Power Plots

Curtis Miller

July 19, 2019

The plots below are power simulations comparing the ability of different test statistics to detect structural change in linear regression models. Here, $\beta = (1, 2)^T$, so the regression model is $y_t = 1 + 2x_t + \epsilon_t$, where $\{\epsilon_t\}_{t\in\mathbb{N}}$ is a mean-zero stationary process and $x_t \sim \mathrm{N}(1, 1)$, when the null hypothesis is true.

The test statistics displayed include the Rényi-type statistic with trimming parameter $t_T = T^{1/2}$ computed with the regression model coefficients directly (———) the Rényi-type statistic with trimming parameter $t_T = T^{1/2}$ computed on the residuals (– – – –), the CUSUM statistic (— — –), and the Hidalgo-Seo statistic (-·-·-·-·). The Rényi-type statistic and CUSUM statistics are the univariate statistics computed on the estimated residuals of the regression model. The Hidalgo-Seo statistic was tailored to the regression context as described in their paper.

These statistics do not all use the same LRV procedure. The Hidalgo-Seo statistic uses the periodogram matrix for LRV estimation. The univariate statistics use the procedure described in [1], which uses the full sample but allows for changes in the mean of the univariate data. The regression-coefficient-based Rényi-type statistic uses either the long-run variance estimation procedure described in [2] with kernel methods (in the case of the ARMA(2, 2) process) or a procedure that does not use kernel methods (for all other cases). The kernel used in the LRV estimation procedures for all statistics using kernel methods (the Rényi-type and CUSUM statistics) is the quadratic spectral kernel, and the bandwidth was selected using the procedure advocated by [3] and implemented in [4].

Two different change point regimes are implemented. The first is an early change point context. For a sample size T, the change point occurs near $T^{3/5}$ if there is a change at all. If it does, the new model the data follows is $y_t = (1 + \delta) + (2 - \delta)x_t + \epsilon_t$, where $\delta \in [-2, 2]$. The second regime has the change occur near $\frac{T}{10}$.

In this document, ϵ_t is either homoskedastic or heteroskedastic. If ϵ_t is homoskedastic, it follows one of four DGPs: independent and identically distributed $N(0, 1)$ random variables; an AR(1) process with autocorrelation coefficient $\phi = \sqrt{0.5} \approx 0.71$; an ARMA(2, 2) process with $\phi_1 = 0.4$, $\phi_2 = -0.03$, $\theta_1 = 0.5$, $\theta_2 = 0.06$, and the variance of the noise parameter is chosen so that the long-run variance is one; and a GARCH(1, 1) process with $\omega = 0.2$, $\alpha = 0.2$, and $\beta = 0.5$ (again the long-run variance is one). The proportional change point is investigated only when the residuals are *i.i.d.* Normal.

If ϵ_t is heteroskedastic, then it follows one of the four DGPs followed by one of the homoskedastic versions for the first half of the sample. In the second half of the sample, the model switches. In all cases, the LRV scales from 1 to 10. If the data was *i.i.d.* Normal or AR(1), that's all that changes. In the ARMA(2, 2) and GARCH(1, 1) cases, though, the structure of the dependency on past values changes as well, in addition to the LRV change. For the ARMA(2, 2) case, the post-change process has parameters $\phi_1 = 0.4$, $\phi_2 = 0.2$, $\theta_1 = -0.1$, $\theta_2 = -0.42$. For the GARCH(1, 1) case, the post-change process has parameters $\omega = 1$, $\alpha = 0.7$, $\beta = 0.2$. (The proportional change point was not investigated in the heteroskedastic case.)

For every combination of sample size, test statistic, δ, and data-generating process for ϵ_t, 20,000 independent test statistics were simulated, and their asymptotically-appropriate p-values were computed. H_0 is rejected for p-values less than $\alpha = 0.05$; the value of α is signified with a dotted line (··········). Values of δ increment by 0.1, and the sample sizes considered are $T \in \{50, 250, 1000\}$.

In the case of homoskedastic residuals, the best-performing test statistic in all plots is the Hidalgo-Seo statistic, but the statistic also always suffers from considerable size inflation even for large sample sizes. The univariate statistics (the CUSUM statistic and the univariate Rényi-type statistic) do not demonstrate good power properties; sometimes they experience size inflation, they have bad power for large sample sizes, and in the case of the ARMA-type residuals models they even have power curves with the wrong shape. The CUSUM statistic is the worst performing while the univariate Rényi-type statistic sometimes has decent power for large sample sizes. On balance, though, the regression-based Rényi-type often has excellent power without suffering from size inflation; in fact, for the ARMA-type residuals,

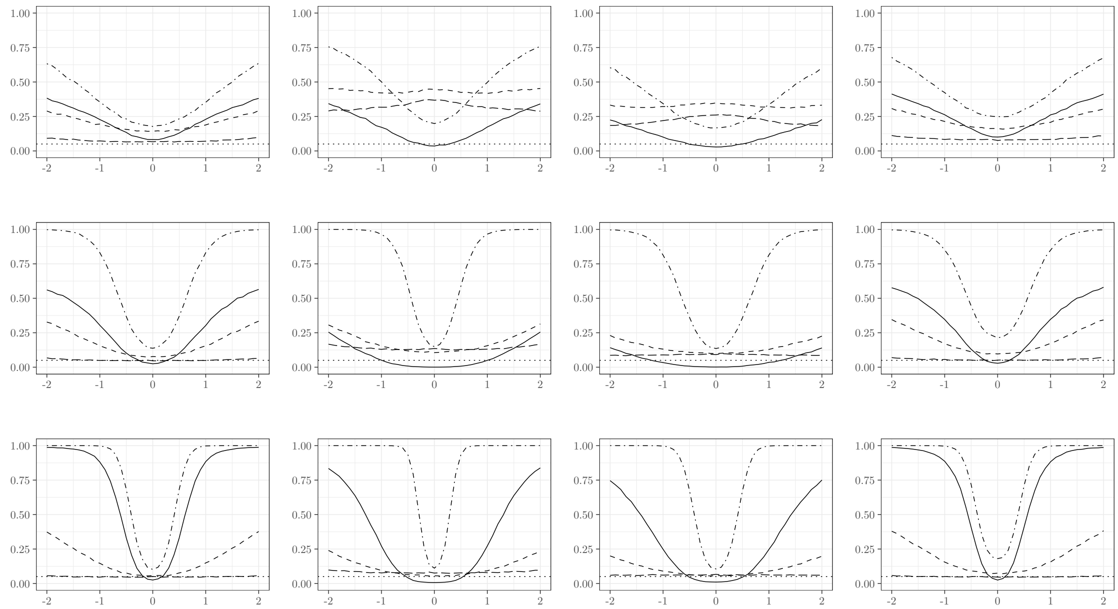

Figure 1: Power plots for when the change occurs near $T^{3/5}$; rows correspond to the sample sizes 50, 250, and 1000 respectively, the first column is for *iid* Normal residuals, the second column AR(1) residuals, the third column ARMA(2, 2) residuals, and the fourth column GARCH(1, 1) residuals. The x-axis tracks values of δ and the y-axis tracks observed power.

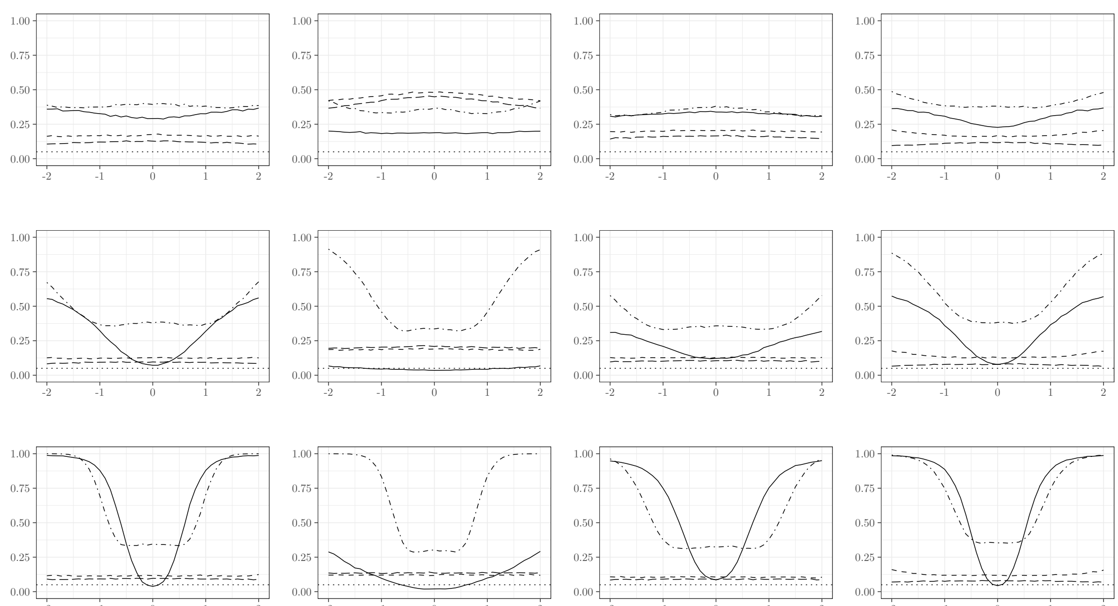

Figure 2: Power plots for change with heteroskedasticity in residuals. The layout is similar to Figure 1.

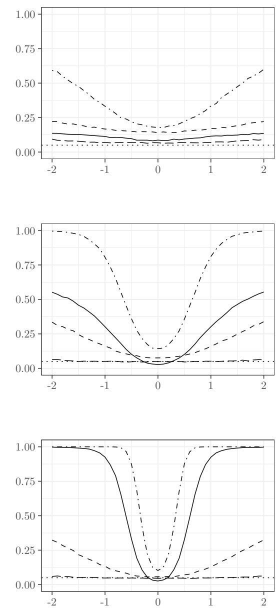

Figure 3: Power plots for when the change occurs near $\frac{T}{10}$; see Figure 1 for description.

the statistic (using kernel methods for LRV estimation) has a smaller size than specified.

In the case of heteroskedastic residuals, the univariate statistics have effectively no power. The Hidalgo-Seo statistic has some power at the cost of massive size inflation. All statistics struggle at smaller sample sizes, but the Rényi-type statistic shines for larger sample sizes, being both appropriately sized and having good power, sometimes even beating the ill-sized Hidalgo-Seo statistic.

Now consider a five-variable regression model $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t} + \epsilon_t$. We wish to see how well the methods considered here can detect changes in this higher-dimensional model. Under the null hypothesis, $\beta = (1, 2, 3, 4, 5, 6)^T$, while under the alternative, $\beta = ((1+\delta), (2-\delta), (3+\delta), (4-\delta), (5+\delta), (6-\delta))^T$, with $\delta \in [-2, 2]$ after the change point. The setup is otherwise the same (including $x_{jt} \sim \mathrm{N}(1, 1)$), but we only consider Normal error terms.

The results are shown in Figure 4. Higher dimension does appear to have an effect on the power of all the statistics considered. Power appears to be worse and often has undesirable properties; all statistics for a sample size of 50 are effectively junk. In all cases, the size inflation of the Hidalgo-Seo statistic appears to have become even worse. But when the sample size is 250, the new statistic appears to have desirable power properties, and for a sample size of 1000 the statistic performs well. These plots should serve as a warning against attempting to detect changes in higher-dimensional regression models when the sample size is not large.

References

[1] L. Horváth, C. Miller, and G. Rice, "A new class of change point test statistics of Rényi type," *Journal of Business & Economic Statistics*, vol. 0, no. 0, pp. 1–19, 2019.

[2] L. Horváth, C. Miller, and G. Rice, "Detecting early or late changes in linear models." draft.

[3] D. W. K. Andrews, "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, vol. 59, pp. 817–858, May 1991.

[4] P. Aschersleben and M. Wagner, *cointReg: Parameter Estimation and Inference in a Cointegrating Regression*, 2016. R package version 0.2.0.

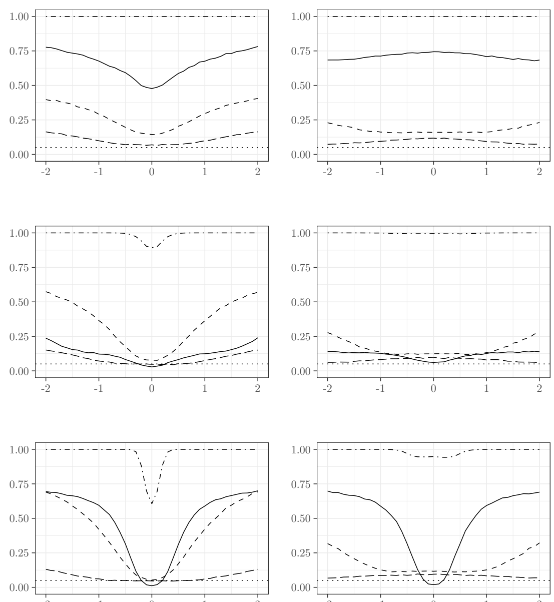

Figure 4: Power plots for change in a five-variable regression model; see Figure 1 for description.