Bootstraps for Dynamic Panel Threshold Models

Woosik Gong∗ Myung Hwan Seo†

September 12, 2024

Abstract

This paper develops valid bootstrap inference methods for the dynamic short panel threshold regression. We demonstrate that the standard nonparametric bootstrap is inconsistent for the first-differenced generalized method of moments (GMM) estimator. The inconsistency is due to an $n^{1/4}$-consistent non-normal asymptotic distribution for the threshold estimate when the parameter resides within the continuity region of the parameter space. It stems from the rank deficiency of the approximate Jacobian of the sample moment conditions on the continuity region. To address this, we propose a grid bootstrap to construct confidence intervals of the threshold, a residual bootstrap to construct confidence intervals of the coefficients, and a bootstrap for testing continuity. They are shown to be valid under uncertain continuity, while the grid bootstrap is additionally shown to be uniformly valid. A set of Monte Carlo experiments demonstrate that the proposed bootstraps perform well in the finite samples and improve upon the standard nonparametric bootstrap. An empirical application to firms’ investment model illustrates our methods.

KEYWORDS: Dynamic Panel Threshold; Kink; Bootstrap; Endogeneity; Identification; Rank Deficiency; Uniformity.

JEL: C12, C23, C24

1 Introduction

Threshold regression models have been widely used by empirical researchers, which have been more fruitful because of their extensions to the panel data context. Estimation and inference methods for the threshold model in non-dynamic panels were developed by Hansen (1999b) and Wang (2015). Dynamic panel threshold models were considered by Seo and Shin (2016), which proposes the generalized method of moments (GMM) estimation by generalizing the Arellano and Bond (1991) dynamic panel estimator. A latent group structure in the parameters of the panel threshold model was investigated by Miao et al. (2020b).

∗University of Wisconsin-Madison. Email: vgong28@wisc.edu. This research was supported by the BK21 FOUR (Fostering Outstanding Universities for Research) funded by the Ministry of Education(MOE, Korea) and National Research Foundation of Korea(NRF). The author is also grateful to the student travel grant award by the IAAE 2022 conference.

†Seoul National University. Email: myunghseo@snu.ac.kr. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A5A2A01033487)
Applications of the panel threshold models cover numerous topics in economics. The effect of debt on economic growth is a well-known example that has been analyzed using the panel threshold models, e.g. Adam and Bevan (2005), Cecchetti et al. (2011) and Chudik et al. (2017). Another example is the threshold effect of inflation on economic growth such as the works by Khan and Senhadji (2001), Rousseau and Wachtel (2002), Bick (2010), and Kremer et al. (2013). The benefit of foreign direct investment to productivity growth that depends on the regime determined by absorptive capacity is studied by Girma (2005) using firm level panel data.

It is common practice to make inference in the threshold regression models based on an assumption about whether the model is continuous or not. Continuous threshold models that have kinks at the tipping points have received active research attention, e.g. Hansen (2017) or Yang et al. (2020). In the literature, the kink threshold models are analyzed for the estimators that impose the continuity restriction as in Chan and Tsay (1998), Hansen (2017), or Zhang et al. (2017). On the other hand, unrestricted estimators are commonly used for the discontinuous threshold models as in Hansen (2000). However, Hidalgo et al. (2019) showed that the unrestricted least squares estimator possesses a different asymptotic property in the absence of the discontinuity. Specifically, while the unrestricted model is not misspecified under continuity, failing to impose the restriction results in incorrect inference without a proper care.

In the empirical research literature, there has been mixed use of the kink/discontinuous threshold models without much consideration on a possible specification error. Among the empirical examples referred previously, Khan and Senhadji (2001) use the continuous threshold model and impose the continuity on their estimation. They claim that the continuous model is desirable to prevent small changes in inflation rate yields different impact around the threshold level. On the other hand, Bick (2010) claims that the discontinuous threshold model is more appropriate for the same research question since overlooking regime-dependent intercept can result in omitted variable bias. However, both of them do not provide an econometric evidence that supports their choice of models.

For the dynamic panel threshold models, asymptotic normality of the GMM estimator is derived by Seo and Shin (2016) under the fixed $T$ scheme. However, the asymptotic normality is valid only for the discontinuous models since it requires the full rank condition on the Jacobian of the population moment, which is violated in the continuous models. Although the continuity-restricted estimator described in Seo et al. (2019) is asymptotically normal, it may be problematic since empirical researchers often do not agree about whether their threshold models should have a kink or a jump at the tipping point as in Khan and Senhadji (2001) and Bick (2010). Therefore, we are focusing on the unrestricted estimator and bootstrap inference methods which do not require any pretest on continuity or prior knowledge about the continuity of the true models.

We first show that when the true model is continuous, the asymptotic normality of the unrestricted GMM estimator breaks down and the convergence rate of the threshold estimator becomes $n^{1/4}$-rate, which is slower than the standard $\sqrt{n}$-rate. Moreover, the standard
nonparametric bootstrap is inconsistent in this case because the Jacobian from the bootstrap
distribution does not degenerate fast enough due to the slow convergence rate of the threshold
estimator.

We propose two different bootstrap methods to obtain confidence intervals of the parameters
that are consistent regardless of whether the true model is continuous or not. One is for the
threshold location, and the other is for the coefficients. The two bootstrap methods achieve
the consistency irrespective of the continuity of the model by adaptively setting the recentering
parameter at the standard residual bootstrap introduced by Hall and Horowitz (1996). This
means that our bootstrap moment function achieves zero not at the sample estimator but at
the parameter values that we propose. In the bootstrap for the threshold location, we employ
a grid bootstrap to fix the recentering parameter. The grid bootstrap was originally proposed
by Hansen (1999a) for inference on an autoregressive parameter and applies the test inversion
principle. In case of the coefficients the recentering parameter is set to adjust the unrestricted
estimator by a data driven criterion on the model’s continuity. We also introduce a bootstrap
method for testing the continuity of the regression models.

Furthermore, we establish the uniform validity of the grid bootstrap for the unknown
(dis)continuity of the threshold model. The importance of uniform validity is well-documented
in the literature, as seen in works by Mikusheva (2007), Andrews and Guggenberger (2009),
and Romano and Shaikh (2012), among others, who have studied the uniformity of resampling
procedures. Mikusheva (2007) showed the uniform validity of the grid bootstrap for linear
autoregressive models. Our work extends the advantage of the grid bootstrap to another non-
standard inference problem.

A set of Monte Carlo simulations demonstrates that the grid bootstrap performs favorably
for inference on the threshold location, not only when the model is continuous but also when it
includes a jump for various jump sizes. However, inference on the coefficients turns out to be
more challenging. Bootstrap confidence intervals based on percentiles of bootstrap distributions
for the coefficient estimators tend to show undercoverage while those based on the symmetric
versions exhibit overcoverage. Nevertheless, our residual bootstrap method improves upon the
standard nonparametric bootstrap in both cases.

We apply our inference methods to the dynamic firm investment model, whose static version
has been studied by Fazzari et al. (1988) or Hansen (1999b). It takes financial constraints into
account via the threshold effect to determine a firm’s investment decision.

In the literature, Dovonon and Renault (2013) and Dovonon and Hall (2018) also deal with
the degeneracy of the Jacobian in the context of the common conditional heteroskedasticity
testing problem. And a bootstrap based test for the common conditional heteroskedasticity
feature was proposed by Dovonon and Goncalves (2017). However, their works do not deal
with a discontinuous criterion function and their null hypothesis of interest always induces the
degeneracy of the first-order derivative. That is, they are only concerned with a hypothesis
testing and do not consider the confidence intervals. So, they do not have to address the
uncertainty associated with the potential degeneracy of the Jacobian.
Meanwhile, there is also a substantial body of literature on singularity-robust inference such as Andrews and Cheng (2012, 2014) and Han and McCloskey (2019), among many others. They are motivated by weak or non-identification problems, where models are not point identified. In constrast, we focus on the inference problem that does not involve identification failure even though the Jacobian of the moment restriction can become singular. Andrews and Guggenberger (2019) study more general singular cases than non-identification, but their approach requires differentiability of sample moments for the subvector inference. Since our model exhibits discontinuity, the method of Andrews and Guggenberger (2019) is not applicable.

This paper is organized as follows. Section 2 explains the dynamic panel threshold model. Section 3 presents the asymptotic distribution theories of the estimators and test statistics related to the threshold location and continuity. Section 4 proposes bootstrap methods. Section 5 reports Monte Carlo simulation results. Section 6 contains an empirical application. Section 7 concludes. The mathematical proofs and technical details are left to the Appendix.

2 Dynamic Panel Threshold Model

We consider the dynamic panel threshold model,

\[ y_{it} = x_{it}'\beta + (1, x_{it}')\delta 1\{q_{it} > \gamma\} + \eta_i + \epsilon_{it}, \]

where \(1 \leq i \leq n, 1 \leq t \leq T,\) and \(x_{it} \in \mathbb{R}^p\) is a regressor vector that includes \(y_{i,t-1}\) and \(q_{it}\). The threshold variable \(q_{it} \in \mathbb{R}\) is allowed to be endogenous and is the last element of \(x_{it}\). Then, we partition \(x_{it}\) and write \(x_{it} = (\xi_{it}', q_{it}')' \in \mathbb{R}^p\).

When \(x_{it}\) consists of the lagged dependent variables, the model becomes the well-known self-exciting threshold autoregressive (TAR) model popularized by Chan and Tong (1985). The static version where the lagged dependent variables are excluded from \(x_{it}\) was considered by Hansen (1999b), while the current dynamic model was studied by Seo and Shin (2016).

The parameter \(\gamma \in \Gamma\) denotes the threshold location, where \(\Gamma\) is a compact set in \(\mathbb{R}\), and \(\alpha = (\beta', \delta')' \in A \subset \mathbb{R}^{2p+1}\) denotes the collection of coefficients. Let \(\theta = (\alpha', \gamma) = (\beta', \delta', \gamma)' \in \Theta = A \times \Gamma\) denote the vector of all the parameters. The fixed effect \(\eta_i\) is constant across time for each individual in the panel data. It is not identified but is eliminated after first-differencing for the GMM estimation. The idiosyncratic error \(\epsilon_{it}\) is independent across individuals.

For the estimation, we use the GMM after the first-difference transformation

\[ \Delta y_{it} = \Delta x_{it}'\beta + 1_{it}(\gamma)'X_{it}\delta + \Delta \epsilon_{it}, \]

where

\[ X_{it} = \begin{pmatrix} (1, x_{it}') \\ (1, x_{it-1}') \end{pmatrix}, \quad \text{and} \quad 1_{it}(\gamma) = \begin{pmatrix} 1\{q_{it} > \gamma\} \\ -1\{q_{i,t-1} > \gamma\} \end{pmatrix}. \]

1Our analysis still holds if researchers have two sets of regressors \(x_{1it}\) and \(x_{2it}\) such that \(y_{it} = x_{1it}'\beta + (1, x_{2it}')\delta 1\{q_{it} > \gamma\} + \eta_i + \epsilon_{it}\) where \(q_{it}\) is an element of \(x_{2it}\). However, this paper sticks to the current form to keep the exposition simple.
Let \( z_{it} \) denote a set of instrumental variables at time \( t \) such that \( E[z_{it}\Delta \epsilon_{it}] \) becomes a zero vector, which may include lagged dependent variables \( y_{it-2}, \ldots, y_{it1} \) and certain lagged variables of covariates \( x_{it} \) and/or \( q_{it} \), depending on the assumptions regarding exogeneity of those variables.

Then, we can define a vector of moment functions for the GMM estimation,

\[
g_i(\theta) = \begin{pmatrix}
  z_{it0}(\Delta y_{it0} - \Delta x'_{it0}\beta - 1_{it0}(\gamma)'X_{it0}\delta) \\
  \vdots \\
  z_{iT}(\Delta y_{iT} - \Delta x'_{iT}\beta - 1_{iT}(\gamma)'X_{iT}\delta)
\end{pmatrix} \in \mathbb{R}^k,
\]

where \( k \geq \dim(\theta) = 2p + 2 \) and \( t_0 \geq 2 \) is the earliest period that the regressor and instrument can be defined. For example, \( k = (T-1)(T-2)/2 \) when \( z_{it} = (y_{it-2}, \ldots, y_{it1})' \) and \( t_0 = 3 \). Denote the population moment by \( g_0(\theta) = E[g_i(\theta)] \) and the sample moment by

\[
\tilde{g}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} g_i(\theta).
\]

We write \( g_i \) instead of \( g_i(\theta_0) \) for simplicity of notations.

We consider the two-stage GMM estimation of the dynamic panel threshold model. In the first stage, we get an initial estimate by \( \hat{\theta}(1) = \arg\min_{\theta \in \Theta} \tilde{g}_n(\theta)'\tilde{g}_n(\theta) \) to compute a weight matrix

\[
W_n = \left( \frac{1}{n} \sum_{i=1}^{n} [g_i(\hat{\theta}(1))g_i(\hat{\theta}(1))'] - \tilde{g}_n(\hat{\theta}(1))\tilde{g}_n(\hat{\theta}(1))' \right)^{-1},
\]

and obtain the second stage estimator

\[
\hat{\theta} = \arg\min_{\theta \in \Theta} \hat{Q}_n(\theta),
\]

where \( \hat{Q}_n(\theta) = \tilde{g}_n(\theta)'W_n\tilde{g}_n(\theta) \). Seo and Shin (2016) proposed averaging of a class of GMM estimators that are constructed from randomized first stage estimators. We do not pursue the averaging since our primary goal is the bootstrap inference.

In practice, the grid search algorithm is employed to compute the estimates. Note that when \( \gamma \) is given, \( \hat{\alpha}(\gamma) = \arg\min_{\alpha \in \Delta} \hat{Q}_n(\alpha, \gamma) \) can be easily computed because the problem becomes the estimation of the linear dynamic panel model. Then, \( \hat{\gamma} \) minimizes the profiled criterion \( \hat{Q}_n(\gamma) = \hat{Q}_n(\hat{\alpha}(\gamma), \gamma) \) over the grid of \( \Gamma \).

Let \( \theta_0 = (\alpha_0, \gamma_0)' = (\beta_0, \delta_0, \gamma_0)' \) denote the true parameter value that lies in the interior of \( \Theta \). For the point identification of \( \theta_0 \), \( g_0(\theta) = 0_k \) should hold if and only if \( \theta = \theta_0 \), where \( 0_k = (0, \ldots, 0)' \in \mathbb{R}^k \). Let

\[
M_{1i} = - \begin{bmatrix} z_{it0}\Delta x_{it0}' \\ \vdots \\ z_{iT}\Delta x_{iT}' \end{bmatrix} \in \mathbb{R}^{k \times p}, \quad M_{2i}(\gamma) = - \begin{bmatrix} z_{it0}1_{it0}(\gamma)'X_{it0} \\ \vdots \\ z_{iT}1_{iT}(\gamma)'X_{iT} \end{bmatrix} \in \mathbb{R}^{k \times (p+1)}.
\]
and $M_i(\gamma) = \begin{bmatrix} M_{i1} & M_{i2}(\gamma) \end{bmatrix}$. Additionally, define $M_0(\gamma) = E[M_i(\gamma)]$, $M_{10} = E[M_{i1}]$, $M_{20}(\gamma) = E[M_{i2}(\gamma)]$, $M_n(\gamma) = n^{-1}\sum_{i=1}^{n} M_i(\gamma)$, $M_{1n} = n^{-1}\sum_{i=1}^{n} M_{i1}$, and $M_{2n}(\gamma) = n^{-1}\sum_{i=1}^{n} M_{i2}(\gamma)$. We write $M_0$, $M_{20}$ and $M_n$ instead of $M_0(\gamma_0)$, $M_{20}(\gamma_0)$ and $M_n(\gamma_0)$, respectively, for simplicity of notation. The identification condition is stated in Theorem 1 that follows.

**Theorem 1.** Let the following two conditions hold:

(i) The matrix $M_0$ is of full column rank.

(ii) For any $\gamma \neq \gamma_0$, $M_{20}\delta_0$ is not in the column space of $M_{20}(\gamma)$.

Then, $\theta_0$ is a unique solution to $g_0(\theta) = 0_k$.

Theorem 1 (i) is the identification condition for the coefficients once the true threshold location is identified. This means that instruments should be relevant to the first differenced regressors appearing in (2) when $\gamma = \gamma_0$.

Theorem 1 (ii) is for the identification of the threshold location, which excludes the possibility of $\delta_0 = 0_{p+1}$. In the standard GMM problem, it is usually assumed that the Jacobian of $g_0(\theta)$ at $\theta_0$ is of full column rank for both the point identification and the asymptotic normality of the GMM estimator. The condition (ii) does not require the full rank condition on the Jacobian, which is related to the presence of a jump in the threshold model, and thus it generalizes the identification conditions in Seo and Shin (2016). When the model is continuous and has a kink at the threshold location, the last column of the Jacobian matrix, which is the first-order derivative with respect to $\gamma$, becomes a zero vector. This degeneracy does not violate the condition (ii), but it fails the asymptotic normality of the standard GMM estimator, which relies on the linearization of $g_0(\theta)$ near $\theta_0$ as in Newey and McFadden (1994).

To define the continuity, recall that $q_{it}$ is the last element of $x_{it}$ such that $x_{it} = (\xi_{it}', q_{it})' \in \mathbb{R}^p$. Accordingly, partition $\delta = (\delta_1, \delta_2, \delta_3)'$, where $\delta_2 \in \mathbb{R}^{p-1}$ and $\delta_1, \delta_3 \in \mathbb{R}$, and $\delta_0 = (\delta_{10}, \delta_{20}, \delta_{30})'$. Hence, $\delta_3$ is the change in the coefficient of the threshold variable when the threshold variable surpasses the tipping point. Likewise, $\delta_2$ and $\delta_1$ are the changes in the coefficients for the other regressors, $\xi_{it}$, and intercept, respectively. The continuity of the dynamic panel threshold model is formally defined in Definition 1.

**Definition 1.** Let $\delta \neq 0_{p+1}$. A dynamic panel threshold model is continuous if $\delta_2 = 0_{p-1}$ and $\delta_1 + \delta_3\gamma = 0$. Otherwise, it is discontinuous.

Note that this definition of continuity requires that $\delta_3 \neq 0$; otherwise, $\delta = 0_{p+1}$.

The rank of the first-order derivative matrix, say $D_1$, of $g_0(\theta)$ at $\theta = \theta_0$ is crucial to the standard asymptotic normality of the GMM estimator. Let $G$ denote the first-order derivative of $g_0(\theta)$ with respect to $\gamma$ at $\theta = \theta_0$. Then,

$$G = \begin{bmatrix} E_{t_0}[z_{t_0}(1, x_{t_0}'(1)|\gamma_0)f_{t_0}(\gamma_0) - E_{t_0-1}[z_{t_0}(1, x_{t_0-1}'(1)|\gamma_0)f_{t_0-1}(\gamma_0)] \\ \vdots \\ E_T[z_{iT}(1, x_{iT}'(1)|\gamma_0)f_{T}(\gamma_0) - E_{T-1}[z_{iT}(1, x_{iT-1}'(1)|\gamma_0)f_{T-1}(\gamma_0)] \end{bmatrix} \times \delta_0 \in \mathbb{R}^k, \quad (5)$$

$$G_0$$

6
where the conditional expectation $E_t[q|\cdot] = E[|q_{it} = q]$ and the density function $f_t(\cdot)$ of $q_{it}$ are assumed to exist. The derivation of $G$ is provided in the proof of Lemma C.1. Note that the first-order derivative of $g_0(\theta)$ with respect to $\theta$ at $\theta = \theta_0$ is $M_0$. The linear independence of $G$ from the other columns in $D_1$ is required for the standard linear approximation

$$g_0(\theta) \approx D_1(\theta - \theta_0) = M_0(\alpha - \alpha_0) + G(\gamma - \gamma_0).$$

Note that the vector $G$ can be written as the product of the matrix $G_0$ and the vector $\delta_0$, and the first and last columns of $G_0$ are linearly dependent since $q_{it} = \gamma_0$ for all $t$ due to the conditioning.

Then, the standard rank condition on the first derivative matrix $D_1$ can follow from a more primitive rank condition on $\begin{bmatrix} M_0 \; G_{0, -(p+1)} \end{bmatrix}$, that is, the linear independence of all the columns in $M_0$ and all but the last column of $G_0$, for the discontinuous case. Even if the primitive condition is met, however, the continuity restriction makes $G = 0_k$ since $E_s[z_{it}(1, x_{is})0|\gamma_0] = (\delta_{t0} + \delta_{t0}\gamma_0)E_s[z_{it}|\gamma_0] = 0$ for $s \leq t$. Additionally, the continuity destroys the asymptotic normality of the GMM estimator.

When the rank condition fails due to the continuity, the expansion becomes

$$g_0(\theta) \approx M_0(\alpha - \alpha_0) + H(\gamma - \gamma_0)^2,$$

where

$$H = \frac{\partial^2 g_0(\theta_0)}{\partial \gamma \partial \gamma} = \frac{\delta_{t0}}{2} \left( \begin{array}{c} E_{t0}[z_{it0}|\gamma_0]f_{t0}(\gamma_0) - E_{t0-1}[z_{it0}|\gamma_0]f_{t0-1}(\gamma_0) \\ \vdots \\ E_T[z_{iT}|\gamma_0]f_T(\gamma_0) - E_{T-1}[z_{iT}|\gamma_0]f_{T-1}(\gamma_0) \end{array} \right) \in \mathbb{R}^k. \quad (6)$$

The detailed derivation is given in the proof of Lemma C.1. It is worth noting that $H$ is identical to the first column of $G_0$ up to a constant multiple. Then, the rank condition on $\begin{bmatrix} M_0 \; H \end{bmatrix}$ is implied by the rank condition on $\begin{bmatrix} M_0 \; G_{0, -(p+1)} \end{bmatrix}$. Thus, the rank condition on $\begin{bmatrix} M_0 \; G_{0, -(p+1)} \end{bmatrix}$ can be viewed as a sufficient condition for both Assumptions LK and LJ in the next section, apart from the continuity restriction on $\theta$. Next section formalizes this discussion and presents the asymptotic distribution of the GMM estimator $\hat{\theta}$ under the continuity.

### 3 Asymptotic theory

This section considers the asymptotic analysis when $T$ is fixed, the data are independent and identically distributed across $i$, and $n \to \infty$. We make the following assumptions.

**Assumption G.** The parameter space $\Theta$ is compact and $\theta_0 \in \text{int } \Theta$. $M_0$ is of full column rank, and $M_0\delta_0$ is not in the column space of $M_0(\gamma)$ for any $\gamma \neq \gamma_0$. $\Omega = E[g_tg_t']$ is positive definite. $E\|z_{it}\|^4$, $E\|x_{it}\|^4$, and $E\|e_{it}\|^4$ are finite for all $t$.

**Assumption D.** For all $t$, (i) $q_{it}$ has a continuous distribution and a bounded density $f_t(\cdot)$, which is continuously differentiable at $\gamma_0$ and $f_t(\gamma_0) > 0$. (ii) $E_t[z_{it}(1, x_{is})|q]$ and
\( E_t[z_{it}(1, x_{it-1})|q] \) are continuous on \( q \in \Gamma \) and continuously differentiable at \( q = \gamma_0 \).

**Assumption LK.** \( D_2 = \left[ \begin{array}{c} M_0 \quad H \end{array} \right] \in \mathbb{R}^{k \times (2p+2)} \) has full column rank.

Assumption G and D are similar to Assumptions 1 and 2 in Seo and Shin (2016) except for the differentiability conditions in Assumption D which allows the second-order derivative of the population moment to be defined. Note that Assumption G includes the conditions in Theorem 1. Assumption LK is a rank condition for a nondegenerate asymptotic distribution when the underlying model is continuous. This condition may be viewed as less restrictive than the standard rank assumption as discussed in the preceding section where \( G \) and \( H \) are defined.

For easy reference, we restate the standard full rank assumption for the asymptotic normality of the GMM estimator for the discontinuous threshold regression below.

**Assumption LJ.** \( D_1 = \left[ \begin{array}{c} M_0 \quad G \end{array} \right] \in \mathbb{R}^{k \times (2p+2)} \) has full column rank.

In a simple model, where \( y_{it} = x'_{it}\beta + (\delta_1 + \delta_3 q_{it})1\{q_{it} > \gamma\} + \eta_i + \epsilon_{it} \), Assumption LK is equivalent to Assumption LJ because \( G = (\delta_{10} + \delta_3 \gamma_{0})G_{01} \) while \( H = \frac{\delta_{00}}{2}G_{01} \), where \( G_{01} \) is the first column of \( G \) in (5).

Theorem 2 below establishes the asymptotic distribution of the GMM estimator when the dynamic panel threshold model is continuous.

**Theorem 2.** When the true model is continuous and Assumptions G, D, and LK hold,

\[
\left( \frac{\sqrt{n}(\hat{\alpha} - \alpha_0)}{\sqrt{n}(\hat{\gamma} - \gamma_0)^2} \right) \overset{d}{\rightarrow} \begin{pmatrix} u - (M_0'\Omega^{-1}M_0)^{-1}M_0'\Omega^{-1}Hv \end{pmatrix},
\]

where \( u \sim N(0, (M_0'\Omega^{-1}M_0)^{-1}) \) and \( v^* \sim N(0, (H'\Xi H)^{-1}) \) are independent of each other, while \( \Xi = \Omega^{-1} - \Omega^{-1}M_0(M_0'\Omega^{-1}M_0)^{-1}\Omega^{-1} \) and \( v = \max(v^*, 0) \).

We observe that the convergence rate of \( \hat{\gamma} \) is \( n^{1/4} \), which is slower than the standard \( \sqrt{n} \)-rate. Meanwhile, Seo and Shin (2016) show the \( \sqrt{n} \)-convergence rate for \( \hat{\gamma} \) when the model is discontinuous. Intuitively, it would be more difficult to detect the precise threshold location when there is a kink than when there is a jump at the tipping point. More technically, when the threshold model is discontinuous and the Jacobian is not singular, the limit of the GMM objective function admits a quadratic approximation with respect to \( \gamma \) at the true value, while the limit admits a quartic approximation for the continuous model. Hence, the limit objective function becomes flatter in \( \gamma \) at the true value resulting in the slower convergence rate. Hidalgo et al. (2019) also showed in the least square context that when the model is continuous, the convergence rate of the threshold estimator slows down to \( n^{1/3} \), while it is superconsistent \( n \)-rate when the model is discontinuous.

Moreover, we can observe that the asymptotic distribution of \( \hat{\alpha} \) is also shifting to a non-normal distribution. Hence, standard inference methods based on the asymptotic normality becomes invalid for both coefficients and threshold location parameters in the continuous dynamic panel threshold model.
The asymptotic distribution of the GMM estimator is identical to the distribution reported in Theorem 1 (b) in Dovonon and Hall (2018), which studies a smooth GMM problem with the degeneracy of the Jacobian. Theorem 2 shows that even though the criterion of our threshold model is discontinuous, the same asymptotic distribution as that of Dovonon and Hall (2018) appears.

The censored normal distribution also appears in Andrews (2002) which studies the estimation of a parameter on a boundary. Heuristically, because our analysis depends on the second-order derivative of $\gamma$ for the local polynomial expansion of $g_0(\theta)$ near $\theta_0$, only the asymptotic distribution of $(\hat{\gamma} - \gamma_0)^2$ can be derived. Since $(\hat{\gamma} - \gamma_0)^2$ should be nonnegative, the asymptotic censored normal distribution appears as in Andrews (2002). Meanwhile, Dovonon and Goncalves (2017) show that the standard nonparametric bootstrap becomes invalid when the Jacobian degenerates. To address this issue, we propose different bootstrap methods in Section 4 for the inference of the parameters.

The asymptotic distribution in Theorem 2 can be used for parameter inference when the true model is continuous, but the estimator is obtained without imposing the continuity restriction. As discussed in Remark 1, $M_0$ and $\Omega$ can be consistently estimated, while $H$ can be nonparametrically estimated similarly to $G$. It is then straightforward to simulate the limit distribution from Theorem 2 by generating random numbers for $u$ and $v$. However, there are several drawbacks to that approach, and hence we do not recommend it. First, empirical researchers might construct confidence intervals based on Theorem 2 when they cannot reject the continuity. However, Leeb and Pötscher (2005) show that confidence intervals after model selection are subject to size-distortion. Second, even if the true model is known to be continuous, the continuity-restricted estimator explained in Seo et al. (2019) is more efficient and asymptotically normal. Therefore, using the continuity-restricted estimator for estimation and inference is preferable. Finally, the nonparametric estimation of $H$ requires a tuning parameter and has a slower convergence rate.

Remark 1. Seo and Shin (2016), in their section 4, derive the asymptotic distribution of the GMM estimator and propose inference methods when the underlying model is discontinuous. When the true model is discontinuous and Assumptions G, D, and LJ hold,

$$\frac{\sqrt{n}(\hat{\alpha} - \alpha_0)}{\sqrt{n}(\hat{\gamma} - \gamma_0)} \xrightarrow{d} N(0, (D_1'\Omega^{-1}D_1)^{-1}).$$

$\Omega$ can be estimated by $\hat{\Omega} = \frac{1}{n} \sum_{i=1}^n [g_i(\hat{\theta})g_i(\hat{\theta})'] - \bar{g}_n(\hat{\theta})\bar{g}_n(\hat{\theta})'$. Note that $D_1 = \begin{bmatrix} M_0 ; G \end{bmatrix}$, and $M_0$ can be estimated by $M_n(\hat{\gamma})$, while the estimation of $G$ involves nonparametric estimation of the conditional means and densities. See section 4 of Seo and Shin (2016) for more details.

3.1 Testing for threshold value

Since the asymptotic distribution of the threshold estimator is not standard, we consider the GMM distance test introduced by Newey and West (1987) for a hypothesis on the location of
the threshold. Let the test statistic for the threshold location at $\gamma$ be

$$D_n(\gamma) = n(\min_{\alpha \in A} \hat{Q}_n(\alpha, \gamma) - \hat{Q}_n(\hat{\theta})),$$

and let $\chi^2_1$ denote the chi-square distribution with 1 degree of freedom.

**Theorem 3.** (i) If $\gamma = \gamma_0$, the true model is continuous, and Assumptions G, D, and LK hold, then

$$D_n(\gamma) \xrightarrow{d} Z_0^2$$

where $Z_0 = \max(0, Z_0^*)$, $Z_0^* \sim N(0, 1)$.

(ii) If $\gamma = \gamma_0$, the true model is discontinuous, and Assumptions G, D, and LJ hold, then

$$D_n(\gamma) \xrightarrow{d} \chi^2_1.$$

(iii) If $\gamma \neq \gamma_0$, then for any $M < \infty$, $\lim_{n \to \infty} P(D_n(\gamma) < M) = 0$.

Theorem 3 (i) presents the asymptotic distribution of the distance statistic under the continuity. Due to the censoring, the asymptotic distribution becomes a mixture of the $\chi^2_1$ distribution with weight 1/2 and zero with weight 1/2. This type of distribution also arises in the context of testing parameters on a boundary; see e.g. Andrews (2001).

Meanwhile, the chi-square limit in Theorem 3 (ii) extends Newey and West (1987) for a discontinuous moment function. Seo and Shin (2016) did not study the distance statistic.

Theorem 3 (iii) shows that the GMM distance test for a threshold location is consistent. It also serves as the consistency of a bootstrap test together with Theorem 5 since the bootstrap statistic is stochastically bounded whether or not the threshold location is true.

Since the limit distribution depends on the continuity of the model, we introduce a bootstrap in Section 4.1, which is valid regardless of the model continuity. Furthermore, Appendix H establishes the uniform validity of the bootstrap inference for the threshold location under some simplifying assumptions.

### 3.2 Testing continuity

We propose a test for the continuity of the threshold model, similar to the approach used by Gonzalo and Wolf (2005) or Hidalgo et al. (2023) in the threshold regression literature. While empirical researchers may employ the test to select a model and do inference based on the selected model, we utilize the test to modify the standard nonparametric bootstrap to make the bootstrap valid irrespective of the model continuity. Details of the use of the continuity test statistic in the bootstrap method are explained in Section 4.2.

The continuity hypothesis is a joint hypothesis. We employ the GMM distance test. Let $\hat{\theta} = \arg \min_{\theta : \delta_2 = 0, \delta_1 = -\dot{\gamma}} \hat{Q}_n(\theta)$ be the continuity-restricted estimator. The GMM distance test statistic is

$$\mathcal{T}_n = n(\hat{Q}_n(\hat{\theta}) - \hat{Q}_n(\hat{\theta})).$$
Theorem 4. (i) When the true model is continuous and Assumptions G, D, and LK hold,

\[ T_n \xrightarrow{d} V_1 - V_2 + V_3, \]

where \( V_1 = Z'\Psi M_{20}(M_{10}'\Psi M_{20})^{-1}M_{10}'\Psi Z, \) \( V_2 = Z'\Psi N_{20}(N_{10}'\Psi N_{20})^{-1}N_{10}'\Psi Z, \) \( V_3 = Z_0^2, \)

\( Z \sim N(0, \Omega), \) \( Z_0 = \max(0, Z_0^*), \) \( Z_0^* \sim N(0, 1), \) \( Z_0 \) and \( Z \) are independent, \( \Psi = \Omega^{-1} - \Omega^{-1}M_{10}(M_{10}'\Omega^{-1}M_{10})^{-1}M_{10}'\Omega^{-1}, \) and \( N_{20} = M_{20}\begin{pmatrix} 1 & 0 \\ -\delta_0' & 1 \end{pmatrix}', \) \( \hat{M}_1 = \bar{M}_{1n}, \) \( \hat{M}_2 = \bar{M}_{2n}(\hat{\gamma}), \) etc. Another way to obtain the critical values is via a bootstrap method, which will be introduced in Section 4.3.

(ii) If the model is discontinuous, then \( \lim_{n \to \infty} P(n^{-m}T_n < M) = 0 \) for any \( m \in [0, 1) \) and \( M < \infty. \)

While the limit distribution in Theorem 4 (i) is non-standard, it can be simulated to obtain critical values for the test using consistent plug-in sample analogue estimators, e.g., \( \hat{\Omega} = \frac{1}{n}\sum_{i=1}^n g_i(\hat{\theta})g_i(\hat{\theta})' - \bar{g}_n(\hat{\theta})\bar{g}_n(\hat{\theta})', \) \( \hat{M}_1 = \bar{M}_{1n}, \) \( \hat{M}_2 = \bar{M}_{2n}(\hat{\gamma}), \) etc. Another way to obtain the critical values is via a bootstrap method, which will be introduced in Section 4.3.

Theorem 4 (ii) shows that the continuity test is consistent. It also implies the consistency of the bootstrap test together with Theorem 7, which shows that the bootstrap test statistic is stochastically bounded even when the true model is not continuous. The divergence rate of \( T_n, \) which is faster than \( n^m \) for any \( 0 \leq m < 1, \) is exploited to modify the standard nonparametric bootstrap for the coefficients as detailed in Section 4.2.

4 Bootstrap

As usual, the superscript “\( \ast \)” denotes the bootstrap quantities or the convergence of bootstrap statistics under the bootstrap probability law conditional on the original sample. For example, \( E^\ast \) denotes the expectation with respect to the bootstrap probability law conditional on the data. “\( d^\ast \),” in \( P^\ast \) denotes the distributional convergence of bootstrap statistics under the bootstrap probability law with probability approaching one. We write “\( \nu_n^\ast = O^\ast_p(1), \)” in \( P^\ast \) if a sequence \( \nu_n^\ast \) is stochastically bounded under the bootstrap probability law with probability approaching one. More details are written in Section B.1. Let \( \hat{F}_n^{\ast-1}(\varphi; S^\ast) \) denote the empirical \( \varphi\)-th quantile of a bootstrap statistic \( S^\ast. \)

This section introduces three different bootstrap schemes. The first bootstrap is for constructing bootstrap confidence interval(CI)s of the threshold, while the second bootstrap is for constructing bootstrap CIs of the coefficients. Both methods aim to provide valid inferences, regardless of whether the model is continuous or not. The third bootstrap is for testing continuity of the threshold model. The three bootstrap methods can be represented by means of Algorithm 1 with suitable choices of \( \theta_0^\ast = (\beta_0^\ast', \delta_0^\ast', \gamma_0^\ast)' \).

In step 1, we resample the regressors, the instruments, and the residuals jointly to maintain the dependence among them, unlike in the usual residual bootstrap. See e.g. Giannerini et al. (2024) for the description of the standard residual bootstrap, which resamples the residuals only, and the wild bootstrap for the testing of linearity in the threshold regression. There could be other ways of resampling not mentioned here and we do not attempt to decide which is the
Algorithm 1 Bootstrap with $\theta_0^*$

1: For $i = 1, \ldots, n$, let $i^*$ be the $i$th i.i.d. random draw from the discrete uniform distribution on $\{1, \ldots, n\}$. Generate a bootstrap sample $\{ (x_{i,t}^*, x_{i,t-1}^*, z_{i,t}^*, \Delta x_{i,t}^*)^T : i = 1, \ldots, n \}$ by setting $(x_{i,t}^*, x_{i,t-1}^*, z_{i,t}^*, \Delta x_{i,t}^*)^T_{t = t_0} = (x_{i,t}, x_{i,t-1}, z_{i,t}, \Delta x_{i,t})^T_{t = t_0}$ for each $i$.

2: Generate $\{(\Delta y_{i,t}^*)_{t = t_0} : i = 1, \ldots, n\}$ using $\theta_0^*$ by

$$\Delta y_{i,t}^* = \Delta x_{i,t}^* \beta_0^* + 1_{i}^*(\gamma_0^*)X_{i,t}^* \delta_0^* + \Delta \epsilon_{i,t}^*,$$

where $\Delta x_{i,t}^* = x_{i,t}^* - x_{i,t-1}^*$.

3: Define the bootstrap moment function $g_i^*(\theta) = (g_{i0}(\theta), \ldots, g_{i,T}(\theta))'$ where $g_i^*(\theta) = z_{i,t}^*(\Delta y_{i,t}^* - \Delta x_{i,t}^* \beta - 1_{i}^*(\gamma)X_{i,t}^*)$.

4: Define the (recentered) bootstrap sample moment

$$\overline{g}_i^*(\theta) = \frac{1}{n} \sum_{i=1}^{n} (g_i^*(\theta) - \overline{g}_i(\theta)).$$

5: Compute the initial estimator $\hat{\theta}_0^* = \arg \min_{\theta} \overline{g}_i^*(\theta)'\overline{g}_i^*(\theta)$ and the weight matrix $W_n^* = (\frac{1}{n} \sum_{i=1}^{n} g_i^*(\hat{\theta}_0^*) g_i^*(\hat{\theta}_0^*)' - \frac{1}{n} \sum_{i=1}^{n} g_i^*(\hat{\theta}_0^*) |\frac{1}{n} \sum_{i=1}^{n} g_i^*(\hat{\theta}_0^*)|')^{-1}$.

6: Define the bootstrap criterion function $\hat{Q}_n^*(\theta) = \overline{g}_i^*(\theta)'W_n^* \overline{g}_i^*(\theta)$, and obtain the bootstrap estimator or the test statistics.

The parameter $\theta_0^*$ is used in step 2 of Algorithm 1 to generate the dependent variables in the bootstrap samples. In step 4, centering of the bootstrap sample moment is done by subtracting $\overline{g}_i(\theta) = (\frac{1}{n} \sum_{i=1}^{n} z_{i,t}^* \Delta \epsilon_{i,t}, \ldots, \frac{1}{n} \sum_{i=1}^{n} z_{i,T}^* \Delta \epsilon_{i,T})'$. Note that the expectation of $\overline{g}_i^*(\theta)$ by the bootstrap probability law conditional on the data becomes zero when $\theta = \theta_0^*$ due to the centering, which can be easily checked from the following equations: $g_i^*(\theta_0^*) = z_{i,t}^*(\Delta y_{i,t}^* - \Delta x_{i,t}^* \beta_0^* - 1_{i}^*(\gamma_0^*)X_{i,t}^* \delta_0^*) = z_{i,t}^* \Delta \epsilon_{i,t}$ and $E[|g_i^*(\theta_0^*)|] = n^{-1} \sum_{i=1}^{n} z_{i,t}^* \Delta \epsilon_{i,t}$ for $t = t_0, \ldots, T$.

A different choice of $\theta_0^*$ leads to a different bootstrap. For example, if $\theta_0^* = \hat{\theta}$, then the bootstrap becomes the standard nonparametric bootstrap in Hall and Horowitz (1996) because $\Delta y_{i,t}^* = \Delta y_{i,t}$ holds true for $i = 1, \ldots, n$ and $t = t_0, \ldots, T$ in step 2. Note that, for $\theta_0^*$ not equal to $\hat{\theta}$, step 2 of Algorithm 1 generates $\Delta y_{i,t}^*$'s that are generally different from $\Delta y_{i,t}$'s. The following subsections detail three different choices of $\theta_0^*$ for three different inference problems.

### 4.1 Grid bootstrap for threshold location

To construct CIs of the threshold location, we propose to employ the grid bootstrap method introduced by Hansen (1999a) for autoregressive models. Let $\Gamma_n = \{ \gamma : \ell = 1, \ldots, L \}$ be a grid of the candidate thresholds. The grid bootstrap constructs the confidence set by inverting the bootstrap threshold location tests over $\Gamma_n$. Specifically, a sequence of hypothesis tests for the hypothesized threshold locations in $\Gamma_n$ are performed by the bootstrap that imposes the
null to generate bootstrap samples.

The null imposed bootstrap at a point $\gamma_\ell \in \Gamma_n$ can be implemented by setting $\theta_0^* = (\hat{\alpha}(\gamma_\ell)', \gamma_\ell)'$ in Algorithm 1, and the bootstrap test statistic is

$$D_n^*(\gamma_\ell) = n(\min_{\alpha \in A} \hat{Q}_n^*(\alpha, \gamma_\ell) - \min_{\theta \in \Theta} \hat{Q}_n^*(\theta)).$$

The null hypothesis $H_0 : \gamma = \gamma_\ell$ is rejected at size $\tau$ if $D_n(\gamma_\ell) > \hat{\nu}_n^{-1}(1 - \tau; D_n^*(\gamma_\ell))$. Consequently, after running the null imposed bootstrap for each point in $\Gamma_n$, we can construct the $\text{CI}_{\text{grid}}^{\text{grid}}_{n, 1 - \tau} = \{\gamma \in \Gamma_n : D_n(\gamma) \leq \hat{\nu}_n^{-1}(1 - \tau; D_n^*(\gamma))\}$.

Note that the confidence set is not necessarily a connected set, even though researchers can convexify the set to get a connected CI. The CI does not become an empty set because $D_n(\hat{\gamma}) = 0$ while $D_n^*(\hat{\gamma}) \geq 0$. The consistency of the grid bootstrap method is implied by Theorem 5 that follows.

**Theorem 5.** For a given $\gamma \in \Gamma$, assume that $D_n^*(\gamma)$ is obtained by Algorithm 1 with $\theta_0^* = (\hat{\alpha}(\gamma)', \gamma)'$.

(i) If $\gamma = \gamma_0$, the true model is continuous, and Assumptions G, D, and LK hold, then

$$D_n^*(\gamma) \overset{d^*}{\rightarrow} Z_0^2 \quad \text{in } P,$$

where $Z_0 = \max(0, Z_0^*)$ and $Z_0^* \sim N(0, 1)$.

(ii) If $\gamma = \gamma_0$, the true model is discontinuous, and Assumptions G, D, and LJ hold, then

$$D_n^*(\gamma) \overset{d^*}{\rightarrow} \chi_1^2 \quad \text{in } P.$$

(iii) If $\gamma \neq \gamma_0$, then $D_n^*(\gamma) = O_p^*(1)$ in $P$.

Theorem 5 (i) and (ii) show that the limit distribution of the bootstrap test statistic, conditional on the data, is identical to that of the sample test statistic regardless of the continuity of the true model. Therefore, the CI of the threshold location by the grid bootstrap, (7), achieves an exact coverage rate for both continuous and discontinuous models asymptotically. Specifically, $\lim_{n \to \infty} P(\gamma_0 \in \text{CI}_{n, 1 - \tau}^{\text{grid}}) = 1 - \tau$ for both cases (i) and (ii). Theorem 5 (iii) says that the bootstrap test statistic is still stochastically bounded, conditionally on the data, under the alternatives. As Theorem 3 (iii) shows that the sample test statistic is stochastically unbounded under the alternatives, the grid bootstrap CI has power against the alternative threshold locations.

### 4.1.1 Uniform validity of grid bootstrap

We extend Theorem 5 to the uniform validity of the grid bootstrap to ensure its good finite sample performance when the model is nearly continuous. We establish the uniform validity for
the following simplified specification for analytical tractability:

\[ y_{it} = x_{it}' \beta + (\delta_1 + \delta_3 q_{it})1\{q_{it} > \gamma\} + \eta_i + \epsilon_{it}, \]

where \( \theta = (\beta', \delta', \gamma)' \) and \( \delta = (\delta_1, \delta_3)' \) in this subsection.

This section briefly states the uniformity result of the grid bootstrap and gives heuristic justification. Our derivation follows Andrews et al. (2020). It is highly complicated and involves more technical conditions, which are stated in Appendix H.

Specifically, we establish in Theorem H.1 that

\[ \liminf_{n \to \infty} \inf_{\phi_0 \in \Phi_0} P_{\phi_0}(\gamma_0 \in CI_{n,1-\tau}^{grid}) = \limsup_{n \to \infty} \sup_{\phi_0 \in \Phi_0} P_{\phi_0}(\gamma_0 \in CI_{n,1-\tau}^{grid}) = 1 - \tau, \]

where \( P_{\phi} \) is the probability law when the model is specified by \( \phi = (\theta, F) \) and \( F \) is the distribution of \( \{\eta_i, y_{i0}, (z_{it}, x_{it}, \epsilon_{it})^T\}_{t=1}^{T} \). The collection of true models \( \Phi_0 \) includes both continuous nor discontinuous threshold models. More detailed discussions of technical assumptions about \( \Phi_0 \) are given in Appendix H.

For the uniformity analysis, we need to consider drifting sequences of true parameters \( \phi_{0n} = (\theta_{0n}, F_{0n}) \) such that \( \theta_{0n} \to \theta_{0,\infty} \) and \( F_{0n} \to F_{0,\infty} \). Here, the distance between \( F_{0n} \) and \( F_{0,\infty} \) is induced by a specific choice of norm that is explained in Appendix H. To show the uniform validity of the grid bootstrap CI, we need to verify that the limit distribution of \( D_n^*(\gamma_{0n}) \) conditional on the data is identical to the limit distribution of \( D_n(\gamma_{0n}) \) under all the above drifting sequences of models. Our analysis finds that the limit distribution of the threshold location test statistic under the true null, i.e., the limit distribution of \( D_n(\gamma_{0n}) \), is determined by \( \zeta = \lim_{n \to \infty} n^{1/4}(\delta_{10n} + \delta_{30n} \gamma_{0n}) \); see Lemma H.1 for details. When \( \zeta = 0 \), the limit distribution of \( D_n(\gamma_{0n}) \) is as described in Theorem 3 (i). In contrast, when \( |\zeta| = \infty \), the limit distribution is the \( \chi^2 \)-distribution as in Theorem 3 (ii). When \( \zeta \) is finite and nonzero, then \( D_n(\gamma_{0n}) \) has a nonstandard limit distribution that depends on \( \zeta \).

Therefore, if \( \theta_{0n} \) comprises a true parameter sequence of a bootstrap scheme, then \( n^{1/4}(\delta_{10n}^* + \delta_{30n}^* \gamma_{0n}^*) \) should consistently estimate \( \zeta \) for the bootstrap to exhibit the same asymptotic behavior as the sample statistic.

Note that the bootstrap test statistic \( D_n^*(\gamma_{0n}) \) is drawn from the bootstrap that imposes the null threshold location \( \gamma_{0n} \) under our grid bootstrap scheme. The true parameter of the bootstrap dgp is \( \theta_{0n}^* = (\hat{\alpha}_n(\gamma_{0n})', \gamma_{0n})' \) where \( \hat{\alpha}_n(\gamma) = (\hat{\beta}_n(\gamma)', \hat{\delta}_1n(\gamma), \hat{\delta}_3n(\gamma))' = \text{arg min}_\alpha \hat{Q}_n^*(\alpha, \gamma) \). The restricted estimator satisfies \( \|\hat{\alpha}(\gamma_{0n}) - \alpha_{0n}\| = O_p(n^{-1/2}) \) as the problem becomes estimating a standard linear dynamic panel model, and hence \( n^{1/4}(\hat{\delta}_{1n}(\gamma_{0n}) + \hat{\delta}_{3n}(\gamma_{0n}) \gamma_{0n}) = \zeta + o_p(1) \). Therefore, \( D_n^*(\gamma_{0n}) \) conditionally converges to the limit distribution of \( D_n(\gamma_{0n}) \), which leads to the uniform validity of the grid bootstrap confidence interval. In contrast, \( \hat{\theta} \) does not satisfy this property for some \( \zeta \) and the bootstrap building on \( \hat{\theta} \) may not be uniformly valid.
4.2 Residual bootstrap for coefficients

The bootstrap CIs of the coefficients can be obtained by applying Algorithm 1 with \( \theta_0^* \) set as

\[
\theta_0^* = w_n \hat{\theta} + (1 - w_n) \bar{\theta}, \quad w_n = \min \left( \frac{T_n}{o(n^{1/4})}, 1 \right),
\]

where \( \hat{\theta} = \arg\min_{\theta \in \Theta; \delta_2 = 0, \delta_1 = -\delta_1, \gamma} \hat{Q}_n(\theta) \) is the continuity-restricted estimator. \( \hat{C} \) is some estimated quantile, such as the 50th percentile, of the limit distribution of the continuity test statistic \( T_n \) when the model is continuous. \( \hat{C} \) can be obtained either by methods in Section 3.2 or Section 4.3. Since \( w_n = O_p(n^{-1/4}) \) if the true model is continuous, and \( w_n = 1 + o_p(1) \) if the model is discontinuous, the true parameter value for the bootstrap adapts to the model continuity.

After collecting the bootstrap estimators

\[
\hat{\theta}^* = (\hat{\alpha}^*, \hat{\gamma}^*)' = \arg\min_{\theta \in \Theta} \hat{Q}_n^*(\theta),
\]

we can construct CIs of the coefficients using the percentiles of either \( \sqrt{n}|\hat{\alpha}_j^* - \alpha_{j0}^*| \) or \( \sqrt{n}|\hat{\gamma}_j - \gamma_{j0}| \). Here, \( \alpha_j^* \) and \( \alpha_{j0}^* \) are the \( j \)-th elements of \( \hat{\alpha}^* \) and \( \alpha_{0}^* \), respectively. The 100(1 - \( \tau \))\% CI of the \( j \)-th element of the coefficients \( \alpha_j \) can be constructed by

\[
\{ \alpha_j \in \mathbb{R} : \hat{F}_n^{-1}(\tau; \sqrt{n}|\hat{\alpha}_j^* - \alpha_{j0}^*|) \leq \sqrt{n}|\hat{\alpha}_j - \alpha_j| \leq \hat{F}_n^{-1}(1 - \tau; \sqrt{n}|\hat{\alpha}_j^* - \alpha_{j0}^*|) \} \quad \text{or} \quad (9)
\]

\[
\{ \alpha_j \in \mathbb{R} : \sqrt{n}|\hat{\alpha}_j - \alpha_j| \leq \hat{F}_n^{-1}(1 - \tau; \sqrt{n}|\hat{\alpha}_j^* - \alpha_{j0}^*|) \}. \quad (10)
\]

The validity of the residual bootstrap CI is implied by Theorem 6 that follows.

We make the following additional assumption to derive the limit distribution of the bootstrap estimator when the true model is discontinuous.

**Assumption P.** The continuity-restricted estimator \( \hat{\theta} = \arg\min_{\theta_2 \in \Theta; \delta_2 = 0, \delta_1 = -\delta_1} \hat{Q}_n(\theta) \) is \( O_p(1) \).

The assumption holds if \( M_0(\gamma) \) has full column rank for all \( \gamma \in \Gamma \). Details are explained in the comment after Lemma D.6.

**Theorem 6.** Let \( \hat{\theta}^* \) be obtained by Algorithm 1 with \( \theta_0^* \) set as (8). (i) When the true model is continuous and Assumptions G, D, and LK hold,

\[
\begin{pmatrix}
\sqrt{n}|\hat{\alpha}_j^* - \alpha_{j0}^*| \\
\sqrt{n}|\hat{\gamma}_j - \gamma_{j0}^*|
\end{pmatrix} \overset{d}{\to} \begin{pmatrix}
(M' \Omega^{-1} M_0)^{-1} M_0' \Omega^{-1} H v \\
v
\end{pmatrix} \quad \text{in } P,
\]

where \( u \) and \( v \) are defined as in Theorem 2.

(ii) When the true model is discontinuous and Assumptions G, D, LJ, and P hold,

\[
\begin{pmatrix}
\sqrt{n}|\hat{\alpha}_j^* - \alpha_{j0}^*| \\
\sqrt{n}|\hat{\gamma}_j - \gamma_{j0}^*|
\end{pmatrix} \overset{d}{\to} N(0, (D_1' \Omega^{-1} D_1)^{-1}) \quad \text{in } P.
\]
The asymptotic distributions of the bootstrap estimators in Theorem 6, conditional on the data, match those of the sample estimators for both continuous and discontinuous cases. Therefore, the residual bootstrap CI becomes asymptotically valid in a pointwise sense, regardless of whether the model is continuous or discontinuous. We acknowledge that Theorem 6 is not about the uniform validity of the bootstrap CI. The difficulty in establishing the uniform validity lies in analyzing asymptotic behaviors of $T_n$ and $w_n$ for drifting sequences of true models. $T_n$ already exhibits an irregular limit distribution even in the pointwise setup, as shown in Theorem 4 (i). This paper does not provide a theoretical analysis regarding the uniformity of the residual bootstrap. Instead, we conduct Monte Carlo experiments for nearly continuous cases in Section 5 and leaves theoretical work on the uniformity of the bootstrap method to future research.

The key motivation for setting $\theta_0^*$, the true parameter of the bootstrap dgp, by (8) is to make $\delta_{10}^* + \delta_{30}^* \gamma_0^*$ degenerate fast enough when the underlying model is continuous. The $n^{1/4}$ convergence rate of the unrestricted estimator $\hat{\gamma}$ to $\gamma_0$ is not sufficiently fast. To see this, let the first-derivative of the population moment with respect to $\gamma$ at $\theta$ be

$$G(\theta) = (\delta_1 + \delta_3 \gamma) \cdot \begin{bmatrix}
E_{t_0}[z_{it_0}\gamma]f_{t_0}(\gamma) - E_{t_0-1}[z_{it_0}\gamma]f_{t_0-1}(\gamma) \\
\vdots \\
E_T[z_{iT}\gamma]f_T(\gamma) - E_{T-1}[z_{iT}\gamma]f_{T-1}(\gamma)
\end{bmatrix}
+ \begin{bmatrix}
E_{t_0}[z_{it_0}z_{it_0}^\prime \delta_2^* | \gamma]f_{t_0}(\gamma) - E_{t_0-1}[z_{it_0}z_{it_0}^\prime - \delta_2^* | \gamma]f_{t_0-1}(\gamma) \\
\vdots \\
E_T[z_{iT}z_{iT}^\prime \delta_2^* | \gamma]f_T(\gamma) - E_{T-1}[z_{iT}z_{iT}^\prime - \delta_2^* | \gamma]f_{T-1}(\gamma)
\end{bmatrix},$$

(11)

for which we recall that $x_{it} = (\xi_{it}^\prime, q_{it}^\prime)$ and that $G(\theta_0) = 0_k$ under continuity. For the validity of a bootstrap method, the degeneracy of the Jacobian should be mimicked by the bootstrap dgp. In our residual bootstrap method, the Jacobian is $G(\theta_0^*) = O_p(n^{-1/2})$. However, it is $G(\hat{\theta}) = O_p(n^{-1/4})$ for the standard nonparametric bootstrap. This fails the standard nonparametric bootstrap. More formal treatment of the invalidity of the standard nonparametric bootstrap is given in Appendix E.

It is not difficult to check $G(\hat{\theta}) = O_p(n^{-1/4})$ but not $o_p(n^{-1/4})$, which is directly implied by $n^{1/4}(\hat{\delta}_1 + \hat{\delta}_3 \hat{\gamma}) = O_p(1)$ but not $o_p(1)$ due to Theorem 2. Meanwhile, in our residual bootstrap method, $\delta_{10}^* + \delta_{30}^* \gamma_0^* = w_n(\hat{\delta}_1 + \hat{\delta}_3 \hat{\gamma}) + o_p(n^{-1/2}) = O_p(n^{-1/2})$ and $\delta_{20}^* = w_n \hat{\delta}_2 = O_p(n^{-3/4})$, which leads to $G(\theta_0^*) = O_p(n^{-1/2})$. The exact formula for $\delta_{10}^* + \delta_{30}^* \gamma_0^*$ is provided in the comment of Lemma D.5.

The $n^{1/4}$ rate of convergence for $w_n$ to zero is not the only possible choice. According to the proof of Theorem 6 in Appendix B, $(\delta_{10}^* + \delta_{30}^* \gamma_0^*) = O_p(n^{-1/2})$ is sufficient for the first-order asymptotic validity. This condition is explicitly stated in the conditions of Lemma D.5.

The idea of shrinking the first-order derivative in our bootstrap is closely related to other bootstrap methods developed for the case when asymptotic distributions of estimators are irregular. For example, Chatterjee and Lahiri (2011) propose a bootstrap method for the lasso
estimator, and Cavaliere et al. (2022) study bootstrap inference on the boundary of a parameter space. Both papers set up the model where the problem appears if the true parameter value is zero, and they obtain true parameters of bootstrap dgps by thresholding unrestricted estimators, i.e., $\theta^*_j = \hat{\theta}_j 1\{\hat{\theta}_j > c_n\}$, where $c_n$ converges to zero in a proper rate.

4.3 Bootstrap for testing continuity

The critical value for the continuity test introduced in Section 3.2 can also be obtained by bootstrapping. Recall that $\tilde{\theta} = \arg\min_{\theta \in \Theta: \delta_2 = 0_{p-1}, \delta_1 = -\delta_3, \gamma} \hat{Q}_n(\theta)$ is the continuity-restricted estimator. By setting $\theta_0^* = \tilde{\theta}$ in Algorithm 1, and collecting the bootstrap test statistic

$$T^*_n = n \left( \min_{\delta \in \delta_2 = 0_{p-1}, \delta_1 = -\delta_3, \gamma} \hat{Q}^*_n(\theta) - \min_{\theta \in \Theta} \hat{Q}^*_n(\theta) \right),$$

we can get the critical value using the empirical quantile of $T^*_n$. To run the bootstrap continuity test at size $\tau$, reject the continuity if $T_n > b_{F^*\gamma}^{-1}(1-\tau; T_n^*)$, where $b_{F^*\gamma}^{-1}(1-\tau; T_n^*)$ is the empirical 100(1 - $\tau$)th percentile of $T_n^*$. The consistency of the bootstrap is implied by Theorem 7 that follows.

**Theorem 7.** Assume that $T_n^*$ is obtained by Algorithm 1 with $\theta_0^* = \tilde{\theta}$.

(i) When the true model is continuous and Assumptions $G$, $D$, and $LK$ hold,

$$T_n^* \overset{d}{\to} V_1 - V_2 + V_3 \quad \text{in } P,$$

where the distributions of $V_1$, $V_2$, and $V_3$ are specified in Theorem 4.

(ii) When the model is discontinuous, then $T_n^* = O_p(1)$ in $P$.

**Theorem 7** (i) shows that the limit distribution of $T_n^*$, conditional on the data, is identical to that of $T_n$ under the null hypothesis. Moreover, **Theorem 7** (ii) says that $T_n^*$ is still stochastically bounded, conditionally on the data, when the true model is discontinuous. As $T_n$ is shown to be stochastically unbounded under the alternative, according to **Theorem 4** (ii), the bootstrap continuity test has power against the alternatives.

5 Monte Carlo results

This section executes Monte Carlo simulations to investigate finite sample performances of our bootstrap methods. The dgp is

$$y_{it} = \beta_2 y_{it-1} + \beta_3 q_{it} + (\delta_1 + \delta_2 y_{it-1} + \delta_3 q_{it})1\{q_{it} > \gamma\} + \sigma e_{it}$$

$$q_{it} = \rho q_{it-1} + u_{it}, \quad e_{it} \overset{iid}{\sim} N(0, 1), \quad u_{it} \overset{iid}{\sim} N(0, 1), \quad E[e_{it} u_{it+1}] = \rho_{cu},$$

with $\beta_2 = 0.6$, $\beta_3 = 1$, $\delta_2 = 0$, $\delta_3 = 2$, $\gamma = 0.25$, $\sigma = 0.5$, $\rho = 0.7$, and $\rho_{cu} = 0.5$. To investigate how coverage rates of the CIs change depending on the continuity, we try different values of
\( \delta_1 \in \{-0.5, -0.4, -0.3, 0, 0.5\} \), which implies different degrees of (dis)continuity \( \delta_1 + \delta_3 \gamma \in \{0, 0.1, 0.2, 0.5, 1\} \). If \( \delta_1 = -0.5 \), then the model is continuous. Otherwise, the model becomes discontinuous. As near continuous designs, we try \( \delta_1 + \delta_3 \gamma = 0.1, 0.2 \) and check if there is any poor performance of CIs. We generate samples of size \( n \in \{400, 800, 1600\} \) and \( T = 6 \). The number of repetitions for the Monte Carlo simulations is 2000. Instruments used for the estimations are the lagged dependent variables that date back from period \( t - 2 \) to period 1 and the lagged threshold variables from period \( t - 1 \) to period 1, i.e., \( z_{it} = (y_{i,t-2}, ..., y_{i1}, q_{i,t-1}, ..., q_{i1})' \). The earliest period used for the estimation is \( t_0 = 3 \), and the total number of the instruments becomes 24.

We begin with examining the finite sample coverage probabilities of the bootstrap CIs for the threshold location. Specifically, the grid bootstrap CI (Grid-B) is compared with both percentile nonparametric bootstrap CI (NP-B) and symmetric percentile nonparametric bootstrap CI (NP-B(S)). To construct \( 100(1 - \tau)\% \) CIs of the \( j \)th element of \( \theta \), the NP-B uses the empirical percentiles of the \( j \)th element of \( \sqrt{n}(\hat{\theta}^* - \bar{\theta}) \) while the NP-B(S) uses those of \( \sqrt{n}|\hat{\theta}^* - \bar{\theta}| \), similarly to the construction of CIs defined by (9) and (10). The number of bootstrap repetitions is set at 500 for each bootstrap method.

Table 1 reports the coverage rates of 95\% confidence intervals for the threshold location. First, it shows that the bootstrap CI by the NP-B is subject to severe undercoverage in all cases. Although the undercoverage is conspicuous for all cases, we can observe that the undercoverage is getting worse as the jump size decreases to zero. This phenomenon can be attributed to the invalidity of the NP-B when the true model is continuous. Meanwhile, the NP-B(S) exhibits extreme over-coverage in all cases. These results by the NP-B and NP-B(S) indicate that the distribution of the bootstrap estimator of the threshold \( \hat{\gamma}^* \) provides a bad approximation to that of the sample estimator \( \hat{\gamma} \), suggesting that \( (\hat{\gamma}^* - \hat{\gamma}) \) is not properly centered at zero for each realized sample.

Table 1: Coverage rates of 95\% CIs for the threshold location. Grid-B denotes the grid bootstrap CIs defined by (7). NP-B and NP-B(S) denote the percentile and the symmetric percentile CIs by the standard nonparametric bootstrap.

| \( \delta_1 + \delta_3 \gamma \) | 0 | 0.1 | 0.2 | 0.5 | 1 |
|---|---|---|---|---|---|
| n | 400 | 0.992 | 0.995 | 0.993 | 0.988 | 0.966 |
| Grid-B | 800 | 0.986 | 0.986 | 0.985 | 0.973 | 0.955 |
| | 1600 | 0.988 | 0.987 | 0.988 | 0.979 | 0.959 |
| NP-B | 400 | 0.484 | 0.491 | 0.494 | 0.524 | 0.631 |
| | 800 | 0.478 | 0.472 | 0.487 | 0.518 | 0.611 |
| | 1600 | 0.471 | 0.468 | 0.476 | 0.521 | 0.642 |
| NP-B(S) | 400 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 |
| | 800 | 1.000 | 1.000 | 1.000 | 0.999 | 0.994 |
| | 1600 | 1.000 | 1.000 | 1.000 | 1.000 | 0.994 |

On the other hand, the Grid-B provides much more reasonable coverage rates. It seems
that a larger jump yields coverage rates closer to nominal level as expected since it is easier to detect a bigger jump. As expected from the uniform validity of the Grid-B against near continuity, coverage rates do not deteriorate as the jump size becomes closer to zero, even though over-coverage occurs when the sample size or the degree of discontinuity are small.

Compared to the Grid-B, the NP-B(S) CIs exhibit higher coverage probability that is one or almost one for all cases. It indicates that the NP-B(S) CIs are overly wide and non-informative. To investigate this further, we examine some power properties as reported in Table 2 below. It shows that the NP-B(S) based tests for the threshold location are trivial for many parametrizations, specifically when the design is continuous or near-continuous or when the alternative is closer to the null. In contrast, the Grid-B tests are more powerful, oftentimes more than twice powerful than the NP-B tests. Here, we report the powers of the tests instead of the lengths of the bootstrap CIs as the Grid-B takes a long time to measure the exact length of a CI in Monte Carlo simulations.

Table 2: Rejection rates of tests on the alternative threshold locations $\gamma = \gamma_0 + c$ are reported. Grid-B denotes the grid bootstrap CIs defined by (7). NP-B(S) denotes the symmetric percentile CIs by the standard nonparametric bootstrap.

| c   | n   | 0    | 0.1  | 0.2  | 0.5  | 1    | 0    | 0.1  | 0.2  | 0.5  | 1    |
|-----|-----|------|------|------|------|------|------|------|------|------|------|
| 0.10| 400 | 0.015| 0.015| 0.015| 0.027| 0.096| 0.000| 0.000| 0.000| 0.004| 0.018|
|     | 800 | 0.011| 0.014| 0.015| 0.038| 0.112| 0.000| 0.000| 0.000| 0.004| 0.017|
|     | 1600| 0.017| 0.020| 0.021| 0.040| 0.125| 0.000| 0.000| 0.002| 0.004| 0.023|
| 0.25| 400 | 0.020| 0.030| 0.042| 0.100| 0.281| 0.002| 0.004| 0.009| 0.043| 0.135|
|     | 800 | 0.020| 0.034| 0.041| 0.112| 0.325| 0.002| 0.003| 0.007| 0.035| 0.154|
|     | 1600| 0.029| 0.034| 0.048| 0.126| 0.351| 0.002| 0.006| 0.007| 0.044| 0.152|
| 0.50| 400 | 0.102| 0.137| 0.172| 0.314| 0.581| 0.062| 0.109| 0.142| 0.274| 0.298|
|     | 800 | 0.114| 0.162| 0.207| 0.362| 0.632| 0.078| 0.117| 0.169| 0.310| 0.327|
|     | 1600| 0.136| 0.186| 0.240| 0.396| 0.652| 0.076| 0.124| 0.189| 0.332| 0.316|

Next, we turn to the coverage probabilities for the regression coefficients by different bootstrap CIs. Table 3 reports the coverage rates of different 95% CIs. We compare the residual bootstrap introduced in this paper with the standard nonparametric bootstrap. For the residual bootstrap, we construct two different CIs; the percentile bootstrap CI (R-B) defined by (9) and the symmetric percentile bootstrap CI (R-B(S)) defined by (10). $\hat{C}$ in (8) is set as the 50th percentile of the bootstrap distribution of the test statistic $T_n$ under the null hypothesis that the model is continuous, using the bootstrap method explained in Section 4.3.

Similarly to the threshold inference case, the CIs of the coefficients by the NP-B exhibit undercoverage while the CIs by the NP-B(S) are conservative. Although the R-B method yields higher coverage rates than the NP-B, these rates still fall short of the nominal 95% level. This indicates that bootstrap CIs based on empirical percentiles of bootstrap estimators may perform poorly for dynamic panel threshold models. The R-B(S) seems to provide coverage rates closer
to 95% while the NP-B(S) is conservative for most cases. To further investigate the differences
between the two bootstraps, Table 4 reports the average lengths of the CIs. When comparing
the percentile CIs, the R-B method results in CIs with a greater average length than the NP-
B. This finding aligns with our observation in Table 3, where the R-B method demonstrates
higher coverage rates than the NP-B. However, when comparing the percentile CIs, the R-B(S)
is slightly longer than for the NP-B(S) in most cases, despite the R-B(S) having lower coverage
rates. This discrepancy may be due to the bootstrap estimators in both standard nonparametric
bootstrap and our residual bootstrap not being well-centered.

6 Empirical example

Our empirical example examines a firm’s investment decision model that incorporates financial
constraints, as in Hansen (1999b) and Seo and Shin (2016). In a perfect financial market,
firms can borrow as much money as they need to finance their investment projects, regardless
of their financial conditions. Therefore, the financial conditions of firms are irrelevant to their
investment decisions. However, in an imperfect financial market, some firms may be restricted in
their access to external financing. These firms are said to be financially constrained. Financially
constrained firms are more sensitive to the availability of internal financing, as they cannot rely
on external financing to fund their investment projects.

Fazzari et al. (1988) argue that firms’ investments are positively related to their cash flow
if they are financially constrained, where those firms are identified by low dividend payments.
Hansen (1999b) applies the threshold panel regression more systematically to show that a more
positive relationship between investment and cash flow is present for firms with higher leverage.

Since there are multiple candidate measures of the financial constraint for the threshold
variable, we compare the following three dynamic panel threshold models:

\[ I_t = \eta_i + \xi'_{t-1} \beta + (\delta_1 + \delta_2 + \delta_3) \{ LEV_{it-1} > \gamma \} + \epsilon_{it} \]  
(12)
\[ I_t = \eta_i + \xi'_{t-1} \beta + (\delta_1 + \delta_2 + TQ_{it-1}) \{ TQ_{it-1} > \gamma \} + \epsilon_{it} \]  
(13)
\[ I_t = \eta_i + \xi'_{t-1} \beta + (\delta_1 + TQ_{it-1}) \{ TQ_{it-1} > \gamma \} + \epsilon_{it} \]  
(14)

where \( \xi_{t-1} = (I_{it-1}, CF_{it-1}, PPE_{it-1}, ROA_{it-2})' \). Here, \( I_t \) is investment, \( CF_{it} \) is cash flow,
\( PPE_{it} \) is property, plant and equipment, and \( ROA_{it} \) is return on assets. \( I_{it}, CF_{it} \), and \( PPE_{it} \)
are normalized by total assets. We have two candidate threshold variables, \( LEV_{it} \) and \( TQ_{it} \),
which are leverage and Tobin’s Q, respectively. Choice of the regressors and threshold variables
is based on previous works like Hansen (1999b) and Lang et al. (1996). Note that the regression
model (14) is nested within (13) and it is closer to a continuous threshold model.

Unlike the previous works, we do not need to assume either continuity or discontinuity for
valid inferences since the bootstrap methods in this paper are adaptive to each case. With
an assumption that the regressors are predetermined, we use the variables dated one period
before as instruments. Hence, the instruments include \( I_{t-2}, CF_{t-1}, PPE_{t-2}, ROA_{t-2} \) added
by \( LEV_{t-2} \) or \( TQ_{t-2} \) for each period.
Table 3: Coverage rates of 95% CIs for the coefficients are shown. R-B and R-B(S) denote the percentile and the symmetric percentile CIs by the residual bootstrap that sets $\theta_0^*$ as (8). NP-B and NP-B(S) denote the percentile and the symmetric percentile CIs by the standard nonparametric bootstrap.

| $\delta_1 + \delta_3 \gamma$ | n     | R-B     | NP-B     |
|-----------------------------|-------|---------|----------|
|                             | $\beta_2$ | $\beta_3$ | $\delta_1$ | $\delta_2$ | $\delta_3$ | $\beta_2$ | $\beta_3$ | $\delta_1$ | $\delta_2$ | $\delta_3$ |
| 0.0                         | 400   | 0.839  | 0.780  | 0.746  | 0.815  | 0.801  | 0.799  | 0.691  | 0.627  | 0.712  | 0.709  |
|                             | 800   | 0.837  | 0.790  | 0.721  | 0.807  | 0.806  | 0.790  | 0.723  | 0.607  | 0.725  | 0.716  |
|                             | 1600  | 0.849  | 0.782  | 0.727  | 0.840  | 0.835  | 0.833  | 0.709  | 0.602  | 0.754  | 0.718  |
| 0.1                         | 400   | 0.837  | 0.784  | 0.749  | 0.813  | 0.799  | 0.794  | 0.714  | 0.607  | 0.725  | 0.716  |
|                             | 800   | 0.830  | 0.779  | 0.724  | 0.803  | 0.800  | 0.786  | 0.714  | 0.599  | 0.720  | 0.710  |
|                             | 1600  | 0.853  | 0.787  | 0.727  | 0.840  | 0.829  | 0.827  | 0.700  | 0.598  | 0.760  | 0.719  |
| 0.2                         | 400   | 0.838  | 0.786  | 0.749  | 0.819  | 0.811  | 0.794  | 0.701  | 0.623  | 0.713  | 0.716  |
|                             | 800   | 0.833  | 0.776  | 0.720  | 0.803  | 0.794  | 0.784  | 0.707  | 0.585  | 0.718  | 0.712  |
|                             | 1600  | 0.855  | 0.789  | 0.728  | 0.846  | 0.832  | 0.830  | 0.707  | 0.606  | 0.764  | 0.722  |
| 0.5                         | 400   | 0.836  | 0.775  | 0.739  | 0.820  | 0.802  | 0.787  | 0.703  | 0.601  | 0.718  | 0.724  |
|                             | 800   | 0.841  | 0.789  | 0.732  | 0.815  | 0.807  | 0.787  | 0.714  | 0.602  | 0.716  | 0.727  |
|                             | 1600  | 0.843  | 0.799  | 0.728  | 0.826  | 0.834  | 0.815  | 0.717  | 0.595  | 0.753  | 0.737  |
| 1.0                         | 400   | 0.858  | 0.815  | 0.745  | 0.832  | 0.805  | 0.800  | 0.741  | 0.627  | 0.741  | 0.743  |
|                             | 800   | 0.858  | 0.827  | 0.749  | 0.846  | 0.820  | 0.808  | 0.731  | 0.620  | 0.741  | 0.738  |
|                             | 1600  | 0.863  | 0.846  | 0.759  | 0.830  | 0.837  | 0.820  | 0.738  | 0.622  | 0.761  | 0.747  |

|                     | R-B(S) | NP-B(S) |
|---------------------|--------|---------|
| 0.0                 | 400    | 0.964  | 0.976  | 0.980  | 0.974  | 0.930  | 0.996  | 0.996  | 0.996  | 0.992  | 0.982  |
|                     | 800    | 0.951  | 0.974  | 0.971  | 0.967  | 0.931  | 0.987  | 0.992  | 0.995  | 0.988  | 0.976  |
|                     | 1600   | 0.955  | 0.972  | 0.964  | 0.961  | 0.931  | 0.983  | 0.994  | 0.995  | 0.980  | 0.977  |
| 0.1                 | 400    | 0.964  | 0.976  | 0.979  | 0.974  | 0.933  | 0.994  | 0.993  | 0.995  | 0.991  | 0.982  |
|                     | 800    | 0.952  | 0.975  | 0.970  | 0.968  | 0.935  | 0.990  | 0.992  | 0.995  | 0.989  | 0.978  |
|                     | 1600   | 0.959  | 0.975  | 0.973  | 0.961  | 0.924  | 0.986  | 0.995  | 0.997  | 0.979  | 0.977  |
| 0.2                 | 400    | 0.963  | 0.974  | 0.978  | 0.977  | 0.939  | 0.995  | 0.993  | 0.997  | 0.993  | 0.986  |
|                     | 800    | 0.959  | 0.972  | 0.977  | 0.974  | 0.929  | 0.992  | 0.994  | 0.996  | 0.987  | 0.978  |
|                     | 1600   | 0.958  | 0.972  | 0.976  | 0.964  | 0.933  | 0.986  | 0.995  | 0.996  | 0.979  | 0.980  |
| 0.5                 | 400    | 0.964  | 0.971  | 0.982  | 0.978  | 0.940  | 0.992  | 0.994  | 0.998  | 0.994  | 0.989  |
|                     | 800    | 0.960  | 0.973  | 0.987  | 0.974  | 0.945  | 0.991  | 0.994  | 0.998  | 0.988  | 0.985  |
|                     | 1600   | 0.957  | 0.977  | 0.985  | 0.970  | 0.945  | 0.985  | 0.996  | 0.998  | 0.981  | 0.987  |
| 1.0                 | 400    | 0.970  | 0.982  | 0.985  | 0.984  | 0.967  | 0.991  | 0.995  | 0.992  | 0.991  | 0.993  |
|                     | 800    | 0.968  | 0.982  | 0.988  | 0.981  | 0.967  | 0.992  | 0.993  | 0.995  | 0.989  | 0.994  |
|                     | 1600   | 0.960  | 0.981  | 0.987  | 0.972  | 0.963  | 0.989  | 0.995  | 0.995  | 0.988  | 0.989  |
Table 4: The ratios of the average lengths of 95% CIs for the coefficients by different bootstrap methods are shown. R-B and R-B(S) denote the percentile and the symmetric percentile CIs by the residual bootstrap that sets $\theta^*_0$ as (8). NP-B and NP-B(S) denote the percentile and the symmetric percentile CIs by the standard nonparametric bootstrap. The R-B / NP-B columns show the ratios of the average lengths of CIs between the R-B and NP-B methods. Similarly, the R-B(S) / NP-B(S) columns show the ratios of the average lengths of CIs between the R-B(S) and NP-B(S) methods.

| $\delta_1 + \delta_3 \gamma$ | n  | $\beta_2$ | $\beta_3$ | $\delta_1$ | $\delta_2$ | $\delta_3$ | $\beta_2$ | $\beta_3$ | $\delta_1$ | $\delta_2$ | $\delta_3$ |
|-----------------------------|----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
|                             | 400| 1.076     | 1.091     | 1.099     | 1.074     | 1.046     | 1.017     | 1.035     | 1.008     | 0.996     | 1.010     |
| 0.0                         | 800| 1.081     | 1.086     | 1.093     | 1.070     | 1.046     | 1.033     | 1.037     | 1.007     | 1.004     | 1.018     |
|                             | 1600| 1.088    | 1.100     | 1.111     | 1.083     | 1.057     | 1.040     | 1.046     | 1.012     | 1.015     | 1.014     |
| 0.1                         | 400| 1.087     | 1.098     | 1.101     | 1.074     | 1.047     | 1.028     | 1.040     | 1.008     | 0.996     | 1.012     |
|                             | 800| 1.080     | 1.082     | 1.090     | 1.075     | 1.043     | 1.032     | 1.033     | 1.000     | 1.004     | 1.015     |
|                             | 1600| 1.086    | 1.102     | 1.111     | 1.077     | 1.057     | 1.039     | 1.047     | 1.011     | 1.020     | 1.016     |
| 0.2                         | 400| 1.080     | 1.088     | 1.097     | 1.074     | 1.047     | 1.022     | 1.035     | 1.003     | 0.996     | 1.012     |
|                             | 800| 1.079     | 1.089     | 1.094     | 1.075     | 1.047     | 1.032     | 1.039     | 1.001     | 1.004     | 1.015     |
|                             | 1600| 1.085    | 1.100     | 1.116     | 1.077     | 1.054     | 1.039     | 1.048     | 1.009     | 1.025     | 1.016     |
| 0.5                         | 400| 1.097     | 1.100     | 1.100     | 1.083     | 1.056     | 1.037     | 1.046     | 0.991     | 1.014     | 1.016     |
|                             | 800| 1.083     | 1.095     | 1.089     | 1.076     | 1.051     | 1.044     | 1.045     | 0.991     | 1.008     | 1.024     |
|                             | 1600| 1.098    | 1.110     | 1.098     | 1.089     | 1.059     | 1.052     | 1.056     | 0.996     | 1.035     | 1.022     |
| 1.0                         | 400| 1.164     | 1.159     | 1.084     | 1.114     | 1.074     | 1.101     | 1.107     | 0.989     | 1.042     | 1.042     |
|                             | 800| 1.158     | 1.159     | 1.079     | 1.109     | 1.076     | 1.096     | 1.111     | 0.988     | 1.039     | 1.052     |
|                             | 1600| 1.158    | 1.177     | 1.084     | 1.109     | 1.079     | 1.115     | 1.136     | 0.996     | 1.051     | 1.048     |
We construct a balanced panel of 1459 U.S. firms, excluding finance and utility firms, from 2010 to 2019 available in Compustat. To deal with extreme values, we drop firms if any of their non-threshold variables’ values fall within the top or bottom 0.5% tails. Moreover, we exclude firms whose Tobin’s Q is larger than 5 for more than 5 years when the threshold variable is Tobin’s Q, leaving 1222 firms in the sample. Meanwhile, Strebulaev and Yang (2013) claims that firms with large CEO ownership or CEO-friendly boards show persistent zero-leverage behavior. To prevent our threshold regression from capturing corporate governance characteristics rather than financial constraints, we exclude firms whose leverage is zero for more than half of the time periods when leverage is the threshold variable, leaving 1056 firms in the sample.

Table 5 reports the estimates and 95% CIs for (12) and (13), and Table 6 for (14). Figure 1 visualizes how the grid bootstrap CIs are obtained. The CIs of the coefficients are constructed by using symmetric percentiles obtained from the residual bootstrap, defined in (10). $\hat{C}$ for the percentile bootstrap is set at the 50th percentile of the bootstrap statistic for the continuity test, explained in Section 4.3. For the threshold locations, the CIs are obtained by the grid bootstrap with convexification. For the grid bootstrap, we make 500 bootstrap draws for each grid point. The grids of the threshold locations have 81 points from the 10th percentile to the 90th percentile of the threshold variables, and there are equal number of observations between two consecutive points. Table 5 and Table 6 also report the bootstrap p-values for the continuity and linearity tests by the bootstrap methods explained in Section 4.3 and Appendix G, respectively. The null hypothesis of the linearity test is $H_0: \delta = (0, ..., 0)^T$, which implies no threshold effects.

Figure 1: Subplots (a), (b), and (c) are for the models (12), (13), and (14), respectively. Black solid lines in each subplot denote the test statistics, red dashed lines denote the 5% size bootstrapped critical values, and horizontal blue arrows visualize the 95% CIs. The regions where the test statistics are below the bootstrapped critical values become the CIs of the threshold locations.

We find supporting evidence for the presence of the threshold effect when the threshold

\footnote{The percentile residual-bootstrap CIs that uses the 0.025th and 0.975th quantiles of $\sqrt{n}(\hat{\alpha}_j - \alpha_{j0})$ returns similar results. We report them in Appendix F.}
variable is Tobin’s Q, but the statistical evidence is not strong for the leverage threshold model. Table 5 and Table 6 report the bootstrap p-values at .135, .011, and .011, for specifications (12) - (14), respectively. The statistical evidence to reject the continuity is not trivial for all specifications and gets stronger when it is the restricted model using Tobin’s Q. The estimated bootstrap p-values are .028 and .004 for the unrestricted and the restricted using Tobin’s Q. Furthermore, the confidence interval for the threshold location is narrower for the restricted model (14) than for the unrestricted model (13).

Table 5: Columns (a) and (b) report results of the models (12) and (13), respectively. The percentile of each threshold location value is shown in parentheses below each value. The significance levels for the coefficients are given by stars: * - 10%, ** - 5% and *** - 1%.

|            | (a)         |            | (b)         |            |
|------------|-------------|------------|-------------|------------|
|            | est. [95% CI] |            | est. [95% CI] |            |
| **Lower regime** |             |            |             |            |
| \(I_{t-1}\)   | 0.778** | 0.319 | 1.237 | \(I_{t-1}\) | 0.252 | -0.242 | 0.746 |
| \(CF_{t-1}\) | 0.047 | -0.041 | 0.135 | \(CF_{t-1}\) | 0.266* | -0.004 | 0.535 |
| \(PPE_{t-1}\) | -0.147 | -0.428 | 0.134 | \(PPE_{t-1}\) | 0.027 | -0.175 | 0.229 |
| \(ROA_{t-1}\) | -0.032 | -0.128 | 0.065 | \(ROA_{t-1}\) | -0.017 | -0.157 | 0.123 |
| \(LEV_{t-1}\) | 0.231 | -1.219 | 1.682 | \(TQ_{t-1}\) | 0.246 | -0.071 | 0.564 |
| **Upper regime** |             |            |             |            |
| \(I_{t-1}\)   | -0.154 | -0.769 | 0.462 | \(I_{t-1}\) | 0.410** | 0.007 | 0.813 |
| \(CF_{t-1}\) | 0.148* | -0.026 | 0.322 | \(CF_{t-1}\) | 0.081* | -0.023 | 0.184 |
| \(PPE_{t-1}\) | -0.291** | -0.566 | -0.015 | \(PPE_{t-1}\) | 0.044 | -0.251 | 0.340 |
| \(ROA_{t-1}\) | 0.013 | -0.076 | 0.102 | \(ROA_{t-1}\) | 0.050 | -0.038 | 0.137 |
| \(LEV_{t-1}\) | -0.081 | -0.216 | 0.054 | \(TQ_{t-1}\) | 0.005 | -0.004 | 0.013 |
| **Difference between regimes** |             |            |             |            |
| intercept     | 0.068 | -0.045 | 0.181 | intercept | 0.236 | -0.083 | 0.554 |
| \(I_{t-1}\)   | -0.932** | -1.803 | -0.061 | \(I_{t-1}\) | 0.158 | -0.542 | 0.857 |
| \(CF_{t-1}\) | 0.101 | -0.117 | 0.319 | \(CF_{t-1}\) | -0.185 | -0.479 | 0.109 |
| \(PPE_{t-1}\) | -0.144 | -0.463 | 0.176 | \(PPE_{t-1}\) | 0.017 | -0.233 | 0.267 |
| \(ROA_{t-1}\) | 0.045 | -0.129 | 0.218 | \(ROA_{t-1}\) | 0.066 | -0.128 | 0.261 |
| \(LEV_{t-1}\) | -0.312 | -1.754 | 1.130 | \(TQ_{t-1}\) | -0.242 | -0.557 | 0.074 |
| **Threshold** |             |            |             |            |
| \(LEV_{t-1}\) | 0.172 | 0.101 | 0.265 | \(TQ_{t-1}\) | 1.298 | 1.169 | 1.386 |
| \(TQ_{t-1}\) | (38%) | (24%) | (58%) | \(TQ_{t-1}\) | (30%) | (21%) | (36%) |
| Testing (p-val) | 0.135 | 0.011 | 0.033 | 0.011 | 0.033 |

A notable finding concerning the coefficients estimates is that the relationship between cash flow and investment is positive and has larger magnitude for the low Tobin’s Q firms and the high leverage firms compared to their other respective regimes, although they are not statistically significant at 5% level but only at 10% level. Even though the sign and magnitude of the estimates align with the observations by Lang et al. (1996) and Hansen (1999b) that a firm is
subject to financial constraints when its Tobin’s Q is low or leverage is high, there is uncertainty in the interpretation of our results due to the lack of statistical significance.

Next, the autoregressive coefficients of the lagged investment are significant at 5% level in both the high Tobin’s Q regime and the low leverage regime. And they are larger than in the other respective regimes. This lends supporting evidence for the presence of asymmetric dynamics in investment, akin to the dynamics of leverage analyzed by Dang et al. (2012). In the meantime, we note that the autoregressive coefficients for the low and high leverage regimes in Column (a) are 0.778 and -0.154, respectively, which appear more extreme than findings of the literature where the estimates are between 0.1 and 0.5, e.g. Blundell et al. (1992). The autoregressive coefficients in the Column (b) are more in line with these estimates. Since the changes of the estimated coefficients in Column (b) are moderate, we also estimate the restricted model (14).

Turning to Table 6, we observe that the differences between the coefficients of the two regimes become significant at 5% level, and the CI of the threshold location becomes narrower while the estimate of the threshold location remains close to the estimate under the unrestricted model. The autoregressive coefficient of the lagged investment and the sensitivity of investment to both cash flow and return on assets are all positive and significant. The effect of Tobin’s Q is both positive and significant for both high and low Tobin’s Q regimes, but it almost disappears once it surpasses the threshold location. This suggests that low Tobin’s Q is related to low investment but higher Tobin’s Q does not cause higher investment once Tobin’s Q reaches some level. The linearity assumption is also rejected at 5% significance level.
7 Conclusion

This paper shows that the asymptotic distribution of the GMM estimator of the dynamic panel threshold models varies depending on whether the true models have a kink or a jump, and that the standard nonparametric bootstrap is inconsistent when the true model is the kink model. Therefore, alternative bootstrap methods are proposed for constructing confidence intervals of the threshold location and the coefficients, which are shown to be consistent independently to the continuity of the true model. Additionally, the grid bootstrap, which constructs confidence intervals of the threshold parameter, is demonstrated to be uniformly valid for the unknown presence of the kink or jump. The proposed bootstrap methods are shown to improve upon the standard nonparametric bootstrap in finite samples even when the model is discontinuous.

There are further extensions of our work that we leave as future research, such as asymptotic analysis when the time series dimension grows, along with other features like some latent group structure or the interactive effect, as explored by Miao et al. (2020b) and Miao et al. (2020a), respectively. Extension of these works to the dynamic threshold model would be valuable. It is also left as a future research to check the uniform validity of the bootstrap for the coefficients theoretically.
References

Adam, C. S. and Bevan, D. L. (2005). Fiscal deficits and growth in developing countries. *Journal of Public Economics*, 89:571–597.

Andrews, D. W., Cheng, X., and Guggenberger, P. (2020). Generic results for establishing the asymptotic size of confidence sets and tests. *Journal of Econometrics*, 218(2):496–531.

Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3):683–734.

Andrews, D. W. K. (2002). Generalized method of moments estimation when a parameter is on a boundary. *Journal of Business & Economic Statistics*, 20(4):530–544.

Andrews, D. W. K. and Cheng, X. (2012). Estimation and Inference With Weak, Semi-Strong, and Strong Identification. *Econometrica*, 80:2153–2211.

Andrews, D. W. K. and Cheng, X. (2014). GMM Estimation and Uniform Subvector Inference With Possible Identification Failure. *Econometric Theory*, 30:287–333.

Andrews, D. W. K. and Guggenberger, P. (2009). Hybrid and size-corrected subsampling methods. *Econometrica*, 77(3):721–762.

Andrews, D. W. K. and Guggenberger, P. (2019). Identification- and singularity-robust inference for moment condition models. *Quantitative Economics*, 10:1703–1746.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58:277–297.

Bick, A. (2010). Threshold effects of inflation on economic growth in developing countries. *Economics Letters*, 108(2):126–129.

Blundell, R., Bond, S., Devereux, M., and Schiantarelli, F. (1992). Investment and Tobin’s Q: Evidence from company panel data. *Journal of Econometrics*, 51:233–257.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Cavaliere, G., Nielsen, H. B., Pedersen, R. S., and Rahbek, A. (2022). Bootstrap inference on the boundary of the parameter space, with application to conditional volatility models. *Journal of Econometrics*, 227(1):241–263.

Cecchetti, S. G., Mohanty, M. S., and Zampolli, F. (2011). The Real Effects of Debt. *BIS Working Paper No. 352*.

Chan, K. S. and Tong, H. (1985). On the use of the deterministic lyapunov function for the ergodicity of stochastic difference equations. *Advances in applied probability*, 17(3):666–678.
Chan, K. S. and Tsay, R. S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika*, 85(2):413–426.

Chatterjee, A. and Lahiri, S. N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association*, 106(494):608–625.

Cheng, G. and Huang, J. Z. (2010). Bootstrap consistency for general semiparametric m-estimation. *The Annals of Statistics*, 38(5):2884–2915.

Chudik, A., Mohaddes, K., Pesaran, M. H., and Raissi, M. (2017). Is There a Debt-Threshold Effect on Output Growth? *The Review of Economics and Statistics*, 99:135–150.

Dang, V. A., Kim, M., and Shin, Y. (2012). Asymmetric capital structure adjustments: New evidence from dynamic panel threshold models. *Journal of Empirical Finance*, 19:465–482.

Dovonon, P. and Goncalves, S. (2017). Bootstrapping the GMM overidentification test under first-order underidentification. *Journal of Econometrics*, 201:43–71.

Dovonon, P. and Hall, A. R. (2018). The asymptotic properties of gmm and indirect inference under second-order identification. *Journal of Econometrics*, 205(1):76–111.

Dovonon, P. and Renault, E. (2013). Testing for Common Conditionally Heteroskedastic Factors. *Econometrica*, 81:2561–2586.

Fazzari, S. M., Hubbard, R. G., Petersen, B. C., Blinder, A. S., and Poterba, J. M. (1988). Financing Constraints and Corporate Investment. *Brookings Papers on Economic Activity*, 1988:141–206.

Giannerini, S., Goracci, G., and Rahbek, A. (2024). The validity of bootstrap testing for threshold autoregression. *Journal of Econometrics*, 239(1):105379.

Gine, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *The Annals of Probability*, 18(2):851 – 869.

Girma, S. (2005). Absorptive Capacity and Productivity Spillovers from FDI: A Threshold Regression Analysis. *Oxford Bulletin of Economics and Statistics*, 67:281–306.

Goncalves, S. and White, H. (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics*, 119(1):199–219.

Gonzalo, J. and Wolf, M. (2005). Subsampling inference in threshold autoregressive models. *Journal of Econometrics*, 127(2):201–224.

Hall, P. and Horowitz, J. L. (1996). Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators. *Econometrica*, 64:891–916.

Han, S. and McCloskey, A. (2019). Estimation and Inference with a (nearly) Singular Jacobian. *Quantitative Economics*, 10:1019–1068.
Hansen, B. E. (1999a). The Grid Bootstrap and the Autoregressive Model. *The Review of Economics and Statistics*, 81:594–607.

Hansen, B. E. (1999b). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics*, 93:345–368.

Hansen, B. E. (2000). Sample Splitting and Threshold Estimation. *Econometrica*, 68:575–603.

Hansen, B. E. (2017). Regression kink with an unknown threshold. *Journal of Business & Economic Statistics*, 35(2):228–240.

Hidalgo, J., Lee, H., Lee, J., and Seo, M. H. (2023). Minimax risk in estimating kink threshold and testing continuity. In *Advances in Econometrics: Essays in Honor of Joon Y. Park: Econometric Theory, Vol. 45A*, pages 233–259.

Hidalgo, J., Lee, J., and Seo, M. H. (2019). Robust Inference for Threshold Regression Models. *Journal of Econometrics*, 210:291–309.

Khan, M. S. and Senhadji, A. S. (2001). Threshold Effects in the Relationship between Inflation and Growth. *IMF Staff Papers*, 48:1–21.

Kremer, S., Bick, A., and Nautz, D. (2013). Inflation and growth: new evidence from a dynamic panel threshold analysis. *Empirical Economics*, 44:861–878.

Lang, L., Ofek, E., and Stulz, R. (1996). Leverage, investment, and firm growth. *Journal of Financial Economics*, 40(1):3–29.

Lee, S., Seo, M. H., and Shin, Y. (2011). Testing for Threshold Effects in Regression Models. *Journal of the American Statistical Association*, 106:220–231.

Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: facts and fiction. *Econometric Theory*, 21(1):21–59.

Miao, K., Li, K., and Su, L. (2020a). Panel threshold models with interactive fixed effects. *Journal of Econometrics*, 219(1):137–170.

Miao, K., Su, L., and Wang, W. (2020b). Panel threshold regressions with latent group structures. *Journal of Econometrics*, 214(2):451–481.

Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica*, 75(5):1411–1452.

Newey, W. K. and McFadden, D. (1994). Chapter 36 Large Sample Estimation and Hypothesis Testing. In *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier.

Newey, W. K. and West, K. D. (1987). Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review*, 28:777–787.
Pakes, A. and Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, 57:1027–1057.

Praestgaard, J. and Wellner, J. A. (1993). Exchangeably Weighted Bootstraps of the General Empirical Process. *The Annals of Probability*, 21(4):2053 – 2086.

Romano, J. P. and Shaikh, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *The Annals of Statistics*, 40(6):2798 – 2822.

Rousseau, P. L. and Wachtel, P. (2002). Inflation thresholds and the finance–growth nexus. *Journal of International Money and Finance*, 21:777–793.

Seo, M. H., Kim, S., and Kim, Y. J. (2019). Estimation of Dynamic Panel Threshold Model Using Stata. *The Stata Journal*, 19:685–697.

Seo, M. H. and Shin, Y. (2016). Dynamic Panels With Threshold Effect and Endogeneity. *Journal of Econometrics*, 195:169–186.

Strebulaev, I. A. and Yang, B. (2013). The mystery of zero-leverage firms. *Journal of Financial Economics*, 109(1):1–23.

van der Vaart, A. W. and Wellner, J. (1996). *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer Series in Statistics. Springer-Verlag, New York.

Wang, Q. (2015). Fixed-effect panel threshold model using stata. *The Stata Journal*, 15(1):121–134.

Yang, L., Zhang, C., Lee, C., and Chen, I.-P. (2020). Panel kink threshold regression model with a covariate-dependent threshold. *The Econometrics Journal*, 24(3):462–481.

Zhang, Y., Zhou, Q., and Jiang, L. (2017). Panel kink regression with an unknown threshold. *Economics Letters*, 157:116–121.
Appendices

Additional Notations. For $k, p \in \mathbb{N}$, $0_{k \times p}$ denotes $k \times p$ a matrix whose elements are all zero. "\( \sim \)" denotes the weak convergence as in section 1.3 of van der Vaart and Wellner (1996). \( \| \cdot \| \) is a norm for either vectors or matrices. For a vector, it is the Euclidean norm. For a matrix, it is the Frobenius norm, i.e., \( \| M \| = \sqrt{\text{tr}(M' M)} \) for a matrix $M$.

A Proofs for Section 3.

A.1 Proof of Theorem 1.

Note that $E[z_{it}(\Delta y_{it} - \Delta x_{it}' \beta - 1_{it}(\gamma)'X_{it}\delta)] = -E[z_{it}\Delta x_{it}'(\beta - \beta_0) - E[z_{it}1_{it}(\gamma)'X_{it}]\delta + E[z_{it}1_{it}(\gamma)'X_{it}\delta_0]$ due to $\Delta y_{it} = \Delta x_{it}' \beta_0 + 1_{it}(\gamma)'X_{it}\delta_0 + \Delta \epsilon_{it}$. Hence, the population moment equation is $g_0(\theta) = M_{10}(\beta - \beta_0) + M_{20}(\gamma)\delta - M_{20}\delta_0 = \begin{bmatrix} M_0(\gamma) & M_{20}\delta_0 \end{bmatrix} \times ((\beta' - \beta_0', \delta'), -1)'$, when $\gamma \neq \gamma_0$. The condition (ii) of Theorem 1 implies that $\begin{bmatrix} M_0(\gamma) & M_{20}\delta_0 \end{bmatrix}$ has full column rank, and hence $g_0(\theta) \neq 0_k$ if $\gamma \neq \gamma_0$. $g_0(\theta) = M_0 \times (\alpha - \alpha_0)$, when $\gamma = \gamma_0$. The condition (i) of Theorem 1 implies that $M_0 \times (\alpha - \alpha_0)$ is not zero if $\alpha \neq \alpha_0$. Therefore, $g_0(\theta) \neq 0_k$ if $\theta \neq \theta_0$, and $g_0(\theta) = 0_k$ if $\theta = \theta_0$, which is the standard identification condition in the literature, e.g. Section 2.2.3 in Newey and McFadden (1994).

A.2 Proof of Theorem 2.

To obtain limit distribution of $\hat{\theta}$, we first establish consistency of $\hat{\theta}$ to $\theta_0$ and rate of $\hat{\theta}$’s convergence. Then, we show asymptotic distribution of the estimates using rescaled versions of the parameters and criterions.

A.2.1 Consistency.

Constrained estimator of the coefficients, $\hat{\alpha}(\gamma) = \arg \min_{\alpha \in A} \hat{Q}_n(\alpha, \gamma)$, given a fixed $\gamma$ can be expressed as

$$\hat{\alpha}(\gamma) = -(\hat{M}_n(\gamma)'W_n\hat{M}_n(\gamma))^{-1}\hat{M}_n(\gamma)'W_n\hat{v}_n$$

where

$$\hat{v}_n = -\hat{M}_n\alpha_0 + u_n, \quad u_n = \frac{1}{n} \sum_{i=1}^{n} \begin{bmatrix} z_{it_0}\Delta \epsilon_{it_0} \\ \vdots \\ z_{iT}\Delta \epsilon_{iT} \end{bmatrix}.$$ 

Therefore,

$$\hat{\alpha}(\gamma) = -(\hat{M}_n(\gamma)'W_n\hat{M}_n(\gamma))^{-1}\hat{M}_n(\gamma)'W_n(-\hat{M}_n\alpha_0 + u_n).$$

Define profiled criterion with respect to $\gamma$ by $\tilde{g}_n(\gamma) = \tilde{g}_n(\hat{\alpha}(\gamma), \gamma)$ and $\tilde{Q}_n(\gamma) = \tilde{g}_n(\gamma)'W_n\tilde{g}_n(\gamma)$. By the law of large numbers (LLN), $u_n \xrightarrow{P} 0$. By the uniform law of large numbers (ULLN) in Lemma C.2, $\hat{M}_n(\gamma) \xrightarrow{P} M_0(\gamma)$
uniformly with respect to $\gamma \in \Gamma$. Hence, $\hat{\gamma} \overset{p}{\to} \gamma_0$ would imply $\hat{M}_n(\hat{\gamma}) \overset{p}{\to} M_0$, and then $\hat{\alpha}(\hat{\gamma}) \overset{p}{\to} \alpha_0$, which completes the proof.

To show consistency of $\hat{\gamma}$ to $\gamma_0$, we apply the argmin/argmax continuous mapping theorem (CMT) as in Theorem 3.2.2 in van der Vaart and Wellner (1996). It is sufficient to check (i) $\hat{Q}_n(\gamma)$ uniformly converges to some function $\hat{Q}_0(\gamma)$ in probability, and (ii) $\hat{Q}_0(\gamma_0) < \inf_{\gamma \notin \mathcal{O}} \hat{Q}_0(\gamma)$ for any open set $\mathcal{O}$. (i) can be shown if $\hat{Q}_0(\gamma)$ is uniquely minimized at $\gamma_0$ and continuous as $\Gamma$ is compact.

The profiled moment can be rewritten as

$$\hat{g}_n(\gamma) = [I - \hat{M}_n(\gamma) (M_n(\gamma)'W_nM_n(\gamma))^{-1} M_n(\gamma)'W_n] (\hat{M}_n\alpha_0 + u_n).$$

Therefore,

$$W_n^{1/2}\hat{g}_n(\gamma) = [I - P_{W_n^{1/2}\hat{M}_n(\gamma)}] (-W_n^{1/2}\hat{M}_n\alpha_0 + W_n^{1/2}u_n),$$

where $P_{W_n^{1/2}\hat{M}_n(\gamma)} = W_n^{1/2}\hat{M}_n(\gamma) (M_n(\gamma)'W_nM_n(\gamma))^{-1} M_n(\gamma)'W_n^{1/2}$ is a projection matrix to the column space of $W_n^{1/2}\hat{M}_n(\gamma)$. The profiled objective can be written as

$$\hat{Q}_n(\gamma) = \| (I - P_{W_n^{1/2}\hat{M}_n(\gamma)}) (-W_n^{1/2}\hat{M}_n\alpha_0 + W_n^{1/2}u_n) \|^2.$$

By $W_n \overset{p}{\to} W$, $u_n \overset{p}{\to} 0$, and $\sup_{\gamma \in \Gamma} \| \hat{M}_n(\gamma) - M_0(\gamma) \| \overset{p}{\to} 0$, we can derive that

$$\hat{Q}_n(\gamma) \overset{p}{\to} \hat{Q}_0(\gamma) = \| (I - P_{W^{1/2}M_0(\gamma)}) W^{1/2}M_0\alpha_0 \|^2$$

uniformly with respect to $\gamma$, where $P_{W^{1/2}M_0(\gamma)} = W^{1/2}M_0(\gamma) (M_0(\gamma)'WM_0(\gamma))^{-1} M_0(\gamma)'W^{1/2}$.

Note that $W = \Omega^{-1}$ in the second stage of the two-step GMM estimation. $W = I$ when we consider the first stage. $\hat{Q}_0(\gamma)$ is uniquely minimized when $\gamma = \gamma_0$. This is because $W$ is positive definite, and the conditions in Theorem 1 implies that $M_0\alpha_0$ does not lie in the column space of $M_0(\gamma)$ whenever $\gamma \neq \gamma_0$. Moreover, $\hat{Q}_0(\gamma)$ is continuous as $M_0(\gamma)$ is continuous with respect to $\gamma$ by Assumption D.

A.2.2 Convergence rate.

$\| W_n - \Omega^{-1} \| \overset{p}{\to} 0$ as the consistency of $\hat{\theta}(1)$ is shown. Our proof follows arguments similar to the proof of Theorem 3.3 by Pakes and Pollard (1989). By the consistency of $\hat{\theta}$ and by Lemma C.3,

$$\sqrt{n} \| \hat{g}_n(\hat{\theta}) - \hat{g}_n(\theta_0) - g_0(\hat{\theta}) \| = o_p(1).$$

By $\| W_n - \Omega^{-1} \| \overset{p}{\to} 0$, we can obtain

$$\sqrt{n} \| W_n^{1/2}\hat{g}_n(\hat{\theta}) - W_n^{1/2}\hat{g}_n(\theta_0) - \Omega^{-1/2}g_0(\hat{\theta}) \| = o_p(1).$$
Apply triangle inequality to get
\[
\sqrt{n}||\Omega^{-1/2}g_0(\hat{\theta})|| \leq o_p(1) + \sqrt{n}||W_n^{-1/2}\hat{g}_n(\theta_0)|| + \sqrt{n}||W_n^{-1/2}\hat{g}_n(\theta)||.
\]

As \(\hat{\theta}\) is the minimizer of the GMM criterion, \(\sqrt{n}||W_n^{-1/2}\hat{g}_n(\hat{\theta})|| \leq o_p(1) + \sqrt{n}||W_n^{-1/2}\hat{g}_n(\theta_0)|| = O_p(1)\). Therefore,
\[
\sqrt{n}||\Omega^{-1/2}g_0(\hat{\theta})|| \leq O_p(1).
\]

\[
\sqrt{n}||\Omega^{-1/2}g_0(\hat{\theta})|| \geq \sqrt{n}||\Omega^{-1/2}D_2(\hat{\alpha}' - \alpha_0', (\hat{\gamma} - \gamma_0)^2)' - \sqrt{n}||\Omega^{-1/2}(g(\hat{\theta}) - D_2(\hat{\alpha}' - \alpha_0', (\hat{\gamma} - \gamma_0)^2)'||,
\]
while \(\sqrt{n}||\Omega^{-1/2}(g(\hat{\theta}) - D_2(\hat{\alpha}' - \alpha_0', (\hat{\gamma} - \gamma_0)^2)'|| \leq o_p(1 + \sqrt{n}||\hat{\alpha}' - \alpha_0', (\hat{\gamma} - \gamma_0)^2)'||)\) by Lemma C.1. Thus,
\[
\sqrt{n}||\hat{\alpha} - \alpha_0|| + (\hat{\gamma} - \gamma_0)^2 \leq O_p(1)
\]
which implies \(||\hat{\alpha} - \alpha_0|| = O_p(n^{-1/2})\) and \((\hat{\gamma} - \gamma_0)^2 = O_p(n^{-1/2})\).

### A.2.3 Asymptotic distribution.

This section derives asymptotic distribution of the estimator through the argmin/argmax continuous mapping theorem (CMT) as in Theorem 3.2.2 in van der Vaart and Wellner (1996).

Introduce a local reparametrization by \(a = \sqrt{n}(\alpha - \alpha_0)\) and \(b = n^{1/2}(\gamma - \gamma_0)\), and let \(\alpha\) consist of subvectors \(a_1 = \sqrt{n}(\beta - \beta_0)\) and \(a_2 = \sqrt{n}(\delta - \delta_0)\). Additionally, define \(\hat{\alpha} = \sqrt{n}(\hat{\alpha} - \alpha_0)\) and \(\hat{b} = n^{1/2}(\hat{\gamma} - \gamma_0)\). Note that \((\hat{\alpha}, \hat{b})^2\) is uniformly tight due to the convergence rate we obtained. Let
\[
S_n(a, b) = n\hat{Q}_n(\alpha_0 + \frac{a}{\sqrt{n}}, \gamma_0 + \frac{b}{n^{1/2}}) = n\hat{g}_n(\alpha_0 + \frac{a}{\sqrt{n}}, \gamma_0 + \frac{b}{n^{1/2}})W_n\hat{g}_n(\alpha_0 + \frac{a}{\sqrt{n}}, \gamma_0 + \frac{b}{n^{1/2}}).
\]
We show that (i) \(S_n\) weakly converges to a stochastic process \(S\) in \(\ell^\infty(\mathbb{K})\) for every compact \(K\) in the Euclidean space, (ii) \(S\) is continuous, and (iii) \(\hat{S}\) possesses an unique optimum not in \(B\) but in its square \(\hat{B}^2\) since \(S(a, b) = S(a, -b)\). Thus, we will establish that \((\hat{a}', \hat{b}')^2\) converges in distribution to \((a'_{0}, b'_{0})^2\) and \((a'_{0}, b')^2\) is shown to be tight.

---

Footnotes:

3A random variable \(X\) is tight if for any \(\epsilon > 0\), there exists a compact set \(K\) such that \(P(X \in K) > 1 - \epsilon\), and \(X_n\) is uniformly tight if for any \(\epsilon > 0\), there exists a compact set \(K\) such that \(P(X_n \in K) > 1 - \epsilon\) for all \(n \in \mathbb{N}\). Note that by the convergence rate we derived, for any \(\epsilon > 0\), there exists a compact set \(K_0\) such that \(\lim_{n \to \infty} P((\sqrt{n}(\hat{\alpha} - \alpha_0)', \sqrt{n}(\hat{\gamma} - \gamma_0)^2) \in K_0) > 1 - \epsilon/2\), and \(N < \infty\) such that \(P((\sqrt{n}(\hat{\alpha} - \alpha_0)', \sqrt{n}(\hat{\gamma} - \gamma_0)^2) \in K_0) > 1 - \epsilon\) if \(n \geq \sigma\). Then, we can define a compact set \(K = (\bigcup_{i=1}^{N-1}K_i) \cup K_0\), where \(K_j\) is a compact set such that \(P((\sqrt{n}(\hat{\alpha} - \alpha_0)', \sqrt{n}(\hat{\gamma} - \gamma_0)^2) \in K_j) > 1 - \epsilon\), which satisfies \(P((\sqrt{n}(\hat{\alpha} - \alpha_0)', \sqrt{n}(\hat{\gamma} - \gamma_0)^2) \in K) > 1 - \epsilon\) for all \(n \in \mathbb{N}\).
The rescaled and reparametrized sample moment can be written as
\[
\sqrt{n} g_n(a_0 + \frac{a}{\sqrt{n}}, \gamma_0 + \frac{b}{n^{\frac{1}{4}}}) = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} z_{i0} \Delta \epsilon_{i0} \right) - \left( \frac{1}{n} \sum_{i=1}^{n} z_{iT} \Delta \epsilon_{iT} \right) a_1 - \left( \frac{1}{n} \sum_{i=1}^{n} z_{i0} X_{i0} \right) a_2 + \left( \frac{1}{n} \sum_{i=1}^{n} z_{iT} X_{iT} \right) \delta_0.
\]

By the central limit theorem (CLT),
\[
\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} z_{i0} \Delta \epsilon_{i0} \right) \xrightarrow{d} \delta \sim N(0, \Omega).
\]

By the LLN,
\[
\left( \frac{1}{n} \sum_{i=1}^{n} z_{iT} \Delta \epsilon_{iT} \right) \xrightarrow{P} \left( E z_{i0} \Delta \epsilon_{i0} \right)
\]

Let \( K < \infty \) be arbitrary. By the ULLN in Lemma C.2,
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} z_{i0} 1_{i0} (\gamma_0 + \frac{b}{n^{\frac{1}{4}}})' X_{i0} \right\| \xrightarrow{P} 0
\]
uniformly with respect to \( b \in [-K, K] \). Then, by continuity of \( \kappa \mapsto E[z_{i0} 1_{i0}(\gamma + \kappa) X_{i0}] \) at \( \kappa = 0 \),
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} z_{i0} 1_{i0} (\gamma_0 + \frac{b}{n^{\frac{1}{4}}})' X_{i0} \right\| \xrightarrow{P} 0
\]
uniformly with respect to \( b \in [-K, K] \). By Lemma C.4,
\[
\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{i0} (1_{i0} (\gamma_0)' - 1_{i0} (\gamma_0 + \frac{b}{n^{\frac{1}{4}}})' X_{i0} \delta_0 \right) \xrightarrow{P} \left( \frac{1}{E_{i0} 1_{i0} (\gamma_0) f_{i0} (\gamma_0) - E_{i0} 1_{i0} (\gamma_0) f_{i0-1} (\gamma_0)} \right) b^2
\]
uniformly with respect to \( b \in [-K, K] \).
Therefore, $S_n(a, b)$ weakly converges to

$$S(a, b) = (M_0a + Hb^2 - e)'\Omega^{-1}(M_0a + Hb^2 - e),$$

in $\ell^\infty(\mathbb{K})$ for any compact $\mathbb{K} \subset \mathbb{R}^{2p+2}$. Then, by the CMT,

$$\left(\hat{a}, \hat{b}^2\right) \xrightarrow{d} \arg \min_{a, b^2} (M_0a + Hb^2 - e)'\Omega^{-1}(M_0a + Hb^2 - e).$$

**Characterization of the minimizers** Next, we characterize the minimizers. The objective function of the minimization problem is strictly convex with respect to $a$ and $b^2$, since $\begin{bmatrix} M_0 & H \end{bmatrix}$ has full column rank and $\Omega^{-1}$ is positive definite. Hence, a solution $(a'_0, b_0^2)'$ can be characterized by the Karush-Kuhn-Tucker (KKT) conditions. See Chapter 5 in Boyd and Vandenberghe (2004) for more details.

The Lagrangian for this problem is

$$\mathcal{L}(a, b, \lambda) = a'M_0\Omega^{-1}M_0a + 2a'M_0\Omega^{-1}Hb^2 + H'\Omega^{-1}Hb^2 - 2a'M_0\Omega^{-1}e - 2H'\Omega^{-1}e.b^2 + e'\Omega^{-1}e - \lambda b^2$$

and the gradient of the Lagrangian with respect to $a$ and $b^2$ should vanish:

\begin{align*}
a & : \quad M_0'\Omega^{-1}M_0a + M_0'\Omega^{-1}Hb^2 - M_0'\Omega^{-1}e = 0 \\
b^2 & : \quad H'\Omega^{-1}Hb^2 + H'\Omega^{-1}M_0a - H'\Omega^{-1}e - \lambda = 0.
\end{align*}

In addition, $\lambda \geq 0$ and $\lambda b^2 = 0$ should hold.

(i) When $\lambda = 0$ and $b^2 \geq 0$, we can obtain

$$b^2 = (H'\Omega^{-1/2}(I - P_{\Omega^{-1/2}M_0})\Omega^{-1/2}H)^{-1}H'\Omega^{-1/2}(I - P_{\Omega^{-1/2}M_0})\Omega^{-1/2}e,$$

where $P_{\Omega^{-1/2}M_0} = \Omega^{-1/2}M_0(M_0'\Omega^{-1}M_0)^{-1}M_0'\Omega^{-1/2}$ is the projection matrix to the column space of $\Omega^{-1/2}M_0$. $H'\Omega^{-1/2}(I - P_{\Omega^{-1/2}M_0})\Omega^{-1/2}H > 0$ because the matrix $\begin{bmatrix} M_0 & H \end{bmatrix}$ has full column rank, and $\Omega^{-1/2}H$ cannot be in the column space of $\Omega^{-1/2}M_0$ and $(I - P_{\Omega^{-1/2}M_0})\Omega^{-1/2}H \neq 0$. Therefore,

$$H'\Omega^{-1/2}(I - P_{\Omega^{-1/2}M_0})\Omega^{-1/2}e \geq 0$$

should hold for the feasibility condition $b^2 \geq 0$.

(ii) When $\lambda > 0$ and $b^2 = 0$, we can obtain

$$a = (M_0'\Omega^{-1}M_0)^{-1}M_0'\Omega^{-1}e.$$

By plugging this into the equation for $b^2$, we get

$$H'\Omega^{-1/2}(I - P_{\Omega^{-1/2}M_0})\Omega^{-1/2}e < 0.$$
Thus,
\[
b_0^2 = \begin{cases} 
[H'\Xi H]^{-1} H' \Xi e & \text{if } H' \Xi e \geq 0 \\
0 & \text{else}
\end{cases}
\]
where \( \Xi = \Omega^{-1/2}(I - P_{\Omega^{-1/2} M_0}) \Omega^{-1/2} \). \( b_0^2 \) follows a normal distribution that is left censored at 0. Then,
\[
a_0 = \begin{cases} 
(M_0'\Omega^{-1} M_0)^{-1} M_0' \Omega^{-1} [I - H[H' \Xi H]^{-1} H' \Xi e] & \text{if } H' \Xi e \geq 0 \\
(M_0'\Omega^{-1} M_0)^{-1} M_0' \Omega^{-1} e & \text{else}.
\end{cases}
\]
Note that the two normal variables \((M_0'\Omega^{-1} M_0)^{-1} M_0' \Omega^{-1} [I - H[H' \Xi H]^{-1} H' \Xi e] \) and \((M_0'\Omega^{-1} M_0)^{-1} M_0' \Omega^{-1} e \) are independent of each other, because \( E[H' \Xi ee' \Omega^{-1} M_0] = H' \Omega^{-1/2}(I - P_{\Omega^{-1/2} M_0}) \Omega^{-1/2} M_0 \) becomes zero.

B Proofs for Section 4

B.1 Preliminaries

The bootstrap methods we consider are Algorithm 1 with different choices of \( \theta_0^* \). There are three bootstrap methods this paper propose: (i) \( \theta_0^* = (\hat{\alpha}(\gamma), \gamma)' \) for \( \gamma \in \Gamma \), (ii) \( \theta_0^* \) set as (8), and (iii) \( \theta_0^* = \hat{\theta} \) which is the continuity-restricted estimator. In Appendix E, we consider the case \( \theta_0^* = \hat{\theta} \) which results in the standard nonparametric bootstrap.

The probability law for the bootstrap is formalized following Goncalves and White (2004). Let \( P \) be the probability measure for data and \( P^* \) be the conditional probability law of bootstrap given observations. \( Z_n^* \xrightarrow{P} 0 \) in \( P \) (\( Z_n^* = a_p^*(1) \) in \( P \)) if for any \( \epsilon, \delta > 0, P(P^*(|Z_n^*| > \epsilon) > \delta) \rightarrow 0 \) as \( n \rightarrow \infty \). \( Z_n^* = O_p^*(1) \) in \( P \) if for any \( \epsilon > 0 \) and \( \delta > 0 \), there exists \( M < \infty \) such that \( \limsup_n P(P^*(|Z_n^*| > M) > \delta) < \epsilon \). \( Z_n^* \xrightarrow{d} Z \) in \( P \) if \( E^* f(Z_n^*) \rightarrow E f(Z) \) in \( P \) for every continuous and bounded function \( f \), where \( E^* \) is the expectation by the bootstrap probability law conditional on observations. \( Z_n^* \xrightarrow{\mathcal{L}} Z \) in \( \mathcal{L}^\infty(\mathbb{K}) \) in \( P \) if \( \sup_{f \in BL_1} |E^* f(Z_n^*) - E f(Z_n)| \xrightarrow{P} 0 \), where \( BL_1 \) is the set of all Lipschitz functions on \( \mathcal{L}^\infty(\mathbb{K}) \) bounded in \([0,1] \) such that \( |f(z_1) - f(z_2)| \leq |z_1 - z_2|_{\mathcal{L}^\infty(\mathbb{K})} = \sup_{x \in \mathbb{K}} |z_1(x) - z_2(x)| \).

The following lemma is useful in analyzing bootstrap stochastic orders.

Lemma B.1.  
(i) If \( A_n = o_p^*(1) \) or \( O_p^*(1) \), then \( A_n = a_p^*(1) \) or \( O_p^*(1) \) in \( P \), respectively.

(ii) Let \( Z_n^* = a_p^*(1) \) in \( P \) and \( W_n^* = O_p^*(1) \) in \( P \). Then, \( Z_n^* \times W_n^* = a_p^*(1) \) in \( P \).

Proof. See Lemma 3 in Cheng and Huang (2010). \( \square \)

Recall that \( W_n^* = \{[\frac{1}{n} \sum_{i=1}^{n} g_i^*(\hat{\theta}_i^*) g_i^*(\hat{\theta}_i^*)'] -[\frac{1}{n} \sum_{i=1}^{n} g_i^*(\hat{\theta}_i^*)]\}^{-1} [\frac{1}{n} \sum_{i=1}^{n} g_i^*(\hat{\theta}_i^*)\}'\}^{-1} \). ||W_n^* - \Omega^{-1}|| = a_p^*(1) in \( P \) when \( \hat{\theta}_i^* \xrightarrow{P} \theta_0 \) in \( P \). This would be the case when \( \hat{\theta}_i^* \xrightarrow{P} 0 \) in \( P \) and \( \|\theta_0^* - \theta_0\| = o_p(1) \) since then \( \|\hat{\theta}_i^* - \theta_0\| \leq \|\theta_0\| + \|\theta_0 - \theta_0\| = o_p(1) \) in \( P \) by Lemma B.1.
B.2 Proof of Theorem 6.

As in the proof of Theorem 2, consistency and convergence rates of the bootstrap estimator should be derived first. These results are summarized in the following proposition, with the proof provided in Online Appendix D.

**Proposition 1.** (i) Under the assumptions of the case (i) in Theorems 5, 6, or 7,
\[
\sqrt{n}(\hat{\alpha}^* - \alpha_0^*) = O_p^*(1) \text{ in } P, \quad \text{and} \quad \sqrt{n}(\hat{\gamma}^* - \gamma_0^*)^2 = O_p^*(1) \text{ in } P.
\]

(ii) Under the assumptions of the case (ii) in Theorems 5 or 6,
\[
\sqrt{n}(\hat{\alpha}^* - \alpha_0^*) = O_p^*(1) \text{ in } P, \quad \text{and} \quad \sqrt{n}(\hat{\gamma}^* - \gamma_0^*) = O_p^*(1) \text{ in } P.
\]

Then, we derive the (conditional) weak convergence limit of the rescaled criterion and apply the CMT to obtain the asymptotic distribution of the bootstrap estimator.

**Asymptotic distribution under continuity.** Based on the convergence rate in Proposition 1, introduce the local reparametrization by \( a = \sqrt{n}(\alpha - \alpha_0^*) \) and \( b = n^{\frac{1}{4}}(\gamma - \gamma_0^*) \), and let \( \alpha \) consist of subvectors \( a_1 = \sqrt{n}(\beta - \beta_0^*) \) and \( a_2 = \sqrt{n}(\delta - \delta_0^*) \).

The asymptotic distributions of the bootstrap estimators can be derived by using the argmin/argmax CMT as in the proof of Theorem 2. Let
\[
S_n^*(a, b) = nQ_n^*(a_0^* + \frac{a}{\sqrt{n}}, \gamma_0^* + \frac{b}{n^{\frac{1}{4}}}) = n\tilde{g}_n(a_0^* + \frac{a}{\sqrt{n}}, \gamma_0^* + \frac{b}{n^{\frac{1}{4}}})W_n^*g_n(a_0^* + \frac{a}{\sqrt{n}}, \gamma_0^* + \frac{b}{n^{\frac{1}{4}}}).
\]

We show that \( S_n^* \xrightarrow{a.s.} S \) in \( \ell^\infty(\mathbb{K}) \) in \( P \) for every compact \( \mathbb{K} \) in the Euclidean space. Recall that \( S(a, b) = (M_0a + Hb^2 - e)'\Omega^{-1}(M_0a + Hb^2 - e) \).

The rescaled and reparametrized bootstrap moment can be written as

\[
\sqrt{n}\tilde{g}_n(a_0^* + \frac{a}{\sqrt{n}}, \gamma_0^* + \frac{b}{n^{\frac{1}{4}}}) = \sqrt{n}\left( \left( \begin{array}{c}
\frac{1}{n} \sum_{i=1}^n z_{i0}^* \tilde{\Delta}_{i0} \\
\frac{1}{n} \sum_{i=1}^n z_{iT}^* \tilde{\Delta}_{iT}
\end{array} \right) - \left( \begin{array}{c}
\frac{1}{n} \sum_{i=1}^n z_{i0}^* \Delta_{i0} \\
\frac{1}{n} \sum_{i=1}^n z_{iT}^* \Delta_{iT}
\end{array} \right) \right) + \frac{1}{n} \sum_{i=1}^n z_{i0}^*X_{i0}^* a_1
\]

\[
- \left( \begin{array}{c}
\frac{1}{n} \sum_{i=1}^n z_{iT}^*X_{iT}^*
\end{array} \right) a_2 + \sqrt{n}\left( \begin{array}{c}
\frac{1}{n} \sum_{i=1}^n z_{i0}^* (1_{i0}^* (\gamma_0^*)' - 1_{i0} (\gamma_0^* + \frac{b}{n^{\frac{1}{4}}})')X_{i0}^* \\
\frac{1}{n} \sum_{i=1}^n z_{iT}^* (1_{iT}^* (\gamma_0^*)' - 1_{iT} (\gamma_0^* + \frac{b}{n^{\frac{1}{4}}})')X_{iT}^*
\end{array} \right) \delta_0^*.
\]
By Lemma D.2,
\[
\sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^{n} z_{i0}^* \Delta \epsilon_{i0}^* \right\} - \left\{ \frac{1}{n} \sum_{i=1}^{n} z_{i0} \Delta \epsilon_{i0} \right\} \xrightarrow{d} -e \sim N(0, \Omega) \quad \text{in } P.
\]

By the bootstrap LLN,
\[
\left( \frac{1}{n} \sum_{i=1}^{n} z_{i0}^* \Delta x_{i0}^* \right) \xrightarrow{P} \left( E z_{i0}^* \Delta x_{i0}^* \right) \quad \text{in } P.
\]

Let \( K < \infty \) be arbitrary. By bootstrap Glivenko-Cantelli, e.g. Lemma 3.6.16 in van der Vaart and Wellner (1996),
\[
\sup_{b: |b| \leq K, \gamma \in \Gamma} \left\| \left( \frac{1}{n} \sum_{i=1}^{n} z_{i0}^* 1_{i0}^* (\gamma + \frac{b}{n^4})' X_{i0}^* \right) \right\|_{\infty} - \left( \frac{1}{n} \sum_{i=1}^{n} z_{i}^* 1_{iT}^* (\gamma + \frac{b}{n^4})' X_{iT}^* \right) \right\|_{\infty} \xrightarrow{P} 0 \quad \text{in } P.
\]

By continuity of \( J(\gamma) := E[z_{d1} \gamma X_{d}] \) at \( \gamma = \gamma_0 \), for any \( c > 0 \), there exists \( h > 0 \) such that \( \| J(\gamma) - J(\gamma_0) \| < c \) if \( |\gamma - \gamma_0| < h \). For any \( h > 0 \), \( P(|\gamma_0 - \gamma_0^* - \frac{b}{n^4}| > h) \to 0 \). Note that \( \{ \| \frac{1}{n} \sum_{i=1}^{n} z_{i0}^* 1_{i0}^* (\gamma_0^* + \frac{b}{n^4})' X_{i0}^* - J(\gamma_0) \| > 2c \} \cup \{ J(\gamma_0^* + \frac{b}{n^4}) - J(\gamma_0) \| > c \} \) and hence \( P^*(\| \frac{1}{n} \sum_{i=1}^{n} z_{i0}^* 1_{i0}^* (\gamma_0^* + \frac{b}{n^4})' X_{i0}^* - J(\gamma_0) \| > 2c) \leq P^*(\| \frac{1}{n} \sum_{i=1}^{n} z_{i0}^* 1_{i0}^* (\gamma_0^* + \frac{b}{n^4})' X_{i0}^* - J(\gamma_0) \| > c) \) with probability approaching 1, while \( \frac{1}{n} \sum_{i=1}^{n} z_{i0}^* 1_{i0}^* (\gamma_0^* + \frac{b}{n^4})' X_{i0}^* - J(\gamma_0) \| > c \) \( \xrightarrow{P} 0 \) uniformly with respect to \( b \in [-K, K] \). Thus,
\[
\left( \frac{1}{n} \sum_{i=1}^{n} z_{i0}^* 1_{i0}^* (\gamma_0^* + \frac{b}{n^4})' X_{i0}^* \right) \xrightarrow{P} \left( E z_{i0}^* 1_{i0}^* (\gamma_0^*)' X_{i0}^* \right) \quad \text{in } P,
\]
both uniformly with respect to \( b \in [-K, K] \). By Lemma D.5,
\[
\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{i0}^* (1_{i0}^* (\gamma_0^* - 1_{i0} (\gamma_0^* + \frac{b}{n^4})') X_{i0}^*) \right) \xrightarrow{P} \frac{\delta_0^*}{2} \begin{pmatrix} E z_{i0}^* [z_{i0}^* | \gamma_0] f_{i0} (\gamma_0) - E z_{i0}^* [z_{i0}^* | \gamma_0] f_{i0-1} (\gamma_0) \\ E z_{iT}^* [z_{iT}^* | \gamma_0] f_{iT} (\gamma_0) - E z_{iT}^* [z_{iT}^* | \gamma_0] f_{iT-1} (\gamma_0) \end{pmatrix} \quad \text{in } P
\]
uniformly with respect to \( b \in [-K, K] \).
Therefore, $S_n^*(a,b) \sim S(a,b)$ in $\ell^\infty(\mathbb{K})$ in $P$ for any compact $\mathbb{K} \subset \mathbb{R}^{2p+2}$. Then, by applying the argmin CMT as in the proof of Theorem 2, we can obtain the limit distribution of the bootstrap estimates conditional on the data.

**Asymptotic distribution under discontinuity.** The proof for the discontinuous model only requires a slight change to the proof for the continuous model. As the convergence rate for the discontinuous model is $\sqrt{n}$ for both coefficients and threshold location estimators, let $a$ be unchanged and $b = \sqrt{n}(\gamma - \gamma_0^*)$ for the local reparametrization. Let

$$S_n^*(a,b) = n\hat{Q}_n^*(\alpha_0^* + \frac{a}{\sqrt{n}}, \gamma_0^* + \frac{b}{\sqrt{n}}) = n\tilde{g}_n^*(\alpha_0^* + \frac{a}{\sqrt{n}}, \gamma_0^* + \frac{b}{\sqrt{n}})/W_n\tilde{g}_n^*(\alpha_0^* + \frac{a}{\sqrt{n}}, \gamma_0^* + \frac{b}{\sqrt{n}}).$$

We can write the rescaled and reparametrized moment as follows:

$$\sqrt{n}\tilde{g}_n^*(\alpha_0^* + \frac{a}{\sqrt{n}}, \gamma_0^* + \frac{b}{\sqrt{n}}) = \sqrt{n}\left\{ \begin{array}{l}
\left( \frac{1}{n} \sum_{i=1}^{n} z_{i0}^* \Delta \epsilon_{i0}^* \right) \\
\left( \frac{1}{n} \sum_{i=1}^{n} z_{iT}^* \Delta \epsilon_{iT}^* \right)
\end{array} \right\} - \left( \begin{array}{l}
\left( \frac{1}{n} \sum_{i=1}^{n} z_{i0}^* (1_{iT}^*(\gamma^*_0) - 1_{ iT}(\gamma^*_0 + \frac{b}{\sqrt{n}})')X_{i0}^* \right) \\
\left( \frac{1}{n} \sum_{i=1}^{n} z_{iT}^* (1_{iT}^*(\gamma^*_0) - 1_{ iT}(\gamma^*_0 + \frac{b}{\sqrt{n}})')X_{iT}^* \right)
\end{array} \right) \delta_0^*.$$  

The limit of $\sqrt{n}\tilde{g}_n^*(\alpha_0^* + \frac{a}{\sqrt{n}}, \gamma_0^* + \frac{b}{\sqrt{n}})$ can be obtained similarly to the continuous model case, except that we use Lemma D.6 instead of Lemma D.5 to get

$$\sqrt{n}\left( \frac{1}{n} \sum_{i=1}^{n} z_{i0}^* (1_{iT}^*(\gamma^*_0) - 1_{iT}(\gamma^*_0 + \frac{b}{\sqrt{n}})')X_{i0}^* \right) \delta_0^*.$$  

uniformly with respect to $b \in [-K,K].$

Then, $S_n^*(a,b)$ conditionally weakly converges to $S_J(a,b) = (M_0a + Gb - e)'\Omega^{-1}(M_0a + Gb - e)$ in $\ell^\infty(\mathbb{K})$ in $P$ for any compact $\mathbb{K} \subset \mathbb{R}^{2p+2}$. And the argmin CMT yields the asymptotic distribution of the bootstrap estimators. The limit distributions of the bootstrap estimators are normal because $(a'_0,b'_0) = \arg\min_{a,b} S_J(a,b) = (D'_1\Omega^{-1}D_1)^{-1}D'_1\Omega^{-1}e.$
Online Supplements for “Bootstraps for Dynamic Panel Threshold Models” (Not for Publication)

Woosik Gong and Myung Hwan Seo

This part of the appendix is only for online supplements. It contains the remaining proofs for Theorem 3, Theorem 4, Proposition 1, Theorem 5, Theorem 7 and additional Lemma’s with proofs. It also presents invalidity of the standard nonparametric bootstrap, percentile bootstrap confidence intervals for empirical application, bootstrap for linearity test, and the uniform validity of the grid bootstrap.

Additional notations  We introduce additional notations as lemmas in this online appendix involve more empirical process theory. Suppose that \((X, A)\) is a measurable space and \(\omega_1, \omega_2, \ldots\) are i.i.d. random elements in \((X, A)\) with probability law \(P\). For a point \(\omega \in X\), let \(\delta_\omega\) be a dirac measure at \(\omega\). The empirical measure of a sample \(\omega_1, \ldots, \omega_n\) is \(P_n = \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i}\), and the empirical process is \(G_n = \sqrt{n}(P_n - P)\). Let \(F\) be a functional class, elements of which are measurable functions from \(X\) to \(\mathbb{R}\). We call a function \(F: X \rightarrow \mathbb{R}\) an envelope of \(F\) if \(|f| \leq F\) for all \(f \in F\). For a stochastic process \(G\) and a functional class \(F\), define \(\|G\|_F := \sup_{f \in F} |Gf|\).

C Proofs of Theorems in Section 3 and Auxiliary Lemmas

C.1 Proof of Theorem 3.

C.1.1 Continuous Model. When \(\gamma = \gamma_0\). Note that the constrained estimator \(\hat{\alpha}(\gamma_0) = \arg \min_{\alpha \in A} \hat{Q}_n(\alpha, \gamma_0)\) is \(\sqrt{n}\)-consistent to \(\alpha_0\), which is identical to the convergence rate of \(\hat{\alpha}\), since the problem becomes a standard linear dynamic panel estimation. Let \(a = \sqrt{n}(\alpha - \alpha_0)\) and \(b = n^{1/4}(\gamma - \gamma_0)\). The distance test statistic can be rewritten as follows:

\[
P_n(\gamma_0) = \inf_a S_n(a, 0) - \inf_{a,b} S_n(a, b) + o_p(1)
\]

\[
\stackrel{d}{=} \inf_a S(a, 0) - \inf_{a,b} S(a, b)
\]

\[
= \min_a (M_0a - e)' \Omega^{-1}(M_0a - e) - \min_{a,b} (M_0a + Hb^2 - e)' \Omega^{-1}(M_0a + Hb^2 - e),
\]

where we apply the CMT. Lee et al. (2011) showed that the difference between the constrained and unconstrained infima is a continuous operator on \(\ell^\infty(\mathbb{K})\).

\[\text{footnote}{\text{4Although we already use } \delta \text{ as the subvector of the parameter } \theta = (\beta', \delta', \gamma)', \text{ we still use } \delta \text{ to represent dirac measure as it is strong convention in the literature. We explicitly mention if } \delta \text{ is used as dirac measure to avoid confusion.}}\]
Note that \( \min_{a,b^2}(M_0a - e)'\Omega^{-1}(M_0a - e) = e'(\Omega^{-1} - \Omega^{-1}M_0(M_0'\Omega^{-1}M_0)^{-1}M_0'\Omega^{-1})e \), while

\[
\min_{a,b^2}(M_0a + Hb^2 - e)'\Omega^{-1}(M_0a + Hb^2 - e) = (M_0a_0 + Hb_0^2 - e)'\Omega^{-1}(M_0a_0 + Hb_0^2 - e)
\]

\[
= (M_0'\Omega^{-1}M_0a_0 + M_0'\Omega^{-1}Hb_0^2)'(M_0'\Omega^{-1}M_0)^{-1}(M_0'\Omega^{-1}M_0a_0 + M_0'\Omega^{-1}Hb_0^2)
\]

\[
+ b_0^2H'(I - P_{\Omega^{-1/2}M_0})\Omega^{-1/2}Hb_0^2 - 2e'(I - P_{\Omega^{-1/2}M_0})\Omega^{-1/2}Hb_0^2 - 2e'(M_0'\Omega^{-1}M_0a_0 + M_0'\Omega^{-1}Hb_0^2)
\]

\[
- 2e'(I - P_{\Omega^{-1/2}M_0})\Omega^{-1/2}Hb_0^2 + e'\Omega^{-1}e,
\]

where \((a_0, b_0^2)\) is the argmin, whose formula is derived in the proof of Theorem 2. By plugging in one of the first order conditions, \(M_0'\Omega^{-1}M_0a_0 + M_0'\Omega^{-1}Hb_0^2 = M_0'\Omega^{-1}e\), and the formula for \(b_0\), we can get

\[
\min_{a,b^2}(M_0a + Hb^2 - e)'\Omega^{-1}(M_0a + Hb^2 - e)
\]

\[
= \begin{cases}
-e'(I - P_{\Omega^{-1/2}M_0})\Omega^{-1/2}Hb_0^2 - 2e'(M_0'\Omega^{-1}M_0a_0 + M_0'\Omega^{-1}Hb_0^2)
\quad \text{if } H'\Omega e \geq 0 \\
-e'(I - P_{\Omega^{-1/2}M_0})\Omega^{-1/2}Hb_0^2 + e'\Omega^{-1}e
\quad \text{else.}
\end{cases}
\]

Therefore, the limit distribution of the test statistic is identical to

\[
\begin{cases}
e'(H'\Omega H)^{-1}H'\Omega e & \text{if } H'\Omega e \geq 0 \\
0 & \text{else.}
\end{cases}
\]

Note that \(e'(H'\Omega H)^{-1}H'\Omega e \sim \chi_1^2\) as \(H'\Omega e \sim N(0, H'\Omega \Omega H)\), and \(H'\Omega \Omega H = H'\Omega H\).

**When** \(\gamma \neq \gamma_0\). We show that \(\mathcal{D}_n(\gamma)\) diverges to infinity in probability. There is a constant \(C_1 \in (0, +\infty)\) such that \(\inf_{a \in A} \|g_0(a, \gamma)\| \geq C_1\). This is because \(g_0(\theta)\) is zero if and only if \(\theta = \theta_0\), by Assumption G and Theorem 1, and continuous on \(\Theta\), by Assumption D, while the restricted parameter set \(\{\theta = (\beta', \delta', \gamma)' : \gamma = c\}\) is closed for all \(c \in \Gamma\). \(\mathcal{G} = \{g(\omega, \theta) : \theta \in \Theta\}\) is shown to satisfy the uniform entropy condition in the proof of Lemma C.3, and hence \(\sup_{\theta \in \Theta} \|g_n(\theta) - g_0(\theta)\| = o_p(1)\) by Glivenko-Cantelli theorem. By triangle inequality, \(C_1 \leq \|g_0(\alpha(\cdot), \gamma)\| \leq \|\hat{g}_n(\alpha(\cdot), \gamma)\| + o_p(1)\). Meanwhile, \(\|\hat{g}_n(\theta)\| = O_p(n^{-1/2})\) because \(\|\hat{g}_n(\theta)\| \leq \|\hat{g}_n(\theta)\| = O_p(n^{-1/2})\). Therefore, there exists \(C_2 \in (0, +\infty)\) such that \(\hat{Q}_n(\alpha(\cdot), \gamma) - \hat{Q}_n(\theta) \geq C_2 + O_p(n^{-1})\), which implies that \(P(\mathcal{D}_n(\gamma) > M) = P(\hat{Q}_n(\alpha(\cdot), \gamma) - \hat{Q}_n(\theta) > M/n) \rightarrow 1\) for any \(M < \infty\).

**C.1.2 Discontinuous Model.**

**When** \(\gamma = \gamma_0\). As in the proof for the continuous model, we apply the CMT to the test statistic. Let \(a = \sqrt{n}(\alpha - \alpha_0)\) and \(b = \sqrt{n}(\gamma - \gamma_0)\). First, we will show that when the model is discontinuous and Assumptions G, D, and LJ are true, \(S_n(a, b) \sim S_f(a, b) = (M_0a + Gb - S_f(a, b))\).
\[ e'\Omega^{-1}(M_0a + Gb - e) \text{ in } \ell^\infty(\mathbb{K}) \text{ for any compact } \mathbb{K} \subset \mathbb{R}^{2p+2}. \] Note that

\[
\sqrt{n} g_n'(a_0 + \frac{a}{\sqrt{n}}, \gamma_0 + \frac{b}{\sqrt{n}}) = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} z_{it_0} \Delta \epsilon_{it_0} \right) - \left( \frac{1}{n} \sum_{i=1}^{n} z_{it} \Delta \epsilon_{iT} \right) a_1 \tag{C.1}
\]

\[
- \left( \frac{1}{n} \sum_{i=1}^{n} z_{it_0} \epsilon_{it_0} (\gamma_0 + \frac{b}{\sqrt{n}})^' X_{it_0} \right) + \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} z_{it_0} (1 - \epsilon_{it_0} (\gamma_0 + \frac{b}{\sqrt{n}})^') X_{it_0} \right) \delta_0. \tag{C.2}
\]

The terms in the first two lines of the right hand side (C.1) and (C.2) converge in distribution to \((M_0a - e)\) uniformly with respect to \(b \in [-K, K]\). Since \(\sup_{|b| \leq K} \sqrt{n} ||g_n'(a_0, \gamma_0 + \frac{b}{\sqrt{n}}) - g_n(a_0, \gamma_0)| - g_0(a_0, \gamma_0, \gamma)| = o_p(1)\) by Lemma C.3,

\[
\sqrt{n} \left[ E[z_{it_0} (1 - \epsilon_{it_0} (\gamma_0 + \frac{b}{\sqrt{n}})^') X_{it_0} \delta_0] \right] - \left[ E[z_{it} (1 - \epsilon_{iT} (\gamma_0 + \frac{b}{\sqrt{n}})^') X_{iT} \delta_0] \right]
\]

converges in probability to zero uniformly with respect to \(b \in [-K, K]\). Suppose \(b > 0\). The result for \(b < 0\) is similar. By application of Taylor expansion,

\[
\sqrt{n} E[z_{it}(1, x_{it}') \delta_0 | \gamma_0 + \frac{b}{\sqrt{n}} \geq q_{it} \geq \gamma_0] \rightarrow E_t[z_{it}(1, x_{it}')] \delta_0 | \gamma_0 | f_t(\gamma_0)b
\]

uniformly with respect to \(b \in [-K, K]\), and similar limit result can be derived for

\[
\sqrt{n} E[z_{it}(1, x_{it-1}') \delta_0 | \gamma_0 + \frac{b}{\sqrt{n}} \geq q_{it-1} \geq \gamma_0].
\]

Hence, we can derive that the term (C.3) converges in probability to \(Gb\) uniformly with respect to \(b \in [-K, K]\).

By the CMT, the test statistic converges in distribution to

\[
\min_{\theta, b}(M_0a - e)' \Omega^{-1}(M_0a - e) - \min_{\theta, b}(M_0a + Gb - e)' \Omega^{-1}(M_0a + Gb - e).
\]

Note that \(\min_{\theta, b}(M_0a - e)' \Omega^{-1}(M_0a - e) = e'(\Omega^{-1} - \Omega^{-1} M_0(M_0' \Omega^{-1} M_0)^{-1} M_0' \Omega^{-1})e\), and \(\min_{\theta, b}(M_0a + Gb - e)' \Omega^{-1}(M_0a + Gb - e) = e'(\Omega^{-1} - \Omega^{-1} D_1(D_1' \Omega^{-1} D_1)^{-1} D_1' \Omega^{-1})e\). Therefore, the limit distribution of the test statistic is identical to the distribution of

\[
e'(\Omega^{-1/2} D_1(D_1' \Omega^{-1} D_1)^{-1} D_1' \Omega^{-1/2} - \Omega^{-1/2} M_0(M_0' \Omega^{-1} M_0)^{-1} M_0' \Omega^{-1/2}) \Omega^{-1/2} e.
\]

The matrix \(\Omega^{-1/2} D_1(D_1' \Omega^{-1} D_1)^{-1} D_1' \Omega^{-1/2} - \Omega^{-1/2} M_0(M_0' \Omega^{-1} M_0)^{-1} M_0' \Omega^{-1/2}\) is idempotent since the column space of \(\Omega^{-1/2} M_0\) lies in the column space of \(\Omega^{-1/2} D_1\). The rank of the
In the proof of Theorem 2, it is shown that $\sqrt{n} \tilde{g}_n \sim N(0, I)$, the chi-square distribution with 1 degree of freedom is the limit distribution.

**When $\gamma \neq \gamma_0$.** The proof showing that $D_n(\gamma)$ diverges when $\gamma \neq \gamma_0$ for the discontinuous model is identical to the proof written for the continuous model.

### C.2 Proof of Theorem 4.

**Under the null hypothesis.** Define a map $T$ such that $T(\psi) = (\beta', -\gamma \delta_3, 0, ..., 0, \delta_3, \gamma)' \in \mathbb{R}^{2p+2}$ if $\psi = (\beta', \delta_3, \gamma)' \in \mathbb{R}^{p+2}$. Let $\psi_0 = (\beta_0', \delta_3, \gamma_0)'$. Note that

$$g_i(T(\psi)) = \begin{pmatrix}
    z_{i0} \{ \Delta y_{i0} - \Delta x_{it0}^{(0)} - [(q_{it0} - \gamma)1 \{ q_{it0} > \gamma \} - (q_{it0-1} - \gamma)1 \{ q_{it0-1} > \gamma \}] \delta_3 \}
    \\
    \vdots
    \\
    z_{it} \{ \Delta y_{iT} - \Delta x_{it}^{(0)} - [(q_{iT} - \gamma)1 \{ q_{iT} > \gamma \} - (q_{iT-1} - \gamma)1 \{ q_{iT-1} > \gamma \}] \delta_3 \}
\end{pmatrix}.$$  

The first-order derivative of $g_0(T(\psi))$ with respect to $\psi$ is

$$D_\psi =
\begin{pmatrix}
    -z_{i0} \Delta x_{it0}^{(0)} & -z_{i0} [(q_{it0} - \gamma)1 \{ q_{it0} > \gamma \} - (q_{it0-1} - \gamma)1 \{ q_{it0-1} > \gamma \}] \delta_3 \\
    \vdots & \vdots \\
    -z_{iT} \Delta x_{iT}^{(0)} & -z_{iT} [(q_{iT} - \gamma)1 \{ q_{iT} > \gamma \} - (q_{iT-1} - \gamma)1 \{ q_{iT-1} > \gamma \}] \delta_3 
\end{pmatrix}.$$  

$D_\psi$ is a matrix that is identical to a binding of the columns of $M_{10}$ and $N_{20}$. If $\hat{\psi} = \arg \min_\psi \hat{Q}_n(T(\psi))$, then $\sqrt{n} (\hat{\psi} - \psi_0) \xrightarrow{d} N(0, (D_\psi^T \Omega D_\psi)^{-1})$ (see Seo et al. (2019)). The continuity test statistic $T_n = n (\hat{Q}_n(\hat{\theta}) - \hat{Q}_n(\hat{\theta}))$ can be rewritten as

$$n(\hat{Q}_n(T(\hat{\psi})) - \hat{Q}_n(\hat{\theta})) = n \left( \min_{(\theta', \psi') : \theta = \theta_0} (\hat{Q}_n(T(\psi)) - \hat{Q}_n(\theta)) - \min_{(\theta', \psi') : \psi = \psi_0} (\hat{Q}_n(T(\psi)) - \hat{Q}_n(\theta)) \right).$$

Reparametrize such that $a = \sqrt{n}(\alpha - \alpha_0)$, $b = n^{1/4}(\gamma - \gamma_0)$, and $r = \sqrt{n}(\psi - \psi_0)$. Define a centered criterion by

$$M_n(a, b, r) = n(\hat{Q}_n(T(\psi_0 + \frac{r}{\sqrt{n}})) - \hat{Q}_n(\alpha_0 + \frac{a}{\sqrt{n}}, \gamma_0 + \frac{b}{n^{1/4}})).$$

We will show that $M_n$ weakly converges to a process $M$ in $\ell^\infty(\mathbb{K})$ for every compact $\mathbb{K} \subset \mathbb{R}^{3p+4}$. Then, by the CMT, the continuity test statistic converges in distribution to

$$\min_{(a', b', r') : (a', b') = (a, b)} M(a, b, r) - \min_{(a', b', r') : r = 0} (-M(a, b, r)).$$

In the proof of Theorem 2, it is shown that $\sqrt{n} \tilde{g}_n (a_0 + \frac{a}{\sqrt{n}}, \gamma_0 + \frac{b}{n^{1/4}}) \sim (M_0 a + H b^2 - e)$ and

$$n\hat{Q}_n(\alpha_0 + \frac{a}{\sqrt{n}}, \gamma_0 + \frac{b}{n^{1/4}}) \sim (M_0 a + H b^2 - e)^\Omega^{-1}(M_0 a + H b^2 - e).$$
Let \( r_1 = \sqrt{n}(\beta - \beta_0) \), \( r_2 = (r_{21}, r_{22})' \), \( r_{21} = \sqrt{n}(\delta_3 - \delta_{30}) \), and \( r_{22} = \sqrt{n}(\gamma - \gamma_0) \). Then,

\[
\sqrt{n} \hat{g}_n(T(\psi_0 + \frac{r}{\sqrt{n}})) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it_0} \Delta \epsilon_{it_0} \right) - \left( \frac{1}{n} \sum_{i=1}^{n} z_{it_0} \Delta x_{it_0} \right) r_1
\]

\[
\sqrt{n} \hat{g}_n(T(\psi_0 + \frac{r}{\sqrt{n}})) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{iT} \Delta \epsilon_{iT} \right) - \left( \frac{1}{n} \sum_{i=1}^{n} z_{iT} \Delta x_{iT} \right) r_{21}
\]

By the CLT and LLN,

\[
\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it_0} \Delta \epsilon_{it_0} \right) - \left( \frac{1}{n} \sum_{i=1}^{n} z_{it_0} \Delta x_{it_0} \right) \xrightarrow{d} (M_{10} r_1 - \epsilon).
\]

By the ULLN (application of Lemma C.2) and continuity of \( \kappa \mapsto E[\hat{z}_{it}(1, \psi_t)1\{\psi_t > \gamma_0 + \kappa\}] \) and \( \kappa \mapsto E[\hat{z}_{it}(1, \psi_{t-1})1\{\psi_{t-1} > \gamma_0 + \kappa\}] \) at \( \kappa = 0 \),

\[
\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it_0} [(q_{it_0} - \gamma_0 - \frac{r_{22}}{\sqrt{n}})1\{q_{it_0} > \gamma_0 + \frac{r_{22}}{\sqrt{n}}\} - (q_{it_0-1} - \gamma_0 - \frac{r_{22}}{\sqrt{n}})1\{q_{it_0-1} > \gamma_0 + \frac{r_{22}}{\sqrt{n}}\}] \right) \]

\[
\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{iT} [(q_{iT} - \gamma_0 - \frac{r_{22}}{\sqrt{n}})1\{q_{iT} > \gamma_0 + \frac{r_{22}}{\sqrt{n}}\} - (q_{iT-1} - \gamma_0 - \frac{r_{22}}{\sqrt{n}})1\{q_{iT-1} > \gamma_0 + \frac{r_{22}}{\sqrt{n}}\}] \right) \xrightarrow{P} \left( E \hat{z}_{it_0} [(q_{it_0} - \gamma_0)1\{q_{it_0} > \gamma_0\} - (q_{it_0-1} - \gamma_0)1\{q_{it_0-1} > \gamma_0\}] \right)
\]

\[
\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{iT} [(q_{iT} - \gamma_0 - \frac{r_{22}}{\sqrt{n}})1\{q_{iT} > \gamma_0 + \frac{r_{22}}{\sqrt{n}}\} - (q_{iT-1} - \gamma_0 - \frac{r_{22}}{\sqrt{n}})1\{q_{iT-1} > \gamma_0 + \frac{r_{22}}{\sqrt{n}}\}] \right) \xrightarrow{P} \left( E \hat{z}_{iT} [(q_{iT} - \gamma_0)1\{q_{iT} > \gamma_0\} - (q_{iT-1} - \gamma_0)1\{q_{iT-1} > \gamma_0\}] \right)
\]
uniformly with respect to $r_{22} \in [-K, K]$. Finally,

$$
\sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^{n} z_{it_0} [(q_{it} - \gamma_0) 1\{ q_{it} > \gamma_0 \} - (q_{it} - \gamma_0 - \frac{r_{22}}{\sqrt{n}}) 1\{ q_{it} > \gamma_0 + \frac{r_{22}}{\sqrt{n}} \}] \right. \\
\left. \vdots \right. \\
\left. \frac{1}{n} \sum_{i=1}^{n} z_{iT} [(q_{iT} - \gamma_0) 1\{ q_{iT} > \gamma_0 \} - (q_{iT} - \gamma_0 - \frac{r_{22}}{\sqrt{n}}) 1\{ q_{iT} > \gamma_0 + \frac{r_{22}}{\sqrt{n}} \}] \right. \\
\left. \vdots \right. \\
\left. \frac{1}{n} \sum_{i=1}^{n} z_{iT} [(q_{iT-1} - \gamma_0) 1\{ q_{iT-1} > \gamma_0 \} - (q_{iT-1} - \gamma_0 - \frac{r_{22}}{\sqrt{n}}) 1\{ q_{iT-1} > \gamma_0 + \frac{r_{22}}{\sqrt{n}} \}] \right. \\
\left. \vdots \right. \\
\left. \mathbb{P} \left( E[z_{it_0} 1\{ q_{it} > \gamma_0 \} - 1\{ q_{it} > \gamma_0 \} ] r_{22} \right. \\
\left. \vdots \right. \\
\left. E[z_{iT} 1\{ q_{iT} > \gamma_0 \} - 1\{ q_{iT} > \gamma_0 \} ] r_{22} \right) \\
\right\} \rightarrow 0
$$

uniformly with respect to $r_{22} \in [-K, K]$. Suppose that $r_{22} > 0$. The case for $r_{22} < 0$ follows similarly. The last uniform convergence holds because Lemma C.3 yields $\sqrt{n}\|\hat{g}_n(T(\beta_0, \delta_{30}, \gamma_0 + \frac{r_{22}}{\sqrt{n}})) - \hat{g}_n(T(\beta_0, \delta_{30}, \gamma_0))\| = o_p(1)$ uniformly with respect to $r_{22} \in [-K, K]$ and the following application of Taylor expansion:

$$
\sqrt{n}E[z_{it} ((q_{it} - \gamma_0) 1\{ q_{it} > \gamma_0 \} - (q_{it} - \gamma_0 - \frac{r_{22}}{\sqrt{n}}) 1\{ q_{it} > \gamma_0 + \frac{r_{22}}{\sqrt{n}} \})] \\
= \sqrt{n}E[z_{it} (q_{it} - \gamma_0) 1\{ \gamma_0 + \frac{r_{22}}{\sqrt{n}} \geq q_{it} > \gamma_0 \}] + r_{22} E[z_{it} 1\{ q_{it} > \gamma_0 + \frac{r_{22}}{\sqrt{n}} \}] \rightarrow E[z_{it} (q_{it} - \gamma_0) |\gamma_0] f_{r_{22}}(\gamma_0) r_{22} + E[z_{it} 1\{ q_{it} > \gamma_0 \}] r_{22} = E[z_{it} 1\{ q_{it} > \gamma_0 \}] r_{22}
$$

uniformly with respect to $r_{22} \in [-K, K]$ as $n \rightarrow \infty$.

In conclusion, $\sqrt{n}\hat{g}_n(T(\psi_0 + \frac{r}{\sqrt{n}})) \rightarrow (D_\psi r - e)$, and

$$
M(a, b, r) = (D_\psi r - e)'\Omega^{-1}(D_\psi r - e) - (M_{0a} + Hb^2 - e)'\Omega^{-1}(M_{0a} + Hb^2 - e) \\
= (M_{10}r_1 + N_{20}r_2 - e)'\Omega^{-1}(M_{10}r_1 + N_{20}r_2 - e) \\
- (M_{10}a_1 + M_{20}a_2 + Hb^2 - e)'\Omega^{-1}(M_{10}a_1 + M_{20}a_2 + Hb^2 - e),
$$

where $a_1 = \sqrt{n}(\beta - \beta_0)$ and $a_2 = \sqrt{n}(\delta - \delta_0)$. By applying the CMT, the continuity test statistic converges in distribution to

$$
\min_{r} \left( M_{10}r_1 + N_{20}r_2 - e \right)'\Omega^{-1}(M_{10}r_1 + N_{20}r_2 - e) \\
- \min_{a, b} \left( M_{10}a_1 + M_{20}a_2 + Hb^2 - e \right)'\Omega^{-1}(M_{10}a_1 + M_{20}a_2 + Hb^2 - e).
$$
By similar computations to the proof of Theorem 3,

\[
\min_{r_1, r_2}(M_{10}r_1 + N_{20}r_2 - e)' \Omega^{-1} (M_{10}r_1 + N_{20}r_2 - e) \\
= e' \Omega^{-1} e - e' \Omega^{-1} M_{10}(M_{10}^{-1} M_{10})^{-1} M_{10}' \Omega^{-1} e - e' \Xi_1 N_{20}(N_{20}' \Xi_1 N_{20})^{-1} N_{20}' \Xi_1 e, \\
\min_{a_1, a_2, b^2} (M_{10}a_1 + M_{20}a_2 + Hb^2 - e)' \Omega^{-1} (M_{10}a_1 + M_{20}a_2 + Hb^2 - e)
\]

\[
\begin{cases}
-e' \Omega^{-1} M_{10}(M_{10}' \Omega^{-1} M_{10})^{-1} M_{10}' \Omega^{-1} e - e' \Xi_1 M_{20}(M_{20}' \Xi_1 M_{20})^{-1} M_{20}' \Xi_1 e & \text{if } H' \Xi_1 e \geq 0 \\
-e' \Xi_1 H(H' \Xi_1 H)^{-1} H' \Xi_1 e + e' \Omega^{-1} e & \text{else}
\end{cases}
\]

where \( \Xi_1 = \Omega^{-1/2}(I - \Omega^{-1/2} M_{10}(M_{10}' \Omega^{-1/2} M_{10})^{-1} M_{10}' \Omega^{-1/2}) \Omega^{-1/2} \) and \( \Xi_1 = \Omega^{-1/2}(I - \Omega^{-1/2} M_{10}(M_{10}' \Xi_1 M_{10})^{-1} M_{10}' \Xi_1) \Xi_1^{-1/2} \). As \( \Xi_1 \Omega \Xi_1 = \Xi_1 \), we can derive \( \Xi_1 H(H' \Xi_1 H)^{-1} H' \Xi_1 e \sim \chi_1^2 \). Since \( E[H' \Xi_1 e M_{10} \Xi_1 M_{20}] = 0, (e' \Xi_1 M_{20}(M_{20}' \Xi_1 M_{20})^{-1} M_{20}' \Xi_1 e, e' \Xi_1 N_{20}(N_{20}' \Xi_1 N_{20})^{-1} N_{20}' \Xi_1 e) \) is independent to \( e' \Xi_1 H(H' \Xi_1 H)^{-1} H' \Xi_1 e \).

**Under the alternative hypothesis.** There is a constant \( C_1 \in (0, +\infty) \) such that \( \inf_{\theta \in \Theta} C_1 \geq C_1 \). This is because \( g_0(\theta) \) is zero if and only if \( \theta = \theta_0 \), by Assumption G and Theorem 1, and continuous on \( \Theta \), by Assumption D, while the restricted parameter set \( \{ \theta = (\beta^0, \delta, \gamma) : \delta = 0_{p-1}, \delta_1 + \delta_3 \gamma = 0 \} \) is closed. \( G = \{ g(\omega, \theta) : \theta \in \Theta \} \) is shown to satisfy the uniform entropy condition in the proof of Lemma C.3, and hence \( \sup_{\theta \in \Theta} \| g_n(\theta) - g_0(\theta) \| = o_p(1) \) by Glivenko-Cantelli theorem. By triangle inequality, \( C_1 \leq \| g_0(\bar{\theta}) \| \leq \| g_n(\bar{\theta}) \| + o_p(1) \). Recall that \( \bar{\theta} \) is the continuity-restricted estimator. Meanwhile, \( \| g_n(\bar{\theta}) \| = O_p(n^{-1/2}) \) because \( \| g_n(\bar{\theta}) \| \leq \| g_n(\theta_0) \| = O_p(n^{-1/2}) \). Therefore, there exists \( C_2 \in (0, +\infty) \) such that \( \bar{Q}_n(\bar{\theta}) - \bar{Q}_n(\hat{\theta}) \geq C_2 + O_p(n^{-1}) \), which implies that \( P(n^{-m} T_n > M) = P(\bar{Q}_n(\bar{\theta}) - \bar{Q}_n(\hat{\theta}) > M/(n^{1-m})) \to 1 \), for any \( m \in (0, 1) \) and \( M < \infty \).

**C.3 Auxiliary Lemmas**

**Lemma C.1.** Suppose that the true model is continuous and Assumptions G, D, and LK are true. For any \( \eta > 0 \), there is a neighborhood \( \Theta_0 \) of \( \theta_0 \) such that the population moment function \( g_0(\theta) \) satisfies

\[
\lim_{n \to \infty} \sup_{\theta \in \Theta_0} \frac{\sqrt{n} \| g_0(\theta) - D_2 (\alpha' - \alpha_0', (\gamma - \gamma_0)^2)' \|}{1 + \sqrt{n} \| (\alpha' - \alpha_0', (\gamma - \gamma_0)^2)' \|} < \eta.
\]

**Proof.** Recall that \( G \), whose formula is (5), is the first-order derivative of \( g_0(\theta) \) with respect to \( \gamma \) at \( \theta = \theta_0 \), and \( H \), whose formula is (6), is a half of the second-order derivative. \( G \) can be
obtained by applying the Leibniz rule as follows:

\[
\frac{d}{d\gamma} E[-z_{it}(1, x'_{it})\delta_0 \{q_{it} > \gamma\}]_{\gamma=\gamma_0} = \frac{d}{d\gamma} \int_{\gamma}^{\infty} -E_t[z_{it}(1, x'_{it})\delta_0[q]f_t(q)dq}_{\gamma=\gamma_0} = E_t[z_{it}(1, x'_{it})\delta_0]\gamma_0 f_t(\gamma_0).
\]

Similarly, we can get

\[
\frac{d}{d\gamma} E_t[z_{it}(1, x'_{it-1})\delta_0 \{q_{it-1} > \gamma\}]_{\gamma=\gamma_0} = -E_{t-1}[z_{it}(1, x'_{it-1})\delta_0]\gamma_0 f_{t-1}(\gamma_0).
\]

This implies the formula (5) for \(G\). \(H\) can also be obtained by the Leibniz rule as follows:

\[
\frac{d}{d\gamma} E_t[z_{it}(1, x'_{it})\delta_0\gamma]f_t(\gamma)_{\gamma=\gamma_0} = \frac{d}{d\gamma} E_t[z_{it}(\delta_{10} + \delta_{30}\gamma)]\gamma f_t(\gamma)_{\gamma=\gamma_0} = \frac{d}{d\gamma} (\delta_{10} + \delta_{30}\gamma) \cdot E_t[z_{it}\gamma]f_t(\gamma)_{\gamma=\gamma_0} = \delta_{30} E_t[z_{it}\gamma]f_t(\gamma) + (\delta_{10} + \delta_{30}\gamma) \frac{d}{d\gamma} E_t[z_{it}\gamma]f_t(\gamma)_{\gamma=\gamma_0} = \delta_{30} E_t[z_{it}\gamma]f_t(\gamma).
\]

Similarly, we can get

\[
\frac{d}{d\gamma} \{-E_{t-1}[z_{it}(1, x'_{it-1})\delta_0\gamma]f_{t-1}(\gamma)\}_{\gamma=\gamma_0} = -\delta_{30} E_{t-1}[z_{it}\gamma]f_{t-1}(\gamma).
\]

This implies the formula (6) for \(H\).

The population moment can be expressed as,

\[
g_0(\alpha, \gamma) = M_0(\gamma)(\alpha - \alpha_0) + H(\gamma - \gamma_0)^2 + o((\gamma - \gamma_0)^2).
\]

Define \(M_{0,G} = \begin{bmatrix} 0_k \times p \end{bmatrix} : M_G \end{bmatrix} \in \mathbb{R}^{k \times (2p+1)}\) where

\[
M_G = \begin{bmatrix} E_{t_0}[z_{it_0}(1, x'_{it_0})|\gamma_0]f_{t_0}(\gamma_0) - E_{t_0-1}[z_{it_0}(1, x'_{it_0-1})|\gamma_0]f_{t_0-1}(\gamma_0) \\
\vdots \\
E_{T}[z_{iT}(1, x'_{iT})|\gamma_0]f_{T}(\gamma_0) - E_{T-1}[z_{iT}(1, x'_{iT-1})|\gamma_0]f_{T-1}(\gamma_0) \end{bmatrix} \in \mathbb{R}^{k \times (p+1)}.
\]

The polynomial expansion \(M_0(\gamma) = M_0 + M_{0,G}(\gamma - \gamma_0) + o(|\gamma - \gamma_0|)\) implies

\[
g_0(\alpha, \gamma) = M_0(\alpha - \alpha_0) + H(\gamma - \gamma_0)^2 + o(\|\alpha - \alpha_0\| + (\gamma - \gamma_0)^2).
\]

Thus, \(\sqrt{m}_1 g_0(\theta) - D_2 (\alpha' - \alpha_0', (\gamma - \gamma_0)^2) = o(\sqrt{m}_1(\|\alpha - \alpha_0\| + (\gamma - \gamma_0)^2))\), which completes the proof. \(\square\)
Lemma C.2. If Assumption G is true, then
\[
\sup_{\gamma \in \Gamma} \|M_n(\gamma) - M_0(\gamma)\| \xrightarrow{p} 0.
\]

Proof. We show that the classes \(\{z_{it}(1, x'_{it})1\{q_{it} > \gamma\} : \gamma \in \Gamma\}\) and \(\{z_{it}(1, x'_{it-1})1\{q_{it-1} > \gamma\} : \gamma \in \Gamma\}\) are P-Glivenko-Cantelli. We focus on the former class since the verification for the latter class is exactly identical. Let \(\omega_i = \{(z_{it}, y_{it}, x_{it}, \epsilon_{it})_{t=1}^T\}\) be a random element in a measurable space \((\mathcal{X}, \mathcal{A})\). A collection of measurable index functions \(\mathcal{G}_{\text{index}} = \{1\{q_{it} > \gamma\} : \gamma \in \Gamma\}\) on \(\mathcal{X}\) is a VC class with a VC index 2. If \(m_{ij}\) is the \((i,j)\)th element of \(z_{it}(1, x'_{it})\), then \(\mathcal{G}_{\text{index} \cdot m_{ij}} = \{g_{\text{index} \cdot m_{ij}} : g_{\text{index}} \in \mathcal{G}_{\text{index}}\}\) is also a VC class as discussed by Lemma 2.6.18 in van der Vaart and Wellner (1996). The envelope for \(\mathcal{G}_{\text{index} \cdot m_{ij}}\) would be \(|m_{ij}|\) since an index function is always bounded by 1. The expectation of the envelope is bounded since \(E\|z_{it}(1, x'_{it})\| \leq \sqrt{E\|z_{it}\|^2 E\|(1, x'_{it})\|^2} < \infty\). In conclusion, \(\mathcal{G}_{\text{index} \cdot m_{ij}}\) is a P-Glivenko-Cantelli for each \((i,j)\), and thus the ULLN for \(\{z_{it}(1, x'_{it})1\{q_{it} > \gamma\} : \gamma \in \Gamma\}\) holds.

Lemma C.3. Let Assumption G hold. If \(h_n \to 0\), then
\[
\sup_{\|\theta_1 - \theta_2\| < h_n} \sqrt{n}\|\bar{g}_n(\theta_1) - \bar{g}_n(\theta_2) - g_0(\theta_1) + g_0(\theta_2)\| = o_p(1).
\]

Proof. Let \(\omega_i = \{(z_{it}, y_{it}, x_{it}, \epsilon_{it})_{t=1}^T\}\) be a random element in a measurable space \((\mathcal{X}, \mathcal{A})\), and \(P\) is the probability measure for \(\omega_i\). Define a functional class \(\mathcal{G} = \{g(\omega_i, \theta) : \theta \in \Theta\}\) on \(\mathcal{X}\) such that
\[
\begin{align*}
g(\omega_i, \theta) &= (g_{t_0}(\omega_i, \theta)', ..., g_T(\omega_i, \theta)')', \\
g_{t}(\omega_i, \theta) &= z_{it}\Delta y_{it} - z_{it}\Delta x'_{it}\beta - z_{it}1_{it}(\gamma)^'X_{it}\delta \\
&= z_{it}\epsilon_{it} - z_{it}\Delta x'_{it}(\beta - \beta_0) - z_{it}1_{it}(\gamma)^'X_{it}(\delta - \delta_0) + z_{it}(1_{it}(\gamma_0)^' - 1_{it}(\gamma)^')(X_{it}\delta_0).
\end{align*}
\]
and \(\mathcal{G}_n = \{g(\omega_i, \theta_1) - g(\omega_i, \theta_2) : \|\theta_1 - \theta_2\| < h, \theta_1, \theta_2 \in \Theta\}\). We need to show that \(P(\|\mathcal{G}_n\|_{\mathcal{G}_n} > x) \to 0\) if \(h \to 0\) as \(n \to \infty\), which is the asymptotic equicontinuity. To show the asymptotic equicontinuity, it is sufficient to show that each element of \(\mathcal{G}\) is P-Donsker, e.g. 2.3.11 Lemma and its corollary in van der Vaart and Wellner (1996), which is implied by the uniform entropy condition:
\[
\int_0^\infty \sup_Q \sqrt{\log N(\varepsilon\|G\|_{Q,2}, \mathcal{G}, L_2(Q))}d\varepsilon < \infty,
\]
where supremum is taken over all probability measures \(Q\) on \((\mathcal{X}, \mathcal{A})\) such that \(QG^2 < \infty\), and \(G\) is an envelope for \(\mathcal{G}\). For more details, see section 2.1 in van der Vaart and Wellner (1996). As we only need to consider each scalar element of \(\mathcal{G}\), it is sufficient to consider the following functional class
\[
\tilde{\mathcal{G}}^{(t)} = \{z_{it}\Delta \epsilon_{it} - z_{it}\Delta x'_{it}\bar{\beta} - z_{it}1_{it}(\gamma_1)^'X_{it}\delta_1 + z_{it}1_{it}(\gamma_2)^'X_{it}\delta_2 : \|\bar{\beta}\| \leq K, \|\delta_1\| \leq K, \|\delta_2\| \leq K, \gamma_1, \gamma_2 \in \Gamma\},
\]

S-9
where $K < \infty$ is a constant such that $\|\theta\| \leq K/2$ if $\theta \in \Theta$. Assume that $z_{it}$ is a scalar without losing of generality. Note that $g_t(\omega_t, \theta) = z_{it}(\Delta y_{it} - \Delta x_{it}^\prime \beta - 1_{it}(\gamma) X_{it} \delta) = z_{it} \Delta c_{it} - z_{it} \Delta x_{it}(\beta - \beta_{0n}) - z_{it} 1_{it}(\gamma)' X_{it} \delta + z_{it} 1_{it}(\gamma_0)' X_{it} \delta_0$ is an element of $\hat{G}(t)$. So it is sufficient to show $\hat{G}(t)$ satisfies the uniform entropy condition.

Let $\mathcal{G}_1 = \{z_{it} \Delta x_{it}^\prime \beta : \|\beta\| \leq K\}$. $\mathcal{G}_1$ is a $p$-dimensional vector space and is a VC class by 2.6.15 Lemma in van der Vaart and Wellner (1996), with an envelope function $G_1(\omega_t) = C \|z_{it} \Delta x_{it}^\prime\|$ for some constant $C < \infty$, and $EG_1^2 < \infty$. Let $\mathcal{G}_2 = \{z_{it}(1, x_{it}^\prime) \delta 1\{q_{it} > \gamma\} : \|\delta\| \leq K, \gamma \in \Gamma\}$, $\mathcal{G}_{2a} = \{z_{it}(1, x_{it}^\prime) \delta : \|\delta\| \leq K\}$, and $\mathcal{G}_{2b} = \{1\{q_{it} > \gamma\} : \gamma \in \Gamma\}$. $G_{2a} = C \|z_{it}(1, x_{it}^\prime)\|$ for some $C < \infty$ and $G_{2b} = 1$ are envelopes for $\mathcal{G}_{2a}$ and $\mathcal{G}_{2b}$, respectively. Note that $\mathcal{G}_2 = G_{2a} \mathcal{G}_{2b}$, i.e., $\mathcal{G}_2$ is a collection of $g_{2a} \cdot g_{2b}$ where $g_{2a} \in \mathcal{G}_{2a}$ and $g_{2b} \in \mathcal{G}_{2b}$. $\mathcal{G}_2$ satisfies the uniform entropy condition as pairwise sum or product of functional classes preserve the uniform entropy condition, e.g. Theorem 2.10.20 in van der Vaart and Wellner (1996). Note that for every $d > 0$,

$$
\int_0^d \sup_Q \sqrt{\log N(\varepsilon \|G_{2a} G_{2b}\|^{1/2} Q, L_2(Q))} d \varepsilon \\
\leq \int_0^d \sup_Q \sqrt{\log N(\varepsilon \|G_{2a}\| Q, L_2(Q))} d \varepsilon + \int_0^d \sup_Q \sqrt{\log N(\varepsilon \|G_{2b}\| Q, L_2(Q))} d \varepsilon,
$$

while $G_{2a} G_{2b}$ is an envelope of $\mathcal{G}_2$. So the uniform entropy condition for $\mathcal{G}_2$ holds. Similarly, we can show that $\mathcal{G}_3 = \{z_{it}(1, x_{it-1}^\prime) \delta 1\{q_{it-1} > \gamma\} : \|\delta\| \leq K, \gamma \in \Gamma\}$ satisfies the uniform entropy condition. Hence, the functional class $(\mathcal{G}_2 - \mathcal{G}_3)$ defined by pairwise sum, which is a set of functions $g_{2a} - g_{3}$ for all $g_{2a} \in \mathcal{G}_2$ and $g_{3} \in \mathcal{G}_3$, also satisfies the uniform entropy condition, e.g. Theorem 2.10.20 in van der Vaart and Wellner (1996). As $(\mathcal{G}_2 - \mathcal{G}_3)$ is a superset of $\{z_{it} 1_{it}(\gamma)' X_{it} \delta : \|\delta\| \leq K, \gamma \in \Gamma\}$, the functional class $\{z_{it} 1_{it}(\gamma)' X_{it} \delta : \|\delta\| \leq K, \gamma \in \Gamma\}$ also satisfies the uniform entropy condition. Thus, $\{z_{it} \Delta c_{it} - \mathcal{G}_1 - (\mathcal{G}_2 - \mathcal{G}_3) + (\mathcal{G}_2 - \mathcal{G}_3), \}$, which is a superset of $\hat{G}(t)$, satisfies the uniform entropy condition by repetitively applying Theorem 2.10.20 in van der Vaart and Wellner (1996), and hence $\hat{G}(t)$ also satisfies the condition.

Note that for some constant $C < \infty$,

$$
\hat{G} = C(\|z_{it} \Delta x_{it}^\prime\| + \|z_{it}(1, x_{it}^\prime)\| + \|z_{it}(1, x_{it-1}^\prime)\|) + \|z_{it} \Delta c_{it}\|
$$

is an envelope for $\hat{G}(t)$, and $E\hat{G}^2 < \infty$ by Assumption G.

\[\]

**Lemma C.4.** When the true model is continuous and Assumptions G, D, and LK are true,

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^n z_{it}(1_{it}(\gamma_0)' - 1_{it}(\gamma_0 + \frac{b}{n^4})' X_{it} \delta_0) \frac{\delta_{30}}{2} \{E_t[z_{it}|\gamma_0] f_t(\gamma_0) - E_{t-1}[z_{it}|\gamma_0] f_{t-1}(\gamma_0)\} b^2
$$

uniformly over $b \in [-K, K]$ for any $K < \infty$. 

S-10
Proof. Note that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}(1_{it}(\gamma_0)' - 1_{it}(\gamma_0 + \frac{b}{n^4})' ) X_{it}\delta_0
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ z_{it}(1_{it}(\gamma_0)' - 1_{it}(\gamma_0 + \frac{b}{n^4})' ) X_{it}\delta_0 - E[z_{it}(1_{it}(\gamma_0)' - 1_{it}(\gamma_0 + \frac{b}{n^4})' ) X_{it}\delta_0] \right\} \quad \text{(C.5)}
\]

\[
+ \sqrt{n} E[z_{it}(1_{it}(\gamma_0)' - 1_{it}(\gamma_0 + \frac{b}{n^4})' ) X_{it}\delta_0]. \quad \text{(C.6)}
\]

The stochastic term (C.5) converges in probability to zero uniformly with respect to \( b \in [-K, K] \). This is because Lemma C.3 shows that when \( h_n \downarrow 0 \), then

\[
\sup_{|\gamma - \gamma_0| < h_n} \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^{n} z_{it}(1_{it}(\gamma_0)' - 1_{it}(\gamma))' X_{it}\delta_0 - E[z_{it}(1_{it}(\gamma_0)' - 1_{it}(\gamma))' X_{it}\delta_0] \right\} = o_p(1)
\]

as it can be expressed as \( \sup_{|\gamma - \gamma_0| < h_n} \| \tilde{g}_n(\alpha_0, \gamma) - \tilde{g}_n(\alpha_0, \gamma_0) - g_0(\alpha_0, \gamma) + g_0(\alpha_0, \gamma_0) \| \).

Suppose \( b > 0 \). The case for \( b < 0 \) follows similarly. As \( n \to \infty \), the deterministic term (C.6) converges as follows:

\[
\sqrt{n} E[z_{it}(1_{it}(\gamma_0)' - 1_{it}(\gamma_0 + \frac{b}{n^4})' ) X_{it}\delta_0]
\]

\[
= \sqrt{n} \left\{ E[z_{it}(\delta_{10} + \delta_{30} q_{it}) 1 \{ \gamma_0 + \frac{b}{n^4} \geq q_{it} > \gamma_0 \}] - E[z_{it}(\delta_{10} + \delta_{30} q_{it-1}) 1 \{ \gamma_0 + \frac{b}{n^4} \geq q_{it-1} > \gamma_0 \}] \right\}
\]

\[
\to \frac{\delta_{30}}{2} \left\{ E_t[z_{it}|\gamma_0] f_t(\gamma_0) - E_{t-1}[z_{it}|\gamma_0] f_{t-1}(\gamma_0) \right\} b^2;
\]

uniformly with respect to \( b \in [-K, K] \). To show that, use the (second-order) derivative of \( \kappa \mapsto E[z_{it}(\delta_{10} + \delta_{30} q_{it}) 1 \{ \gamma_0 + \kappa \geq q_{it} > \gamma_0 \}] \) and derive the Taylor expansion

\[
\sqrt{n} E[z_{it}(\delta_{10} + \delta_{30} q_{it}) 1 \{ \gamma_0 + \frac{b}{n^4} \geq q_{it} > \gamma_0 \}]
\]

\[
= \frac{b^2}{2} \left( \delta_{30} E_t[z_{it}|\gamma_n,b] f_t(\gamma_{n,b}) + (\delta_{10} + \delta_{30} \gamma_{n,b}) \frac{d}{d\gamma} E_t[z_{it}|\gamma] f_t(\gamma)|_{\gamma=\gamma_{n,b}} \right),
\]

where \( \gamma_{n,b} \in [\gamma_0, \gamma_0 + \frac{b}{n^4}] \). Note that \( |\gamma_{n,b} - \gamma_0| \to 0 \) uniformly with respect to \( b \in [-K, K] \). Since \( E_t[z_{it}|\gamma] \) and \( f_t(\gamma) \) are continuously differentiable at \( \gamma_0 \) by Assumption D, both \( \frac{d}{d\gamma} E_t[z_{it}|\gamma] f_t(\gamma)|_{\gamma=\gamma_{n,b}} \to \frac{d}{d\gamma} E_t[z_{it}|\gamma] f_t(\gamma)|_{\gamma=\gamma_0} \) and \( (\delta_{10} + \delta_{30} \gamma_{n,b}) \to 0 \) hold uniformly with respect to \( b \in [-K, K] \). On the other hand, \( E_t[z_{it}|\gamma_n,b] f_t(\gamma_{n,b}) \to E_t[z_{it}|\gamma_0] f_t(\gamma_0) \) uniformly with respect to \( b \in [-K, K] \). Hence, \( \sqrt{n} E[z_{it}(\delta_{10} + \delta_{30} q_{it}) 1 \{ \gamma_0 + \frac{b}{n^4} \geq q_{it} > \gamma_0 \}] \) converges to \( \frac{\delta_{30}}{2} E_t[z_{it}|\gamma_0] f_t(\gamma_0) b^2 \) uniformly with respect to \( b \in [-K, K] \) as \( n \to \infty \). We can derive the similar result for \( \sqrt{n} E[z_{it}(\delta_{10} + \delta_{30} q_{it-1}) 1 \{ \gamma_0 + \frac{b}{n^4} \geq q_{it-1} > \gamma_0 \}] \). 

\( \square \)
D Proofs of Theorems in Section 4 and Auxiliary Lemmas

D.1 Preliminaries

Proofs in this section are regarding bootstrap results, and hence we explain empirical process framework for our bootstrap analysis. Let \( \omega_1^*, \ldots, \omega_n^* \) be i.i.d. resampling draws from a given sample \( \{\omega_i : 1 \leq i \leq n\} \). We set \( \omega_i = \{(z_{it}, y_{it}, x_{it}, \epsilon_{it})\}_{t=1}^{T_i} \) as in the proofs of Lemmas C.2 and C.3. An important functional class for our bootstrap analysis is \( \mathcal{G} = \{g(\omega_i, \theta) : \theta \in \Theta\} \) where \( g(\omega_i, \theta) \) is defined as in (C.4).

Be mindful that \( g_i^*(\theta) \) that appears in Section 4 is different from \( g(\omega_i^*, \theta) \). This is because

\[
g_i^*(\theta) = (g_{i0}^*(\theta)', \ldots, g_{iT}^*(\theta)')'
\]

where

\[
g_{it}^*(\theta) = z_{it}^*(\Delta y_{it}^* - \Delta x_{it}^* \beta - 1_{it}^*(\gamma)' X_{it}^* \delta)
\]

Recall that \( \Delta y_{it}^* \) is not an i.i.d. resampling draw from \( \{\Delta y_{it} : 1 \leq i \leq n\} \) but is generated using resampled regressors and residuals with regression equation using \( \theta_0^* \). The formula for \( \Delta y_{it}^* \) is used to derive the equality in (D.1) (see Step 2 in Algorithm 1). Instead, \( g_{it}^*(\theta) = g_{it}(\omega_i^*, \theta) - g_{it}(\omega_i^*, \theta_0^*) + g_{it}(\omega_i^*, \hat{\theta}) \). To be more precise, \( I \) in (D.1) is \( g_{it}(\omega_i^*, \theta) - g_{it}(\omega_i^*, \theta_0^*) \), and \( II \) in (D.1) is \( g_{it}(\omega_i^*, \hat{\theta}) \).

D.2 Proof of Proposition 1

Consistency of the bootstrap estimator. The bootstrap sample moment can be rewritten by

\[
\bar{G}_n^*(\theta) = \frac{1}{n} \sum_{i=1}^{n} (g_i^*(\theta) - \bar{g}_n(\bar{\theta}))
\]

\[
= \left( \frac{1}{n} \sum_{i=1}^{n} z_{i0}^* x_{i0}^* \Delta \epsilon_{i0} \right) - \left( \frac{1}{n} \sum_{i=1}^{n} z_{iT}^* x_{iT}^* \Delta \epsilon_{iT} \right) - \left( \frac{1}{n} \sum_{i=1}^{n} z_{iT}^* \Delta x_{iT}^* \right) (\beta - \beta_0^*)
\]

\[
= \left( \frac{1}{n} \sum_{i=1}^{n} z_{i0}^* x_{i0}^* (1_{i0}^*(\gamma)' X_{i0}^*) \right) (\delta - \delta_0^*) + \left( \frac{1}{n} \sum_{i=1}^{n} z_{iT}^* (1_{iT}^*(\gamma)' X_{iT}^*) \right) \delta_0^*.
\]

We additionally define

\[
\psi_1^* \equiv \begin{pmatrix}
z_{i0}^* \Delta y_{i0}^* \\
\vdots \\
z_{iT}^* \Delta y_{iT}^*
\end{pmatrix} - \begin{pmatrix}
\frac{1}{n} \sum_{i=1}^{n} z_{i0}^* \Delta \epsilon_{i0} \\
\vdots \\
\frac{1}{n} \sum_{i=1}^{n} z_{iT}^* \Delta \epsilon_{iT}
\end{pmatrix},
M_1^*(\gamma) = - \begin{pmatrix}
\frac{1}{n} \sum_{i=1}^{n} z_{i0}^* (1_{i0}^*(\gamma)' X_{i0}^*) \\
\vdots \\
\frac{1}{n} \sum_{i=1}^{n} z_{iT}^* (1_{iT}^*(\gamma)' X_{iT}^*)
\end{pmatrix},
\]

\[
\frac{1}{n} \sum_{i=1}^{n} z_{i0}^* (\Delta x_{iT}^* 1_{iT}^*(\gamma)' X_{iT}^*).
\]

S-12
\[ \hat{\alpha}^*(\gamma) = -(M_n^*(\gamma)W_n^*\bar{M}_n^*(\gamma))^{-1}\bar{M}_n^*(\gamma)W_n^*\bar{v}_n \]

where

\[
\bar{v}_n = -\bar{M}_n^*(\gamma_0^*)\alpha_0^* + \bar{u}_n; \quad \bar{u}_n = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} z_{it0}^* \Delta \epsilon_{it0} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^{n} z_{iT}^* \Delta \epsilon_{iT} \end{pmatrix}.
\]

Let \( \tilde{Q}_n^*(\gamma) = \tilde{Q}_n^*(\hat{\alpha}^*(\gamma), \gamma) \) be a profiled criterion and \( \hat{\gamma}^* = \arg \min_{\gamma \in \Gamma} \tilde{Q}_n^*(\gamma) \). \( \bar{u}_n^* = o_p(1) \) in \( P \) by Lemma D.1. By Lemma D.3, \( \sup_{\gamma \in \Gamma} \| \bar{Q}_n^*(\gamma) - M_0(\gamma) \| = o_p(1) \) in \( P \). Therefore, if \( |\hat{\gamma}^* - \gamma_0^*| \xrightarrow{p} 0 \) in \( P \), then \( \| \hat{\alpha}^*(\hat{\gamma}^*) - \alpha_0^* \| \xrightarrow{p} 0 \) in \( P \), which completes the proof.

Let \( \tilde{g}_n^*(\gamma) = \tilde{g}_n^*(\hat{\alpha}^*(\gamma), \gamma) \) which can be expressed as

\[
\tilde{g}_n^*(\gamma) = [I - \bar{M}_n^*(\gamma)(\bar{M}_n^*(\gamma)W_n^*\bar{M}_n^*(\gamma))^{-1}\bar{M}_n^*(\gamma)W_n^*] (-\bar{M}_n^*(\gamma_0^*)\alpha_0^* + \bar{u}_n^*). \]

Therefore,

\[
W_n^{1/2}\tilde{g}_n^*(\gamma) = [I - P_{W_n^{1/2}\bar{M}_n^*(\gamma)}] \left(-W_n^{1/2}\bar{M}_n^*(\gamma_0^*)\alpha_0^* + W_n^{1/2}\bar{u}_n^*\right),
\]

and

\[
\sup_{\gamma \in \Gamma} \| \tilde{Q}_n^*(\gamma) - \left[I - P_{W_n^{1/2}M_0(\gamma)}\right] \left(-W_n^{1/2}M_0(\gamma)\alpha_0\right) \|^2 = o_p(1) \] in \( P \)

when \( \|W_n* - W\| = o_p(1) \) in \( P \) and \( \theta_0^* \xrightarrow{P} \theta_0 \). Note that \( W \) is the identity matrix if it is for the first step estimation and \( \Omega^{-1} \) if it is for the second step estimation and the first step estimator is consistent. Since the uniform probability limit of \( \tilde{Q}_n^*(\gamma) \) conditional on the data is minimized when \( \gamma = \gamma_0 \), the argmin CMT implies \( \hat{\gamma}^* - \gamma_0 = o_p(1) \) in \( P \). Recall that \( \theta_0^* \) is set as \( (\hat{\alpha}(\gamma_0), \gamma_0)' \) in Theorem 5, (8) in Theorem 6, and \( \hat{\theta} \) in Theorem 7. For both cases (i) and (ii) of the proposition, \( \gamma_0^* \xrightarrow{P} \gamma_0 \) which implies \( \gamma_0^* - \gamma_0 = o_p(1) \) in \( P \) by Lemma B.1. Therefore, we can derive that \( \hat{\gamma}^* - \gamma_0^* = (\hat{\gamma}^* - \gamma_0) - (\gamma_0^* - \gamma_0) = o_p(1) \) in \( P \).

**Convergence rate under continuity.** By bootstrap equicontinuity, Lemma D.4, and the consistency of \( \hat{\theta}^* \) to \( \theta_0^* \),

\[
\sqrt{n} \| \bar{g}_n^*(\hat{\theta}^*) - \bar{g}_n(\theta_0^*) - \bar{g}_n(\theta_0^*) + \bar{g}_n(\theta_0^*) \| = o_p(1) \] in \( P \).

\[
\| W_n^* - W_n \| \xrightarrow{p} 0 \] in \( P \) since \( \| W_n - \Omega^{-1} \| = o_p(1) \) in \( P \) and \( \| W_n^* - \Omega^{-1} \| = o_p(1) \) in \( P \). The condition \( \| W_n^* - \Omega^{-1} \| = o_p(1) \) in \( P \) is implied by \( \hat{\theta}(1) \xrightarrow{P} \theta_0 \) in \( P \), as \( |\hat{\theta}(1) - \theta_0| \xrightarrow{p} 0 \) and
Therefore, \( \sqrt{n} \| W_n^{s1/2} \hat{g}_n(\hat{\theta}^*) - W_n^{s1/2} \bar{g}_n(\theta_0^*) - W_n^{s1/2} \tilde{g}_n(\theta_0^*) + W_n^{s1/2} \bar{g}_n(\theta_0^*) \| = o_p(1) \) in \( P \).

Apply triangle inequality to get

\[
\sqrt{n} \| W_n^{s1/2} \bar{g}_n(\hat{\theta}^*) - W_n^{s1/2} \bar{g}_n(\theta_0^*) \| \leq o_p(1) + \sqrt{n} \| W_n^{s1/2} \bar{g}_n(\theta_0^*) \| + \sqrt{n} \| W_n^{s1/2} \bar{g}_n(\hat{\theta}^*) \|
\]

where \( o_p(1) \) holds in \( P \). As \( \hat{\theta}^* \) is the minimizer of the bootstrap criterion, \( \sqrt{n} \| W_n^{s1/2} \bar{g}_n(\hat{\theta}^*) \| \leq \sqrt{n} \| W_n^{s1/2} \bar{g}_n(\theta_0^*) \| = O_p(1) \) in \( P \) where the last equality is implied by Lemma D.2. Therefore,

\[
\sqrt{n} \| W_n^{s1/2} \bar{g}_n(\hat{\theta}^*) - W_n^{s1/2} \bar{g}_n(\theta_0^*) \| \leq O_p(1) \) in \( P \).

By Lemma C.3, \( \sqrt{n} \| W_n^{s1/2} \bar{g}_n(\hat{\theta}^*) - W_n^{s1/2} \bar{g}_n(\theta_0^*) - \Omega^{-1/2} g_0(\hat{\theta}^*) + \Omega^{-1/2} g_0(\theta_0^*) \| = o_p(1) \), so it is \( o_p(1) \) in \( P \) by Lemma B.1. Hence,

\[
\sqrt{n} \| \Omega^{-1/2} g_0(\hat{\theta}^*) - \Omega^{-1/2} g_0(\theta_0^*) \| \leq O_p(1) \) in \( P \).

By Lemma C.1, \( \sqrt{n} \| \Omega^{-1/2} g_0(\hat{\theta}^*) - \Omega^{-1/2} g_0(\theta_0^*) \| \geq \sqrt{n} \| \Omega^{-1/2} M_0(\hat{\alpha}^* - \alpha_0^*) + \Omega^{-1/2} H \{ (\hat{\gamma}^* - \gamma_0)^2 - (\gamma_0^* - \gamma_0)^2 \} \| + o_p(1) \) in \( P \). Therefore, \( \sqrt{n} \| \hat{\alpha}^* - \alpha_0^* \| = O_p(1) \) in \( P \) and \( \sqrt{n} (\hat{\gamma}^* - \gamma_0)^2 = O_p(1) \) in \( P \). Suppose that \( \sqrt{n} (\gamma_0^* - \gamma_0)^2 = O_p(1) \) in \( P \). Then, \( \sqrt{n} (\hat{\gamma}^* - \gamma_0)^2 = O_p(1) \) in \( P \) since \( \sqrt{n} (\hat{\gamma}^* - \gamma_0)^2 \leq 2 \sqrt{n} (\gamma_0^* - \gamma_0)^2 + (\gamma_0^* - \gamma_0)^2 = O_p(1) \) in \( P \).

The condition, \( \sqrt{n} (\gamma_0^* - \gamma_0)^2 = O_p(1) \) in \( P \), is true if \( \sqrt{n} (\gamma_0^* - \gamma_0)^2 = O_p(1) \) by Lemma B.1. This is true for \( \gamma_0^* = \gamma_0 \) (Theorem 5 (i)), \( \gamma_0^* = w_n \hat{\gamma} + (1 - w_n) \tilde{\gamma} \) (Theorem 6 (i)), or \( \gamma_0^* = \tilde{\gamma} \) (Theorem 7 (i)). It is also the case for the standard nonparametric bootstrap as \( \sqrt{n} (\hat{\gamma} - \gamma_0)^2 = O_p(1) \) by Theorem 2.

**Convergence rate under discontinuity.** Identically to the proof for the continuous model, we can get

\[
\sqrt{n} \| \Omega^{-1/2} g_0(\hat{\theta}^*) - \Omega^{-1/2} g_0(\theta_0^*) \| \leq O_p(1) \) in \( P \).

Meanwhile, \( \sqrt{n} \| \Omega^{-1/2} g_0(\hat{\theta}^*) - \Omega^{-1/2} g_0(\theta_0^*) \| \geq C \sqrt{n} \| \hat{\theta}^* - \theta_0^* \| + o_p(1 + \sqrt{n} \| \hat{\theta}^* - \theta_0^* \|) \) for some \( C < \infty \) in \( P \) when the true model is discontinuous and Assumption LJ holds. This is because \( g_0(\theta) = D_1(\theta - \theta_0) + o(\| \theta - \theta_0 \|) \) by Assumption LJ and

\[
o(1) = \frac{\| g_0(\theta) - D_1(\theta - \theta_0) \|}{\| \theta - \theta_0 \|} \geq \frac{\sqrt{n} \| g_0(\theta) - D_1(\theta - \theta_0) \|}{1 + \sqrt{n} \| \theta - \theta_0 \|}.
\]

Therefore, \( \sqrt{n} \| \hat{\theta}^* - \theta_0^* \| \leq O_p(1) \) in \( P \).

**D.3 Proof of Theorem 5.**

In the grid bootstrap at \( \gamma \), \( \theta_0^* = (\hat{\alpha}(\gamma)' , \gamma)' \).
When $\gamma = \gamma_0$. The proof of Theorem 6 still holds, and $S_n^*(a, b)$ conditionally weakly converges to either $S$ or $S_J$ in $\ell^\infty(K)$ in $P$ for every compact $K$. The limit is $S$ for the Theorem 5 (i) case, and $S_J$ for the Theorem 5 (ii) case. By following the similar steps to the proof of Theorem 3, we can derive the asymptotic distributions of $D_n^*(\gamma)$.

When $\gamma \neq \gamma_0$. Note that $\tilde{g}_n^*(\hat{\alpha}(\gamma), \gamma) = O_p(n^{-1/2})$. It will be shown that $\|W_n^*\| = O_p(1)$ in $P$. Then, $\min_\alpha \tilde{Q}_n^*(\alpha, \gamma) \leq \tilde{Q}_n(\hat{\alpha}(\gamma), \gamma) = \tilde{g}_n^*(\hat{\alpha}(\gamma), \gamma)W_n^*\tilde{g}_n^*(\hat{\alpha}(\gamma), \gamma) = O_p(n^{-1})$, and $D_n^*(\gamma) \leq \min_\alpha \tilde{Q}_n^*(\alpha, \gamma) = O_p(1)$ in $P$, which completes the proof.

Recall that

$$ W_n^* = \left\{ \frac{1}{n} \sum_{i=1}^{n} [g_i^*(\hat{\alpha}_i^*)g_i^*(\hat{\theta}_i^*)'] - \frac{1}{n} \sum_{i=1}^{n} g_i^*(\hat{\theta}_i^*) \frac{1}{n} \sum_{i=1}^{n} g_i^*(\hat{\theta}_i^*)' \right\}^{-1}, $$

while $g_i^*(\theta) = g(\omega_i^*, \theta) - g(\theta_0^*, \theta_0^*) + g(\omega_i^*, \hat{\theta})$ as explained in Online Appendix D.1. The functional class $\mathcal{G} = \{g(\omega, \theta) : \theta \in \Theta\}$ is shown to satisfy the uniform entropy condition in the proof of Lemma C.3, and pairwise sum or product of functional classes preserve the uniform entropy condition by Theorem 2.10.20 in van der Vaart and Wellner (1996). Hence, by applying the bootstrap Glivenko-Cantelli theorem, e.g. Lemma 3.6.16 in van der Vaart and Wellner (1996),

$$ \sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} [g_i^*(\theta)g_i^*(\theta)'] - \frac{1}{n} \sum_{i=1}^{n} g_i^*(\theta) \frac{1}{n} \sum_{i=1}^{n} g_i^*(\theta)' \right\| $$

is $o_p(1)$ in $P$. Furthermore,

$$ \frac{1}{n} \sum_{i=1}^{n} \left[ \{g_i(\theta) - g_i(\theta_1) + g_i(\theta_2)\}\{g_i(\theta) - g_i(\theta_1) + g_i(\theta_2)\}' \right] $$

$$ - \frac{1}{n} \sum_{i=1}^{n} \left[ \{g_i(\theta) - g_i(\theta_1) + g_i(\theta_2)\} \frac{1}{n} \sum_{i=1}^{n} \{g_i(\theta) - g_i(\theta_1) + g_i(\theta_2)\}' \right] \xrightarrow{P} E \left[ \{g_i(\theta) - g_i(\theta_1) + g_i(\theta_2)\}\{g_i(\theta) - g_i(\theta_1) + g_i(\theta_2)\}' \right] $$

uniformly with respect to $\theta, \theta_1,$ and $\theta_2$. As $\hat{\theta}$ and $\hat{\theta}_0^*$ are consistent to $\theta_0$,

$$ \frac{1}{n} \sum_{i=1}^{n} [g_i^*(\theta)g_i^*(\theta)'] - \frac{1}{n} \sum_{i=1}^{n} g_i^*(\theta) \frac{1}{n} \sum_{i=1}^{n} g_i^*(\theta)' \xrightarrow{P} E \left[ g_i(\theta)g_i(\theta)' \right] - E[g(\theta)]E[g(\theta)'] $$

uniformly with respect to $\theta$. By the compactness of $\Theta$, the minimum eigenvalue of $\{E[g_i(\theta)g_i(\theta)'] - E[g(\theta)]E[g(\theta)']\}$ is bounded below by some constant $c > 0$. Therefore,
Lemma D.2. If Assumption G holds and

\[ W_n^*(\theta) = O_p^*(1) \] in P where

\[ W_n^*(\theta) = \left\{ \frac{1}{n} \sum_{i=1}^n [g_i^*(\theta)g_i^*(\theta)'] - \frac{1}{n} \sum_{i=1}^n g_i^*(\theta) \frac{1}{n} \sum_{i=1}^n g_i^*(\theta)' \right\}^{-1}. \]

As \( W_n^* = W_n^*(\hat{\theta}_n^*) \), we can conclude that \( \|W_n^*\| = O_p^*(1) \).

D.4 Proof of Theorem 7.

In the bootstrap for continuity test, \( \theta_n^* = \hat{\theta} \), where \( \hat{\theta} \) is the continuity-restricted estimator.

Under the null hypothesis. When the true model is continuous, the proof of Theorem 6 still holds. \( S_n^*(a, b) \) conditionally weakly converges to \( S \) in \( \ell^\infty(K) \) in P for every compact \( K \). By following the similar steps to the proof of Theorem 4, we can derive the asymptotic distribution of \( T_n^* \).

Under the alternative hypothesis. Let the true model be discontinuous. Note that \( \tilde{g}_n^*(\hat{\theta}) = O_p^*(n^{-1/2}) \). Meanwhile, \( \|W_n^*\| = O_p^*(1) \) in P, by the same logic used in the proof of Theorem 5 when \( \gamma \neq \gamma_0 \). Then, \( \min_{\theta \in \Theta; K_2=0_p, \delta_1=-\delta, \gamma} \hat{Q}_n^*(\theta) \leq \hat{Q}_n^*(\hat{\theta})' W_n^* \tilde{g}_n^*(\hat{\theta}) = O_p^*(n^{-1}) \). Therefore, \( T_n^* \leq n \min_{\theta \in \Theta; K_2=0_p, \delta_1=-\delta, \gamma} \hat{Q}_n^*(\theta) = O_p^*(1) \) in P, which completes the proof.

D.5 Lemmas

Lemma D.1. If Assumption G holds,

\[ \hat{u}_n^* = \left( \frac{1}{n} \sum_{i=1}^n z_{i0}^* \tilde{\xi}_{i0}^* \right) - \left( \frac{1}{n} \sum_{i=1}^n z_{iT}^* \tilde{\xi}_{iT}^* \right) \xrightarrow{P} 0 \text{ in } P. \]

Proof. Let \( u_n^*(\theta) = \frac{1}{n} \sum_{i=1}^n [g(\omega_i^*, \theta) - \frac{1}{n} \sum_{i=1}^n g(\omega_i, \theta)] \) where \( g(\omega, \theta) \) is defined by (C.4), and \( \omega_i^* \) is a resampling draw from \( \{\omega_i : i = 1, \ldots, n\} \). See Online Appendix D.1 for more explanation.

Lemma D.2. If Assumption G holds and \( \hat{\theta} \xrightarrow{P} \theta_0 \), then

\[ \sqrt{n} \left\{ \left( \frac{1}{n} \sum_{i=1}^n z_{i0}^* \tilde{\xi}_{i0}^* \right) - \left( \frac{1}{n} \sum_{i=1}^n z_{i0} \tilde{\xi}_{i0} \right) \right\} \xrightarrow{d} N(0, \Omega) \text{ in } P. \]
Proof. Note that \( g^*_i(\theta_1) - g^*_i(\theta_2) = g(\omega^*_i, \theta_1) - g(\omega^*_i, \theta_2) \) for any \( \theta_1 \) and \( \theta_2 \) where \( g(\omega_i, \theta) \) is defined by (C.4), and \( \omega^*_i \) is a resampling draw from \( \{\omega_i : i = 1, \ldots, n\} \). See Online Appendix D.1 for more explanation. Hence, \( \tilde{g}_n^*(\theta) = \frac{1}{n} \sum_{i=1}^n [g(\omega_i^*, \theta) - \frac{1}{n} \sum_{i=1}^n g(\omega_i, \theta)] - \frac{1}{n} \sum_{i=1}^n [g(\omega_i^*, \theta_0) - \frac{1}{n} \sum_{i=1}^n g(\omega_i, \theta_0)] \). Furthermore,

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^n g(\omega_i^*, \hat{\theta}) - \frac{1}{n} \sum_{i=1}^n g(\omega_i, \hat{\theta}) = \sqrt{n} \left\{ \left( \frac{1}{n} \sum_{i=1}^n z_{it_0}^* \Delta \epsilon_{it_0} \right) - \left( \frac{1}{n} \sum_{i=1}^n z_{IT}^* \Delta \epsilon_{IT} \right) \right\}.
\]

By Lemma D.4, \( \sqrt{n}||\tilde{g}_n^*(\hat{\theta}) - \tilde{g}_n^*(\theta_0) - \tilde{g}_n(\hat{\theta}) + \tilde{g}_n(\theta_0)|| = \sqrt{n}||\frac{1}{n} \sum_{i=1}^n [g(\omega_i^*, \hat{\theta}) - \frac{1}{n} \sum_{i=1}^n g(\omega_i, \hat{\theta})] - \frac{1}{n} \sum_{i=1}^n [g(\omega_i^*, \theta_0) - \frac{1}{n} \sum_{i=1}^n g(\omega_i, \theta_0)] || = o^*_n(1) \) in \( P \). By the bootstrap CLT (e.g. Gine and Zinn (1990)),

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^n g(\omega_i^*, \theta_0) - \frac{1}{n} \sum_{i=1}^n g(\omega_i, \theta_0) \xrightarrow{d^*} N(0, \Omega) \text{ in } P.
\]

By applying the Slutsky theorem, we can derive \( \frac{1}{\sqrt{n}} \sum_{i=1}^n [g(\omega_i^*, \hat{\theta}) - \frac{1}{n} \sum_{i=1}^n g(\omega_i, \hat{\theta})] \xrightarrow{d^*} N(0, \Omega) \) in \( P \).

Recall that \( \tilde{M}_n^*(\gamma) = \frac{1}{n} \sum_{i=1}^n M_i^*(\gamma) \) where

\[
M_i^*(\gamma) = -\begin{bmatrix} z_{it_0}^*(\Delta x_{it_0}^*, 1_{it_0}(\gamma)'X_{it_0}^*) \\ \vdots \\ z_{IT}^*(\Delta x_{IT}^*, 1_{IT}(\gamma)'X_{IT}^*) \end{bmatrix}.
\]

**Lemma D.3.** If Assumption G is true, then

\[
\sup_{\gamma \in \Gamma} ||\tilde{M}_n^*(\gamma) - M_0(\gamma)|| \xrightarrow{P^*} 0 \text{ in } P.
\]

**Proof.** It is shown that the classes \( \{z_{it}(1, x_{it}^*1\{q_{it} > \gamma\} : \gamma \in \Gamma \} \text{ and } \{z_{it}(1, x_{it-1}^*1\{q_{it-1} > \gamma\} : \gamma \in \Gamma \} \text{ are P-Glivenko-Cantelli in the proof of Lemma C.2. Then, by bootstrap Glivenko-Cantelli theorem, e.g. Lemma 3.6.16 in van der Vaart and Wellner (1996), the result of this lemma holds.}

**Lemma D.4.** Let Assumption G hold. If \( h_n \to 0 \), then

\[
\sup_{||\theta_1 - \theta_2|| < h_n} \sqrt{n}||\tilde{g}_n^*(\theta_1) - \tilde{g}_n^*(\theta_2) - \tilde{g}_n(\theta_1) + \tilde{g}_n(\theta_2)|| = o^*_n(1) \text{ in } P.
\]

**Proof.** Note that \( g^*_i(\theta_1) - g^*_i(\theta_2) = g(\omega^*_i, \theta_1) - g(\omega^*_i, \theta_2) \) for any \( \theta_1 \) and \( \theta_2 \) where \( g(\omega_i, \theta) \) is defined by (C.4), and \( \omega^*_i \) is a resampling draw from \( \{\omega_i : i = 1, \ldots, n\} \). Hence, \( \tilde{g}_n^*(\theta_1) - \tilde{g}_n^*(\theta_2) - \tilde{g}_n(\theta_1) + \tilde{g}_n(\theta_2) = \frac{1}{n} \sum_{i=1}^n [g(\omega_i^*, \theta_1) - \frac{1}{n} \sum_{i=1}^n g(\omega_i, \theta_1)] - \frac{1}{n} \sum_{i=1}^n [g(\omega_i^*, \theta_2) - \frac{1}{n} \sum_{i=1}^n g(\omega_i, \theta_2)] \).

By bootstrap version of stochastic equicontinuity, e.g. C2 in the proof of Theorem 2.1 in
Praestgaard and Wellner (1993), the result of this lemma holds if \( \{g(\omega, \theta) : \theta \in \Theta\} \) satisfies the uniform entropy condition and has a square integrable envelope function, which are verified in the proof of Lemma C.3.

**Lemma D.5.** Suppose that Assumptions G, D, and LK hold, and the true model is continuous. If \( \delta_{20}^* = O_p(n^{-1/2}), \delta_{30}^* - \delta_{30} = O_p(n^{-1/2}), \gamma_0^* - \gamma_0 = O_p(n^{-1/4}), \) and \( \delta_{10}^* + \delta_{30}^* \gamma_0^* = O_p(n^{-1/2}), \) then

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}(1_{it}(\gamma_0^*)' - 1_{it}(\gamma_0^* + \frac{b}{n^2})')X_{it}^\delta_0 \to_{\mathbb{P}} \frac{\delta_{30}^*}{2} \left\{ E_t[z_{it}|\gamma_0]f_t(\gamma_0) - E_{t-1}[z_{it}|\gamma_0]f_{t-1}(\gamma_0) \right\} b^2,
\]

in \( \mathbb{P} \) uniformly with respect to \( b \in [-K, K] \) for any \( K < \infty \).

The conditions for \( \delta_0^* \) and \( \gamma_0^* \) hold if (i) \( \theta_0^* = (\hat{\alpha}(\gamma_0^*), \gamma_0^*) \), (ii) \( \theta_0^* \) is set as (8), and (iii) \( \delta_0^* = \hat{\delta} \), which is the continuity-restricted estimator in Section 3.2, under the assumptions of this lemma. For (i), \( \sqrt{n}(\hat{\alpha}(\gamma_0^*) - \alpha_0) \) is asymptotically normal, and \( \hat{\delta}(\gamma_0) - \delta_{10}(\gamma_0) - \delta_{30}(\gamma_0) - \gamma_0 = O_p(n^{-1/2}) \). For (ii), note that \( w_n = O_p(n^{-1/4}) \). \( \delta_{10}^* + \delta_{30}^* \gamma_0^* = w_n(\hat{\delta}(\gamma_0) + \delta_{30}(\gamma_0)) = O_p(n^{-1/2}), (1 - w_n)(\hat{\delta}(\gamma_0) + \delta_{30}(\gamma_0)) = 0, \) and \( (1 - w_n)w_n(\hat{\delta}(\gamma_0) + \delta_{30}(\gamma_0)) = O_p(n^{-1/4})O_p(n^{-1/2})O_p(n^{-1/4}) \). \( \delta_{20}^* = w_n\delta_{20} = O_p(n^{-3/4}), \) and \( \delta_{30}^* - \delta_{30} = w_n(\delta_{30}(\gamma_0) - \delta_{30}(\gamma_0)) = O_p(n^{-3/4})O_p(n^{-1/4}) + O_p(n^{-1/2}) \). \( \gamma_0^* - \gamma_0 = w_n(\gamma_0 - \gamma_0) + (1 - w_n)(\gamma_0 - \gamma_0) = O_p(n^{-1/4})O_p(n^{-1/4}) + O_p(n^{-1/2}) \) also holds. For (iii), Seo et al. (2019) showed that \( \hat{\theta} - \theta_0 = O_p(n^{-1/2}) \), while \( \hat{\delta} + \delta_{30}(\gamma_0) = 0 \) and \( \delta_{20} = 0_p - 1 \) by definition.

**Proof.** Note that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}(1_{it}(\gamma_0^*)' - 1_{it}(\gamma_0^* + \frac{b}{n^2})')X_{it}^\delta_0 \to_{\mathbb{P}} \frac{\delta_{30}^*}{2} \left\{ E_t[z_{it}|\gamma_0]f_t(\gamma_0) - E_{t-1}[z_{it}|\gamma_0]f_{t-1}(\gamma_0) \right\} b^2,
\]

in \( \mathbb{P} \) uniformly with respect to \( b \in [-K, K] \) for any \( K < \infty \).
is $o_p^*(1)$ in $P$. Hence, by plugging in $\theta_0^*$ to the place of $\theta$ in the last display, we can derive that (D.2) is $o_p^*(1)$ in $P$ uniformly with respect to $b \in [-K, K]$.

Next, we show that (D.3) term converges to a deterministic limit. As $\{z_{it}1_{it}(\gamma')X_{it}\delta : \theta \in \Theta, |\theta| \leq K\}$ satisfies the uniform entropy condition and has a square integrable envelope function, we can derive the following asymptotic equicontinuity:

$$\sup_{b \in [-K,K]} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}(1_{it}(\gamma')' - 1_{it}(\gamma + \frac{b}{n^{\frac{1}{4}}})'X_{it}\delta - \sqrt{n}E[z_{it}(1_{it}(\gamma')' - 1_{it}(\gamma + \frac{b}{n^{\frac{1}{4}}})')X_{it}\delta] \right\|$$

is $o_p(1)$, and hence $o_p^*(1)$ in $P$ by Lemma B.1. Therefore,

$$\sup_{b \in [-K,K]} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}^*(1_{it}(\gamma')' - 1_{it}^*(\gamma + \frac{b}{n^{\frac{1}{4}}})'X_{it}\delta - \sqrt{n}E[z_{it}(1_{it}(\gamma')' - 1_{it}(\gamma + \frac{b}{n^{\frac{1}{4}}})')X_{it}\delta] \right\|$$

is $o_p^*(1)$ in $P$.

Let $J_n(\delta, \gamma, b) = \sqrt{n}E[z_{it}(1_{it}(\gamma')' - 1_{it}(\gamma + \frac{b}{n^{\frac{1}{4}}})')X_{it}\delta]$. By assumption, we can reparametrize such that $\delta_{20}^* = \frac{r_\delta}{\sqrt{n}}$, $\delta_{30}^* = \delta_3 + \frac{r_\delta}{\sqrt{n}}$, $\gamma_0^* = \gamma_0 + \frac{r_\gamma}{n^{\frac{1}{4}}}$, and $\delta_{10}^* = -\delta_{10}^* - \frac{r_{\delta} + r_{\gamma}}{\sqrt{n}} = \delta_{10} - \delta_{30} r_{\gamma} - \gamma_0 \frac{r_{\delta}}{\sqrt{n}} = \frac{r_{\delta}}{\sqrt{n}} - \frac{r_{\delta} + r_{\gamma}}{n^{\frac{1}{4}}\sqrt{n}}$. Then, we can reparametrize the function $J_n$ such that

$$\tilde{J}_n(r_{\delta_1 + \delta_3 \gamma}, r_{\delta_2}, r_{\delta_3}, r_{\gamma}, b) = J_n(\delta_{10} - \delta_{30} r_{\gamma} - \gamma_0 \frac{r_{\delta}}{\sqrt{n}} - \frac{r_{\delta} + r_{\gamma}}{n^{\frac{1}{4}}\sqrt{n}}, \delta_3 + \frac{r_\delta}{\sqrt{n}}, \gamma_0 + \frac{r_\gamma}{n^{\frac{1}{4}}})$$

(D.4)

Let $r = (r_{\delta_1 + \delta_3 \gamma}, r_{\delta_2}, r_{\delta_3}, r_{\gamma})$ which lies in a compact set $\mathcal{R} = \{r \in \mathbb{R}^{p+2} : \|r\| \leq K\}$ for an arbitrary $K < \infty$.

To prove the lemma, it will be shown below that

$$\tilde{J}_n(r_{\delta_1 + \delta_3 \gamma}, r_{\delta_2}, r_{\delta_3}, r_{\gamma}, b) \to \frac{\delta_{30}}{2} \{E[z_{it}(1_{it}(\gamma_0')' - 1_{it}(\gamma_0 + \frac{b}{n^{\frac{1}{4}}})')X_{it}\delta_0^*] = E_0[z_{it}(1_{it}(\gamma_0')' - 1_{it}(\gamma_0 + \frac{b}{n^{\frac{1}{4}}})')X_{it}\delta_0^*] - J_n(\delta_0^*, \gamma_0, b) \}$$

uniformly with respect to $r \in \mathcal{R}$ and $b \in [-K, K]$, which in turn implies

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}^*(1_{it}(\gamma_0')' - 1_{it}(\gamma_0 + \frac{b}{n^{\frac{1}{4}}})')X_{it}\delta_0^* \to \frac{\delta_{30}}{2} \{E[z_{it}(1_{it}(\gamma_0')' - 1_{it}(\gamma_0 + \frac{b}{n^{\frac{1}{4}}})')X_{it}\delta_0^*] - J_n(\delta_0^*, \gamma_0, b) \}$$

in $P$ uniformly with respect to $b \in [-K, K]$ since

$$\sup_{b \in [-K,K]} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}^*(1_{it}(\gamma_0')' - 1_{it}(\gamma_0 + \frac{b}{n^{\frac{1}{4}}})')X_{it}\delta_0^* - J_n(\delta_0^*, \gamma_0, b) \right\| = o_p^*(1)$$

in $P$.

Suppose $b > 0$. The case for $b < 0$ follows similarly. Note that

$$\sqrt{n}E[z_{it}(1_{it}(\gamma')' - 1_{it}(\gamma + \frac{b}{n^{\frac{1}{4}}})')X_{it}\delta] = \sqrt{n}E[z_{it}(1_{it}(x_{it})')\delta 1\{\gamma + \frac{b}{n^{\frac{1}{4}}} \geq q_{it} > \gamma\}]$$

$$- \sqrt{n}E[z_{it}(1_{it}(x_{it-1})')\delta 1\{\gamma + \frac{b}{n^{\frac{1}{4}}} \geq q_{it-1} > \gamma\}]$$

S-19
We focus on the first term on the right hand side \(\sqrt{n}E[z_{it}(1, x'_{it})\delta 1\{\gamma + \frac{b}{n} \geq q_{it} > \gamma}\) since the limit of the second term can be analyzed similarly, and redefine \(J_n(\delta, \gamma, b) = \sqrt{n}E[z_{it}(1, x'_{it})\delta 1\{\gamma + \frac{b}{n} \geq q_{it} > \gamma}\) and \(\tilde{J}_n\), accordingly. Let \(x_{it} = (\xi_{it}', q_{it})'\) where \(\xi_{it} \in \mathbb{R}^{p-1}\). Then, \(J_n(\delta, \gamma, b) = J_{1n}(\delta, \gamma, b) + J_{2n}(\delta, \gamma, b)\) where

\[
J_{1n}(\delta, \gamma, b) = \sqrt{n}E[z_{it}\xi'_{it}\delta 1\{\gamma + \frac{b}{n} \geq q_{it} > \gamma}\}], \quad \text{and}
\[
J_{2n}(\delta, \gamma, b) = \sqrt{n}E[z_{it}(\delta_1 + \delta_3 q_{it})1\{\gamma + \frac{b}{n} \geq q_{it} > \gamma}\}].
\]

Similarly to \(\tilde{J}_n\) in (D.4), we define reparametrized function \(\tilde{J}_{1n}\) and \(\tilde{J}_{2n}\).

**Limit of \(\tilde{J}_{1n}\):** We can derive the Taylor expansion

\[
\tilde{J}_{1n}(r, b) = E[z_{it}\xi'_{it}r_{\delta 2}1\{\gamma + \frac{b+r_{\gamma}}{n} \geq q_{it} > \gamma + \frac{r_{\gamma}}{n}\}] = E_t[z_{it}\xi'_{it}r_{\delta 2}\gamma, b]f_t(\gamma, b),
\]

where \(\gamma, b \in [\gamma + \frac{r_{\gamma}}{n}, \gamma + \frac{r_{\gamma}}{n}]\). As both \(r,\gamma, b\) are in compact spaces, \(\gamma, b \to \gamma_0\) uniformly with respect to \(r, \gamma\) and \(b\). By Assumption D, \(E_t[z_{it}\xi'_{it}\gamma]f_t(\gamma)\) is bounded and continuous on a neighborhood \(\mathcal{O}\) of \(\gamma_0\). Therefore, \(E_t[z_{it}\xi'_{it}\gamma, b]f_t(\gamma, b) \to E_t[z_{it}\xi'_{it}\gamma]f_t(\gamma)\). Since \(r_{\delta 2} \to 0\), we can derive \(\tilde{J}_{1n}(r, b) \to 0\) uniformly in \(r\) and \(b\).

**Limit of \(\tilde{J}_{2n}\):** We can derive the Taylor expansion

\[
\tilde{J}_{2n}(r, b) = \sqrt{n}E[z_{it}(\delta_1 - \delta_3 q_{it}) + (\delta_3 q_{it} - \delta_3)\gamma, b]f_t(\gamma, b),
\]

\[
= \frac{r_{\delta 1} + r_{\delta 3}}{n^{3/4}} E_t[z_{it}\gamma + \frac{r_{\gamma}}{n\delta}]f_t(\gamma + \frac{r_{\gamma}}{n\delta})b + \frac{b^2}{2} E_t[z_{it}(\gamma, b - \delta_3)\gamma, b]f_t(\gamma, b), \quad (D.5)
\]

\[
= \frac{r_{\delta 1} + r_{\delta 3}}{n^{3/4}} E_t[z_{it}\gamma + \frac{r_{\gamma}}{n\delta}]f_t(\gamma + \frac{r_{\gamma}}{n\delta})b + \frac{b^2}{2} E_t[z_{it}(\gamma, b - \delta_3)\gamma, b]f_t(\gamma, b), \quad (D.6)
\]

where \(\gamma, b \in [\gamma + \frac{r_{\gamma}}{n\delta}, \gamma + \frac{r_{\gamma}}{n\delta}]\).

First, we can observe that (D.5) converges to zero uniformly with respect to \(r_{\delta 1}, \delta_3\gamma, r_{\gamma}, b\). This is because \(\gamma, b \to \gamma_0\) uniformly with respect to \(r, \gamma\) and \(b\), which implies \(E_t[z_{it}\gamma + \frac{r_{\gamma}}{n\delta}]f_t(\gamma + \frac{r_{\gamma}}{n\delta})b \to 0\). Next, we check that (D.6) converges to zero uniformly with respect to \(r_{\delta 1}, \delta_3\gamma, r_{\gamma}, b\). By Assumption D, \(\frac{d}{d\gamma}(E_t[z_{it}\gamma]f_t(\gamma))\) is bounded and continuous on a neighborhood \(\mathcal{O}\) of \(\gamma_0\). As \(\gamma, b \to \gamma_0\) uniformly with respect to \(r, \gamma, b\), \(\frac{d}{d\gamma}(E_t[z_{it}\gamma]f_t(\gamma))|_{\gamma = \gamma, b} \to \frac{d}{d\gamma}(E_t[z_{it}\gamma]f_t(\gamma))|_{\gamma = \gamma_0}\) and \((\frac{r_{\delta 1} + r_{\delta 3}}{n^{3/4}} + (\delta_3 + \delta_3)\gamma, b - \delta_3)\gamma, b) \to 0\), which implies the convergence of (D.6) to zero.

Finally, we obtain the limit of (D.7). Since \(E_t[z_{it}\gamma, b]f_t(\gamma, b) \to E_t[z_{it}\gamma]f_t(\gamma)\) and \(\frac{r_{\delta 3}}{n} \to 0\), (D.7) converges to \(\frac{3}{2} E_t[z_{it}\gamma]f_t(\gamma)b^2\) uniformly with respect to \(r \in \mathcal{R}\) and \(b \in [-K, K]\).
In conclusion,

\[ \tilde{J}_n(r, b) \rightarrow \frac{\delta_{30}}{2} E_t[z_{it}|\gamma_0]f_t(\gamma_0)b^2 \]

uniformly with respect to \( r \in \mathcal{R} \) and \( b \in [-K, K] \), and hence

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}^*(1, x_{it}')\delta_0^* \{ \gamma_0^* + \frac{b}{\sqrt{n}} \geq q_{it}' > \gamma_0^* \} \xrightarrow{p^*} \frac{\delta_{30}}{2} E_t[z_{it}|\gamma_0]f_t(\gamma_0)b^2 \quad \text{in } P \]

uniformly with respect to \( b \in [-K, K] \). Similarly, we can show that

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}^*(1, x_{it-1}')\delta_0^* \{ \gamma_0^* + \frac{b}{\sqrt{n}} \geq q_{it-1}' > \gamma_0^* \} \xrightarrow{p^*} \frac{\delta_{30}}{2} E_{t-1}[z_{it}|\gamma_0]f_{t-1}(\gamma_0)b^2 \quad \text{in } P \]

uniformly with respect to \( b \in [-K, K] \).

**Lemma D.6.** Suppose that Assumptions G, D, and LJ hold, and the true model is discontinuous. If \( \delta_0^* - \delta_0 = O_p(n^{-1/2}) \) and \( \gamma_0^* - \gamma_0 = O_p(n^{-1/2}) \), then

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}^*(1, x_{it}')\delta_0^* \{ \gamma_0^* + \frac{b}{\sqrt{n}} \geq q_{it}' > \gamma_0^* \} X_{it}^* \delta_0^* \xrightarrow{p} \left\{ E_t[z_{it}(1, x_{it}')\delta_0^*|\gamma_0]f_t(\gamma_0) - E_{t-1}[z_{it}(1, x_{it-1}')\delta_0^*|\gamma_0]f_{t-1}(\gamma_0) \right\} b, \]

in \( P \) uniformly with respect to \( b \in [-K, K] \) for any \( K < \infty \).

The conditions for \( \delta_0^* \) and \( \gamma_0^* \) hold if (i) \( \theta_0^* = (\hat{\alpha}(\gamma_0)', \gamma_0)' \) or (ii) \( \theta_0^* \) is set as (8) under the assumptions of this lemma. Note that \( \delta_0^* = w_n \delta + (1 - w_n)\hat{\delta} = \delta_0 + O_p(n^{-1/2}) \) since \( w_n \xrightarrow{P} 1 \), \( \hat{\delta} = \delta_0 + O_p(n^{-1/2}) \), and \( \hat{\delta} = O_p(1) \) by Assumption P.

If \( M_0(\gamma) \) has full column rank for all \( \gamma \in \Gamma \), then Assumption P holds. Let \( \tilde{M}_n(\gamma) = \left[ \tilde{M}_{1n}(\gamma) \; \tilde{M}_{2n}(\gamma) \right] \), \( \tilde{M}_{2n}(\gamma) = \tilde{M}_{2n}(\gamma) (-\gamma, 0_{p-1}, 1)' \), \( \tilde{M}_0(\gamma) = \left[ \tilde{M}_{10}(\gamma) \; \tilde{M}_{20}(\gamma) \right] \), and \( \tilde{M}_{20}(\gamma) = M_{20}(\gamma) (-\gamma, 0_{p-1}, 1)' \). Note that \( \tilde{\alpha}(\gamma) = -(\tilde{M}_n(\gamma)'W_nM_n(\gamma))^{-1}\tilde{M}_n(\gamma)'W_nv_n, \) where \( \tilde{\alpha}(\gamma) = \arg \min_{\alpha, \delta_0} \delta_0^* \rightarrow_{P} -M_0\alpha_0 \) and \( \sup_{\gamma \in \Gamma} \| \tilde{M}_n(\gamma) - M_0(\gamma) \| \rightarrow_{P} 0 \) by Lemma C.2, \( \tilde{\alpha}(\gamma) \rightarrow_{P} (M_0(\gamma)'\Omega^{-1}M_0(\gamma))^{-1}M_0(\gamma)\Omega^{-1}M_0\alpha_0 \) uniformly with respect to \( \gamma \). Since \( \Gamma \) is compact, there exists \( C < \infty \) such that \( \sup_{\gamma \in \Gamma} \| (M_0(\gamma)'\Omega^{-1}M_0(\gamma))^{-1}M_0(\gamma)\Omega^{-1}M_0\alpha_0 \| < C \). As \( \hat{\gamma} \in \Gamma, P(\| \hat{\alpha} \| \geq C) \rightarrow 0 \) holds, which implies \( \hat{\delta} = O_p(1) \).

**Proof.** By similar arguments used in the proof of **Lemma D.5**, we can derive that

\[ \sup_{b \in [-K, K]} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}^*(1, x_{it}')\delta_0^* \{ \gamma_0^* + \frac{b}{\sqrt{n}} \geq q_{it}' > \gamma_0^* \} X_{it}^* \delta - \sqrt{n}E[z_{it}(1, x_{it}') - 1_{it}(\gamma + \frac{b}{\sqrt{n}}')X_{it}] \right\| \]

is \( o_p(1) \) in \( P \).

Let \( J_n(\delta, \gamma, b) = \sqrt{n}E[z_{it}(1, x_{it}') - 1_{it}(\gamma + \frac{b}{\sqrt{n}}')X_{it}] \). By assumption, we can reparametrize such that \( \delta_0^* = \delta_0 + \frac{\tilde{\alpha}_0}{\sqrt{n}} \) and \( \gamma_0^* = \gamma_0 + \frac{\tilde{\alpha}_0}{\sqrt{n}} \). Then, we can reparametrize the function \( J_n \)
such that \( \tilde{J}_n(r, \gamma) = J_n(\delta_0 + \frac{r}{\sqrt{n}}, \gamma_0 + \frac{r}{\sqrt{n}}) \). Let \( r = (r_\delta, r_\gamma) \) which lies in a compact set \( \mathcal{R} = \{ r \in \mathbb{R}^{p+2} : \| r \| \leq \mathcal{K} \} \) for an arbitrary \( \mathcal{K} < \infty \).

To prove the lemma, it will be shown that

\[
\tilde{J}_n(r_\delta, r_\gamma, b) \rightarrow \{ E_i[z_{it}(1, x'_{it})\delta_0|\gamma_0]f_1(\gamma_0) - E_{t-1}[z_{it}(1, x'_{it-1})\delta_0|\gamma_0]f_{t-1}(\gamma_0) \} b
\]

uniformly with respect to \( r \in \mathcal{R} \) and \( b \in [-K, K] \), which in turn implies

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}'(1_\delta(\gamma_0)' - 1_\delta(\gamma_0) + \frac{b}{\sqrt{n}})X_{it}\delta_0
\]

\[
p^* \{ E_i[z_{it}(1, x'_{it})\delta_0|\gamma_0]f_1(\gamma_0) - E_{t-1}[z_{it}(1, x'_{it-1})\delta_0|\gamma_0]f_{t-1}(\gamma_0) \} b \quad \text{in } P
\]

uniformly with respect to \( b \in [-K, K] \) since

\[
\sup_{b \in [-K, K]} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}'(1_\delta(\gamma_0)' - 1_\delta(\gamma_0) + \frac{b}{\sqrt{n}})X_{it}\delta_0 - J_n(\delta_0, \gamma_0, b) \right\| = o_p(1) \quad \text{in } P.
\]

Suppose \( b > 0 \). The case for \( b < 0 \) follows similarly. Then,

\[
\sqrt{n}E[z_{it}(1, x'_{it})' - 1_\delta(\gamma) + \frac{b}{\sqrt{n}})X_{it}\delta] = \sqrt{n}E[z_{it}(1, x'_{it})\delta(\gamma + \frac{b}{\sqrt{n}} \geq q_\delta > \gamma)]
\]

\[
- \sqrt{n}E[z_{it}(1, x'_{it-1})\delta(\gamma + \frac{b}{\sqrt{n}} \geq q_\delta > \gamma)]
\]

We focus on the first term of the right hand side \( \sqrt{n}E[z_{it}(1, x'_{it})\delta(\gamma + \frac{b}{\sqrt{n}} \geq q_\delta > \gamma)] \) as the limit of the second term can be derived identically, and redefine \( J_n(\delta_0, \gamma_0, b) = \sqrt{n}E[z_{it}(1, x'_{it})\delta(\gamma + \frac{b}{\sqrt{n}} \geq q_\delta > \gamma)] \) and \( \tilde{J}_n \), accordingly.

We can derive the following Taylor expansion:

\[
\tilde{J}_n(r, b) = \sqrt{n}E[z_{it}(1, x'_{it})\delta(\gamma_0 + \frac{r_\delta}{\sqrt{n}}, \gamma_0 + \frac{r_\gamma}{\sqrt{n}} \geq q_\delta > \gamma_0 + \frac{r_\gamma}{\sqrt{n}})] = E_i[z_{it}(1, x'_{it})\delta(\gamma_0 + \frac{r_\delta}{\sqrt{n}}, \gamma_0, b)] f_1(\gamma_0, b),
\]

where \( \gamma_{n, b} \in [\gamma_0 + \frac{r_\delta}{\sqrt{n}}, \gamma_0 + \frac{b + r_\gamma}{\sqrt{n}}] \). As \( \gamma_{n, b} \rightarrow \gamma_0 \) uniformly with respect to \( r \in \mathcal{R} \) and \( b \in [-K, K], E_i[z_{it}(1, x'_{it})\delta(\gamma_0 + \frac{r_\delta}{\sqrt{n}})|\gamma_{n, b}, f_1(\gamma_{n, b}) b \rightarrow E_i[z_{it}(1, x'_{it})\delta_0|\gamma_0]f_1(\gamma_0) b \) uniformly, and hence \( \tilde{J}_n(r, b) \rightarrow E_i[z_{it}(1, x'_{it})\delta_0|\gamma_0]f_1(\gamma_0) b \) uniformly.

In conclusion,

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}'(1, x'_{it})\delta_0 1\{ \gamma_0 + \frac{b}{\sqrt{n}} \geq q_\delta > \gamma_0 \} \xrightarrow{p^*} E_i[z_{it}(1, x'_{it})\delta_0|\gamma_0]f_1(\gamma_0) b \quad \text{in } P
\]

uniformly with respect to \( b \in [-K, K] \). Similarly, we can show that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}'(1, x'_{it-1})\delta_0 1\{ \gamma_0 + \frac{b}{\sqrt{n}} \geq q_\delta-1 > \gamma_0 \} \xrightarrow{p^*} E_{t-1}[z_{it}(1, x'_{it-1})\delta_0|\gamma_0]f_{t-1}(\gamma_0) b \quad \text{in } P
\]

uniformly with respect to \( b \in [-K, K] \).
E Invalidity of standard nonparametric bootstrap

In this section, we explain why the bootstrap estimators of the standard bootstrap does not have the asymptotic distribution in Theorem 2 when the true model is continuous. Note that the bootstrap explained by Algorithm 1 becomes the standard nonparametric bootstrap when \( \theta_0^* = \hat{\theta} \). The consistency and convergence rate derivations in the proof of Proposition 1 can still be followed, and hence \( \sqrt{n}(\hat{\gamma} - \gamma) = O_p(1) \) and \( \sqrt{n}(\hat{\gamma} - \gamma)^2 = O_p(1) \) both in \( P \). However, the conditions for Lemma D.5 do not hold for the standard nonparametric bootstrap as \( n^{1/4}(\hat{\gamma}_1 + \hat{\gamma}_3) \) is not \( O_P(1) \) as explained in Section 4.2. Therefore, the rescaled versions of the criterion converges to a different limit. Specifically,

\[
\sqrt{n}\gamma_n^*(\hat{\alpha} + a_n\gamma + b_n\gamma) - n^{1/4}G(\hat{\theta})b \Rightarrow M_0a + Hb^2 - e
\]

in \( \ell^\infty(\mathbb{K}) \) in \( P \) for every compact \( \mathbb{K} \) in the Euclidean space, where \( G(\theta) \) is defined by (11).

Recall that \( n^{1/4}G(\hat{\theta}) \neq O_P(1) \) as shown in Section 4.2. The conditional weak convergence, \( \Rightarrow^* \), in the last display comes from applying the following Lemma E.1 in the place of Lemma D.5 used in the proof of Theorem 6.

**Lemma E.1.** Suppose that Assumptions G, D, LK are true and that the true model is continuous. Then,

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i^*(\gamma_i)' - 1_i(\gamma + \frac{b}{n^{1/4}})'X_i^*\delta - \sqrt{n}E[z_i(\gamma_i)' - 1_i(\gamma + \frac{b}{n^{1/4}})']X_i\delta - \frac{30}{2} E[z_i\gamma_i]f_i(\gamma_0) - E[z_i\gamma_i]f_i(\gamma_0) \to 0
\]

in \( P \) uniformly with respect to \( b \in [-K, K] \) for any \( K < \infty \).

**Proof.** By similar arguments used in the proof of Lemma D.5, we can derive that

\[
\sup_{b \in [-K, K]} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i^*(\gamma_i)' - 1_i(\gamma + \frac{b}{n^{1/4}})'X_i^*\delta - \sqrt{n}E[z_i(\gamma_i)' - 1_i(\gamma + \frac{b}{n^{1/4}})']X_i\delta \right\|
\]

is \( o_P^*(1) \) in \( P \).

Suppose that \( b > 0 \). The \( b < 0 \) case can be analyzed similarly. Let \( J_n(\delta, \gamma, b) = \sqrt{n}E[z_i(1, x_i')^\delta_1 \{ \gamma + \frac{b}{n^{1/4}} \geq q_i \} \gamma] - n^{1/4}(\delta_1 + \delta_3 \gamma)E[z_i\gamma_i]f_i(\gamma_0)b \). Reparametrize such that \( \hat{\gamma} = \gamma_0 + \frac{r_\gamma}{\sqrt{n}} \) and \( \hat{\delta} = \delta_0 + \frac{r_\delta}{\sqrt{n}} \). Let the set of \( r = (r_\delta, r_\gamma) \) be \( \mathcal{R} = \{ r \in \mathbb{R}^{p+2} : \| r \| \leq K \} \) for arbitrary \( K < \infty \). Let \( \hat{J}_n(r, b) = J_n(\delta_0 + \frac{r_\delta}{\sqrt{n}}, \gamma_0 + \frac{r_\gamma}{\sqrt{n}}, b) \).

We will show that \( \hat{J}_n(r, b) \to \frac{\delta_0}{2} E[z_i\gamma_i]f_i(\gamma_0)b^2 \) uniformly with respect to \( r \in \mathcal{R} \) and \( b \in [-K, K] \), which implies

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i^*(1, x_i') \hat{\delta}_1 \{ \gamma + \frac{b}{n^{1/4}} \geq q_i \} \gamma - n^{1/4}(\hat{\delta}_1 + \hat{\delta}_3 \gamma)E[z_i\gamma_i]f_i(\gamma_0)b \to \frac{\delta_0}{2} E[z_i\gamma_i]f_i(\gamma_0)b^2
\]
in $P$ uniformly with respect to $b \in [-K, K]$, because

$$
\sup_{b \in [-K, K]}\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}(1, x_{it}') \delta \hat{\gamma} + \frac{b}{n^{1/4}} \geq q_{it} > \hat{\gamma} \right| = o_p^*(1) \text{ in } P.
$$

Note that $J_n(\delta, \gamma, b) = J_{1n}(\delta, \gamma, b) + J_{2n}(\delta, \gamma, b)$ where

\begin{align*}
J_{1n}(\delta, \gamma, b) &= \sqrt{n}E[|z_{it}\delta_{i'}|\delta_{2}1\{\gamma + \frac{b}{n^{1/4}} \geq q_{it} > \gamma\}], \text{ and} \\
J_{2n}(\delta, \gamma, b) &= \sqrt{n}E[|z_{it}(\delta_{10} + \delta_{30}q_{it})|\delta_{2}1\{\gamma + \frac{b}{n^{1/4}} \geq q_{it} > \gamma\}] - n^{1/4}(\delta_{1} + \delta_{3}\gamma)E_{t}[z_{it}\gamma_{0}]f_t(\gamma_{0})b.
\end{align*}

Let $\tilde{J}_{1n}$ and $\tilde{J}_{2n}$ denote the reparametrized version of $J_{1n}$ and $J_{2n}$, respectively.

$\tilde{J}_{1n}(r, b)$ converges to zero uniformly, for which we recall that it is identical to $\tilde{J}_{1n}$ that appears in the proof of Lemma D.5.

$\tilde{J}_{2n}(r, b) = \tilde{J}_{2an}(r, b) + \tilde{J}_{2bn}(r, b)$ where

\begin{align*}
\tilde{J}_{2an}(r, b) &= E[z_{it}(\delta_{i} + r_{i}q_{it})1\{\gamma_{0} + \frac{r_{i}q_{it}}{n^{1/4}} \geq q_{it} > \gamma_{0} + \frac{r_{i}q_{it}}{n^{1/4}}\}], \text{ and} \\
\tilde{J}_{2bn}(r, b) &= \sqrt{n}E[|z_{it}(\delta_{10} + \delta_{30}q_{it})|\delta_{2}1\{\gamma_{0} + \frac{r_{i}q_{it}}{n^{1/4}} \geq q_{it} > \gamma_{0} + \frac{r_{i}q_{it}}{n^{1/4}}\}] \\
&\quad - n^{1/4}(\delta_{1} + \delta_{3}\gamma)E_{t}[z_{it}\gamma_{0}]f_t(\gamma_{0})b.
\end{align*}

It can be easily checked that $\tilde{J}_{2an}(r, b)$ converges to zero uniformly. It will be shown in the next paragraph that $\tilde{J}_{2bn}(r, b) \to \frac{b^{2}n^{1/2}}{2}E_{t}[z_{it}|\gamma_{0}]f_t(\gamma_{0})b^{2}$ uniformly, which implies $\tilde{J}_{n}(r, b) \to \frac{b^{2}n^{1/2}}{2}E_{t}[z_{it}|\gamma_{0}]f_t(\gamma_{0})b^{2}$ uniformly.

By Taylor expansion,

\begin{align*}
\tilde{J}_{2bn}(r, b) &= \sqrt{n}E[|z_{it}(\delta_{10} + \delta_{30}q_{it})|\delta_{2}1\{\gamma_{0} + \frac{r_{i}q_{it}}{n^{1/4}} \geq q_{it} > \gamma_{0} + \frac{r_{i}q_{it}}{n^{1/4}}\}] - n^{1/4}(\delta_{1} + \delta_{3}\gamma)E_{t}[z_{it}\gamma_{0}]f_t(\gamma_{0})b \\
&= \delta_{30}\gamma_{0}E_{t}[z_{it}|\gamma_{0}]f_t(\gamma_{0})b - \frac{b^{2}n^{1/2}}{2}(\delta_{10} + \delta_{30}\gamma_{n, b})d\frac{d}{d\gamma}(E_{t}[z_{it}|\gamma]f_t(\gamma))|_{\gamma = \gamma_{n, b}} \quad \text{and} \quad \delta_{30}E_{t}[z_{it}|\gamma_{n, b}]f_t(\gamma_{n, b}) \quad (E.1) \\
&+ \frac{b^{2}}{2}E_{t}[z_{it}|\gamma_{0}]f_t(\gamma_{0})b^{2} \quad (E.2)
\end{align*}

where $\gamma_{n, b} \in [\gamma_{0} + \frac{r_{i}q_{it}}{n^{1/4}}, \gamma_{0} + \frac{b_{i}r_{i}q_{it}}{n^{1/4}}]$. By continuity of $E_{t}[z_{it}|\gamma]f_t(\gamma)$ at $\gamma = \gamma_{0}$, (E.1) converges to 0 uniformly with respect to $r \in R$ and $b \in [-K, K]$. As $\gamma_{n, b} \to \gamma_{0}$ uniformly, we can derive that (E.2) converges to $\frac{b^{2}n^{1/2}}{2}E_{t}[z_{it}|\gamma_{0}]f_t(\gamma_{0})b^{2}$ uniformly.

By similar manner, we can derive

\begin{align*}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}^{*}(1, x_{it-1}') \delta \hat{\gamma} + \frac{b}{n^{1/4}} \geq q_{it-1}^{*} > \hat{\gamma} \right| - n^{1/4}(\delta_{1} + \delta_{3}\hat{\gamma})E_{t-1}[z_{it}|\gamma_{0}]f_{t-1}(\gamma_{0})b
\end{align*}

S-24
in $P$ uniformly with respect to $b \in [-K,K]$.

## F Percentile bootstrap confidence intervals for empirical application

In this section, we report the percentile residual-bootstrap confidence intervals of the coefficients for the empirical application. Table 7 and Table 8 correspond to Table 5 and Table 6 in Section 6, respectively.

Table 7: The 95% percentile bootstrap confidence intervals that use the 0.025th and 0.975th quantiles of each element of $\sqrt{n}(\hat{\alpha} - \alpha_0^*)$ are reported. Columns (a) and (b) report results of the models (12) and (13), respectively. The percentile of each threshold location value is shown in parentheses below each value. The significance levels for the coefficients are given by stars: * - 10%, ** - 5% and *** - 1%.

|               | (a)                          | (b)                          |
|---------------|------------------------------|------------------------------|
|               | est. [95% CI] |               | est. [95% CI] |               |
| **Lower regime** |                  | **Upper regime** |                  |
| $I_{t-1}$     | 0.778** 0.124 1.154 | $I_{t-1}$     | 0.252 -0.258 0.724 |
| $CF_{t-1}$    | 0.047 -0.034 0.145 | $CF_{t-1}$    | 0.266* -0.003 0.535 |
| $PPE_{t-1}$   | -0.147 -0.385 0.171 | $PPE_{t-1}$   | 0.027 -0.103 0.264 |
| $ROA_{t-1}$   | -0.032 -0.132 0.047 | $ROA_{t-1}$   | -0.017 -0.180 0.090 |
| $LEV_{t-1}$   | 0.231 -0.843 1.849 | $TQ_{t-1}$    | 0.246* -0.031 0.577 |
| **Upper regime** |                  | **Difference between regimes** |                  |
| $I_{t-1}$     | -0.154 -0.717 0.551 | $I_{t-1}$     | 0.410 -0.049 0.751 |
| $CF_{t-1}$    | 0.148 -0.015 0.326 | $CF_{t-1}$    | 0.081** 0.021 0.200 |
| $PPE_{t-1}$   | -0.291* -0.519 0.015 | $PPE_{t-1}$   | 0.044 -0.214 0.398 |
| $ROA_{t-1}$   | 0.013 -0.066 0.113 | $ROA_{t-1}$   | 0.050* -0.019 0.153 |
| $LEV_{t-1}$   | -0.081 -0.234 0.037 | $TQ_{t-1}$    | 0.005 -0.004 0.012 |
| **Difference between regimes** |                  | **Difference between regimes** |                  |
| intercept     | 0.068 -0.024 0.200 | intercept     | 0.236* -0.014 0.580 |
| $I_{t-1}$     | -0.932** -1.830 -0.097 | $I_{t-1}$     | 0.158 -0.559 0.843 |
| $CF_{t-1}$    | 0.101 -0.107 0.322 | $CF_{t-1}$    | -0.185 -0.479 0.108 |
| $PPE_{t-1}$   | -0.144 -0.519 0.134 | $PPE_{t-1}$   | 0.017 -0.227 0.275 |
| $ROA_{t-1}$   | 0.045 -0.111 0.232 | $ROA_{t-1}$   | 0.066 -0.074 0.287 |
| $LEV_{t-1}$   | -0.312* -1.893 0.792 | $TQ_{t-1}$    | -0.242* -0.573 0.038 |
Table 8: The 95% percentile bootstrap confidence intervals that use the 0.025th and 0.975th quantiles of each element of $\sqrt{n}(\hat{\alpha}^* - \alpha_0^*)$ are reported. Results of the model (14) are reported. The percentile of each threshold location value is shown in parentheses below each value. The significance levels for the coefficients are given by stars: * - 10%, ** - 5% and *** - 1%.

| Coefficients | est.  | [95% CI] |
|--------------|------|---------|
| $I_{t-1}$    | 0.392*** | 0.304 0.539 |
| $CF_{t-1}$   | 0.122*** | 0.084 0.154 |
| $PPE_{t-1}$  | 0.076   | -0.027 0.271 |
| $ROA_{t-1}$  | 0.027*** | 0.006 0.046 |
| $TQ_{t-1}\{TQ_{t-1} \leq \gamma\}$ | 0.298** | 0.073 0.571 |
| $TQ_{t-1}\{TQ_{t-1} > \gamma\}$ | 0.008** | 0.001 0.015 |
| Difference between regimes | 0.275** | 0.074 0.566 |
| $TQ_{t-1}$   | -0.290** | -0.566 -0.061 |

G Bootstrap for linearity test

We explain the bootstrap for linearity test based on sup-Wald statistic, explained in Seo and Shin (2016). Null hypothesis of the test is $\delta = 0_{p+1}$. The sup-Wald test statistic is

$$
\sup_{\gamma \in \Gamma} \{n\hat{\delta}(\gamma)'[B'(M_n(\gamma)'W_n(\gamma)\hat{M}_n(\gamma))^{-1}M_n(\gamma)'W_n(\gamma)\hat{\Omega}(\gamma)W_n(\gamma)\hat{M}_n(\gamma)(M_n(\gamma)'W_n(\gamma)\hat{M}_n(\gamma))^{-1}B]^{-1}\hat{\delta}(\gamma)\},
$$

where $B = \left[0_{(p+1) \times p}^T; I_{p+1}\right] \in \mathbb{R}^{(p+1) \times (2p+1)}$, $W_n(\gamma)$ is the weight matrix obtained by the initial estimator with the restriction that the threshold location is $\gamma$, $\hat{\delta}(\gamma)$ is a subvector of the restricted estimator $\hat{\alpha}(\gamma) = (\hat{\beta}(\gamma)', \hat{\delta}(\gamma)')'$, and $\hat{\Omega}(\gamma) = (\frac{1}{n} \sum_{i=1}^n [g_i(\hat{\alpha}(\gamma), \gamma)g_i(\hat{\alpha}(\gamma), \gamma)'] - \frac{1}{n} \sum_{i=1}^n g_i(\hat{\alpha}(\gamma), \gamma)[\frac{1}{n} \sum_{i=1}^n g_i(\hat{\alpha}(\gamma), \gamma)']).$

The bootstrap for the linearity test can be implemented by setting

$$
\beta_0^* = \hat{\beta}, \quad \delta_0^* = 0_{p+1}
$$

in Algorithm 1. Note that $\gamma_0^*$ does not matter in this case as $\delta_0^* = 0_{p+1}$. The critical value for $\tau$-size test is obtained by using the $(1 - \tau)$th quantile of the bootstrapped sup-Wald test statistics, defined analogously to (G.1).

H Uniform validity of the grid bootstrap

In this section, we show the uniform validity of the grid bootstrap given in Section 4.1. As discussed in Section 4.1.1, the following simplified specification is analyzed for the clarity of exposition:

$$
y_{it} = x_{it}'\beta + (\delta_1 + \delta_3q_{it})\{q_{it} > \gamma\} + \eta_i + \epsilon_{it}, \quad t = 1, \ldots, T,
$$
where \( \theta = (\alpha', \gamma)' = (\beta', \delta', \gamma) \), \( \alpha = (\beta', \delta)' \), and \( \delta = (\delta_1, \delta_2)' \in \mathbb{R}^2 \). \( x_{it} = (\xi_{it}', q_{it}') \) still includes the threshold variable. The goal here is to show the uniform validity of the sup-norm over the space of distribution functions as sup conditions for \( \Phi \). Let \( \Phi \) be the square root of the minimum and maximum eigenvalues of \( A' A \), and \( \sigma_{\text{max}}(A) \) be the square root of the minimum and maximum eigenvalues of \( A' A \), respectively. Let the parameter space for \( \phi_{0n} \) be

\[
\Phi_0 = \left\{ \phi_0 \in \Phi : (\delta_{10} + \delta_{30}\gamma_0)^2 + \delta_{30}^2 \geq c_1, \right. \\
c_2 \leq \sigma_{\text{min}}(\Omega) \leq \sigma_{\text{max}}(\Omega) \leq c_3, \\
c_4 \leq E|z_{it}|^{4+r} \leq c_5, \ c_4 \leq E|x_{it}|^{4+r} \leq c_5, \ c_4 \leq E|c_{it}|^{4+r} \leq c_5, \\
\left. f_t(\cdot) \text{ is continuously differentiable at } [\gamma_0 - c_6, \gamma_0 + c_6], \\
c_7 = \min_{q \in [\gamma_0 - c_6, \gamma_0 + c_6]} f_t(q) \leq \max_{q \in [\gamma_0 - c_6, \gamma_0 + c_6]} f_t(q) \leq c_8, \\
\min_{q \in [\gamma_0 - c_6, \gamma_0 + c_6]} |f_t'(q)| \leq c_9, \\
E_t[|z_{it}|^q] \text{ and } E_{t-1}[|z_{it}|^q] \text{ are continuously differentiable at } [\gamma_0 - c_{10}, \gamma_0 + c_{10}], \\
\max_{q \in [\gamma_0 - c_{10}, \gamma_0 + c_{10}]} \|E_t[z_{it}|^q]\| \leq c_{11}, \\
\max_{q \in [\gamma_0 - c_{10}, \gamma_0 + c_{10}]} \|E_{t-1}[z_{it}|^q]\| \leq c_{11}, \\
\max_{q \in [\gamma_0 - c_{10}, \gamma_0 + c_{10}]} \| \frac{d}{dq} (E_t[z_{it}|^q])_{q=q} \| \leq c_{11}, \\
\max_{q \in [\gamma_0 - c_{10}, \gamma_0 + c_{10}]} \| \frac{d}{dq} (E_{t-1}[z_{it}|^q])_{q=q} \| \leq c_{11}, \\
c_{12} \leq \sigma_{\text{min}} \left( \left\| M_0 \ ; \ H \right\| \right) \leq \sigma_{\text{max}} \left( \left\| M_0 \ ; \ H \right\| \right) \leq c_{13}, \\
E_t[|z_{it}|^{1+r} |\gamma_0] \leq c_{14}, \ E_{t-1}[|z_{it}|^{1+r} |\gamma_0] \leq c_{14}, \ \forall t = 1, \ldots, T \right\},
\]

where \( c_1, \ldots, c_{14}, \) and \( r \) are some positive constants. Note that \( (\delta_{10} + \delta_{30}\gamma_0)^2 + \delta_{30}^2 \geq c_1 \) is to prevent \( (\delta_{10n} + \delta_{30n}\gamma_0, \delta_{30n})' \) from (having a subsequence) converging to zero.\(^6\) The remaining conditions for \( \Phi_0 \) other than \( E_t[|z_{it}|^{1+r} |\gamma_0] \leq c_{14} \) and \( E_{t-1}[|z_{it}|^{1+r} |\gamma_0] \leq c_{14} \) imply that Assumptions

\(^5\)That means \( d(F_1, F_2) = \sup_{x \in \mathbb{R}^d} |f_1(x) - f_2(x)| \), where \( f_1 \) and \( f_2 \) are densities of the distribution functions \( F_1 \) and \( F_2 \), and \( d_s \) is a dimension of the random vectors whose distributions are \( F_1 \) or \( F_2 \). It is a stronger norm than the sup-norm over the space of distribution functions as \( \sup_{x \in \mathbb{R}^d} |f_n(x) - f_0(x)| \to 0 \) implies \( \sup_{x \in \mathbb{R}^d} |F_n(x) - F_0(x)| \to 0 \).

\(^6\)This implies that our threshold model has a strong threshold effect which excludes the diminishing or small threshold effect as in Hansen (2000).

S-27
D, G, and LK/LJ hold uniformly. The condition $E_t[\|z_{it}\|^{1+r} | \gamma_0] \leq c_{14}$, $E_{t-1}[\|z_{it}\|^{1+r} | \gamma_0] \leq c_{14}$ is a uniform integrability condition for the distribution of $z_{it}$ conditional on $q_{it}$ or $q_{it-1}$. Its role will be explained after introducing the drifting sequence framework.

Because of the nonlinearity and discontinuity of our dynamic model, it is not trivial to answer what primitive conditions for the parameter and distributions of random variables, such as initial value $y_0$ or individual fixed effect $\eta_i$, are sufficient for $\Phi_0$. This paper does not investigate this issue so that we can focus on uniformity analysis with respect to degeneracy of the Jacobian of nonlinear GMM.

For $n = 1, 2, \ldots$, let $\{\eta_{itn}, y_{0itn}, (z_{itn}, x_{itn}, \epsilon_{itn})_{i=1}^n\}$ be drawn from distribution $F_{0itn}$. For a function or random variable $u$, e.g. $u = z, x$ or $\Delta \epsilon$, we often write $u_{itn}$ and $u_{it-1, n}$ to indicate more explicitly that indices in subscript are $((i, t), n)$ or $((i, t - 1), n)$, while $n$ is the new index introduced in this section. Suppose that

$$y_{itn} = x_{itn}' \beta_0 + (\delta_{10n} + \delta_{30n} q_{itn}) 1\{q_{itn} > \gamma_0\} + \eta_{itn} + \epsilon_{itn},$$

$$E[z_{itn} \Delta \epsilon_{itn}] = 0, \quad \text{where } \Delta \epsilon_{itn} = \epsilon_{itn} - \epsilon_{it-1, n}.$$  

As in Section 2, we define

$$M_{in} = - \begin{bmatrix} z_{itn} \Delta x_{itn} \\ z_{iTn} \Delta x_{iTn} \end{bmatrix} \in \mathbb{R}^{k \times p}, \quad M_{in}(\gamma) = - \begin{bmatrix} z_{itn} 1_{itn}(\gamma)' X_{itn} \\ z_{iTn} 1_{iTn}(\gamma)' X_{iTn} \end{bmatrix} \in \mathbb{R}^{k \times 2},$$

where $\Delta y_{itn} = y_{itn} - y_{it-1, n}$, $\Delta x_{itn} = x_{itn} - x_{it-1, n}$,

$$X_{itn} = \begin{pmatrix} (1, q_{itn}) \\ (1, q_{it-1, n}) \end{pmatrix}, \quad 1_{itn}(\gamma) = \begin{pmatrix} 1 \{q_{itn} > \gamma\} \\ -1 \{q_{it-1, n} > \gamma\} \end{pmatrix}.$$  

Let $M_n(\gamma) = \begin{bmatrix} M_{1in} \\ M_{2in}(\gamma) \end{bmatrix}$, and $M_{0n}(\gamma) = E[M_n(\gamma)]$, $M_{10n} = E[M_{1in}]$, $M_{20n}(\gamma) = E[M_{2in}(\gamma)]$, $M_n(\gamma) = \frac{1}{n} \sum_{i=1}^n M_{in}(\gamma)$, $\tilde{M}_n = \frac{1}{n} \sum_{i=1}^n M_{in}$, and $\tilde{M}_n(\gamma) = \frac{1}{n} \sum_{i=1}^n M_{in}(\gamma)$. We write $M_{0n}$, $M_{20n}$, and $\tilde{M}_n$ instead of $M_{0n}(\gamma)$, $M_{20n}(\gamma)$ and $\tilde{M}_n(\gamma)$. Define

$$\tilde{H}_n = \begin{pmatrix} E_{0n}[z_{itn} \gamma_{0n}] f_{0n}(\gamma_{0n}) - E_{0-1n}[z_{itn} \gamma_{0n}] f_{0-1n}(\gamma_{0n}) \\ \vdots \\ E_{Tn}[z_{iTn} \gamma_{0n}] f_{Tn}(\gamma_{0n}) - E_{T-1n}[z_{iTn} \gamma_{0n}] f_{T-1n}(\gamma_{0n}) \end{pmatrix},$$

where $E_{\cdot}[q]$ and $f_{\cdot}(-)$ are the conditional expectation $E[\cdot | q_{itn} = q]$ and the density of $q_{itn}$, respectively.

Suppose that a sequence $\{\phi_{0n}\}$ (or its subsequence $\{\phi_{0pn}\}$) converges so that $\theta_{0n} \rightarrow \theta_{0, \infty} = (\alpha_{0, \infty}' , \gamma_{0, \infty})'$ and $F_{0n} \rightarrow F_{0, \infty}$, i.e., $\phi_{0n}$ (or $\phi_{0pn}$) $\rightarrow \phi_{0, \infty}$. Note that the density of the distribution $F_{0n}$ converges to the density of $F_{0, \infty}$ uniformly by our choice of norm in $\Phi_F$, and sup_v $\|F_{0n}(v) - F_{0, \infty}(v)\| \rightarrow 0.$

S-28
Note that $M_{0,\infty}(\gamma) = E[M_{i,\infty}(\gamma)] = \lim_{n \to \infty} M_{0n}(\gamma)$ as each element of $M_{in}(\gamma)$ is uniformly integrable by $\max\{E\|z_{itn}\|^{1+r}, E\|z_{itn}\|^{1+r}, E\|\epsilon_{tn}\|^{1+r}\} \leq c_5 < \infty$ for all $n$ while $F_{0n}$ converges to $F_{0,\infty}$. Hence, $M_{10,\infty} = E[M_{i,\infty}] = \lim_{n \to \infty} M_{10n}$ and $M_{20,\infty}(\gamma) = E[M_{2i,\infty}(\gamma)] = \lim_{n \to \infty} M_{20n}(\gamma)$ also hold. Furthermore, $\tilde{H}_n = \lim_{n \to \infty} H_n$, where

$$\tilde{H}_n = \begin{bmatrix}
E_{t_0,\infty}[z_{it_0}\gamma_{0,\infty}]f_{t_0}(\gamma_{0,\infty}) - E_{t_0,\infty}[z_{it_0-1}\gamma_{0,\infty}]f_{t_0-1}(\gamma_{0,\infty}) \\
E_{T,\infty}[z_{iT}\gamma_{0,\infty}]f_{T}(\gamma_{0,\infty}) - E_{T,\infty}[z_{IT-1}\gamma_{0,\infty}]f_{T-1}(\gamma_{0,\infty}) \\
\vdots \\
E_{T,\infty}[z_{iTn}\gamma_{0,\infty}]f_{Tn}(\gamma_{0,\infty}) - E_{T,\infty}[z_{ITn-1}\gamma_{0,\infty}]f_{Tn-1}(\gamma_{0,\infty})
\end{bmatrix}.$$ 

This is because $f_{tn} \to f_{t,\infty}$ uniformly by our definition of norm in $\Phi_F$, and it is straightforward to derive $z_{itn}\gamma_{q_{isn}} = \gamma_{0n} \xrightarrow{d} z_{it,\infty}\gamma_{q_{is,\infty}} = \gamma_{0,\infty}$ for $s = t, t-1$, which implies $E_s[z_{itn}\gamma_{0n}] \to E_s[z_{it,\infty}\gamma_{0,\infty}]$ due to the uniform integrability $E_s[z_{it,\infty}\gamma_{0,\infty}]$ for all $s = t, t-1$. Furthermore, $\|M_{0n} - M_{0,\infty}\| \to 0$ as $n \to \infty$ because $\|M_{0n}(\gamma_{0,\infty}) - M_{0,\infty}(\gamma_{0,\infty})\| \to 0$, and $\|M_{0n} - M_{0n}(\gamma_{0,\infty})\| \leq \|\delta_n(\gamma_n)\|\|\gamma_n - \gamma_{0,\infty}\|$, where

$$\delta_n(\gamma_n) = \begin{bmatrix}
E_{t_0n}[z_{it_0n}(1,\gamma)]f_{tn}(\gamma) - E_{t_0-1,\gamma}[z_{it_0n}(1,\gamma)]f_{t_0-1}(\gamma) \\
E_{Tn}[z_{ITn}(1,\gamma)]f_{Tn}(\gamma) - E_{T-1,\gamma}[z_{ITn}(1,\gamma)]f_{Tn-1}(\gamma)
\end{bmatrix},$$

and $\gamma_n$ is between $\gamma_{0n}$ and $\gamma_{0,\infty}$. Note that $\|\delta_n(\gamma_n)\| < C$ for some nonnegative $C < \infty$ for sufficiently large $n$ as $(\theta_{0n}, F_{0n}) \in \Phi_0$.

Let $\omega_{in} = \{z_{itn}, y_{itn}, x_{itn}, \epsilon_{itn}\}_{t=1}^{n}$ and $g(\omega_{in}, \theta) = (g_{0i}(\omega_{in}, \theta)^{\prime}, \ldots, g_T(\omega_{in}, \theta)^{\prime})^{\prime}$, where $g_t(\omega_{in}, \theta) = z_{itn}(\Delta y_{itn} - \Delta x_{itn}^{\prime} \beta - 1 h_T(\gamma) X_{itn}^{\prime})$. Let $\Omega_n = E[g(\omega_{in}, \theta_{0n})g(\omega_{in}, \theta_{0n})^{\prime}]$, and $\Omega_{\infty} = E[g(\omega_{in}, \theta_{0n})g(\omega_{in}, \theta_{0n})^{\prime}] = \lim_{n \to \infty} \Omega_n$. Let $\bar{g}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} g(\omega_{in}, \theta)$, $\bar{Q}_n(\theta) = \bar{g}_n(\theta)^{\prime} W_n \bar{g}_n(\theta)$, and $g_{0n}(\theta) = E[g(\omega_{in}, \theta)]$, while $W_n = \frac{1}{n} \sum_{i=1}^{n} g(\omega_{in}, \theta_{0n})g(\omega_{in}, \theta_{0n})^{\prime} - \bar{g}_n(\theta_{0n})^{\prime} \bar{g}_n(\theta_{0n})^{-1}$ and $\theta_{0n} = \text{arg min}_\theta \bar{g}_n(\theta)^{\prime} \bar{g}_n(\theta)$ is the initial estimator. $\theta_n = (\hat{\alpha}_n^{\prime}, \gamma_n) = \text{arg min}_\theta \bar{Q}_n(\theta)$ and $\bar{D}_n(\gamma) = n(\min_{\alpha \in A} \bar{Q}_n(\alpha, \gamma) - \bar{Q}_n(\hat{\theta}_n))$.

Let $\omega_{in}^*$ be an i.i.d. draw along the index $i$ from $\{\omega_{in} : i = 1, \ldots, n\}$. Let

$$g_{in}^*(\theta) = (g_{itn}^*(\theta)^{\prime}, \ldots, g_{ITn}^*(\theta)^{\prime})^{\prime}$$

$$g_{in}^*(\theta) = g_t(\omega_{in}^*, \theta) = g_t(\omega_{in}^*, \theta_{0n}) + g_t(\omega_{in}^*, \hat{\theta}_n)$$

$$= -z_{itn}^{\prime} \Delta x_{itn}^{\prime} (\beta - \beta_{0n}^*) - z_{itn}^{\prime} 1_{itn}(\gamma)^{\prime} X_{itn}^* (\delta - \delta_{0n}) + z_{itn}^{\prime} 1_{itn}(\gamma^*)^{\prime} X_{itn}^* \delta_{0n} + z_{itn}^{\prime} \bar{\epsilon}_{itn}^*,$$ (H.1)

where $\theta_{0n}^* = (\hat{\alpha}_n(\gamma_{0n}^\prime)^{\prime}, \gamma_{0n}^\prime)^{\prime}$ and $\hat{\alpha}_n(\gamma) = \text{arg min}_\theta \bar{Q}_n(\alpha, \gamma)$. For the justification of the representation (H.1), please refer to (D.1) and description in Section D.1. Note that $\bar{g}_{in}^*(\theta) = \frac{1}{n} \sum_{i=1}^{n} [g_{in}^*(\theta) - \bar{g}_n(\hat{\theta}_n)]$ becomes the bootstrap sample moment from the grid bootstrap. Then, let $Q_n^*(\theta) = \bar{g}_{in}^*(\theta)^{\prime} W_n^* \bar{g}_{in}^*(\theta)$, $W_n^* = \frac{1}{n} \sum_{i=1}^{n} [g_{in}^*(\hat{\theta}_{0n}) g_{in}^*(\hat{\theta}_{0n})^{\prime}] - \frac{1}{n} \sum_{i=1}^{n} [g_{in}^*(\hat{\theta}_{1n}) g_{in}^*(\hat{\theta}_{1n})^{\prime}]^{-1}$, $\hat{\theta}_{1n} = \text{arg min}_\theta \bar{g}_{in}^*(\theta)^{\prime} \bar{g}_{in}^*(\theta)$, $\theta_n^* = \text{arg min}_\theta Q_n^*(\theta)$, and $D_n^*(\gamma) = n(\min_{\alpha} Q_n^*(\alpha, \gamma) - Q_n^*(\hat{\theta}_n))$. Recall that in Section 4.1 the 100(1 - \gamma)\% grid
For continuous case, $D$ the following convergences hold:

\[
\{ \pi \in \Gamma : D_n(\pi) \leq \hat{F}_n^{-1}(1 - \tau; D_n^*(\pi)) \}.
\]

Define a mapping $\pi_n : \Phi_0 \to \Pi$, where $\Pi = [-\infty, \infty] \times \mathbb{R} \times \Phi_0$ such that

\[
\pi_n(\phi) = \begin{pmatrix} n^{1/4}(\delta_1 + \delta_2\gamma) \\ (\delta_1 + \delta_2\gamma) \\ \phi \end{pmatrix}.
\]

This is because the limits of $n^{1/4}(\delta_1 + \delta_2\gamma)$ and $(\delta_1 + \delta_2\gamma)$ characterize the asymptotic behaviors of the test statistic used in the grid bootstrap.

**Theorem H.1.** For any subsequence $\{p_n\}$ of $\{n : n \in \mathbb{N}\}$ and any sequence $\{\phi_{0p_n} \in \Phi_0 : n \geq 1\}$ s.t. $\pi_{p_n}(\phi_{0p_n}) \to (\zeta_1, \zeta_2, \phi_{0,\infty}) \in \Pi$, $P_{\phi_{0p_n}}(\gamma_{0p_n} \in CI_{p_n,1-\tau}^{grid}) \to 1 - \tau$,

where $P_{\phi_{0p_n}}(\cdot)$ is the probability law under $\phi_{0p_n} = (\theta_{0p_n}, F_{0p_n})$. Moreover,

\[
\lim inf_{n \to \infty} \inf_{\phi_0 \in \Phi_0} P_{\phi_0}(\gamma_0 \in CI_{n,1-\tau}^{grid}) = \lim sup_{n \to \infty} \sup_{\phi_0 \in \Phi_0} P_{\phi_0}(\gamma_0 \in CI_{n,1-\tau}^{grid}) = 1 - \tau,
\]

which establishes the uniform validity of the grid bootstrap confidence interval.

Note that the last statement of Theorem H.1 follows from the theorem’s preceding statement, as the latter verifies Assumption B* from Andrews et al. (2020). Let $\{\pm \infty\} = \{-\infty, +\infty\}$. To show Theorem H.1, we consider the following four cases:

(i) continuous: $\zeta_1 = 0$ and $\zeta_2 = 0$. 

(ii) semi-continuous: $\zeta_1 \in \mathbb{R} \setminus \{0\}$ and $\zeta_2 = 0$. 

(iii) semi-discontinuous: $\zeta_1 \in \{-\infty, 0\}$ and $\zeta_2 = 0$. 

(iv) discontinuous: $\zeta_1 \in \{\pm \infty\}$ and $\zeta_2 \neq 0$. 

The following lemma implies Theorem H.1.

**Lemma H.1.** For all sequences $\{\phi_{0p_n} \in \Phi_0 : n \geq 1\}$ for which $\pi_{p_n}(\phi_{0p_n}) \to (\zeta_1, \zeta_2, \phi_{0,\infty}) \in \Pi$, the following convergences hold (in $\overset{d^*}{\to}$ in $P$ denotes the probability of $\{\omega_{ip_n} : 1 \leq i \leq p_n, n = 1, 2, \ldots\}$):

(i) For continuous case, $D_{p_n}(\gamma_{0p_n}) \overset{d}{\to} Z_0^2$, and $D_{p_n}^*(\gamma_{0p_n}) \overset{d^*}{\to} Z_0^*$ in $P$, where $Z_0 = \max\{Z_0^*, 0\}$ and $Z_0^* \sim N(0, 1)$.

(ii) For semi-continuous case, $D_{p_n}(\gamma_{0p_n}) \overset{d}{\to} D_{\infty}$, and $D_{p_n}^*(\gamma_{0p_n}) \overset{d^*}{\to} D_\infty$ in $P$, where

\[
D_\infty = \begin{cases} 
\left( \frac{U}{\sqrt{\hat{H}'_\infty \Xi \hat{H}_\infty}} \right)^2 & \text{if } U \geq \frac{-\epsilon_1^2}{2|\delta(30, \epsilon_1)|} \hat{H}'_\infty \Xi \hat{H}_\infty \\
\left( \frac{-\epsilon_1^2}{2|\delta(30, \epsilon_1)|} \right)^2 \hat{H}'_\infty \Xi \hat{H}_\infty + 2 \frac{-\epsilon_1^2}{2|\delta(30, \epsilon_1)|} U & \text{if } U < \frac{-\epsilon_1^2}{2|\delta(30, \epsilon_1)|} \hat{H}'_\infty \Xi \hat{H}_\infty.
\end{cases}
\]
\[ U \sim N(0, \tilde{H'}_{n0} \Xi_{n0} \tilde{H}_{n0}), \text{ and } \Xi_{n0} = \Omega_{n0}^{-1} - \Omega_{n}^{-1} M_{n0} \Omega_{n0}^{-1} M_{n0} \Omega_{n0}^{-1}. \]

(iii) For semi-discontinuous and discontinuous cases, \( D_{pn}(\gamma_{pn}) \xrightarrow{d} \chi_1^2 \), and \( D_{pn}^\ast(\gamma_{pn}) \xrightarrow{d} \chi_1^2 \) in \( P \).

Remark 2. Note that the distribution of \( D_{n0} \) is (first-order) stochastically dominated by the \( \chi_1^2 \) distribution. This is because
\[ f_1(Z_0) := \left( \frac{Z_0}{\sqrt{H'_{n0} \Xi_{n0} H_{n0}}} \right)^2 = \left( -\frac{\gamma_0^T}{2|\gamma_{30}|} \right)^2 \tilde{H'}_{n0} \Xi_{n0} \tilde{H}_{n0} + 2 \frac{\gamma_0^T}{2|\gamma_{30}|} Z_0 =: f_2(Z_0) \text{ when } Z_0 = \frac{-\gamma_0^T}{2|\gamma_{30}|} \tilde{H'}_{n0} \Xi_{n0} \tilde{H}_{n0} < 0, \text{ and } f_1'(Z_0) < f_2'(Z_0) \text{ when } Z_0 < \frac{-\gamma_0^T}{2|\gamma_{30}|} \tilde{H'}_{n0} \Xi_{n0} \tilde{H}_{n0}, \]

which implies \( f_1(Z_0) > f_2(Z_0) \) for \( Z_0 < \frac{-\gamma_0^T}{2|\gamma_{30}|} \tilde{H'}_{n0} \Xi_{n0} \tilde{H}_{n0} \).

Proof of Lemma H.1. We prove the result for sequence \( \{n\} \) rather than \( \{p_n\} \) to ease notation. Then, we can replace \( \{n\} \) by \( \{p_n\} \) to complete the proof.

First, we derive the consistency, convergence rates, and asymptotic distributions of \( \hat{\theta}_n \), and then we derive the asymptotic distributions of \( D_{n0}(\gamma_{0n}) \), depending on the regimes determined by \( \zeta_1 \) and \( \zeta_2 \). Then, the same results are derived for bootstrap estimator and test statistic for each case.

Consistency of estimator Define \( \hat{\alpha}_n(\gamma) = \arg \min_{\alpha \in A} \hat{Q}_n(\alpha, \gamma) \), which is
\[ \hat{\alpha}_n(\gamma) = - (\tilde{M}_n(\gamma)' W_n \tilde{M}_n(\gamma))^{-1} \tilde{M}_n(\gamma)' W_n \bar{v}_n, \]
\[ \bar{v}_n = - \tilde{M}_n \alpha_{0n} + u_n, \quad u_n = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} z_{it0n} \Delta \epsilon_{it0n} \\ \vdots \\ z_{iTn} \Delta \epsilon_{iTn} \end{pmatrix}. \]

Therefore, \( \hat{\alpha}_n(\gamma) = - (\tilde{M}_n(\gamma)' W_n \tilde{M}_n(\gamma))^{-1} \tilde{M}_n(\gamma)' W_n (- \tilde{M}_n \alpha_{0n} + u_n) \).

Note that \( u_n \xrightarrow{P} 0 \) by the WLLN for triangular array which holds as \( \sup_{n \in \mathbb{N}} E \| z_{itn} \Delta \epsilon_{itn} \|^2 \leq \sup_{n \in \mathbb{N}} (E \| z_{itn} \|^4)^{1/2} (E \| \Delta \epsilon_{itn} \|^4)^{1/2} < \infty. \) Furthermore, \( \sup_{\gamma \in \Gamma} \| \tilde{M}_n(\gamma) - M_{0n}(\gamma) \| \xrightarrow{P} 0 \) by Lemma H.3. Thus, \( \sup_{\gamma \in \Gamma} \| \hat{\alpha}_n(\gamma) - (M_{0n}(\gamma)' W M_{0n}(\gamma))^{-1} M_{0n}(\gamma)' W M_{0n} \alpha_{0n} \| \xrightarrow{P} 0 \) so that \( \| \hat{\alpha}_n(\gamma_n) - \alpha_{0n} \| \xrightarrow{P} 0 \) if \( \gamma_n = \arg \min_{\gamma \in \Gamma} \hat{Q}_n(\gamma) \), where \( \hat{Q}_n(\gamma) = \hat{Q}_n(\hat{\alpha}_n(\gamma), \gamma) \), is consistent such that \( |\gamma_n - \gamma_{0n}| \xrightarrow{P} 0. \)

If \( \hat{\theta}_{(1)n} \) is consistent, then \( \| W_n - \Omega_{n}^{-1} \| \xrightarrow{P} 0 \) by Lemma H.4. Then,
\[ \sup_{\gamma \in \Gamma} \left| \hat{Q}_n(\gamma) - \| (I - P_{\Omega_{n}^{-1/2} M_{0n}(\gamma)}) (\Omega_{n}^{-1/2} M_{0n} \alpha_{0n}) \| \right| \xrightarrow{P} 0. \]

Since \( \sigma_{\min} \left( \begin{bmatrix} M_{20n} & \tilde{H}_{n0} \end{bmatrix} \right) \geq c_2 \) for all \( n \), \( M_{20n} \hat{\alpha}_{0n} \) is not in the column space of \( M_{20n}(\gamma) \), and \( \gamma_{0n} \) is the unique minimizer of \( \| (I - P_{\Omega_{n}^{-1/2} M_{0n}(\gamma)}) (\Omega_{n}^{-1/2} M_{0n} \alpha_{0n}) \|. \) By applying the argmin CMT as in the proof of Theorem 2, \( |\gamma_n - \gamma_{0n}| \xrightarrow{P} 0 \) can be derived. Derivation of the consistency of \( \hat{\theta}_{(1)n} \) is straightforward if we replace \( \Omega_{n}^{-1/2} \) by the identity matrix.

S-31
Convergence rate of estimator. By Lemma H.5 and \( \| \hat{\theta}_n - \theta_0 \| \xrightarrow{p} 0 \), \( \sqrt{n} \| \bar{g}_n(\hat{\theta}_n) - \bar{g}_n(\theta_0) - g_0 n(\hat{\theta}_n) \| = o_p(1) \). As \( \| W_n - \Omega_n^{-1} \| \xrightarrow{p} 0 \),

\[
\sqrt{n} \| W_n^{-1/2} \bar{g}_n(\hat{\theta}_n) - W_n^{-1/2} \bar{g}_n(\theta_0) - \Omega_n^{-1/2} g_0 n(\hat{\theta}_n) \| = o_p(1).
\]

By triangle inequality, \( \sqrt{n} \| \Omega_n^{-1/2} g_0 n(\hat{\theta}_n) \| \leq \sqrt{n} \| W_n^{-1/2} \bar{g}_n(\hat{\theta}_n) \| + \sqrt{n} \| W_n^{-1/2} \bar{g}_n(\theta_0) \| + o_p(1) \). As \( \hat{\theta}_n \) minimizes \( \| W_n^{-1/2} \bar{g}_n(\theta) \| \), \( \sqrt{n} \| W_n^{-1/2} \bar{g}_n(\hat{\theta}_n) \| \leq \sqrt{n} \| W_n^{-1/2} \bar{g}_n(\theta_0) \| \). Note that \( \sqrt{n} \| W_n^{-1/2} \bar{g}_n(\theta_0) \| = O_p(1) \) because \( \| W_n \| = O_p(1) \), while the CLT for triangular array implies \( \frac{1}{\sqrt{n}} \sum_{i=1}^n z_{itn} \Delta_{itn} \xrightarrow{d} N(0, \lim_{n \to \infty} E[z_{itn}^2 \Delta_{itn}^2]) \). The CLT holds by combination of Lyapunov condition and Cramér-Wold if \( \lim_{n \to \infty} \frac{E[(X_{z_{itn}})^2 r^2 \Delta_{itn}^2]}{n} = 0 \) for some \( r > 0 \) and for any \( \lambda \in \mathbb{R}^{\dim(z_{itn})} \), which holds as \( \inf_{n \in \mathbb{N}} \sigma_{\min}(\Omega_n) > 0 \) and \( \sup_{n \in \mathbb{N}} \max\{E[z_{itn}^4 + 2r^2], (E[\Delta_{itn}^4 + 2r^2])^{1/2}\} < \infty \) for some \( r > 0 \). Therefore,

\[
\sqrt{n} \| \Omega_n^{-1/2} g_0 n(\hat{\theta}_n) \| \leq \sqrt{n} \| W_n^{-1/2} \bar{g}_n(\hat{\theta}_n) \| + \sqrt{n} \| W_n^{-1/2} \bar{g}_n(\theta_0) \| + o_p(1)
\]

\[
\leq 2 \sqrt{n} \| W_n^{-1/2} \bar{g}_n(\theta_0) \| + o_p(1)
\]

\[
= O_p(1),
\]

while \( \sqrt{n} \| \Omega_n^{-1/2} g_0 n(\hat{\theta}_n) \| \geq \sqrt{n} \| \Omega_n^{-1/2} M_0 n(\hat{\alpha}_n - \alpha_0) + \Omega_n^{-1/2} \bar{H}_n (\hat{\delta}_{10n} + \delta_{30n} \gamma_0n) (\hat{\gamma}_n - \gamma_0) + \frac{\delta_{400}}{2} (\hat{\gamma}_n - \gamma_0)^2 \| + o(\sqrt{n}(\| \hat{\alpha}_n - \alpha_0 \| + (\| \delta_{10n} + \delta_{30n} \gamma_0n \| (\gamma_n - \gamma_0) + (\gamma_n - \gamma_0)^2)) \) by Lemma H.2.

In conclusion,

\[
\sqrt{n}(\| \hat{\alpha}_n - \alpha_0 \| + (\| \delta_{10n} + \delta_{30n} \gamma_0n \| (\gamma_n - \gamma_0) + (\gamma_n - \gamma_0)^2) \leq O_p(1).
\]

It implies that \( \sqrt{n}(\| \hat{\alpha}_n - \alpha_0 \| = O_p(1) \) for any values of \( \zeta_1 = \lim_n n^{1/4}(\delta_{10n} + \delta_{30n} \gamma_0n) \) and \( \zeta_2 = \lim_n (\delta_{10n} + \delta_{30n} \gamma_0n) \), while for \( \hat{\gamma}_n \),

(i) \( n^{1/4}(\hat{\gamma}_n - \gamma_0n) = O_p(1) \) if \( \zeta_1 = \zeta_2 = 0 \)

(ii) \( n^{1/4}(\hat{\gamma}_n - \gamma_0n) = O_p(1) \) if \( \zeta_1 \in \mathbb{R} \setminus \{0\}, \zeta_2 = 0 \)

(iii) \( \sqrt{n}(\delta_{10n} + \delta_{30n} \gamma_0n)(\gamma_n - \gamma_0) = O_p(1) \) if \( |\zeta_1| = \infty, \zeta_2 = 0 \)

(iv) \( \sqrt{n}(\gamma_n - \gamma_0) = O_p(1) \) if \( |\zeta_1| = \infty, \zeta_2 \neq 0 \).

Asymptotic distribution of estimator and test statistic. We only consider (ii) semi-continuous and (iii) semi-discontinuous cases since the proofs for (i) continuous and (iv) discontinuous cases are almost identical to the proof of continuous and discontinuous cases in Theorem 3.

Case (ii): Let \( a = \sqrt{n}(\alpha - \alpha_0n) \) and \( b = n^{1/4}(\gamma - \gamma_0n) \). Additionally, define \( \hat{\alpha}_n = \sqrt{n}(\hat{\alpha} - \alpha_0n) \) and \( \hat{\gamma}_n = n^{1/4}(\hat{\gamma} - \gamma_0n) \). Let

\[
S_n(a, b) = n \hat{Q}_n(\alpha_0n + \frac{a}{\sqrt{n}}, \gamma_0n + \frac{b}{n^{1/2}}) = n \bar{g}_n(\alpha_0n + \frac{a}{\sqrt{n}}, \gamma_0n + \frac{b}{n^{1/2}}) \sqrt{W_n} \bar{g}_n(\alpha_0n + \frac{a}{\sqrt{n}}, \gamma_0n + \frac{b}{n^{1/2}}).
\]
The rescaled and reparametrized sample moment can be written as

\[
\sqrt{n} \hat{g}_n(a_0 n + a \frac{a}{\sqrt{n}}, \gamma_0 n + b \frac{a}{n^2}) = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} z_{it0n} \Delta \epsilon_{it0n} \right) - \left( \frac{1}{n} \sum_{i=1}^{n} z_{it0n} \Delta x'_{it0n} \right) a_1 \\
- \left( \frac{1}{n} \sum_{i=1}^{n} z_{iTn} \Delta \epsilon_{iTn} \right) \left( \frac{1}{n} \sum_{i=1}^{n} z_{iTn} \Delta x'_{iTn} \right) a_2 \\
+ \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} z_{it0n} (1_{it0n}(\gamma_{it0n} + \frac{b}{n^2}))' X_{it0n} \right) \\
- \left( \frac{1}{n} \sum_{i=1}^{n} z_{iTn} 1_{iTn}(\gamma_{iTn} + \frac{b}{n^2})' X_{iTn} \right) \delta_{0n}.
\]

By the CLT for triangular array,

\[
\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} z_{it0n} \Delta \epsilon_{it0n} \right) \to -e \sim N(0, \Omega_\infty).
\]

Note that the CLT holds by combination of Lyapunov condition and Cramér-Wold device if \( \lim_{n \to \infty} E[(\lambda z_{it0n})^{2 + r} \Delta \epsilon_{it0n}^{2 + r}] = 0 \) for some \( r > 0 \) for any \( \lambda \in \mathbb{R}^k \), which holds as \( \inf_{n \in \mathbb{N}} \sigma_{\min}(\Omega_n) > 0 \) and \( \sup_{n \in \mathbb{N}} \max\{(E\|z_{it0n}\|^{4+2r})^{1/2}, (E\|\Delta \epsilon_{it0n}\|^{4+2r})^{1/2}\} < \infty \) for some \( r > 0 \). By the WLLN for triangular array,

\[
\left( \frac{1}{n} \sum_{i=1}^{n} z_{it0n} \Delta x'_{it0n} \right) \to (E z_{it0n, \infty} \Delta x'_{it0n, \infty}), \\
\left( \frac{1}{n} \sum_{i=1}^{n} z_{iTn} \Delta x'_{ iTn} \right) \to (E z_{iT, \infty} \Delta x'_{iT, \infty}),
\]

which holds as \( \sup_{n \in \mathbb{N}} E\|z_{it0n} \Delta x_{it0n}\|^2 \leq \sup_{n \in \mathbb{N}} (E\|z_{it0n}\|^4)^{1/2} (E\|\Delta x_{it0n}\|^4)^{1/2} < \infty \). Let \( K < \infty \) be some constant. By the ULLN in Lemma H.3,

\[
\left( \frac{1}{n} \sum_{i=1}^{n} z_{it0n} 1_{it0n}(\gamma_{it0n} + \frac{b}{n^2})' X_{it0n} \right) - \left( E z_{it0n, \infty} 1_{it0n, \infty}(\gamma_{it0n, \infty} + \frac{b}{n^2})' X_{it0n, \infty} \right) \to 0,
\]

uniformly with respect to \( b \in [-K, K] \). Then, by the continuity of \( \kappa \mapsto E[z_{it, \infty} 1_{it, \infty}(\gamma_{it, \infty} + \frac{b}{n^2})' X_{it, \infty}] \).
\[ \kappa) X_{it, \infty} \] at \( \kappa = 0, \]

\[
\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it_0} \frac{1}{1 - \lambda_i} (\gamma_{0, \infty} + \frac{b}{n^2}) X_{it_0, \infty} \right) \quad \rightarrow_p \quad \left( E z_{it_0} \frac{1}{1 - \lambda_i} (\gamma_{0, \infty})^t X_{it_0, \infty} \right)
\]

uniformly with respect to \( b \in [-K, K] \). By Lemma H.6,

\[
\sqrt{n} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it_0} (1 - \lambda_i) \left( \gamma_{0, \infty} + \frac{b}{n^2} \right)^t X_{it_0, \infty} \right) \quad \rightarrow_p \quad \left( E z_{it_0} (1 - \lambda_i) \left( \gamma_{0, \infty} \right)^t X_{it_0, \infty} \right)
\]

uniformly with respect to \( b \in [-K, K] \). Therefore, \( S_n(a, b) \) weakly converges to

\[ S(a, b) = (M_{0, \infty} a + \tilde{H}_\infty (\zeta_1 b + \frac{\delta_{30, \infty}}{2} b^2) - e)^t \Omega_{\infty}^{-1} (M_{0, \infty} a + \tilde{H}_\infty (\zeta_1 b + \frac{\delta_{30, \infty}}{2} b^2) - e), \]

in \( \ell^\infty(K) \) for any compact \( K \subset \mathbb{R}^{2p+2} \).

Let \( \bar{b} = \zeta_1 b + \frac{\delta_{30, \infty}}{2} b^2 \) and \( \hat{b}_n = \zeta_1 \hat{b}_n + \frac{\delta_{30, \infty}}{2} \hat{b}_n^2 \). We consider \( \delta_{30, \infty} > 0 \) so that \( \hat{b} \geq -\frac{\zeta_1^2}{2\delta_{30, \infty}} \).

When \( \delta_{30, \infty} < 0 \), derivations are almost identical and lead to the same limit distribution of the test statistic. Let \( \tilde{b} = -\frac{\zeta_1^2}{2\delta_{30, \infty}} \). Then, by the CMT,

\[ (\tilde{a}_n, \hat{b}_n) \xrightarrow{d} (a_0, \hat{b}_0) = \arg \min_{a, b \geq \tilde{b}} \left( M_{0, \infty} a + \tilde{H}_\infty \tilde{b} - e \right)^t \Omega_{\infty}^{-1} (M_{0, \infty} a + \tilde{H}_\infty \tilde{b} - e) \]

KKT conditions, as in the proof of Theorem 2, imply

\[
M'_{0, \infty} \Omega_{\infty}^{-1} M_{0, \infty} a_0 + M'_{0, \infty} \Omega_{\infty}^{-1} \tilde{H}_\infty \tilde{b}_0 - M'_{0, \infty} \Omega_{\infty}^{-1} e = 0,
\]

\[ \tilde{H}_\infty \Omega_{\infty}^{-1} \tilde{H}_\infty \tilde{b}_0 + \tilde{H}_\infty \Omega_{\infty}^{-1} M_{0, \infty} a_0 - \tilde{H}_\infty \Omega_{\infty}^{-1} e - \lambda = 0, \]

\[ \lambda \geq 0, \tilde{b}_0 \geq \tilde{b}, \text{ and } \lambda (\tilde{b}_0 - \tilde{b}) = 0 \text{ should hold. Then, we can get} \]

\[ \tilde{b}_0 = \begin{cases} \left[ \tilde{H}'_{\infty} \Xi_{\infty} \tilde{H}_{\infty} \right]^{-1} \tilde{H}'_{\infty} \Xi_{\infty} e & \text{if } \left[ \tilde{H}'_{\infty} \Xi_{\infty} \tilde{H}_{\infty} \right]^{-1} \tilde{H}'_{\infty} \Xi_{\infty} e \geq \tilde{b} \\ \tilde{b} & \text{else} \end{cases} \]

where \( \Xi_{\infty} = \Omega_{\infty}^{-1/2} \left( I - P_{\Omega_{\infty}^{-1/2} M_{0, \infty}} \right) \Omega_{\infty}^{-1/2} \). \( \tilde{b}_0 \) follows a normal distribution that is left censored
at \( b \). Then,

\[
a_0 = \begin{cases} 
(M_0, \infty \Omega_{\infty}^{-1} M_0, \infty)^{-1} M_0, \infty \Omega_{\infty}^{-1} (I - \tilde{H}_{\infty}^\prime \Xi_{\infty} \tilde{H}_{\infty})^{-1} \tilde{H}_{\infty}^\prime \Xi_{\infty} e \quad & \text{if } [\tilde{H}_{\infty}^\prime \Xi_{\infty} \tilde{H}_{\infty}]^{-1} \tilde{H}_{\infty}^\prime \Xi_{\infty} e \geq \tfrac{b}{2} \\
(M_0, \infty \Omega_{\infty}^{-1} M_0, \infty)^{-1} M_0, \infty \Omega_{\infty}^{-1} (e - \tilde{H}_{\infty}^\prime b) \quad & \text{else.}
\end{cases}
\]

Asymptotic distribution of the test statistic \( D_n(\gamma_{0n}) \) can be derived by

\[
D_n(\gamma_{0n}) \overset{d}{=} \min_{a, b \geq \frac{b}{2}} (M_0, \infty a + e' \Omega_{\infty}^{-1} (M_0, \infty M_0, \infty)^{-1} \Omega_{\infty}^{-1} (M_0, \infty a + e' \Omega_{\infty}^{-1} (M_0, \infty b - e)
\]

where we apply the CMT. Note that \( \min_{a, b \geq \frac{b}{2}} (M_0, \infty a - e' \Omega_{\infty}^{-1} (M_0, \infty a - e) = e' \Omega_{\infty}^{-1/2} (I - P_{\Omega_{\infty}^{-1/2} M_0, \infty}) \Omega_{\infty}^{-1/2} e 
\)

By plugging in the formula for \((a_0, \tilde{b}_0)\) (note that \(M_0, \infty \Omega_{\infty}^{-1} M_0, \infty a_0 + M_0, \infty \Omega_{\infty}^{-1} \tilde{H}_{\infty} b_0 = M_0, \infty \Omega_{\infty} e \)) we can get

\[
\min_{a, b \geq \frac{b}{2}} (M_0, \infty a + \tilde{H}_{\infty}^\prime \tilde{b} - e) \Omega_{\infty}^{-1} (M_0, \infty a + \tilde{H}_{\infty}^\prime \tilde{b} - e)
\]

Therefore, the limit distribution of the test statistic is identical to

\[
\left\{ 
\begin{array}{ll}
 e' \Xi_{\infty} \tilde{H}_{\infty}^\prime \Xi_{\infty} \tilde{H}_{\infty}^{-1} \tilde{H}_{\infty} \Xi_{\infty} e & \text{if } [\tilde{H}_{\infty}^\prime \Xi_{\infty} \tilde{H}_{\infty}]^{-1} \tilde{H}_{\infty}^\prime \Xi_{\infty} e \geq \tfrac{b}{2} \\
 -(\tilde{H}_{\infty}^\prime \Xi_{\infty} \tilde{H}_{\infty})^2 + 2(e' \Xi_{\infty} \tilde{H}_{\infty}) \tilde{b} & \text{else.}
\end{array}
\right.
\]

Case (iii): Let \( a = \sqrt{n} \delta_{0n} \) and \( b = \sqrt{n} (\delta_{30n} \gamma_{0n}) (\gamma - \gamma_{0n}) \). The rescaled and
reparametrized sample moment can be written as

\[
\sqrt{n} \tilde{g}_n(a_{0n} + \frac{a}{\sqrt{n}}, \gamma_{0n} + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n} \gamma_{0n})}) = \\
\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} z_{i0n} \Delta \epsilon_{i0n} \right) - \left( \frac{1}{n} \sum_{i=1}^{n} z_{i0n} \Delta x'_{i0n} \right) a_1 \\
- \left( \frac{1}{n} \sum_{i=1}^{n} z_{i0n} 1_{i0n} (\gamma_{0n} + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n} \gamma_{0n})})' X_{i0n} \right) a_2 \\
+ \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} z_{i0n} (1_{i0n} (\gamma_{0n})' - 1_{i0n} (\gamma_{0n} + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n} \gamma_{0n})})' X_{i0n}) \right) \delta_{0n}.
\]

By the CLT for triangular array,

\[
\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} z_{i0n} \Delta \epsilon_{i0n} \right) \xrightarrow{d} -e \sim N(0, \Omega_{\infty}).
\]

By the WLLN for triangular array,

\[
\left( \frac{1}{n} \sum_{i=1}^{n} z_{i0n} \Delta x'_{i0n} \right) \xrightarrow{p} \left( \frac{E z_{i0n,\infty} \Delta x'_{i0n,\infty}}{E z_{i,\infty} \Delta x'_{i,\infty}} \right).
\]

By the ULLN in Lemma H.3,

\[
\left\| \frac{1}{n} \sum_{i=1}^{n} z_{i0n} 1_{i0n} (\gamma_{0n} + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n} \gamma_{0n})})' X_{i0n} \right\| \\
= \left\| \frac{1}{n} \sum_{i=1}^{n} z_{i0n} 1_{i0n} (\gamma_{0n} + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n} \gamma_{0n})})' X_{i0n} \right\| \\
\xrightarrow{p} 0
\]
uniformly with respect to $b \in [-K, K]$, which implies
\[
\left( \frac{1}{n} \sum_{i=1}^{n} z_{it_0, \infty} 1_{i t_0, \infty} (\gamma_0, \infty) + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_0)} \right) X_{it_0, \infty} \xrightarrow{p} \left( E z_{it_0, \infty} 1_{i t_0, \infty} (\gamma_0, \infty) X_{it_0, \infty} \right)
\]
\[
\left( \frac{1}{n} \sum_{i=1}^{n} z_{iT, \infty} 1_{i T}, \infty (\gamma_0, \infty) + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_0)} \right) X_{iT, \infty} \xrightarrow{p} \left( E z_{iT, \infty} 1_{i T}, \infty (\gamma_0, \infty) X_{iT, \infty} \right)
\]
uniformly with respect to $b \in [-K, K]$. By Lemma H.7,
\[
\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} z_{it_0, \infty} 1_{i t_0, \infty} (\gamma_0, \infty) - 1_{i t_0, \infty} (\gamma_0, \infty) + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_0)} \right) X_{it_0, \infty} \delta_0(n) \xrightarrow{p} \left( E_{t_0, \infty} [z_{it_0, \infty}] (\gamma_0, \infty) f_{t_0, \infty} (\gamma_0, \infty) - E_{t_0-1, \infty} [z_{i T, \infty}] (\gamma_0, \infty) f_{t_0-1, \infty} (\gamma_0, \infty) \right)
\]
\[
\left( \frac{1}{n} \sum_{i=1}^{n} z_{iT, \infty} 1_{i T}, \infty (\gamma_0, \infty) - 1_{i T}, \infty (\gamma_0, \infty) + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_0)} \right) X_{iT, \infty} \delta_0(n) \xrightarrow{p} \left( E_{T, \infty} [z_{iT, \infty}] (\gamma_0, \infty) f_{T, \infty} (\gamma_0, \infty) - E_{T-1, \infty} [z_{iT, \infty}] (\gamma_0, \infty) f_{T-1, \infty} (\gamma_0, \infty) \right)
\]
uniformly with respect to $b \in [-K, K]$. Therefore, $S_n(a, b) = n \tilde{Q}_n(\alpha_0 n + \frac{a}{\sqrt{n}}, \gamma_0 n + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_0)})$ weakly converges to
\[
S(a, b) = (M_{0, \infty} a + \tilde{H}_\infty b - e) \Omega_{\infty}^{-1} (M_{0, \infty} a + \tilde{H}_\infty b - e),
\]
in $\ell^\infty(\mathbb{K})$ for any compact $\mathbb{K} \subset \mathbb{R}^{2^{p+2}}$. Then, $\tilde{a}_n = \sqrt{n}(\tilde{a}_n - \alpha_0 n)$ and $\tilde{b}_n = \sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_0)(\tilde{\gamma}_n - \gamma_0 n)$ converges in distribution to
\[
(a_0, b_0) = \arg\min_{a,b} (M_{0, \infty} a + \tilde{H}_\infty b - e) \Omega_{\infty}^{-1} (M_{0, \infty} a + \tilde{H}_\infty b - e).
\]
by the argmin CMT. KKT conditions, as in the proof of Theorem 2, imply
\[
M'_{0, \infty} \Omega_{\infty}^{-1} M_{0, \infty} a_0 + M'_{0, \infty} \Omega_{\infty}^{-1} \tilde{H}_\infty b_0 - M_{0, \infty} \Omega_{\infty}^{-1} e = 0
\]
\[
\tilde{H}'_{\infty} \Omega_{\infty}^{-1} \tilde{H}_\infty b_0 + \tilde{H}'_{\infty} \Omega_{\infty}^{-1} M_{0, \infty} a_0 - \tilde{H}'_{\infty} \Omega_{\infty}^{-1} e = 0.
\]
Then, we can get
\[
b_0 = [\tilde{H}'_{\infty} \Xi_{\infty} \tilde{H}_\infty]^{-1} \tilde{H}'_{\infty} \Xi_{\infty} e,
\]
where $\Xi_{\infty} = \Omega_{\infty}^{-1/2} (I - P_{\Omega_{\infty}^{-1/2} M_{0, \infty}}) \Omega_{\infty}^{-1/2}$, and
\[
a_0 = (M'_{0, \infty} \Omega_{\infty}^{-1} M_{0, \infty})^{-1} M'_{0, \infty} \Omega_{\infty}^{-1} (I - \tilde{H}_\infty [\tilde{H}'_{\infty} \Xi_{\infty} \tilde{H}_\infty]^{-1} \tilde{H}'_{\infty} \Xi_{\infty}) e
\]
Asymptotic distribution of the test statistic $D_n(\gamma_{0n})$ can be derived by

$$D_n(\gamma_{0n}) \xrightarrow{d} \min_{a,b}(M_{0,\infty}a - e)\Omega^{-1}_\infty(M_{0,\infty}a - e)$$

$$- \min_{a,b}(M_{0,\infty}a + \tilde{H}_\infty b - e)\Omega^{-1}_\infty(M_{0,\infty}a + \tilde{H}_\infty b - e),$$

where we apply the CMT. Note that $\min_{a}(M_{0,\infty}a - e)\Omega^{-1}_\infty(M_{0,\infty}a - e) = e\Omega^{-1/2}(I - P_{\Omega^{-1/2}M_{0,\infty}})\Omega^{-1/2}e$, while

$$\min_{a,b}(M_{0,\infty}a + \tilde{H}_\infty b - e)\Omega^{-1}_\infty(M_{0,\infty}a + \tilde{H}_\infty b - e)$$

$$= (M_{0,\infty}a_0 + \tilde{H}_\infty b_0 - e)\Omega^{-1}_\infty(M_{0,\infty}a_0 + \tilde{H}_\infty b_0 - e)$$

$$= (M'_{0,\infty}\Omega^{-1}_\infty M_{0,\infty}a_0 + M'_{0,\infty}\Omega^{-1}_\infty \tilde{H}_\infty b_0)'(M'_{0,\infty}\Omega^{-1}_\infty M_{0,\infty}a_0 + M'_{0,\infty}\Omega^{-1}_\infty \tilde{H}_\infty b_0)$$

$$+ b_0\tilde{H}'_\infty\Omega^{-1/2}(I - P_{\Omega^{-1/2}M_{0,\infty}})\Omega^{-1/2}\tilde{H}_\infty b_0$$

$$- 2e\Omega^{-1/2}(M'_{0,\infty}\Omega^{-1}_\infty M_{0,\infty})^{-1}(M'_{0,\infty}\Omega^{-1}_\infty M_{0,\infty}a_0 + M'_{0,\infty}\Omega^{-1}_\infty \tilde{H}_\infty b_0)$$

$$- 2e\Omega^{-1/2}(I - P_{\Omega^{-1/2}M_{0,\infty}})\Omega^{-1/2}\tilde{H}_\infty b_0 + e\Omega^{-1/2}e.$$ 

By plugging in the formula for $(a_0, b_0)$ (note that $M'_{0,\infty}\Omega^{-1}_\infty M_{0,\infty}a_0 + M'_{0,\infty}\Omega^{-1}_\infty \tilde{H}_\infty b_0 = M'_{0,\infty}\Omega^{-1}_\infty e$), we can get

$$\min_{a,b}(M_{0,\infty}a + \tilde{H}_\infty b - e)\Omega^{-1}_\infty(M_{0,\infty}a + \tilde{H}_\infty b - e)$$

$$= e\Omega^{-1/2}(I - P_{\Omega^{-1/2}M_{0,\infty}})\Omega^{-1/2}e - e\Xi\tilde{H}_\infty(\tilde{H}'_\infty\Xi\tilde{H}_\infty)^{-1}\tilde{H}_\infty\Xi e$$

Therefore, the limit distribution of the test statistic is identical to

$$e\Xi\tilde{H}_\infty(\tilde{H}'_\infty\Xi\tilde{H}_\infty)^{-1}\tilde{H}_\infty\Xi e,$$

which has the $\chi^2_1$ distribution.

**Limit distribution of bootstrap estimator and test statistic** The derivation of the limit distributions of the bootstrap estimator and test statistic is almost identical to that of the asymptotic distributions of the sample estimator and test statistic. We need to replace $\delta_{0n}$ by $\delta^*_{0n} = \delta_{0n}(\gamma_{0n}), \{\Delta e_{itn}\}$ by $\{\tilde{\Delta} e_{itn}\}$, and sample moments by bootstrap moments in the previous part of the proof regarding asymptotic analysis. Be mindful that we do not need to replace $\gamma_{0n}$ in the previous part of the proof as we focus on the grid bootstrap when $\gamma^*_{0n} = \gamma_{0n}$ to show that the grid bootstrap CI provides correct coverage rate. Lemmas H.10, H.11, H.12, and H.13 are applied instead of Lemmas H.3, H.5, H.6, and H.7 in the places where the latter are used in the previous part of the proof. Moreover, Lemmas H.8 and H.9 are applied instead of the WLLN and CLT for triangular array applied to $\{z_{itn}\Delta e_{itn} : 1 \leq i \leq n, n \in \mathbb{N}\}$ in the places where the latter are used in the previous part of the proof. 

\[\square\]
H.1 Auxiliary Lemmas

Lemma H.2. Let \( \phi_{0n} \in \Phi_0 : n \geq 1 \) and \( \pi_n(\phi_{0n}) \to (\zeta_1, \zeta_2, \phi_{0, \infty}) \in \Pi \). For any \( \eta > 0 \), there is \( h > 0 \) such that

\[
\lim_{n \to \infty} \sup_{\|\theta - \theta_{0n}\| < h} \frac{\sqrt{n} \left\| g_{0n}(\theta) - M_{0n}(\alpha - \alpha_{0n}) - \tilde{H}_n \left[ (\beta_{10n} + \delta_{30n}\gamma_{0n})(\gamma - \gamma_{0n}) + \frac{\delta_{0n}}{2}(\gamma - \gamma_{0n})^2 \right] \right\|}{1 + \sqrt{n} \left\| \alpha - \alpha_{0n} \right\| + \left\| (\delta_{10n} + \delta_{30n}\gamma_{0n})(\gamma - \gamma_{0n}) + (\gamma - \gamma_{0n})^2 \right\|} < \eta.
\]

Proof. Note that \( g_{0n}(\theta) - M_{0n}(\alpha - \alpha_{0n}) = M_{0n}(\gamma)\alpha - M_{0n}\alpha_{0n} - M_{0n}(\alpha - \alpha_{0n}) = (M_{0n}(\gamma) - M_{0n})\alpha = (M_{20n}(\gamma) - M_{20n})\delta + (M_{20n}(\gamma) - M_{20n})[\delta_{10n} + (\delta - \delta_{0n})]. \)

First, we derive a bound for \( (M_{20n}(\gamma) - M_{20n})\delta_{0n} \) which is

\[
\begin{aligned}
E[z_{itn}(\delta_{10n} + \delta_{30n}q_{itn})1\{\gamma \geq \gamma_{itn} > \gamma_{0n}\}] - E[z_{itn-1,n}(\delta_{10n} + \delta_{30n}q_{itn-1,n})1\{\gamma \geq \gamma_{itn-1,n} > \gamma_{0n}\}] & \\
\vdots & \\
E[z_{iTn}(\delta_{10n} + \delta_{30n}q_{iTn})1\{\gamma \geq \gamma_{iTn} > \gamma_{0n}\}] - E[z_{iT-1,n}(\delta_{10n} + \delta_{30n}q_{iT-1,n})1\{\gamma \geq \gamma_{iT-1,n} > \gamma_{0n}\}].
\end{aligned}
\]

Suppose \( \gamma > \gamma_{0n} \), and the other case can be analyzed similarly. By Taylor expansion,

\[
E[z_{itn}(\delta_{10n} + \delta_{30n}q_{itn})1\{\gamma \geq \gamma_{itn} > \gamma_{0n}\}] = E[z_{itn}|\gamma_{0n}]f_{tn}(\gamma_{0n}) \left\{ (\delta_{10n} + \delta_{30n}\gamma_{0n}) \cdot (\gamma - \gamma_{0n}) + \frac{\delta_{30n}}{2}(\gamma - \gamma_{0n})^2 \right\} + R_n,
\]

where

\[
R_n = \frac{1}{2} \frac{d}{d\gamma} \left( E[z_{itn}|\gamma]f_{tn}(\gamma) \right) |_{\gamma = \gamma_{0n}} \times (\delta_{10n} + \delta_{30n}\gamma_{0n})(\gamma - \gamma_{0n})^2
\]

\[
+ \frac{1}{2} E[z_{itn}|\gamma_{0n}]f_{tn}(\gamma_{0n}) - E[z_{itn}|\gamma_{0n}]f_{tn}(\gamma_{0n}) \right\} (\gamma - \gamma_{0n})^2,
\]

and \( \gamma_{0n} \in [\gamma_{0n}, \gamma] \). Suppose \( |\gamma - \gamma_{0n}| \leq h_1 \). For sufficiently small \( h_1 > 0 \), there is \( N \) such that if \( n > N \), then \( \frac{d}{d\gamma} \left( E[z_{itn}|\gamma]f_{tn}(\gamma) \right) |_{\gamma = \gamma_{0n}} \leq C_1 < \infty \) for some \( C_1 < \infty \). There also exists \( C_2 < \infty \) such that \( \delta_{10n} + \delta_{30n}\gamma_{0n} \leq (\delta_{10n} + \delta_{30n}\gamma_{0n}) + \frac{\delta_{30n}}{h_1} \leq (\delta_{10n} + \delta_{30n}\gamma_{0n}) + C_2 h_1 \), and hence \( \frac{d}{d\gamma} \left( E[z_{itn}|\gamma]f_{tn}(\gamma) \right) |_{\gamma = \gamma_{0n}} \times (\delta_{10n} + \delta_{30n}\gamma_{0n})(\gamma - \gamma_{0n})^2 \leq C_1((\delta_{10n} + \delta_{30n}\gamma_{0n}) + C_2 h_1) h_1^2 \) for sufficiently large \( n \). Moreover, there exists \( C_3 < \infty \) such that \( E[z_{itn}|\gamma_{0n}]f_{tn}(\gamma_{0n}) - E[z_{itn}][\gamma_{0n}]f_{tn}(\gamma_{0n}) \leq \sup_{|\gamma_{0n} - \gamma| \leq h_1 \gamma_{0n}} \frac{d}{d\gamma} \left( E[z_{itn}|\gamma]f_{tn}(\gamma) \right) |_{\gamma = \gamma} \| h_1 \leq C_3 h_1 \) for sufficiently small \( h_1 > 0 \) and sufficiently large \( n \). Hence, \( \|R_n\| < C((\delta_{10n} + \delta_{30n}\gamma_{0n}) h_1^2 + h_1^3) \) for some \( C < \infty \) and for sufficiently small \( h_1 > 0 \) and sufficiently large \( n \). Therefore, there exists \( h_1 > 0 \) such that if \( |\gamma - \gamma_{0n}| \leq h_1 \), then

\[
\|E[z_{itn}(\delta_{10n} + \delta_{30n}q_{itn})1\{\gamma \geq \gamma_{itn} > \gamma_{0n}\}] - E[z_{itn}|\gamma_{0n}]f_{tn}(\gamma_{0n})
\]

\[
\times \left\{ (\delta_{10n} + \delta_{30n}\gamma_{0n}) \cdot (\gamma - \gamma_{0n}) + \frac{\delta_{30n}}{2}(\gamma - \gamma_{0n})^2 \right\} \| < C((\delta_{10n} + \delta_{30n}\gamma_{0n}) h_1^2 + h_1^3)
\]

for some \( C < \infty \) and for sufficiently large \( n \). By similar computations for \( E[z_{itn}(\delta_{10n} + \)
\[ \delta_{3nq_{t-1.n}}\{\gamma \geq q_{t-1,n} > \gamma_{0n}\}, \]

we can derive that there exists \( h_1 > 0 \) such that if \( |\gamma - \gamma_{0n}| \leq h_1 \), then

\[ \| (M_{20n}(\gamma) - M_{20n})\delta_{0n} - \tilde{H}_n[\{\delta_{10n} + \delta_{30n}\gamma_{0n}\}(\gamma - \gamma_{0n}) + \frac{\delta_{0n}}{2}(\gamma - \gamma_{0n})^2] \| < C(\delta_{10n} + \delta_{30n}\gamma_{0n})h_1^2 + h_1^3) \] for some \( C < \infty \) and for sufficiently large \( n \).

Meanwhile, there exist \( h_1, h_2 > 0 \) such that if \( |\gamma - \gamma_{0n}| \leq h_1 \) and \( \|\alpha - \alpha_{0n}\| \leq h_2 \), then

\[ \| (M_{20n}(\gamma) - M_{20n})(\delta - \delta_{0n}) \| < Ch_2h_1 \] for some \( C < \infty \) and for sufficiently large \( n \). This is because for sufficiently small \( h_1 > 0 \), \( \|M_{20n}(\gamma) - M_{20n}\| < \sup_{|\gamma - \gamma_{0n}| \leq h_1} \|S_n(\gamma)\| / h_1 \), where

\[ S_n(\gamma) = \begin{pmatrix} E_{10n}[z_{10n}(1, \gamma)]f_{10n}(\gamma) - E_{10-1,n}[z_{10n}(1, \gamma)]f_{10-1,n}(\gamma) \\ \vdots \\ E_{Tn}[z_{Tn}(1, \gamma)]f_{Tn}(\gamma) - E_{T-1,n}[z_{Tn}(1, \gamma)]f_{T-1,n}(\gamma) \end{pmatrix} , \]

Note that if \( h_1 \) is sufficiently small, \( \sup_{|\gamma - \gamma_{0n}| \leq h_1} \|S_n(\gamma)\| \) is bounded above by some nonnegative constant \( C < \infty \), and \( \|M_{20n}(\gamma) - M_{20n}\| < Ch_1 \).

Hence, for any \( \eta > 0 \), there exist \( h_1, h_2 > 0 \) such that if \( |\gamma - \gamma_{0n}| \leq h_1 \) and \( \|\alpha - \alpha_{0n}\| \leq h_2 \), then

\[ \| (M_{20n}(\gamma) - M_{20n})[\delta_{0n} + (\delta - \delta_{0n})] - \tilde{H}_n[\{\delta_{10n} + \delta_{30n}\gamma_{0n}\}(\gamma - \gamma_{0n}) + \frac{\delta_{0n}}{2}(\gamma - \gamma_{0n})^2] \| < C(h_1h_2 + (\delta_{10n} + \delta_{30n}\gamma_{0n})h_1^2 + h_1^3), \]

for some nonnegative \( C < \infty \) and sufficiently large \( n \). Therefore, for any \( \eta > 0 \), we can set \( h_1 \) and \( h_2 \) sufficiently small such that \( \sup_{|\gamma - \gamma_{0n}| \leq h_1, \|\alpha - \alpha_{0n}\| \leq h_2} \|g_{0n}(\theta) - M_{0n}(\alpha - \alpha_{0n}) - \tilde{H}_n[\{\delta_{10n} + \delta_{30n}\gamma_{0n}\}(\gamma - \gamma_{0n}) + \frac{\delta_{0n}}{2}(\gamma - \gamma_{0n})^2]\| \leq \sqrt{h_2 + (\delta_{10n} + \delta_{30n}\gamma_{0n})h_1 + h_1^2} \eta \) for sufficiently large \( n \), which completes the proof.

\[ \square \]

**Lemma H.3.** Let \( \{\phi_{0n} \in \Phi_0 : n \geq 1\} \) and \( \pi_n(\phi_{0n}) \to (\zeta_1, \zeta_2, \phi_{0,\infty}) \in \Pi \). Then,

\[ \sup_{\gamma \in \Gamma} \| M_n(\gamma) - M_0(\gamma) \| \xrightarrow{P} 0. \]

**Proof.** We show that the classes \( \{z_{it}(1, q_{it})1\{q_{it} > \gamma\} : \gamma \in \Gamma\} \) and \( \{z_{it}(1, q_{it-1})1\{q_{it-1} > \gamma\} : \gamma \in \Gamma\} \) are Glivenko-Cantelli uniformly in \( \{P_n : n = 1, 2, \ldots\} \), where \( P_n \) is the probability law of \( \omega_{in} = \{z_{itn}, y_{itn}, x_{itn}, \epsilon_{itn}\}_{i=1}^{T} \). We focus on the former class since the verification for the latter class is exactly identical. As it is sufficient to show that each element of \( \{z_{it}(1, q_{it})1\{q_{it} > \gamma\} : \gamma \in \Gamma\} \), we additionally restrict our focus on \( G_{m-index} = \{z_{it}q_{it}1\{q_{it} > \gamma\} : \gamma \in \Gamma\} \) and assume that \( z_{it} \) is scalar without losing of generality. By Theorem 2.8.1 in van der Vaart and Wellner (1996), \( G_{m-index} \) is Glivenko-Cantelli uniformly in \( \{P_n\} \) if

\[ \sup_{n \in \mathbb{N}} E|G_{m-index}(\omega_{in})|^{1+r} < \infty \text{ for some } r > 0, \text{ and} \]

\[ \sup_{Q} \log N(\varepsilon\|G_{m-index}\|_{Q,1}, G_{m-index}, L_{1}(Q)) < \infty \text{ for all } \varepsilon > 0, \]

S-40
where supremum is taken over all probability measures $Q$ such that $QG^\text{m-index} < \infty$, and $G^\text{m-index} = |z_{it}q_{it}|$ is an envelope of $G^\text{m-index}$. The first condition holds because 
\[ \sup_{n \in \mathbb{N}} E[z_{itn}q_{itn}]^{1+r} \leq \sup_{n \in \mathbb{N}} (E[z_{itn}^{2+2r}]^{1/2}/E[q_{itn}]^{2+2r})^{1/2} < C \text{ for some } C < \infty \text{ and } r > 0. \]

The second condition holds as we have shown in the proof of Lemma C.2 that $G^\text{m-index}$ is a VC class that satisfies the uniform entropy condition. Therefore, the ULLN with triangular array holds for $\{z_{it}q_{it}1\{q_{it} > \gamma\} : \gamma \in \Gamma\}$. 

\[ \square \]

**Lemma H.4.** Let $\{\phi_0 \in \Phi_0 : n \geq 1\}$ and $\pi_n(\phi_0) \rightarrow (\zeta_1, \zeta_2, \phi_{0, \infty}) \in \Pi$. Suppose that $|\hat{\theta}(1)_n - \theta_0| \overset{P}{\rightarrow} 0$. Then,

\[ \left\| \left\{ \frac{1}{n} \sum_{i=1}^{n} [g(\omega_{in}, \hat{\theta}(1)_n)g(\omega_{in}, \hat{\theta}(1)_n)'] - \bar{g}_n(\hat{\theta}(1)_n)\bar{g}_n(\hat{\theta}(1)_n)'/2 \right\} - \Omega_n \right\| \overset{P}{\rightarrow} 0, \]

where $\Omega_n = E[g(\omega_{in}, \theta_0)g(\omega_{in}, \theta_0)']/2 - g_0(\theta_0)g_0(\theta_0)'$.

**Proof.** We need to show $\left\| \bar{g}_n(\hat{\theta}(1)_n) - g_0(\theta_0) \right\| \overset{P}{\rightarrow} 0$ and $\left\| \frac{1}{n} \sum_{i=1}^{n} g(\omega_{in}, \hat{\theta}(1)_n)g(\omega_{in}, \hat{\theta}(1)_n)'/E[g(\omega_{in}, \theta_0)g(\omega_{in}, \theta_0)'] \overset{P}{\rightarrow} 0$. $G = \{g(\omega, \theta) : \theta \in \Theta\}$ is Glivenko-Cantelli class uniformly with respect to $\{P_n : n = 1, 2, \ldots\}$, where $P_n$ is the probability law of $\omega_{in} = \{z_{in}, y_{itn}, x_{itn}, \epsilon_{itn}\}_{i=1}^{n}$, as the proof of Lemma H.5 shows that the class is uniformly Donsker and pre-Gaussian. Therefore, $\left\| \bar{g}_n(\hat{\theta}(1)_n) - g_0(\theta_0) \right\| \overset{P}{\rightarrow} 0$ when $|\hat{\theta}(1)_n - \theta_0| \overset{P}{\rightarrow} 0$.

Let $G^2 = \{g(\omega, \theta)g(\omega, \theta) : \theta \in \Theta\}$. If $G^2$ is Glivenko-Cantelli class uniformly with respect to $\{P_n\}$, then $\sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} g(\omega_{in}, \theta)g(\omega_{in}, \theta)' - E[g(\omega_{in}, \theta)g(\omega_{in}, \theta)'] \right\| \overset{P}{\rightarrow} 0$. Then, $\left\| \frac{1}{n} \sum_{i=1}^{n} g(\omega_{in}, \hat{\theta}(1)_n)g(\omega_{in}, \hat{\theta}(1)_n)'/E[g(\omega_{in}, \theta_0)g(\omega_{in}, \theta_0)'] \overset{P}{\rightarrow} 0$ as $|\hat{\theta}(1)_n - \theta_0| \overset{P}{\rightarrow} 0$. By Theorem 2.8.1 in van der Vaart and Wellner (1996), $G^2$ is Glivenko-Cantelli uniformly in $\{P_n\}$ if

\[ \sup_{n \in \mathbb{N}} E[G^2(\omega_{in})]^{1+r} < \infty \text{ for some } r > 0, \]

\[ \sup_{Q} \log N(\varepsilon ||G^2||_{Q,1}, G^2, L_1(Q)) < \infty \text{ for all } \varepsilon > 0, \]

where supremum is taken over all probability measures $Q$ such that $QG^2 < \infty$, and $G^2 = \left\{ \sum_{i=1}^{T} C(\|z_{it}^2\|, \|z_{it}^2\|, \|z_{it}(1, q_{it}')\|, \|z_{it}(1, q_{it} - 1)\|, \|z_{it}(1, q_{it}' - 1)\|) \right\}^2$ for some $C < \infty$ is an envelope of $G^2$ as $G$ is an envelope of $G$ as shown in the proof of Lemma C.3. The first condition $\sup_{n \in \mathbb{N}} E[G^2(\omega_{in})]^{1+r} < \infty$ holds because $\sup_{n \in \mathbb{N}} \max \left\{ (E||z_{itn}^2||^{4+2r})^{1/2}, (E||z_{itn}^2||^{4+2r})\right\}^{1/2}, (E||\epsilon_{itn}^2||^{4+2r})^{1/2}, (E||\Delta\epsilon_{itn}^2||^{4+2r})^{1/2} < \infty \text{ for some } r > 0$. The second condition holds because $G$ satisfies the uniform entropy condition (see the proof of Lemma C.3) while pairwise product preserves uniform entropy condition, e.g. Theorem 2.10.20 in van der Vaart and Wellner (1996).

\[ \square \]

**Lemma H.5.** Let $\{\phi_0 \in \Phi_0 : n \geq 1\}$ and $\pi_n(\phi_0) \rightarrow (\zeta_1, \zeta_2, \phi_{0, \infty}) \in \Pi$. If $h_n \rightarrow 0$, then

\[ \sup_{\theta_1 - \theta_2 < h_n} \sqrt{n} \left\| \bar{g}_n(\theta_1) - \bar{g}_n(\theta_2) - g_0(\theta_1) + g_0(\theta_2) \right\| = o_p(1). \]

S-41
Proof. Let \( P_n \) be a probability law of \( \omega_{in} = \{(z_{itn}, y_{itn}, x_{itn}, \epsilon_{itn})^T_{i=1}\} \). We show that the class \( \mathcal{G} = \{g(\omega_n, \theta) : \theta \in \Theta\} \) is pre-Gaussian uniformly in \( \{P_n : n = 1, 2, \ldots\} \) (see Section 2.8.2 in van der Vaart and Wellner (1996) for its definition), which implies asymptotic equicontinuity uniform in \( \{P_n\} \). That is, for any \( \epsilon > 0 \), \( \sup_{m \in \mathbb{N}} P_m(\|G_n\|_G > \epsilon) \to 0 \) if \( h \to 0 \) and \( n \to \infty \), while \( \mathcal{G}_h = \{g(\omega_n, \theta_1) - g(\omega_n, \theta_2) : \|\theta_1 - \theta_2\| < h\} \). Let \( G \) be an envelope of \( \mathcal{G} \). By Theorem 2.8.3 in van der Vaart and Wellner (1996), it is sufficient to show that

\[
\sup_{n \in \mathbb{N}} \mathbb{E}|G(\omega_{in})|^2 + r < \infty \text{ for some } r > 0, \quad \int_0^\infty \sup_{Q} \log N(\|G\|_{Q, 2}, \mathcal{G}, L_2(Q))d\epsilon < \infty,
\]

where \( Q \) ranges over all finitely discrete probability measures, which implies that \( \mathcal{G} \) is Donsker and uniformly pre-Gaussian in \( \{P_n\} \).

Let \( \tilde{G}^{(t)} = \{z_{it} \Delta x_{it} - z_{it} \Delta x_{it} \beta - z_{it} 1_{it}(\gamma_i)'X_{it} \delta_1 + z_{it} 1_{it}(\gamma_i)'X_{it} \delta_2 : \|\beta\| \leq K, \|\delta_1\| \leq K, \|\delta_2\| \leq K, \gamma_i, \delta_2 \in \Gamma\} \). Suppose that \( z_{it} \) is a scalar without losing of generality as it is sufficient to show the conditions hold for each element of \( \mathcal{G} \). Note that \( \tilde{g}(\omega_n, \theta) = z_{it}(\Delta y_{it} - \Delta x_{it} \beta - 1_{it}(\gamma_i)'X_{it} \delta) = z_{it}(\Delta u_{it} - \Delta x_{it} (\beta - \beta_0) - z_{it} 1_{it}(\gamma_i)'X_{it} \delta + z_{it} 1_{it}(\gamma_i)'X_{it} \delta_0) \) is an element of \( \tilde{G}^{(t)} \) for any \( \theta_0 \in \Theta \). So it is sufficient to show \( \tilde{G}^{(t)} \) is pre-Gaussian uniformly in \( \{P_n\} \) instead of each element of \( \mathcal{G} \).

\( \tilde{G}(\omega) = C(\|z_{it} \Delta x_{it}\| + \|z_{it}(1_{it} - \beta)\| + \|z_{it}(1_{it} - \beta_0)\|) = \tilde{G}^{(t)} \) for some \( C < \infty \). The first condition for the uniform pre-Gaussianity \( \sup_{n \in \mathbb{N}} \mathbb{E}|\tilde{G}(\omega_{in})|^2 + r < \infty \) holds as \( \sup_{n \in \mathbb{N}} \max \{\mathbb{E}|\tilde{g}(\omega_n, \theta)|^{2+2r}, \mathbb{E}|\tilde{g}(\omega_n, \theta)|^{2+2r} / 2, \mathbb{E}|\tilde{g}(\omega_n, \theta)|^{2+2r} / 2, \mathbb{E}|\tilde{g}(\omega_n, \theta)|^{2+2r} / 2\} < \infty \) for some \( r > 0 \). The second condition holds as \( \tilde{G}^{(t)} \) is shown to satisfy the uniform entropy condition in the proof of Lemma C.3.

\[\square\]

Lemma H.6. Let \( \{\phi_{0n} \in \Phi_0 : n \geq 1\} \) and \( \pi_n(\phi_{0n}) \to (\zeta_1, \zeta_2, \phi_{0, \infty}) \in \Pi \), and suppose that \( \zeta_1 \neq \{\pm \infty\} \), and \( \zeta_2 = 0 \), i.e., it is (i) continuous or (ii) semi-continuous. Then,

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^n z_{itn}(1_{itn}(\gamma_{0n})' - 1_{itn}(\gamma_{0n} + \frac{b}{n^q})')X_{itn} \delta_{0n} \xrightarrow{P} \{E_{t, \infty}[z_{it, \infty}|\gamma_{0, \infty}]f_{t, \infty}(\gamma_{0, \infty}) - E_{t-1, \infty}[z_{it, \infty}|\gamma_{0, \infty}]f_{t-1, \infty}(\gamma_{0, \infty})\}[\zeta_1 b + \frac{\delta_{00, \infty} b^2}{2}]
\]

uniformly over \( b \in [-K, K] \) for any \( K < \infty \).

Proof. Note that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^n z_{itn}(1_{itn}(\gamma_{0n})' - 1_{itn}(\gamma_{0n} + \frac{b}{n^q})')X_{itn} \delta_{0n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ z_{itn}(1_{itn}(\gamma_{0n})' - 1_{itn}(\gamma_{0n} + \frac{b}{n^q})')X_{itn} \delta_{0n} - E[z_{itn}(1_{itn}(\gamma_{0n})' - 1_{itn}(\gamma_{0n} + \frac{b}{n^q})')X_{itn} \delta_{0n}] \right\}
\]

(H.2)

\[+ \sqrt{n}E[z_{itn}(1_{itn}(\gamma_{0n})' - 1_{itn}(\gamma_{0n} + \frac{b}{n^q})')X_{itn} \delta_{0n}]\] (H.3)
The stochastic term (H.2) converges in probability to zero uniformly with respect to \( b \in [-K, K] \). This is because Lemma H.5 shows that when \( h_n \downarrow 0 \), then

\[
\sup_{\vert \gamma - \gamma_0 \vert < h_n} \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^{n} z_{itn}(1_{\gamma}(\gamma_0 n) - 1_{\gamma}(\gamma))'X_{itn}\delta_{0n} - E[z_{itn}(1_{\gamma}(\gamma_0 n) - 1_{\gamma}(\gamma))'X_{itn}\delta_{0n}] \right\} = o_p(1)
\]

as it can be expressed as \( \sup_{\vert \gamma - \gamma_0 \vert < h_n} \| \tilde{g}_n(\alpha_0 n, \gamma) - \tilde{g}_n(\alpha_0 n, \gamma_0) - g_0(\alpha_0 n, \gamma) + g_0(\alpha_0 n, \gamma_0) \| \).

Suppose \( b > 0 \). The case for \( b < 0 \) follows similarly. We will show that (H.3) converges as follows:

\[
\sqrt{n}E[z_{itn}(1_{\gamma}(\gamma_0 n) - 1_{\gamma}(\gamma_0 n + \frac{b}{n^\alpha})')X_{itn}\delta_{0n} = \sqrt{n} \left\{ E[z_{itn}(\delta_{10n} + \delta_{30n}q_{itn}) 1\{\gamma_0 n + \frac{b}{n^\alpha} \geq q_{itn} > \gamma_0 n\}] - E[z_{itn}(\delta_{10n} + \delta_{30n}q_{itn-1,n}) 1\{\gamma_0 n + \frac{b}{n^\alpha} \geq q_{itn-1,n} > \gamma_0 n\}] \right\}
\]

\[
\rightarrow \{ E_{t,\infty}[z_{it,\infty}f_{t,\infty}(\gamma_0,\gamma) - E_{t-1,\infty}[z_{it,\infty}f_{t-1,\infty}(\gamma_0,\gamma)] \} [\gamma b + \frac{\delta_{30n} b^2}{2}],
\]

uniformly with respect to \( b \in [-K, K] \).

Let

\[
R_{n,b} = \left( \sqrt{n}E[z_{itn}(\delta_{10n} + \delta_{30n}q_{itn}) 1\{\gamma_0 n + \frac{b}{n^\alpha} \geq q_{itn} > \gamma_0 n\}] - E_{t}z_{itn}\gamma_0 n]f_{t}(\gamma_0 n)](n^{1/4}(\delta_{10n} + \delta_{30n}\gamma_0 n)b + \frac{\delta_{30n} b^2}{2}) \right),
\]

which will be shown to converge to zero uniformly with respect to \( b \in [-K, K] \). By Taylor expansion, its formula can be derived as follows:

\[
R_{n,b} = \left( \delta_{30n}\{E_{t}[z_{itn}\gamma_n,b]f_{t}(\gamma_n,b) - E_{t}[z_{itn}\gamma_0 n]f_{t}(\gamma_0 n)\}ight.
\]

\[
+ (\delta_{10n} + \delta_{30n}\gamma_n,b) \frac{d}{d\gamma}E_{t}[z_{itn}\gamma]f_{t}(\gamma) \vert_{\gamma = \gamma_n,b} b^2 \frac{1}{2},
\]

where \( \gamma_n,b \in [\gamma_0 n, \gamma_0 n + \frac{b}{n^\alpha}] \). Note that \( |\gamma_n,b - \gamma_0 n| \rightarrow 0 \) uniformly with respect to \( b \in [-K, K] \).

Hence, for sufficiently large \( n \), \( \| \frac{d}{d\gamma}E_{t}[z_{itn}\gamma]f_{t}(\gamma) \vert_{\gamma = \gamma_n,b} \| \leq C \) for some \( C < \infty \). Moreover, \( \delta_{10n} + \delta_{30n}\gamma_n,b \rightarrow 0 \) and \( E_{t}z_{itn}\gamma_0 n]f_{t}(\gamma_0 n) - E_{t}z_{itn}\gamma_n,b]f_{t}(\gamma_n,b) \rightarrow 0 \) uniformly with respect to \( b \in [-K, K] \). Therefore, \( \| R_{n,b} \| \rightarrow 0 \) uniformly with respect to \( b \in [-K, K] \), i.e.,

\[
\left( \sqrt{n}E[z_{itn}(\delta_{10n} + \delta_{30n}q_{itn}) 1\{\gamma_0 n + \frac{b}{n^\alpha} \geq q_{itn} > \gamma_0 n\}] - \{E_{t}[z_{itn}\gamma_0 n]f_{t}(\gamma_0 n)](n^{1/4}(\delta_{10n} + \delta_{30n}\gamma_0 n)b + \frac{\delta_{30n} b^2}{2}) \right) \rightarrow 0
\]

uniformly with respect to \( b \in [-K, K] \). We can derive a similar result for \( \sqrt{n}E[z_{itn}(\delta_{10n} + \frac{z_{itn}(\delta_{10n} + \delta_{30n}q_{itn}) 1\{\gamma_0 n + \frac{b}{n^\alpha} \geq q_{itn} > \gamma_0 n\}] - \{E_{t}[z_{itn}\gamma_0 n]f_{t}(\gamma_0 n)](n^{1/4}(\delta_{10n} + \delta_{30n}\gamma_0 n)b + \frac{\delta_{30n} b^2}{2}) \right) \rightarrow 0
\]

uniformly with respect to \( b \in [-K, K] \).
\[ \delta_{30nq_{it-1,n}}1\{\gamma_0 + \frac{b}{n^\frac{1}{4}} \geq q_{it-1,n} > \gamma_0\} \]

that leads to

\[ \left\| \sqrt{n}E\{z_{itn}(\gamma_0)' - 1_{itn}(\gamma_0 + \frac{b}{n^\frac{1}{4}})'\}X_{itn}\delta_{0n} \right\| - \{E_{itn}[z_{itn}(\gamma_0)]f_{tn}(\gamma_0) - E_{it-1,n}[z_{itn}(\gamma_0)f_{t-1,n}(\gamma_0)]\} \left[ n^{1/4}(\delta_{10n} + \delta_{30n}\gamma_0)b + \frac{\delta_{30n}}{2}b^2 \right] \to 0, \]

uniformly with respect to \( b \in [-K, K] \). As \( \pi_n(\phi_{0n}) \to (\zeta_1, \zeta_2, \phi_{0,\infty}) \),

\[ \{E_{itn}[z_{itn}(\gamma_0)]f_{tn}(\gamma_0) - E_{it-1,n}[z_{itn}(\gamma_0)f_{t-1,n}(\gamma_0)]\} \left[ n^{1/4}(\delta_{10n} + \delta_{30n}\gamma_0)b + \frac{\delta_{30n}}{2}b^2 \right] \]

\[ \to \{E_{t,\infty}[z_{it,\infty}|\gamma_0,\infty]f_{t,\infty}(\gamma_0,\infty) - E_{t-1,\infty}[z_{it,\infty}|\gamma_0,\infty]f_{t-1,\infty}(\gamma_0,\infty)\} \left[ \zeta_1b + \frac{\delta_{30,\infty}}{2}b^2 \right], \]

which completes the proof. \( \Box \)

**Lemma H.7.** Let \( \{\phi_{0n} \in \Phi_0 : n \geq 1\} \) and \( \pi_n(\phi_{0n}) \to (\zeta_1, \zeta_2, \phi_{0,\infty}) \in \Pi \), and suppose that \( \zeta_1 = \{\pm \infty\} \) and \( \zeta_2 = 0 \), i.e., it is (iii) semi-discontinuous. Then,

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{itn}(1_{itn}(\gamma_0)' - 1_{itn}(\gamma_0 + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_0)})')X_{itn}\delta_{0n} \]

\[ \overset{p}{\to} \{E_{t,\infty}[z_{it,\infty}|\gamma_0,\infty]f_{t,\infty}(\gamma_0,\infty) - E_{t-1,\infty}[z_{it,\infty}|\gamma_0,\infty]f_{t-1,\infty}(\gamma_0,\infty)\} \left( \zeta_1b + \frac{\delta_{30,\infty}}{2}b^2 \right) \]

uniformly over \( b \in [-K, K] \) for any \( K < \infty \).

**Proof.** Note that

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{itn}(1_{itn}(\gamma_0)' - 1_{itn}(\gamma_0 + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_0)})')X_{itn}\delta_{0n} \]

\[ = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ z_{itn}(1_{itn}(\gamma_0)' - 1_{itn}(\gamma_0 + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_0)})')X_{itn}\delta_{0n} \right. \]

\[ - E\left[z_{itn}(1_{itn}(\gamma_0)' - 1_{itn}(\gamma_0 + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_0)})')X_{itn}\delta_{0n}\right] \left( \zeta_1b + \frac{\delta_{30,\infty}}{2}b^2 \right) \]

The stochastic term (H.4) converges in probability to zero uniformly with respect to \( b \in [-K, K] \) by **Lemma H.5**, by an argument similar to the proof of **Lemma H.6** that shows (H.2) converges to zero.

Suppose \( b > 0 \). The case for \( b < 0 \) follows similarly. We will show that (H.5) converges as
follows:

\[
\sqrt{n}E[z_{itn}(\gamma_{0n})'] - 1_{itn}(\gamma_{0n} + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_{0n})})'X_{itn}\delta_{0n} \\
= \sqrt{n}\left\{E[z_{itn}(\delta_{10n} + \delta_{30n}q_{itn})1(\gamma_{0n} + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_{0n})} \geq q_{itn} > \gamma_{0n})] \\
- E[z_{itn}(\delta_{10n} + \delta_{30n}q_{it-1,n})1(\gamma_{0n} + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_{0n})} \geq q_{it-1,n} > \gamma_{0n})]\right\} \\
\to \{E_{t,\infty}[z_{it,\infty}\gamma_{0,\infty}|f_{t,\infty}(\gamma_{0,\infty})] - E_{t-1,\infty}[z_{it,\infty}|\gamma_{0,\infty}]f_{t-1,\infty}(\gamma_{0,\infty})\}b,
\]

uniformly with respect to \(b \in [-K, K]\).

Let

\[
R_{n,b} = \left(\sqrt{n}E[z_{itn}(\delta_{10n} + \delta_{30n}q_{itn})1(\gamma_{0n} + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_{0n})} \geq q_{itn} > \gamma_{0n})] - \{E_{tn}[z_{itn}|\gamma_{0n}]f_{tn}(\gamma_{0n})\}b\right),
\]

which will be shown to converge to zero uniformly with respect to \(b \in [-K, K]\). By Taylor expansion, its formula can be derived as follows:

\[
R_{n,b} = \frac{1}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_{0n})^2}\delta_{30n}\{E_{tn}[z_{itn}|\gamma_{n,b}]f_{tn}(\gamma_{n,b})\} \\
+ (\delta_{10n} + \delta_{30n}\gamma_{0n})\frac{d}{d\gamma}\{E_{tn}[z_{itn}|\gamma]f_{tn}(\gamma)\}|_{\gamma = \gamma_{n,b}}\frac{b^2}{2},
\]

where \(\gamma_{n,b} \in [\gamma_{0n}, \gamma_{0n} + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_{0n})}]\). Note that \(|\gamma_{n,b} - \gamma_{0n}| \to 0\) uniformly with respect to \(b \in [-K, K]\). Hence, for sufficiently large \(n\), \(\|E_{tn}[z_{itn}|\gamma_{n,b}]f_{tn}(\gamma_{n,b})\| \leq C\) and \(\|\frac{d}{d\gamma}\{E_{tn}[z_{itn}|\gamma]f_{tn}(\gamma)\}\|_{\gamma = \gamma_{n,b}} \leq C\) for some \(C < \infty\). Moreover, \((\delta_{10n} + \delta_{30n}\gamma_{0n})^2 \to \infty\), \(\|R_{n,b}\| \to 0\) uniformly with respect to \(b \in [-K, K]\), i.e.,

\[
\left(\sqrt{n}E[z_{itn}(\delta_{10n} + \delta_{30n}q_{itn})1(\gamma_{0n} + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_{0n})} \geq q_{itn} > \gamma_{0n})] - \{E_{tn}[z_{itn}|\gamma_{0n}]f_{tn}(\gamma_{0n})\}b\right) \to 0
\]

uniformly with respect to \(b \in [-K, K]\). We can derive a similar result for \(\sqrt{n}E[z_{itn}(\delta_{10n} + \delta_{30n}q_{it-1,n})1(\gamma_{0n} + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_{0n})} \geq q_{it-1,n} > \gamma_{0n})]\) that leads to

\[
\left\|\sqrt{n}E[z_{itn}(\gamma_{0n})' - 1_{itn}(\gamma_{0n} + \frac{b}{\sqrt{n}(\delta_{10n} + \delta_{30n}\gamma_{0n})})')X_{itn}\delta_{0n} \\
- \{E_{tn}[z_{itn}|\gamma_{0n}]f_{tn}(\gamma_{0n}) - E_{t-1,n}[z_{itn}|\gamma_{0n}]f_{t-1,n}(\gamma_{0n})\}b\right\| \to 0,
\]

uniformly with respect to \(b \in [-K, K]\). As \(\pi_n(\phi_{0n}) \to (\zeta_1, \zeta_2, \phi_{0,\infty})\),

\[
\{E_{tn}[z_{itn}|\gamma_{0n}]f_{tn}(\gamma_{0n}) - E_{t-1,n}[z_{itn}|\gamma_{0n}]f_{t-1,n}(\gamma_{0n})\}b \\
\to \{E_{t,\infty}[z_{it,\infty}|\gamma_{0,\infty}]f_{t,\infty}(\gamma_{0,\infty}) - E_{t-1,\infty}[z_{it,\infty}|\gamma_{0,\infty}]f_{t-1,\infty}(\gamma_{0,\infty})\}b,
\]

which completes the proof. \(\square\)
Lemma H.8. Let \( \{ \phi_{0n} \in \Phi_0 : n \geq 1 \} \) and \( \pi_n(\phi_{0n}) \to (\zeta_1, \zeta_2, \phi_{0,\infty}) \in \Pi \). Then,

\[
\hat{u}_n^* = \left( \frac{1}{n} \sum_{i=1}^{n} z_{it0n}^* \Delta_{it0n}^* \right) - \left( \frac{1}{n} \sum_{i=1}^{n} z_{iTn}^* \Delta_{iTn}^* \right) \xrightarrow{p^*} 0 \text{ in } P.
\]

Proof. Note that \( \hat{u}_n^* = \frac{1}{n} \sum_{i=1}^{n} [g(\omega_{in}, \hat{\theta}_n) - E[g(\omega_{in}, \hat{\theta}_n)]] - \frac{1}{n} \sum_{i=1}^{n} [g(\omega_{in}, \hat{\theta}_n) - E[g(\omega_{in}, \hat{\theta}_n)]]\). Let \( P_n \) be the probability law of \( \omega_{in} = \{(\hat{\eta}_{itn}, y_{itn}, x_{itn}, \epsilon_{itn})\}_{i=1}^{T} \). As \( \mathcal{G} = \{g(\omega, \theta) : \theta \in \Theta\} \) is Glivenko-Cantelli uniformly in \( \{P_n\} \), which is shown in the proof of Lemma H.5, \( \frac{1}{n} \sum_{i=1}^{n} [g(\omega_{in}, \hat{\theta}_n) - E[g(\omega_{in}, \hat{\theta}_n)]] \) is \( o_p(1) \), and hence \( o_p^*(1) \) in \( P \) by Lemma B.1. By Proposition 2, \( \frac{1}{n} \sum_{i=1}^{n} [g(\omega_{in}, \hat{\theta}_n) - E[g(\omega_{in}, \hat{\theta}_n)]] \) is also \( o_p^*(1) \) in \( P \), which completes the proof. \( \square \)

Lemma H.9. Let \( \{ \phi_{0n} \in \Phi_0 : n \geq 1 \} \) and \( \pi_n(\phi_{0n}) \to (\zeta_1, \zeta_2, \phi_{0,\infty}) \in \Pi \). Then,

\[
\sqrt{n} \hat{u}_n^* = \sqrt{n} \left\{ \left( \frac{1}{n} \sum_{i=1}^{n} z_{it0n}^* \Delta_{it0n}^* \right) - \left( \frac{1}{n} \sum_{i=1}^{n} z_{iTn}^* \Delta_{iTn}^* \right) \right\} \xrightarrow{d^*} N(0, \Omega_\infty) \text{ in } P.
\]

Proof. Note that \( \sqrt{n} \hat{u}_n^* = \sqrt{n} \{\hat{\theta}_n - \theta_0\} - \{\hat{\theta}_n - \theta_0\} + \sqrt{n} \{\hat{g}_n^{*}(\theta_{0n}) - \hat{g}_n(\theta_{0n})\}. \) As \( \|\hat{\theta}_n - \theta_0\| = o_p(1) \) and \( o_p^*(1) \) in \( P \) by Lemma B.1, \( \sqrt{n} \{\hat{g}_n^{*}(\theta_{0n}) - \hat{g}_n(\theta_{0n})\} \) is \( o_p^*(1) \) in \( P \). By applying Lemma H.18, \( \sqrt{n} \lambda' \{\hat{g}_n^{*}(\theta_{0n}) - \hat{g}_n(\theta_{0n})\} \xrightarrow{d^*} N(0, \lambda' \Omega_\infty \lambda) \) in \( P \) for any real vector \( \lambda \). By Cramér-Wold, \( \sqrt{n} \{\hat{g}_n^{*}(\theta_{0n}) - \hat{g}_n(\theta_{0n})\} \xrightarrow{d^*} N(0, \Omega_\infty) \) in \( P \), and applying Slutsky theorem completes the proof. \( \square \)

The Lemma H.10 states uniform bootstrap probability limit of the following matrix:

\[
\hat{M}_n^* (\gamma) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} z_{it0n}^* \Delta_{it0n}^* \cdot X_{it0n}^* \\ \vdots \\ z_{iTn}^* \Delta_{iTn}^* \cdot X_{iTn}^* \end{pmatrix}.
\]

Lemma H.10. Let \( \{ \phi_{0n} \in \Phi_0 : n \geq 1 \} \) and \( \pi_n(\phi_{0n}) \to (\zeta_1, \zeta_2, \phi_{0,\infty}) \in \Pi \). Then,

\[
\sup_{\gamma \in \Gamma} \| \hat{M}_n^* (\gamma) - M_n(\gamma) \| \xrightarrow{p^*} 0 \text{ in } P.
\]

Proof. We apply Proposition 2 to prove the result. First, we need to show that \( \{ z_{it(1,q_{it})\{q_{it} > \gamma \}} : \gamma \in \Gamma \} \) and \( \{ z_{it(1,q_{it-1})\{q_{it-1} > \gamma \}} : \gamma \in \Gamma \} \) are Glivenko-Cantelli uniformly in \( \{P_n : n = 1, 2, ...\} \), where \( P_n \) is the probability law of \( \omega_{in} = \{(\hat{\eta}_{itn}, y_{itn}, x_{itn}, \epsilon_{itn})\}_{i=1}^{T} \). It is shown in Lemma H.3 that the functional classes are Glivenko-Cantelli uniformly in \( \{P_n\} \). Second, the condition for envelope holds as \( \sup_{n \in \mathbb{N}} E[\| z_{it(1,q_{it})\{q_{it} > \gamma \} \| + \| z_{it(1,q_{it-1},n)} \| < \infty \), which is implied by \( \sup_{n \in \mathbb{N}} \max \{ (E[\| z_{itn} \|^{2+r})^{1/2}, (E[\| q_{itn} \|^{2+r})^{1/2}, (E[\| q_{it-1,n} \|^{2+r})^{1/2} < \infty \) for some \( r > 0 \). \( \square \)
Lemma H.11. Let \( \{ \phi_{0n} \in \Phi_0 : n \geq 1 \} \) and \( \pi_n(\phi_{0n}) \to (\zeta_1, \zeta_2, \phi_{0,\infty}) \in \Pi \). If \( h_n \to 0 \), then
\[
\sup_{\|\theta_1 - \theta_2\| < h_n} \sqrt{n} \| \hat{g}_n^* (\theta_1) - \hat{g}_n^* (\theta_2) - \bar{g}_n (\theta_1) + \bar{g}_n (\theta_2) \| = o_p^* (1) \text{ in } P.
\]

Proof. Note that \( \hat{g}_n^* (\theta_1) - \hat{g}_n^* (\theta_2) = \frac{1}{n} \sum_{i=1}^{n} (g(\omega_{in}, \theta_1) - g(\omega_{in}, \theta_2)) \) because \( g_n^* (\theta) = g(\omega_{in}^*, \theta) - g(\omega_{in}^*, \theta_{0n}) + g(\omega_{in}^*, \delta_n) \), see (H.1). Therefore, \( \sqrt{n} \| \hat{g}_n^* (\theta_1) - \hat{g}_n^* (\theta_2) - \bar{g}_n (\theta_1) + \bar{g}_n (\theta_2) \| = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ g(\omega_{in}^*, \theta_1) - g(\omega_{in}^*, \theta_2) - g(\omega_{in}, \theta_1) + g(\omega_{in}, \theta_2) \} \). Let \( \bar{G}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\delta_{\omega_{in}^*} - \bar{P}_n) \) and \( \bar{P}_n = n^{-1} \sum_{i=1}^{n} \delta_{\omega_{in}} \), where \( \delta_{\omega_{in}^*} \) and \( \delta_{\omega_{in}} \) are dirac masses at \( \omega_{in}^* \) and \( \omega_{in} \). Then, it is sufficient to prove \( \| \bar{G}_n \|_{G_h} = o_p^* (1) \) in \( P \) if \( h \to 0 \) and \( n \to \infty \).

For \( h > 0 \), let \( G_h = \{ g(\omega_i, \theta_1) - g(\omega_i, \theta_2) : \| \theta_1 - \theta_2 \| \leq h \} \) and \( G_h \) be its envelope. Let \( \tilde{N}_1, \tilde{N}_2, \ldots \) be symmetrized Poisson random variables with parameter 1/2. By Lemma H.14,
\[
E^* \| \hat{G}_n \|_{G_h} \leq 4 E_{\tilde{N}} \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{N}_i \delta_{\omega_{in}} \|_{G_h}
\]
conditionally on \( \{ \omega_{in} : 1 \leq i \leq n \} \). For all \( 1 \leq n_0 \leq n \), the last display is stochastically bounded up to constant by
\[
(n_0 - 1) E_{\tilde{N}} \max_{1 \leq i \leq n} \frac{\tilde{N}_i}{\sqrt{n}} P G(\omega_{in}) + \| \tilde{N}_1 \|_{2,1} \max_{n_0 \leq j \leq n} E \| \frac{1}{\sqrt{j}} \sum_{i=n_0}^{j} \varepsilon_i \delta_{\omega_{in}} \|_{G_h},
\]
by Lemma H.16, where \( G(\cdot) \) is an envelope function of \( G \). The first term is bounded above by \( (n_0 - 1)2 \sqrt{2} n^{-1/4} \), which converges to zero for any \( n_0 \) as \( n \to \infty \), and \( \| \tilde{N}_1 \|_{2,1} \leq 2 \sqrt{2} \) (see proof of Theorem 3.6.3 in van der Vaart and Wellner (1996)). By triangle inequality,
\[
\max_{n_0 \leq j \leq n} E \| \frac{1}{\sqrt{j}} \sum_{i=n_0}^{j} \varepsilon_i \delta_{\omega_{in}} \|_{G_h} \leq \max_{n_0 \leq j \leq n} E \left( \| \frac{1}{\sqrt{j}} \sum_{i=1}^{j} \varepsilon_i \delta_{\omega_{in}} \|_{G_h} + \| \frac{1}{\sqrt{j}} \sum_{i=1}^{n_0-1} \varepsilon_i \delta_{\omega_{in}} \|_{G_h} \right) \\
\leq 2 \max_{n_0-1 \leq j \leq n} E \| \frac{1}{\sqrt{j}} \sum_{i=1}^{j} \varepsilon_i \delta_{\omega_{in}} \|_{G_h},
\]
and the last display is bounded up to constant by
\[
\max_{n_0-1 \leq j \leq n} \left( E \sup_{\|\theta_1 - \theta_2\| \leq h} \| \frac{1}{\sqrt{j}} \sum_{i=1}^{j} \varepsilon_i (g(\omega_{in}, \theta_1) - g(\omega_{in}, \theta_2) - E[g(\omega_{in}, \theta_1)] + E[g(\omega_{in}, \theta_2)]) \| \\
+ E \sup_{\|\theta_1 - \theta_2\| \leq h} \| \frac{1}{\sqrt{j}} \sum_{i=1}^{j} \varepsilon_i (E[g(\omega_{in}, \theta_1)] - E[g(\omega_{in}, \theta_2)]) \| \right).
\]
For each $j$, by Lemma H.15,

$$
E \sup_{\|\theta_1 - \theta_2\| \leq h} \left\| \frac{1}{\sqrt{j}} \sum_{i=1}^{j} \varepsilon_i (g(\omega_i, \theta_1) - g(\omega_i, \theta_2)) - E[g(\omega_i, \theta_1)] + E[g(\omega_i, \theta_2)] \right\| 
\leq 2E \sup_{\|\theta_1 - \theta_2\| \leq h} \left\| \frac{1}{\sqrt{j}} \sum_{i=1}^{j} (g(\omega_i, \theta_1) - g(\omega_i, \theta_2)) - E[g(\omega_i, \theta_1)] + E[g(\omega_i, \theta_2)] \right\|.
$$

The right hand side of the last display converges to zero uniformly with respect to $n$ as $j \to \infty$ and $h \to 0$ since the functional class $\mathcal{G}$ is shown to be pre-Gaussian uniformly in $\{P_n\}$ in the proof of Lemma H.5.

For each $j$,

$$
E \sup_{\|\theta_1 - \theta_2\| \leq h} \left\| \frac{1}{\sqrt{j}} \sum_{i=1}^{j} \varepsilon_i (E[g(\omega_i, \theta_1)] - E[g(\omega_i, \theta_2)]) \right\| \leq E \left\| \frac{1}{\sqrt{j}} \sum_{i=1}^{j} \varepsilon_i \right\| (E\|G_h(\omega_i)\|),
$$

and $E \frac{1}{\sqrt{j}} \sum_{i=1}^{j} \varepsilon_i < \infty$ by Hoeffding’s inequality, e.g. Lemma 2.2.7 in van der Vaart and Wellner (1996). The following paragraph shows that $E\|G_h(\omega_i)\| \to 0$ as $h \to 0$ and $n \to \infty$.

As it is sufficient to consider each element of $\mathcal{G}$, we focus on $g_t(\omega_i, \theta)$, the $t$th term of $g(\omega_i, \theta)$, and assume that $g_t(\omega_i, \theta)$ is a scalar without losing of generality. Note that

$$
g_t(\omega_i, \theta_1) - g_t(\omega_i, \theta_2) = -z_{it} \Delta x_i'(\beta_1 - \beta_2) - z_{it} 1_{it}(\gamma_1')X_{it}(\delta_1 - \delta_2) + z_{it} 1_{it}(\gamma_2') - 1_{it}(\gamma_1')X_{it}\delta_2.
$$

Without losing of generality, let $\gamma_1 \geq \gamma_2$, and $K$ be a constant such that $\|\theta\| \leq K/2$ for $\theta \in \Theta$. Set

$$
G_{h,t}(\omega_i) = \|z_{it} \Delta x_i'\| \cdot h + (\|z_{it}(1,q_{it})\| + \|z_{it}(1,q_{it-1})\|) \cdot h + K(\|z_{it}(1,q_{it})\{\gamma_1 \geq q_{it} > \gamma_2\}\| + \|z_{it}(1,q_{it-1})\{\gamma_1 \geq q_{it-1} > \gamma_2\}\|),
$$

which is an envelope of $\{g_t(\omega_i, \theta_1) - g_t(\omega_i, \theta_2) : \|\theta_1 - \theta_2\| < h\}$. \sup_{n \in \mathbb{N}} E[\|z_{itn} \Delta x_{itn}'\| + \|z_{itn}(1,q_{itn})\| + \|z_{itn}(1,q_{itn-1})\|] < \infty$. Furthermore,

$$
E\|z_{itn}(1,q_{itn})\{\gamma_1 \geq q_{itn} > \gamma_2\}\| \leq (E\|z_{itn}(1,q_{itn})\|^2)^{1/2}(E1\{\gamma_1 \geq q_{itn} > \gamma_2\})^{1/2},
$$

while \sup_{n \in \mathbb{N}}(E\|z_{itn}(1,q_{itn})\|^2)^{1/2} < \infty, and

$$
E1\{\gamma_1 \geq q_{itn} > \gamma_2\} = \int_{\gamma_2}^{\gamma_1} f_{itn}(q)dq = (\gamma_1 - \gamma_2)f_{itn}(\bar{\gamma})
$$

for some $\bar{\gamma} \in [\gamma_2, \gamma_1]$. Hence, $E1\{\gamma_1 \geq q_{itn} > \gamma_2\} < Ch$ for some $C < \infty$ uniformly over all $n$. Therefore, $E|G_{h,t}(\omega_{itn})| < C\sqrt{n}$ for some $C < \infty$ and converges to zero as $h \to 0$.

Recall that the first term in (H.6) goes to zero for any fixed $n_0$ when $n \to \infty$. The second
term in (H.6) is bounded by \(2\sqrt{2}\max_{n_0 \leq j \leq n} Z_{jn}\), where \(Z_{jn} = E \| \frac{1}{\sqrt{j}} \sum_{i=n_0}^{j} \varepsilon_i \delta_{\omega_i} \|_{\mathcal{G}_n} \). It is shown in the previous paragraph that \(Z_{jn} \to 0\) uniformly with respect to \(n\) as \(j \to \infty\) and \(h \to 0\). Therefore, for any \(\epsilon > 0\), there exists \(n_0 < \infty\) such that \(\max_{n_0 \leq j \leq n} Z_{jn} < \epsilon/2\) for all \(n > n_0\). Then, there exists \(N(n_0)\) large enough such that the first term in (H.6) is bounded by \(\epsilon/2\) for \(n > N(n_0)\). In conclusion, \(E^*\|\hat{\mathcal{G}}_n\|_{\mathcal{G}_n} \to 0\) if \(h \to 0\) and \(n \to \infty\). By applying the Markov inequality, we can complete the proof.

**Lemma H.12.** Let \(\{\phi_n \in \Phi_0 : n \geq 1\}\) and \(\pi_n(\phi_0) \to (\zeta_1, \zeta_2, \phi_{0,\infty}) \in \Pi\), and suppose that \(\zeta_1 \not\in \{\pm \infty\}\), and \(\zeta_2 = 0\), i.e., it is (i) continuous or (ii) semi-continuous. Then, for any \(K < \infty\),

\[
\left\{ \sup_{b \in [-K,K]} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z^*_n(1_{itn}(\gamma_{0n})' - 1_{ite}(\gamma_{0n} + \frac{b}{n^{\alpha}})' \right) X^*_l(\delta^*_n - \delta(n)) \} \right\} = o^*_p(1) \quad \text{in} \quad P.
\]

**Proof.** As the proof is quite similar to the proofs of Lemma D.5 and Lemma H.6, we just explain direction of the proof heuristically. As \(\delta^*_n = \delta_n(\gamma_{0n})\) is consistent to \(\delta(n)\),

\[
\sup_{b \in [-K,K]} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z^*_n(1_{itn}(\gamma_{0n})' - 1_{ite}(\gamma_{0n} + \frac{b}{n^{\alpha}})' \right) X^*_l(\delta^*_n - \delta(n)) \} = o^*_p(1) \quad \text{in} \quad P.
\]

By Lemma H.11,

\[
\sup_{b \in [-K,K]} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z^*_n(1_{itn}(\gamma_{0n})' - 1_{ite}(\gamma_{0n} + \frac{b}{n^{\alpha}})' \right) X^*_l(\delta^*_n - \delta(n)) \} = o^*_p(1) \quad \text{in} \quad P,
\]

as the last display can be expressed by \(\sqrt{n} \| \tilde{g}_n(\alpha_0, \gamma_0 + \frac{b}{n^{\alpha} \tau}) - \tilde{g}_n(\alpha_0, \gamma_0) - \tilde{g}_n(\alpha_0, \gamma_0 + \frac{b}{n^{\alpha} \tau}) + \tilde{g}_n(\alpha_0, \gamma_0) \| \). Hence,

\[
\sup_{b \in [-K,K]} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z^*_n(1_{itn}(\gamma_{0n})' - 1_{ite}(\gamma_{0n} + \frac{b}{n^{\alpha}})' \right) X^*_l(\delta^*_n - \delta(n)) \} = o^*_p(1) \quad \text{in} \quad P,
\]

and applying Lemma H.6 completes the proof.
Lemma H.13. Let \( \{\phi_0 \in \Phi_0 : n \geq 1\} \) and \( \pi_n(\phi_0) \to (\zeta_1, \zeta_2, \phi_0, \infty) \in \Pi \), and suppose that \( \zeta_1 = \{\pm \infty\} \) and \( \zeta_2 = 0 \), i.e., it is (iii) semi-discontinuous. Then, for any \( K < \infty \),

\[
\sup_{b \in [-K,K]} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z^*_n(1_{itn}(\gamma_0n)' - 1_{itn}(\gamma_0n + \delta_{0n}^{\gamma_0n} + \delta_{0n}^{\gamma_0n} + \delta_{0n}^{\gamma_0n})) \right\| \rightarrow \infty
\]

is \( o_p^*(1) \) in \( P \).

Proof. We omit the proof as it is almost identical to the proof of Lemma H.12. \( \square \)

The following proposition is bootstrap Glivenko-Cantelli theorem uniform in underlying probability measures \( P \in \{P_1, P_2, \ldots\} \).

Proposition 2. Let \( \{X_i : 1 \leq i \leq n, n = 1, 2, \ldots\} \) be a triangular array of random elements in a measurable space \((X, A)\) while \( X_i \)'s are independent to each other with probability law \( P_n \), and \( F \) be a class of functions on \((X, A)\) with an envelope \( F \). Suppose that \( F \) is a Glivenko-Cantelli class uniformly in \( P \in \{P_n\} \), and \( \sup_{n \in \mathbb{N}} P_n F < \infty \). For each \( n \), let \( W = (W_1, \ldots, W_n) \) be an exchangeable nonnegative random vector independent of \( X_1, X_2, \ldots, X_n \) such that \( \sum_{i=1}^{n} W_i = 1 \) and \( \max_{1 \leq i \leq n} |W_i| \) converges to zero in probability. Then, for every \( \epsilon > 0 \) and \( \eta > 0 \), as \( n \to \infty \),

\[
P_n \left( \left\| \sum_{i=1}^{n} W_i (\delta_{X_i} - P_n) \right\|_F > \epsilon \right) > \eta \to 0,
\]

where \( \delta_{X_i} \) is a dirac measure at \( X_i \).

Let \( W = (W_1, \ldots, W_n) \) be a multinomial vector divided by \( n \) with parameters \( n \) and probabilities \( (1/n, \ldots, 1/n) \), which satisfies \( \sum_{i=1}^{n} W_i = 1 \) and \( \max_{1 \leq i \leq n} |W_i| \) converges to zero in probability. Suppose that \( \tilde{X}_1, \ldots, \tilde{X}_n \) are i.i.d. resampling draws from \( \{X_1, \ldots, X_n\} \). Then, \( \frac{1}{n} \sum_{i=1}^{n} (\delta_{\tilde{X}_i} - P_n) = \frac{1}{n} \sum_{i=1}^{n} W_i (\delta_{X_i} - P_n) \), and the probability law of \( W \) can be identified with the probability law of the empirical bootstrap conditional on the data.

Proof. Let \( Z_{in} = (\delta_{X_i} - P_n) \). By Lemma H.17,

\[
E_W \left\| \sum_{i=1}^{n} W_i Z_{in} \right\|_F \leq 2(n_0 - 1) \frac{1}{n} \|Z_{in}\|_F E_W \max_{1 \leq i \leq n} |W_i| + 2n \|W_1\|_{2,1} \max_{n_0 \leq k \leq n} E_R \| \frac{1}{k} \sum_{i=n_0}^{k} Z_{R,n} \|_F. \quad (H.7)
\]

Note that \( \frac{1}{n} \sum_{i=1}^{n} \|Z_{in}\|_F \leq \frac{1}{n} \sum_{i=1}^{n} Z_{in}(F) \leq (P_n - P_n)F + 2P_n F \), while \( (P_n - P_n)F \to 0 \) and \( \limsup_n \|P_n\|_F \leq \limsup P_n F < \infty \). Moreover, \( E_W \max_{1 \leq i \leq n} |W_i| \to 0 \) by dominated convergence theorem because \( |W_i| \leq 1 \). Hence, the first term in the right hand side of (H.7)
converges to zero in probability for fixed \( n_0 \) as \( n \to \infty \). That is, for any \( \epsilon > 0 \) and \( n_0 < \infty \),

\[
P_n \left( \left( n_0 - 1 \right) \sum_{i=1}^{n} \| \left( W_{in} \right)_{F} \right)_{1} \leq \left( \delta_{x_i} \right)_{i=1}^{n} > \epsilon \right) \to 0 \text{ as } n \to \infty.
\]

Note that \( n\|W_{1n}\|_{2,1} \leq n(EW_{1n}) = 1 \) (see the proof of Theorem 3.6.16 in van der Vaart and Wellner (1996)). Finally, we need to show \( \max_{n_0 \leq k \leq n} E_{R} \left[ \left( \frac{1}{k} \sum_{i=n_0}^{k} Z_{Rn} \right)_{F} \right] \to 0 \). By triangle inequality,

\[
\max_{n_0 \leq k \leq n} E_{R} \left[ \frac{1}{k} \sum_{i=n_0}^{k} Z_{Rn} \right]_{F} \leq \max_{n_0 \leq k \leq n} \left\{ E_{R} \left[ \frac{1}{k} \sum_{i=1}^{k} Z_{Rn} \right]_{F} + E_{R} \left[ \frac{1}{k} \sum_{i=1}^{n_0-1} Z_{Rn} \right]_{F} \right\} \\
\leq \max_{n_0-1 \leq k \leq n} 2E_{R} \left[ \frac{1}{k} \sum_{i=1}^{k} Z_{Rn} \right]_{F} \\
= \max_{n_0-1 \leq k \leq n} 2 \left( \frac{1}{k} \sum_{i=1}^{k} Z_{in} \right)_{F}.
\]

The equality comes from \( R \) being independent of \( Z_{in} \). Note that \( \sup_{n \in \mathbb{N}} P_n \left[ \left( \frac{1}{k} \sum_{i=1}^{k} Z_{in} \right)_{F} > \epsilon \right] \to 0 \) as \( k \to \infty \) since \( \mathcal{F} \) is Glivenko-Cantelli uniformly in \( \{P_n\} \). Hence, the second term in the right hand side of (H.7) converges to zero in probability as \( n_0 \to \infty \). That is, for any \( \epsilon > 0 \),

\[
\sup_{n \geq n_0} P_{n} \left( n\|W_{1n}\|_{2,1} \max_{n_0 \leq k \leq n} E_{R} \left[ \left( \frac{1}{k} \sum_{i=n_0}^{k} Z_{Rn} \right)_{F} \right] > \epsilon \right) \to 0 \text{ as } n_0 \to \infty.
\]

Therefore, for any \( \epsilon > 0 \),

\[
P_{n} \left( E_{W} \left[ \sum_{i=1}^{n} W_{in} Z_{in} \right]_{F} > \epsilon \right) \to 0 \text{ as } n \to \infty.
\]

By applying the Markov inequality as follows, we can complete the proof:

\[
P_{n} \left( P_{W} \left( \left\| \sum_{i=1}^{n} W_{in} Z_{in} \right\|_{F} > \eta \right) > \eta \right) \leq P_{n} \left( E_{W} \left[ \sum_{i=1}^{n} W_{in} Z_{in} \right]_{F} > \eta \epsilon \right).
\]

\[\square\]

**Lemma H.14** (Lemma 3.6.6 van der Vaart and Wellner (1996)). For fixed elements \( x_1, ..., x_n \) of a set \( \mathcal{X} \), let \( \hat{X}_1, ..., \hat{X}_k \) be an i.i.d. sample from \( \mathbb{P}_n = n^{-1} \sum_{i=1}^{n} \delta_{x_i} \), where \( \delta_{x_i} \) is a dirac measure at \( x_i \). Then,

\[
E_{\hat{X}} \left[ \sum_{j=1}^{k} \left( \delta_{\hat{X}_j} - \mathbb{P}_n \right) \right]_{F} \leq 4E_{N,N'} \sum_{i=1}^{n} (N_i - N'_i) \delta_{x_i} \|_{F}
\]

for every class \( \mathcal{F} \) of functions \( f : \mathcal{X} \to \mathbb{R} \) and i.i.d. Poisson variables \( N_1, N'_1, ..., N_n, N'_n \) with
mean $\frac{1}{2}k/n$.

**Lemma H.15** (Lemma 2.3.6 van der Vaart and Wellner (1996)). Let $Z_1, ..., Z_n$ be independent stochastic processes with mean zero. Then,

$$E\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i Z_i \right\|_F \leq 2E\left\| \sum_{i=1}^{n} Z_i \right\|_F$$

for i.i.d. Rademacher random variables $\varepsilon_1, ..., \varepsilon_n$ and any functional class $\mathcal{F}$.

**Lemma H.16** (Lemma 2.9.1 van der Vaart and Wellner (1996)). Let $Z_1, ..., Z_n$ be i.i.d. stochastic processes with $E\left\| Z_i \right\|_F < \infty$ independent of the Rademacher variables $\varepsilon_1, ..., \varepsilon_n$. Then, for every i.i.d. sample $\xi_1, ..., \xi_n$ of mean-zero and symmetrically distributed random variables independent of $Z_1, ..., Z_n$ and $1 \leq n_0 \leq n$,

$$E\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\xi_i Z_i}{\sqrt{n}} \right\|_F \leq (n_0 - 1)E\left\| Z_i \right\|_F E\xi \max_{1 \leq i \leq n} \frac{\xi_i}{\sqrt{n}} + \frac{E\xi}{\sqrt{n}} \max_{n_0 \leq k \leq n} \frac{1}{\sqrt{k}} E\left\| \sum_{i=n_0}^{k} \frac{\varepsilon_i Z_i}{\sqrt{n}} \right\|_F,$$

where $\cdot \left\| \cdot \right\|_2$ is $L_{2,1}$ norm such that $\left\| \xi \right\|_2,1 = \int_0^\infty \sqrt{\mathbb{P}(|\xi| > x)} dx$ for a random variable $\xi$.

**Lemma H.17** (Lemma 3.6.7 van der Vaart and Wellner (1996)). For arbitrary stochastic processes $Z_1, ..., Z_n$, every exchangeable random vector $(\xi_1, ..., \xi_n)$ that is independent of $Z_1, ..., Z_n$, and any $1 \leq n_0 \leq n$,

$$E\xi \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i \right\|_F \leq 2(n_0 - 1)\frac{1}{n} \sum_{i=1}^{n} \left\| Z_i \right\|_F E\xi \max_{1 \leq i \leq n} \frac{\xi_i}{\sqrt{n}} + 2\left\| \xi \right\|_2,1 \max_{n_0 \leq k \leq n} \frac{1}{\sqrt{k}} E\xi E_R \frac{1}{\sqrt{k}} \sum_{i=n_0}^{k} Z_i \right\|_F,$$

where $(R_1, ..., R_n)$ is a random vector uniformly distributed on the set of all permutations of $\{1, ..., n\}$ and independent of $Z_1, ..., Z_n$. $\cdot \left\| \cdot \right\|_2,1$ is $L_{2,1}$ norm such that $\left\| \xi \right\|_2,1 = \int_0^\infty \sqrt{\mathbb{P}(|\xi| > x)} dx$ for a random variable $\xi$.

**Lemma H.18** (Lemma 3.6.15 van der Vaart and Wellner (1996)). For each $n$, let $(a_1, ..., a_n)$ and $(B_1, ..., B_n)$ be a vector of numbers and exchangeable random vector such that

$$\frac{1}{n} \sum_{i=1}^{n} (a_i - \bar{a})^2 \to \sigma^2, \quad \lim_{M \to \infty} \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} a_i^2 \{ |a_i| > M \} = 0,$$

$$\frac{1}{n} \sum_{i=1}^{n} (B_i - \bar{B})^2 \to \alpha^2, \quad \frac{1}{n} \sum_{1 \leq i \leq n} (B_i - \bar{B})^2 \to 0,$$

where $\bar{a} = \frac{1}{n} \sum_{i=1}^{n} a_i$ and $\bar{B} = \frac{1}{n} \sum_{i=1}^{n} B_i$. Then, $n^{-1/2} \sum_{i=1}^{n} (a_i B_i - \bar{a} B) \overset{d}{\to} N(0, \alpha^2 \sigma^2)$.

Let $B = (B_1, ..., B_n)$ be a multinomial vector with parameters $(1/n, ..., 1/n)$. Then, $\bar{B} = 1$, and conditions for $B$ in Lemma H.18 hold.