

Structural Breaks in Time Series*

Alessandro Casini †

Pierre Perron‡

Boston University

Boston University

May 9, 2018

Abstract

This chapter covers methodological issues related to estimation, testing and computation for models involving structural changes. Our aim is to review developments as they relate to econometric applications based on linear models. Substantial advances have been made to cover models at a level of generality that allow a host of interesting practical applications. These include models with general stationary regressors and errors that can exhibit temporal dependence and heteroskedasticity, models with trending variables and possible unit roots and cointegrated models, among others. Advances have been made pertaining to computational aspects of constructing estimates, their limit distributions, tests for structural changes, and methods to determine the number of changes present. A variety of topics are covered. The first part summarizes and updates developments described in an earlier review, Perron (2006), with the exposition following heavily that of Perron (2008). Additions are included for recent developments: testing for common breaks, models with endogenous regressors (emphasizing that simply using least-squares is preferable over instrumental variables methods), quantile regressions, methods based on Lasso, panel data models, testing for changes in forecast accuracy, factors models and methods of inference based on a continuous records asymptotic framework. Our focus is on the so-called off-line methods whereby one wants to retrospectively test for breaks in a given sample of data and form confidence intervals about the break dates. The aim is to provide the readers with an overview of methods that are of direct usefulness in practice as opposed to issues that are mostly of theoretical interest.

JEL Classification: C22

Keywords: Change-point, linear models, testing, confidence intervals, trends, stationary and integrated regressors, factor models, Lasso, forecasts.

*Prepared for the *Oxford Research Encyclopedia of Economics and Finance*. We thank Palgrave Macmillan for granting permission to use parts of the following prior review: Perron, P. “Structural Change” in *The New Palgrave Dictionary of Economics*, 2nd ed, S. Durlauf and L. Blume (eds.), 2008, Palgrave Macmillan.

†Economics Department, Boston University, 270 Bay State Rd., Boston MA 02215 (acasini@bu.edu).

‡Economics Department, Boston University, 270 Bay State Rd., Boston MA 02215 (perron@bu.edu).

This chapter covers methodological issues related to estimation, testing and computation for models involving structural changes. The amount of work on this subject is truly voluminous and any survey is bound to focus on specific aspects. Our aim is to review developments as they relate to econometric applications based on linear models. Substantial advances have been made to cover models at a level of generality that allow a host of interesting practical applications. These include models with general stationary regressors and errors that can exhibit temporal dependence and heteroskedasticity, models with trending variables and possible unit roots and cointegrated models, among others. Advances have been made pertaining to computational aspects of constructing estimates, their limit distributions, testing and methods to determine the number of changes present. The first part summarizes and updates developments described in an earlier review, Perron (2006), with the exposition following heavily that of Perron (2008). Additions are included for recent developments: testing for common breaks, models with endogenous regressors (emphasizing that simply using least-squares is preferable over instrumental variables methods), quantile regressions, methods based on Lasso, panel data models, testing for changes in forecast accuracy, factors models and methods of inference based on a continuous records asymptotic framework. Our focus is solely on linear models and deals with so-called off-line methods whereby one wants to retrospectively test for breaks in a given sample of data and form confidence intervals about the break dates. Given the space constraint, our review is obviously selective. The aim is to provide an overview of methods that are of direct usefulness in practice as opposed to issues that are mostly of theoretical interest.

The basic setup. We consider the linear regression with m breaks (or $m + 1$ regimes):

$$y_t = x_t' \beta + z_t' \delta_j + u_t, \quad t = T_{j-1} + 1, \dots, T_j, \quad (1)$$

for $j = 1, \dots, m + 1$, following Bai and Perron (1998) (henceforth BP). In this model, y_t is the observed dependent variable; x_t ($p \times 1$) and z_t ($q \times 1$) are vectors of covariates and β and δ_j ($j = 1, \dots, m + 1$) are the corresponding vectors of coefficients; u_t is the disturbance. The break dates (T_1, \dots, T_m) are unknown (the convention that $T_0 = 0$ and $T_{m+1} = T$ is used). The aim is to estimate the regression coefficients and the break points when T observations on (y_t, x_t, z_t) are available. This is a partial structural change model since the parameter vector β is not subject to shifts. When $p = 0$, we obtain a pure structural change model with all coefficients subject to change. A partial structural change model can be beneficial in terms of obtaining more precise estimates and more powerful tests. The method of estimation is standard least-squares (OLS), i.e., minimizing the overall sum of squared residuals (SSR)

$\sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} [y_t - x'_t \beta - z'_t \delta_i]^2$. Let $\hat{\beta}(\{T_j\})$ and $\hat{\delta}(\{T_j\})$ denote the estimates based a partition (T_1, \dots, T_m) denoted $\{T_j\}$. Substituting these in the objective function and denoting the resulting SSR as $S_T(T_1, \dots, T_m)$, the estimated break points are

$$(\hat{T}_1, \dots, \hat{T}_m) = \operatorname{argmin}_{(T_1, \dots, T_m)} S_T(T_1, \dots, T_m), \quad (2)$$

with the minimization taken over a set of admissible partitions (see below). The parameter estimates are those associated with the partition $\{\hat{T}_j\}$, i.e., $\hat{\beta} = \hat{\beta}(\{\hat{T}_j\})$, $\hat{\delta} = \hat{\delta}(\{\hat{T}_j\})$. Since estimation is based on OLS, even if changes in the variance of u_t are allowed, provided they occur at the same dates (T_1, \dots, T_m) , they are not exploited to increase the precision of the break date estimators unless a quasi-likelihood framework is adopted, see below.

The theoretical framework, the assumptions and their relevance. To obtain theoretical results about the consistency and limit distribution of the estimates of the break dates, some conditions need to be imposed on the asymptotic framework, the regressors, the errors, the set of admissible partitions and the break dates. By far the most common asymptotic framework is one whereby as T increases the total span of the data increases such that the length of the regimes increases proportionately, which implies that the break dates are asymptotically distinct, i.e., $T_i^0 = [T\lambda_i^0]$, where $0 < \lambda_1^0 < \dots < \lambda_m^0 < 1$ (a recent alternative framework proposed is to let the span fixed and increase the number of observations by letting the sampling interval decrease; see below). To our knowledge, the most general set of assumptions in the case of weakly stationary, or mixing, regressors and errors are those in Perron and Qu (2006). Along with the asymptotic framework, this implies that what is relevant for inference about a break date is only the neighborhood around the break date considered. Some conditions are technical while others restrict the potential applicability of the results. The assumptions on the regressors specifies that for $w_t = (x'_t, z'_t)'$, $(1/l_i) \sum_{t=T_i^0+1}^{T_i^0+[l_i v]} w_t w'_t \rightarrow_p Q_i(v)$ a non-random positive definite matrix uniformly in $v \in [0, 1]$. It allows their distribution to vary across regimes but requires the data to be weakly stationary stochastic processes. This can be relaxed on a case by case basis though the proofs then depend on the nature of the relaxation. For instance the scaling used forbids trending regressors, unless they are of the form $\{1, (t/T), \dots, (t/T)^p\}$, say, for a polynomial trend of order p . Casting trend functions in this form can deliver useful results in many cases. However, there are instances where specifying trends in unscaled form, i.e., $\{1, t, \dots, t^p\}$, can deliver much better results, especially if level and trend slope changes occur jointly. Results using unscaled trends with $p = 1$ are presented in Perron and Zhu (2005). A comparison of their results with other trend specifications is presented in Deng and Perron (2006). A

generalization with fractionally integrated errors can be found in Chang and Perron (2016). Another important restriction is implied by the requirement that the limit be a fixed, as opposed to stochastic, matrix. This precludes integrated processes as regressors (i.e., unit roots). In the single break case, this has been relaxed by Bai et al. (1998) who consider structural changes in cointegrated relationships in a system of equations. Kejriwal and Perron (2008) provide results for multiple structural changes in a single cointegrating vector. Consistency still applies but the rate of convergence and limit distributions are different.

The assumptions on u_t and $\{w_t u_t\}$ impose mild restrictions and permit a wide class of potential correlation and heterogeneity (including conditional heteroskedasticity) and lagged dependent variables. It rules out errors with unit roots, which can be of interest; for example when testing for a change in the deterministic component of the trend function for an integrated series (Perron and Zhu, 2005). The set of conditions is not the weakest possible. For example, Lavielle and Moulines (2000) allow the errors to be long memory processes but consider only the case of multiple changes in the mean. It is also assumed that the minimization problem is taken over all partitions such that $T_i - T_{i-1} \geq \epsilon T$ for some $\epsilon > 0$. This is not restrictive in practice since ϵ can be small. Under these conditions, the break fractions λ_i^0 are consistently estimated, i.e., $\hat{\lambda}_i \equiv (\hat{T}_i/T) \rightarrow_p \lambda_i^0$ and that the rate of convergence is T . The estimates of the break dates are not consistent themselves. The estimates of the other parameters have the same distribution as would prevail if the break dates were known. Kejriwal and Perron (2008) obtain similar results with $I(1)$ regressors for a cointegrated model subject to multiple changes, using the static regression or a dynamic regression augmented with leads and lags of the first differences of the $I(1)$ regressors.

Allowing for restrictions on the parameters. Perron and Qu (2006) consider a broader framework whereby linear restrictions on the parameters can be imposed. The class of models considered is $y_t = z_t' \delta_j + u_t$ ($t = T_{j-1} + 1, \dots, T_j$) where $R\delta = r$, with R a k by $(m+1)q$ matrix with rank k and r , a k dimensional vector of constants. The assumptions are the same as discussed above. There is no need for a distinction between variables whose coefficients are allowed to change and those whose coefficients are not allowed to change. Restricting some coefficients to be identical across regimes can yield a partial structural change model. This is a useful generalization since it permits a wider class of models of practical interests; e.g., a model with a specific number of states less than the number of regimes, or one where a subset of coefficients may be allowed to change over only a limited number of regimes. They show that the same consistency and rate of convergence results hold. Moreover, the limit distributions of the estimates of the break dates are unaffected by the imposition of valid

restrictions, but improvements can be obtained in finite samples. The main advantages of imposing restrictions are that more powerful tests and more precise estimates are obtained.

Method to Compute Global Minimizers. To estimate the model, we need global minimizers of the objective function (2). A standard grid search requires least squares operations of order $O(T^m)$, which is prohibitive when $m > 2$. Bai and Perron (2003a), henceforth BP-2003a, discuss a method based on a dynamic programming algorithm that is very efficient (see also Hawkins, 1976). Indeed, the additional computing time needed to estimate more than two break dates is marginal compared to the time needed to estimate a two break model. Consider the case of a pure structural change model. The basic idea is that the total number of possible segments is at most $T(T + 1)/2$ and is therefore of order $O(T^2)$. One then needs a method to select which combination of segments yields a minimal value of the objective function. This is achieved efficiently using a dynamic programming algorithm. For models with restrictions (including a partial structural change model), an iterative procedure is available, which in most cases requires very few iterations. Hence, even with large samples, the computing cost is small. If the sample is very large, various methods have been proposed that are of order $O(T)$ in computation time. Given that they are of less importance in economics, such procedures are not reviewed. Also, in the context of very large data sets, methods using Lasso, discussed later, appear more promising.

The limit distribution of the estimates of the break dates. With the assumptions on the regressors and errors, and given the asymptotic framework, the limit distributions of the estimates of the break dates are independent, so that the analysis is the same as for a single break. This holds because the distance between each break increases at rate T , and the mixing conditions on the regressors and errors impose a short memory property so that distant events are independent. The main results for this case are those of Bai (1997a) for the single break case and the extension of BP (1998) for multiple breaks. The limit distribution depends on: a) the magnitude of the change in coefficients (larger changes leading to higher precision), b) the limit sample moment matrices of the regressors for the pre and post break segments; c) the so-called ‘long-run’ variance of $\{w_t u_t\}$, which accounts for serial correlation; d) whether the regressors are trending or not. In all cases, the nuisance parameters can be consistently estimated and confidence intervals constructed. For a change of fixed magnitude the limit distribution depends on the finite sample distribution of the errors. To get rid of this dependence, the asymptotic framework is modified with the change in parameters getting smaller as T increases, but slowly enough for the estimated break fraction to remain consistent. The limit distribution obtained in Bai (1997a) and BP (1998) applies to the

case with no trending regressors. With trending regressors, a similar result is still possible (assuming trends of the form (t/T)); see Bai (1997a) when z_t is a polynomial time trend. For an unscaled linear trend, see Perron and Zhu (2005). Deng and Perron (2006) show that the shrinking shift asymptotic framework leads to a poor approximation for a change in a linear trend and that the limit distribution based on a fixed magnitude of shift is preferable. In a cointegrating regression with $I(1)$ variables, Kejriwal and Perron (2008) show that the estimated break fractions are asymptotically dependent so that confidence intervals need to be constructed jointly. If only the intercept and/or the coefficients of the stationary regressors are allowed to change, the estimates of the break dates are asymptotically independent.

Besides the original asymptotic arguments used by Bai (1997a) and BP (1998), Elliott and Müller (2007) propose to invert Nyblom’s (1989) statistic to construct confidence sets, while Eo and Morley (2015) generalized Siegmund’s (1988) method thereby inverting the likelihood-ratio statistic of Qu and Perron (2007); henceforth ILR. The latter methods were mainly motivated by the fact that the empirical coverage rates of the confidence intervals obtained from Bai’s (1997) method are below the nominal level with small breaks. The method of Elliott and Müller (2007) delivers the most accurate coverage rates, though at the expense of increased average lengths of the confidence sets especially with large breaks. The length can be very large (e.g., the whole sample) even with very large breaks; e.g., with serially correlated errors or with lagged dependent variables. Yamamoto (2016) propose a modification of the long-run variance estimator that alleviates this problem, though the method does not apply when lagged dependent variables are present. The ILR-based confidence sets display a coverage probability often above the nominal level and this results in an average length larger than with Bai’s (1997a) method. See Chang and Perron (2017) for a review. Kurozumi and Yamamoto (2015) propose confidence sets obtained by inverting a test that maximizes some weighted average power function. Overall, the findings suggest a need for a method that provides over a wide range of empirically relevant models both good coverage probabilities and reasonable lengths of confidence sets, especially for all break sizes, whether large or small. See below for a recent alternative using a continuous time asymptotic framework and the concept of Highest Density Regions (Casini and Perron, 2017a).

Estimating Breaks one at a time. Bai (1997b) and BP (1998) show that one can consistently estimate all break fractions sequentially. When estimating a single break model in the presence of multiple breaks, the estimate of the break fraction will converge to one of the true break fractions, the one that allows the greatest reduction in the SSR. Then, allowing for a break at the estimated value, a second one break model can be applied which

will consistently estimate the second dominating break, and so on. Interestingly, Yang (2017) shows that this result fails to hold for breaks in a linear trend model. Bai (1997b) considers the limit distribution of the estimates and shows that they are not the same as those obtained when estimating all break dates simultaneously. Except for the last break, the limit distributions depend on the parameters in all segments. He suggests a repartition procedure, which re-estimates each break date conditional on the adjacent break dates. The limit distribution is then the same as when the break dates are estimated simultaneously.

Estimation in a system of regressions. Substantial efficiency gains can be obtained by casting the analysis in a system of regressions. Bai et al. (1998) consider estimating a single break date in multivariate time series allowing stationary or integrated regressors as well as trends. Bai (2000) considers a segmented stationary VAR model with breaks occurring in the parameters of the conditional mean, the covariance matrix of the error term or both. The most general framework is that of Qu and Perron (2007) who consider models of the form $y_t = (I \otimes z_t')S\beta_j + u_t$, for $T_{j-1} + 1 \leq t \leq T_j$ ($j = 1, \dots, m + 1$), where y_t is an n -vector of dependent variables and z_t is a q -vector that includes the regressors from all equations, and $u_t \sim (0, \Sigma_j)$. The matrix S is of dimension nq by p with full column rank (usually a selection matrix that specifies which regressors appear in each equation). They also allow for the imposition of a set of r restrictions of the form $g(\beta, \text{vec}(\Sigma)) = 0$, where $\beta = (\beta_1', \dots, \beta_{m+1}')$, $\Sigma = (\Sigma_1, \dots, \Sigma_{m+1})$ and $g(\cdot)$ is an r dimensional vector. Both within- and cross-equation restrictions are allowed, and in each case within or across regimes. The assumptions on the regressors and errors u_t are similar to those discussed above. Hence, the framework permits a wide class of models including VAR, SUR, linear panel data, change in means of a vector of stationary processes, etc. Models with integrated regressors are not permitted. Allowing for restrictions on β_j and Σ_j permits a wide range of cases of practical interest: partial structural change models, block partial structural change models where only a subset of the equations are subject to change; changes in only some element of the covariance matrix Σ_j ; changes in only the covariance matrix Σ_j , while β_j is the same for all segments; models where the breaks occur in a particular order across subsets of equations; etc.

The method of estimation is again QML (based on Normal errors) subject to the restrictions. They derive the consistency, rate of convergence and the limit distributions of the estimated break dates. Though only root- T consistent estimates of (β, Σ) are needed to construct asymptotically valid confidence intervals, more precise estimates will lead to better finite sample coverage rates. Hence, it is recommended to use the estimates obtained imposing the restrictions even though imposing restrictions does not have a first-order effect

on the limiting distributions of the estimates of the break dates. To make estimation possible in practice, they present an algorithm which extends the one discussed in BP (2003a) using an iterative GLS procedure to construct the likelihood function for all possible segments.

Qu and Perron (2007) also consider “locally ordered breaks”. This applies when the breaks across different equations are “ordered” based on prior knowledge and are “local” since the time span between them is short. Hence, the breaks cannot be viewed as occurring simultaneously nor as asymptotically distinct. An estimation algorithm is presented and a framework to analyze the limit distribution of the estimates is introduced. Unlike the case with asymptotically distinct breaks, the distributions of the estimates of the break dates need to be considered jointly. Their analysis has been considerably extended to cover models with trends and integrated regressors in Li and Perron (2017).

Tests that allow for a single break. To test for a structural change at an unknown date, Quandt (1960) suggests the likelihood ratio test evaluated at the break date that maximizes it. This problem was treated under various degrees of specificity that culminated in the general treatment by Andrews (1993). The basic method is to use the maximum of the likelihood ratio test over all possible values of the parameter in some pre-specified set. For a single change, the statistic is $\sup_{\lambda_1 \in \Lambda_\epsilon} LR_T(\lambda_1)$, where $LR_T(\lambda_1)$ denotes the likelihood ratio evaluated at $T_1 = [T\lambda_1]$ and the maximization is restricted over break fractions in the set $\Lambda_\epsilon = [\epsilon_1, 1 - \epsilon_2]$ with $\epsilon_1, \epsilon_2 > 0$. The limit distribution depends on Λ_ϵ . Andrews (1993) also considers tests based on the maximal value of the Wald and LM tests and shows that they are asymptotically equivalent under a sequence of local alternatives. This does not mean, however, that they all have the same properties in finite samples. The simulations of Vogelsang (1999), for a change in mean with serially correlated errors, showed the $\sup LM_T$ test to be seriously affected by the problem of non monotonic power, in the sense that, for a fixed T , the power of the test can decrease to zero as the change in mean increases.

For Model (1) with *i.i.d.* errors, the LR and Wald tests have similar properties, so we shall discuss the Wald test. For a single change, it is defined by (up to a scaling by q):

$$\sup_{\lambda_1 \in \Lambda_\epsilon} W_T(\lambda_1; q) = \sup_{\lambda_1 \in \Lambda_\epsilon} \hat{\delta}' H' (H(\bar{Z}' M_X \bar{Z})^{-1} H')^{-1} H \hat{\delta} / [SSR_k / (T - 2q - p)] \quad (3)$$

where $\bar{Z} = \text{diag}(Z_1, \dots, Z_{m+1})$ with $Z_i = (z_{T_{i-1}+1}, \dots, z_{T_i})'$, H is such that $(H\delta)' = (\delta'_1 - \delta'_2)$ and $M_X = I - X(X'X)^{-1}X'$. Here SSR_k is the SSR under the alternative hypothesis. The break point that maximizes the Wald test is the same as the estimate obtained by minimizing the SSR provided the minimization problem (2) is restricted to the set Λ_ϵ , i.e., $\sup_{\lambda_1 \in \Lambda_\epsilon} W_T(\lambda_1; q) = W_T(\hat{\lambda}_1; q)$. When serial correlation and/or heteroskedasticity in the

errors is permitted, the Wald test must be adjusted to account for this, i.e.,

$$W_T^*(\lambda_1; q) = (T - 2q - p)\hat{\delta}' H'(H\hat{V}(\hat{\delta})H')^{-1}H\hat{\delta}, \quad (4)$$

where $\hat{V}(\hat{\delta})$ is robust to serial correlation and heteroskedasticity; a consistent estimate of

$$V(\hat{\delta}) = \text{plim}_{T \rightarrow \infty} T(\bar{Z}'M_X\bar{Z})^{-1}\bar{Z}'M_X\Omega M_X\bar{Z}(\bar{Z}'M_X\bar{Z})^{-1}. \quad (5)$$

Note that it can be constructed allowing identical or different distributions for the regressors and the errors across segments. This is important because if an unaccounted variance shift occurs at the same time inference can be distorted (Pitarakis, 2004). The computation of the robust version of the Wald test (4) can be involved. Since the estimate of λ_1 is T -consistent even with correlated errors, an asymptotically equivalent version is to first take the supremum of the original Wald test (3) to obtain the break points, i.e. imposing $\Omega = \sigma^2 I$. The robust version is obtained by evaluating (4) and (5) at these estimates, i.e., using $W_T^*(\hat{\lambda}_1; q)$ instead of $\sup_{\lambda_1 \in \Lambda_\epsilon} W_T^*(\lambda_1; q)$. An issue of concern for such tests is the adequacy of the asymptotic distribution as an approximation to the finite sample distribution. The tests can exhibit size distortions, especially when the regressors and/or the errors are strongly serially correlated. A potential solution is to use a bootstrap method (e.g., Prodan, 2008, Chang and Perron, 2017). Alternatively, given that the estimation of the long-run variance and, consequently, the choice of the bandwidth, play an essential role, Cho and Vogelsang (2017) propose a fixed-bandwidth theory. It is shown to improve upon the standard asymptotic distribution theory, whereby the bandwidth is negligible relative to T . However, while their results are convincing for a given choice of the bandwidth, when the latter is chosen endogenously, e.g., using Andrews' (1991) method, the improvements are not as important.

The vast majority of tests considered in the econometrics literature imposes some trimming that does not consider the possibility of a break occurring near the beginning or end of the sample. This is not so in the statistics literature. Such tests lead to a different limiting distribution, mostly inducing a log-log rate of divergence that needs to be accounted for (e.g., Csörgő and Horváth, 1997). These tests usually have poor finite sample properties. An application for general models in econometrics is Hidalgo and Seo (2013). Their results are, however, restricted to LM tests in order to have decent size properties in finite samples.

Optimal tests. Andrews and Ploberger (1994) consider a class of tests that maximize a weighted average local asymptotic power function. They are weighted functions of the standard Wald, LM or LR statistics for all permissible break dates. Using either of the three basic statistics leads to tests that are asymptotically equivalent and we proceed with

the Wald test. Assuming equal weights are given to all break fractions in some interval $[\epsilon_1, 1 - \epsilon_2]$, the optimal test for distant alternatives is the so-called Exp-type test: $Exp-W_T = \log(T^{-1} \sum_{T_1=[T\epsilon_1]+1}^{T-[T\epsilon_2]} \exp((1/2)W_T(T_1/T)))$. For alternatives close to the null value of no change the optimal test is the Mean- W_T test: $Mean-W_T = T^{-1} \sum_{T_1=[T\epsilon_1]+1}^{T-[T\epsilon_2]} W_T(T_1/T)$. Kim and Perron (2009) approach the optimality issue from a different perspective using the approximate Bahadur measure of efficiency. They show that tests based on the Mean functional are inferior to those based on the Sup and Exp (which are as efficient) when using the same base statistic. When considering tests that incorporate a correction for potential serial correlation in the errors: a) for a given functional, using the LM statistic leads to tests with zero asymptotic relative efficiency compared to using the Wald statistic; b) for a given statistic the Mean-type tests have zero relative efficiency compared to using the Sup and Exp versions, which are as efficient. Hence, the preferred test should be the Sup or Exp-Wald tests. Any test based on the LM statistic should be avoided. Such results, and more discussed below, call into question the usefulness of a local asymptotic criterion to evaluate the properties of testing procedures; on this issue, see also Deng and Perron (2008).

Non monotonicity in power. The issue of non-monotonicity of the power function of structural change tests was first analyzed in Perron (1991) for changes in a trend function. In more general contexts, the Sup-Wald and Exp-Wald tests have monotonic power when only one break occurs under the alternative. As shown in Vogelsang (1999), the Mean-Wald test can exhibit a non-monotonic power function, though the problem has not been shown to be severe. All of these, however, suffer from important power problems when the alternative involves two breaks (Vogelsang, 1997). This suggests that a test will exhibit a non monotonic power function if the number of breaks present is greater than the number accounted for. Hence, though a single break test is consistent against multiple breaks, power gains can result using tests for multiple structural changes (e.g., the *UDmax* test of BP (1998); see below). Crainiceanu and Vogelsang (2001) also show how the problem is exacerbated when using estimates of the long-run variance that allow for correlation. Accordingly, there are problems with tests based on some local asymptotic arguments (e.g., LM or CUSUM) or that do not try to model the breaks explicitly. An example is the test of Elliott and Müller (2006) which is deemed optimal for a class of models involving “small” non-constant coefficients. The suggested procedure does not explicitly model breaks and the test is then akin to a ‘partial sums type’ test. Perron and Yamamoto (2016) shows that, their test suffers from severe non-monotonic power, while offering only modest gains for small breaks. Methods to overcome non-monotonic power problems have been suggested by Altissimo and Corradi (2003) and

Juhl and Xiao (2009). They suggest using non-parametric methods for the estimation of the mean. The resulting estimates and tests are, however, very sensitive to the bandwidth used. There is currently no reliable method to appropriately choose this parameter in the context of structural changes. Kejriwal (2009) propose to use the residuals under the alternative to select the bandwidth and those under the null to compute the long-run variance for the case of a change in mean. Yang and Vogelsang (2011) show that this can be viewed as an LM test with a long-run variance constructed with a constrained small bandwidth. They provide asymptotic critical values based on the fixed-bandwidth asymptotics of Kiefer and Vogelsang (2005). None of these remedies works when lagged dependent variables are present.

Tests for multiple structural changes. A problem with the *Mean- W_T* and *Exp- W_T* tests is that they require computations of order $O(T^m)$. Consider instead the Sup-Wald test. With *i.i.d.* errors, maximizing the Wald statistic is equivalent to minimizing the SSR, which can be solved efficiently and the Wald test for k changes is:

$$W_T(\lambda_1, \dots, \lambda_k; q) = ((T - (k + 1)q - p)/k) \hat{\delta}' H' (H(\bar{Z}' M_X \bar{Z})^{-1} H')^{-1} H \hat{\delta} / SSR_k$$

with H such that $(H\delta)' = (\delta'_1 - \delta'_2, \dots, \delta'_k - \delta'_{k+1})$. The sup-Wald test is $\sup_{(\lambda_1, \dots, \lambda_k) \in \Lambda_{k, \epsilon}} W_T(\lambda_1, \dots, \lambda_k; q) = W_T(\hat{\lambda}_1, \dots, \hat{\lambda}_k; q)$, where $\Lambda_\epsilon = \{(\lambda_1, \dots, \lambda_k); |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_1 \geq \epsilon, \lambda_k \leq 1 - \epsilon\}$ and $(\hat{\lambda}_1, \dots, \hat{\lambda}_k) = (\hat{T}_1/T, \dots, \hat{T}_k/T)$, with $(\hat{T}_1, \dots, \hat{T}_k)$ obtained minimizing the SSR over Λ_ϵ . With serial correlation and heteroskedasticity in the errors, the test is

$$W_T^*(\lambda_1, \dots, \lambda_k; q) = ((T - (k + 1)q - p)/k) \hat{\delta}' H' (H\hat{V}(\hat{\delta})H')^{-1} H \hat{\delta},$$

with $\hat{V}(\hat{\delta})$ as defined by (5). Again, the asymptotically equivalent version with the Wald test evaluated at the estimates $(\hat{\lambda}_1, \dots, \hat{\lambda}_k)$ is used to make the problem tractable. Critical values are presented in BP (1998) and Bai and Perron (2003b). The importance of the choice of ϵ for the size and power of the test is discussed in BP (2003a) and Bai and Perron (2006). Often, one may not wish to pre-specify a particular number of breaks. Then a test of the null hypothesis of no structural break against an unknown number of breaks given some upper bound M can be used. These are called ‘double maximum tests’. The first is an equal-weight version defined by $UD \max W_T(M, q) = \max_{1 \leq m \leq M} W_T(\hat{\lambda}_1, \dots, \hat{\lambda}_m; q)$. The second test applies weights to the individual tests such that the marginal p-values are equal across values of m denoted $WD \max F_T(M, q)$ (see BP, 1998). The choice $M = 5$ should be sufficient for most applications. In any event, the critical values vary little as M is increased beyond 5. The Double Maximum tests are arguably the most useful to determine if structural changes are present: 1) there are types of multiple structural changes that are difficult to

detect with a single break test change (e.g., two breaks with the first and third regimes the same); 2) the non-monotonic power problem when the number of changes is greater than specified is alleviated; 3) the power of the double maximum tests is almost as high as the best power that can be achieved using a test with the correct number of breaks (BP, 2006).

Sequential tests. BP (1998) also discuss a test of ℓ versus $\ell + 1$ breaks, which can be used to estimate the number of breaks using a sequential testing procedure. For the model with ℓ breaks, the estimated break points denoted by $(\hat{T}_1, \dots, \hat{T}_\ell)$ are obtained by a global minimization of the SSR. The strategy proceeds by testing for the presence of an additional break in each of the $(\ell + 1)$ segments obtained using the partition $\hat{T}_1, \dots, \hat{T}_\ell$. We reject in favor of a model with $(\ell + 1)$ breaks if the minimal value of the SSR over all segments where an additional break is included is sufficiently smaller than that from the ℓ breaks model. The break date selected is the one associated with this overall minimum. The limit distribution of the test is related to that of a test for a single change. Bai (1999) considers the same problem allowing the breaks to be global minimizers of the SSR under both the null and alternative hypotheses. The limit distribution of the test is different. A method to compute the asymptotic critical values is discussed and the results extended to the case of trending regressors. These tests can form the basis of a sequential testing procedure by applying them successively starting from $\ell = 0$, until a non-rejection occurs. The simulation results of BP (2006) show that such estimate of the number of breaks is better than those obtained using information criteria as suggested by, e.g., Liu et al. (1997) (see also, Perron, 1997). But this sequential procedure should not be applied mechanically. In several cases, it stops too early. The recommendation is to first use a double maximum test to ascertain if any break is at all present. The sequential tests can then be used starting at some value greater than 0. Kurozumi and Tuvaandorj (2011) consider useful information criteria explicitly tailored to structural change problems, which should complement a sequential testing procedure.

Tests for restricted structural changes. Consider testing the null hypothesis of 0 break versus an alternative with k breaks in a model which imposes the restrictions $R\delta = r$. In this case, the limit distribution of the Sup-Wald test depends on the nature of the restrictions so that it is not possible to tabulate critical values valid in general. Perron and Qu (2006) discuss a simulation algorithm to compute the relevant critical values given some restrictions. Imposing valid restrictions results in tests with much improved power.

Tests for structural changes in multivariate systems. Bai et al. (1998) consider a Sup-Wald test for a single common change in a multivariate system. Qu and Perron (2007) extend the analysis to multiple structural changes. They consider the case where only a

subset of the coefficients is allowed to change, whether in the parameters of the conditional mean, the covariance matrix of the errors, or both. The tests are based on the maximized likelihood ratio over permissible partitions assuming *i.i.d.* errors. They can be corrected for serial correlation and heteroskedasticity when testing for changes in the parameters of the conditional mean assuming no change in the covariance matrix of the errors.

An advantage of the framework of Qu and Perron (2007) is that it allows studying changes in the variance of the errors in the presence of simultaneous changes in the parameters of the conditional mean, thereby avoiding inference problem when changes in variance are studied in isolation. Also, it allows for the two types of changes to occur at different dates, thereby avoiding problems related to tests for changes in the parameters when a change in variance occurs at some other date. Their results are, however, only valid in the case of normally distributed errors when testing for changes in variances (or covariances). This problem was remedied by Perron and Zhou (2008) who propose tests for changes in the variances of the errors allowing for changes in the parameters of the regression in the context of a single equation model. They also consider various extensions including testing for changes in the parameter allowing for change in variances and testing for joint changes. These tests are especially important in light of Hansen's (2000) analysis. First note that the limit distribution of the tests in a single equation system are valid under the assumption that the regressors and the variance of the errors have distributions that are stable across the sample. He shows that when this condition is not satisfied the limit distribution changes and the test can be distorted. If the errors are homoskedastic, the size distortions are quite mild but they can be severe when a change in variance occurs. Both problems of changes in the distribution of the regressors and the variance of the errors can be handled using the framework of Qu and Perron (2007) and Perron and Zhou (2008). If a change in the variance of the residuals is a concern, one can perform a test for no change in some parameters of the conditional model allowing for a change in variance. If changes in the marginal distribution of some regressors is a concern, one can use a multi-equations system with equations for these regressors.

Tests valid with $I(1)$ regressors. With $I(1)$ regressors, a case of interest is a system of cointegrated variables. For testing, Hansen (1992) considers the null hypothesis of no change in both coefficients and proposed Sup and Mean-LM tests for a one time change. He also considers a version of the LM test directed against the alternative that the coefficients are random walk processes. Kejriwal and Perron (2010a) provide a comprehensive treatment of issues related to testing for multiple structural changes at unknown dates in cointegrated regression models using the Sup-Wald test. They allow both $I(0)$ and $I(1)$ variables and

derive the limiting distribution of the Sup-Wald test for a given number of cointegrating regimes. They also consider the double maximum tests and provide critical values for a wide variety of models that are expected to be relevant in practice. The asymptotic results have important implications for inference. It is shown that in models involving both $I(1)$ and $I(0)$ variables, inference is possible as long as the intercept is allowed to change across regimes. Otherwise, the limiting distributions of the tests depend on nuisance parameters. They propose a modified Sup-Wald test that has good size and power properties. Note, however, that the Sup and Mean-Wald test will also reject when no structural change is present and the system is not cointegrated. Hence, the application of such tests should be interpreted with caution. No test is available for the null hypothesis of no change in the coefficients allowing the errors to be $I(0)$ or $I(1)$. This is because when the errors are $I(1)$, we have a spurious regression and the parameters are not identified. To be able to properly interpret the tests, they should be used in conjunction with tests for the presence or absence of cointegration allowing shifts in the coefficients (see, Perron, 2006).

Tests valid whether the errors are $I(1)$ or $I(0)$. The issue of testing for structural changes in a linear model with errors that are either $I(0)$ or $I(1)$ is of interest when the regression is a polynomial time trend (e.g., testing for a change in the slope of a linear trend). The problem is to devise a procedure that has the same limit distribution in both the $I(0)$ and $I(1)$ cases. The first to provide such a solution is Vogelsang (2001). He also accounts for correlation with an autoregressive approximation so that the Wald test has a non-degenerate limit distribution in both the $I(0)$ and $I(1)$ cases. The novelty is that he weights the statistic by a unit root test scaled by some parameter. For any given significance level, a value of this scaling parameter can be chosen so that the asymptotic critical values will be the same. His simulations show, however, the test to have little power in the $I(1)$ case so that he resorts to advocating the joint use of that test and a normalized Wald test that has good properties in the $I(1)$ case but has otherwise very little power in the $I(0)$ case. Perron and Yabu (2009b) and Harvey et al. (2009) have independently proposed procedures that achieve the same goal and that were shown to have better size and power than that of Vogelsang (2001). The approach of Harvey et al. (2009) builds on the work of Harvey et al. (2007). It is based on a weighted average of the regression t-statistic for a change in the slope of the trend appropriate for the case of $I(0)$ and $I(1)$ errors. In the former case a regression in levels is used while in the latter a regression in first-differences is used. With an unknown break date, the supremum over a range of possible break dates is taken. As in Vogelsang (2001), a correction is required to ensure that, for a given significance level, the

weighted test has the same asymptotic critical value in both the $I(0)$ and $I(1)$ cases.

Perron and Yabu (2009b) builds on Perron and Yabu (2009a) who analyzed hypothesis testing on the slope coefficient of a linear trend model. The method is based on a Feasible Quasi Generalized Least Squares approach that uses a superefficient estimate of the sum of the autoregressive parameters α when $\alpha = 1$. The estimate of α is the OLS estimate from an autoregression applied to detrended data and is truncated to take a value 1 whenever it is in a $T^{-\delta}$ neighborhood of 1. This makes the estimate “super-efficient” when $\alpha = 1$. Theoretical arguments and simulation evidence show that $\delta = 1/2$ is the appropriate choice. Perron and Yabu (2009b) analyze the case of testing for changes in level or slope of the trend function of a univariate time series. When the break dates are unknown, the limit distribution is nearly the same in the $I(0)$ and $I(1)$ cases using the Exp-Wlad test. Hence, it is possible to have tests with nearly the same size in both cases. To improve the finite sample properties, use is made of a bias-corrected version of the OLS estimate. The Perron-Yabu test has greater power overall; see Chun and Perron (2013). Kejriwal and Perron (2010b) extend the results to show that the test of Perron and Yabu (2009b) can be applied in a sequential manner using the same critical values. An alternative perspective was provided by Sayginsoy and Vogelsang (2011) who use a fixed-bandwidth asymptotic theory. Extensions that allow the errors to be fractionally integrated have been considered by Iacone et al. (2013a,b).

Testing for common breaks. Oka and Perron (2017) consider testing for common breaks across or within equations in a multivariate system. The framework is very general and allows stationary or integrated regressors and trends. The null hypothesis is that breaks in different parameters occur at common locations or are separated by some positive fraction of the sample size. Under the alternative hypothesis, the break dates are not the same and need not be separated by a positive fraction of the sample size across parameters. A quasi-likelihood ratio test assuming normal errors is used. The quantiles of the limit distribution need to be simulated and an efficient algorithm is provided. Kim, Oka, Estrada and Perron (2017) extend this work to cover the case of testing for common breaks in a system of equations involving joint-segmented trends. The motivation was spurred by a need to test whether the breaks in the slope of the trend functions of temperatures and radiative forcing, occurring in 1960, are common (see, Estrada, Perron, Martinez-Lopez, 2013).

Band spectral regressions and low frequency changes. Perron and Yamamoto (2013) consider the issue of testing for structural change using a band-spectral analysis. They allow changes over time within some frequency bands, permitting the coefficients to be different across frequency bands. Using standard assumptions, the limit distributions obtained are

similar to those in the time domain counterpart. They show that when the coefficients change only within some frequency band (e.g., the business cycle) we can have increased efficiency of the estimates of the break dates and increased power for the tests provided, of course, that the user chosen band contains the band at which the changes occur. They also discuss a very useful application in which the data is contaminated by some low frequency process and that the researcher is interested in whether the original non-contaminated model is stable. For example, the dependent variable may be affected by some random level shift process (a low frequency contamination) but at the business cycle frequency the model of interest is otherwise stable. They show that all that is needed to obtain estimates of the break dates and tests for structural changes that are not affected by such low frequency contaminations is to truncate a low frequency band that shrinks to zero at rate $\log(T)/T$. Simulations show that the tests have good sizes for a wide range of truncations. The exact truncation does not really matter, as long as some of the very low frequencies are excluded. Hence, the method is quite robust. They also show that the method delivers more precise estimates of the break dates and tests with better power compared to using filtered series via a band-pass filter or with a Hodrick-Prescott (1997) filter. This work is related to a recent strand in the literature that attempts to deliver methods robust to low frequency contaminations. One example pertains to estimation of the long-memory parameter. It is by now well known that spurious long-memory can be induced by level shifts or various kinds of low frequency contaminations. Perron and Qu (2007, 2010), Iacone (2010), McCloskey and Perron (2013) and Hou and Perron (2015) exploit the fact low frequency contaminations will produce peaks in the periodograms at a very few low frequencies, and suggest robust procedures eliminating such low frequencies. Tests for spurious versus genuine long-memory have been proposed by Qu (2011). McCloskey and Hill (2017) provide a method applicable to various time series models, such as ARMA, GARCH and stochastic volatility models.

Endogenous regressors. Consider a model with errors correlated with the regressors:

$$y = \bar{X}\delta + u, \tag{6}$$

where $\bar{X} = \text{diag}(X_1, \dots, X_{m+1})$, a T by $(m+1)p$ matrix with $X_i = (x_{T_{i-1}+1}, \dots, x_{T_i})'$ ($i = 1, \dots, m+1$). This includes partial structural change models imposing $R\delta = r$ with R a k by $(m+1)p$ matrix. When the regressors are correlated with the errors, we assume a set of q variables z_t that can serve as instruments, and define the T by q matrix $Z = (z_1, \dots, z_T)'$. We consider a reduced form linking Z and X that itself exhibits m_z changes, so that $X = \bar{W}^0\theta^0 + v$, with $\bar{W}^0 = \text{diag}(W_1^0, \dots, W_{m_z+1}^0)$, the diagonal partition of W at the break dates

$(T_1^{z0}, \dots, T_{m_z}^{z0})$ and $\theta^0 = (\theta_1^0, \dots, \theta_{m_z+1}^0)$. Also, $v = (v_1, \dots, v_T)'$ is a T by q matrix, which can be correlated with u_t but not with z_t . Given estimates $(\hat{T}_1^z, \dots, \hat{T}_{m_z}^z)$ obtained using the method of BP (2003a), one constructs $\hat{W} = \text{diag}(\hat{W}_1, \dots, \hat{W}_{m_z+1})$, a T by $(m_z + 1)q$ matrix with $\hat{W}_l = (w_{\hat{T}_{l-1}^z+1}, \dots, w_{\hat{T}_l^z})'$ for $l = 1, \dots, m_z + 1$. Let $\hat{\theta}$ be the OLS estimate in a regression of X on \hat{W} . The instruments are $\hat{X} = \hat{W}\hat{\theta} = \text{diag}(\hat{X}'_1, \dots, \hat{X}'_{m_z+1})'$ where $\hat{X}_l = \hat{W}_l(\hat{W}'_l\hat{W}_l)^{-1}\hat{W}'_l\tilde{X}_l$ with $\tilde{X}_l = (x_{\hat{T}_{l-1}^z+1}, \dots, x_{\hat{T}_l^z})'$. The instrumental variable (IV) regression is

$$y = \bar{X}^*\delta + \tilde{u}, \quad (7)$$

subject to the restrictions $R\delta = r$, where $\bar{X}^* = \text{diag}(\hat{X}'_1, \dots, \hat{X}'_{m+1})$, a T by $(m + 1)p$ matrix with $\hat{X}_j = (\hat{x}_{T_{j-1}+1}, \dots, \hat{x}_{T_j})'$ ($j = 1, \dots, m + 1$). Also, $\tilde{u} = (\tilde{u}_1, \dots, \tilde{u}_T)'$ with $\tilde{u}_t = u_t + \eta_t$ where $\eta_t = (x'_t - \hat{x}'_t)\delta_j$ for $T_{j-1}^0 + 1 \leq t \leq T_j^0$. The estimates of the break dates are $(\hat{T}_1, \dots, \hat{T}_{m_x}) = \arg \min_{T_1, \dots, T_m} SSR_T^R(T_1, \dots, T_m)$, where $SSR_T^R(T_1, \dots, T_m)$ is the SSR from (7) evaluated at $\{T_1, \dots, T_m\}$. Perron and Yamamoto (2014) provide a simple proof of the consistency and limit distribution of the estimates of the break dates showing that using generated (or second stage) regressors implies that the assumptions of Perron and Qu (2006) are satisfied. For an earlier, more elaborate though less comprehensive treatment, see Hall et al. (2012) and Boldea et al. (2012). Hence, all results of BP (1998) carry through, but care must be applied when the structural and reduced forms contain non-common break.

Of substantive interest is that the IV approach is not necessary as discussed in Perron and Yamamoto (2015); one can simply still use OLS applied to (6). First, except for a knife-edge case, changes in the true parameters imply a change in the probability limits of the OLS estimates, which is equivalent in the leading case of regressors and errors having a homogenous distribution across segments. Second, one can reformulate the model with those limits as the basic parameters so that the regressors and errors are uncorrelated. We are then back to the standard framework. Importantly, using OLS involves the original regressors while IV the second stage regressors, which have less quadratic variation since $\|P_Z X\| \leq \|X\|$. Hence, in most cases, a given change in the parameters will cause a larger change in the conditional mean of the dependent variable using OLS compared with IV. It follows that using OLS delivers consistent estimates of the break fractions and tests with the usual limit distributions and also improves on the efficiency of the estimates and the power of the tests in most cases. Also, OLS avoids weak identification problems inherent when using IV methods. Some care must, however, be exercised. Upon a rejection, one should verify that the change in the probability limit of the OLS parameter estimates is not due to a change in the bias terms. In most applications, there will be no change in the bias but

still one should be careful to assess the source of the rejection. This is easily done since after obtaining the OLS-based estimates of the break dates one would estimate the structural model based on such estimates. The relevant quantities needed to compute the change in bias across segments are then directly available. To elaborate, assume known break dates and let $p \lim_{T \rightarrow \infty} (\Delta T_i^0)^{-1} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+[s\Delta T_i^0]} x_t x_t' = Q_{XX}^i$ and $p \lim_{T \rightarrow \infty} E(X_i u) = \phi_i$ ($i = 1, \dots, m+1$), the probability limit of the OLS estimate is $\delta^* = \delta^0 + [(Q_{XX}^1)^{-1} \phi_1, \dots, (Q_{XX}^{m+1})^{-1} \phi_{m+1}]'$. Any change in δ^0 imply a change in δ^* , except for a knife-edge case when the change in the bias exactly offsets the change in δ^0 . Hence, one can still identify parameter changes using OLS and a change in δ^0 will, in general, cause a larger change in the conditional mean of y_t . Consider writing (6) as

$$y = \bar{X}_0 \delta^0 + P_{\bar{X}_0} u + (I - P_{\bar{X}_0}) u = \bar{X}_0 [\delta^0 + (\bar{X}_0' \bar{X}_0)^{-1} \bar{X}_0' u] + (I - P_{\bar{X}_0}) u = \bar{X}_0 \delta_T^* + u^*,$$

where $u^* = (I - P_{\bar{X}_0}) u$ and $\delta_T^* = [\delta^0 + (\bar{X}_0' \bar{X}_0)^{-1} \bar{X}_0' u]$ for which $\delta_T^* \rightarrow_p \delta^*$. So we can consider a regression in terms of the population value of the parameters. Now, \bar{X}_0 is uncorrelated with u^* so that the OLS estimate, say $\hat{\delta}^*$, is consistent for δ^* . This suggests estimating the break dates by minimizing the SSR from the regression $y = \bar{X} \delta^* + u^*$. OLS dominates IV except for a narrow case, unlikely in practice. The loss in efficiency when using 2SLS can be especially pronounced when the instruments are weak as is often the case. Of course, the ultimate goal is not to get estimates of the break dates per se but of the parameters within each regime, one should then use an IV regression but conditioning on the estimates of the break dates obtained using the OLS-based procedure. Their limit distributions will, as usual, be the same as if the break dates were known. Using the same logic, tests for structural change are more powerful when based on OLS rather than IV. This idea was used by Kurozumi (2017) to show that, with endogenous regressors, using OLS is better when monitoring online for a structural change using a CUSUM-type method.

Quantile Regressions. Following Oka and Qu (2011), assume that the τ th conditional quantile function of y_t given z_t is linear in the parameters and given by $Q_{y_t}(\tau|z_t) = F_{y_t|z_t}^{-1}(\tau|z_t) = z_t \beta(\tau)$. The population coefficient $\beta(\tau)$ is known to be the minimizer of the criterion function $Q_\tau(\beta) = E[\rho_\tau(y - z'\beta)]$ where $\rho_\tau(a)$ is the check function given by $\rho_\tau(a) = a(\tau - 1(a < 0))$; see Koenker (2005). Without structural changes, the quantile regression estimator of $\beta(\tau)$ is the minimizer of the empirical analog $\hat{Q}_\tau(\beta) = \sum_{t=1}^T \rho_\tau(y_t - z_t'\beta)$. Suppose that the τ th quantile has m structural changes, occurring at unknown dates (T_1, \dots, T_m) , such that

$$Q_{y_t}(\tau|z_t) = z_t' \theta_j(\tau), \quad t = T_{j-1} + 1, \dots, T_j \quad (8)$$

for $j = 1, \dots, m + 1$. The vector $\theta_j(\tau)$ are the quantile dependent unknown parameters, with possible restrictions to allow for partial structural changes. Qu (2008) proposes a fluctuation type statistic based on the subgradient and a Wald type statistic based on comparing parameter estimates from different subsamples. They can be used to test for changes occurring in a prespecified quantile, or across quantiles. The limiting distributions under the null is nuisance parameter free and can be simulated. Oka and Qu (2011) consider the estimation of multiple structural changes at unknown dates in one or multiple conditional quantile functions. A procedure to determine the number of breaks is also discussed. The method can deliver more informative results than the analysis of the conditional mean function alone.

Lasso. A growing literature uses Lasso-type methods to address structural change problems, which can estimate the location and number of breaks simultaneously. Estimating structural changes can be viewed as a variable selection problem and Lasso estimates the regression coefficients by minimizing the usual SSR with a penalty for model complexity through the sum of the absolute values of the coefficients. Assume, for simplicity, $p = 0$ and $q = 1$, i.e., a pure structural change model with a single regressor. When $m = 1$, the model is:

$$y_t = z_t \delta_1 + (\delta_2 - \delta_1) z_t \mathbf{1}(t \geq T_1 + 1) + u_t.$$

When there is a break at $t = T_1$, then $\delta_2 \neq \delta_1 \neq 0$, otherwise $\delta_2 = \delta_1$. Using a set of regressors $g(z_t) = \{z_t \mathbf{1}(t \geq \tilde{t}), \forall \tilde{t} \in [\underline{t}, \bar{t}], 1 < \underline{t} < \bar{t} < T\}$, we can express the model as:

$$\begin{aligned} y_t &= z_t \delta_1 + b_1 z_t \mathbf{1}(t \geq \underline{t}) + \dots + b_{\bar{t}-\underline{t}} z_t \mathbf{1}(t \geq \bar{t}) + u_t \\ &= \underbrace{z_t}_{w_{0,t}} \delta_1 + \underbrace{(z_t \mathbf{1}(t \geq \underline{t}))}_{w_{1,t}} b_1 + \dots + \underbrace{(z_t \mathbf{1}(t \geq T_1 + 1))}_{w_{(T_1-\underline{t}+2),t}} \underbrace{b_{T_1-\underline{t}+2}}_{\delta_2-\delta_1} + \dots + \underbrace{(z_t \mathbf{1}(t \geq \bar{t}))}_{z_{(\bar{t}-\underline{t}+1),t}} b_{\bar{t}-\underline{t}+1} + u_t \\ &= w_{0,t} \theta_0 + g(w_{0,t}) \theta_1 + u_t = w'_t \theta + u_t \end{aligned}$$

with $b_i = 0$, $i = 1, \dots, T_1 - \underline{t} + 1, T_1 - \underline{t} + 3, \dots, \bar{t} - \underline{t} + 1$ and $b_{T_1-\underline{t}+2} = |\delta_2 - \delta_1| \neq 0$. Denote by $g(x_{0,t}) = \{w_{1,t}, \dots, w_{(\bar{t}-\underline{t}+1),t}\}$ the transformed regressors generated from $w_{0,t} = z_t$ whose coefficient is subject to change. Then $w_t = \{w_{0,t}, g(w_{0,t})\}$ is the complete set of regressors. If we can consistently estimate the coefficients associated with $g(x_{0,t})$ that are greater than zero, we can date the break point. OLS would not provide consistent estimates because the number of regressors is too large. The method is flexible; e.g., if one has some prior knowledge that a change has occurred at some date, dummy variable can be added without the associated generated regressors. A model with multiple structural changes in many regressors can be obtained as an extension. One simply let $w_{0,t}$ be the vector of regressors whose coefficients are subject to change and $g(w_{0,t})$ be the vector of artificially

constructed regressors obtained from the original ones. In general, the number of regressors is $n = p + q(\bar{r} + 1)$, where $\bar{r} = \bar{t} - \underline{t} + 1$ is the number of transformed regressors associated with the original regressors whose coefficients are allowed to change (fewer are possible if there is prior information about where the breaks cannot occur). Hence, n can be very large, much larger than T . The structural break model has a sparse pattern since few coefficients are non-zero, namely $s = p + q(m + 1)$. The Lasso estimator for sparse models is $\hat{\theta} = \arg \min_{\theta \in \mathbb{R}^q} \hat{Q}(\theta) + (\lambda/T) \|\hat{Y}\theta\|_1$, where λ is the penalty level, $\hat{Q}(\theta) = \sum_{t=1}^T (y_t - w_t'\theta)^2$, $\hat{Y} = \text{diag}(\hat{\gamma}_1, \dots, \hat{\gamma}_p)$ is a diagonal matrix with the penalty loadings and $\|\hat{Y}\theta\|_1 = \sum_{j=1}^p |\hat{\gamma}_j \theta_j|$ is the ℓ_1 -norm. Ideally, these are adapted to information about the error term, which is not feasible since u_t is not observed. In practice, we can use the estimated residuals and proceed via iterations or simply assume homoskedastic Gaussian errors, in which case \hat{Y} is the identity matrix. Often, additional thresholding is applied to remove regressors with small estimated coefficients which may have been included due to estimation error. Then, the thresholded Lasso estimator is $\hat{\theta}(t_L) = (\hat{\theta}_j 1\{|\hat{\theta}_j| > t_L\}, j = 1, \dots, q)$ where $t_L \geq 0$ is the threshold level. The problem is in the choice of λ and t_L . Results are well established for a random sample of data. With serially correlated series, things are more complex.

The following is a partial list of some relevant papers using Lasso for models with structural changes. Harchaoui and Lévy-Leduc (2010) propose a total variation penalty to estimate changes in the mean of a sequence of independent and identically distributed (*i.i.d.*) Gaussian random variables. Bleakley and Vert (2011) propose a group fused Lasso method for changes in the mean of a vector of *i.i.d.* Gaussian random variables assumed to share common break points. Chan, Yau, and Zhang (2014) consider using a group Lasso method for changes in an autoregressive model with heteroskedastic errors. They suggest a two steps method involving the use of an information criterion to select the number of break points. Ciuperca (2014) considers multiple changes in a linear regression model with *i.i.d.* errors using Lasso with an information criterion or adaptive Lasso. Rojas and Wahlberg (2014) use penalized maximum likelihood estimator for changes in the mean of a sequence of *i.i.d.* Gaussian random variables. Aue, Cheung, Lee and Zhong (2014) consider structural breaks in conditional quantiles using the minimum description length principle. While the framework is quite general, the method cannot consistently estimate the number of breaks jointly with their location. Qian and Su (2016) consider estimation and inference of common breaks in panel data models with endogenous regressors. The regressors and errors are, however, restricted to be *i.i.d.* processes, a common feature in this literature up to now. Allowing for general mixing regressors and errors has not, to our knowledge, been achieved. Work is

needed to achieve the level of generality available using standard procedures. Nevertheless, it remains a promising approach, especially in the context of very large datasets.

Factors. The issue of structural breaks in factor models has recently received considerable attention; for a more detailed survey see Bai and Han (2016). We first discuss in some details the methods of Baltagi, Kao and Wang (2017) and Bai, Han and Shi (2017) and then briefly mention other works. The high dimensional factor model with m changes in the factor loadings considered by Baltagi, Kao and Wang (2017) is given by

$$y_{it} = f_t' \phi_i + f_{b,t}' \phi_{i,j} + u_{it}, \quad (t = T_{j-1}^0 + 1, \dots, T_j^0)$$

for $j = 1, \dots, m + 1$, $i = 1, \dots, N$ and $t = 1, \dots, T$ where f_t and $f_{b,t}$ are vectors of factors without and with changes in the loadings, respectively, ϕ_i and $\phi_{i,j}$ are the factor loadings of unit i corresponding to f_t and $f_{b,t}$ in the j -th regime, respectively, and u_{it} is a disturbance term which can have serial and cross-sectional dependence as well as heteroskedasticity. The problem is to estimate the break points, determine the number of factors, and estimate the factors and loadings in each regime. We first discuss a procedure when the number of breaks is known. It first estimates the break points using a simultaneous or sequential method, which leads to consistent estimates for $\lambda_b^0 = T_b^0/T$, though not for T_b^0 . Secondly, it involves plugging-in the break points estimates and estimating the number of factors and the factor space in each regime. Since the factors are latent, one has to determine the number of pseudo factors which is akin to selecting moment conditions whereas in BP (1998) the model is parametric and the moment conditions are known a priori. Baltagi, Kao and Wang (2017) propose to convert the statistical problem from estimating multiple changes in the loadings to estimating changes in the pseudo factors. This relies on the fact that the mean of the second moment matrix of the pseudo factors have changes at the same dates as the loadings. After this conversion, the data become fixed dimensional with observable regressors and conceptually the problem becomes similar to that of Qu and Perron (2007).

Estimation of the break points involve three steps: (1) ignoring breaks, use any consistent estimator \tilde{r} to estimate the number of factors; (2) estimate the first \tilde{r} factors \tilde{g}_t using the Principal Component (PC) method; (3) for any partition (T_1, \dots, T_m) , split the sample into $m + 1$ subsamples, estimate $\tilde{\Sigma}_j = (T_j - T_{j-1})^{-1} \sum_{t=T_{j-1}+1}^{T_j} \tilde{g}_t \tilde{g}_t'$ and calculate the SSR,

$$\tilde{S}(T_1, \dots, T_m) = \sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} [\text{vech}(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}_j)]' [\text{vech}(\tilde{g}_t \tilde{g}_t' - \tilde{\Sigma}_j)].$$

The estimates of the break points are the minimizer of $\tilde{S}(T_1, \dots, T_m)$. The motivation is that the second moment matrix of g_t has breaks at the same dates as the factor loadings.

The results are valid under general high-level conditions on the errors and the factors are allowed to be dynamic and to include lags. To estimate the number of factors, these can be correlated with the errors. The limiting distributions of $\hat{T}_b - T_b^0$ ($b = 1, \dots, m$) have the same form as the one for the single break case derived in Baltagi, Kao and Wang (2016).

Consider now testing for multiple changes in the factor loadings. Following BP (1998), Baltagi, Kao and Wang (2017) propose two different tests: no change versus a fixed number of changes; l versus $l + 1$ changes. The limit distributions follow those in BP (1998). The first test loses power when the number of changes is misspecified. They propose adapting the UDmax and WDmax tests of BP (1998) allowing an unknown number of changes (up to some upper bound). For to the test of $m = l$ versus $m = l + 1$, one first estimates the break points and once they are plugged in, testing for $m = l$ versus $m = l + 1$ changes becomes equivalent to testing no change versus a single change in each regime jointly. The null limiting distribution is obtained by simulations and depends on the number of factors in each regime. When this number is stable, it is similar to that in BP (1998).

Bai, Han and Shi (2017) study the properties of the least-squares estimator of the single break point in a high dimensional factor model. The model is given by

$$y_{it} = \phi'_{ij} f_t + u_{it}, \text{ for } t = 1, \dots, T_b^0, \text{ if } j = 1 \text{ and } t = T_b^0 + 1, \dots, T, \text{ if } j = 2,$$

for $i = 1, \dots, N$ where $T_b^0 = T\lambda_b^0$ is the unknown common break point. The estimation of T_b^0 involves the estimated latent factors. The model in matrix format is:

$$\begin{bmatrix} Y_{T_b^0}^{(1)} \\ Y_{T_b^0}^{(2)} \end{bmatrix} = \begin{bmatrix} F^{(1)} & 0_{T_b^0 \times r} \\ 0_{(T-T_b^0) \times r} & F^{(2)} \end{bmatrix} \begin{bmatrix} \Lambda'_1 \\ \Lambda'_2 \end{bmatrix} + \begin{bmatrix} u_{T_b^0}^{(1)} \\ u_{T_b^0}^{(2)} \end{bmatrix} \equiv G\Theta' + u \quad (9)$$

Model (9) is an observationally equivalent factor model, where the number of factors is doubled and the factor loadings are time invariant. Under (9), the factor process has structural breaks, which was the basis for the framework of Baltagi, Kao and Wang (2017). The estimates of the break points are the minimizer of

$$SSR(T_b, \tilde{F}) = \sum_{i=1}^N \sum_{t=1}^{T_b} (y_{it} - \tilde{f}'_t \tilde{\phi}_{i1})^2 + \sum_{i=1}^N \sum_{t=T_b+1}^T (y_{it} - \tilde{f}'_t \tilde{\phi}_{i2})^2.$$

over $T\lambda_1 \leq T_b \leq T\lambda_2$, where \tilde{f} are estimates of f_t using a principal component method. Since the factors are estimated by a PC method, they are not efficient. However, the efficiency loss relative to the maximum likelihood estimator, vanishes as $N, T \rightarrow \infty$. The errors are assumed to be independent from the factors and loadings, however, u_{it} can be weakly

correlated in both cross-sectional and time dimensions. There can be dependence between $\{f_t\}$ and $\{\phi_{i1}, \phi_{i2}\}$ and the break magnitudes $\phi_{i2} - \phi_{i1}$ can be dependent on the factors. Theoretical results are provided under both large and small breaks where the latter are modeled in two ways: 1) the magnitude of the change in each factor loading is of order $N^{\frac{\nu-1}{2}}$ for some $0 < \nu \leq 1$; 2) the magnitude of the change is fixed but only $O(N^\nu)$ units have a break for some $0 < \nu \leq 1$. The case $\nu = 1$ corresponds to large breaks studied by Chen, Dolado and Gonzalo (2014), Han and Inoue (2015), Cheng, Liao and Schorfheide (2015), and Baltagi, Kao and Wang (2015, 2017). Discussing separately large and small breaks is useful given the asymptotic properties of the PC estimator of the factors. For small breaks, \tilde{T}_b is consistent for T_b^0 as $N, T \rightarrow \infty$ provided conditions on the ratio N/T are satisfied. The consistency result for \tilde{T}_b is strong and different from the univariate case for which only the break fraction is consistently estimated. For large breaks, T_b^0 is not consistently estimable. The framework under small breaks allows a non-degenerate asymptotic distribution similar to that of the OLS break point estimator in panel models (cf. Bai, 2010, and the section on panels below). It does not depend on the exact distribution of the errors but depends on the distribution of the factors f_t , being the same whether f_t is observable or not. Hence, evaluating the limit distribution using the plug-in approach, replacing population quantities with consistent estimates, is not applicable due to rotations. Bai, Han and Shi (2017) propose a bootstrap method, which, however, lacks robustness to cross-sectional correlation.

Additional remarks on factor models follow. First, Bai, Han and Shi's (2017) analysis assumes a known number of factors. This is a strong restriction which future research should relax. Su and Wang (2017) developed an innovative adaptive group Lasso estimator that can determine the number of factors and the break fraction simultaneously but it is valid only under large breaks in the loading matrix. Therefore, joint estimation of the break points and the number of factors remains an open issue even from a computational perspective. It is also necessary to consider alternative inference methods for the break points because the bootstrap procedure does not work very well for small breaks. Other authors have proposed alternative methods for estimating factor models with breaks. Cheng, Liao and Schorfheide (2016) developed a shrinkage method that can consistently estimate the break fraction. Chen (2015) considers a least-squares estimator of the break points and proves the consistency of estimated break fractions while Massacci (2015) studies the least-squares estimation of structural changes in factor loadings under a threshold model setup. Additional tests for structural changes in factor models have been proposed; see Chen, Dolado and Gonzalo (2014), Corradi and Swanson (2014), Han and Inoue (2015), Su and Wang (2015)

and Cheng, Liao and Schorfheide (2016). Breitung and Eickmeier (2011) propose a test for dynamic factor models. Yamamoto and Tanaka (2015) show that their test has nonmonotonic power and propose a modified version that solves the problem. The major restriction for most studies is their focus on testing for a common break date in the factor loadings. While a common break date is sometimes relevant, one cannot exclude the possibility that some of the loadings have breaks at different dates. Additional work is needed in that direction.

Panels. Panel data studies have become increasingly popular including inference about breaks. The literature on estimating panel structural breaks can be categorized as assuming whether the parameters of interest are allowed to be heterogenous across units or not. We focus on heterogeneous panels since they are more relevant in practice and refer the reader to De Wachter and Tzavalis (2012) and Qian and Su (2014) for corresponding methods for homogeneous panels. Bai (2010) considers the problem of estimating a common break point in a panel with N units and T observations for each unit. The model takes the form: $y_{it} = \mu_{ij} + u_{it}$, for $t = 1, \dots, T_b^0$, if $j = 1$ and $t = T_b^0 + 1, \dots, T$, if $j = 2$, where $i = 1, \dots, N$, and u_{it} is a disturbance term. The common break specification means that each unit has a break point at T_b^0 . The model allows for heterogeneous means and break magnitudes $\mu_{i2} - \mu_{i1}$. Bai (2010) provides results for both fixed T and T going to infinity. Unlike in the univariate model where the break point is assumed to correspond to a positive fraction of the total sample size, the panel setup allows one to consider also the case where T_b^0 can take any value in $[1, T - 1]$. The latter case can be studied under the asymptotics framework with $N \rightarrow \infty$ and T fixed or $N, T \rightarrow \infty$ such that $T/N \rightarrow 0$. Importantly, only consistency can be established under the latter scenario. For the derivation of the limiting distribution one needs the standard assumption $T_b^0 = T\lambda_0$. Turning to the assumptions on the errors, Bai (2010) requires stationarity of $\{u_{it}\}$ in the time dimension and independence over i . It is argued that it can be relaxed without affecting the consistency result, though for the asymptotic distribution one requires the cross-sectional dependence to be not too strong. The assumption on the break sizes is $\lim_{N \rightarrow \infty} N^{-1/2} \sum_{i=1}^N (\mu_{i2} - \mu_{i1})^2 = \infty$. To understand, note that if $\mu_{i2} - \mu_{i1}$ were *i.i.d.* random variables with positive variance then the above limit with N^{-1} replacing $N^{-1/2}$ should converge to a positive constant. Thus, the condition does not require every unit to have a break. The estimation method involves least-squares. For a given $1 \leq T_b \leq T - 1$, define $\bar{y}_{i1} = T_b^{-1} \sum_{t=1}^{T_b} y_{it}$ and $\bar{y}_{i2} = (T - T_b)^{-1} \sum_{t=T_b+1}^T y_{it}$ which are the estimates of μ_{i1} and μ_{i2} , respectively. Define the sum of squared residuals for the i^{th} equation as $S_{iT}(T_b) = \sum_{t=1}^{T_b} (y_{it} - \bar{y}_{i1})^2 + \sum_{t=T_b+1}^T (y_{it} - \bar{y}_{i2})^2$ for $T_b = 1, \dots, T - 1$ and $S_{iT}(T_b) = \sum_{t=1}^T (y_{it} - \bar{y}_i)^2$ for $T_b = T$ where \bar{y}_i is the whole sample mean. The least-

squares estimator of the break point is defined as $\hat{T}_b = \operatorname{argmin}_{1 \leq T_b \leq T-1} SSR(T_b)$, where $SSR(T_b) = \sum_{i=1}^N S_{iT}(T_b)$. The availability of panel data leads to a stronger result about the rate of convergence. In the univariate case, we have $\hat{T}_b = T_b^0 + O_p(1)$. For panel data, under either fixed T or $T \rightarrow \infty$, $\hat{T}_b = T_b^0 + o_p(1)$ and results show that with $N \rightarrow \infty$ the common break point can be estimated precisely even with a regime having a single observation. The limiting distribution is derived under a small shifts assumption with $\mu_{i2} - \mu_{i1} = N^{-1/2}\Delta_i$, $\Delta_i > 0$. Construction of the confidence intervals requires simulations of the derived limiting distribution. Unlike in the univariate case, a change in variable allowing to express it as a function of quantities that can be consistently estimated cannot be immediately carried out. Nonetheless, noting that a Gaussian random walk and a standard Wiener process have the same distribution at integer times, one can apply the change in variable argument leading to the same inference procedure as in the univariate case, which now only holds approximately, so that inference works as in the univariate setting. Finally, Bai (2010) further considers the case of (possibly) simultaneous break in mean and variance and proposes a QML method.

Kim (2011) studies least-squares estimation of a common deterministic time trend break in large panels with a break either in the intercept, slope or both with a general dependence structure in the errors. He models cross-sectional dependence with a common factor structure and allows the errors to be serially correlated in each equation: $u_{it} = \gamma_i' F_t + e_{it}$ where F_t is a vector of latent common factors, γ_i is a factor loading and e_{it} is a unit specific error. Under joint asymptotics $(T, N) \rightarrow \infty$, serial correlation and cross-sectional dependence are shown to affect the rate of convergence and the limiting distribution of the break point estimator. As in Bai (2010), \hat{T}_b can be consistent for T_b^0 even when T is fixed, though it only holds when the u_{it} are independent across both i and t . When the e_{it} are serially correlated and there are no common factors, the rate of convergence depends on N and is faster than in the univariate case. With common factors generating strong cross sectional dependence, the rate of convergence does not depend on N and reduces to the univariate case (cf. Perron and Zhu, 2005). For fixed shifts, the limiting distribution depends on many elements; e.g., the form of the break, the presence of common factors, stationary versus integrated errors. For a joint broken trend, it can be normal. To obtain a limit theory not depending on the exact distribution of the errors, the break magnitudes need to converge to zero at a rate $N^{-1/2}$, in which case asymptotically valid confidence intervals can be computed via simulations.

Results pertaining to regression models using stationary panels were obtained by Baltagi, Feng and Kao (2016). They consider large heterogeneous panels with common correlated effects (CCE) and allowed for unknown common structural breaks in the slopes. The CCE

setting takes the following form: $y_{it} = x'_{it}\beta(T_b^0) + u_{it}$, where $u_{it} = \gamma'_i F_t + e_{it}$, $x_{it} = \Gamma'_i F_t + v_{it}$, e_{it} , v_{it} are idiosyncratic errors and some or all components of $\beta(T_b^0)$ differ pre and post-break. Due to the correlation between x_{it} and u_{it} , least-squares for each cross-sectional regression could be inconsistent. They use a least-squares method using augmented data and confirm the result in Bai (2010) that the break point T_b^0 can be consistently estimated as both N and T go to infinity. Common breaks in panels were also considered by Qian and Su (2016) and Li, Qian and Su (2016) who study estimation and inference with and without interactive fixed effects using Lasso methods. Kim (2014) generalizes Kim (2011) by allowing a factor structure in the error terms. Baltagi, Feng and Kao (2016) study structural breaks in a heterogeneous large panel with interactive fixed effects. They show the consistency of the estimated break fraction and break date under some conditions.

Most of the work on structural breaks in panels focused on common breaks, in which case T_b^0 itself can be consistently estimated and not only λ_b^0 . One may infer that simply adding a cross-sectional dimension yields more information and precise estimates. This is misleading because the result crucially relies on the assumption that the break is common to all units. Although this may be relevant in practice, the results should be interpreted carefully.

Continuous record asymptotics. Casini and Perron (2017a) consider an asymptotic framework based on a continuous-time approximation, i.e., T observations with a sampling interval h over a fixed time span $[0, N]$, where $N = Th$, with $T \rightarrow \infty$ and $h \rightarrow 0$ with N fixed. Liang et a. (2016) consider a similar, though different, framework for the simple case of a change in mean and they do not provide feasible versions for the construction of the confidence sets, hence we follow the general approach of Casini and Perron (2017a) who consider the following partial structural change model with a single break point:

$$Y_t = D'_t \pi^0 + Z'_t \delta_1^0 + e_t, \quad (t \leq T_b^0); \quad Y_t = D'_t \pi^0 + Z'_t \delta_2^0 + e_t, \quad (t > T_b^0). \quad (10)$$

where $\{D_s, Z_s, e_s\}_{s \geq 0}$ are continuous-time processes and we observe realizations at discrete points of time, namely $\{Y_{kh}, D_{kh}, Z_{kh}; k = 0, \dots, T = N/h\}$. For any process X , we denote its ‘‘increments’’ by $\Delta_h X_k = X_{kh} - X_{(k-1)h}$. For $k = 1, \dots, T$, let $\Delta_h D_k = \mu_{D,k} h + \Delta_h M_{D,k}$ and $\Delta_h Z_k = \mu_{Z,k} h + \Delta_h M_{Z,k}$, where $\mu_{D,t}$, $\mu_{Z,t}$ are the ‘‘drifts’’ and $M_{D,k}$, $M_{Z,k}$ are continuous local martingales. They consider the least-squares estimator of the break point and the analysis is valid for general time series regression models including errors with correlation and/or heteroskedasticity and lagged dependent variables. Under fixed shifts, $\hat{T}_b - T_b^0 = O_p(1)$. Besides the usual small shifts assumption, the limiting distribution is derived under the additional assumption of increasing local variances around the true break date. The

continuous record asymptotic distribution of the least-squares estimator is then given by

$$Th(\hat{\lambda}_b - \lambda_0) \Rightarrow \underset{v \in [-N_b^0 / (\|\delta^0\|^{-2} \bar{\sigma}^2), (N - N_b^0) / (\|\delta^0\|^{-2} \bar{\sigma}^2)]}{\operatorname{argmax}} \{- (\delta^0)' \langle Z_\Delta, Z_\Delta \rangle (v) \delta^0 + 2 (\delta^0)' W(v)\}, \quad (11)$$

where $\langle Z_\Delta, Z_\Delta \rangle (v)$ is the predictable quadratic variation process of Z_t , $W(v)$ is a two-sided centered Gaussian process and $\bar{\sigma}^2$ is the limit of an estimate of the error innovation variance over $[0, N]$. The results (11) is defined on a new “fast time scale”. The latter provides a better approximation to the properties of the finite-sample distribution when h is actually fixed. It is shown that the continuous record asymptotic distribution provides a much better approximation to the finite-sample distribution of the least-squares estimator. The former is highly non-standard and captures the main properties of the latter such as tri-modality, asymmetry and skewness. Thus, basing inference on the continuous record asymptotic theory results in inference procedures about the break date that perform better than existing methods. As shown in Elliott and Müller (2007) and Chang and Perron (2017a), Bai’s (1997) method for constructing confidence intervals for the break date displays a coverage probability far from the nominal level when the magnitude of the break is small. Casini and Perron (2017a) propose constructing confidence sets by computing the Highest Density Regions (HDR) of the density of the continuous record limiting distribution. Their confidence sets are shown to provide accurate converge rates and relatively short length of the confidence sets across different break magnitudes and break locations when compared with those of Bai (1997), Elliot and Müller (2007) and Eo and Morley (2015). In addition, Casini and Perron (2017b) investigate a GL estimation and inference method, which involves transforming the least-squares objective function into a proper distribution (i.e., a Quasi-posterior) and minimizing the expected risk under a given loss function. The analysis is carried out under continuous record asymptotics. This yields a new estimator of the break shown to be more accurate than the original least-squares estimator. The proposed confidence sets, which use the HDR concept, also have coverage rates close to the nominal level and of relatively small length whether the break is small or large. The GL estimator based on the least-squares estimate was also considered under large- T asymptotics by Casini and Perron (2017c) who also relate it to the distribution theory of Bayesian change-point estimators.

Forecasting. We first discuss the concept of forecast failure (or breakdown) and describe methods proposed to detect changes in the forecasting performance over time. Second, we discuss techniques to compare the relative predictive ability of two competing forecast models in an unstable environment. It is useful to clarify the purpose of forecast breakdown tests.

The aim is to assess retrospectively whether a given forecasting model provides forecasts which show evidence of changes (improvements or deterioration) with respect to some loss function. Since the losses can change because of changes in the variance of the shocks (e.g., good luck), detection of a forecast failure does not necessarily mean that a forecast model should be abandoned. Care must be exercised to assess the source of the changes. But if a model is shown to provide stable forecasts, it can more safely be applied in real time. In practice, such forecasts are made at the time of the last available data, using a fixed, recursive or rolling window. Hence, there is a natural separation between the in-sample and out-of-sample periods simply dictated by the last data point. Such is not the case when trying to assess retrospectively whether a given model provides stable forecasts. There is then the need for a somewhat artificial separation between the in and out-of-sample periods at some date labelled T_m , say. This separation date should be such that the model in the in-sample period is stable in some sense, e.g., yielding stable forecasts. This can, however, create problems; e.g., one needs a truncation point T_m to assess forecast failures but the choice of this value is itself predicated on some knowledge of stability.

The forecast failure test of Giacomini and Rossi (2009), GR (2009) hereafter, is a global and retrospective test which compares the in-sample average with the out-of-sample average of the sequence of forecast losses. Adopting the same notation as in the previous section, we have $k = 1, \dots, T$ observations with a sampling frequency h over the time span $[0, N]$ with $N = Th$. We recover the setting of GR (2009) by setting $h = 1$ in what follows. Define at time $(k + \tau)h$ a surprise loss given by the deviation between the time- $(k + \tau)h$ out-of-sample loss and the average in-sample loss: $SL_{(k+\tau)h}(\hat{\beta}_k) = L_{(k+\tau)h}(\hat{\beta}_k) - \bar{L}_{kh}(\hat{\beta}_k)$, for $k = T_m, \dots, T - \tau$, where $\bar{L}_{kh}(\hat{\beta}_k)$ is the average in-sample loss computed according to the specific forecasting scheme, where $\hat{\beta}_k$ is some estimator of the model parameters and T_m is the in-sample size. One can then define the average of the out-of-sample surprise losses $\overline{SL}_{N_0}(\hat{\beta}_k) = N_0^{-1} \sum_{k=T_m}^{T-\tau} SL_{(k+\tau)h}(\hat{\beta}_k)$, where $N_0 = N - N_{in} - h$ denotes the time span of the out-of-sample window and $N_{in} = T_m h$. GR (2009) observed that under the hypothesis of no forecast instability \overline{SL}_{N_0} should have zero mean (i.e., no systematic surprise losses in the out-of-sample window). Under the null hypothesis of no forecast failure, the GR (2009) test $t_{GR} = N_0^{1/2} \overline{SL}_{N_0}(\hat{\beta}_k) / \hat{\sigma}_{T_m, T_n}$ follows asymptotically a standard normal distribution.

Casini (2017) extends the analysis by considering a continuous-time asymptotic framework and partitioning the out-of-sample into $m_T = \lfloor T_n/n_T \rfloor$ blocks each containing n_T observations. Let $B_{h,b} = n_T^{-1} \sum_{j=1}^{n_T} SL_{T_m+\tau+bn_T+j-1}(\hat{\beta})$ and $\bar{B}_{h,b} = n_T^{-1} \sum_{j=1}^{n_T} L_{(T_m+\tau+bn_T+j-1)h}(\hat{\beta})$ for $b = 0, \dots, \lfloor T_n/n_T \rfloor - 1$ with T_n the out-of-sample size. The test statistic is, $Q_{max,h}(T_n, \tau) =$

$\nu_L^{-1} \max_{b=0, \dots, [T_n/n_T]-2} |B_{h,b+1} - B_{h,b}|$ where ν_L is the square root of the asymptotic variance of the test. The test partitions the out-of-sample window into m_T blocks of asymptotically vanishing length $[bn_T h, (b+1)n_T h]$ and $B_{h,b}$ is a local average of the surprise losses within the block b . The test $Q_{max,h}(T_n, \tau)$ takes on a large value if there is a large deviation $B_{h,b+1} - B_{h,b}$, which suggests a discontinuity or non-smooth shift in the surprise losses close to time $bn_T h$ and thus it provides evidence against the null. Simulations show that the test of GR (2009) and Casini (2017) both have good power properties when the instability is long-lasting while the latter performs better when the instability is short-lived.

Perron and Yamamoto (2017) adapt the classical structural change tests to the forecast failure context. First, they recommend that all tests should be carried with a fixed scheme to have best power, which ensures the maximum difference between the fitted in and out-of sample means of the losses. There are contamination issues under the rolling and recursive scheme that induce power losses. With such a fixed scheme, GR's (2009) test is simply a Wald test for a one-time change in the mean of the total (the in-sample plus out-of-sample) losses at a known break date T_m . To alleviate this problem, which leads to important losses in power when the break in forecasting performance is not exactly at T_m , one can follow Inoue and Rossi (2012) and consider maximizing the GR (2009) test over all possible values of T_m within a pre-specified range. This then corresponds to a sup-Wald test for a single change at some date constrained to be the separation point between the in and out-of-sample periods. The test is still not immune to non-monotonic power problems when multiple changes occur. Hence, Perron and Yamamoto (2017) propose a Double sup-Wald test which for each $T_m \in [T_0, T_1]$ performs a sup-Wald test for a change in the mean of the out-of-sample losses and takes the maximum of such tests over the range $T_m \in [T_0, T_1]$: $DSW = \max_{T_m \in [T_0, T_1]} SW_{L^\circ(T_m)}$, where $SW_{L^\circ(T_m)}$ is the sup-Wald test for a change in the mean of the out-of-sample loss series $L_t^\circ(\hat{\beta})$ for $t = T_m + \tau, \dots, T$, defined by

$$SW_{L^\circ(m)} = \max_{T_b(T_m) \in [T_m + \epsilon T, T_m + (1-\epsilon)T]} [SSR_{L^\circ(m)} - SSR(T_b(T_m))_{L^\circ(T_m)}] / \hat{V}_{L^\circ(T_m)},$$

where $SSR_{L^\circ(T_m)}$ is the unrestricted sum of squares, $SSR(T_b(T_m))_{L^\circ(T_m)}$ is the sum of squared residuals assuming a one-time change at time $T_b(T_m)$, and $\hat{V}_{L^\circ(T_m)}$ is the long-run variance estimate of the out-of-sample loss series. In addition, Perron and Yamamoto (2017) propose to work directly with the total loss series $L(T_m)$ to define the Total Loss Sup-Wald test (TLSW) and the Total Loss UDmax test (TLUD). Using extensive simulations, based on the original design of GR (2009) which involves single and multiple changes in the regression parameters and/or the variance of the errors, they show that with forecasting models poten-

tially involving lagged dependent variables, the only tests having a monotonic power function for all DGPs are the Double sup-Wald and Total Loss UDmax tests, constructed with a fixed forecasting window scheme (Casini's (2017) test was not included in the simulations).

Next, we turn to testing for forecast comparisons in unstable environment. Here, the goal is to determine the relative out-of-sample predictive ability between two competing models in the presence of possible breaks. Giacomini and Rossi (2010) propose two tests: the Fluctuation test and the One-time Reversal test. The former tests whether the local relative forecasting performance equals zero at each point in time whereas the One-time Reversal tests the null hypothesis that the two models perform equally well at each point in time against the alternative that there is a break in the relative performance. Here we discuss the Fluctuation test only. Suppose we compare two τ -step ahead forecast models for the scalar y_k . The first model is characterized by a parameter θ and the second model by a parameter γ . The relative performance is evaluated by a sequence of out-of-sample loss differences $\{\Delta L_k(\hat{\theta}_{k-\tau, T_m}, \hat{\gamma}_{k-\tau, T_m})\}_{k=T_m+\tau}^T$, where $\Delta L_k(\hat{\theta}_{k-\tau, T_m}, \hat{\gamma}_{k-\tau, T_m}) = L^{(1)}(y_k, \hat{\theta}_{k-\tau, T_m}) - L^{(2)}(y_k, \hat{\gamma}_{k-\tau, T_m})$. The expressions for the estimators $\hat{\theta}_{k-\tau, T_m}$ and $\hat{\gamma}_{k-\tau, T_m}$ depend on the forecasting scheme. The Fluctuation test is $\text{Fluct}_{k, T_m}^o = m^{-1} \hat{\sigma}^{-1} \sum_{k=t-m/2}^{k+m/2-1} \Delta L_k(\hat{\theta}_{k-\tau, m}, \hat{\gamma}_{k-\tau, m})$, for $k = T_n + \tau + m/2, \dots, T - m/2 + 1$, where $\hat{\sigma}^2$ is an estimate of the long-run variance of the sequence of out-of-sample losses and m is the size of the window. The asymptotic null distribution is non-standard and critical values are computed by simulations. The test has good finite-sample properties under serially uncorrelated losses when scaled by an estimate of the variance instead of the long-run variance. However, Martins and Perron (2016) show that the test suffers from non-monotonic power when constructed with the long-run variance estimate, as it should be, whether or not the sequence of loss differences exhibit serial correlation. They propose using simple structural change tests such as the sup-Wald test of Andrews (1993) and the UDmax test of Bai and Perron (1998). These are preferred since they have the highest monotonic power even with a long-run variance constructed with a constrained small bandwidth. Finally, Fossati (2017) notices that when the predictive ability is state-dependent (e.g., recessions versus expansions), then taking account of such property by using a test based on Markov Regime-Switching can be a useful alternative.

Additional work and surveys that relate to the issue of testing for structural changes in forecasting and/or forecasting allowing for possible in and out-of-sample changes include, among others, Clements and Hendry (1998a,b, 2006), Pesaran, Pettenuzzo, and Timmermann (2006), Banerjee, Marcellino, and Masten (2008), Rossi (2013a), Giacomini (2015), Giacomini and Rossi (2015) and Xu and Perron (2017).

References

- Altissimo, F., & Corradi, V. (2003). Strong rules for detecting the number of breaks in a time series. *Journal of Econometrics*, 117, 207-244.
- Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61, 821-856 (Corrigendum, 71, 395-397).
- Andrews, D. W. K., & Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62, 1383-1414.
- Antoch, J., Hanousek, J., Horvath, L., Huskova, M. H., & Wang, S. (2017). Structural breaks in panel data: large number of panels and short time series. CEPR Discussion Paper No. 11891.
- Aue, A., Cheung, R. C. Y., Lee, T. C. M., & Zhong, M. (2014). Segmented model selection in quantile regression using the minimum description length principle. *Journal of the American Statistical Association*, 109, 1241-1256.
- Bai, J. (1997a). Estimation of a change point in multiple regression models. *Review of Economic and Statistics*, 79, 551-563.
- Bai, J. (1997b). Estimating multiple breaks one at a time. *Econometric Theory*, 13, 315-352.
- Bai, J. (1999). Likelihood ratio tests for multiple structural changes. *Journal of Econometrics*, 91, 299-323.
- Bai, J. (2000). Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices. *Annals of Economics and Finance*, 1, 303-339.
- Bai, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics*, 157, 78-92.
- Bai, J., & Han, X. (2016). Structural changes in high dimensional factor models. *Frontiers of Economics in China*, 11, 9-39.
- Bai, J., Han, X., & Shi, Y. (2017). Estimation and inference of structural changes in high dimensional factor models. Unpublished Manuscript, Department of Economics, Columbia University.
- Bai, J., Lumsdaine, R. L., Stock, J. H. (1998). Testing for and dating breaks in multivariate time series. *Review of Economic Studies*, 65, 395-432.
- Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66, 47-78.
- Bai, J., & Perron, P. (2003a). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18, 1-22.
- Bai, J., & Perron, P. (2003b). Critical values for multiple structural change tests. *Econometrics Journal*, 6, 72-78.

- Bai, J., & Perron, P. (2006). Multiple structural change models: a simulation analysis. In *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, D. Corbea, S. Durlauf and B.E. Hansen (eds.), Cambridge University Press, 212-237.
- Baltagi, B., Feng, Q., & Kao, C. (2016). Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics*, 191, 176-195.
- Baltagi, B., Kao, C., & Wang, F. (2015). Change point estimation in large heterogeneous panels. Unpublished manuscript, Department of Economics, Syracuse University.
- Baltagi, B., Kao, C., & Wang, F. (2016). Identification and estimation of a large factor model with structural instability. *Journal of Econometrics*, forthcoming.
- Baltagi, B., Kao, C., & Wang, F. (2017). Estimating and testing high dimensional factor models with multiple structural changes. Unpublished manuscript, Department of Economics, Syracuse University.
- Banerjee, A., Marcellino, M., & Masten, I. (2008). Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. In *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, Rapach, D. E., & Wohar, M. E. (Eds.). Emerald Group Publishing Limited, 149-194.
- Bleakley, K., & Vert, J.-P. (2011). The group fused Lasso for multiple change-point detection. arXiv preprint arXiv:1106.4199.
- Boldea, O., Hall, A. R., & Han, S. (2012). Asymptotic distribution theory for break point estimators in models estimated via 2SLS. *Econometric Reviews*, 31, 1-33.
- Breitung, J., & Eickmeier, S., (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, 163, 71-74.
- Casini, A., 2017. Tests for forecast instability and forecast failure under a continuous record asymptotic framework. Unpublished manuscript, Department of Economics, Boston University.
- Casini, A., & Perron, P. (2017a). Continuous record asymptotics for structural change models. Unpublished manuscript, Department of Economics, Boston University.
- Casini, A., & Perron, P. (2017b). Continuous record Laplace-based inference about the break date in structural change models. Unpublished manuscript, Department of Economics, Boston University.
- Casini, A., & Perron, P. (2017c). Generalized Laplace inference in multiple structural change models. Unpublished manuscript, Department of Economics, Boston University.
- Chan, N. H., Yau, C. Y., & Zhang, R.-M. (2014). Group lasso for structural break time series. *Journal of the American Statistical Association*, 109, 590-599.
- Chang, S. Y., & Perron, P. (2016). Inference on a structural break in trend with fractionally integrated errors. *Journal of Time Series Analysis*, 37, 555-574.
- Chang, S. Y., & Perron, P. (2017). A comparison of alternative methods to construct confidence intervals for the estimate of a break date in linear regression models. Forthcoming in *Econometric Reviews*.

- Chen, L. (2015). Estimation the common break date in large factor models. *Economics Letters*, 131, 70-74.
- Chen, L., Dolado, J. J., & Gonzalo, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics*, 180, 3-48.
- Cheng, X., Liao, Z., & Schorfheide, F., (2015). Shrinkage estimation of high-dimensional factor models with structural instabilities. *Review of Economic Studies*, forthcoming.
- Cho, C. K., & Vogelsang, T. J. (2017). Fixed-b inference for testing structural change in a time series regression. *Econometrics*, 5(1), Article 2.
- Chun, S., & Perron, P. (2013). Comparisons of robust tests for shifts in trend with an application to trend deviations of real exchange rates in the long run. *Applied Economics*, 45, 3512-3528.
- Ciuperca, G. (2014). Model selection by LASSO methods in a change-point model. *Statistical Papers*, 55, 349-374.
- Clements, M., & Hendry, D. (1998a). *Forecasting Economic Time Series*. Cambridge University Press.
- Clements, M., & Hendry, D. F. (1998b). Forecasting economic processes. *International Journal of Forecasting*, 14, 111-131.
- Clements, M. & Hendry, D.F. (2006). Forecasting with breaks. In Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Elsevier Science, Amsterdam, 605-657.
- Corradi, V., & Swanson, N.R., (2014). Testing for structural stability of factor augmented forecasting models. *Journal of Econometrics*, 182, 100-118.
- Crainiceanu, C. M., & Vogelsang, T. J. (2007). Nonmonotonic power for tests of a mean shift in a time series. *Journal of Statistical Computation and Simulation*, 77, 457-476.
- Csörgő, M., & Horváth, L. (1997). *Limit Theorems in Change-point Analysis*. Wiley Series in Probability and Statistics, New York: John Wiley.
- Deng, A., & Perron, P. (2006). A comparison of alternative asymptotic frameworks to analyze structural change in a linear time trend. *Econometrics Journal*, 9, 423-447.
- De Wachter, S., & Tzavalis, E. (2012) Detection of structural breaks in linear dynamic panel data models. *Computational Statistics and Data Analysis* 56, 11, 3020-3034.
- Deng, A., & Perron, P. (2008). A non-local perspective on the power properties of the CUSUM and CUSUM of squares tests for structural change. *Journal of Econometrics*, 142, 212-240.
- Elliott, G., & Müller, U. K. (2006). Efficient tests for general persistent time variation in regression coefficients. *Review of Economic Studies*, 73, 907-940.
- Elliott, G., Müller, U. K. (2007). Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics*, 141, 1196-1218.

- Eo, Y., & Morley, J. (2015). Likelihood-ratio-based confidence sets for the timing of structural breaks. *Quantitative Economics*, 6, 463-497.
- Estrada, F., Perron, P., & Martínez-López, B. (2013). Statistically-derived contributions of diverse human influences to 20th century temperature changes. *Nature Geoscience*, 6, 1050-1055.
- Fossati, S. (2017). Testing for state-dependent predictive ability. Unpublished manuscript, Department of Economics, University of Alberta.
- Giacomini, R. (2015). Economic theory and forecasting: lessons from the literature. *The Econometrics Journal*, 18, C22-C41.
- Giacomini, R., & Rossi, B. (2009). Detecting and predicting forecast breakdowns. *The Review of Economic Studies*, 76, 669-705.
- Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25, 595-620.
- Giacomini, R., & Rossi, B. (2015). Forecasting in nonstationary environments: What works and what doesn't in reduced-form and structural models. *Annual Review of Economics*, 7, 207-229.
- Hall, A. R., Han, S., & Boldea, O. (2012). Inference regarding multiple structural changes in linear models with endogenous regressors. *Journal of Econometrics*, 170, 281-302.
- Han, X., & Inoue, A., (2015). Tests for parameter instability in dynamic factor models. *Econometric Theory*, 31, 05, 1117-1152.
- Hansen, B. E. (1992). Tests for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics*, 10, 321-335.
- Hansen, B. E. (2000). Testing for structural change in conditional models. *Journal of Econometrics*, 97, 93-115.
- Harchaoui, Z., & Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association*, 105, 1480-1493.
- Harvey, D. I., Leybourne, S. J., & Taylor, A. M. R. (2007). A simple, robust and powerful test of the trend hypothesis. *Journal of Econometrics*, 141, 1302-30.
- Harvey, D. I., Leybourne, S. J., & Taylor, A. M. R. (2009). Simple, robust, and powerful tests of the breaking trend hypothesis. *Econometric Theory*, 25, 995-1029.
- Hawkins, D. M. (1976). Point estimation of the parameters of piecewise regression models. *Applied Statistics*, 25, 51-57.
- Hidalgo, J., & Seo, M. H. (2013). Testing for structural stability in the whole sample. *Journal of Econometrics*, 175, 84-93.
- Hodrick R. J., & Prescott, E. C., 1997. Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit, and Banking*, 29, 1-16.

- Hou, J., & Perron, P. (2014). Modified local Whittle estimator for long memory processes in the presence of low frequency (and other) contaminations. *Journal of Econometrics*, 182, 309-328.
- Iacone F. (2010). Local Whittle estimation of the memory parameter in presence of deterministic components. *Journal of Time Series Analysis*, 31, 37-49.
- Iacone, F., Leybourne, S. J., & Taylor, A. M. R. (2013a). Testing for a break in trend when the order of integration is unknown. *Journal of Econometrics*, 176, 30-45.
- Iacone, F., Leybourne, S. J., & Taylor, A. M. R. (2013b). On the behavior of fixed-b trend break tests under fractional integration. *Econometric Theory*, 29, 393-418.
- Inoue, A. & Rossi, B. (2012) Out-of-sample forecast test robust to the window size choice *Journal of Business and Economic Statistics* 30, 432-453.
- Juhl, T., & Xiao, Z. (2009). Testing for changing mean with monotonic power. *Journal of Econometrics*, 148, 14-24.
- Kejriwal, M. (2009). Tests for a shift in mean with good size and monotonic power. *Economics Letters*, 102, 78-82.
- Kejriwal, M., & Perron, P. (2008). The limit distribution of the estimates in cointegrated regression models with multiple structural changes. *Journal of Econometrics*, 146, 59-73.
- Kejriwal, M., & Perron, P. (2010a). Testing for multiple structural changes in cointegrated regression models. *Journal of Business and Economic Statistics*, 28, 503-522.
- Kejriwal, M., & Perron, P. (2010b). A sequential procedure to determine the number of breaks in trend with an integrated or stationary noise component. *Journal of Time Series Analysis*, 31, 305-328.
- Kiefer, N. M., & Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21, 1130-1164.
- Kim, D. (2011). Estimating a common deterministic time trend break in large panels with cross sectional dependence. *Journal of Econometrics*, 164, 310-330.
- Kim, D. (2014). Common breaks in time trends for large panel data with a factor structure. *The Econometrics Journal*, 17, 301-337.
- Kim, D., & Perron, P. (2009). Assessing the relative power of structural break tests using a framework based on the approximate Bahadur slope. *Journal of Econometrics*, 149, 26-51.
- Kim, D., Oka, T., Estrada, F., & Perron, P. (2017). Inference related to common breaks in a multivariate system with joined segmented trends with applications to global and hemispheric temperatures. Unpublished manuscript, Department of Economics, Boston University.
- Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.
- Kurozumi, E. (2017). Monitoring parameter constancy with endogenous regressors. *Journal of Time Series Analysis*, 38, 791-805.

- Kurozumi, E., & Tuvaandorj, P. (2011). Model selection criteria in multivariate models with multiple structural changes. *Journal of Econometrics*, 164, 218-238.
- Kurozumi, E., & Yamamoto, Y. (2015). Confidence sets for the break date based on optimal tests. *The Econometrics Journal*, 18, 412-435.
- Lavielle, M., & Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, 21, 33-59.
- Li, D., Qian, J., & Su, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association* 111, 1804-1819.
- Li, Y., & Perron, P. (2017). Inference on locally ordered breaks in multiple regressions. *Econometric Reviews*, 36, 289-353.
- Liang, J., Wang, X., & Yu, J. (2016). New distribution theory for the estimation of structural break point in mean. *Journal of Econometrics*, forthcoming.
- Liu, J., Wu, S., & Zidek, J. V. (1997). On segmented multivariate regressions. *Statistica Sinica*, 7, 497-525.
- Martins, L. P., & Perron, P. (2016). Improved tests for forecast comparisons in the presence of instabilities. *Journal of Time Series Analysis*, 37, 650-659.
- Massacci, D. (2015). Least squares estimation of large dimensional threshold factor models. Unpublished Manuscript.
- McCloskey, A., & Hill, J. B. (2017). Parameter estimation robust to low-frequency contamination. *Journal of Business & Economic Statistics*, 35, 598-610.
- McCloskey, A., & Perron, P. (2013). Memory parameter estimation in the presence of level shifts and deterministic trends. *Econometric Theory*, 29, 1196-1237.
- Nyblom, J. (1989). Testing for constancy of parameters over time. *Journal of the American Statistical Association*, 84, 223-230.
- Oka, T., & Perron, P. (2017). Testing for common breaks in a multiple equations system. Unpublished manuscript, Department of Economics, Boston University.
- Oka, T., & Qu, Z. (2011). Estimating structural changes in regression quantiles. *Journal of Econometrics*, 162, 248-267.
- Perron, P. (1991). A test for changes in a polynomial trend function for a dynamic time series. Unpublished manuscript, Université de Montréal, Centre de recherche et développement en économique.
- Perron, P. (1997). L'estimation de modèles avec changements structurels multiples. *Actualité Économique*, 73, 457-505.
- Perron, P. (2006). Dealing with structural breaks. In *Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory*, K. Patterson and T.C. Mills (eds.), Palgrave Macmillan, 278-352.

- Perron, P. (2008). Structural change. In *The New Palgrave Dictionary of Economics*, 2nd ed, S. Durlauf and L. Blume (eds.).
- Perron, P., & Qu, Z. (2006). Estimating restricted structural change models. *Journal of Econometrics*, 134, 373-399.
- Perron P., & Qu, Z. (2007). An analytical evaluation of the log-periodogram estimate in the presence of level shifts. Unpublished Manuscript, Boston University.
- Perron P., & Qu, Z. (2010). Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business and Economic Statistics*, 28, 275-290.
- Perron, P., & Yabu, T. (2009a). Estimating deterministic trends with an integrated or stationary noise component. *Journal of Econometrics*, 151, 56-69.
- Perron, P., & Yabu, T. (2009b). Testing for shifts in trend with an integrated or stationary noise component. *Journal of Business and Economic Statistics*, 27, 369-396.
- Perron, P., & Yamamoto, Y. (2013). Estimating and testing multiple structural changes in linear models using band spectral regressions. *Econometrics Journal*, 16, 400-429.
- Perron, P., & Yamamoto, Y. (2014). A note on estimating and testing for multiple structural changes in models with endogenous regressors via 2SLS. *Econometric Theory*, 30, 491-507.
- Perron, P., & Yamamoto, Y. (2015). Using OLS to estimate and test for structural changes in models with endogenous regressors. *Journal of Applied Econometrics*, 30, 119-144.
- Perron, P., & Yamamoto, Y. (2016). On the usefulness or lack thereof of optimality criteria for structural change tests. *Econometric Reviews*, 35, 782-844.
- Perron, P., & Yamamoto, Y. (2017). Testing for changes in forecasting performance. Unpublished manuscript, Department of Economics, Boston University.
- Perron, P., & Zhou, J. (2008). Testing jointly for structural changes in the error variance and coefficients of a linear regression model. Unpublished manuscript, Department of Economics, Boston University.
- Perron, P., & Zhu, X. (2005). Structural breaks with stochastic and deterministic trends. *Journal of Econometrics*, 129, 65-119.
- Pesaran, M. H., Pettenuzzo, D., & Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies*, 73, 1057-1084.
- Pitarakis, J.-Y. (2004). Least squares estimation and tests of breaks in mean and variance under misspecification. *Econometrics Journal*, 7, 32-54.
- Prodan, R. (2008). Potential pitfalls in determining multiple structural changes with an application to purchasing power parity. *Journal of Business & Economic Statistics*, 26, 50-65.
- Qian, J., & Su, L. (2016). Shrinkage estimation of common breaks in panel data models via adaptive group Lasso. *Journal of Econometrics*, 191, 86-109.

- Qu, Z. (2008). Testing for structural change in regression quantiles. *Journal of Econometrics*, 146, 170-184.
- Qu Z. (2011). A test against spurious long memory. *Journal of Business and Economic Statistics*, 29, 423-438.
- Qu, Z., & Perron, P. (2007). Estimating and testing multiple structural changes in multivariate regressions. *Econometrica*, 75, 459-502.
- Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, 55, 324-330.
- Rojas, C. R., & Wahlberg, B. (2014). On change point detection using the fused lasso method. arXiv preprint arXiv:1401.5408.
- Rossi, B. (2013). Advances in forecasting under instability. In G. Elliott, & A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Ch. 21, Elsevier Science, Amsterdam, 1203-1324.
- Sayginsoy, Ö., & Vogelsang, T. J. (2011). Testing for a shift in trend at an unknown date: a fixed-b analysis of heteroskedasticity autocorrelation robust OLS-based tests. *Econometric Theory*, 27, 992-1025.
- Siegmund, D. (1988). Confidence sets in change-point problems. *International Statistical Review* 56, 31-48.
- Su, L., & Wang, X., (2017). On time-varying factor models: estimation and inference. *Journal of Econometrics* 198, 84-101.
- Vogelsang, T. J. (1997). Wald-type tests for detecting breaks in the trend function of a dynamic time series. *Econometric Theory*, 13, 818-849.
- Vogelsang, T. J. (1999). Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *Journal of Econometrics*, 88, 283-299.
- Vogelsang, T. J. (2001). Testing for a shift in trend when serial correlation is of unknown form. Unpublished manuscript. Department of Economics, Cornell University.
- Xu, J., & Perron, P. (2017). Forecasting in the presence of in and out of sample breaks. Unpublished Manuscript, Department of Economics, Boston University.
- Yamamoto, Y. (2016). A modified confidence set for the structural break date in linear regression models. Forthcoming in *Econometric Reviews*.
- Yamamoto, Y., & Tanaka, S. (2015). Testing for factor loading structural change under common breaks. *Journal of Econometrics*, 189, 187-206.
- Yang, J. (2017). Consistency of trend break point estimator with underspecified break number. *Econometrics* 5 (4), 1-19.
- Yang, J., & Vogelsang, T. J. (2011). Fixed-b analysis of LM-type tests for a shift in mean. *Econometrics Journal*, 14, 438-456.