Threshold Effects in Non-Dynamic Panels: Estimation, Testing, and Inference

Bruce E. Hansen[*]
Boston College[†]

May 1997
Revised: February 1998

Abstract

Threshold regression methods are developed for non-dynamic panels with individual-specific fixed effects. Least squares estimation of the threshold and regression slopes is proposed using fixed-effects transformations. A non-standard asymptotic theory of inference is developed which allows construction of confidence intervals and testing of hypotheses. The methods are applied to a 15-year sample of 565 U.S. firms to test whether financial constraints affect investment decisions.

[†]Department of Economics, Boston College, Chestnut Hill MA 02167-3806. bruce.hansen@bc.edu. Homepage: http://www2.bc.edu/~hansenb/

1 Introduction

Are regression functions identical across all observations in a sample, or do they fall into discrete classes? This question may be addressed using threshold regression techniques. Threshold regression models specify that individual observations can be divided into classes based on the value of an observed variable. Despite their intuitive appeal, econometric techniques have not been well developed for threshold regression.

This paper introduces econometric techniques appropriate for threshold regression with panel data. Least squares estimation methods are described. An asymptotic distribution theory is derived which is used to construct confidence intervals for the parameters. A bootstrap method to assess the statistical significance of the threshold effect is also described. The methods are similar to those developed in earlier work by the author (Hansen, 1996, 1997).

The methods are used to investigate whether financial constraints affect the investment practices of firms. The classical theory of the firm suggests that financing should have no allocative effects (e.g., the Modigliani-Miller theorem). Investment decisions should only be based on the marginal Q of a specific project, since banks will be willing to extend finance. In the context of imperfect information, external financing may be limited, and debt-constrained firms may need to finance investment out of cash flow. If this is the case, investment will be correlated with cash flow for constrained firms. This observation led Fazzari, Hubbard and Petersen (1988) to divide a sample of U.S. firms into classes based on their degree of financial constraints and estimate the differing effects of cash flow on investment among these classes. Their analysis suffered from two problems. First, they used an endogenous variable (dividend to income ratio) rather than an exogenous variable to form their sample splits. Second, they used an ad hoc method to select their sample splits. We repeat their analysis on an analogous data set using appropriate econometric techniques and find qualitatively similar results.

Other authors have investigated the implications of non-linear q models of investment. Abel and Eberly (1994) propose a model which implies that the response of investment to q may be non-linear in q. Abel and Eberly (1996) use panel data to estimate a similar model, and find evidence for non-linearities in the investment function. Barnett and Sakellaris (1995) find similar results using a threshold regression approach. Hu and Schiantarelli (1998) use a

switching regression framework to study the same problem. Our paper extends and reinforces this growing literature.

The next section introduces the model and notation. Section 3 discusses estimation by fixed effects. Section 4 outlines our asymptotic theory of inference. A distribution theory is developed for the threshold estimate and the slope coefficients. Section 5 reports the empirical application to firms' investment decisions. Section 6 concludes. Proofs of the asymptotic theory are provided in the appendix. GAUSS programs and data which replicate the empirical work are available from the author's homepage.

2 Model

The observed data are from a balanced[1] panel $\{y_{it}, q_{it}, x_{it} : 1 \leq i \leq n, 1 \leq t \leq T\}$. The subscript i indexes the individual and the subscript t indexes time. The dependent variable y_{it} is scalar, the threshold variable q_{it} is scalar, and the regressor x_{it} is a k vector. The structural equation of interest is

$$y_{it} = \mu_i + \beta'_1 x_{it} I (q_{it} \leq \gamma) + \beta'_2 x_{it} I (q_{it} > \gamma) + e_{it}. \tag{1}$$

where $I(\cdot)$ is the indicator function. An alternative intuitive way of writing (1) is

$$y_{it} = \begin{cases} \mu_i + \beta'_1 x_{it} + e_{it}, & q_{it} \leq \gamma \\ \mu_i + \beta'_2 x_{it} + e_{it}, & q_{it} > \gamma. \end{cases}$$

Another compact representation of (1) is to set

$$x_{it}(\gamma) = \begin{cases} x_{it} I (q_{it} \leq \gamma) \\ x_{it} I (q_{it} > \gamma) \end{cases}$$

and $\beta = (\beta'_1 \quad \beta'_2)'$ so that (1) equals

$$y_{it} = \mu_i + \beta' x_{it}(\gamma) + e_{it}. \tag{2}$$

The observations are divided into two "regimes" depending on whether the threshold variable q_{it} is smaller or larger than the threshold γ. The regimes are distinguished by differing regression slopes, β_1 and β_2. For the identification of β_1 and β_2, it is required that the elements of x_{it} are not time-invariant. We also assume that the threshold variable q_{it} is not time-invariant. The error e_{it} is assumed to be independent and identically distributed (iid) with mean zero and finite variance σ^2. The iid assumption excludes lagged dependent variables from x_{it}. It is unclear how to extend the results to allow for dynamic models and/or heteroskedastic errors. The analysis is asymptotic with fixed T as $n \to \infty$.

3 Estimation

3.1 Least-Squares Estimation

One traditional method to eliminate the individual effect μ_i is to remove individual-specific means. While straightforward in linear models, the non-linear specification (1) calls for a

---

[1]It is unknown if the results extend to unbalanced panels.

more careful treatment. Note that taking averages of (1) over the time index t produces

$$\overline{y}_i = \mu_i + \beta' \overline{x}_i(\gamma) + \overline{e}_i \tag{3}$$

where $\overline{y}_i = T^{-1} \sum_{t=1}^T y_{it}$, $\overline{e}_i = T^{-1} \sum_{t=1}^T e_{it}$, and

$$
\begin{aligned}
\overline{x}_i(\gamma) &= \frac{1}{T} \sum_{t=1}^T x_{it}(\gamma) \\
&= \begin{pmatrix} \dfrac{1}{T} \sum_{t=1}^T x_{it} I\left(q_{it} \leq \gamma\right) \\ \dfrac{1}{T} \sum_{t=1}^T x_{it} I\left(q_{it} > \gamma\right) \end{pmatrix}.
\end{aligned}
$$

Taking the difference between (2) and (3) yields

$$y_{it}^* = \beta' x_{it}^*(\gamma) + e_{it}^* \tag{4}$$

where

$$y_{it}^* = y_{it} - \overline{y}_i,$$

$$x_{it}^*(\gamma) = x_{it}(\gamma) - \overline{x}_i(\gamma),$$

and

$$e_{it}^* = e_{it} - \overline{e}_i.$$

Let

$$
y_i^* = \begin{bmatrix} y_{i2}^* \\ \vdots \\ y_{iT}^* \end{bmatrix}, \qquad
x_i^*(\gamma) = \begin{bmatrix} x_{i2}^*(\gamma)' \\ \vdots \\ x_{iT}^*(\gamma)' \end{bmatrix}, \qquad
e_i^* = \begin{bmatrix} e_{i2}^* \\ \vdots \\ e_{iT}^* \end{bmatrix}
$$

denote the stacked data and errors for an individual, with one time period deleted. Then let Y^*, $X^*(\gamma)$ and e^* denote the data stacked over all individuals, for example

$$
X^*(\gamma) = \begin{bmatrix} x_1^*(\gamma) \\ \vdots \\ x_i^*(\gamma) \\ \vdots \\ x_n^*(\gamma) \end{bmatrix}
$$

Using this notation, (4) is equivalent to

$$Y^* = X^*(\gamma)\beta + e^*. \tag{5}$$

4

For any given γ, the slope coefficient β can be estimated by ordinary least squares (OLS). That is,

$$\hat{\beta}(\gamma) = \left(X^*(\gamma)'X^*(\gamma)\right)^{-1}X^*(\gamma)'Y^*. \tag{6}$$

The vector of regression residuals is

$$\hat{e}^*(\gamma) = Y^* - X^*(\gamma)\hat{\beta}(\gamma)$$

and the sum of squared errors is

$$
\begin{aligned}
S_1(\gamma) &= \hat{e}^*(\gamma)'\hat{e}^*(\gamma) \\
&= Y^{*\prime}\left(I - X^*(\gamma)'\left(X^*(\gamma)'X^*(\gamma)\right)^{-1}X^*(\gamma)'\right)Y^*.
\end{aligned} \tag{7}
$$

Chan (1993) and Hansen (1997) recommend estimation of γ by least-squares. This is easiest to achieve by minimization of the concentrated sum of squared errors (7). Hence the least-squares estimators of γ is

$$\hat{\gamma} = \operatorname*{argmin}_{\gamma} S_1(\gamma). \tag{8}$$

It is undesirable for a threshold $\hat{\gamma}$ to be selected which sorts too few observations into one or the other regime. This possibility can be excluded by restricting the search in (8) to values of γ such that a minimal percentage of the observations (say, 1% or 5%) lie in both regimes.

Once $\hat{\gamma}$ is obtained, the slope coefficient estimate is $\hat{\beta} = \hat{\beta}(\hat{\gamma})$. The residual vector is $\hat{e}^* = \hat{e}^*(\hat{\gamma})$ and residual variance

$$\hat{\sigma}^2 = \frac{1}{n(T-1)}\hat{e}^{*\prime}\hat{e}^* = \frac{1}{n(T-1)}S_1(\hat{\gamma}). \tag{9}$$

3.2   Computation Issues

The computation of the least squares estimate of the threshold γ involves the minimization problem (8). Since the sum of squared error function $S_1(\gamma)$ depends on γ only through the indicator functions $I(q_{it} \le \gamma)$, the sum of squared error function is a step function with at most nT steps, with the steps occurring at distinct values of the observed threshold variable q_{it}. Thus the minimization problem (8) can be reduced to searching over values of γ equalling the (at most nT) distinct values of q_{it} in the sample.

To implement the minimization, the following approach may be taken. Sort the distinct values of the observations on the threshold variable q_{it}. Eliminate the smallest and largest $\eta\%$ for some $\eta > 0$. The remaining N values constitute the values of γ which can be searched for $\hat{\gamma}$. For each of these N values, the regressions (6) are estimated yielding the sum of squared errors (7). The smallest value of the latter yields the estimate $\hat{\gamma}$.

In practice, N may be a very large number, and the optimization search describe above may be numerically intensive. A simplifying shortcut which yields nearly identical results is to restrict the search to a smaller set of values of γ. Instead of searching over all values of q_{it} (between the $\eta\%$ and $(1-\eta)\%$ quantile) the search may be limited to specific quantiles, perhaps integer-valued. This greatly reduces the number of regressions performed in the search. The estimates from such an approximation are likely to be sufficiently precise for most applications of interest. For the empirical work reported in section 4, we used the grid $\{1.00\%, 1.25\%, 1.50\%, 1.75\%, 2\%, ..., 99.0\%\}$ which contains 393 quantiles.

4   Inference

4.1   Testing for a Threshold

It is important to determine whether the threshold effect is statistically significant. The hypothesis of no threshold effect in (1) can be represented by the linear constraint

$$H_0 : \beta_1 = \beta_2.$$

Under H_0 the threshold γ is not identified, so classical tests have non-standard distributions. This is typically called the "Davies' Problem" (see Davies (1977, 1987)) and has been recently investigated by Andrews and Ploberger (1994) and Hansen (1996). The fixed-effects equations (4) fall in the class of models considered by Hansen (1996) who suggested a bootstrap to simulate the asymptotic distribution of the likelihood ratio test.

Under the null hypothesis of no threshold, the model is

$$y_{it} = \mu_i + \beta_1' x_{it} + e_{it}. \tag{10}$$

After the fixed-effect transformation is made, we have

$$y_{it}^* = \beta_1' x_{it}^* + e_{it}^*. \tag{11}$$

The regression parameter β_1 is estimated by OLS, yielding estimate $\tilde{\tilde{\beta}}_1$, residuals $\tilde{e}_{it}^*$ and sum of squared errors $S_0 = \tilde{e}^{*\prime} \tilde{e}^*$. The likelihood ratio test of H_0 is based on

$$F_1 = \frac{S_0 - S_1(\hat{\gamma})}{\hat{\sigma}^2}. \tag{12}$$

The asymptotic distribution of F_1 is non-standard, and strictly dominates the χ_k^2 distribution. Unfortunately, it appears to depend in general upon moments of the sample and

thus critical values cannot be tabulated. Hansen (1996) shows that a bootstrap procedure attains the first-order asymptotic distribution, so p-values constructed from the bootstrap are asymptotically valid[2]. Given the panel nature of the data we recommend the following implementation of the bootstrap. Treat the regressors x_{it} and threshold variable q_{it} as given, holding their values fixed in repeated bootstrap samples. Take the regression residuals $\hat{e}_{it}^*$, and group them by individual: $\hat{e}_i^* = (\hat{e}_{i1}^*, \hat{e}_{i2}^*, ..., \hat{e}_{iT}^*)$. Treat the sample $\{\hat{e}_1^*, \hat{e}_2^*, ..., \hat{e}_n^*\}$ as the empirical distribution to be used for bootstrapping. Draw (with replacement) a sample of size n from the empirical distribution and use these errors to create a bootstrap sample under H_0. (Notice that the test statistic F_1 does not depend on the parameter β_1 under H_0, so any value of β_1 may be used.) Using the bootstrap sample, estimate the model under the null (11) and alternative (4) and calculate the bootstrap value of the likelihood ratio statistic F_1 (12). Repeat this procedure a large number of times and calculate the percentage of draws for which the simulated statistic exceeds the actual. This is the bootstrap estimate of the asymptotic p-value for F_1 under H_0. The null of no threshold effect is rejected if the p-value is smaller than the desired critical value.

4.2 Asymptotic Distribution of Threshold Estimate

When there is a threshold effect ($\beta_1 \neq \beta_2$) Chan (1993) and Hansen (1997) have shown that $\hat{\gamma}$ is consistent for γ_0 (the true value of γ) and that the asymptotic distribution is highly non-standard. Hansen (1997) argues that the best way to form confidence intervals for γ is to form the "no-rejection region" using the likelihood ratio statistic for tests on γ. To test the hypothesis $H_0 : \gamma = \gamma_0$, the likelihood ratio test is to reject for large values of $LR_1(\gamma_0)$ where

$$LR_1(\gamma) = \frac{S_1(\gamma) - S_1(\hat{\gamma})}{\hat{\sigma}^2}. \tag{13}$$

Note that the statistic (13) is testing a different hypothesis from the statistic (12) introduced in the previous section. $LR_1(\gamma_0)$ is testing $H_0 : \gamma = \gamma_0$ while F_1 is testing $H_0 : \beta_1 = \beta_2$.

Theorem 1 *Under Assumptions 1-8 given in the Appendix, and $H_0 : \gamma = \gamma_0$,*

$$LR_1(\gamma) \rightarrow_d \xi$$

---

[2]Since the asymptotic distribution is non-pivotal, bootstrap size will not have an accelerated rate of convergence relative to conventional asymptotic approximations. A referee suggested that pre-pivoting as in Beran (1987) may improve the convergence rate. This is an interesting suggestion and would be a constructive subject for future research.

as $n \to \infty$, where ξ is a random variable with distribution function

$$P\left(\xi \le x\right) = \left(1 - \exp(-x/2)\right)^2 . \tag{14}$$

Theorem 1 shows that the asymptotic distribution of the likelihood ratio statistic is non-standard yet free of nuisance parameters. The technical assumptions include the rather unusual condition that $(\beta_2 - \beta_1) \to 0$ as $n \to \infty$, and is borrowed from the changepoint literature. The condition means that the difference in the slopes between the two regimes is "small" relative to sample size. Its practical relevance is that the asymptotic approximation implied by Theorem 1 is likely to hold better for cases where $\beta_2 - \beta_1$ is small than for cases where $\beta_2 - \beta_1$ is large. If the threshold effect is large, however, the threshold will be quite precisely estimated.

Since the asymptotic distribution in Theorem 1 is pivotal, it may be used to form valid asymptotic confidence intervals. Furthermore, the distribution functional (14) has the inverse

$$c(\alpha) = -2 \log\left(1 - \sqrt{1 - \alpha}\right), \tag{15}$$

from which it is easy to calculate critical values. For example, the 10% critical value is 6.53, the 5% is 7.35 and the 1% is 10.59. A test of $H_0 : \gamma = \gamma_0$ rejects at the asymptotic level α if $LR_1(\gamma_0)$ exceeds $c(\alpha)$.

To form an asymptotic confidence interval for γ, the "no-rejection region" of confidence level $1 - \alpha$ is the set of values of γ such that $LR_1(\gamma) \le c(\alpha)$, where $LR_1(\gamma)$ is defined in (13) and $c(\alpha)$ is defined in (15). This is easiest to find by plotting $LR_1(\gamma)$ against γ and drawing a flat line at $c(\alpha)$.

One of the convenient features of this confidence region is that it is a natural by-product of model estimation. In order to find the LS estimate $\hat{\gamma}$, the sequence of sum of squared errors $S_1(\gamma)$ were calculated. The likelihood ratio sequence $LR_1(\gamma)$ is a simple re-normalization of these numbers, and require no further computation.

4.3   Asymptotic Distribution of Slope Coefficients

The estimator $\hat{\beta} = \hat{\beta}(\hat{\gamma})$ depends on the threshold estimate $\hat{\gamma}$, which appears to complicate inference on β. Chan (1993) and Hansen (1997) show that the dependence on the threshold estimate is not of first-order asymptotic importance, so inference on β can proceed as if the threshold estimate $\hat{\gamma}$ were the true value. Hence $\hat{\beta}$ is asymptotically normal with a covariance matrix which can be estimated by

$$\hat{V} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^*(\hat{\gamma}) x_{it}^*(\hat{\gamma})'\right)^{-1} \hat{\sigma}^2 .$$

8

While we need the assumption that the errors are iid for the purposes of constructing confidence intervals for γ, it would seem appropriate to relax this assumption when constructing confidence intervals for the slope coefficients. If the errors are allowed to be conditionally heteroskedastic, the natural covariance matrix estimator for $\hat{\beta}$ is

$$\hat{V}_h = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^*(\hat{\gamma}) x_{it}^*(\hat{\gamma})' \right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^*(\hat{\gamma}) x_{it}^*(\hat{\gamma})' \left(\hat{e}_{it}^*\right)^2 \right) \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}^*(\hat{\gamma}) x_{it}^*(\hat{\gamma})' \right)^{-1}.$$

5 Multiple Thresholds

Model (1) has a single threshold. In some applications there may be multiple thresholds. For example, the double threshold model takes the form

$$y_{it} = \mu_i + \beta_1' x_{it} I\left(q_{it} \leq \gamma_1\right) + \beta_2' x_{it} I\left(\gamma_1 < q_{it} \leq \gamma_2\right) + \beta_3' x_{it} I\left(\gamma_2 < q_{it}\right) + e_{it}. \qquad (16)$$

where the thresholds are ordered so that $\gamma_1 < \gamma_2$. We will focus on this double-threshold model since the methods extend in a straightforward manner to higher-order threshold models. We discuss three relevant statistical issues: (1) Estimation; (2) Testing for the presence of a double threshold; (3) Construction of confidence intervals for the threshold parameters γ_1 and γ_2.

5.1 Estimation

For given (γ_1, γ_2), (16) is linear in the slopes $(\beta_1, \beta_2, \beta_3)$ so OLS estimation is appropriate. Thus for given (γ_1, γ_2) the concentrated sum of squared errors $S(\gamma_1, \gamma_2)$ is straightforward to calculate (as in the single threshold model). The joint LS estimates of (γ_1, γ_2) are by definition the values which jointly minimize $S(\gamma_1, \gamma_2)$. While these estimates might seem desirable, they may be quite cumbersome to implement in practice. A grid search over (γ_1, γ_2) requires approximately $N^2 = (nT)^2$ regressions which may be prohibitively expensive.

A remarkable insight allows us to escape this computational burden. It has been found (Chong (1994), Bai (1997), Bai-Perron (1998)) in the multiple changepoint model that sequential estimation is consistent. The same logic appears to apply to the multiple threshold model. The method works as follows. In the first stage, let $S_1(\gamma)$ be the single threshold sum of squared errors as defined in (7) and let $\hat{\gamma}_1$ be the threshold estimate which minimizes $S_1(\gamma)$. The analysis of Chong and Bai suggests that $\hat{\gamma}_1$ will be consistent[3] for either γ_1 or γ_2 (depending on which effect is "stronger").

---

[3]The reason why $\hat{\gamma}_1$ is consistent is because the single-threshold sum of squared errors function $S_1(\gamma)$ asymptotically converges to a limit function which has two local minima at γ_1 and γ_2.

Fixing the first-stage estimate $\hat{\gamma}_1$, the second-stage criterion is

$$S_2^r(\gamma_2) = \begin{cases} S(\hat{\gamma}_1, \gamma_2) & \text{if } \hat{\gamma}_1 < \gamma_2 \\ S(\gamma_2, \hat{\gamma}_1) & \text{if } \gamma_2 < \hat{\gamma}_1 \end{cases} \tag{17}$$

and the second-stage threshold estimate is

$$\hat{\gamma}_2^r = \underset{\gamma_2}{\text{argmin}} \; S_2^r(\gamma_2). \tag{18}$$

Since it is undesirable to have a small number of observations in any given "regime", we can restrict the search in (18) so that a minimum number of observations fall in each of the three regimes.

Bai (1997) has shown that $\hat{\gamma}_2^r$ is asymptotically efficient, but $\hat{\gamma}_1$ is not. This is because the estimate $\hat{\gamma}_1$ was obtained from a sum of squared errors function which was contaminated by the presence of a neglected regime. The asymptotic efficiency of $\hat{\gamma}_2^r$ suggests that $\hat{\gamma}_1$ can be improved by a third-stage estimation. Bai (1997) suggests the following *refinement* estimator. Fixing the second-stage estimate $\hat{\gamma}_2^r$, define the refinement criterion

$$S_1^r(\gamma_1) = \begin{cases} S(\gamma_1, \hat{\gamma}_2^r) & \text{if } \gamma_1 < \hat{\gamma}_2^r \\ S(\hat{\gamma}_2^r, \gamma_1) & \text{if } \hat{\gamma}_2^r < \gamma_1 \end{cases} \tag{19}$$

and the refinement estimate

$$\hat{\gamma}_1^r = \underset{\gamma_1}{\text{argmin}} \; S_1^r(\gamma_1). \tag{20}$$

Bai (1997) shows that the refinement estimator $\hat{\gamma}_1^r$ is asymptotically efficient in changepoint estimation, and we expect similar results to hold in threshold regression.

5.2 Determining Number of Thresholds

In the context of model (16), there are either no thresholds, one threshold, or two thresholds. In section 3.1 we introduced F_1 as a test of no thresholds against one threshold, and suggested a bootstrap to approximate the asymptotic p-value. If F_1 rejects the null of no threshold, in the context of model (16) we need a further test to discriminate between one and two thresholds.

The minimizing sum of squared errors from the second-stage threshold estimate is $S_2^r(\hat{\gamma}_2^r)$ with variance estimate $\hat{\sigma}^2 = S_2^r(\hat{\gamma}_2^r)/n(T-1)$. Thus an approximate likelihood ratio test of one versus two thresholds can be based on the statistic

$$F_2 = \frac{S_1(\hat{\gamma}_1) - S_2^r(\hat{\gamma}_2^r)}{\hat{\sigma}^2}.$$

The hypothesis of one threshold is rejected in favor of two thresholds if F_2 is large.

Since the null asymptotic distribution of the likelihood ratio test is non-pivotal[4] we suggest using a bootstrap procedure to approximate the sampling distribution. To generate the bootstrap samples, hold the regressors x_{it} and threshold variable q_{it} fixed in repeated bootstrap samples. The bootstrap errors will be drawn from the residuals calculated under the alternative hypothesis, so should be the residuals from LS estimation of model (16). Group the regression residuals $\hat{e}_{it}^*$ by individual: $\hat{e}_i^* = (\hat{e}_{i1}^*, \hat{e}_{i2}^*, ..., \hat{e}_{iT}^*)$, and treat the sample $\{\hat{e}_1^*, \hat{e}_2^*, ..., \hat{e}_n^*\}$ as an empirical distribution. Draw (with replacement) error samples from the empirical distribution. Let $e_i^\#$ denote a generic $T \times 1$ draw. The dependent variable y_{it} should be generated under the null hypothesis of a single threshold (1), so use the equation

$$y_{it}^\# = \hat{\beta}_1' x_{it} I\left(q_{it} \leq \hat{\gamma}\right) + \hat{\beta}_2' x_{it} I\left(q_{it} > \hat{\gamma}\right) + e_{it}^\#, \tag{21}$$

which depends on the parameter values $\hat{\beta}_1, \hat{\beta}_2$, and $\hat{\gamma}$, the least-squares estimates from the single threshold model. From the bootstrap sample, the test statistic F_2 may be calculated, and this procedure repeated multiple times to calculate the bootstrap p-value.

From equation (21) it is clear that unlike the sampling distribution of LR_1, which did not depend on the regression parameters, the sampling distribution of F_2 depends on the regression parameters $\beta_1 - \beta_2$ and γ. (The distribution only depends on the difference $\beta_2 - \beta_1$ since the model can be reparameterized to include the linear regressors x_{it} under both the null and alternative.) Thus the sampling distribution of F_2 is not pivotal, and the bootstrap is not guaranteed to produce an excellent approximation.

5.3   Confidence Region Construction

We finally consider the construction of confidence intervals for the two threshold parameters $\{\gamma_1, \gamma_2\}$. Bai (1997) showed (for the analogous case of change-point models) that the refinement estimators of section 5.1 have the same asymptotic distributions as the threshold estimate in a single threshold model. This suggests that we can construct confidence intervals in the same way as in section 4.2.

Let

$$LR_2^r(\gamma) = \frac{S_2^r(\gamma) - S_2^r(\hat{\gamma}_2^r)}{\hat{\sigma}^2}$$

---

[4]It is important to remember that this differs from the changepoint case (see Chong (1994), Bai (1997) and Bai-Perron (1998)), where the asymptotic distribution of F_2 is known and pivotal.

and

$$LR_1^r(\gamma) = \frac{S_1^r(\gamma) - S_1^r(\hat{\gamma}_1^r)}{\hat{\sigma}^2},$$

where $S_2^r(\gamma)$ and $S_1^r(\gamma)$ are defined in (17) and (19), respectively. Our asymptotic $(1-\alpha)\%$ confidence intervals for γ_2 and γ_1 are the set of values of γ such that $LR_2^r(\gamma) \leq c(\alpha)$ and $LR_1^r(\gamma) \leq c(\alpha)$, respectively.

6   Investment and Financing Constraints

Classical models of the firm assume the existence of perfect financial markets on which firms can borrow the needed resources for investment projects. Alternative models of financing place restrictions on the extent of external financing. An important empirical question is whether or not there exist firms which behave as though they are subject to such constraints.

A well-cited paper which explored the empirical implications of financing constraints is Fazzari, Hubbard and Petersen (1988), henceforth FHP. These authors argue that the presence of financing constraints implies that a firm's cash flow will be positively related to its investment rate only when the firm faces constraints on external financing. If a firm is free to borrow on external financial markets, cash flow will be irrelevant for investment. This distinction motivated FHP to test for financing constraints by estimating separate investment regressions for "constrained" and "unconstrained" firms to see if there are differing effects of contemporaneous cash flow. To distinguish constrained and unconstrained firms, they used the dividend to income ratio, as their theory suggests that a financially constrained firm will choose to retain earnings rather than pay dividends. Hence the firms which have low levels of dividend payments are the financially constrained firms.

FHP divide their sample into three classes, depending on whether the dividend to income ratio was less than 0.1 for 10 years in the sample, between 0.1 and 0.2 for over 10 years, and all other firms. Thus they are estimating a double-threshold regression on panel data, where q_{it} is the largest dividend-income ratio over the 10-year period, and the thresholds γ are set at 0.1 and 0.2.

There are two obvious problems with the FHP regression. First, it treats the dividend-income ratio as exogenous, while their theory explicitly treats dividend payments as decision variables. The use of an endogenous threshold variable may bias their results. Second, they select their threshold levels arbitrarily, rather than estimating these parameters from the sample. In this section we explore whether our methods allow for a re-appraisal of FHP's analysis.

Table 1: Summary Statistics

	Minimum	25% Quantile	Median	75% Quantile	Maximum
I_{it}	0.001	0.049	0.076	0.113	1.66
Q_{it-1}	0.021	0.371	0.675	1.31	111.8
CF_{it-1}	-0.94	0.12	0.22	0.32	8.71
D_{it-1}	0.000	0.089	0.206	0.320	4.67

The original data used by FHP is no longer available. We use a similar dataset, extracted from the dataset used by Hall and Hall (1993), which is an unbalanced panel of U.S. firms originally taken from Compustat. Our methods are designed for balanced panels, so we took the subset of 565 firms which are observed for the years 1973-1987.

The threshold variable should be an exogenous indicator of a firm's access to external financing. A natural candidate is the existing debt level. It seems reasonable to believe that banks will be reluctant to lend money to debt-heavy firms. This choice is similar to that of Hu and Schiantarelli (1998) who estimate a switching regression using the debt-asset ratio as one variable in their switching equation.

To fix notation, let I_{it} be the ratio of investment to capital; Q_{it} be the ratio of total market value to assets; CF_{it} be the ratio of cash flow to assets; and D_{it} be the ratio of long term debt to assets, where stock variables are defined at the end of year. Summary statistics of the four variables are given in Table 1.

We use the multiple threshold regression model

$$
\begin{aligned}
I_{it} \quad = \quad & \mu_i + \theta_1 Q_{it-1} + \theta_2 Q_{it-1}^2 + \theta_3 Q_{it-1}^3 + \theta_4 D_{it-1} \\
& + \theta_5 Q_{it-1} D_{it-1} + \beta_1 CF_{it-1} I\left(D_{it-1} \leq \gamma_1\right) \\
& + \beta_2 CF_{it-1} I\left(\gamma_1 < D_{it-1} \leq \gamma_2\right) + \beta_3 CF_{it-1} I\left(\gamma_2 < D_{it-1}\right) + e_{it},
\end{aligned}
\tag{22}
$$

where (22) represents a double threshold model for illustration. Model (22) falls in the class of models (1) setting $q_{it} = D_{it-1}$ and $x_{it} = CF_{it-1}$. There are also the additional regressors $(Q_{it-1}, Q_{it-1}^2, Q_{it-1}^3, D_{it-1}, Q_{it-1} D_{it-1})$. The latter can be viewed as a special case of (1) by constraining the slope coefficients on these variables to be the same in the two regimes, which has no effect on the distribution theory. The reason model (22) has only the slope coefficient on cash flow switch between regimes is to focus attention on this key variable of interest. The non-linear terms in the regression (namely, Q_{it-1}^2, Q_{it-1}^3, and $Q_{it-1} D_{it-1}$) were included to reduce the possibility of spurious correlations due to omitted variables bias. The

Table 2: Tests for Threshold Effects

Test for Single Threshold	
F_1	32.6
P-Value	.003
(10%, 5%, 1% critical values)	(12.4, 14.8, 26.2)
Test for Double Threshold	
F_2	25.8
P-Value	.017
(10%, 5%, 1% critical values)	(12.3, 14.9, 42.9)
Test for Triple Threshold	
F_3	4.2
P-Value	.723
(10%, 5%, 1% critical values)	(10.9, 13.3, 22.9)

Table 3: Threshold Estimates

	Estimate	95% Confidence Interval	99% Confidence Interval
$\hat{\gamma}_1^r$	.0157	[.0139, .0181]	[.0120, .0239]
$\hat{\gamma}_2^r$	.5362	[.5305, .5629]	[.5190, .5693]

choice of the particular non-linear terms was data-based, as the variables D_{it-1}^2 and D_{it-1}^3 were insignificant and omitted to reduce computation costs.

To determine the number of thresholds, model (22) was estimated by least squares, allowing for (sequentially) zero, one, two, and three thresholds. The test statistics F_1, F_2 and F_3, along with their bootstrap[5] p-values, are shown in Table 2. We find that the test for a single threshold F_1 is highly significant with a bootstrap p-value of 0.003, and the test for a double threshold F_2 is also strongly significant, with a bootstrap p-value of 0.017. On the other hand, the test for a third threshold F_3 is not close to being statistically significant, with a bootstrap p-value of .723. We conclude that there is strong evidence that there are two thresholds in the regression relationship. For the remainder of the analysis we work with this double threshold model.

The point estimates of the two thresholds and their asymptotic 95% and 99% confidence intervals are reported in Table 3. The estimates are .016 and .536, which are very small (and

---

[5]300 bootstrap replications were used for each of the three bootstrap tests.

Table 4: Percentage of Firms in Each Regime By Year

Firm Class	Year													
	1974	'75	'76	'77	'78	'79	'80	'81	'82	'83	'84	'85	'86	'87
$D_{it-1} \leq .0157$	16	14	14	15	15	13	13	11	10	10	10	10	10	11
$.0157 < D_{it-1} \leq .5362$	78	79	78	81	81	84	82	85	86	85	84	82	77	73
$.5362 < D_{it-1}$	6	7	8	5	4	4	5	4	4	5	6	8	13	16

very large) values in the empirical distribution of the debt/assets threshold variable. Thus the three classes of firms indicated by the point estimates are those with "very low debt," "very high debt" and "other". The asymptotic confidence intervals for the threshold are very tight, indicating little uncertainty about the nature of this division. More information can be learned about the threshold estimates from plots of the concentrated likelihood ratio function $LR_1(\gamma)$, $LR_2^r(\gamma)$ and $LR_1^r(\gamma)$ in Figures 1, 2, and 3 (corresponding to the first-stage estimate $\hat{\gamma}_1$ and the refinement estimators $\hat{\gamma}_2^r$ and $\hat{\gamma}_1^r$). The point estimates are the value of γ at which the likelihood ratio hits the zero axis, which is in the far left part of the graph. The 95% confidence intervals for γ_2 and γ_1 can be found from $LR_2^r(\gamma)$ and $LR_1^r(\gamma)$ by the values of γ for which the likelihood ratio lies beneath the dotted line.

It is interesting to examine the unrefined first-step likelihood ratio function $LR_1(\gamma)$, which is computed when estimating a single threshold model. The first-step threshold estimate is the point where the $LR_1(\gamma)$ equals zero, which occurs at $\hat{\gamma}_1 = .0157$. There is a second major dip in the likelihood ratio around the second-step estimate .53. Thus the single threshold likelihood conveys information that suggests that there is a second threshold in the regression.

Table 4 reports the percentage of firms which fall into the three regimes each year. We see that the percentage of firms in the "very low debt" category ranges from 10% to 16% of the sample over the years. The "very high debt" firms range from 4% to 16% of the sample in a given year. It is interesting to note that the last two years of the sample (1986 and 1987) saw a large increase in the number of firms with very high debt ratios.

The regression slope estimates, conventional OLS standard errors, and White-corrected standard errors are displayed in Table 5. We see that Q_{it-1} and its powers are statistically significant, indicating a positive (and very slightly non-linear) relationship between q and investment. The debt level D_{it-1} has a negative and significant effect on investment, and there is no apparent interaction effect between q and the debt level.

The coefficients of primary interest are those on cash flow. The point estimates suggest that investment is positively related to cash flow, with "very low debt" firms having a lower

Table 5: Regression Estimates: Double Threshold Model

Regression Estimates			
Regressor	Coefficient Estimate	OLS S.E.	White S.E.
Q_{it-1}	.010	.001	.002
$Q_{it-1}^2/10^3$	-.198	.026	.064
$Q_{it-1}^3/10^6$	1.047	.199	.448
D_{it-1}	-.016	.005	.009
$Q_{it-1}D_{it-1}$	.001	.001	.002
$CF_{it-1}I\left(D_{it-1} \leq .0157\right)$	.063	.006	.014
$CF_{it-1}I\left(.0157 < D_{it-1} \leq .5362\right)$	.098	.006	.010
$CF_{it-1}I\left(.5362 < D_{it-1}\right)$	.039	.012	.031

coefficient (about one-third smaller in magnitude) than the typical firm. What is quite unexpected is that the firms with the highest debt levels have the smallest coefficient of 0.039. The White standard error on this last coefficient, however, is quite high, indicating that there is still considerable uncertainty in the estimate.

The conventional OLS standard errors and the White-corrected standard errors are considerably different, with the White-corrected ones roughly twice as big. This is evidence in favor of heteroskedasticity, which violates one of the maintained assumptions of our asymptotic analysis. Based on the theory (Hansen, 1997) for least-squares threshold regression (the model without fixed effects), we would expect the threshold estimates to be consistent and the distribution theory of Theorem 1 to be correct up to a scale effect, so that asymptotic confidence intervals would still take the form given in Table 3, but would require a different critical value.

7 Conclusion

This paper has developed new empirical methods for panel data. We have defined a threshold regression model with individual-specific effects, and shown that the model is rather straightforward to estimate using a fixed-effects transformation. The asymptotic theory is non-standard, but confidence intervals for the threshold can be constructed by inverting the likelihood ratio statistic, and this construction is a natural by-product of the estimation method.

The methods are applied to the investment decisions of a panel of 565 U.S. firms for the period 1973-1987. We find overwhelming evidence of a double threshold effect which separates the firms based on their debt to asset ratio. The estimates are somewhat consistent with the theory of financing constraints. The notable difference between our work and that of Fazzari, Hubbard and Petersen (1988) is that we are also able to quantify the extent of financing constraints in the economy rather than assuming the degree of such constraints.

Several extensions of our methods would be desirable, including allowing for heteroskedasticity, lagged dependent variables, endogenous variables, and random effects. These would be useful subjects for future research.

References

[1] Abel, A. B. and J.C. Eberly (1994): "A unified model of investment under uncertainty," *American Economic Review*, 84, 1369-1384.

[2] Abel, A. B. and J.C. Eberly (1996): "Investment and q with fixed costs: An empirical analysis," working paper, University of Pennsylvania.

[3] Andrews, D.W.K. and W. Ploberger (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica,* 62, 1383-1414.

[4] Bai, J. (1997): "Estimating multiple breaks one at a time," *Econometric Theory.* 13, 315-352.

[5] Bai, J. and P. Perron (1998): "Estimating and testing linear models with multiple structural changes," Econometrica, 66, 47-78.

[6] Barnett, S.A. and P. Sakellaris (1995): "Non-linear response of firm investment to q: Testing a model of convex and non-convex adjustment costs.

[7] Beran, R. (1987): "Prepivoting to reduce the level error of confidence sets," *Biometrika*, 74, 457-468.

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

[9] Chan, K.S. and R.S. Tsay (1996): "Limiting properties of the least squares estimator of a continuous threshold autoregressive model," working paper, University of Iowa.

[10] Chong, T. T-L. (1994): "Consistency of change-point estimators when the number of change-points in structural change models is underspecified," working paper, Chinese University of Hong Kong.

[11] Davies, R.B. (1977): "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 64, 247-254.

[12] Davies, R.B. (1987): "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 74, 33-43.

[13] Fazzari, S.M., R. Glenn Hubbard and B.C. Petersen (1988): "Financing constraints and corporate investment," *Brookings Papers on Economic Activity*, 141-195.

[14] Hall, B.H. and R.E. Hall (1993): "The value and performance of U.S. corporations," *Brookings Papers on Economic Activity*, 1-34.

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

[16] Hansen, B.E. (1997): "Sample splitting and threshold estimation," working paper, Boston College.

[17] Hu, X. and F. Schiantarelli (1998): "Investment and financing constraints: A switching regression approach for U.S. panel data," *Review of Econometrics and Statistics*, forthcoming.

8   Appendix: Mathematical Proofs

We need the following technical assumptions. Let γ_0 denote the true value of γ. Let $\theta = \beta_2 - \beta_1$ and $C = n^\alpha \theta$, where $\alpha \in (0, 1/2)$. Let $f_t(\gamma)$ denote the density function of q_{it}, set $z_{it} = C' x_{it}$,

$$D(\gamma) = \sum_{t=1}^{T} E\left(z_{it}^2 \mid q_{it} = \gamma\right) f_t(\gamma),$$

and $D = D(\gamma_0)$. Let $f_{k|t}(\gamma_1 \mid \gamma_2)$ denote the conditional density of q_{ik} given q_{it}.

Assumptions

1. For each t, (q_{it}, x_{it}, e_{it}) are independent and identically distributed (iid) across i;

2. For each i, e_{it} is iid over t, is independent of $\left\{(x_{ij}, q_{ij})_{j=1}^{T}\right\}$, and $E(e_{it}) = 0$;

3. For each $j = 1, ..., k$, $P\left(x_{i1}^j = x_{i2}^j = \cdots = x_{iT}^j\right) < 1$, where x_{it}^j is the j'th element of x_{it}.

4. For some $s > 2$, $E\left|x_{it}\right|^s < \infty$ and $E\left|e_{it}\right|^s < \infty$;

5. For some fixed $C < \infty$ and $0 < \alpha < 1/2$, $\theta = n^{-\alpha} C$;

6. $D(\gamma)$ is continuous at $\gamma = \gamma_0$;

7. $0 < D < \infty$;

8. For $k > t$, $f_{k|t}(\gamma_0 \mid \gamma_0) < \infty$.

Assumptions 1, 2, 3, and 4 are standard for fixed effect panel models with strictly exogenous regressors. Assumption 5 is more unusual, setting $\theta = n^{-\alpha} C \rightarrow 0$ as $n \rightarrow \infty$. The renormalization is to force θ to be "small," reducing the information in the sample concerning the threshold and hence slowing down the rate of convergence of the threshold estimate. This assumption need not be viewed as very restrictive since the rate at which θ decreases to zero can be set quite low. It does suggest, however, that the asymptotic approximation is more likely to provide good approximations when θ is small relative to the case where θ is large.

Assumption 6 excludes threshold effects which occur simultaneously in the marginal distribution of the regressors and in the regression function. Assumption 7 excludes continuous threshold models (see Chan and Tsay (1997)), and requires that the threshold variable q_{it} be continuously distributed with positive support at the threshold γ_0. Assumption 8 excludes the possibility that $q_{it} = \gamma_0$ for $t = 1, ..., T$.

The proof of the Theorem is based on the following Lemma. Let " $\Rightarrow$ " denote weak convergence with respect to the uniform metric, and let $\lambda_n = n^{1-2\alpha}$.

Lemma 1 *As $n \to \infty$, uniformly over $v \in [-\overline{v}, \overline{v}]$,*

$$S_1(\gamma_0) - S_1(\gamma_0 + v/\lambda_n) \Rightarrow q(v),$$

where
$$q(v) = -D_T |v| + 2\sqrt{\sigma^2 D_T} W(v)$$

$D_T = D\left(1 - \frac{1}{T}\right)$, *and $W(v)$ is a double-sided standard Brownian motion on $(-\infty, \infty)$.*

Proof: Let
$$\nabla z_{it}(\gamma) = z_{it} I\left(q_{it} \leq \gamma\right) - z_{it} I\left(q_{it} \leq \gamma_0\right).$$

The regression equation (1) holds when $\gamma = \gamma_0$, the true value. For values of $\gamma \neq \gamma_0$, note that (1) can be re-written as

$$
\begin{aligned}
y_{it} &= \mu_i + \beta_1' x_{it} I\left(q_{it} \leq \gamma_0\right) + \beta_2' x_{it} I\left(q_{it} > \gamma_0\right) + e_{it} \\
&= \mu_i + \beta_1' x_{it} I\left(q_{it} \leq \gamma\right) + \beta_2' x_{it} I\left(q_{it} > \gamma\right) \\
&\quad - \beta_1' x_{it} \left[I\left(q_{it} \leq \gamma\right) - I\left(q_{it} \leq \gamma_0\right)\right] - \beta_2' x_{it} \left[I\left(q_{it} > \gamma\right) - I\left(q_{it} > \gamma_0\right)\right] + e_{it} \\
&= \mu_i + \beta' x_{it}(\gamma) + (\beta_2 - \beta_1)' x_{it} \left[I\left(q_{it} \leq \gamma\right) - I\left(q_{it} \leq \gamma_0\right)\right] + e_{it} \\
&= \mu_i + \beta' x_{it}(\gamma) + n^{-\alpha} \nabla z_{it}(\gamma) + e_{it}
\end{aligned}
\tag{23}
$$

(23) makes explicit the regression error for $\gamma \neq \gamma_0$.

The fixed effect transformation is linear, so can be applied to (23) to yield

$$y_i^* = \beta' x_{it}^*(\gamma) + n^{-\alpha} \nabla z_{it}^*(\gamma) + e_{it}^*$$

which is the correct representation of (4) for $\gamma \neq \gamma_0$. Hansen (1997) shows that the asymptotic distribution of $\hat{\gamma}$ is not affected by the estimation of β, and this holds in our environment as well. We can thus simplify matters by assuming that β is known and only γ is estimated, so that the regression residual (for fixed γ) is

$$\hat{e}_{it}(\gamma) = n^{-\alpha} \nabla z_{it}^*(\gamma) + e_{it}^*. \tag{24}$$

Using (24),

$$
\begin{aligned}
S(\gamma_0) - S(\gamma) &= \sum_{i=1}^{n}\sum_{t=1}^{T}\hat{e}_{it}(\gamma_0)^2 - \sum_{i=1}^{n}\sum_{t=1}^{T}\hat{e}_{it}(\gamma)^2 \\
&= \sum_{i=1}^{n}\sum_{t=1}^{T}e_{it}^{*2} - \sum_{i=1}^{n}\sum_{t=1}^{T}\left(n^{-\alpha}\nabla z_{it}^{*}(\gamma) + e_{it}^{*}\right)^2 \\
&= -n^{-2\alpha}\sum_{i=1}^{n}\sum_{t=1}^{T}\nabla z_{it}^{*}(\gamma)^2 - 2n^{-\alpha}\sum_{i=1}^{n}\sum_{t=1}^{T}\nabla z_{it}^{*}(\gamma)e_{it}^{*}. \qquad (25)
\end{aligned}
$$

We now show that as $n \to \infty$, uniformly over $v \in [-\overline{v}, \overline{v}]$,

$$
n^{-2\alpha}\sum_{i=1}^{n}\sum_{t=1}^{T}\nabla z_{it}^{*}(\gamma_0 + v/\lambda_n)^2 \Rightarrow D_T\,|v|. \qquad (26)
$$

We prove (26) for the case $v \in [0, \overline{v}]$. We will show that for $\gamma = \gamma_0 + v/\lambda_n$,

$$
E\left(n^{-2\alpha}\sum_{i=1}^{n}\sum_{t=1}^{T}\nabla z_{it}^{*}(\gamma)^2\right) = \lambda_n \sum_{t=1}^{T} E\left(\nabla z_{it}^{*}(\gamma)\right)^2 \to D_T\,|v|. \qquad (27)
$$

Arguments similar to those in the proof of Lemma A.2 of Hansen (1997) show that (27) implies (26) under the assumptions. Expansion of the quadratic yields

$$
\sum_{t=1}^{T} E\left(\nabla z_{it}^{*}(\gamma)\right)^2 = \sum_{t=1}^{T} E\left(\nabla z_{it}(\gamma)\right)^2 - \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{T} E\left(\nabla z_{it}(\gamma)\nabla z_{ik}(\gamma)\right). \qquad (28)
$$

Consider the first sum on the right-hand-side of (28). Observe that since $\gamma = \gamma_0 + v/\lambda_n \to \gamma_0$,

$$
\begin{aligned}
\lambda_n P\left(\gamma_0 < q_{it} \leq \gamma\right) &= v\frac{P\left(q_{it} \leq \gamma\right) - P\left(q_{it} \leq \gamma_0\right)}{\gamma - \gamma_0} \\
&\to v f_t(\gamma_0) \qquad (29)
\end{aligned}
$$

as $n \to \infty$. Thus

$$
\begin{aligned}
\lambda_n \sum_{t=1}^{T} E\left(\nabla z_{it}(\gamma)\right)^2 &= \lambda_n \sum_{t=1}^{T} E\left(z_{it}^2 I\left(\gamma_0 < q_{it} \leq \gamma\right)\right) \\
&= \sum_{t=1}^{T} E\left(z_{it}^2 \mid \gamma_0 < q_{it} \leq \gamma\right)\lambda_n P\left(\gamma_0 < q_{it} \leq \gamma\right) \\
&\to \sum_{t=1}^{T} vE\left(z_{it}^2 \mid q_{it} = \gamma_0\right) f_t(\gamma_0) = vD. \qquad (30)
\end{aligned}
$$

21

Next consider the double-sum on the right-hand-side of (28). By Assumption 8, for $k > t$,

$$
\begin{aligned}
P\left(\gamma_0 < q_{ik} \leq \gamma \mid \gamma_0 < q_{it} \leq \gamma\right) &= v\lambda_n^{-1} \frac{\left[P\left(q_{ik} \leq \gamma \mid \gamma_0 < q_{it} \leq \gamma\right) - P\left(q_{ik} \leq \gamma_0 \mid \gamma_0 < q_{it} \leq \gamma\right)\right]}{\gamma - \gamma_0} \\
&= v\lambda_n^{-1}\left(f_{k|t}(\gamma_0 \mid \gamma_0) + o(1)\right) \\
&\rightarrow 0,
\end{aligned}
$$

and combined with (29), for $k > t$,

$$
\lambda_n P\left(\gamma_0 < q_{it} \leq \gamma, \gamma_0 < q_{ik} \leq \gamma\right) = \lambda_n P\left(\gamma_0 < q_{ik} \leq \gamma\right) P\left(\gamma_0 < q_{it} \leq \gamma \mid \gamma_0 < q_{ik} \leq \gamma\right) \rightarrow 0.
\tag{31}
$$

(31) also holds for $k < t$ by symmetry.

Hence

$$
\begin{aligned}
\lambda_n \frac{1}{T} \sum_{t=1}^{T} \sum_{k=1}^{T} E\left(\nabla z_{it}(\gamma)\nabla z_{ik}(\gamma)\right) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{T}\lambda_n E\left(z_{it}z_{ik}I\left(\gamma_0 < q_{it} \leq \gamma\right)I\left(\gamma_0 < q_{ik} \leq \gamma\right)\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{T} E\left(z_{it}^2 \mid \gamma_0 < q_{it} \leq \gamma, \gamma_0 < q_{ik} \leq \gamma\right) \\
&\quad \cdot \lambda_n P\left(\gamma_0 < q_{it} \leq \gamma, \gamma_0 < q_{ik} \leq \gamma\right) \\
&= \frac{1}{T}\sum_{t=1}^{T} E\left(z_{it}^2 \mid \gamma_0 < q_{it} \leq \gamma\right)\lambda_n P\left(\gamma_0 < q_{it} \leq \gamma\right) + o(1) \\
&\rightarrow \frac{1}{T}\sum_{t=1}^{T} E\left(z_{it}^2 \mid q_{it} = \gamma_0\right) v f_t(\gamma_0) = \frac{1}{T}v D.
\end{aligned}
\tag{32}
$$

by (31). (28), (30) and (32) imply (27) and hence (26).

Next, we wish to show that uniformly over $v \in [-\overline{v}, \overline{v}]$,

$$
n^{-\alpha}\sum_{i=1}^{n}\sum_{t=1}^{T}\nabla z_{it}^*(\gamma_0 + v/\lambda_n)e_{it}^* \Rightarrow \sqrt{\sigma^2 D_T}W(v).
\tag{33}
$$

By the properties of least squares projection, $\sum_{t=1}^{T}\nabla z_{it}^*(\gamma)e_{it}^* = \sum_{t=1}^{T}\nabla z_{it}^*(\gamma)e_{it}$, and since the e_{it} are iid,

$$
\begin{aligned}
E\left(n^{-\alpha}\sum_{i=1}^{n}\sum_{t=1}^{T}\nabla z_{it}^*(\gamma)e_{it}^*\right)^2 &= \lambda_n E\left(\sum_{t=1}^{T}\nabla z_{it}^*(\gamma)e_{it}\right)^2 \\
&= \lambda_n \sum_{t=1}^{T}E\left(\nabla z_{it}^*(\gamma)\right)^2\sigma^2 \\
&\rightarrow D_T|v|\sigma^2
\end{aligned}
\tag{34}
$$

22

by (27). This establishes that the finite dimensional distributions of the stochastic process are those of the stated double-sided Brownian motion. By arguments identical to those in the proof of Lemma A.3 of Hansen (1997), (34) and Assumption 1 are sufficient to establish (33). (26) and (33) combine with (25) to yield the stated result. ∎

Proof of Theorem 1. Since $\hat{\gamma}$ minimizes $S(\gamma)$,

$$
\begin{aligned}
LR_1(\gamma_0) &= \max_{\gamma} \left(\frac{S_1(\gamma_0) - S_1(\gamma)}{\hat{\sigma}^2} \right) \\
&= \max_{v} \left(\frac{S_1(\gamma_0) - S_1(\gamma_0 + v/\lambda_n)}{\hat{\sigma}^2} \right).
\end{aligned}
\tag{35}
$$

where the final equality makes the change-of-variables $\gamma = \gamma_0 + v/\lambda_n$,

Hansen (1997) shows that under Assumption 1,

$$
\hat{v} \equiv \lambda_n \left(\hat{\gamma} - \gamma_0 \right) = O_p(1).
\tag{36}
$$

We will not repeat the proof of (36) here. The stochastic boundedness of (36) shows that for any $\eta > 0$, there is some $\bar{v} < \infty$ such that

$$
P\left(|\hat{v}| \leq \bar{v} \right) \geq 1 - \eta.
\tag{37}
$$

Let

$$
\begin{aligned}
\widetilde{LR} &= \max_{|v| \leq \bar{v}} \left(\frac{S_1(\gamma_0) - S_1(\gamma_0 + v/\lambda_n)}{\hat{\sigma}^2} \right) \\
&\Rightarrow \sigma^{-2} \max_{|v| \leq \bar{v}} q(v)
\end{aligned}
$$

where the stated weak convergence follows from Lemma 1 and the continuous mapping theorem.

(37) shows that

$$
P\left(LR_1(\gamma_0) = \widetilde{LR} \right) \geq 1 - \eta.
$$

Since η is arbitrary we conclude that

$$
LR_1(\gamma_0) \Rightarrow \sigma^{-2} \max_{-\infty < v < \infty} q(v) = \xi,
$$

say. Hansen (1997, Proof of Theorem 2) shows that the distribution function of ξ is $P\left(\xi \leq x \right) = \left(1 - \exp(-x/2) \right)^2.$ ∎

Figure 1
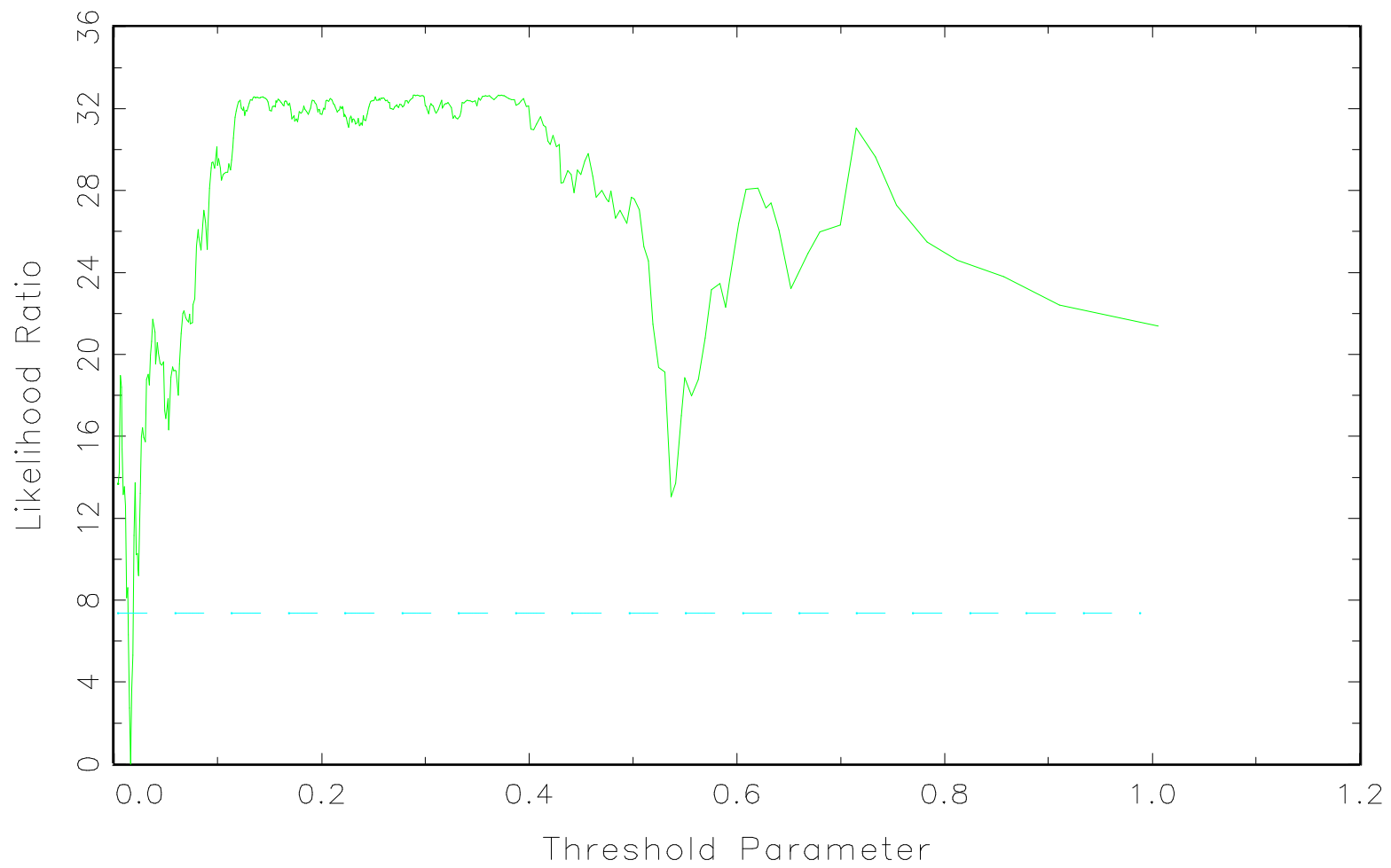Confidence Interval Construction in Single Threshold Model

Figure 2
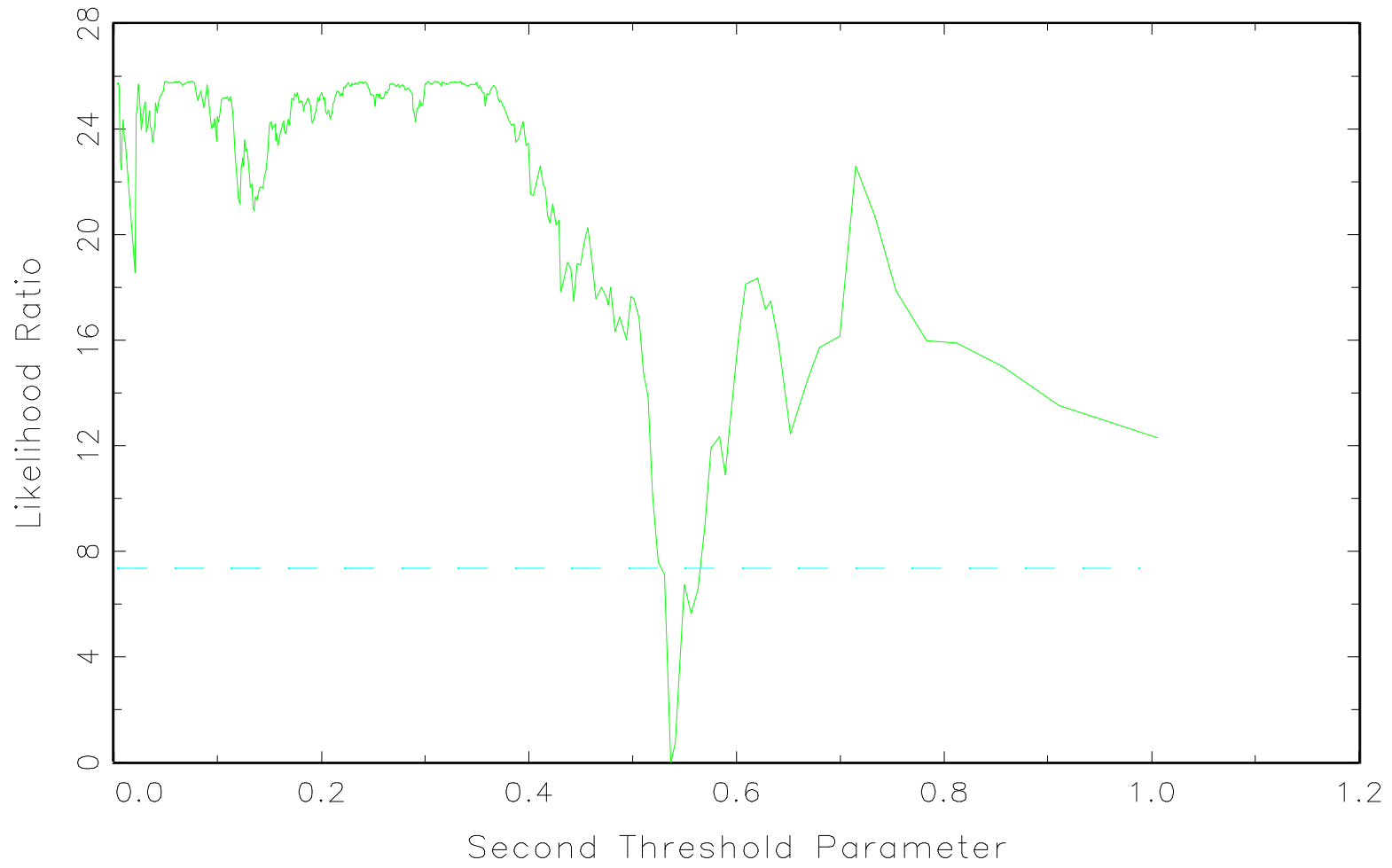Confidence Interval Construction in Double Threshold Model

Figure 3
Confidence Interval Construction in Double Threshold Model