12-2014

Structural change estimation in time series regressions with endogenous variables

Junhui QIAN
Shanghai Jiao Tong University

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Published in Economics Letters, 2014 December, volume 125, issue 3, 415-421.
http://dx.doi.org/10.1016/j.econlet.2014.10.021

Structural change estimation in time series regressions with endogenous variables

Junhui Qian [a], Liangjun Su [b,*]

[a] *Antai College of Economics and Management, Shanghai Jiao Tong University, China*
[b] *School of Economics, Singapore Management University, Singapore*

HIGHLIGHTS

- We propose to adopt the group fused lasso to estimate regression models with endogeneity and an unknown number of breaks.
- We propose an information criterion to determine the number of breaks correctly with probability approaching one.
- Post-Lasso GMM estimation can be conducted as usual.
- Simulations are done in comparison with some existing tests.
- We apply our method to study the forward-looking monetary policy rule of the US from 1960 to 2012 and detect two breaks.

Abstract

We propose to apply the group fused Lasso to estimate time series models with endogenous regressors and an unknown number of breaks. It can correctly determine the number of breaks and estimate the break dates asymptotically. Simulations and applications are given.

JEL classification:
C13
C22

Keywords:
Group fused Lasso
Multiple breaks
Penalized least squares
Penalized GMM
Structural change

1. Introduction

Consider the following time series regression with an unknown number of breaks in parameters,

$$y_t = x_t'\beta_t + u_t, \quad t = 1, \ldots, T, \tag{1.1}$$

where x_t is a $p \times 1$ vector of regressors with at least one element being endogenous. We assume that the coefficients $\{\beta_t\}$ are time-varying but subject to the restriction that the sequential changes in β_t are sparse:

$$\beta_t = \alpha_j \quad \text{for } t = T_{j-1}, \ldots, T_j - 1 \text{ and } j = 1, \ldots, m+1,$$

where $m \ll T$ and we adopt the convention that $T_0 = 1$ and $T_{m+1} = T + 1$. Here, m denotes the number of break dates that is typically unknown and $\left\{\alpha_j\right\}_{j=1}^{m+1}$ denotes the set of regime-specific parameters. Let z_t be a $q \times 1$ vector of instrumental variables (IVs) for x_t such that $q \geq p$. Andrews (1993) proposes three tests under the GMM framework for a one-time break and Andrews and Ploberger (1994) consider optimal tests of one-time structural change in nonlinear models with stationary observations. In principle, one can extend Andrews' tests to deal with multiple breaks by repeatedly applying them to sub-samples. However, as substantial trimming (e.g., 15%) is a common practice to enhance power, the approach can easily miss true break points not only at the boundaries but also in the middle of the sample. Recently, tests for multiple breaks based on 2sls are proposed by Hall et al. (2012), Boldea et al. (2012), and Perron and Yamamoto (2013, 2014). All

* Correspondence to: School of Economics, Singapore Management University, 90 Stamford Road, 178903, Singapore. Tel.: +65 6828 0386.
E-mail address: ljsu@smu.edu.sg (L. Su).

these papers assume that the number of breaks is known and their procedures also require trimming.

To the best of our knowledge, no method is available to directly estimate an unknown number of breaks with the presence of endogeneity in the regression. In this paper we apply the group fused Lasso (GFL) developed in Qian and Su (2014, QS hereafter) to the above model.[1] QS's GFL procedure is penalized-least-squares based and developed for linear regression models without endogeneity. Interestingly, we find that it also works for models with endogenous regressors under suitable conditions despite the fact that it does not utilize any IVs. Perron and Yamamoto (2013) similarly find that least-squares-based tests work for regressions with endogeneity and that sometimes even outperform tests based on 2sls. With a BIC-type model selection procedure, GFL can correctly determine the unknown number of breaks with probability approaching one (w.p.a.1) and estimate the break dates accurately as in Bai and Perron (1998). After obtaining the break dates, one can apply the usual GMM method to estimate the regime-specific parameters for each estimated regime and conduct statistical inference as usual. Simulations indicate such a procedure is comparable in terms of finite sample performance with the testing procedures of Andrews (1993) and Andrews and Ploberger (1994) when there is no break or a single break in the regression but our procedure can also correctly identify multiple breaks with high probability. We apply our GFL procedure to study the forward-looking monetary policy rule of the US from 1960Q1 to 2012Q4 and identify three regimes for this period.

The rest of the paper is organized as follows. We present the GFL procedure in Section 2 and report the asymptotic properties of the procedure in Section 3. We report the Monte Carlo evidence in Section 4 and apply our method to the estimation of the US monetary policy rule from 1960 to 2012 in Section 5.

2. Estimation

When x_t is endogenous but can be instrumented with z_t, one may be tempted to solve a penalized GMM (PGMM) problem for $\{\beta_t\}$:

$$\min_{\{\beta_t\}} \left[\frac{1}{T}\sum_{t=1}^{T} z_t(y_t-\beta_t' x_t)\right]' W_T \left[\frac{1}{T}\sum_{t=1}^{T} z_t(y_t-\beta_t' x_t)\right] + \lambda\sum_{t=2}^{T}\|\beta_t-\beta_{t-1}\|, \tag{2.1}$$

where W_T is a $q\times q$ symmetric weighting matrix that is positive definite asymptotically, $\lambda=\lambda_T\to 0_+$ is a tuning parameter, and $\|\cdot\|$ denotes the Frobenius norm. However, this formulation only works for the over-identification case. In the exact identification case, the resulting PGMM estimators of $\{\beta_t\}$ are given by

$$\hat{\beta}_t^{gmm} = \left(\sum_{s=1}^{T} z_s x_s'\right)^{-1}\sum_{s=1}^{T} z_s y_s \quad \text{for } t=1,\ldots,T,$$

which ensures the objective function in (2.1) to always take the value zero, regardless of the choices of W_T and λ. That is, the PGMM estimators of $\{\beta_t\}$ remain as a constant no matter whether there is a break in the data or not. So we cannot apply PGMM to estimate the number of breaks in the exact identification case. In the over-identification case, one might attempt to follow QS and derive the asymptotic properties of the PGMM estimators. But we find that the proof strategy in QS for penalized least squares (PLS) regressions cannot be extended to the PGMM framework because of the difference in the two types of objective functions and it remains unknown how to derive the asymptotic properties of the PGMM estimators in the over-identification case.

Given the above disadvantage and difficulty of applying the PGMM method to regression models with an unknown number of breaks, we follow QS and consider the following PLS estimation of $\{\beta_t\}$:

$$\min_{\{\beta_t\}} \frac{1}{T}\sum_{t=1}^{T}(y_t-\beta_t' x_t)^2 + \lambda\sum_{t=2}^{T}\|\beta_t-\beta_{t-1}\|, \tag{2.2}$$

where λ is defined as above. Let $\{\hat{\beta}_t\}$ denote the solution to the above PLS problem. Despite the fact that $\hat{\beta}_t$'s are generally inconsistent for β_t's, we will argue that such estimates can help identify the unknown number of breaks and estimate the unknown break dates under some suitable conditions. After obtaining such estimates, we can apply the usual post-Lasso GMM estimation for each estimated regime to obtain consistent estimates of the regime-specific structural parameters $\{\alpha_j\}$.

Given the solution $\{\hat{\beta}_t\}$ to (2.2), we obtain estimates of the break dates, $\hat{T}_1,\ldots,\hat{T}_{\hat{m}_\lambda}$, such that

$$\hat{\beta}_t=\hat{\beta}_s \quad \text{for all } t,s\in\left[\hat{T}_{j-1},\hat{T}_j-1\right] \quad \text{and}$$

$$\hat{\beta}_{\hat{T}_j}\neq\hat{\beta}_{\hat{T}_{j-1}}, \quad j=1,\ldots,\hat{m}_\lambda+1,$$

where $\hat{T}_0=1$ and $\hat{T}_{\hat{m}_\lambda+1}=T+1$. That is, $\hat{m}_\lambda$ and $\hat{\mathcal{T}}_{\hat{m}_\lambda}\equiv(\hat{T}_1,\ldots,\hat{T}_{\hat{m}_\lambda})$ denote the estimated number of breaks and estimated set of break points, respectively. Let $\hat{\alpha}_j=\hat{\beta}_{\hat{T}_{j-1}}$ for $j=1,\ldots,\hat{m}_\lambda+1$. Note that we have suppressed the dependence of $\hat{\beta}_t$ and $\hat{\alpha}_j$ on λ.

3. Asymptotic properties

We first introduce a set of assumptions and then study the asymptotic properties of the PLS estimators $\{\hat{T}_j\}$ and $\{\hat{\alpha}_j\}$.

3.1. Assumptions

We denote the true value of a parameter with a superscript ^{0}. In particular, $m^0,\ \beta_t^0,\ \alpha_j^0$, and T_j^0 denote the true values of $m,\ \beta_t,\ \alpha_j$, and T_j, respectively. Let $I_j^0=T_j^0-T_{j-1}^0$ for $j=1,\ldots,m^0+1$. Define

$$I_{\min}=\min_{1\le j\le m^0+1}\left|I_j^0\right|, \qquad J_{\min}=\min_{1\le j\le m^0}\left\|\alpha_{j+1}^0-\alpha_j^0\right\|, \quad \text{and}$$

$$J_{\max}=\max_{1\le j\le m^0}\left\|\alpha_{j+1}^0-\alpha_j^0\right\|.$$

For simplicity, we assume that m^0 is a fixed constant, $I_{\min}$ is proportional to T, $J_{\max}=O(1)$, and $J_{\min}^{-1}=O(1)$. These conditions can be relaxed as in QS at the cost of more complicated proofs and less transparent assumptions than specified below.

Let $\beta_t^*=\left[E\left(x_t x_t'\right)\right]^{-1}E\left(x_t y_t\right)$ and $u_t^*=y_t-\beta_t^{*\prime}x_t$. To study the consistency of the GFL procedure, we follow QS and make the following assumptions.

Assumption A1. (i) $\{(x_t,u_t),\ t=1,2,\ldots\}$ is a strong mixing process with mixing coefficients $\alpha(\cdot)$ satisfying $\alpha(\tau)\le c_\alpha\rho^\tau$ for some $c_\alpha>0$ and $\rho\in(0,1)$.

(ii) Either one of the following two conditions is satisfied: (a) $\sup_{t\ge1}E\|x_t\|^{4q}<\infty$ and $\sup_{t\ge1}E\left|u_t^*\right|^{4q}<\infty$ for some $q>1$; (b) there exist some constants c_{xx} and c_{xu} such that

[1] In related works, Kolar et al. (2009) adopts the fused Lasso to study piecewise constant varying-coefficient models and Bleakley and Vert (2011) apply the GFL to detect change points in a multidimensional signal.

$\sup_{t\geq 1} E[\exp(c_{xx}\|x_t\|^{2\gamma})] \leq C_{xx} < \infty$ and $\sup_{t\geq 1} E\left[\exp\left(c_{xu}\|x_t u_t^*\|^{\gamma}\right)\right] \leq C_{xu} < \infty$ for some $\gamma \in (0, \infty]$.

(iii) $b_t \equiv \left[E\left(x_t x_t'\right)\right]^{-1} E\left(x_t u_t\right) = b^0$ for some finite vector b^0 for $t = 1, \ldots, T$.

Assumption A2. (i) There exist two positive constants $\underline{c}_{xx}$ and $\bar{c}_{xx}$ and a positive sequence $\{\delta_T\}$ declining to zero as $T \to \infty$ such that

$$\underline{c}_{xx} \leq \inf_{1\leq s<r\leq T+1, r-s\geq T\delta_T} \mu_{\min}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} E\left(x_t x_t'\right)\right) \leq \sup_{1\leq s<r\leq T+1, r-s\geq T\delta_T} \mu_{\max}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} E\left(x_t x_t'\right)\right) \leq \bar{c}_{xx}.$$

(ii) $T\delta_T$ satisfies one of the following two conditions: (a) $T\delta_T \geq c_v T^{1/q}$ for some $c_v > 0$ if A1(ii.a) is satisfied; (b) $T\delta_T \geq c_v (\log T)^{(2+\gamma)/\gamma}$ for some $c_v > 0$ if A1(ii.b) is satisfied.

Assumption A3. As $T \to \infty$, $\delta_T \to 0$, $T\delta_T/(\log T)^{c_\delta} \to \infty$, and $\lambda/\delta_T \to 0$.

Assumptions A1–A3 parallel A1, A2 and A3** in QS. The only difference is that QS require that $E\left(x_t u_t\right) = 0$ in their Assumption A1(i) and hence rule out endogeneity. Here, we require that $\left[E\left(x_t x_t'\right)\right]^{-1} E\left(x_t u_t\right)$ remains a constant over time in our A1(iii), which includes QS's exogenous case as a special case. Apparently, $\beta_t^* = \beta_t^0 + b_t$ and b_t represents the bias term by using the OLS method to estimate the structural parameter β_t^0. The time-invariance of b_t can be satisfied under various cases, e.g., when $\{x_t, u_t\}$ is a covariance stationary process (in contrast with the strict stationarity assumption in Andrews and Ploberger, 1994). It can be relaxed along at least two directions. One is to allow for locally covariance-stationary process $\{x_t, u_t\}$ by requiring the local deviation of b_t from the mean value to be well controlled. The other is to restrict the variation of b_t directly such that it does not change the piecewise constancy of $\left\{\beta_t^0\right\}$, i.e., $\beta_t^0 = \beta_{t-1}^0$ if and only if $\beta_t^* = \beta_{t-1}^*$ for $t = 2, \ldots, T$. For simplicity and clarification, we maintain A1(iii) below. We refer readers to QS for the discussion on the other assumptions.

3.2. The consistency of $\hat{T}_j$

With A1–A3, we can establish the consistency of $\{\hat{T}_j\}$ and $\left\{\hat{\alpha}_j\right\}$ conditional on the event $\hat{m}_\lambda = m^0$.

Theorem 3.1. *Suppose that Assumptions A1–A3 hold. If $\hat{m}_\lambda = m^0$, then*

(i) $P\left(\max_{1\leq j\leq m^0}\left|\hat{T}_j - T_j^0\right| \leq T\delta_T\right) \to 1$ *as* $T \to \infty$,

(ii) $\hat{\alpha}_j - (\alpha_j^0 + b^0) = O_P\left(T^{-1/2} + \lambda + \delta_T\right)$ *for each* $j = 1, \ldots, m^0 + 1$.

Proof. The proof follows essentially from that of Theorem 3.1 in QS. Let $\hat{\theta}_1 = \hat{\beta}_1$ and $\hat{\theta}_t = \hat{\beta}_t - \hat{\beta}_{t-1}$ for $t = 2, \ldots, T$. The Karush–Kuhn–Tucker optimality conditions for the PLS problem imply that

(a) $\frac{1}{T}\sum_{r=\hat{T}_j}^{T} x_r(y_r - x_r'\hat{\beta}_r) = \frac{\lambda}{2}\hat{\theta}_{\hat{T}_j}/\left\|\hat{\theta}_{\hat{T}_j}\right\|$ for $j = 1, \ldots, \hat{m}$;

(b) $\frac{1}{T}\left\|\sum_{r=t}^{T} x_r(y_r - x_r'\hat{\beta}_r)\right\| \leq \frac{\lambda}{2}$ for $t = 1, \ldots, T$.

In addition, Lemma A.3 in QS continues to hold under our Assumptions A1 and A2 and Lemma A.4 in QS holds with u_t replaced by u_t^*. By using these results repeatedly and following the proof of Theorem 3.1 in QS, we can readily prove (i)–(ii).

Let $\kappa_j^0 = T_j^0/T$ and $\hat{\kappa}_j = \hat{T}_j/T$ for $j = 1, \ldots, m^0$. Theorem 3.1(i) suggests that $\max_{1\leq j\leq m^0}\left|\hat{\kappa}_j - \kappa_j^0\right| = O_p(\delta_T)$, i.e., the break ratios κ_j^0, $1 \leq j \leq m^0$, can be estimated at rate δ_T provided that $\hat{m}_\lambda = m^0$. Theorem 3.1(ii) indicates that $\hat{\alpha}_j$ is a consistent estimator of $\alpha_j^0 + b^0$ and its convergence rate depends on the choice of λ and the convergence rate of $\hat{\kappa}_j$. Below we propose a BIC-type information criterion (IC) that helps to determine the tuning parameter to ensure $\hat{m}_\lambda$ to be equal to m^0 w.p.a.1.

Given the break date estimators, one can apply the GMM method to each estimated regime to obtain the post-Lasso GMM estimators of regime-specific parameters. To ensure $\delta_T = o(T^{-1/2})$ so that the estimation of break dates has no effect on the first order asymptotic distribution of these regime-specific estimators, we need $q > 2$ in Assumption A1(ii.a) so that both x_t and u_t^* have finite eight plus moments. In this case, we can set $\delta_T \propto T^{(1-q)/q}$ and $\lambda = \delta_T/\log T$ so that all conditions in Assumption A3 can also be satisfied. If Assumption A1(ii.b) is satisfied, then by setting $\delta_T \propto (\log T)^{(2+\gamma)/\gamma}/T$ and $\lambda = \log T/T$, we can ensure that Assumptions A2(ii) and A3 are simultaneously satisfied.

3.3. Selection of the tuning parameter λ

Given $\hat{\mathcal{T}}_{\hat{m}_\lambda}$, we can obtain the post-Lasso OLS estimators of the $\{\alpha_j\}$ by solving

$$\min_{\{\alpha_j\}} \sum_{j=1}^{\hat{m}_\lambda+1}\sum_{t=\hat{T}_{j-1}}^{\hat{T}_j-1}(y_t - \alpha_j' x_t)^2.$$

We denote the estimators as $\hat{\alpha}_j^p(\lambda)$, $j = 1, \ldots, \hat{m}_\lambda + 1$, which are generally inconsistent with the structural parameters $\left\{\alpha_j\right\}$. We propose to minimize the following BIC-type information criterion to determine λ

$$\mathrm{IC}(\lambda) = \log(\hat{\sigma}_\lambda^2) + \rho_T p(\hat{m}_\lambda + 1), \tag{3.1}$$

where $\hat{\sigma}_\lambda^2 = \frac{1}{T}\sum_{j=1}^{\hat{m}_\lambda+1}\sum_{t=\hat{T}_{j-1}}^{\hat{T}_j-1}(y_t - \hat{\alpha}_j^p(\lambda)'x_t)^2$. Let $\hat{\lambda} = \arg\min_{\lambda\in\Lambda}\mathrm{IC}(\lambda)$ where Λ is a properly chosen set of tuning parameters. In practice, one may conduct grid search in an interval $\Lambda = [\lambda^{\min}, \lambda^{\max}]$, where $\lambda^{\min}$ and $\lambda^{\max}$ are selected by trials so that $\lambda^{\max}$ would lead to zero break and $\lambda^{\min}$ would lead to many breaks.

Assumption A4. $\rho_T \to 0$ and $\delta_T^{-1}\rho_T \to \infty$ as $T \to \infty$.

We can prove the following result.

Theorem 3.2. *Suppose that Assumptions A1–A2 and the first two parts of Assumption A3 hold. Suppose λ in Λ satisfies the last condition on λ in Assumption A3. Then $P\left(\hat{m}_{\hat{\lambda}} = m^0\right) \to 1$ as $T \to \infty$.*

Proof. It suffices to prove the theorem by showing (i) $P\left(\hat{m}_{\hat{\lambda}} < m^0\right) \to 0$ as $T \to \infty$ and (ii) $P\left(\hat{m}_{\hat{\lambda}} > m^0\right) \to 0$ as $T \to \infty$. The proofs of (i) and (ii) are analogous to those of Theorems 3.3 and 3.4 in QS and thus omitted. ■

Theorem 3.2 ensures the GFL to yield the correct number of breaks w.p.a.1. Note that with the data-driven choice of the tuning parameter, $\hat{\lambda}$, the results in Theorem 3.1 continue to hold. Given the accurate estimates of the break dates, we then apply the post-Lasso GMM procedure to estimate the regime-specific parameters. As remarked above, as long as the condition $\delta_T = o(T^{-1/2})$ is ensured, these post-Lasso GMM estimators of the regime-specific parameters are asymptotically equivalent to the corresponding GMM estimators obtained in the case of known true break dates.

Table 1
Simulation results for the no-break case: % of falsely detecting breaks when $m = 0$.

b	σ	T	GFL			supLM			avgLM			expLM		
			$a = 0.8$	0.5	0.2	0.8	0.5	0.2	0.8	0.5	0.2	0.8	0.5	0.2
0		100	8.6	1	0	2	0.6	0.6	3.2	1	2.2	2.6	1.4	1.8
	0.5	200	1.8	0	0	2.2	3.2	3.2	2.6	5.6	4.8	2.8	5.4	4.6
		500	0.4	0	0	4.6	4.8	4.6	5.8	5.2	4.6	5.8	6.2	4.2
		100	8	0.6	0.6	2.2	0.8	1.4	3.4	3.4	4.6	2	2.2	3
	1	200	2.6	0	0	2.6	2.8	2.8	3.2	5.4	4.2	2.4	4	4.2
		500	0	0	0	2.8	3.8	4.2	4.2	4.8	6.6	4.4	4.4	5.2
0.5		100	16.2	3.4	1.6	1.8	1.8	2.6	4	5.4	4.8	2.6	4.2	4
	0.5	200	5.8	0.4	0	2.8	2.8	1.6	3.6	4.6	4.6	4.4	5.6	4.4
		500	0.2	0	0	3.6	5.2	3.2	6.2	5.2	4.6	5.4	6.4	4
		100	16.6	2.8	1.8	1	0.2	1.4	5	2.6	4.4	3.2	1.2	3.6
	1	200	6	0.8	0.8	3	2.4	2.2	4.4	5.4	6.2	2.8	4.6	5.6
		500	0.2	0.2	0	2.6	5.4	4.2	4.2	5.8	6.8	3.2	5.8	6.8
0.8		100	16.8	5	2	1.8	0.8	0.4	5.8	3.6	4.4	4	1.8	3
	0.5	200	6.2	0.6	0	3	3.4	3.2	5.6	5	5.8	4.4	5.4	6
		500	0.4	0	0	4.6	3.6	4	5.4	4.2	4.2	5.8	3.6	4.6
		100	16.8	6	2	1.6	2.2	2	3.6	7.4	5	2.2	5.2	3.8
	1	200	4.6	1.2	1	1.8	2.8	4.2	5	5.2	7.2	4.2	4.2	6.2
		500	0.4	0	0	3.8	5.8	3.6	4.6	7	5.4	4.4	7	5.2

4. Simulations

In this section we conduct a small set of Monte Carlo simulations. We generate the data $\{y_t, x_t, t = 1, \ldots, T\}$ as follows

$$y_t = \beta_t + \beta_t x_t + \sigma u_t,$$
$$x_t = \frac{1}{(1+\rho^2)^{1/2}}(\rho u_t + \xi_t),$$
$$u_t = v_t + b v_{t-1}, \quad \text{with } v_t \sim \text{i.i.d. } N(0, 1/(1+b^2)),$$
$$\xi_t = a\xi_{t-1} + \epsilon_t, \quad \text{with } \epsilon_t \sim \text{i.i.d. } N(0, 1-a^2),$$
$$z_t = (\xi_t + \eta_t)/\sqrt{2}, \quad \text{with } \eta_t \sim \text{i.i.d. } N(0, 1),$$

where y_t is the dependent variable, $(1, x_t)'$ is the regressor vector, and σu_t is the error term, $\{v_t\}$, $\{\epsilon_t\}$ and $\{\eta_t\}$ are mutually independent. Apparently, u_t is a moving average of order one (MA (1)) process with the MA coefficient b, and the constant ρ controls the correlation between x_t and u_t. z_t is a valid IV for x_t, which is only used in the post Lasso GMM estimation and Andrews' GMM-based tests, to which we compare our GFL procedure.

For the no-break case, we let $\beta_t = 1$. For a one-time break, we let $\beta_t = \mathbf{1}\{t \le T/2\}$, where $\mathbf{1}\{\cdot\}$ is an indicator function. For the case of two breaks, we let $\beta_t = 1 - \mathbf{1}\{\lfloor T/3 \rfloor < t \le \lfloor 2T/3 \rfloor\}$, where $\lfloor \cdot \rfloor$ denotes the integer part of $\cdot$. For the GMM-based tests, we consider supLM in Andrews (1993), avgLM and expLM in Andrews and Ploberger (1994), all of which test for an unknown one-time break in the alternative hypothesis. We use a trimming size of 15% in the calculation of each statistic and the 5% asymptotic critical value for each test. If the no-break null hypothesis is rejected, the break date that maximizes the LM statistic is taken to be the estimated break date. We set $\rho = 1/2$, corresponding to $\text{corr}(x_t, u_t) \approx 0.45$. And we let $T = 100$, 200, or 500, and $\sigma = 0.5$ or 1, corresponding to a signal-to-noise ratio (in the no-break case) of 4 or 1, respectively. We follow QS and choose the tuning parameter λ to minimize the IC in (3.1) with $\rho_T = T^{-1/2}$. The number of repetitions is 500.

Table 1 shows the percentages of falsely detecting breaks when no break exists ($m^0 = 0$) for GFL and the empirical size of the supLM, avgLM, and expLM tests. GFL enjoys small probabilities of false detection of breaks when the persistence level of x_t is low or moderate ($a = 0.2$ or 0.5), even when the sample size is relatively small. When x_t is highly persistent ($a = 0.8$), however, we see relatively large probability of false detection when the sample size is small. As the sample size increases, the probability of false detection declines rapidly to nearly zero. When the error is serially correlated ($b = 0.5$ or 0.8), we witness a larger probability of false detection, especially when at the same time x_t is highly persistent. In large samples (say, $T = 500$), the performance of GFL does not appear sensitive to serial correlation in the error. In comparison, the GMM-based tests generally have good size performances, especially in large samples.

Table 2 tabulates the probabilities of correctly identifying one break when it indeed exists, and Table 3 reports the mean absolute error of the estimated break ratio multiplied by 100, conditional on detecting one break. Note that we do not report results on parameter estimation, since once break dates are determined, one can apply standard GMM to estimate the regime-specific parameters. The behavior of parameter estimators depends crucially on the accuracy of break-date estimation as well as the quality of IVs. For a better comparison between our approach and the GMM-based tests, the empirical powers for the latter are size-corrected. We may conclude from the results: (i) GFL generally enjoys a high probability of correctly detecting the break. Its performance is only slightly affected by the increased persistence in x_t or the increased serial correlation in the error. (ii) When the sample size is relatively small, GFL substantially outperforms GMM-based tests. The empirical powers of the latter are sensitive to the persistence level of x_t in small samples. avgLM and expLM generally have better power performance than supLM. (iii) Once a break is detected, GFL also yields more accurate estimation of break-dates, as shown in Table 3.

A distinguishing characteristic of our method is that we directly allow more than one breaks. This is born out in Table 4, which shows results for the case where two breaks exist in the DGP. The left half of the table gives the percentages of correctly detecting two breaks, and the right half shows the average Hausdorff distance[2]between the estimated and the true sets of break dates in the percentages of T (i.e., $\overline{\text{HD}}(\hat{\mathcal{T}}, \mathcal{T})/T \cdot 100$). Note that the average Hausdorff errors are comparable to the mean absolute error in Table 3, since the Hausdorff distance between two singleton sets reduces to the absolute error. Here we set $b = 0.5$ and experiment on different values of ρ, which controls the correlation between x_t and u_t, in addition to variation in the noise level (σ) and

[2] The Hausdorff distance between any two sets A and B is defined as $\text{HD}(A, B) = \max\{\mathcal{D}(A, B), \mathcal{D}(B, A)\}$, where $\mathcal{D}(A, B) \equiv \sup_{b \in B} \inf_{a \in A} |a - b|$.

Table 2
Simulation results for the one-break case: % of correctly detecting one break.

b	σ	T	GFL			supLM			avgLM			expLM		
			$a = 0.8$	0.5	0.2	0.8	0.5	0.2	0.8	0.5	0.2	0.8	0.5	0.2
0	0.5	100	96.4	99.4	100	79.8	98.2	99.8	91.6	99	99.6	91.2	99.2	99.8
		200	98.2	100	100	100	100	100	99.8	100	100	100	100	100
		500	99.2	99.8	100	100	100	100	100	100	100	100	100	100
	1	100	89	94.6	95.8	64	79.6	87.8	71.8	87.2	89.4	74.8	89.6	91.8
		200	94.8	97	99	99.2	99.4	99.8	99.4	98.8	100	99.4	99.4	100
		500	96.2	97.6	99.4	100	100	100	100	100	100	100	100	100
0.5	0.5	100	95.6	98.6	99.4	76	87.4	97.4	83.4	90.8	98.4	82	96	99.4
		200	97.8	99.8	100	99.4	100	100	99.4	100	100	99.6	100	100
		500	99	100	99.6	100	100	100	100	100	100	100	100	100
	1	100	84	92.8	93.6	45.6	65	71	55.8	75.2	81.6	57	76.4	80.8
		200	93	96.2	97.6	91.2	98	98.2	94.6	99	98.4	95.2	99	98.8
		500	95.4	97.2	97.6	100	100	100	100	100	100	100	100	100
0.8	0.5	100	94.2	98.4	99.8	69.2	90.6	98.6	79	95.6	97.4	82.6	95.8	99
		200	97.8	99.6	100	99.4	100	100	99.4	100	100	100	100	100
		500	99.4	99.8	100	100	100	100	100	100	100	100	100	100
	1	100	86.8	91.4	95.2	45.6	50.4	67	62.6	57	76.8	60.6	57.8	77.4
		200	89.8	96.4	96.6	91.4	95.8	97.4	92	97.2	97.2	94	97.4	98
		500	95.8	97.8	97.4	100	100	100	100	100	100	100	100	100

Table 3
Simulation results for the one-break case: mean absolute errors of estimated break ratios $\times 100$, conditional on $\hat{m}_\lambda = 1$.

b	σ	T	GFL			supW		
			$a = 0.8$	0.5	0.2	0.8	0.5	0.2
0	0.5	100	1.1286	0.6439	0.592	5.5639	3.6069	2.1483
		200	0.5316	0.296	0.295	2.634	1.195	0.825
		500	0.1802	0.1134	0.1124	1.0536	0.4108	0.3096
	1	100	3.2292	1.9514	1.4489	8.5188	5.9799	5.0023
		200	1.9599	0.9979	0.798	4.373	2.6207	1.983
		500	0.6491	0.3947	0.3247	1.44	0.8144	0.7096
0.5	0.5	100	1.2427	0.7627	0.5674	5.7132	3.6682	3.1663
		200	0.5706	0.3858	0.305	2.993	1.441	0.988
		500	0.2311	0.1416	0.1133	1.03	0.564	0.3624
	1	100	3.8119	2.4159	2.0513	9.114	8.3692	6.8028
		200	2.1441	1.1164	0.9426	5.9901	3.7469	3.4582
		500	0.7962	0.5103	0.3643	2.1796	1.0752	0.9908
0.8	0.5	100	1.293	0.7805	0.6232	5.8613	3.6336	2.6673
		200	0.5818	0.3303	0.302	3.4688	1.675	1.025
		500	0.2016	0.1343	0.1208	1.1836	0.4768	0.3904
	1	100	3.7327	2.5689	2.3298	10.7368	6.6746	6.991
		200	2.3196	1.2604	0.9617	5.895	3.9593	3.3419
		500	0.8526	0.4953	0.4296	2.062	1.2272	1.0712

Table 4
Simulation results for the two-break case: % of detecting two breaks ($\hat{m}_\lambda = 2$) and break-date estimation error in terms normalized Hausdorff distance (conditional on $\hat{m}_\lambda = 2$).

σ	T	% of $\hat{m} = 2$						$\overline{\mathrm{HD}}(\hat{\mathcal{T}}, \mathcal{T})/T \cdot 100$					
		$\rho = 1/2$			$\rho = 1$			$\rho = 1/2$			$\rho = 1$		
		$a = 0.8$	0.5	0.2	0.8	0.5	0.2	0.8	0.5	0.2	0.8	0.5	0.2
0.5	100	88.6	98.6	99	91.2	99.8	99.8	2.10	1.41	1.14	1.19	0.92	0.83
	200	96.6	99	99.8	97.6	99.8	100	1.12	0.66	0.51	0.57	0.45	0.41
	500	99.2	99.6	99.6	99.4	100	100	0.31	0.22	0.19	0.21	0.15	0.16
1	100	49	53.8	52.4	71	80.4	85.2	4.36	3.03	2.52	4.05	2.30	2.08
	200	70	77.6	81.6	85	94.6	95.2	2.81	1.71	1.52	1.99	1.30	1.06
	500	89	94.2	94.8	95.4	96.2	98.4	1.12	0.63	0.59	0.88	0.54	0.47

the persistence level of x_t. We find that as in the one-break case, the probability of correctly estimating the number of breaks approaches one as T increases. At the lower noise level, GFL performs well even when the sample size is small. Finally, in this particular DGP, the increased correlation between x_t and u_t leads to better performance in small samples, in terms of higher percentage of correctly estimating the number of breaks and accuracy of break-date estimation.

5. Empirical application to regime shifts in the monetary policy rule

In monetary economics, the celebrated "Taylor's rule" (Taylor, 1993) may be formulated as follows,

$$r_t = \bar{r} + \beta(\pi_t - \pi^*) + \gamma y_t + \varepsilon_t,$$

where r_t is the policy rate (e.g., federal funds rate), π_t is the rate of inflation, π^* is the inflation target rate, y_t is the GDP gap, $\bar{r}$ is the

Table 5
Summary statistics (sample size =215).

Variables	GDP gap	Fed funds rate	Inflation	Inflation in PPI	M2 growth	Spread
Mean	−0.8925	5.5179	3.3836	3.4483	6.6566	1.4859
Std	2.826	3.4987	2.3448	7.0203	3.3776	1.2554
Min	−8.063	0.07	−0.668	−47.612	−1.205	−2.65
Max	6.202	17.78	11.61	27.7	21.441	4.42

Table 6
Empirical results: estimated regimes of the US monetary policy rule.

$\hat{m}_\lambda$	Regime	α	β	γ	$\bar{r}$	p value	IC
	1960Q3–1979Q2	0.9095*	0.9419***	0.2458***	2.7933***		
2	1979Q3–1991Q4	4.8625***	1.0275***	−0.1875**	6.9175***	0.4168	1.4508
	1992Q1–2012Q3	5.0775***	−0.3042	0.6953***	4.4691***		

Note: $\bar{r}$ is calculated assuming $\pi^* = 2$.
* Denote significance at 10% level.
** Denote significance at 5% level.
*** Denote significance at 1% level.

"natural" policy rate, and ε_t is a random error. It is now a widely accepted principle that the central bank can stabilize the economy by adjusting the policy interest rate more than one-to-one with inflation ($\beta > 1$). Many authors including Taylor (1999) and Clarida et al. (2000) have argued that the violation of this principle can be blamed for the macroeconomic instabilities in the 1960s and 1970s and that the "Great Moderation" since 1980s is partly due to the "good" monetary policy that satisfies the principle.

It is thus a plausible conjecture that there are structural changes in the evolution of the US monetary policy rule, assuming that the simplified rule is an acceptable characterization of the US monetary policy history. In this empirical exercise we study the period from 1960Q1 to 2012Q4. We assume that the US monetary policy follows a time-varying forward-looking rule,

$$r_t = \bar{r}_t + \beta_t(E_t\pi_{t+1} - \pi_t^*) + \gamma_t E_t y_{t+1} + \varepsilon_t, \tag{5.1}$$

where E_t represents expectation conditional on information available at time t, and where the coefficients (β_t and γ_t), the target inflation (π_t^*), and the "natural" policy rate ($\bar{r}_t$) are all allowed to be time-varying, but only in the form of sparse breaks. Breaks may occur because of the appointment of a new Fed chairman or the change in political and economic conditions. We assume that the policy rule is stable between two consecutive break points and that breaks are sparse relative to the time period we consider. This is in contrast to Kim and Nelson (2006), who assume that the coefficients behave as a random walk. Furthermore, we take an agnostic view on how one regime may switch to another or how many "states" there are in the evolution of monetary history. This is in contrast to Sims and Zha (2006), who use the Bayesian VAR framework and assume the Markov switching among a finite number of states.

We may rewrite (5.1) as the following regression,

$$r_t = \alpha_t + \beta_t\pi_{t+1} + \gamma_t y_{t+1} + u_t, \tag{5.2}$$

where $\alpha_t = (\bar{r}_t - \beta_t\pi_t^*)$ and $u_t = \varepsilon_t - \beta_t(\pi_{t+1} - E_t\pi_{t+1}) - \gamma_t(y_{t+1} - E_t y_{t+1})$. It is clear that both regressors in (5.2) may be correlated with the error u_t. This justifies the use of IV's in the post-lasso GMM estimation.

5.1. Data

The data are quarterly time series spanning the period 1960Q1–2012Q4. We obtain the potential GDP data from the Congressional Budget Office (CBO) and the remaining data from the Federal Reserve Economic Data (FRED), whose mnemonics we follow. The policy rate is taken to be the US federal funds rate (FEDFUNDS). The rate of inflation is measured by the annualized rate of change of the GDP deflator (GDPDEF) between two subsequent quarters, as in Taylor (1993) and Clarida et al. (2000). The GDP gap is calculated by $y = 100(\text{Actual GDP}/\text{Potential GDP} - 1)$. We use as IV's the lagged observations of inflation, GDP gap, inflation in producer's price index (PPIACO), M2 growth (M2SL), and the "spread" between the 10-year bond rate (GS10) and the 3-month Treasury Bill rate (TB3MS). The federal funds rate, inflation, inflation in PPI, and M2 growth are all taken to be the quarterly averages. The interest rate spread is taken to be the end-of-period observations. All numbers are in percentage terms. Table 5 shows the summary statistics of the data.

5.2. Empirical results

Table 6 summarizes the empirical results based on Hansen's (1982) optimal GMM estimation for the estimated three regimes. Note that the time span starts from 1960Q3, since at most two lags are used in IV, and ends at 2012Q3, since regressors are one-quarter-ahead. In addition to the estimated parameters in (5.2), we also calculate the "natural" policy rate $\bar{r}_t$ in (5.1), assuming that π_t^* is a constant $\pi^* = 2$ for identification.

We identify two breaks, or three regimes, in the evolution of the US monetary policy since 1960. The first break occurs in 1979Q3, obviously relating to the appointment of Paul Volcker as the FR chairman in August 1979. The second break occurs in 1992Q1, five years into Alan Greenspan's tenure as the FR chairman. The post-Lasso over-identification test points to a p-value of 0.4168, indicating that the moment conditions (i.e., the choice of IV's) are valid.

The regime shift of monetary policy at 1979 is well studied in the literature (e.g., Taylor, 1999 and Clarida et al., 2000). In contrast to the received wisdom that the success of the new regime is solely due to a more than one-to-one response of policy rate to inflation, our results suggest that the dramatic change in the monetary authority's stance on the "natural" policy rate should be credited. While the results do confirm that $\beta_t < 1$ in the pre-Volcker regime and $\beta_t > 1$ after the break, the change in β_t is far less dramatic than the change in $\bar{r}_t$, which is from 2.7933 to 6.9175, or from 0.7933 to 4.9175 in real terms. The second break at 1992Q1 is much less documented in the literature. It may be connected with the dramatic transformation of the US banking industry after the financial deregulation in the 1980s or the end of 12-year Republican rule in the White House. Compared with the second regime (1979Q3–1991Q4), the third regime (1992Q1–2012Q3) is less hawkish to inflation and much more sensitive to GDP gap. What is also evident is the loosening of the monetary authority's stance on the "natural" policy rate during this last period.

Acknowledgments

The authors would like to thank the editor Badi H. Baltagi and an anonymous referee for helpful comments. They gratefully acknowledge the Singapore Ministry of Education for Academic Research Fund under Grant number MOE2012-T2-2-021.

References

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

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

Bai, J., Perron, P., 1998. Estimating and testing liner models with multiple structural changes. Econometrica 66, 47–78.

Bleakley, K., Vert, J.-P., 2011. The group fused Lasso for multiple change point detection. Working Paper, INRIA Saclay, Orsay, France.

Boldea, O., Hall, A.R., Han, S., 2012. Asymptotic distribution theory for break point estimators in models estimated via 2SLS. Econometric Rev. 31, 1–33.

Clarida, R., Galí, J., Gertler, M., 2000. Monetary policy rules and macroeconomic stability: evidence and some theory. Quart. J. Econ. 115, 147–180.

Hall, A.R., Han, S., Boldea, O., 2012. Inference regarding multiple structural changes in linear models with endogenous regressors. J. Econometrics 170, 281–302.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Kim, C.J., Nelson, C.R., 2006. Estimation of a forward-looking monetary policy rule: a time-varying parameter model using ex post data. J. Monetary Econ. 53, 1949–1966.

Kolar, M., Song, L., Xing, E.P., 2009. Sparsistent learning of varying-coefficient models with structural changes. Adv. Neural Inf. Process. Syst. 22, 1006–1014.

Perron, P., Yamamoto, Y., 2013. Using OLS to estimate and test for structural changes in models with endogenous regressors. J. Appl. Econometrics http://dx.doi.org/10.1002/jae.2320.

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.

Qian, J., Su, L., 2014. Shrinkage estimation of regression models with multiple structural changes. Working Paper, School of Economics, Singapore Management University.

Sims, C.A., Zha, T., 2006. Were there regime switches in US monetary policy? Amer. Econ. Rev. 96, 54–81.

Taylor, J.B., 1993. Discretion versus policy rules in practice. In: Carnegie–Rochester Conference Series on Public Policy, vol. 39, pp. 195–214.

Taylor, J.B., 1999. A historical analysis of monetary policy rules. In: Taylor, John. B. (Ed.), Monetary Policy Rules. In: NBER Conference Report Series, University of Chicago Press, Chicago, pp. 319–341.