Regression Kink with an Unknown Threshold

Bruce E. Hansen*
University of Wisconsin†

February 2015
Revised: June 2015

Abstract

This paper explores estimation and inference in a regression kink model with an unknown threshold. A regression kink model (or continuous threshold model) is a threshold regression constrained to be everywhere continuous with a kink at an unknown threshold. We present methods for estimation, to test for the presence of the threshold, for inference on the regression parameters, and for inference on the regression function. A novel finding is that inference on the regression function is non-standard since the regression function is a non-differentiable function of the parameters. We apply recently developed methods for inference on non-differentiable functions. The theory is illustrated by an application to the growth & debt problem introduced by Reinhart and Rogoff (2010), using their long-span time-series for the United States.

*Research supported by the National Science Foundation. I thank the Associate Editor, two referees, Zheng Fang, Han Hong, Jessie Li, Tzu-Chi Lin, and Andres Santos for very helpful comments, suggestions, and correspondence which were critical for the results developed here.

†Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706.
behansen@wisc.edu

1 Introduction

The regression kink model was popularized by Card, Lee, Pei and Weber (2012) as a modification of the regression discontinuity model. In the regression kink model, the regression function is continuous but the slope has a discontinuity at a threshold point, hence a "kink". This model has gained much empirical attention, including applications by Landais (2014) and Ganong and Jager (2014). The traditional regression discontinuity model assumes that the threshold is known, but in some cases (such as Card, Mas and Rothstein, 2008) it is unknown and must be estimated. Our paper concerns this latter case of an unknown threshold.

Regression discontinuity models are similar to threshold regression models. The latter were introduced by Howell Tong (1983, 1990) in the context of nonlinear autoregression, but can be applied to many nonlinear regression contexts. Most of the literature and methods focus on the discontinuous (unconstrained) threshold model, where the regression model is split into two (or more) groups based on a threshold indicator. A notable exception is the continuous threshold model introduced by Chan and Tsay (1998), which is identical to a regression kink model with piecewise linear regression segments. One economic application of the continuous threshold model is Cox, Hansen, and Jimenez (2004). The regression kink model may be appealing for empirical applications where the threshold effect focuses on one variable and there is no reason to expect a discontinuous regression response at the threshold.

This paper extends the theory of Chan and Tsay (1998), considering the problems of testing for a threshold effect, inference on the regression parameters, and inference on the regression function. As in Chan and Tsay (1998) we confine attention to the context where the regression segments are linear rather than nonparametric. This is appropriate in contexts of moderate sample sizes where nonparametric methods may be inappropriate.

There is a large literature on discontinuous threshold regression. For the problem of testing for a threshold effect, relevant contributions include Chan (1990, 1991), Chan and Tong (1990), Hansen (1996) and Lee, Seo and Shin (2011). For inference on the regression parameters, relevant papers include Chan (1993), Hansen (2000), and Seo and Linton (2007). Panel data methods have been developed by Hansen (1999) and Ramirez-Rondan (2013). Instrumental variable methods have been developed by Caner and Hansen (2004). An estimation and inference theory for regression discontinuity with unknown thresholds has been developed by Porter and Yu (2014).

In the statistics literature, related classes of models include two-phase regression, segmented regression, the broken stick model, and the bent cable model. Important contributions to this literature include Hinkley (1969, 1971) and Feder (1975). See also the references in Chiu, Lockhart, and Routledge (2006).

For illustration, we apply the regression kink model to the growth & debt problem made famous by Reinhart and Rogoff (2010). These authors argued that there is a nonlinear effect of aggregate debt on economic growth, specifically that as the ratio of debt to GDP increases above some threshold, aggregate economic growth will tend to slow. This can be formalized as a regression kink model, where GDP growth is the dependent variable and the debt/GDP ratio is the key regressor

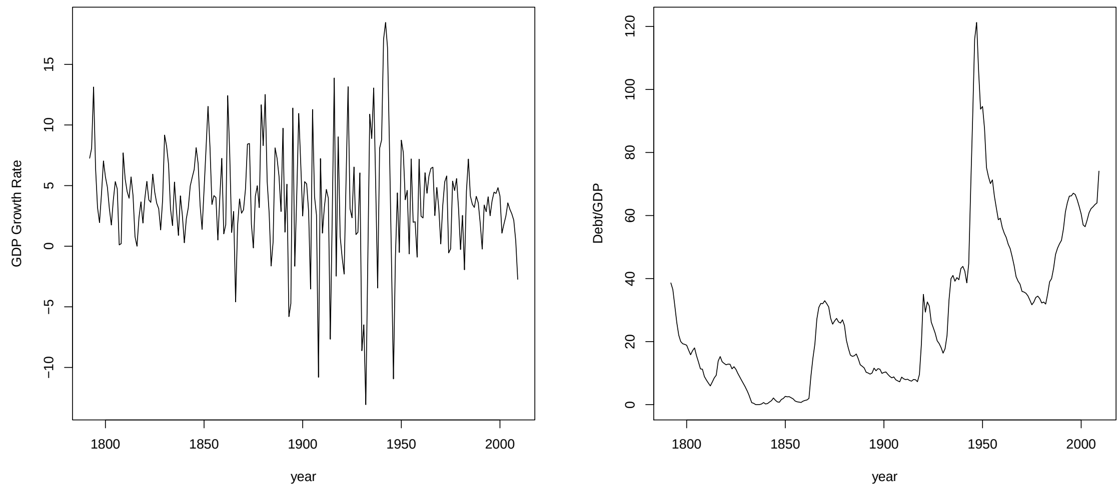


Figure 1: Annual U.S. Real GDP Growth Rate and GDP/Debt Ratio, 1791-2009

and threshold variable. Econometric analysis of their proposition using threshold regression tools has been pursued by several authors, including Caner, Grennes and Koehler (2010), Cecchetti, Mohanty, and Zampolli (2011), Lin (2014), and Lee, Park, Seo and Shin (2014). These papers investigate the Reinhart-Rogoff proposition using a variety of cross-section, panel, and time-series regressions, but all have focused on the discontinuous threshold regression model. We add to this literature by a small investigation using long-span U.S. time-series data.

During our investigation we encounter one novel technical issue. While the parameter estimates in the regression kink model are asymptotically normal (as shown by Chan and Tsay (1998)), the estimates of the regression function itself are not asymptotically normal, since the regression function is a non-differentiable function of the parameter estimates. Consequently, conventional inference methods cannot be applied to the regression function. To address this issue we use recently developed inference methods by Fang and Santos (2014) and Hong and Li (2015), and present a new non-normal distribution theory for the regression function estimates, and shown how to construct numerical delta method bootstrap confidence intervals for the regression function.

Our organization is as follows. Section 2 introduces the model. Section 3 describes least-squares estimation of the model parameters. Section 4 discusses testing for a threshold effect in the context of the model. Section 5 presents an asymptotic distribution theory for the parameter estimates and discusses confidence interval construction. Section 6 discusses inference on the regression function. A formal proof of Theorem 2 is presented in the Appendix.

The data and R code for the empirical and simulation work reported in the paper is available on the author's webpage http://www.ssc.wisc.edu/~bhansen/

2 Model

Our regression kink model takes the form

$$y_t = \beta_1(x_t - \gamma)_- + \beta_2(x_t - \gamma)_+ + \beta_3' z_t + e_t \tag{1}$$

where y_t, x_t and e_t are scalars, and z_t is an ℓ-vector which includes an intercept. The variables (y_t, x_t, z_t) are observed for $t = 1, ..., n$. The parameters are β_1, β_2, β_3, and γ. We use $(a)_- = \min[a, 0]$ and $(a)_+ = \max[a, 0]$ to denote the "negative part" and "postive part" of a real number a. In model (1) the slope with respect to the variable x_t equals β_1 for values of x_t less than γ, and equals β_2 for values of x_t greater than γ, yet the regression function is continuous in all variables.

The model (1) is a regression kink model because the regression function is continuous in the variables x and z, but the slope with respect to x is discontinuous (has a kink) at $x = \gamma$. The model (1) specifies the regression segments to be linear, but this could be modified to any parametric form (such as a polynomial). The conventional regression kink design assumes that the threshold point γ is known. (This is suitable in many policy-oriented applications where the threshold is known and determined by policy.) Instead, we treat the parameter γ as an unknown to be estimated. Thus our methods are appropriate when the threshold is either not determined by the policy, or when the researcher wishes to investigate the robustness of this assumption.

The model (1) has $k = 3 + \ell$ parameters. $\beta = (\beta_1, \beta_2, \beta_3)$ are the regression slopes and are generally unconstrained so that $\beta \in \mathbb{R}^{k-1}$. The parameter γ is called the "threshold", "knot" or "kink point". The model (1) only makes sense if the threshold is in the interior of the support of the threshold variable x_t. We thus assume that $\gamma \in \Gamma$ where Γ is compact and strictly in the interior of the support of x_t.

The regression kink model is nested within the discontinuous threshold model. Therefore one could imagine testing the assumption of continuity within the threshold model class. This is a difficult problem, one to which we are unaware of a solution, and therefore is not pursued in this paper. We simply assume that regression function is continuous and do not explore the issue of testing the assumption of continuity.

As an empirical example, consider the growth/debt regression problem of Reinhart and Rogoff (2010). They argued that economic growth tends to slow when the level of government debt relative to GDP exceeds a threshold. To write this as a regression, we set y_t to be the real GDP growth rate in year t and x_t to the debt to GDP percentage from the previous year (so that it is plausibly pre-determined). We set $z_t = (y_{t-1} \ 1)'$ so that the regression contains a lagged dependent variable to ensure that the error e_t is approximately serially uncorrelated.

We focus on the United States and use the long span time series for the years 1792-2009 gathered by Reinhart and Rogoff and posted on their website, so that there are $n = 218$ observations. We display time-series plots of the two series in Figure 1. We follow Lin (2014) and focus on time-series estimates for a single country rather than cross-country or panel estimation, so to not impose parameter homogeneity assumptions.

In Figure 2 we display a scatter plot of (y_t, x_t) along with the fitted regression line and pointwise 90% confidence intervals. (We will discuss estimation in the next section and confidence intervals in Section 6.) We can see that the fitted regression shows a small positive slope for low debt ratios, with a kink (threshold) around 44% (displayed as the red square), switching to a negative slope for debt ratios above that value.

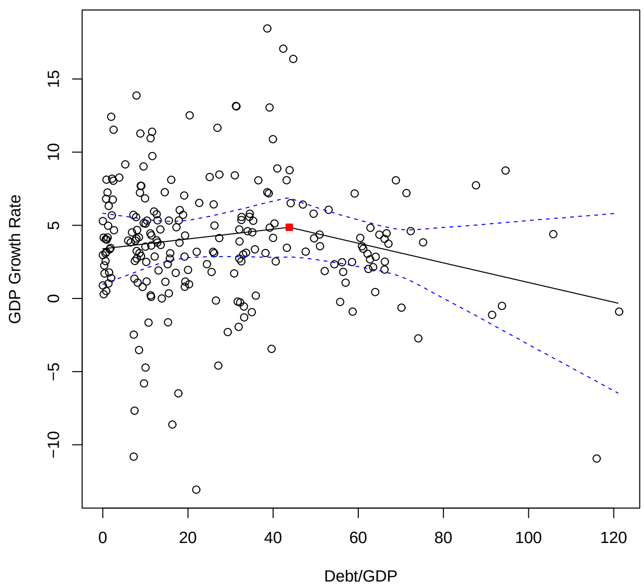


Figure 2: Scatter Plot of Real GDP growth and Debt/GDP, with Estimated Regression Kink Model, and 90% Confidence Intervals

We would like to consider inference in the context of this model, specifically focusing on the following questions: (1) Is the threshold regression statistically different from a linear regression? (2) What is the asymptotic distribution of the parameter estimates, and can we construct confidence intervals for the parameters? (3) Can we construct confidence intervals for the regression line?

3 Estimation

If the model (1) is interpreted as a conditional mean then the natural estimator of the parameters is least-squares. It will be convenient to write

$$x_t(\gamma) = \begin{pmatrix} (x_t - \gamma)_- \\ (x_t - \gamma)_+ \\ z_t \end{pmatrix} \tag{2}$$

so that (1) can be written as $y_t = \beta' x_t(\gamma) + e_t$. The least-squares criterion is

$$S_n(\beta, \gamma) = \frac{1}{n} \sum_{t=1}^{n} \left(y_t - \beta' x_t(\gamma) \right)^2. \tag{3}$$

The least-squres estimator $\left(\widehat{\beta}, \widehat{\gamma}\right)$ is the joint minimizer of $S_n(\beta, \gamma)$:

$$\left(\widehat{\beta}, \widehat{\gamma}\right) = \underset{\beta \in \mathbb{R}^{k-1}, \gamma \in \Gamma}{\operatorname{argmin}} S_n(\beta, \gamma) \tag{4}$$

The criterion function $S_n(\beta, \gamma)$ is quadratic in β but non-convex in γ. Thus it is computationally convenient to use a combination of concentration and grid search, as is typical in the threshold literature. Specifically, notice that by concentration we can write

$$\begin{aligned} \widehat{\gamma} &= \underset{\gamma \in \Gamma}{\operatorname{argmin}} \min_{\beta \in \mathbb{R}^{k-1}} S_n(\beta, \gamma) \\ &= \underset{\gamma \in \Gamma}{\operatorname{argmin}} S_n^*(\gamma) \end{aligned} \tag{5}$$

where $\widehat{\beta}(\gamma)$ are the least-squares coefficients from a regression of y_t on the variables $x_t(\gamma)$ for fixed γ, and

$$S_n^*(\gamma) = S_n\left(\widehat{\beta}(\gamma), \gamma\right) = \frac{1}{n} \sum_{t=1}^{n} \left(y_t - \widehat{\beta}(\gamma)' x_t(\gamma) \right)^2$$

is the concentrated sum-of-squared errors function. The solution to (5) can be found numerically by a grid search over γ. After $\widehat{\gamma}$ is found the coefficient estimates $\widehat{\beta}$ are obtained by standard least squares of y_t on $x_t(\widehat{\gamma})$. We write the fitted regression function as

$$y_t = \widehat{\beta}' x_t(\widehat{\gamma}) + \widehat{e}_t. \tag{6}$$

In (6), $\widehat{e}_t$ are the (non-linear) least-squares residuals. An estimate of error variance $\sigma^2 = Ee_i^2$ is

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^{n} \widehat{e}_t^2 = S_n^*(\widehat{\gamma}).$$

To illustrate, consider the U.S. growth regression for 1792-2009. First, we set the parameter space Γ for the threshold parameter as $\Gamma = [10, 70]$, so that at least 5% of the sample and 10% of the support of the threshold variable are below the lower bound and above the upper bound. We then approximate the minimization (5) by computing $S_n^*(\gamma)$ on a discrete grid with increments 0.1 (which has 601 gridpoints). At each gridpoint for γ we estimated the least-squares coefficients and computed the least-squares criterion $S_n^*(\gamma)$. This criterion is plotted as a function of γ in Figure 3. We observe that the function is reasonably smooth and has a well defined global minimum, but the criterion is not well described as quadratic. The relative smoothness of the plot suggests that our choice for the grid evaluation is sufficiently fine to obtain the global minimum. The criterion is minimized at $\widehat{\gamma} = 43.8$. Interestingly, this threshold estimate is very close to that found by Lin (2014) for the U.S. using a different data window and quite different empirical framework, and also very close to the estimate found by Lee, Park, Seo and Shin (2014) using a median cross-country regression.

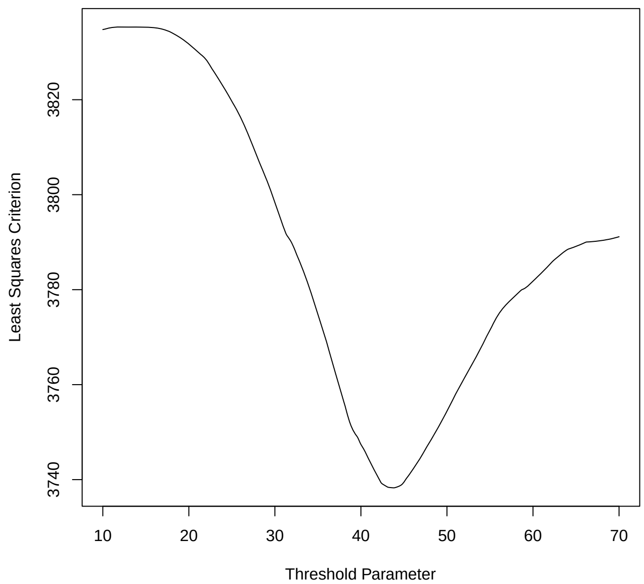


Figure 3: Concentrated Least Squares Criterion for Threshold Parameter

The parameter estimates from this fitted regression kink model are as follows.

$$y_t = \underset{(0.026)}{0.033} \left(x_{t-1} - \underset{(12.1)}{43.8} \right)_- - \underset{(0.048)}{0.067} \left(x_{t-1} - \underset{(12.1)}{43.8} \right)_+ + \underset{(0.09)}{0.28}\, y_{t-1} + \underset{(0.69)}{3.78} + \widehat{e}_t$$

$$\widehat{\sigma}^2 = 17.2$$

The standard errors will be discussed in Section 5. The estimates show a positive first-order autocorrelation coefficient of 0.28, which is standard for time-series estimates for growth rates. The estimates suggest that high debt ratios (those above 44) induce a moderate slowdown in expected GDP growth rates, consistent with the Reinhart-Rogoff hypothesis.

4 Testing for a Threshold Effect

A reasonable question is whether or not the threshold model (1) is "significant" relative to the linear model

$$y_t = \beta_1 x_t + \beta_3' z_t + e_t. \tag{7}$$

The linear model (7) is nested in the threshold model (1), as (7) holds under the restriction $\beta_1 = \beta_2$ (with a redefinition of the intercept). Under this hypothesis the threshold γ disappears and we see the familiar problem of an unidentified parameter under the null hypothesis. This requires non-standard testing methods; fortunately such issues are well understood and easily applied.

An appropriate estimator of the linear model is least squares. We can write these estimates as

$$y_t = \widetilde{\beta}_1 x_t + \widetilde{\beta}_3' z_t + \widetilde{e}_t. \tag{8}$$

The error variance estimate is

$$\widetilde{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n} \widetilde{e}_t^2.$$

For example, the estimates of the linear model for our empirical example are as follows.

$$\begin{aligned} y_t = -\underset{(0.013)}{0.008}\; x_{t-1} + \underset{(0.09)}{0.30}\; y_{t-1} + \underset{(0.62)}{2.90} + \widetilde{e}_t \\ \widetilde{\sigma}^2 = 17.6 \end{aligned} \tag{9}$$

The estimates show a postive first-order autocorrelation coefficient, and a near zero coefficient on the debt ratio.

A standard test for the null hypothesis of the linear model (7) against the threshold model (1) is to reject $H_0 : \beta_1 = \beta_2$ for large values of the F-type statistic

$$T_n = \frac{n\left(\widetilde{\sigma}^2 - \widehat{\sigma}^2\right)}{\widehat{\sigma}^2}.$$

For example, for our empirical example we can compute that $T_n = 5.66$.

To describe the distribution theory for the test statistic, we need to be precise about the stochastic assumptions on the model. We require the following regularity conditions.

Assumption 1 *For some* $r > 1$,

1. (y_t, z_t, x_t) *is strictly stationary, ergodic, and absolutely regular with mixing coefficients* $\eta(m) = O(m^{-A})$ *for some* $A > r/(r-1)$
2. $E\left|y_t\right|^{4r} < \infty$, $E\left|x_t\right|^{4r} < \infty$, *and* $E\left\|z_t\right\|^{4r} < \infty$
3. $\inf_{\gamma\in\Gamma} \det Q(\gamma) > 0$ *where* $Q(\gamma) = E\left(x_t(\gamma)x_t(\gamma)'\right)$

4. x_t *has a density function* $f(x)$ *satisfying* $f(x) \leq \overline{f} < \infty$

5. $\gamma \in \Gamma$ *where* Γ *is compact*

Assumption 1.1 & 1.2 are standard weak dependence conditions which are sufficient for a central limit theorem. The choice of r involves a trade-off between the allowable degree of serial dependence and the number of finite moments. For independent observations we can set r arbitrarily close to one. Assumption 1.3 is an identification condition, requiring that the projection coefficients be well defined for all values of γ in the parameter space Γ. Assumption 1.4 requires that the threshold variable has a bounded density function.

In particular, Assumption 1 allows the regressors x_t and z_t to include lagged dependent variables.

The following is implied by Theorem 3 of Hansen (1996). The only difference is that the latter theorem focused on a heteroskedasticity-robust Wald statistic, while for simplicity we consider the simpler homoskedastic form of the test statistic.

Theorem 1 *Suppose that Assumption 1 holds and* e_t *is a martingale difference sequence. Under* $H_0 : \beta_1 = \beta_2$,

$$T_n \rightarrow_d \sup_{\gamma \in \Gamma} G(\gamma)' Q(\gamma)^{-1} G(\gamma)/\sigma^2 \tag{10}$$

where $G(\gamma)$ *is a zero-mean Gaussian process with covariance kernel*

$$E\left(G(\gamma_1)G(\gamma_2)\right) = E\left(x_t(\gamma_1)x_t(\gamma_2)' e_t^2\right). \tag{11}$$

Theorem 1 shows that the asymptotic null distribution of the threshold F statistic can be written as the supremum of a stochastic process. In addition to Assumption 1, Theorem 1 adds the regularity condition that the error e_t is a martingale difference sequence. This is a sufficient condition, and convenient, but is not essential. What is important is that the best fitting approximation in the regression kink model is the linear model so that the central limit theorem applies for all γ, and that the regression scores are uncorrelated so that the covariance kernel takes the simple form (11) rather than a HAC form.

The limiting distribution (10) is non-standard and cannot be tabulated. However, as shown by Hansen (1996), it is simple to simulate approximations to (10) using a multiplier bootstrap, and thus asymptotically valid p-values can be calculated. The following is his recommended algorithm. (Theorem 3 of Hansen (2006) shows that the algorithm produces asymptotically first-order correct p-values under the conditions of Theorem 1.)

Algorithm 1: Testing for a Regression Kink with an Unknown Threshold

1. Generate n iid draws u_t from the $N(0, 1)$ distribution.

2. Set $y_t^* = \widetilde{e}_t u_t$ where $\widetilde{e}_t$ are the OLS residuals from the fitted linear regression (8).

3. Using the observations (y_t^*, z_t, x_t), estimate the linear model (8) and the regression kink model (6) and compute the error variance estimates $\widetilde{\sigma}^{*2}$ and $\widehat{\sigma}^{*2}$ and the F-statistic

$$T_n^* = \frac{n\left(\widetilde{\sigma}^{*2} - \widehat{\sigma}^{*2}\right)}{\widehat{\sigma}^{*2}}.$$

4. Repeat this B times, so as to obtain a sample $T_n^*(1), ..., T_n^*(B)$ of simulated F statistics.

5. Compute the p-value as the percentage of simulated F statistics which exceed the actual value:

$$p_n = \frac{1}{B}\sum_{b=1}^{B} 1\left(T_n^*(b) \geq T_n\right)$$

6. If desired, compute the level α critical value c_α as the empirical $1-\alpha$ quantile of the simulated F statistics $T_n^*(1), ..., T_n^*(B)$.

7. Reject H_0 in favor of H_1 at significance level α if $p_n < \alpha$, or equivalently if $T_n > c_\alpha$.

The number of bootstrap replications B should be set fairly large to ensure accuracy of the p-value p_n. For example, if $B = 10{,}000$ then p_n is approximately within 0.006 of its limiting (large B) value. Fortunately, the computational cost is minimal. For example, my office computer computed all the empirical calculations reported in this paper in just one minute, including two separate bootstrap simulations using 10,000 replications each. (The calculations were performed in R. The code implements the multiplier bootstrap efficiently by executing all 10,000 regressions simultaneously, exploiting the fact that the regressors are common across the boostrap replications.)

Returning to the U.S. GDP example, as we said earlier the empirical value of the F statistic is $T_n = 5.66$. The multiplier bootstrap p-value is 0.15, and the bootstrap estimate of the 10% critical value is 7.1. Thus the test does not reject the null hypothesis of linearity in favor of the regression kink model at the 10% level.

The multiplier bootstrap method does not account for the time-series nature of the observations and thus can be expected to exhibit some finite sample distortions. In particular, Assumption 1 does not allow for non-stationary or near-non-stationary regressors. As argued by Stambaugh (1999) and others, highly persistent regressors will distort normal distribution theory. This is particularly relevant for our example as can be seen in Figure 1, the debt/GDP level displays considerable serial dependence.

To investigate this possibility we present a simple simulation experiment calibrated on the empirical example. Our data generating process is

$$\begin{aligned} y_t &= \beta_1 (x_t - \gamma)_- + \beta_2 (x_t - \gamma)_+ + \beta_3 y_{t-1} + \beta_4 + e_t \\ e_t &\sim N(0, \sigma^2) \end{aligned}$$

To evaluate size, we set $\beta_1 = \beta_2 = 0$, $\beta_3 = 0.3$, $\beta_4 = 3$, and $\sigma^2 = 16$ to match the estimates (9). We fix x_t to equal the empirical values of debt/GDP for the U.S, and set $n = 218$ as in the empirical example. Setting x_t to equal the sample values is done precisely to force the simulation to evalulate the performance in a setting with the serial correlation properties of the observed debt/GDP series. We generated 10,000 samples from this process. To speed computation, for all simulations in the paper we evaluate $\Gamma = [10, 70]$ using a grid with increments of 1, which reduces the number of gridpoints to 61. The number of bootstrap replications for each simulated sample was set at $B = 1000$.

We first evalualated the size of the threshold test. At the nominal 10% significance level, we found the simulation size to be 10.6%. Thus, the test exhibits no meaningful size distortion.

Second, we evaluated the power of the test. We used the same parameterization as above, but now set $\beta_1 = 0$ and $\gamma = 40$, and vary β_2 from 0 to -0.16 in steps of 0.02. The power is presented in Table 1. What we can see is that the test has increasing power in β_2, and had reasonable ability to detect changes as small as 0.08. Examining the point estimates of the threshold model, we see that the estimate of the difference in regression slopes is $\widehat{\beta}_1 - \widehat{\beta}_2 = 0.10$, where the simulation suggests the power should be about 70%, which is reasonable but modest. It follows that our empirical p-value of 15% could be due to the modest power of the test. We conclude that the threshold test is inconclusive regarding the question of whether or not there is a regression kink effect in GDP growth due to high debt.

Table 1: Power of Threshold F test with Multiplier Bootstrap, Nominal Size 10%

	β_2							
	-0.02	-0.04	-0.06	-0.08	-0.10	-0.12	-0.14	-0.16
Power	0.14	0.24	0.38	0.54	0.72	0.85	0.93	0.98

5 Inference on the Regression Coefficients

In this section we consider the distribution theory for the least-squares estimates of the regression kink model with unknown threshold under the assumption that the threshold is identified.

In an ideal context we might consider (1) as the true regression function, so that the error has conditional mean zero. To allow extra generality, we will not impose this condition. Instead, we will view (1) as the best approximation in the sense that it minimizes squared error loss. We define the best approximation as (pseudo)-true values (β_0, γ_0) which minimize the squared loss

$$L(\beta, \gamma) = E\left(y_t - \beta' x_t(\gamma)\right)^2 . \qquad (12)$$

As in our analysis of estimation, we can define the minimizers by concentration. Let $\beta(\gamma)$ be the minimizer of $L(\beta, \gamma)$ over β for fixed γ, this is $\beta(\gamma) = E\left(x_t(\gamma)x_t(\gamma)'\right)^{-1} E\left(x_t(\gamma)y_t\right)$. Under Assumption 1.3, $\beta(\gamma)$ is uniquely defined for all $\gamma \in \Gamma$. The concentrated squared loss is then

$$L^*(\gamma) = L(\beta(\gamma), \gamma) = E\left(y_t - \beta(\gamma)' x_t(\gamma)\right)^2 .$$

By concentration, γ_0 is the minimizer of $L^*(\gamma)$ and $\beta_0 = \beta(\gamma_0)$.

We will require that the minimizer γ_0 is unique. This excludes the case of a best-fitting linear model (in which case $L^*(\gamma)$ is a constant function) and the case of multiple best-fitting threshold parameters γ.

To simplify our proof of consistency, we also impose that the parameter space for β is compact, though this assumption could be relaxed by a more detailed argument.

Assumption 2

1. $\gamma_0 = \operatorname{argmin}_{\gamma\in\Gamma} L^*(\gamma)$ *is unique.*
2. $\beta \in B \subset \mathbb{R}^{k-1}$ *where* B *is compact.*

Chan and Tsay (1998) showed that the least-squares estimates in the continuous threshold autoregressive model, including both the slope and threshold coefficients, are jointly asymptotically normal. We extend their distribution theory to the regression kink model with unknown threshold. Set $\theta = (\beta, \gamma)$, $\theta_0 = (\beta_0, \gamma_0)$,

$$H_t(\theta) = -\frac{\partial}{\partial\theta}\left(y_t - \beta' x_t(\gamma)\right) = \begin{pmatrix} x_t(\gamma) \\ -\beta_1 1(x_t < \gamma) - \beta_2 1(x_t > \gamma) \end{pmatrix}$$

and $H_t = H_t(\theta_0)$.

Theorem 2 *Under Assumption 1*

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \to_d N(0, V)$$

where $V = Q^{-1} S Q^{-1}$, $S = \sum_{j=-\infty}^{\infty} E\left(H_t H_{t+j}' e_t e_{t+j}\right)$ *and*

$$Q = E\left(H_t H_t'\right) + E\begin{pmatrix} 0 & 0 & 0 & e_t 1(x_t < \gamma_0) \\ 0 & 0 & 0 & e_t 1(x_t > \gamma_0) \\ 0 & 0 & 0 & 0 \\ e_t 1(x_t < \gamma_0) & e_t 1(x_t > \gamma_0) & 0 & 0 \end{pmatrix}$$

A formal proof of Theorem 2 is presented in the Appendix.

Notice that the slope and threshold estimates are jointly asymptotically normal with $\sqrt{n}$ convergence rate, and the slope and threshold estimates have a non-zero asymptotic covariance. In contrast, in the conventional non-continuous threshold model the threshold estimate $\widehat{\gamma}$ is rate n consistent with a non-standard asymptotic distribution, and the slope coefficient estimates are asymptotically independent of the threshold estimate. The difference in the regression kink model is because the regression function is continuous. Consequently, the least-squares criterion is continuous in the parameters (though not differentiable) and asymptotically locally quadratic.

Since the threshold estimate $\widehat{\gamma}$ has only a $\sqrt{n}$ convergence rate, we should expect its precision to be less accurate than threshold estimates in the non-continuous case. Thus it is important to take its sampling distribution into account when constructing confidence intervals.

The asymptotic distribution does not require the model to be correctly specified, so the error e_t need not be a martingale difference sequence. Thus (in general) the covariance matrix S takes a HAC form. When the regression is dynamically well specified (by appropriate inclusion of lagged variables) then the matrix will simplify to $S = E\left(H_t H_t' e_t^2\right)$. As our application includes a lagged dependent variable, we use this simplification in practice.

The second term in the definition of Q is zero when the threshold model is correctly specified so that $E(e_t|x_t) = 0$. However, in the case of model misspecification (so that the regression is a best approximation) then the second term can be non-zero.

We suggest the following estimate of the asymptotic covariance matrix (assuming uncorrelatedness). Set

$$\widehat{H}_t = \left(\begin{array}{c} x_t(\widehat{\gamma}) \\ -\widehat{\beta}_1 1\left(x_t < \widehat{\gamma}\right) - \widehat{\beta}_2 1\left(x_t > \widehat{\gamma}\right) \end{array} \right),$$

$\widehat{V} = \widehat{Q}^{-1}\widehat{S}\widehat{Q}^{-1}$, $\widehat{S} = \frac{1}{n-k}\sum_{t=1}^{n} \widehat{H}_t \widehat{H}_t' \widehat{e}_t^2$, and

$$\widehat{Q} = \frac{1}{n}\sum_{t=1}^{n}\left(\widehat{H}_t\widehat{H}_t' + \left(\left[\begin{array}{cccc} 0 & 0 & 0 & \widehat{e}_t 1\left(x_t < \widehat{\gamma}\right) \\ 0 & 0 & 0 & \widehat{e}_t 1\left(x_t > \widehat{\gamma}\right) \\ 0 & 0 & 0 & 0 \\ \widehat{e}_t 1\left(x_t < \widehat{\gamma}\right) & \widehat{e}_t 1\left(x_t > \widehat{\gamma}\right) & 0 & 0 \end{array}\right]\right)\right).$$

We divide by $n-k$ rather than n for the definition of $\widehat{S}$ as an ad hoc degree-of-freedom adjustment. Given $\widehat{V}$, standard errors for the coefficient estimates are found by taking the square roots of the diagonal elements of $n^{-1}\widehat{V}$. If the regression is dynamically misspecified (for example, if no lagged dependent variable is included) then $\widehat{S}$ could be formed using a standard HAC estimator.

Asymptotic confidence intervals for the coefficients could then be formed using the conventional rule, e.g. for a 95% interval for β_1, $\widehat{\beta}_1 \pm 1.96 s(\widehat{\beta}_1)$. Theorem 2 shows that such confidence intervals have asymptotic correct coverage.

In small samples, however, the asymptotic confidence intervals may have poor coverage. As shown in Figure 2, the least-squares critierion is non-quadratic with respect to the threshold parameter γ, meaning that quadratic (e.g. normal) approximations may be poor unless sample sizes are quite large. In this context better coverage can be obtained by test-inversion confidence sets. This is particularly convenient for the threshold parameter γ, as a test inversion confidence region is a natural by-product of the computation of the least-squares minimization. Specifically, to test the hypothesis $H_0 : \gamma = \gamma_0$ against $H_1 : \gamma \neq \gamma_0$, the criterion-based test is to reject for large values of the F-type statistic $F_n(\gamma_0)$ where

$$F_n(\gamma) = \frac{n\left(\widehat{\sigma}^2(\gamma) - \widehat{\sigma}^2\right)}{\widehat{\sigma}^2} \tag{13}$$

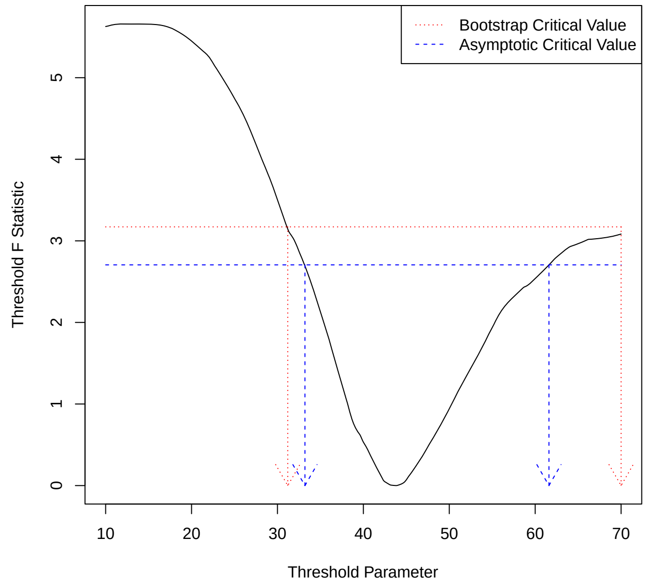


Figure 4: Confidence Interval Construction for Threshold

and $\widehat{\sigma}^2(\gamma) = S_n^*(\gamma)$ is the estimator of the error variance when γ is fixed. Given the asymptotic normality of Theorem 2, this test has an asymptotic χ_1^2 distribution under H_0. Thus, for a nominal level α test we can take the critical value $c_{1-\alpha}$ from the χ_1^2 distribution. A nominal $1-\alpha$ asymptotic confidence interval for γ can then be formed by test inversion: the set of γ for which $F_n(\gamma)$ is smaller the χ_1^2 critical value:

$$C_\gamma = \{\gamma : F_n(\gamma) \leq c_{1-\alpha}\}.$$

Given that $\widehat{\sigma}^2(\gamma)$ has already been calculated on a grid (for estimation), we have $F_n(\gamma)$ on the same grid as a by-product. The interval C_γ is then obtained as the set of γ gridpoints for which $F_n(\gamma)$ is smaller than $c_{1-\alpha}$.

To illustrate, examine Figure 4. Here we have plotted the statistic $F_n(\gamma)$ from (13) as a function of γ. By construction, the statistic is non-negative and hits 0 at $\gamma = \widehat{\gamma}$. We have drawn in the asymptotic 90% critical value $c_{.9} = 2.7$ using the blue dashed line. The points of intersection indicate the ranges for the asymptotic confidence interval, and shown on the graph by the blue dashed arrows to the horizontal axes.

Further improvements in coverage accuracy can be obtained via a bootstrap. Since the regressor x_t displays considerable serial dependence, we recommend a wild bootstrap which conditions on the values of x_t. This will preserve the time-series properties better than alternative bootstrap methods such as a block bootstrap which is more suited for cases of mild serial dependence. A model-based bootstrap could also be used, but then an explicit model would be required for the debt/GDP series. To avoid these challenges we adopt a wild bootstrap. Here are the steps.

Algorithm 2: Wild Bootstrap Confidence Intervals for Parameters

1. Generate n iid draws u_t from the $N(0,1)$ distribution.

2. Set $e_t^* = \widehat{e}_t u_t$ where $\widehat{e}_t$ are the LS residuals from the fitted regression kink model (6).

3. Set $y_t^* = \widehat{\beta}' x_t(\widehat{\gamma}) + e_t^*$, where (z_t, x_t) are the sample observations, and $(\widehat{\beta}, \widehat{\gamma})$ are the least-squares estimates.

4. Using the observations (y_t^*, z_t, x_t), estimate the the regression kink model (6), parameter estimates $(\widehat{\beta}^*, \widehat{\gamma}^*)$, and $\widehat{\sigma}^{*2} = n^{-1} \sum_{t=1}^n \widehat{e}_t^{*2}$, where $\widehat{e}_t^* = y_t^* - \widehat{\beta}^{*\prime} x_t(\widehat{\gamma}^*)$.

5. Calculate the F-statistic for γ
$$F_n^* = \frac{n\left(\widehat{\sigma}^{*2}(\widehat{\gamma}) - \widehat{\sigma}^{*2}\right)}{\widehat{\sigma}^{*2}}$$
where $\widehat{\sigma}^{*2}(\widehat{\gamma}) = n^{-1} \sum_{t=1}^n \widehat{e}_t^*(\widehat{\gamma})^2$ and $\widehat{e}_t^*(\widehat{\gamma}) = y_t^* - \widehat{\beta}^*(\widehat{\gamma})' x_t(\widehat{\gamma})$.

6. Repeat this B times, so as to obtain a sample of simulated coefficient estimates $(\widehat{\beta}^*, \widehat{\gamma}^*)$ and F statistics F_n^*.

7. Create $1-\alpha$ bootstrap confidence intervals for the coefficients β_1, β_2 and β_3 by the symmetric percentile method: the coefficient estimates plus and minus the $(1-\alpha)$ quantile of the absolute centered bootstrap estimates. E.g., for β_1 the interval is $\widehat{\beta}_1 \pm q_1^*$ where q_1^* is the $(1-\alpha)$ quantile of $|\widehat{\beta}_1^* - \widehat{\beta}_1|$.

8. Calculate the $1-\alpha$ quantile $c_{1-\alpha}^*$ of the simulated F statistics F_n^*.

9. Create a $1-\alpha$ bootstrap confidence interval for γ as the set of γ for which the empirical F statistics $F_n(\gamma)$ (13) are smaller than the bootstrap critical value $c_{1-\alpha}^*$.
$$C_\gamma^* = \left\{\gamma : F_n(\gamma) \leq c_{1-\alpha}^*\right\}.$$

This wild bootstrap algorithm can be computed concurrently with the multiplier bootstrap used for the threshold test, resulting in efficient computation. We do not have a formal distribution theory which justifies the use of this (or any other) bootstrap method for confidence interval construction. However, we see no reason for the bootstrap to fail given the asymptotic normality of Theorem 2.

Again for illustration the confidence interval construction can be seen via Figure 4, where the statistic $F_n(\gamma)$ is plotted against γ. The bootstrap 90% critical value 3.3 (calculated with $B = 10,000$ bootstrap replications) is plotted as the red dotted line. The points of intersection indicate the ranges of the confidence interval, and are indicated on the figure by the red dotted arrows to the horizontal axis. Since the bootstrap critical value is larger than the asymptotic critical value, the asymptotic interval is a subset of the bootstrap confidence interval.

Before presenting the empirical results we report results from our simulation experiment to assess the performance of the methods. The data generating process is identical to that used in the previous section. As before, β_2 is the key free parameter, controlling the strength of the threshold effect, and the remaining parameters and variables are set to match the empirical data. In particular, we fixed the regressor x_t to equal the empirical values of debt/GDP for the U.S, so to precisely preserve its strong serial dependence properties. As before, we generated $10,000$ simulated samples and used $B = 1000$ bootstrap replications for each sample. In Tables 2 and 3 we report the coverage rates of nominal 90% intervals for the parameters β_2 and γ.

Table 2: Coverage rates of nominal 90% confidence intervals for β_2

	β_2							
	-0.02	-0.04	-0.06	-0.08	-0.10	-0.12	-0.14	-0.16
$\widehat{\beta}_2 \pm 1.645 s(\widehat{\beta}_2)$	0.81	0.80	0.81	0.81	0.81	0.80	0.81	0.81
Percentile Bootstrap	0.81	0.83	0.85	0.87	0.86	0.86	0.86	0.86
Inverse Percentile Bootstrap	0.84	0.84	0.85	0.83	0.82	0.82	0.82	0.83
Symmetric Percentile Bootstrap	0.86	0.86	0.87	0.88	0.87	0.86	0.86	0.86

The confidence intervals for β_2 are reported in Table 2. The first row reports the coverage rates for the asymptotic "plus or minus" standard error confidence intervals. We see that the confidence intervals undercover, with rates approximately 81%. The second row report the coverage rates of the percentile confidence interval. These rates are better, especially for moderate values of β_2. In the third row we report the coverage rates for the "Inverse Percentile Bootstrap, whose endpoints are twice the coefficient estimates, less the $1 - \alpha/2$ and $\alpha/2$ quantiles of the simulated coefficient estimates $\widehat{\beta}_2^*$. Its performance is not superior to the percentile interval. In the fourth row we report the coverage rates for the "Symmetric Percentile Bootstrap" whose endpoints equal the coefficient estimates, plus and minus the $1 - \alpha$ quantile of the simulated statistics $|\widehat{\beta}_2^* - \widehat{\beta}_2|$. These intervals only slightly uncover, with coverage rates about 86%-88%. These are our recommended confidence intervals for the slope coefficients.

In Table 3 we report coverage rates of nominal 90% intervals for the parameter γ. The first row reports the coverage rates for the asymptotic "plus or minus" standard error confidence intervals. These severely undercover, with coverage rates as low as 67%. The second and third rows report coverage rates of the conventional percentile interval and the inverse percentile interval. Neither performs well, with the percentile interval substantially overcovering and the inverse percentile interval severely undercovering for small β_2. The fourth row reports the coverage rates of the symmetric percentile interval. It's performance is better than the other percentile intervals, but coverage rates are sensitive to the value of β_2. The fifth and sixth rows report the coverage rates for the test-inversion confidence intervals C_γ and C_γ^*. These two have quite good coverage, especially C_γ^* which uses the bootstrap critical values. The interval only mildly under-covers, and has roughly uniform coverage across β_2. The invervals C_γ^* are our recommended confidence intervals for γ.

Table 3: Coverage rates of nominal 90% confidence intervals for γ

	β_2							
	-0.02	-0.04	-0.06	-0.08	-0.10	-0.12	-0.14	-0.16
$\widehat{\gamma} \pm 1.645 s(\widehat{\gamma})$	0.67	0.68	0.69	0.70	0.71	0.73	0.74	0.75
Percentile Bootstrap	0.99	0.99	0.97	0.96	0.94	0.93	0.92	0.91
Inverse Percentile Bootstrap	0.33	0.40	0.50	0.58	0.67	0.74	0.77	0.81
Symmetric Percentile Bootstrap	0.96	0.94	0.91	0.89	0.87	0.86	0.85	0.86
C_γ	0.90	0.88	0.87	0.86	0.84	0.85	0.84	0.85
C_γ^*	0.91	0.89	0.88	0.88	0.87	0.88	0.87	0.88

Our simulation results show that the recommended wild bootstrap confidence intervals work reasonably well, but with some moderate undercoverage. This is likely because our simulated data are a time-series autoregression (and thus have time-series dependence) while the wild bootstrap does not account for the time-series dependence. The reason we recommend a wild bootstrap (rather than a model-based bootstrap) is that the wild bootstrap is able to condition on the observed regressor processes and (most importantly) nonparametrically handle conditional heteroskedasticity. These are important advantages and conveniences.

We now present in Table 4 the confidence intervals for the coefficient estimates from the empirical regression, computed with our recommended bootstrap methods with $B = 10,000$ bootstrap replications. In our assessment the two most important coefficients are β_2 (the slope effect of debt on growth for high debt ratios) and γ (the threshold level). The confidence interval for β_2 is $[-0.18, -0.01]$. This range is sufficiently wide that we cannot infer with precision the magnitude of the impact of debt on expected growth. The 90% confidence interval for the threshold is 31% to 70%, also indicating substantial uncertainty.

The wide confidence interval for the threshold could be a feature of the $\sqrt{n}$ convergence rate for the threshold estimate, or it could be due to the small sample size.

Table 4: Coefficient Estimates and Bootstrap 90% Confidence Intervals

	Estimate	s.e.	Interval
β_1	0.033	0.035	$[-0.002, 0.136]$
β_2	-0.067	0.037	$[-0.178, -0.006]$
y_{t-1}	0.28	0.088	$[0.14, 0.43]$
Intercept	3.78	0.79	$[2.58, 4.94]$
γ	43.8	12.3	$[30.8, 70.0]$

6 Inference on the Regression Kink Function

In this section we consider inference on the regression kink function

$$g(\theta) = \beta' x(\gamma) \tag{14}$$

where

$$x(\gamma) = \begin{pmatrix} (x-\gamma)_- \\ (x-\gamma)_+ \\ z \end{pmatrix}.$$

Our theory will treat (x, z) as fixed, even though we will display the estimates as a function of x. (Thus we focus on pointwise confidence intervals for the regression function.)

The plug-in estimate of $g(\theta)$ is $g(\widehat{\theta}) = \widehat{\beta}' x(\widehat{\gamma})$. In our empirical example, we display in Figure 2 $g(\widehat{\theta})$ as a function of x, with z fixed at the sample mean of z_t.

Since Theorem 2 shows that $\widehat{\theta}$ is asymptotically normal, it might be conjectured that $g(\widehat{\theta})$ will be as well. This conjecture turns out to be incorrect. While $g(\theta)$ is a continuous function of θ, it is not differentiable at $\gamma = x$. As discussed in a recent series of papers (Hirano and Porter (2012), Woutersen and Ham (2013), Fang and Santos (2014), Fang (2014) and Hong and Li (2015), the non-differentiability means that $g(\widehat{\theta})$ will not be asymptotically normal at $\gamma_0 = x$, and asymptotic normality is likely to be a poor approximation for γ_0 close to x.

While the regression kink function $g(\theta)$ is not differentiable at $\gamma = x$, it is directionally differentiable at all points, meaning that both left and right derivatives are defined. The directional derivative of a function $\phi(\theta) : \mathbb{R}^k \to \mathbb{R}$ in the direction $h \in \mathbb{R}^k$ is

$$\phi_\theta(h) = \lim_{\varepsilon \downarrow 0} \frac{\phi(\theta + h\varepsilon) - \phi(\theta)}{\varepsilon}. \tag{15}$$

(See Shapiro (1990).) The primary difference with the conventional notion of a derivative is that $\phi_\theta(h)$ is allowed to depend on the direction h in which the derivative is taken. For continuously differentiable functions the directional derivative is linear in h. For example, if $\phi(\theta) = x'\theta$ is linear, then $\phi_\theta(h) = x'h$. If $\phi(\theta)$ is continuously differentiable in θ then $\phi_\theta(h) = \left(\frac{\partial}{\partial\theta}\phi(\theta)\right)' h$. However, if $\phi(\theta)$ is continuous but not continuously differentiable the directional derivative will be non-linear in h.

In the case of (14), we can calculate that the directional derivative of $g(\theta)$ in the direction $h = (h_\beta, h_\gamma)$ is

$$g_\theta(h) = x(\gamma)' h_\beta + g_\gamma(h_\gamma) \tag{16}$$

where

$$g_\gamma(h_\gamma) = \begin{cases} -\beta_1 h_\gamma, & \text{if} \quad x < \gamma \\ -\beta_1 h_\gamma 1\left(h_\gamma < 0\right) - \beta_2 h_\gamma 1\left(h_\gamma \geq 0\right), & \text{if} \quad x = \gamma \\ -\beta_2 h_\gamma, & \text{if} \quad x > \gamma \end{cases} \tag{17}$$

The directional derivative $g_\gamma(h_\gamma)$ is linear for $x \neq \gamma$ but non-linear for $x = \gamma$, with a slope to the left of $-\beta_1$ and a slope to the right of $-\beta_2$.

Conventional asymptotic theory for functions of the form $\phi(\widehat{\theta})$ requires that the function $\phi(\theta)$ be continuously differentiable. Shapiro (1991, Theorem 2.1) and Fang and Santos (2014, Corollary 2.1) have generalized this to the case of directional differentiability. They have established that if $\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \rightarrow_d \mathrm{Z} \sim N(0, V)$ and $\phi_\theta(h)$ is continuous in h then $\sqrt{n}\left(\phi(\widehat{\theta}) - \phi(\theta_0)\right) \rightarrow_d \phi_\theta(\mathrm{Z})$. In the classic case where $\phi(\theta)$ is continuously differentiable, the limiting distribution specializes to $\phi_\theta(\mathrm{Z}) = \left(\frac{\partial}{\partial\theta}\phi(\theta)\right)'\mathrm{Z} \sim N\left(0, \left(\frac{\partial}{\partial\theta}\phi(\theta)\right)' V \left(\frac{\partial}{\partial\theta}\phi(\theta)\right)\right)$, so their result is a strict generalization of the Delta Method.

Since (16)-(17) is continuous in h, we can immediately deduce the asymptotic distribution of the regression estimate.

Theorem 3 $\sqrt{n}\left(g(\widehat{\theta}) - g(\theta_0)\right) \rightarrow_d g_\theta(\mathrm{Z})$ *where* $\mathrm{Z} \sim N(0, V)$ *and* $g_\theta(h)$ *is defined in (16)-(17).*

For $x \neq \gamma_0$ the asymptotic distribution in Theorem 3 is normal, but for $x = \gamma_0$ it is a non-linear transformation of a normal random vector. At $x = \gamma_0$ the asymptotic distribution will be biased, with the direction of bias depending on the relative magnitudes of β_1 and β_2. For example, if $\beta_1 = 1$ and $\beta_2 = -1$ then $g_\gamma(h_\gamma) = |h_\gamma|$ so the asymptotic distribution in Theorem 3 will have a positive mean while if the signs of the coefficients are reversed then $g_\gamma(h_\gamma) = -|h_\gamma|$ and the asymptotic distribution in Theorem 3 will have a negative mean.

This non-normality implies that classical confidence intervals will have incorrect asymptotic coverage. We illustrate this in Figure 5 by a continuation of our simulation experiment. Using the same simulation design as in the previous section, and varying β_2 from -0.04 to -0.16 in steps of 0.04, we plot the coverage probability of classical (naive) nominal 90% pointwise confidence intervals, plotting the coverage as a function of x. The asymptotic theory predicts that the confidence intervals will have asymptotically correct 90% coverage for $x \neq \gamma_0$ but not for $x = \gamma_0$, so we should expect coverage to be better for x distinct from $\gamma = 40$, but deteriorating for x near $\gamma = 40$. Indeed, we see that the coverage rates vary from near 90% for small x to approximately 77% at $x = 40$. The distortion from the nominal coverage is sensitive to β_2, with the distortion steepening as β_2 increases. We can also see that the coverage rates are less than the nominal 90% for large values of x, which is not predicted from the asymptotic theory, and these distortions are more severe for small values of β_2. This appears to be a small-sample issue (a finite sample bias in the regression estimate; there are only seven observations where $x_t \geq 80$) and thus unrelated to the asymptotic non-normality of Theorem 3.

We might hope that bootstrap methods would improve the coverage probabilities, but this is a false hope. As shown by Fang and Santos (2014, Corollary 3.1), the non-normality of Theorem 3 implies that the conventional bootstrap will be inconsistent. Indeed, we investigated the coverage probabilities of confidence intervals constructed using the percentile and inverse percentile methods, and their coverage rates are similar to that shown in Figure 5 (though less distortion for large x), so are not displayed here.

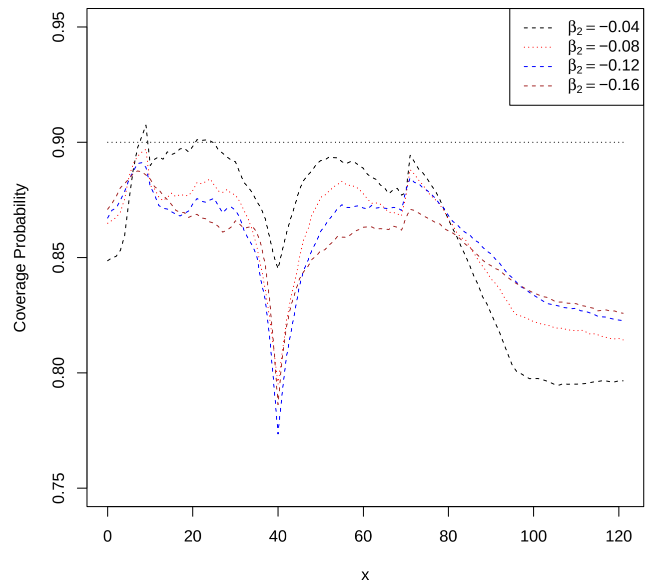


Figure 5: Coverage Probabilities of Nominal 90% Naive Asymptotic Confidence Intervals for Regression Kink Function

Constructively, Fang and Santos (2014) suggest an alternative bootstrap method which is consistent for the asymptotic distribution. Their suggestion is to approximate the distribution of $\sqrt{n}\left(g(\widehat{\theta}) - g(\theta_0)\right)$ by that of $\widehat{g}_\theta\left(\sqrt{n}\left(\widehat{\theta}^* - \widehat{\theta}\right)\right)$, where $\widehat{\theta}^*$ is the bootstrap distribution of $\widehat{\theta}$ and $\widehat{g}_\theta(h)$ is an estimate of $g_\theta(h)$. Confidence intervals can be constructed from this alternative bootstrap distribution.

The Fang-Santos alternative bootstrap requires that the directional derivative $g_\theta(h)$ be either known or consistently estimated. Fortunately, the problem of estimating the directional derivative has been solved in a recent paper by Hong and Li (2015). Their suggestion is to estimate (15) with a discrete analog. Specifically, for some sequence $\varepsilon_n > 0$ satisfying $\varepsilon_n \to 0$ and $\sqrt{n}\varepsilon_n \to \infty$, estimate the directional derivative $\phi_\theta(h)$ of $\phi(\theta)$ by

$$\widehat{\phi}_\theta(h) = \frac{\phi(\widehat{\theta} + h\varepsilon_n) - \phi(\widehat{\theta})}{\varepsilon_n}.$$

Hong and Li (2015) show that this produces a consistent estimate of $\phi_\theta(h)$ when $\phi(\theta)$ is Lipschitz continuous. They call $\widehat{\phi}_\theta(h)$ "the numerical delta method."

In our case the regression function $g(\theta)$ is Lipschitz continuous in θ and the directional derivative $g_\theta(h)$ takes the semi-linear form (16), so we only need to estimate the non-linear component (17). The Hong-Li estimate is

$$\widehat{g}_\gamma(h_\gamma) = \frac{g(\widehat{\beta}, \widehat{\gamma} + h_\gamma\varepsilon_n) - g(\widehat{\beta}, \widehat{\gamma})}{\varepsilon_n}$$

where we have written $g(\widehat{\beta},\widehat{\gamma})=g(\widehat{\theta})$. Thus our estimate of the full directional derivative is

$$\widehat{g}_{\theta}(h)=x(\widehat{\gamma})'h_{\beta}+\frac{g(\widehat{\beta},\widehat{\gamma}+h_{\gamma}\varepsilon_n)-g(\widehat{\beta},\widehat{\gamma})}{\varepsilon_n}.$$

Evaluated at $h=(h_{\beta},h_{\gamma})=\left(\sqrt{n}\left(\widehat{\beta}^{*}-\widehat{\beta}\right),\sqrt{n}\left(\widehat{\gamma}^{*}-\widehat{\gamma}\right)\right)$, the bootstrap estimate of the distribution of $g(\widehat{\theta})-g(\theta_0)$ is

$$r^{*}=x(\widehat{\gamma})'\left(\widehat{\beta}^{*}-\widehat{\beta}\right)+\frac{g(\widehat{\beta},\widehat{\gamma}+\sqrt{n}\varepsilon_n\left(\widehat{\gamma}^{*}-\widehat{\gamma}\right))-g(\widehat{\beta},\widehat{\gamma})}{\sqrt{n}\varepsilon_n}.$$

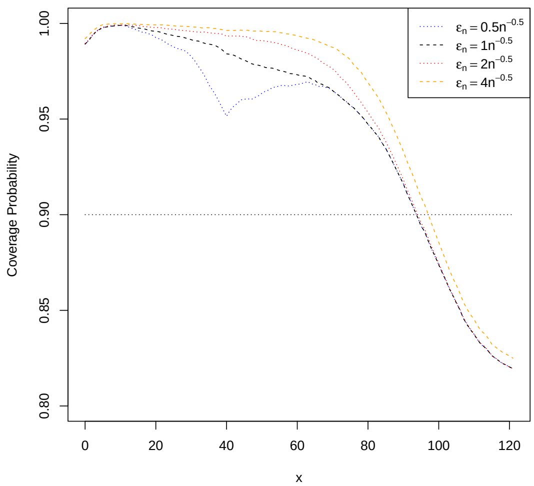
Hong and Li (2015) show that if the function $\phi(\theta)$ is a convex function of θ then upper one-sided confidence intervals for θ constructed using r^{*} uniformly control size, but lower one-sided confidence intervals will not, and the converse holds when $\phi(\theta)$ is a concave function. In our case, the function $g(\theta)$ is a convex function of γ if $\beta_1\leq\beta_2$ but is concave if $\beta_1\geq\beta_2$. Thus neither one-sided confidence intervals will uniformly control size. Their suggestion is to instead use two-sided symmetric confidence intervals, as these will be asymptotically conservative.

Let $q^{*}_{1-\alpha}$ denote the $(1-\alpha)^{th}$ quantile of the variable $|r^{*}|$. The symmetric bootstrap confidence interval for $g(\theta_0)$ is $g(\widehat{\theta})\pm q^{*}_{1-\alpha}$. We call this the numerical delta method bootstrap interval.

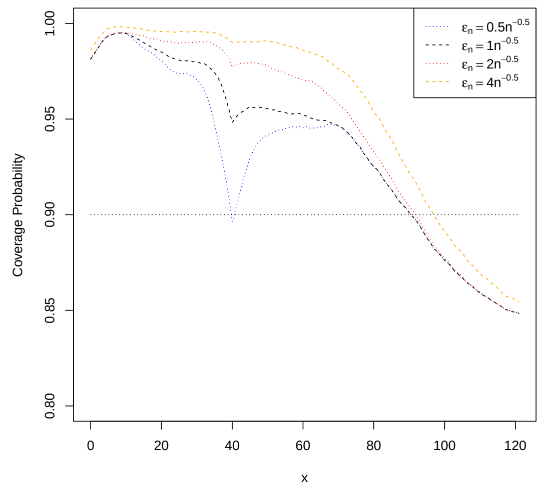
An important issue is setting ε_n, the increment for the numerical derivative. While the general theory of Hong and Li (2015) requires $\sqrt{n}\varepsilon_n\to\infty$, in the context of this model they recommend $\varepsilon_n=cn^{-1/2}$, though they provide no guidance for selection of c. We follow their advice and set $\varepsilon_n=cn^{-1/2}$, and assess the choice of the constant c via simulation. We implemented the numerical delta method bootstrap, and calculated in the same simulation as Figure 5 the pointwise coverage probabilities of nominal 90% confidence intervals with $c=0.5$, $c=1$, $c=2$, and $c=4$. The coverage probabilities are plotted in Figure 6, where each panel corresponds to a value of β_2.

Figure 6 shows that the coverage probabilities are monotonically increasing in c, so a more conservative (and larger) confidence interval can be found by increasing c. The coverage probabilities are also non-uniform in the regression argument x, with over-coverage for small x and under-coverage for large x. There is also a downward spike in coverage probabilities at $x=40$, which is the location of the true threshold and is consistent with the non-normal distribution theory. This downward spike is lessened for larger c but is also increasing with β_2. The simulations indicate the appropriate choice of c depends on β_2 thus making a generic choice difficult. In our application, our point estimate for $\beta_2-\beta_1$ (which is the appropriate analog) is -0.10, which is roughly intermediate between panels (b) and (c). If we focus on obtaining correct coverage at $x=40$, the results suggest $c=1$ is reasonable, with over coverage for $\beta_2=-0.08$ and mild undercoverage for $\beta_2=-0.12$. We thus set $c=1$ and hence $\varepsilon_n=n^{-1/2}$. For other applications, we suggest evaluating the coverage via simulation.

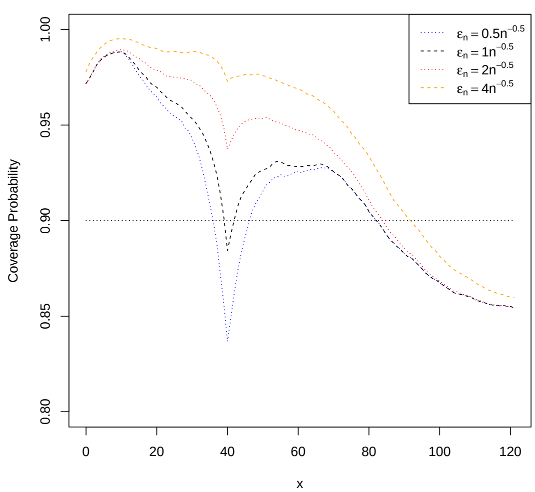
The theoretical argument for the numerical delta method bootstrap is that it has asymptotically conservative coverage for all values of x, even at $x=\gamma_0$. For values of x distant from γ_0, however, it is possible that the classical bootstrap will have better (e.g. less conservative) coverage. There



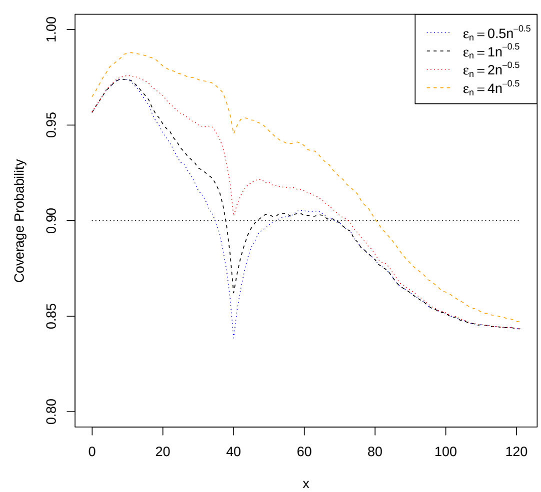
(a) $\beta_2 = -0.04$



(b) $\beta_2 = -0.08$



(c) $\beta_2 = -0.12$



(d) $\beta_2 = -0.16$

Figure 6: Coverage Probabilities of Nominal 90% Numerical Delta Method Bootstrap Intervals

is no clear way, however, to mix bootstrap methods, and the value of γ_0 is unknown in practice, so we recommend using the numerical delta method bootstrap for confidence intervals for all values of x.

We now summerize the steps for the numerical delta method bootstrap.

Algorithm 3: Numerical Delta Method Bootstrap Confidence Intervals for Regression Kink Function at a fixed value of (x, z)

1. Follow steps 1-3 of the wild bootstrap of Algorithm 2
2. Set $x(\widehat{\gamma}) = ((x - \widehat{\gamma})_-, (x - \widehat{\gamma})_+, z')'$
3. Set $\varepsilon_n = cn^{-1/2}$
4. Set $r^* = x(\widehat{\gamma})' \left(\widehat{\beta}^* - \widehat{\beta}\right) + \left(g(\widehat{\beta}, \widehat{\gamma} + \sqrt{n}\varepsilon_n (\widehat{\gamma}^* - \widehat{\gamma})) - g(\widehat{\beta}, \widehat{\gamma})\right) / \sqrt{n}\varepsilon_n$
5. Repeat B times, so as to obtain a sample of simulated estimates r^*
6. Calculate $q^*_{1-\alpha}$, the $(1-\alpha)^{th}$ quantile of the simulated $|r^*|$
7. Set the $1 - \alpha$ confidence interval for $g(\theta_0)$ as $[g(\widehat{\theta}) - q^*_{1-\alpha}, g(\widehat{\theta}) + q^*_{1-\alpha}]$

This is numerically quite simple to implement, although step 4 is unusual for a bootstrap procedure. For confidence interval bands, steps 2 through 7 need to be repeated for each value of (x, z) considered.

Formally, the merging of the Hong-Li numerical delta method with this specific wild bootstrap for serially dependent data has not been studied. More work would be welcome to understand its properties.

We implemented the numerical delta method bootstrap using the rule $\varepsilon_n = cn^{-1/2}$ on the U.S. growth regression for $c = 0.5$, $c = 1$, $c = 2$, and $c = 4$, and plotted the pointwise 90% confidence intervals in Figure 7. Numerically, this means that Algorithm 3 was applied with z set at the sample mean of z_t and x evaluated on a grid from 1 to 120. The confidence intervals widen as ε_n increases, except at $x = \widehat{\gamma}$.

It is important to interpret these confidence intevals as pointwise in x (not uniform). There have been recent advances in developing uniform inference methods for nonparametric regression by Belloni, et. al. (2015), but these methods do not apply to the numerical delta method bootstrap.

Following the recommendation of our simulation study we take the second estimator (with $\varepsilon_n = n^{-1/2}$) as our preferred choice, and this is plotted in Figure 2 with the dashed blue lines. The confidence intervals are sufficiently wide that it is unclear if the true regression function is flat or downward sloping. The confidence intervals reveal that by using the U.S. longspan data alone, the estimates of the regression kink model are not sufficiently precise to make a strong conclusion about whether or not there is a negative effect of debt levels on GDP growth rates.

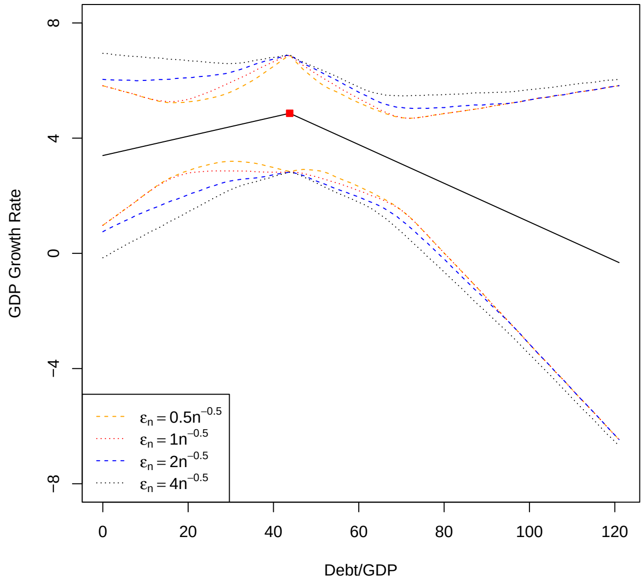


Figure 7: Estimated Regression Kink Function and 90% Numerical Delta Method Confidence Intervals

7 Conclusion

This paper developed a theory of estimation and inference for the regression kink model with an unknown threshold and applied it to study the growth & debt threshold problem of Reinhart and Rogoff (2010). An interesting theoretical contribution is the finding that the estimate of the regression function is non-normal due to non-differentiability, and confidence intervals can be formed using the recent inference methods of Fang and Santos (2014) and Hong and Li (2015).

We apply the method to the long-span time-series U.S. data developed by Reinhart and Rogoff (2010). Our point estimates are consistent with the Reinhart-Rogoff hypothesis of a growth slowdown when debt levels exceed a threshold. However, the formal evidence for the presence of the threshold effect is inconclusive, and our confidence intervals for the regression function are sufficiently wide that the effect of debt on growth is difficult to detect.

An important cavaet is that our empirical results are based only on a single time-series (the United States), thus ignoring the information in other nations' experiences. This has the advantage of not imposing homogeneity, but also may reduce precision and power. It would be useful to extend the results here to panel data analysis.

8 Appendix

Proof of Theorem 2:

By the definition (4), $\widehat{\theta}$ minimizes $S_n(\theta)$ defined in (3), which we can write as $S_n(\theta) =$

$\frac{1}{n}\sum_{t=1}^{n} e_t(\theta)^2$ with $e_t(\theta) = y_t - \beta' x_t(\gamma)$. Thus $\widehat{\theta}$ approximately solves the first-order condition $\frac{1}{n}\sum_{t=1}^{n} m_t(\widehat{\theta}) = 0$ where $m_t(\theta) = H_t(\theta)e_t(\theta)$. The pseudo-true value θ_0 minimizes $L(\theta) = L(\beta, \gamma)$ and thus solves $m(\theta_0)$ where $m(\theta) = \frac{1}{2}\frac{\partial}{\partial\theta}L(\theta) = Em_t(\theta)$. Define $Q(\theta) = -\frac{\partial}{\partial\theta'}m(\theta)$.

Andrews (1994), Section 3.2, shows that Theorem 2 will hold under the following conditions.

Condition 1 $\widehat{\theta} \to_p \theta_0$

Condition 2 $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} H_t e_t \to_d N(0, S)$

Condition 3 *$Q(\theta)$ is continuous in θ and $Q(\theta_0) = Q$*

Condition 4 *$v_n(\theta) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(m_t(\theta) - m(\theta))$ is stochastically equicontinuous*

We first establish Condition 1. Since $x_t(\gamma)$ is continuous in γ, $e_t(\theta)$ and $e_t(\theta)^2$ are continuous in θ. Recalling definition (2) we have the simple bound $\|x_t(\gamma)\|^2 = z_t' z_t + (x_t - \gamma)^2 \leq \|z_t\|^2 + x_t^2 + \gamma^2$. Then by the C_r and Cauchy-Schwarz inequalities,

$$e_t(\theta)^2 \leq 2y_t^2 + 2\left|\beta' x_t(\gamma)\right|^2 \leq 2y_t^2 + 2\overline{\beta}^2\left(\|z_t\|^2 + x_t^2 + \overline{\gamma}^2\right) \tag{18}$$

where $\overline{\beta} = \sup\{\|\beta\| : \beta \in B\}$ and $\overline{\gamma} = \sup\{|\gamma| : \gamma \in \Gamma\}$ which are finite under Assumption 2.2. The bound in (18) has finite expectation under Assumption 1.2. Then by Lemma 2.4 of Newey and McFadden (1994) (which is based on Lemma 1 of Tauchen (1985)), $L(\theta) = Ee_t(\theta)^2$ is continuous and $\sup_{\theta \in B\times\Gamma}|S_n(\theta) - L(\theta)| \to_p 0$ as $n \to \infty$. Lemma 2.4 of Newey and McFadden is stated for i.i.d. observations, but the result only requires application of a weak law of large numbers, which holds under Assumption 1 by the Ergodic theorem.

Given the compactness of $B \times \Gamma$ and the uniqueness of the minimum θ_0 (by Assumptions 1.2 and 1.1, respectively), Theorem 2.1 of Newey and McFadden (1994) establishes that $\widehat{\theta} \to_p \theta_0$ as $n \to \infty$, which is Condition 1.

Condition 2 follows by the Herrndorf's (1984) central limit theorem for strong mixing processes which holds under Assumption 1.1-1.2.

The following bound will be useful to establish Conditions 3 and 4. Let w_t be any vector whose elements are pairwise products of the variables (y_t, x_t, z_t), and set $q = r/(r-1)$. Note that Assumption 1.2 implies $\left(E\|w_t\|^{2r}\right)^{1/r} \leq C$ for some $C < \infty$. Let $F(\gamma)$ denote the distribution function of x_t, which satisfies $F(\gamma_2) - F(\gamma_1) \leq \overline{f}|\gamma_2 - \gamma_1|$ under Assumption 1.4. By Hölder's inequality, we find

$$\begin{aligned} E\|w_t 1(\gamma_1 \leq x_t \leq \gamma_2)\|^2 &\leq \left(E\|w_t\|^{2r}\right)^{1/r}\left(E|1(\gamma_1 \leq x_t \leq \gamma_2)|^q\right)^{1/q} \\ &\leq C\left(F(\gamma_2) - F(\gamma_1)\right)^{1/q} \\ &\leq C\overline{f}^{1/q}|\gamma_2 - \gamma_1|^{1/q} \end{aligned} \tag{19}$$

One implication of this bound is that $E(w_t 1(x_t \leq \gamma))$ is uniformly continuous in γ.

We now establish Condition 3. Note that

$$Q(\theta) = -\frac{\partial}{\partial\theta'} E\left(H_t(\theta)e_t(\theta)\right)$$
$$= E\left(H_t(\theta)H_t(\theta)'\right) + E\begin{pmatrix} 0 & 0 & 0 & e_t(\theta)1\left(x_t \le \gamma\right) \\ 0 & 0 & 0 & e_t(\theta)1\left(x_t > \gamma\right) \\ 0 & 0 & 0 & 0 \\ e_t(\theta)1\left(x_t \le \gamma\right) & e_t(\theta)1\left(x_t > \gamma\right) & 0 & 0 \end{pmatrix}$$

The elements of this matrix are quadratic functions of β, and functions of γ through moments of the form $E\left(w_t 1\left(x_t \le \gamma\right)\right)$ where w_t is a product of the variables (y_t, x_t, z_t). Since (19) holds we see that $Q(\theta)$ is continuous in θ. Evaluated at θ_0 we find $Q(\theta_0) = Q$. Thus Condition 3 holds.

We now establish Condition 4 by appealing to Theorem 1, Application 4, case (2.15) of Doukan, Massart, and Rio (1995). Notice that $m_t(\theta)$ is a quadratic function of β, so the only issue is establishing stochastic equicontinuity with respect to γ. We thus so we simplify notation by writing $m_t(\theta)$ as $m_t(\gamma)$. Notice that we can write $m_t(\gamma) = w_t 1\left(x_t \le \gamma\right)$ where w_t is vector whose elements are products of the variables (y_t, x_t, z_t). Under Assumption 1.2, $m_t(\gamma)$ has a bounded $2r^{th}$ moment and thus satisfies the needed envelope condition.

For any $\delta > 0$ set $N(\delta) = \delta^{-2/q}$ and set γ_k, $k = 1, ..., N$, to be an equally spaced grid on Γ. Note that the distance between the gridpoints is $O(N(\delta)^{-1})$. Define $m^*_{tk} = \min[m_t(\gamma_{k-1}), m_t(\gamma_k)]$ and $m^{**}_{tk} = \max[m_t(\gamma_{k-1}), m_t(\gamma_k)]$. Then for each $\gamma \in \Gamma$ there is a $k \in [1, ..., N]$ such that $m^*_{tk} \le m_t(\gamma) \le m^{**}_{tk}$. Thus $[m^*_{tk}, m^{**}_{tk}]$ brackets $m_t(\gamma)$. By their construction, the fact that $m_t(\gamma) = w_t 1\left(x_t \le \gamma\right)$, and applying (19),

$$\begin{aligned} E\left\|m^{**}_{tk} - m^*_{tk}\right\|^2 &= E\left\|m_t(\gamma_k) - m_t(\gamma_{k-1})\right\|^2 \\ &\le E\left\|w_t 1\left(\gamma_{k-1} \le x_t \le \gamma_k\right)\right\|^2 \\ &\le C\overline{f}^{1/q}\left|\gamma_k - \gamma_{k-1}\right|^{1/q} \\ &\le O\left(N(\delta)^{-q}\right) = O\left(\delta^2\right). \end{aligned}$$

This means that $N(\delta) = \delta^{-2/q}$ are the L^2 bracketing numbers and $H_2(\delta) = \ln N(\delta) = |\log\delta|$ is the metric entropy with bracketing for the class $\{m_t(\gamma) : \gamma \in \Gamma\}$. This and Assumption 1.1 imply equation (2.15) of Doukan, Massart, and Rio (1995) and thus their Theorem 1, establishing stochastic equicontinuity of $v_n(\theta)$ and hence Condition 4.

We have established that Conditions 1-4 hold under Assumptions 1 and 2. As discussed above, this is sufficient to establish Theorem 2. ■

References

[1] Andrews, Donald W.K. (1994): "Empirical process methods in econometrics," *Handbook of Econometrics, Vol IV*, R.F. Engle and D.L McFadden, eds., 2247-2294. New York: Elsevier.

[2] Belloni, Alexandre, Victor Chernozhukov, Denis Chetverikov and Kengo Kata (2015): "On the asymptotic theory for least squares series: Pointwise and uniform results," *Journal of Econometrics*, forthcoming.

[3] Caner, Mehmet, Thomas Grennes, and Fritzi Koehler-Geib (2010): "Finding the tipping point- when sovereign debt turns bad," Policy Research Working Paper Series 5391, The World Bank.

[4] Caner, Mehmet and Bruce E. Hansen (2004): "Instrumental Variable Estimation of a Threshold Model," *Econometric Theory*, 20, 813-843

[5] Card, David, David Lee, Zhuan Pei, Andrea Weber (2012): "Nonlinear policy rules and the identification and estimation of causal effects in a generalized regression kind design," NBER Working Paper 18564.

[6] Card, David, Alexandre Mas and Jesse Rothstein (2008): "Tipping and the dynamics of segregation," *Quarterly Journal of Economics*, 123, 177-218.

[7] Cecchetti, Stephen G, M. S. Mohanty, and Fabrizio Zampolli (2011). "The real effects of debt," *Economic Symposium Conference Proceedings: Federal Reserve Bank of Kansas City*, 145-196.

[8] Chan, Kung-Sik (1990): "Testing for threshold autoregression," *The Annals of Statistics* 18, 1886-1894.

[9] Chan, Kung-Sik (1991): "Percentage points of likelihood ratio tests for threshold autoregression," *Journal of the Royal Statistical Society, Series B*, 53, 691-696.

[10] Chan, Kung-Sik (1993): "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model," *The Annals of Statistics*, 21, 520-533.

[11] Chan, Kung-Sik and Howell Tong (1990): "On likelihood ratio tests for threshold autoregression," *Journal of the Royal Statistical Society B*, 52, 469-476.

[12] Chan, Kung-Sik and Ruey S. Tsay (1998): "Limiting properties of the least squares estimator of a continuous threshold autoregressive model," *Biometrika*, 45, 413-426.

[13] Chiu, Grace, Richard Lockhart, and Richard Routledge (2006): "Bent-cable regression theory and applications," *Journal of the American Statistical Association*, 101, 542-553.

[14] Cox, Donald, Bruce E. Hansen and Emmanuel Jimenez (2004): "How responsive are private transfers to income?" *Journal of the Public Economics*, 88, 2193-2219.

[15] Doukhan, P., P. Massart and E. Rio (1995): "Invariance principles for absolutely regular empirical processes," *Annales de l'Institut H. Poincare, Probabiltes et Statistiques,* 31, 393-427.

[16] Fang, Zheng (2014) "Optimal plug-in estimators of directionally differentiable functionals," working paper, UCSD.

[17] Fang, Zheng and Andres Santos (2014): "Inference on directionally differentiable functions," working paper, UCSD.

[18] Feder, Paul I. (1975): "On asymptotic distribution theory in segmented regression problems – identified case," *Annals of Statistics*, 3, 49-83.

[19] Ganong, Peter and Simon Jager (2014): "A permutation test and estimation alternatives for the regression kink design," working paper, Harvard University.

[20] Hansen, Bruce E. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, 64, 413-430.

[21] Hansen, Bruce E. (1999): "Threshold effects in non-dynamic panels: Estimation, testing and inference," *Journal of Econometrics*, 93, 345-368

[22] Hansen, Bruce E. (2000): "Sample splitting and threshold estimation," *Econometrica*, 68, 575-603.

[23] Herrndorf, N. (1984): "A functional central limit theorem for weakly dependent sequences of random variables," *The Annals of Probability*, 12, 141-153.

[24] Hinkley, David V. (1969): "Inference about the intersection in two-phase regression," *Biometrika*, 56, 495-504.

[25] Hinkley, David V. (1971): "Inference in two-phase regression," *Journal of the American Statistical Association*, 66, 736-743.

[26] Hirano, Keisuke and Jack Porter (2012): "Impossibility results for nondifferentiable functionals," *Econometrica*, 80, 1769-1790.

[27] Hong, Han and Jessie Li (2015): "The numerical directional delta method," working paper, Stanford University.

[28] Landais, Camille (2014): "Assessing the welfare effects of unemployment benefits using the regression kink design," *American Economic Review: Economic Policy*, forthcoming.

[29] Lee, Sokbae, Hyunmin Park, Myung Hwan Seo, and Youngki Shin (2014): "A contribution to the Reinhart and Rogoff debate: not 90 percent but maybe 30 percent," CEMMAP working paper.

[30] Lee, Sokbae, Myung Hwan Seo, and Youngki Shin (2011): "Testing for threshold effects in regression models," *Journal of the American Statistical Association*, 106, 220-231.

[31] Lin, Tzu-Chi (2014): "High-dimensional threshold quantile regression with an application to debt overhang and economic growth," working paper, University of Wisconsin.

[32] Newey, Whitney K. and Daniel McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of Econometrics, Vol IV*, R.F. Engle and D.L McFadden, eds., 2113-2245. New York: Elsevier.

[33] Porter, Jack and Ping Yu (2014): "Regression discontinuity designs with unknown discontinuity points: Testing and estimation," working paper, University of Wisconsin.

[34] Ramirez-Rondan, Nelson (2013): "Maximum likelihood estimation of a dynamic panel threshold model," working paper, Central Bank of Peru, https://sites.google.com/site/teltoramirezrondan/.

[35] Reinhart, Carmen M. and Kenneth S. Rogoff (2010): "Growth in a time of debt," *American Economic Review: Papers and Proceedings*, 100, 573-578.

[36] Seo, Myung Hwan and Oliver Linton (2007): "A smoothed least squares estimator for threshold regression models," *Journal of Econometrics*, 141 704-735.

[37] Shapiro, Alexander (1990): "On concepts of directional differentiability," *Journal of Optimization Theory and Applications*, 66, 477-487.

[38] Shapiro, Alexander (1991): "Asymptotic analysis of stochastic programs," *Annals of Operations Research*, 30, 169-186.

[39] Stambaugh, Robert F. (1999): "Predictive regressions," *Journal of Financial Economics*, 54, 375-421.

[40] Tauchen, George (1985): "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415-443.

[41] Tong, Howell (1983): *Threshold Models in Non-linear Time Series Analysis. Lecture Notes in Statistics*, 21, Berlin: Springer.

[42] Tong, Howell (1990): *Non-Linear Time Series: A Dynamical System Approach*, Oxford University Press, Oxford.

[43] Woutersen, Tiemen and John C. Ham (2013): "Calculating confidence intervals for continuous and discontinuous functions of parameters," working paper, University of Arizona.