict using the price-dividend ratio.

Overall, van Norden and Schaller provide strong results in favor of switching behaviour in stock market returns. These findings are robust to a variety of different specifications — switching in means, switching in variances, or switching in both means and variances.

4. Conclusion

There are different understandings of the terms "change point" and "regime switch". Table 2 gives an overview of the type of change the models discussed in this paper cover.

The majority of the models presented focus on changes in linear relationships between a dependent variable and one or more independent variables. Dias and Embrechts allow a more general relationship in their copula model but, nevertheless, the model regresses on exogenous variables. Only Box and Jenkins, with their Intervention Analysis, and Hamilton, with his Regime Switching Model, focus on autoregressive time series. It is noticeable, that the term "change point" is mainly used only for non-autoregressive models.

Another point, where the considered models differ, is the respective identification of change. In most models the test for change is a, if not the, main focus. Table 3 gives an overview of implementation of identification of change. The models which offer formal tests for change still differ at least partly from each other. Hansen, e.g., develops a test statistic, which allows non-stationary regressors, in contrast to the other tests. Moreover, Bai and Perron differentiate their model by allowing multiple changes.

It is worth remarking, that the two autoregressive models by Box and Jenkins and Hamilton do not contain such formal tests. Box and Jenkins assume that there is exactly one change point, which is known

a priori and the model only revolves around the type of change. Hamilton, on the other hand, allows multiple changes but does not offer a formal test. Thus, the number of regimes has to be known a priori.

One more property, where the considered models differ, is the estimation method used. Most models use ordinary least squares (OLS) or maximum likelihood (ML) estimation techniques, if possible. But only the Bayesian model of Western and Kleykamp and Hamilton's regime switching model take the insecurity about the change point into account. Western and Kleykamp compare their model to Hamilton's as follows: "The Markov-switching model is similar to the Bayesian change point model. Both models draw a probability distribution over the space of possible regimes. Like the change-point model, structural coefficients in the switching model are a stochastic mixture where the mixture weights are given by the probability of being in a particular regime."

Table 4 shows an overview of the estimation techniques used:

The change point models regarded in this paper have, to our knowledge, so far mainly been applied in social and economic context rather than finance. This state does not have to remain though. Dias and Embrechts e.g. give a good illustration of how to apply their model to pricing in different market environments. The regime switching model on the other hand already has many applications in finance, as van Norden and Schaller show. These two topics revolve about the same question, however, i.e. parameter stability in time series models.

Concluding, we assume that change point and regime switching models will remain to be researched. There are possibilities for further development, which is demonstrated by recent approaches to change point modeling in autoregressive processes that are worth studying: "Change-Point Estimation in Long Memory Nonparametric Models with Applications" [9] considers discontinuity in a fractional ARIMA

Model	Identification of Change
Isaac and Griffin	Diagnostic identification
Box and Jenkins	Change point known a priori
Hansen	Formal test for single change and non-stationary regressors
Dias and Embrechts	Formal test for single change
Bai and Perron	Formal test for multiple changes
Western and Kleykamp	Formal test for single change
Hamilton	Change point known a priori

TABLE 3. Identification of Change Implemented in each Model

Model	Estimation of Parameters
Isaac and Griffin	None
Box and Jenkins	OLS
Hansen	OLS
Dias and Embrechts	ML
Bai and Perron	OLS
Western and Kleykamp	Gibbs sampler, considering probability
Hamilton	ML, considering probability

TABLE 4. Identification of Parameters Implemented in each Model

process. "Estimation Mean Change-Point in ARCH Models with Heavy-Tailed Innovations" [7] provides a test for a change point in an ARCH model. "Full predictivistic modeling of stock market data: Application to change point problems" [10] is especially interesting in finance, since it identifies multiple change points in the mean and variance of a stock market return time series.

REFERENCES

[1] Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47 – 78, 1998.

[2] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time Series Analysis: Forecasting and Control*, chapter 10: Transfer Function Model Building. Prentice Hall PTR, Upper Saddle River, NJ, USA, 3rd edition, 1994.

[3] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time Series Analysis: Forecasting and Control*, chapter 12: Intervention Analysis Methods and Outlier Detection. Prentice Hall PTR, Upper Saddle River, NJ, USA, 3rd edition, 1994.

[4] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time Series Analysis: Forecasting and Control*, chapter 11: Identification, Fitting, and Checking of Transfer Function Models. Prentice Hall PTR, Upper Saddle River, NJ, USA, 3rd edition, 1994.

[5] Alexandra Dias and Paul Embrechts. Change-point analysis for dependence structures in finance and insurance, 2002.

[6] Bruce E. Hansen. Testing for structural change in conditional models. *Journal of Econometrics*, 97(1):93 – 115, 2000.

[7] Jin Hao, Tian Zheng, and Yang Yunfeng. Estimation mean change-point in arch models with heavy-tailed innovations. *Communications in Statistics: Simulation & Computation*, 39(2):390 – 404, 2010.

[8] Larry W. Isaac and Larry J. Griffin. A historicism in time-series analyses of historical process: Critique, redirection, and illustrations from u.s. labor history. *American Sociological Review*, 54(6):873 – 890, 1989.

[9] Wang Lihong. Change-point estimation in long memory nonparametric models with applications. *Communications in Statistics: Simulation & Computation*, 37(1):48 – 61, 2008.

[10] R.H. Loschi, P.L. Iglesias, R.B. Arellano-Valle, and F.R.B. Cruz. Full predictivistic modeling of stock market data: Application to change point problems. *European Journal of Operational Research*, 180(1):282 – 291, 2007.
[11] Huntley Schaller and Simon Van Norden. Regime switching in stock market returns. *Applied Financial Economics*, 7(2):177 – 191, 1997.
[12] Bruce Western and Meredith Kleykamp. A bayesian change point model for historical time series analysis, 2004.

ETS Institut für Wirtschaftstheorie und Statistik
Gebäude 20.12 (Kollegium am Schloss)
D-76128 Karlsruhe
E-mail address, Paul Weskamp: paul.weskamp@student.kit.edu
E-mail address, Markus Höchstötter: markus.hoechstoetter@kit.edu
URL: http://statistik.ets.kit.edu/