Testing for Parameter Instability across Different Modeling Frameworks*

Francesco Calvori[a], Drew Creal[b], Siem Jan Koopman[c,d], André Lucas[c]

(a) *R&D, UnipolSai Assicurazioni, Bologna, Italy*

(b) *Booth School of Business, University of Chicago, USA*

(c) *Vrije Universiteit Amsterdam and Tinbergen Institute, The Netherlands*

(d) *CREATES, Aarhus University, Denmark*

Abstract

We develop a new parameter instability test that generalizes the seminal ARCH-*LM* test of Engle (1982) for a constant variance against the alternative of autoregressive conditional heteroskedasticity to settings with nonlinear time-varying parameters and non-Gaussian distributions. We investigate the performance of the new test relative to both classic and recently proposed parameter instability tests, including tests against structural breaks and parameter-driven dynamics. We find that the recent test of Müller and Petalas (2010) performs best across a wide range of alternatives, particularly if parameter instability is slow. For time-varying parameters that exhibit more mean reversion, our new test has higher power. We provide an application to a heavily unbalanced panel of losses given default for U.S. corporations from 1982 to 2010 and provide evidence of significant parameter instability in the parameters of a static beta distributed model.

Key words: time-varying parameters; observation-driven and parameter-driven models; structural breaks; generalized autoregressive score model; regime switching; credit risk.
JEL classifications: C12, C52, C22.

*We thank Bob Brugman for excellent research assistance. We have benefited from the insightful comments of participants at the Workshop on Score-Driven Models in Amsterdam. Creal thanks the William Ladany Faculty Scholar Fund at Booth School of Business for financial support. Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. Lucas thanks the Dutch National Science Foundation (NWO Grant VICI453-09-005) for financial support.

Email addresses: calvori@disia.unifi.it, drew.creal@chicagobooth.edu, s.j.koopman@vu.nl, a.lucas@vu.nl.

1 Introduction

A key concern in the empirical modeling of financial time series is parameter instability. Hansen (2001) provides an overview of a large number of different parameter instability tests, including standard tests such as the Chow (1960) break test, the supremum F-tests of Andrews (1993), and the weighted F-tests by Andrews and Ploberger (1994). Many more tests have been developed since then. When testing for parameter instability, we can consider different model specifications under the alternative. For example, there might be one or more deterministic structural breaks in the parameters as in, for example, Vogelsang and Perron (1998), Bai and Perron (2003), Perron (2006), and Qu and Perron (2007); the parameters may exhibit regular regime switches as in Hamilton (1989) or random regime switches as in Chib (1998) and Pesaran et al. (2006); or the parameters may evolve continuously over time, either in a parameter-driven (state space) framework such as Harvey (1989), Bauwens and Veredas (2004), Shephard (2005), Hafner and Manner (2012), and Durbin and Koopman (2012), or an observation-driven framework such as Engle (1982), Bollerslev (1986), Engle and Russell (1998), Davis et al. (2003), Patton (2006), Creal et al. (2013), and Harvey (2013).

The aim of this research project is twofold. First, we develop a new test for parameter instability in nonlinear and non-Gaussian models against the alternative of the generalized autoregressive score (GAS) dynamics of Creal et al. (2013). Score-driven models provide a flexible framework for the introduction of time-varying parameters and include many well-known cases, such as the generalized autoregressive conditional heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1986), the autoregressive conditional duration (ACD) model of Engle and Russell (1998), and the dynamic Poisson model of Davis et al. (2003). A test against the observation-driven score-based alternative might thus complement powerful tests against other alternatives, such as parameter-driven alternatives as in the testing approach of Müller and Petalas (2010). Second, we perform a full-scale investigation of the finite sample properties of the new test against those of the point optimal test of Müller and Petalas (2010). Their test sets a strong benchmark compared to more classical tests such as the sup-LM test of Andrews (1993) against structural breaks, or the Nyblom (1989) test against the alternative of random walk parameters.

Score-driven time varying parameter models are characterized by a parametric conditional observation density. Since these models are observation-driven, their likelihood function is available in closed-form, and parameter estimation and inference are straightforward. A variety of empirical studies have illustrated that the score-driven approach leads to a range of interesting, new empirical models. Examples are given by Creal et al. (2011) and Lucas et al. (2014) for multivariate volatility and correlation models, Creal et al. (2014) for mixed measurement factor models, Harvey and Luati (2014) for location and scale models, and De Lira Salvatierra and Patton (2015) for copula models.[1] Interestingly, score-driven models have information theoretic optimality properties; see Blasques et al. (2015). Moreover, score-driven models have good out-of-sample forecasting accuracy compared to nonlinear non-Gaussian (parameter-driven) state space models, even in cases where the latter are correctly specified; see Koopman et al. (2015). A parameter instability test against the alternative of score-driven dynamics can thus provide a useful signal that a static model is too simplistic and needs to be augmented. In particular, given that forecasting performances of score-driven and parameter-driven models are similar, a test against a score-driven alternative might have similarly strong properties as the test of Müller and Petalas (2010), which centers around a parameter-driven alternative.

We consider linear as well as nonlinear models where computational efficiency becomes a concern. The estimation of parameters in an extensively nonlinear model is typically cumbersome and computationally demanding. Hence we favor the use of Lagrange Multiplier (LM) based test statistics. Our new LM test is highly intuitive and practical. It tests for non-zero serial dependence in the score function of ℓ_t, where ℓ_t is the tth contribution to the log-likelihood function $\sum_{t=1}^{T} \ell_t$ in a static model for a time series of length T. We view our test as an omnibus diagnostic tool and as a generalization of the familiar ARCH-LM test of Engle (1982) to settings beyond time-varying volatility. The asymptotic distribution of the test is standard and follows by familiar results from White (1987). Similar to most omnibus LM diagnostic tests, our test can easily be computed by means of an auxiliary regression.

In our simulations, we also extend the testing approach of Müller and Petalas (2010) to include a data-driven choice of the key tuning parameter in the test.[2] Originally, Müller and Petalas (2010) introduced their the test with a fixed tuning parameter. However, depending on the empirical setting and depending on the deviation from the null of constant parameters, a fixed tuning parameter becomes inconvenient. We use their suggested likelihood-based smoothing weights as a means to determine the data-driven tuning parameter. In particular, we use the tuning constant that receives highest weight in their procedure to estimate the time-varying parameter path from the data.

We obtain several interesting insights. In cases where parameters change infrequently or slowly, the Müller and Petalas (2010) test is the best. The optimality of the Müller-Petalas test for time-variation close to the null of constant parameters follows immediately from their analytical results. If, however, parameter-instabilty is less sticky and displays more mean

[1] An extensive compilation of papers on score-driven models is presented at `http://www.gasmodel.com`.

[2] We thank a referee for this suggestion.

reverting behavior, the new test against a score-driven alternative performs best. This holds particularly if the signal-to-noise ratio is small, such as for instance for financial return data with a possibly time-varying expected return. The power performance of the new test is surprisingly robust over alternative specifications of the data generating process. This is less so for a number of the competing test procedures. For both the Müller-Petalas test and the new test, we find that data-driven choices of tuning parameters work well.

We can explain the different behavior of the Müller-Petalas and the new test by looking at how the tests are constructed. The score-based LM test relies directly on the serial dependence in the scores of the log-likelihood from the static model. These serial correlations can typically be estimated accurately even when the data generating process of the time-varying parameter is erratic and moves quickly. Hence we find a good performance of our test even in settings with many breaks or with strong mean reverting parameter dynamics. The test of Müller and Petalas (2010), by contrast, uses the unconditional volatility of the path of the time-varying parameter as its main ingredient. This path is estimated under the assumption of a highly persistent parameter process. When the true time-varying parameter moves quickly, the estimated path typically becomes almost constant, thus reducing the estimated volatility of the path and the power of the test. The problem is alleviated by a data-driven choice of the test's tuning constant, but remains present, particularly if the signal-to-noise ratio is small as the approach tries to balance the stickiness of the parameter variation under the alternative and the signal-to-noise ratio with just one tuning constant. The phenomenon of a bad estimate of the path of the time-varying parameter also applies to the score-based LM test, but has less of an effect since it is not based on the volatility of the estimated path.

In our empirical illustration, we test for the instability of parameters in loss given default (LGD) models that are used for credit risk analyses. Loss given default is the fraction of the outstanding amount of a loan or bond that is lost if the company defaults. It is a key ingredient of current models for financial risk management and regulation. Many financial industry credit risk models for LGDs use static parameters. A prime example is the use of a static beta distribution for modeling LGD fractions. The use of static parameters in this application may be inappropriate since financial conditions vary over time and affect the LGDs accordingly. For example, losses may be higher on average in situations where default risk is also higher, thus exacerbating the total loss experience. If the parameters of a model for LGDs are actually time-varying, the regulator may require higher capital requirements for financial institutions to mitigate financial stability concerns.

We analyse a panel data set of LGDs for corporate bond data obtained from Moody's to test for the presence of time-varying parameters. The data set is non-standard and provides an interesting example of the flexibility of our testing approach. The number of LGD observations for each quarter varies over time because LGDs can only be observed when a default occurs, and the number of defaults evidently varies over time. Assuming that the LGDs are drawn from a beta distribution with possibly time-varying parameters, all tests strongly confirm that the distributional properties of LGDs vary over time. In particular, we find that LGDs have been on average very low compared to the static model during the period leading up to the 2008 financial crisis. It suggests that the abundance of liquidity during this period has not only prevented firms from defaulting, but has also mitigated the losses for those cases in which a default was unavoidable.

The remainder of this paper is organized as follows. In Section 2 we describe our new test statistic as well as the main alternative tests from the literature. In Section 3 we provide simulation results. Section 4 discusses the empirical application. Section 5 concludes. The online Appendix presents additional simulation results.

2 Testing frameworks for time-varying parameters

We consider a dependent variable $y_t \in \mathbb{R}^m$ for $t = 1, \ldots, T$, where T denotes the sample size; a vector of time-varying parameters $f_t \in F \subset \mathbb{R}^k$; and a vector of static parameters $\delta \in D \subset \mathbb{R}^n$, where F and D denote the parameter space of the time-varying and static parameter vectors, respectively. The dimensions m, k and n are positive integers; for a univariate model with a single time-varying parameter and four static parameters, we have $m = k = 1$ and $n = 4$.

2.1 Observation-driven time-variation

In an observation-driven framework, the time-varying parameter f_t is driven by a deterministic function of lagged dependent variables and contemporaneous or lagged exogenous variables. The observation-driven modeling framework has the advantage that the likelihood is available in closed-form and can easily be evaluated. It leads to estimation and inference procedures that can be easily implemented. The main challenge however is to determine the function of the observations that drives the parameter f_t through time. A general approach encompassing many popular nonlinear and non-Gaussian dynamic models is the generalized autoregressive score (GAS) model of Creal et al. (2013); see also Harvey (2013). In the GAS(p, q) model, the observations y_t have the dynamic specification

$$y_t \sim p(y_t | f_t;\ \delta) \tag{1}$$

$$f_{t+1} = (\mathrm{I} - B_1 - \ldots - B_p)\omega + \sum_{i=1}^{q} A_i s_{t-i+1} + \sum_{j=1}^{p} B_j f_{t-j+1}, \tag{2}$$

where the elements of the vector ω and of the matrices A_i and B_j are static parameters for $i = 1, \ldots, p$ and $j = 1, \ldots, q$, with

$$s_t := S_t \cdot \nabla_{f,t}, \qquad \nabla_{f,t} := \frac{\partial \ln p(y_t | f_t;\ \delta)}{\partial f_t}, \tag{3}$$

where $\nabla_{f,t}$ is the score of the conditional observation density with respect to the time-varying parameter f_t. The $k \times k$ matrix $S_t = S(f_t; \delta)$ is the scaling matrix for the score. For example, we can consider a power of the Fisher information matrix of the conditional observation density to account for the curvature of the score; see Creal et al. (2013) for more details.

The defining feature of the GAS model is its use of the score of the conditional observation density to drive the parameter f_t through time. At each time period, the dynamics of the time-varying parameter can be interpreted as a steepest-ascent or Gauss-Newton step, where the local fit of the model is improved by using the information in the most recent observation and its distribution. Such steps locally improve the Kullback-Leibler divergence of the model even in cases where the model is mis-specified; see Blasques et al. (2015). The GAS framework encompasses as special cases the GARCH model of Engle (1982) and Bollerslev (1986), the ACD and ACI models of Engle and Russell (1998) and Russell (2001), the MEM model of Engle and Gallo (2006) and Cipollini et al. (2013), the models for Poisson counts of Davis et al. (2003), and the Beta-t-GARCH model of Harvey (2013), among many others.

To introduce the GAS-*LM* test, we draw the analogue with the ARCH-*LM* test of Engle (1982) or the GARCH-*LM* test of Lee (1991). The ARCH(1)-*LM* test of Engle for the model $y_t = x_t'\beta + \sigma_t \varepsilon_t$ where ε_t has mean zero and variance one, tests the null of a constant variance against the alternative

$$\sigma^2_{t+1} = (1 - A)\omega + A\varepsilon_t^2 = (1 - A)\omega + A(\varepsilon_t^2 - \sigma_t^2) + A\sigma_t^2, \tag{4}$$

with parameter $|A| < 1$, for $t = 1, \ldots, T$. Under the alternative hypothesis, the variance σ^2_{t+1} varies around the static level ω as driven by the scaled score $\varepsilon_t^2 - \sigma_t^2$ of a Gaussian density with respect to the parameter σ_t^2. The null hypothesis is $\mathrm{H}_0 :\ A = 0$.

The *LM* test against a GAS alternative takes the same perspective as the ARCH-*LM* test (4), but generalizes the observation density and allows the time-varying parameter f_t to characterize a different distributional property than the variance. Formally, we test the null hypothesis of no parameter variation against the GAS alternative

$$f_{t+1} = (\mathrm{I} - A_1 - \ldots - A_q)\omega + \sum_{i=1}^{q} A_i s_{t-i+1} + \sum_{i=1}^{q} A_i f_{t-i+1}, \tag{5}$$

where the dynamics of f_t are driven by the scaled scores s_t from (3). Similar to (4), under the alternative the time-varying parameter f_t varies around its static level ω. We can use the same arguments as in Lee (1991) to allow for different coefficients B_i (rather than A_i only) for the lags of f_{t-i+1} under the alternative, for $i = 1, \ldots, q$.

To define the LM test statistic, let $\ell_t(\delta, \omega, a) = \ln p(y_t | f_t;\ \delta)$ be the contribution at time t of the log-likelihood function, where $a = \mathrm{vec}(A_1, \ldots, A_q)$. We suppress the dependence of f_t on the static parameters δ, ω. Define $\bar{s}_{p,t} = \mathrm{vec}(s_t, \ldots, s_{t-p+1})$ and let $\mathcal{G}_t' = (\nabla_{\delta,t}', \nabla_{\omega,t}', \nabla_{\omega,t}' \otimes \bar{s}_{p,t-1}')$ where $\otimes$ is the Kronecker product and with $\nabla_{\delta,t}$ and $\nabla_{\omega,t}$ denoting the derivatives of ℓ_t with respect to δ and ω, respectively. Following White (1987), the LM test for $\mathrm{H}_0 : a = 0$ versus the alternative $\mathrm{H}_1 : a \neq 0$, is given by

$$LM = \mathcal{G}'\mathcal{H}^{-1}\mathcal{G}, \quad \mathcal{G} = \sum_{t=1}^{T} \mathcal{G}_t, \quad \mathcal{H} = \sum_{t=1}^{T} \mathcal{G}_t \mathcal{G}_t', \tag{6}$$

where derivatives are evaluated at the maximum likelihood estimates under the null hypothesis. The covariance matrix $\mathcal{H}$ can be replaced by a robust long-run covariance matrix, that is

$$\tilde{\mathcal{H}} = \sum_{t=1}^{T} \sum_{\tau=1}^{t} w_{T,t-\tau} \left(\mathcal{G}_t \mathcal{G}_\tau' + \mathcal{G}_\tau \mathcal{G}_t' \right),$$

for some kernel weights $w_{T,t-\tau}$; see Andrews (1991).

Under standard regularity conditions, the GAS-LM test converges under the null to a χ^2 distributed random variable with $\dim(a)$ degrees of freedom; see White (1987). An appealing feature of the new test is thus that its asymptotic statistical theory is entirely standard. This contrasts with the asymptotic behavior of some of the other tests that are considered later.

Following Davidson and MacKinnon (1990), the LM test statistic can be written as the explained sum of squares of the auxiliary least squares regression

$$\begin{aligned} 1 &= \mathcal{G}_t' \beta_{LM} + \text{residual} \\ &= \nabla_{\delta,t}' \beta_{LM}^{\delta} + \nabla_{\omega,t}' \beta_{LM}^{\omega} + (\nabla_{\omega,t} \otimes \bar{s}_{p,t-1})' \beta_{LM}^{a} + \text{residual}, \end{aligned} \tag{7}$$

where $\beta_{LM}' = (\beta_{LM}^{\delta\prime}, \beta_{LM}^{\omega\prime}, \beta_{LM}^{a\prime})$ is a vector of regression parameters and all derivatives in $\mathcal{G}_t$ are evaluated under the null. The regression interpretation of the GAS test makes it easy to compute in standard statistical software. The first derivatives of the conditional observation density at each time t can be obtained either analytically or numerically.

The GAS-LM test has an intuitive interpretation. The key term on the right-hand side of the auxiliary regression (7) is $\nabla_{\omega,t} \otimes \bar{s}_{p,t-1}$. The elements of this vector are $\mathrm{vec}(S_{t-i} \nabla_{\omega,t-i} \nabla_{\omega,t}')$, for $i = 1, \ldots, q$, because, under the null, the score of the conditional density with respect to f_t is equal to the score with respect to ω. Hence, the LM test against the GAS alternative verifies whether there is any serial dependence in the scores $\nabla_{f,t}$ of the *static* model. If serial dependence exists, the actual autocorrelations in the likelihood scores can be exploited to improve the fit of the model by using the scores to drive the time-varying parameter f_t. This is exactly what the dynamics of the GAS model in (2) achieve.

Although the above LM test has been derived with the GAS alternative in mind, we expect it to have power against other forms of parameter instability as well. The test can therefore be regarded as a generic portmanteau test against model misspecification. The same holds for tests against structural break alternatives and against parameter-driven time variation. Next we discuss these more general applications of our test.

2.2 Parameter-driven time-variation

For parameter-driven models, the time-varying parameter f_t is a stochastic process that is subject to its own source of error. Important examples of this class of models are unobserved components time series models as in Harvey (1989), stochastic volatility models as reviewed in Shephard (2005), stochastic conditional duration models as in Bauwens and Veredas (2004), and stochastic copula models as in Hafner and Manner (2012) and Creal and Tsay (2015). The randomness in both f_t and y_t lead to challenging problems in parameter estimation and testing. The likelihood function is not available in closed form except in cases such as linear Gaussian state space models and discrete-state hidden Markov models; see Durbin and Koopman (2012) and Hamilton (1989), respectively. In all other cases likelihood-based inference, including parameter estimation, signal extraction and testing, requires the use of approximation and/or simulation methods; see, for example, Creal (2012) and Durbin and Koopman (2012) for general discussions on these methods.

The research conducted by Müller and Petalas (2010), which we denote as MP(10) hereafter, provide an elegant and general framework for testing parameter instability. Their approach encompasses nonlinear and non-Gaussian models with moderately time-varying parameters. If the time variation vanishes asymptotically at the appropriate rate, MP(10) show that we can address the inference problem of parameter instability by considering a linear Gaussian state space model where the observations are replaced by the likelihood scores of the static model at time t. Moreover, they prove that such an approach is not only asymptotically optimal against the alternative of (local) parameter-driven time-varying parameters, but it is also optimal against a much wider range of alternative (local) parameter dynamics. The MP(10) test stands in a long tradition of point optimal tests against local alternatives, such as random walk parameters; see, for example, the contributions by Nyblom and Mäkeläinen (1983), Franzini and Harvey (1983), King and Hillier (1985), Nyblom (1989), and Elliott and Müller (2006).

The key intuition of the MP(10) test comes from a (pseudo-)linear Gaussian state space model with observation and transition equations given by

$$\begin{aligned} \mathcal{H}\mathcal{V}^{-1}\nabla_{\omega,t} &= \mathcal{S}^{-1}(f_t - \bar{f}) + \nu_t, & \nu_t &\sim \mathrm{N}(0, \mathcal{S}^{-1}) \\ (f_{t+1} - \bar{f}) &= (1 - cT^{-1})(f_t - \bar{f}) + \tilde{\nu}_t, & \tilde{\nu}_t &\sim \mathrm{N}(0, c^2T^{-2}\mathcal{S}), \end{aligned} \tag{8}$$

where

$$\mathcal{H} = T^{-1}\sum_{t=1}^{T} \partial^2 \ln p(y_t|\bar{f};\ \delta)/\partial f_t^2, \qquad \mathcal{V} = T^{-1}\sum_{t=1}^{T} \nabla_{\omega,t}\nabla_{\omega,t}', \qquad \mathcal{S} = \mathcal{H}^{-1}\mathcal{V}\mathcal{H}^{-1}.$$

The log-likelihood score $\nabla_{\omega,t}$ is the same as for the observation-driven model and the state variable f_t follows a nearly-integrated process with fixed tuning parameter c. The parameter $\bar{f}$ is a fixed benchmark level for f_t, for $t = 1, \dots, T$. It follows that f_t is a persistent process with local time variation. As the sample size grows, time variation in f_t vanishes as the autoregressive parameter converges to unity and the variance of the transition equation in (8) converges to zero.

The variable $\mathcal{H}\mathcal{V}^{-1}\nabla_{\omega,t}$ can be viewed as a pseudo-observation. Its linear Gaussian state space model (8) is the result of applying Laplace transformations to nonlinear and non-Gaussian models. A similar technique is used for the likelihood-based approaches of Shephard and Pitt (1997) and Durbin and Koopman (1997, 2000) where they adopt an approximating linear, Gaussian model to implement importance sampling or MCMC methods. Such simulation-based methods for the estimation of parameters in nonlinear non-Gaussian state space models are used extensively in econometrics. MP(10) point out that the key difference between the approximating model used by Shephard and Pitt (1997) and Durbin and Koopman (1997, 2000) and their approximating model is the use of the global Hessian $\mathcal{H}$ rather than the local Hessian of the conditional observation density at time t.

Müller and Petalas (2010) construct a point optimal (likelihood ratio) test of the null $c = 0$ versus the alternative $c = 10$. We discuss the impact of alternative values of c below. Although the theory in MP(10) is somewhat advanced, the proposed test statistic is surprisingly straightforward to compute using regression methods. An algorithm is provided in their paper. The point optimality of the test allows a likelihood ratio test interpretation. As a result, the test has a power advantage compared to an LM test. This stems from the fact that we actually obtain an approximate fit of the model under the (local) parameter-driven alternative $c = 10$ based on the regressions used to compute the test statistic. Consequently, the test captures part of the gain of the likelihood ratio compared to the Lagrange multiplier test, just as if parameters of a model would have been estimated under both the null and the alternative rather than under the null only.

Based on similar regressions as those used to obtain the test statistic, MP(10) also propose an estimator for the path of the time-varying parameter f_t. Their estimator is a weighted average risk based combination of the estimated paths for different values of c, namely $c = 0, 5, 10, \ldots, 50$. We consider this estimator in our empirical application.

The crucial ingredient of the MP(10) test is the sum of $(f_t - \bar{f}) \cdot \nabla_{\omega,t}$. The test uses the variability of the estimated path f_t around its static counterpart $\bar{f}$ for $c = 10$. The differences $(f_t - \bar{f})$ are weighted by the likelihood score with respect to the, possibly, time-varying parameter. If the estimated path f_t is relatively constant, or if the likelihood is not very sensitive with respect to f_t, the resulting test statistic is small. The smoothed estimate of the path for the test is obtained under $c = 10$, which implies a high degree of persistence for sufficiently large T. This can become problematic if there is rapid time variation in f_t under the alternative, such as in the case of regular regime switches or strongly mean reverting parameter changes. In these cases, the estimated path of f_t can become close to a constant, resulting in a low value of the test statistic and a low power of the corresponding testing procedure.

Compared to the GAS-LM test of Section 2.1, the MP(10) test has three main differences. First, due to the choice of $c = 10$ and the structure of the auxiliary regressions, the MP(10) test statistic weighs both present and future autocovariances of the score. By contrast, the GAS-LM test only uses past autocovariances. Second, the GAS-LM test allows the user to include an explicit number of autocovariances through the choice of the parameter q. The MP(10) test, by contrast, takes all autocovariances into account, but implicitly defines their weight through the choice of the tuning parameter $c = 10$. Third, the distributions of the GAS and MP(10) tests under the null differ profoundly. The GAS-LM test follows the standard χ^2 asymptotics of White (1987) while the MP(10) test follows the asymptotic distribution as derived in Elliott and Müller (2006).

Müller and Petalas (2010) recommended choosing a value of $c = 10$ under the time-varying parameter alternative when calculating their test. Their primary applications were quarterly or monthly macroeconomic time series with T ranging between 100 to 450 implying a range of $(0.9, 0.978)$ for the autoregressive parameter $(1 - cT^{-1})$ in (8). For applications with larger sample sizes, alternative values of c may perform better than $c = 10$. In our Monte Carlo study and in the empirical application, we implement their test for larger values of c, including a simple data-based procedure. Specifically, we calculate their test as well as the weights used for their path estimator for $c = 0, 5, 10, \ldots, 500$. We then choose the test statistic associated with the largest weight. We label this test MP($*$).

2.3 Structural breaks

Andrews (1993) proposes a general parameter instability test for nonlinear parametric models against alternatives with a one-time break in (a subset of) the parameters. Generalizations to multiple breaks are possible, but are typically computer intensive unless the structure of the model is sufficiently simple; see Bai and Perron (2003). The tests against a structural change alternative are based on partial-sample GMM estimators and can be of the supremum Wald, Lagrange multiplier (LM), and likelihood ratio types. Modifications of these tests that use weighted averages rather than the supremum of the tests over all possible break points are proposed by Ploberger et al. (1989) and Andrews and Ploberger (1994). Here we focus on the LM based version of the test. This precludes the need to estimate a possibly nonlinear model over many different subsamples. It can lead to a time-consuming process with many computations, in particular during the exploratory modeling phase.

Let $\pi \subset (0,1)$ and let $\lfloor \pi T \rfloor + 1$ denote the breakpoint of the parameter f_t, where $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$. The null and alternative hypotheses for the Andrews' sup-LM test are given by

$$\mathrm{H}_0 \quad : \quad f_t = \bar{f}_0 \;\; \forall\, t \geq 1 \text{ and some } \bar{f}_0 \in F \subset \mathbb{R}^k, \tag{9}$$

$$\mathrm{H}_1 \quad : \quad \bigcup_{\pi \in \Pi} \mathrm{H}_{1,T}(\pi) \text{ for some } \Pi \subset (0,1), \tag{10}$$

$$\mathrm{H}_{1,T}(\pi) \quad : \quad f_t = \begin{cases} \bar{f}_1(\pi), & \text{for } t = 1, \ldots, \lfloor \pi T \rfloor, \\ \bar{f}_2(\pi), & \text{for } t = \lfloor \pi T \rfloor + 1, \ldots, T, \end{cases} \tag{11}$$

for constants $\bar{f}_1(\pi), \bar{f}_2(\pi) \in F$. The test is designed for a single break at an unknown date. However, the test also has good power properties against a range of more general alternatives; see, for example, the survey of Hansen (2001). The distribution of the Andrews' sup-LM test is the supremum of the square of a tied down Bessel process as derived in Theorem 3 of Andrews (1993), where one can also find the critical values of the test.

In contrast to the tests described in Sections 2.1 and 2.2, the Andrews' sup-LM test does not build on the autocorrelations of the score of the likelihood, but rather on the average level of the score before and after the break. The crucial ingredients of the test are scaled versions of $\sum_{t=1}^{\lfloor \pi T \rfloor} \nabla_{\omega,t}$ and $\sum_{t=\lfloor \pi T \rfloor + 1}^{T} \nabla_{\omega,t}$. This follows from the alternative, which is a structural break at an unknown time point. When regular switches between alternative values of the parameter occur, the Andrews' test may have difficulty in identifying such a pattern. The sample means of moment conditions before and after any particular tentative breakpoint may not be sufficiently different in small sample sizes. We expect to observe this problem when the true time-varying parameters are subject to regular regime switches or to strongly mean reverting dynamics.

2.4 Martingale type time–variation

Our final benchmark is the parameter instability test of Nyblom (1989). This "all-purpose" test is based on the assumption that under the alternative the time-varying parameter follows a martingale process. Nyblom (1989) argues that his test encompasses the case of one or more structural breaks. The Nyblom test is therefore related to both LM tests of Sections 2.2 and 2.3.

The key ingredient of the Nyblom test is the partial sum of the likelihood scores. Hence the test is close to the partial sums in the Andrews' test. However, the Nyblom test does not take a supremum, it relies on the average of the squares of partial sums. We therefore expect that the Nyblom test has an inferior performance compared to the other three tests in most settings.

3 Monte Carlo study

3.1 Design of study

In this section we report a subselection of results from a larger Monte Carlo simulation study. More results can be found in the working paper version and the web appendix of this paper. We consider a range of different data generating processes (DGPs). For each DGP, we perform $N = 10,000$ replications by generating time series of length $T = 200, 500, 1000$, or 2000 with different forms of time-varying parameters. Next, for each time series, we compute the GAS-$LM(1)$ test, the test of Müller and Petalas (2010), the sup-LM test of Andrews (1993), and the test of Nyblom (1989). Müller and Petalas advise a tuning constant $c = 10$ for their test. We call this the MP(10) test. We also use two data-driven ways to set the tuning parameters of the GAS-LM test and the MP test. For the GAS-LM test, we use the procedure of Escanciano and Lobato (2009) to select an optimal number of lags. We denote this as the GAS-$LM(*)$ test. For the MP test, we use the procedure as described at the end of Section 2.2 to select the tuning constant in a data-driven way. This is denoted as the MP($*$) test. For the implementation of the Andrews sup-LM test, we use common cut-off values and consider all possible breakpoints between 15% and 85% of the sample size. All tests are implemented at a nominal level of 5%.

We consider different types of dynamics for the time-varying parameter: switching models, models with random structural breaks, and state space models. We also consider different types for time-varying parameters: means, variances, dependence parameters, and parameters describing other moments of the obervation density. Here, we mainly focus on the results of time-varying means. The results for variances and dependence parameters are highly similar.

In addition, we consider a Beta distribution with time-varying parameters in line with our empirical application in Section 4.

For the setting with a time-varying mean, our observation equation is

$$y_t = f_t + \varepsilon_t, \qquad \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2). \tag{12}$$

Results for other distributions such as Student's t are highly similar; see the working paper version of this paper. We consider the following DGPs for f_t.

State space models: We assume that f_t follows an autoregressive process of order one

$$f_{t+1} = \phi f_t + \sigma_\eta \eta_t, \tag{13}$$

where η_t is normally distributed with zero mean and unit variance. We set $\sigma_\eta^2 = \delta/T$, with $\delta = 0, \ldots, 200$, denoting the deviation from the null of constant parameters.

Regime switches or random breaks: We assume that f_t follows a discrete Markov chain with $s_1 = 0$ and

$$\mathrm{P}[s_{t+1} = j \mid s_t = j] = 1 - \mathrm{P}[s_{t+1} = j + 1 \mid s_t = j] = \pi, \tag{14}$$

and $f_t = \tilde{f}_j$, with $\tilde{f}_j$ an i.i.d. sequence of $\mathrm{N}(0, \delta/T)$ random variables. This creates random stretches of f_t at different levels as generated by $\tilde{f}_t$. We consider several values of $\pi \in [0, 1]$. This set-up of random regimes aligns well with the set-up adopted in Chib (1998) and Pesaran et al. (2006).

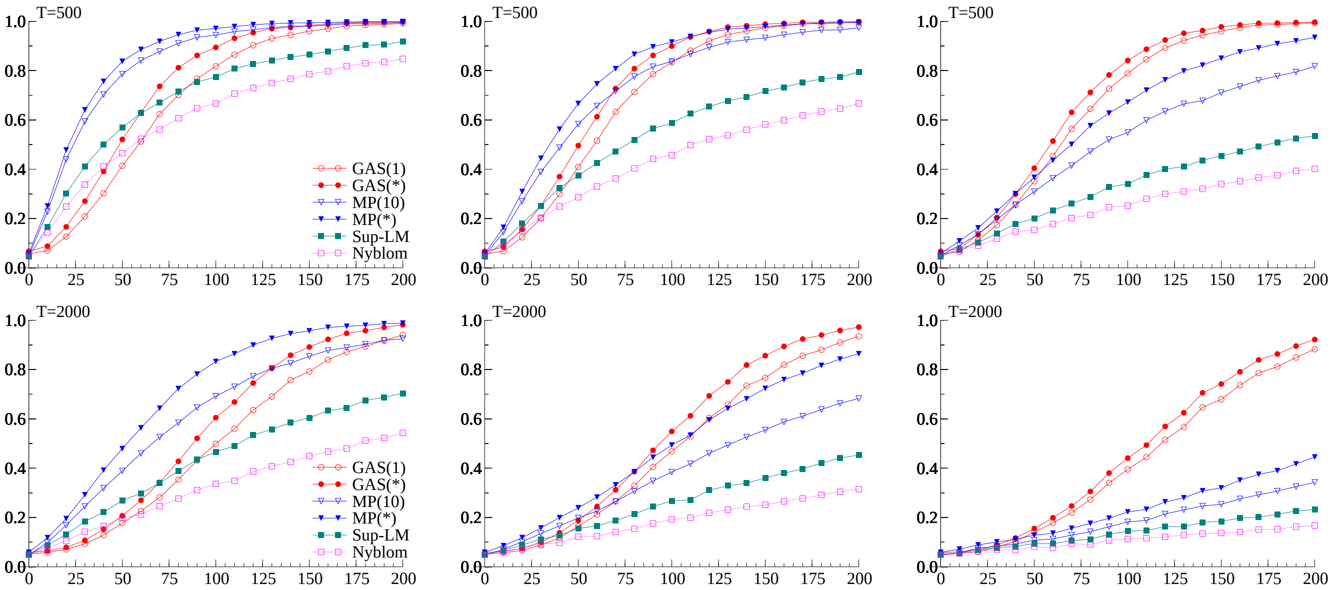
Apart from the DGP in (12), we also use a simulation set-up close to our empirical application from Section 4. A full discussion of the model is deferred until Section 4. Here, it suffices to state that we consider a measurement equation of the form

$$y_t \mid f_t \quad \overset{\text{i.i.d.}}{\sim} \quad B(\alpha_t, \beta_t), \tag{15}$$

where $B(\alpha_t, \beta_t)$ denotes the Beta distribution with parameters α_t and β_t. We consider two versions of this model. In the first version, we use the dynamics in equation (13) and set

$$\alpha_t = \bar{f} \times \frac{\exp(f_t)}{1 + \exp(f_t)}, \qquad \beta_t = \frac{\bar{f}}{1 + \exp(f_t)}, \qquad \bar{f} > 0, \tag{16}$$

so that mean $\mu_t = \exp(f_t)/(1+\exp(f_t))$ and variance $\mu_t(1-\mu_t)/(1+\bar{f})$ of the beta distribution are both time-varying. The mean lies in the (0,1) range by construction, irrespective of the value of f_t. The variance automatically tends to zero if the mean tends to either 0 or 1, which is natural for the beta distribution. The test statistics in this setting and the next setting test whether f_t is constant or not. The constant $\bar{f} > 0$ determines the additional extent of concentration of the distribution.



Panel (a): $\phi = 0.95$ Panel (b): $\phi = 0.90$ Panel (c): $\phi = 0.80$

Figure 1: Empirical rejection frequencies for state-space DGP (13) for a mean (12) with $\sigma_\varepsilon^2 = (1-\phi^2)^{-1}$ and $\sigma_\eta^2 = \delta/T$, where δ is on the horizontal axis. The tests are the GAS-based LM tests (with 1 or optimal ($*$) lags according to Escanciano and Lobato (2009)), the Müller-Petalas test (with tuning parameter 10 (prescribed) or data-driven ($*$)), the Andrew's sup-LM test, and the Nyblom test. Vertical axis holds rejection frequencies for a 5% nominal level critical value

The second version of the model uses the empirical estimates from Section 4 to generate the data. We set

$$\begin{aligned} \alpha_t &= \exp(f_{1,t}), \qquad f_{1,t+1} = \omega_1 + \phi_1 f_{1,t} + \sigma_\eta \eta_t, \qquad \sigma_\eta^2 = \delta/T, \\ \beta_t &= \exp(f_{2,t}), \qquad f_{2,t+1} = \omega_2 + \phi_2 f_{2,t} + \sigma_\eta \eta_t, \end{aligned} \tag{17}$$

with $\omega_1 = 0.0559$, $\omega_2 = -0.0023$, $\phi_1 = k \cdot 0.8571$, and $\phi_2 = k \cdot 0.9235$, and $k = 0.90, 1.00, 1.05$. For $k = 1$ we obtain the setting with the empirically relevant parameters from Section 4. For $k > 1$, there is more persistence in the time-varying parameters, and for $k < 1$ there is less.

3.2 Results

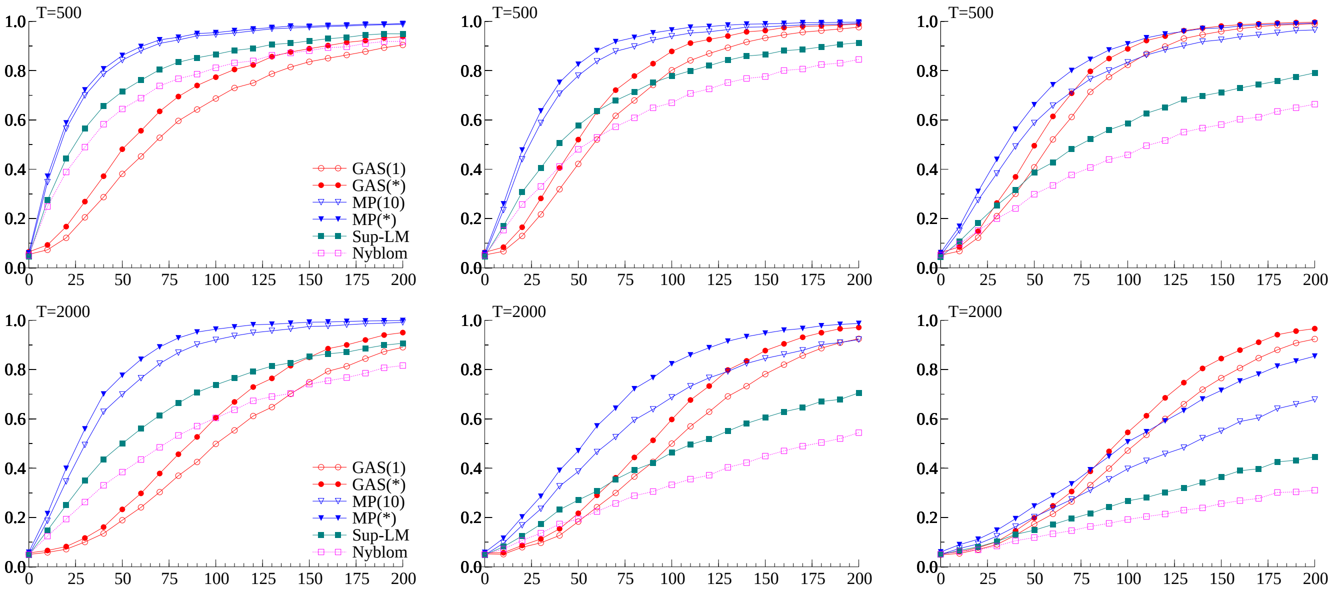
We start our discussion with the results for the state space DGP in (13). The key results are presented in Figure 1. We set $(1-\phi^2)\sigma_\varepsilon^2$ to a fixed constant, such that the magnitude of the signal-to-noise ratio is comparable along the horizontal axis in all three panels varying from 0% for $\sigma_\eta^2 = 0$ to 17% for $\sigma_\eta^2 = 0.06$. Such signal-to-noise ratio values are similar to those typically obtained for stock return data with possibly time-varying expected returns. The time-variation in the mean is rather subtle in those cases compared to the noise in the return data. This is also the case in our simulation settings.

Given the choice of variances $\sigma_\varepsilon^2 = (1-\phi^2)^{-1}$ and $\sigma_\eta^2 = \delta/T$, the signal to noise ratio is δ/T. We see that the power of most tests quickly decreases if the persistence (ϕ) of the time-varying parameters decreases. This is particularly clear in larger samples ($T = 2,000$). Only the GAS-LM test has a robust power performance, irrespective of the stickiness of the time-varying parameter dynamics. Because of this, the GAS based tests have the best power performance if the time-varying parameters are strongly mean reverting, while the MP tests perform best if the parameters are rather sticky. These results are in line with the (point) optimality of the MP test close to the null of constant parameters. The results are also in line with equation (8). A fixed value c for the MP test determines both the stickiness of the time-varying parameter process and the size of its innovations. It does so in such a way that the signal-to-noise ratio is always constant. However, in Figure 1, the signal-to-noise ratio varies with δ as we move to the right along the horizontal axis. Conversely, for a fixed position on the horizontal axis, we therefore expect the power of the MP test to rise as c grows smaller, i.e., as we move from the right-hand to the left-hand plot. This is precisely what happens.

It is also clear from Figure 1 that the data-driven selections of the tuning parameters work well. Both for the MP(*) test and the GAS-LM(*) test, power increases if we select the tuning parameters in a data-driven manner. The increased power does not appear to be caused by substantial size distortions.

These results are confirmed for the DGP with random regime switches as presented in Figure 2. The higher the value of π, the stickier the time-varying parameter process in the DGP. Again, as the stickiness of the time-varying parameter decreases, the power of the MP test as well as of the Andrews sup LM test and Nyblom test quickly deteriorate. The performance of the GAS-LM test, by contrast, remains highly robust, irrespective of the value of π.

The robustness of the GAS-LM test may look somewhat surprising at first sight. The explanation, however, is rather straightforward. The GAS test is based on the autocorrelation in the likelihood scores with respect to the time-varying parameter. These likelihood scores are evaluated under the null hypothesis. If the true DGP has a time-varying parameter, the scores are autocorrelated and the first order (nonzero) autocorrelation can be easily estimated consistently. The persistence of the time-varying parameter will have little effect on the estimate of the first order autocorrelation, which explains the robust performance of GAS-LM in both Figures 1 and 2. The MP(10) test, by contrast, estimates the path of the time varying parameter using a persistence parameter of $1 - 10/T$. If the true DGP has a quickly mean reverting time-varying parameter, the imposed persistence parameter is too high to adequately capture the time-varying parameter path. As a result, the unconditional variance of this path is estimated too low, and the MP test, being based on this unconditional variance, loses power. The data-



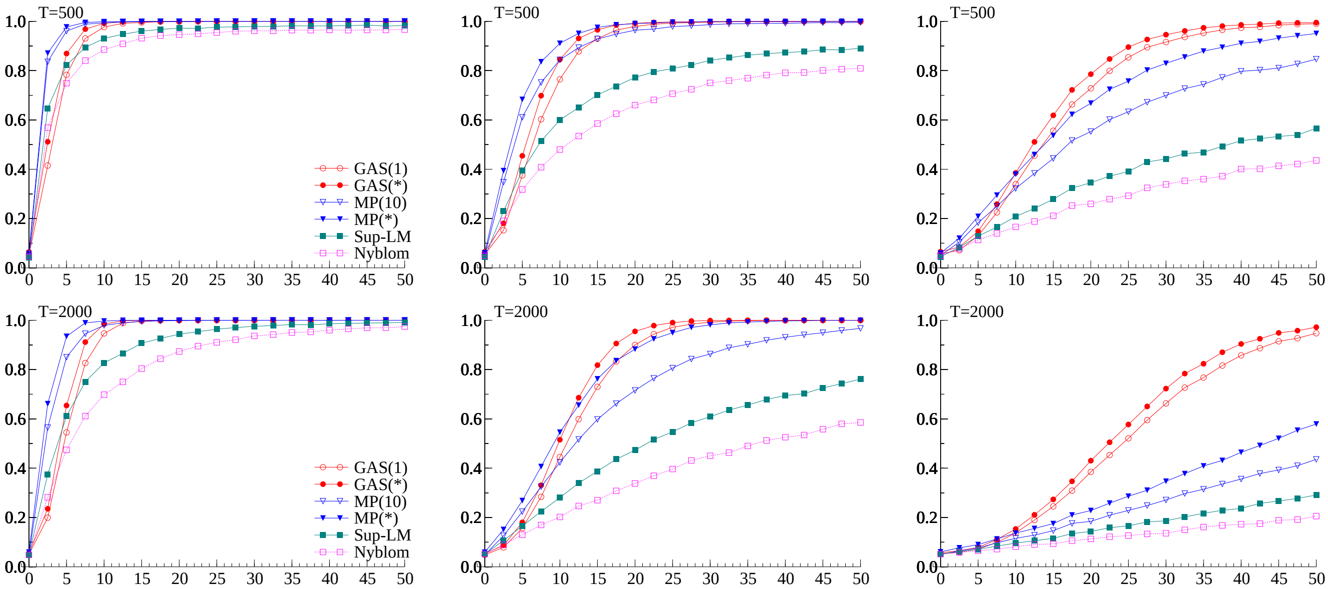
Panel (a): $\pi = 0.98$ Panel (b): $\pi = 0.95$ Panel (c): $\pi = 0.90$

Figure 2: Empirical rejection frequencies (random regime switches alternative) for the GAS-based LM tests (with 1 or optimal $(*)$ lags according to Escanciano and Lobato (2009)), the Müller-Petalas test (with tuning parameter 10 (prescribed) or data-driven $(*)$), the Andrew's sup-LM test, and the Nyblom test. Vertical axis holds rejection frequencies for a 5% level critical value, while the horizontal axis holds the value of δ in (14). The measurement volatility is $\sigma_\varepsilon^2 = 1$.

driven choice of the constant c in the MP$(*)$ test alleviates this problem to some extent, but the root of the problem remains.

Finally, we present a subset of the simulation results for the Beta distribution. The data generating process is based on the model of Section 4 with its parameter values being close to the reported empirical estimates; see equation (17). The results of these simulations are presented in Figure 3.

For $k = 1$, the persistence parameters coincide with those from the empirical setting (middle panels). In this case, it appears that the GAS test and the MP test are about equally powerful, with a slight advantage for the MP test in smaller samples ($T = 500$) and a slight advantage for the GAS test in larger samples ($T = 2000$). Again, we see the familiar pattern: if the persistence in the time-varying parameter processes becomes stickier ($k > 1$), the MP tests behave better. By contrast, if there is more mean reversion, the GAS test performs best and the other tests drop substantially in terms of power. A joint use of the MP and GAS tests thus appears to have power against a wider range of alternatives than each of these tests in isolation.



Panel (a): $k = 1.05$ Panel (b): $k = 1.00$ Panel (c): $k = 0.90$

Figure 3: Empirical rejection frequencies for the Beta distribution with empirically estimated parameters as in (17), with $\sigma_\eta^2 = \delta/T$, where δ is on the horizontal axis. The tests are the GAS-based LM tests (with 1 or optimal ($*$) lags according to Escanciano and Lobato (2009)), the Müller-Petalas test (with tuning parameter 10 (prescribed) or data-driven ($*$)), the Andrew's sup-LM test, and the Nyblom test. The vertical axis holds the rejection frequencies for a 5% nominal level critical value. For $k = 1$, the persistence parameters are as estimated empirically. There is more persistence for $k > 1$, and more mean reversion for $k < 1$.

4 Empirical application

In our empirical application, we consider quarterly observations of loss given defaults (LGDs) on corporate bonds. The data are obtained from Moody's and cover the first quarter of 1982 to the first quarter of 2010. The fraction of losses is measured as the percentage price drop in the value of the corporate bond from the-day-before to 20-days-after the announcement of default. The percentage price drop is also known as market implied LGD and can become negative, for example, after a timely restructuring of a firm or after a merger announcement. We censor negative observed LGDs to 1 basis point, 0.01%. The censoring affects only 14 out of 1125 observations, that is, 1.25% of our data set.

Our aim is to test whether there is significant time variation in the distributional characteristics of LGDs. Understanding the possible time-varying nature of (average) LGDs is important for credit risk modeling and financial stability research: credit portfolio losses could be severely underestimated if default risk and LGD risk exacerbate one another; see, for example, Creal et al. (2014). There are several lines of literature highlighting the possible state dependence of LGDs and its co-variation with macroeconomic conditions. Particularly for financial sector firms, time-varying LGDs have been a major focus since the 2008 financial crises. A possible cause for this time-variation is the occurrence of fire sales: prices of collateral can deteriorate quickly in case many firms liquidate the same type of assets at the same times, e.g., during a period of stress; see Shleifer and Vishny (2010) for an overview. Such deteriorating prices lead to higher LGDs particularly at times when default risk is also higher and observed defaults are more frequent. Helwege (2010) argues that default cascades and fire sales may be over-emphasized and that common factors in collateral values of diversified portfolios are likely to play a more important role. This is in line with earlier models by for example Frye (2000) and Jokivuolle and Peura (2003) that study the effect of systematic components in LGDs; see also the overview of Altman et al. (2005). The argument is that during bad economic states, collateral values may drop because there are fewer possible buyers for these assets. Either way, there are a number of arguments why (average) LGDs may not be constant over time, and the testing procedures discussed in the previous sections can be used to check this on the empirical data.

Our LGD data are at an (equidistant) quarterly frequency. Given the large number of companies rated by Moody's, we observe at least 1 LGD in each quarter. Apart from this standard format, the data display several non-standard features. First, LGDs are measured as percentage losses so that they are bounded to the interval $[0, 1]$. We therefore assume that the LGDs are drawn from a beta distribution with possibly time-varying parameters as explained further below. Second, the number of observed LGDs varies per quarter. If there are more defaults in a particular quarter, we also observe more LGDs. Hence the dimension of the observation vector is typically different for each quarter. Such features have to be accounted for in the testing methodology. Particularly during stressed periods, we observe more defaults and more LGDs, as clearly shown in Figure 4. Combining the observation period with the varying number of LGDs per quarter, we have 1125 LGD-quarter observations, with the number of LGDs per quarter varying from 1 in 1982 to a maximum of 58 in 2009.

Let $y_{i,t}$ denote the ith observation at time t with $i = 1, \ldots, K_t$, where K_t represents the number of LGD observations at time t. We take K_t as given and model $y_{i,t}$ at time t as independent draws from a beta distribution with time-varying parameters $\alpha_t = \exp(f_{1,t})$ and $\beta_t = \exp(f_{2,t})$ where $f_t = (f_{1,t}, f_{2,t})'$. Define $y_t = (y_{1,t}, \ldots, y_{K_t,t})'$. Then the log conditional observation density of y_t is given by

$$\ln p(y_t|f_t) = \sum_{i=1}^{K_t} \ln\Gamma(\alpha_t + \beta_t) - \ln\Gamma(\alpha_t) - \ln\Gamma(\beta_t) + (\alpha_t - 1)\ln y_{i,t} + (\beta_t - 1)\ln(1 - y_{i,t}), \quad (18)$$

where Γ denotes the gamma function. The conditional score and information matrix for (18)

Table 1: Test statistics and critical values for the corporate LGD data, 1982Q1–2010Q1

	Stat	10%	5%	1%
LM GAS(0,1)	16.20	4.60	5.99	9.21
LM GAS(0,5)	22.84	15.99	18.31	23.21
MP(10)	-27.34	-12.80	-14.32	-17.57
MP(10)*	-40.78	-12.80	-14.32	-17.57
Andrews	18.56	10.01	11.79	15.51
Nyblom	1.63	0.61	0.75	1.07

are given by

$$\nabla_{f,t} = \sum_{i=1}^{K_t} \begin{pmatrix} \big(\Psi(\alpha_t + \beta_t) - \Psi(\alpha_t) + \ln y_{i,t}\big) \times \alpha_t \\ \big(\Psi(\alpha_t + \beta_t) - \Psi(\beta_t) + \ln(1 - y_{i,t})\big) \times \beta_t \end{pmatrix}, \tag{19}$$

and

$$\mathcal{I}_t = K_t \times \begin{pmatrix} \alpha_t^2(\Psi'(\alpha_t) + \Psi'(\alpha_t + \beta_t)) & -\alpha_t\beta_t\Psi'(\alpha_t + \beta_t) \\ -\alpha_t\beta_t\Psi'(\alpha_t + \beta_t) & \beta_t^2(\Psi'(\beta_t) + \Psi'(\alpha_t + \beta_t)) \end{pmatrix}, \tag{20}$$

where Ψ denotes the digamma function that is defined as $\Psi(x) = \mathrm{d}\ln\Gamma(x)/\mathrm{d}\,x$. We set the GAS scaling matrix to the inverse information matrix, $S_t = \mathcal{I}_t^{-1}$, to account for the curvature of the score; see Creal et al. (2013). Using these definitions, we are able to compute the considered test statistics. The results are presented in Table 1.

All test statistics clearly reject the null hypothesis of constant parameters. We have slightly modified the Muller-Petalas test (denoted as MP(10)*) to account for the fact that the number of observations K_t varies over time. In the original MP(10) paper, the Hessian is estimated unconditionally over the full sample since the number of observations for each period is constant. In our setting of a varying number of observations K_t, we treat K_t as given but multiply the Hessian in the algorithm of MP(10) at time t by $K_t/\bar{K}$, where $\bar{K}$ is the average of K_t in the full sample. This modification follows from a similar derivation that is used for the information matrix in (20). It corrects the steps in the MP(10) algorithm for periods when there are either many or few LGD observations in the cross section.

Next we confirm the test results from Table 1 by estimating the path of the time-varying parameter f_t in two alternative ways. First, we estimate f_t based on the GAS(1,1) model as given by

$$f_{t+1} = \omega + As_t + Bf_t, \quad s_t = S_t\nabla_t, \quad A = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix}. \tag{21}$$

The model with full rather than diagonal matrices for A and B produces similar results and is

Table 2: GAS(1, 1) coefficients estimation results

Coeff.	Estimate	Std Err	t–stat
ω_1	0.0559	(0.0088)	6.3449
ω_2	-0.0023	(0.0073)	-0.3075
a_{11}	0.1943	(0.0220)	8.8156
a_{22}	0.1836	(0.0361)	5.0770
b_{11}	0.8571	(0.0233)	36.8217
b_{22}	0.9235	(0.0355)	26.0252

therefore omitted. The parameter estimates are presented in Table 2. All parameter estimates are strongly significant, except ω_2. We find that there is strong persistence in both α_t and β_t, as both b_{11} and b_{22} are relatively high. Interestingly, the persistence in α_t is not as strong as that in β_t, that is $b_{11} < b_{22}$. Since α_t and β_t characterize the mean of the beta distribution when it is close to 0 and 1, respectively, the higher persistence of β_t indicates that the higher LGDs are more persistent than low LGDs. Such differences do not appear between a_{11} and a_{22}.

Our second estimate of f_t is obtained as a by-product of the MP(10) algorithm. It is based on the Weighted Average Risk estimate of the path f_t for several local alternatives as explained in Section 2.2. We use the same method as for MP(10)* to correct for the time-varying number of observations K_t when estimating the path. The results are presented in Figure 4.

The LGD observations range from close to zero to almost one for given cross sections. In Figure 4 we also plotted the mean of the beta distribution $\alpha_t/(\alpha_t + \beta_t)$. The MP(10)* and GAS estimates of the mean capture the salient features of the data. There are clear peaks in average credit losses around the 1991 recession, the 2000-2001 burst of the dotcom bubble, and the most recent financial crisis. The peaks clearly defy the assumption of constant parameters. The MP estimate appears to lead the GAS estimate. We point out that the GAS estimate is a filter (produces a one-sided estimate) and the MP estimate is a smoother (produces a two-sided estimate). In the latter case, future observations are also taken into account. We also observe that the two estimates differ substantially in the period before the 2008 financial crisis. The GAS estimate reveals a more moderate trough than the MP estimate.

We further conclude that the MP estimate is rather successful in extracting the mean signal throughout the sample while it is designed for *local* time variation only. The smoothed path of f_t in Figure 4 for MP(10)* is constructed by a weighted average of 10 different paths, corresponding to the autoregressive coefficients $b_{11} = b_{22} = 1 - c/T$ for $c = 0, 5, 10, \ldots, 50$, with $T = 113$. The largest weights are assigned to the paths corresponding to $c = 30, \ldots, 50$ while

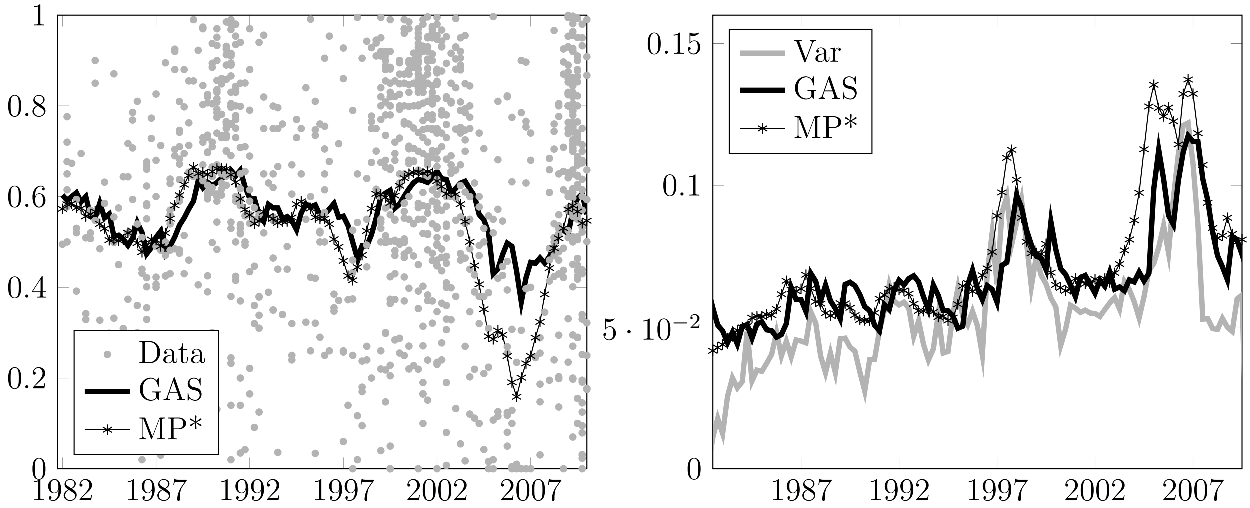


Figure 4: Corporate Loss Given Default (LGD) data, 1982–2010

This figure contains the market implied LGDs of corporate bonds over the period 1982Q1–2010Q1 as observed by Moody's, left panel. The left panel also contains the mean of the fitted beta distribution, $\alpha_t/(\alpha_t + \beta_t)$, for the GAS model from Table 2 and the MP(10)* smoothed parameter path of Müller and Petalas (2010). The right hand curve provides the estimates of the variance, $\alpha_t/((\alpha_t + \beta_t)^2(1 + \alpha_t + \beta_t))$, for both methods, as well as a 1 year rolling window estimate of the variance (Var).

the mode weight is at $c = 40$; it corresponds to an autoregressive decay of $1 - 40/113 \approx 0.65$. This autoregressive coefficient is lower than those of the GAS model, b_{11} and b_{22}, in Table 2. Moreover, the autoregressive coefficient in the MP(10) method is the same for α_t and β_t, in contrast to the GAS model. A smaller persistence parameter in MP(10)* is counterbalanced by a higher innovation variance in order to match the unconditional variance. The two effects lead to an MP(10)* estimate that is more sensitive to the small LGD values in the period leading up to the 2008 credit crisis. We emphasize that the MP(10)* test is not influenced by the less persistent paths. The MP(10)* test is based on the local alternative $c = 10$ that corresponds to a persistence parameter of $1 - 10/113 \approx 0.91$. This value is closer to the estimated persistence parameters b_{11} and b_{22} in the GAS model.

The right hand panel of Figure 4 presents the estimates of the variance of our beta model as given by $\alpha_t/((\alpha_t+\beta_t)^2(1+\alpha_t+\beta_t))$. The time-varying variance is slightly trending upwards. The variation in more recent LGD percentages is somewhat larger than in the early 1980s. We observe two peaks in the variance. These are linked to periods when LGD observations are sparse and the corresponding relative dispersions are high. The variance estimates in the MP(10)* and GAS frameworks are roughly similar. The main differences are in the periods around 1997 and around 2004–2006. In the latter period we find that the lower mean for MP(10)* in the left panel of Figure 4 is partly compensated by the higher variance. The MP(10)* test is not affected because it is based on the autoregressive coefficient of approximately 0.91 for MP(10)*, rather than 0.65 (implied by the mode $c = 40$) for the smoothed estimates reported in Figure 4.

The results clearly suggest that there is a sticky, time-varying systematic risk component in observed LGDs. A next step in the modeling process would then be to extend the standard, static Beta distribution to include explanatory variables in the specification for (the logit of) the mean, f_t, in equation (16),

$$f_t = b_0 + b_1 x_{1,t} + \ldots + b_k x_{k,t},$$

where $x_{1,t}, \ldots, x_{k,t}$ denotes a set of covariates, and $b_0, \ldots, b_k$ a set of fixed, unknown parameters. Typical variables $x_{i,t}$ include business cycle indicators, bank lending climate variables, and monetary conditions, as also used to model the time-variation in probabilities of default; see for example Duffie et al. (2007), and Koopman et al. (2009, 2012). This would account for part of the systematic cross-sectional dependence of the LGDs during each quarter, particularly during periods of stress. For empirical results using such an approach, we refer to the overview of Altman et al. (2005). An extended model could subsequently be tested for any remaining time-variation in the parameters by performing a test on the intercept b_0 in the new specification for f_t. We leave such an extension for future work.

5 Conclusions

We proposed a new omnibus misspecification test for parameter instability in general nonlinear non-Gaussian time series models. By adopting the generalized autoregressive score (GAS) model of Creal et al. (2013) as the time-varying parameter process under the alternative, we were able to derive a Lagrange Multiplier test for the null of constant parameters against the alternative of parameter instability. In an extensive Monte Carlo study we found that the new test had a remarkably robust behavior to the persistence of the time-varying parameter process. The test kept its power if the parameters mean-revert more quickly. This stands in sharp contrast to other tests from the literature, such as the test of Müller and Petalas (2010) against a parameter-driven alternative. If the time-varying parameters vary sufficiently slowly, the Müller-Petalas test has the best behavior. For time-varying parameter processes with less persistence, however, the GAS-LM test developed in this paper may have a better behavior. This result holds despite a data-driven choice of the tuning constants for both the GAS-LM test and the Müller-Petalas test.

We also applied all tests to an empirical panel data set consisting of loss given default percentages for U.S. corporate bonds. We have shown how the tests can be used in a practical setting and how they can be adapted in cases with a time-varying number of observations. An interesting finding is that the smoothing approach of Müller and Petalas (2010) can also be useful in cases of non-local time variation in the parameters. The estimated paths from their algorithm produce similar results as the estimated path from the GAS model in our empirical application. It illustrates that the two testing paradigms can provide complementary as well as mutually reinforcing evidence in empirical studies.

References

Altman, E. I., B. Brady, A. Resti, and A. Sironi (2005). The link between default and recovery rates: Theory, empirical evidence, and implications. *Journal of Business 78*(6), 2203–2227.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 817–858.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–856.

Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica 62*(6), 1383–1414.

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*(1), 1–22.

Bauwens, L. and D. Veredas (2004). The stochastic conditional duration model: A latent factor model for the analysis of financial durations. *Journal of Econometrics 119*(2), 381–412.

Blasques, F., S. J. Koopman, and A. Lucas (2015). Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika*, forthcoming.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics 31*(3), 307–327.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics 86*(2), 221–241.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica 28*(3), 591–605.

Cipollini, F., R. F. Engle, and G. M. Gallo (2013). Semiparametric vector MEM. *Journal of Applied Econometrics 28*(7), 1067–1086.

Creal, D. D. (2012). A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews 31*(3), 245–296.

Creal, D. D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics 29*(4), 552–563.

Creal, D. D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Creal, D. D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). Observation driven mixed-measurement dynamic factor models with an application to credit risk. *The Review of Economics and Statistics 96*(5), 898–915.

Creal, D. D. and R. S. Tsay (2015). High-dimensional dynamic stochastic copula models. *Journal of Econometrics 189*(2), 335–345.

Davidson, R. and J. G. MacKinnon (1990). Specification tests based on artificial regressions. *Journal of the American Statistical Association 85*(409), 220–227.

Davis, R. A., W. T. M. Dunsmuir, and S. Streett (2003). Observation driven models for Poisson counts. *Biometrika 90*(4), 777–790.

De Lira Salvatierra, I. and A. J. Patton (2015). Dynamic copula models and high frequency data. *Journal of Empirical Finance 30*, 120–135.

Duffie, D., L. Saita, and K. Wang (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics 83*(3), 635–665.

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space models. *Biometrika 84*, 669–684.

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *Journal of the Royal Statistical Society (Series B) 62*, 3–56.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (2nd ed.). Oxford: Oxford University Press.

Elliott, G. and U. K. Müller (2006). Efficient tests for general persistent time variation in regression coefficients. *The Review of Economic Studies 73*(4), 907–940.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*(4), 987–1007.

Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics 131*(1-2), 3–27.

Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: A new model for volatility using intra-daily data. *Econometrica 66*(5), 1127–1162.

Escanciano, J. C. and I. N. Lobato (2009). An automatic Portmanteau test for serial correlation. *Journal of Econometrics 151*, 140–149.

Franzini, L. and A. C. Harvey (1983). Testing for deterministic trend and seasonal components in time series models. *Biometrika 70*(3), 673–682.

Frye, J. (2000). Collateral damage: A source of systematic credit risk. *Risk 13*(4), 91–94.

Hafner, C. and H. Manner (2012). Dynamic stochastic copula models: Estimation, inference and applications. *Journal of Applied Econometrics 27*(2), 269–295.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357–384.

Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in us labor productivity. *Journal of Economic perspectives 15*(4), 117–128.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Series Monographs. Cambridge University Press.

Harvey, A. C. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association 109*, 1112–1122.

Helwege, J. (2010). Financial firm bankruptcy and systemic risk. *Journal of International Financial Markets, Institutions and Money 20*(1), 1–12.

Jokivuolle, E. and S. Peura (2003). Incorporating collateral value uncertainty in loss given default estimates and loan-to-value ratios. *European Financial Management 9*(3), 299–314.

King, M. L. and G. H. Hillier (1985). Locally best invariant tests of the error covariance matrix of the linear regression model. *Journal of the Royal Statistical Society. Series B (Methodological) 47*(1), 98–102.

Koopman, S. J., R. Kräussl, A. Lucas, and A. B. Monteiro (2009). Credit cycles and macro fundamentals. *Journal of Empirical Finance 16*(1), 42–54.

Koopman, S. J., A. Lucas, and M. Scharth (2015). Predicting time-varying parameters with parameter-driven and observation-driven models. *The Review of Economics and Statistics 98*(1), 97–110.

Koopman, S. J., A. Lucas, and B. Schwaab (2012). Dynamic factor models with macro, frailty, and industry effects for us default counts: the credit crisis of 2008. *Journal of Business & Economic Statistics 30*(4), 521–532.

Lee, J. (1991). A Lagrange multiplier test for GARCH models. *Economics Letters 37*(3), 265–271.

Lucas, A., B. Schwaab, and X. Zhang (2014). Conditional euro area sovereign default risk. *Journal of Business & Economic Statistics 32*(2), 271–284.

Müller, U. K. and P. E. Petalas (2010). Efficient estimation of the parameter path in unstable time series models. *The Review of Economic Studies 77*, 1508–1539.

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association 84*(405), 223–230.

Nyblom, J. and T. Mäkeläinen (1983). Comparisons of tests for the presence of random walk coefficients in a simple linear model. *Journal of the American Statistical Association 78*(384), 856–864.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review 47*(2), 527–556.

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

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006). Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies 73*(4), 1057–1084.

Ploberger, W., W. Krämer, and K. Kontrus (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics 40*(2), 307–318.

Qu, Z. and P. Perron (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica 75*(2), 459–502.

Russell, J. R. (2001). Econometric modeling of multivariate irregularly-spaced high-frequency data. Unpublished manuscript, University of Chicago, Graduate School of Business.

Shephard, N. (2005). *Stochastic Volatility: Selected Readings.* Oxford: Oxford University Press.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–667.

Shleifer, A. and R. W. Vishny (2010). Fire sales in finance and macroeconomics. *NBER Working Paper No. 16642*.

Vogelsang, T. J. and P. Perron (1998). Additional tests for a unit root allowing for a break in the trend function at an unknown time. *International Economic Review*, 1073–1100.

White, H. (1987). *Advances in Econometrics, Fifth World Congress, T. F. Bewley (Ed.)*, Volume 1, Chapter Specification testing in dynamic models, pp. 1–58. Cambridge University Press.