NEW CONTROL FUNCTION APPROACHES IN THRESHOLD REGRESSION WITH ENDOGENEITY

PING YU
University of Hong Kong

QIN LIAO
University of Hong Kong

PETER C. B. PHILLIPS
Yale University, University of Auckland, University of Southampton, and Singapore Management University

This paper studies control function (CF) approaches in endogenous threshold regression where the threshold variable is allowed to be endogenous. We first use a simple example to show that the structural threshold regression (STR) estimator of the threshold point in Kourtellos, Stengos and Tan (2016, *Econometric Theory* 32, 827–860) is inconsistent unless the endogeneity level of the threshold variable is low compared to the threshold effect. We correct the CF in the STR estimator to generate our first CF estimator using a method that extends the two-stage least squares procedure in Caner and Hansen (2004, *Econometric Theory* 20, 813–843). We develop our second CF estimator which can be treated as an extension of the classical CF approach in endogenous linear regression. Both these approaches embody threshold effect information in the conditional variance beyond that in the conditional mean. Given the threshold point estimates, we propose new estimates for the slope parameters. The first is a by-product of the CF approach, and the second type employs generalized method of moment (GMM) procedures based on two new sets of moment conditions. Simulation studies, in conjunction with the limit theory, show that our second CF estimator and confidence interval for the threshold point together with the associated second GMM estimator and confidence interval for the slope parameter dominate the other methods. We further apply the new estimation methodology to an empirical application from international trade to illustrate its usefulness in practice.

1065

1. INTRODUCTION

Endogenous threshold regression (ETR) has attracted much attention in recent econometric practice. The interest is explained partly by the recognition that economic relationships may shift over time, partly by the empirical relevance of thresholding in the design of government tax and welfare programs, and partly by the growing use of merit or need based awards in various social and educational programs. In all such cases, shifts in the associated regression equations may be driven by a companion variable which may itself be endogenous. For one example, in the empirical growth models used in Papageorgiou (2002) and Tan (2010), the corresponding threshold variables, the trade share to GDP in 1985 and the average expropriation risk from 1984 to 1997, are endogenous, as argued in Frankel and Romer (1999) and Acemoglu, Johnson and Robinson (2001), respectively. For another example, our empirical application later in the paper explores the effects of exports on real output growth, where the real exchange rate volatility, as the threshold variable, is endogenous as argued in Tenreyro (2007). For recent discussions of research on thresholding, including endogenously determined thresholds and overviews of various time series and cross-sectional applications (see Hansen, 2011; Yu and Phillips, 2018 [YP hereafter] and Yu, Liao and Phillips, 2018 [YLP hereafter]).

The usual modeling setup of the ETR has the following form:

$$
\begin{aligned}
y &= \left(\mathbf{x}'\beta_1 + \sigma_1 u\right) 1\left(q \leq \gamma\right) + \left(\mathbf{x}'\beta_2 + \sigma_2 u\right) 1\left(q > \gamma\right) \\
&=: \mathbf{x}'\beta_2 + \sigma_2 u + \left(\mathbf{x}'\delta_\beta + \delta_\sigma u\right) 1\left(q \leq \gamma\right) \text{ with } \mathbb{E}[u|\mathbf{x}] \neq 0,
\end{aligned}
\tag{1}
$$

where $\mathbf{x} = \left(x', q\right)' \in \mathbb{R}^d$, the parameter of interest is $\theta = \left(\gamma, \beta'\right)'$ with $\beta = \left(\beta_1', \beta_2'\right)'$ or equivalently, $\theta = \left(\gamma, \beta_2', \delta_\beta'\right)'$ with $\delta_\beta = \beta_1 - \beta_2$ being the threshold effect in conditional mean, and $\delta_\sigma = \sigma_1 - \sigma_2$ is the threshold effect in conditional variance. Quite often, the ETR literature assumes $\delta_\sigma = 0$ and absorbs σ_2 into u (see, e.g., Caner and Hansen, 2004 [CH hereafter]; Kourtellos, Stengos and Tan, 2016 [KST hereafter]), but allowing $\delta_\sigma \neq 0$ is often more realistic and plays an important role in this paper. The regressor x may contain some exogenous regressors such as an intercept, q may be a component of $\mathbf{x}$ (or some discussions in this paper can be simplified), and both x and q may be treated as endogenous. This setup is similar to endogenous linear regression (ELR) except that the regression coefficients and error variances depend on whether the threshold variable q crosses the threshold point γ.

As shown in YP, identification of γ does not require any instruments. Nevertheless, all the estimation procedures discussed in this paper involve instruments. In this introductory discussion, the data are assumed to be sampled as i.i.d. but weak dependence is permitted in the formal discussion later. When q is exogenous, CH use a two-stage least squares (2SLS) method to estimate γ. Suppose the first-stage regression is

$$
\mathbf{x} = \Pi'\mathbf{z} + \mathbf{v},
\tag{2}
$$

where the instruments $\mathbf{z}$ contain both included (in $\mathbf{x}$) exogenous regressors such as 1 and q, and excluded exogenous regressors, $\mathbb{E}[\mathbf{v}|\mathbf{z}] = \mathbf{0}$ and $\mathbb{E}[u|\mathbf{z}] = 0$. Then by plugging (2) in (1) and taking the conditional expectation of y given $\mathbf{z}$, we can show that

$$\mathbb{E}[y|\mathbf{z}] = (\Pi_0'\mathbf{z})' \beta_{10} 1 (q \leq \gamma_0) + (\Pi_0'\mathbf{z})' \beta_{20} 1 (q > \gamma_0) =: g_{CH}(\mathbf{z}; \theta_0), \tag{3}$$

where $\theta = (\theta', \Pi')'$, and the subscript 0 is used to emphasize the true value of a parameter. The CH estimator of γ minimizes the sample analog of

$$\mathbb{E}\left[(y - g_{CH}(\mathbf{z}; \theta))^2\right]$$

with $g_{CH}(\mathbf{z}; \theta)$ defined in (3).

When q is endogenous, KST call (1) a structural threshold regression (STR) model and use a type of control function (CF) approach to estimate γ (see Navarro, 2008, for an introduction and Wooldridge, 2015, for a recent summary of CF approaches in regular models). Their CF approach can be treated as an extension of CH's 2SLS method. Specifically, they derive under some assumptions that

$$\mathbb{E}[y|\mathbf{z}, q \leq \gamma_0] = g_1(\mathbf{z}; \gamma_0, \beta_{10}, \kappa_0) \text{ and } \mathbb{E}[y|\mathbf{z}, q > \gamma_0] = g_2(\mathbf{z}; \gamma_0, \beta_{20}, \kappa_0), \tag{4}$$

where $\mathbf{z}$ contains the instruments which do not include q, κ_0 is a nuisance parameter indicating the endogeneity level of q, and g_1 and g_2 are some parametric functions of $\mathbf{z}$. Then KST's STR estimator of γ, say $\widehat{\gamma}$, minimizes the sample analog of

$$\mathbb{E}\left[(y - g_1(\mathbf{z}; \gamma, \beta_1, \kappa) 1 (q \leq \gamma) - g_2(\mathbf{z}; \gamma, \beta_2, \kappa) 1 (q > \gamma))^2\right],$$

for which a detailed construction is given in Section 2.1. KST develop their asymptotic theory of $\widehat{\gamma}$ in the small-threshold-effect framework as in CH, especially, in which both δ_β and κ are $O(n^{-\alpha})$ with $\alpha \in (0, 1/2)$. We demonstrate with a simple example in Section 2.2 that $\widehat{\gamma}$ is not generically consistent unless κ is relatively small compared to δ_β and that there does not exist a nonzero κ value such that $\widehat{\gamma}$ is consistent for all possible δ_β values. We further point out that an extension of KST using copula theory by Christopoulos, McAdam and Tzavalis (2021) (CMT hereafter) suffers a similar problem. Moreover, if $\widehat{\gamma}$ is not consistent, the corresponding β estimator, $\widehat{\beta}$, is also inconsistent, thereby leading to failure in the whole estimation procedure. Section 2.3 shows that the key problem in the KST estimator is that when the objective function takes the form of a sum of squares, the conditioning set in (4) cannot depend on the unknown parameter γ_0.

As an alternative methodology, Section 3 puts forward two CF approaches for γ, labeled CF-I and CF-II. These two CF approaches differ from KST and do not require the degree of endogeneity to shrink to zero asymptotically. Method CF-I corrects the CF in KST so that it can be treated as an extension of CH and thereby includes CH as a special case. CF-II extends the classical CF approach in ELR to the current context. The approach is new in the ETR literature and covers the standard CH model (i.e., with q exogenous) as a special case. The difference

between these two approaches lies in using $\mathbb{E}[y|\mathbf{z}, q]$ or $\mathbb{E}[y|\mathbf{z}, \mathbf{x}]$ to identify γ. Specifically, we first derive

$$\mathbb{E}[y|\mathbf{z}, q] = g_I(\mathbf{z}, q; \theta_0) \text{ in CF-I and } \mathbb{E}[y|\mathbf{z}, \mathbf{x}] = g_{II}(\mathbf{z}, \mathbf{x}; \theta_0) \text{ in CF-II}$$

under some auxiliary assumptions (which are standard in the usual CF approach), noting that CH also derive $\mathbb{E}[y|\mathbf{z}, q]$ since $q \in \mathbf{z}$, so CF-I is indeed an extension of CH. Then we estimate γ by minimizing in generic form the sample analogs of

$$\mathbb{E}\left[(y - g_I(\mathbf{z}, q; \theta))^2\right] \text{ and } \mathbb{E}\left[(y - g_{II}(\mathbf{z}, \mathbf{x}; \theta))^2\right],$$

for "*I*" and "*II*," respectively, where we use θ to collect all unknown parameters which need not be the same in CF-I and CF-II. Note that the conditioning sets in $\mathbb{E}[y|\mathbf{z}, q]$ and $\mathbb{E}[y|\mathbf{z}, \mathbf{x}]$ do not depend on any unknown parameter, which explains why the new estimators are consistent and the KST estimator is not.

Interestingly, although both CF-I and CF-II use the conditional mean of y to identify γ, these approaches also utilize some threshold effect information in the conditional variance (i.e., $\delta_\sigma \neq 0$). This methodology differs from the usual least squares (LS) estimator (without endogeneity) and CH's 2SLS estimator (with endogenous x but exogenous q), where only the threshold effect information in the conditional mean can be explored even if there is a threshold effect in conditional variance. We derive the asymptotic distribution of these two CF estimators of γ in the small-threshold-effect framework, and also suggest inverting the likelihood-ratio (LR) statistics to construct confidence intervals (CIs) for γ. Unlike the Heckit model,

$$y = \begin{cases} y^*, & \text{if } \mathbf{z}'\gamma + v > 0 \text{ or } d = 1, \\ 0, & \text{if } \mathbf{z}'\gamma + v \leq 0 \text{ or } d = 0, \end{cases} \tag{5}$$
$$\text{with } y^* = \mathbf{x}'\beta + u, \ (u, v) \sim \mathcal{N}(0, (\sigma^2, \rho\sigma; \rho\sigma, 1)) \text{ and } \rho \neq 0,$$

where a CF is added to the LS objective function based on the observed data, the extra randomness from the generated regressors in our objective functions does not affect the asymptotic distribution of the CF estimators of γ, say, $\widehat{\gamma}$.

Recently, Kourtellos, Stengos and Sun (2022) extended $g_I(\mathbf{z}, q; \theta)$ in CF-I to a semiparametric setup where the CFs take nonparametric forms, whereas the other components of $g_I(\mathbf{z}, q; \theta)$ retain parametric forms; their approach can also be applied to CF-II. Section 3.2 provides detailed comparisons between the parametric setup of this paper and their semiparametric setup. We emphasize here that CF-II is the preferable approach because it has a firmer theoretical foundation as discussed in Section 3 and is found to perform much better than CF-I in our simulations of Section 5. CF-I is included in our analysis only to show the correct objective function for the KST framework.

All the estimators of γ discussed above are *M*-estimators, and another type of estimator with instruments employed are generalized method of moment (GMM) estimators (or *Z*-estimators). YLP propose the 2SLS estimator in the cross-sectional data case and Seo and Shin (2016) (SS hereafter) propose the FD-GMM in the dynamic panel data case (Appendix A of SS also extends CH's estimator

above to the dynamic panel data case). More specifically, such GMM estimators use the moment conditions like

$$\mathbb{E}\left[\mathbf{z}\left(y - \mathbf{x}'\beta_2 - \mathbf{x}'\delta_\beta 1\left(q \le \gamma\right)\right)\right] = \mathbf{0}$$

to identify γ. Although GMM estimators are important and natural to handle endogeneity, compared with M-estimators, they suffer from at least three drawbacks as discussed in YLP. First, GMM changes the nature of γ from a threshold point (which is nonregular) to a quantile of q (which is regular), which implies the convergence rate of $\widehat{\gamma}$ is $n^{1/2}$, much slower than the convergence rate n of M-estimators. Second, γ is not always identifiable by GMM, for example, when q is independent of the rest of the system such as the time index in structural change models, γ cannot be identified by GMM. Third, GMM requires more instruments than our CF estimators for identification, which implies that GMM may have less applicability since good instruments are hard to find in practice. Our simulations show that the performance of GMM estimators is much worse than our CF-II estimator. Of course, GMM estimators require only unconditional moments to identify γ, whereas M-estimators require conditional moments; however, for identification in nonlinear models conditional moments are routinely assumed in the literature. Also, it should be emphasized that these drawbacks of GMM are not due to the use of unconditional moments; as argued in YLP, even if conditional moments (which are equivalent to uncountably many unconditional moments) are employed, the first two drawbacks of GMM do not disappear. In summary, GMM has several less desirable features for estimate of γ in ETR.

Given $\widehat{\gamma}$, we propose two estimation procedures for β in Section 4. The first type estimates β by $\widehat{\beta} = \left(\widehat{\beta}_1\left(\widehat{\gamma}\right)', \widehat{\beta}_2\left(\widehat{\gamma}\right)'\right)'$ and is therefore a by-product of the CF approaches for γ, where $\widehat{\beta}\left(\gamma\right) = \left(\widehat{\beta}_1\left(\gamma\right)', \widehat{\beta}_2\left(\gamma\right)'\right)'$ is the concentrated version of the estimator of β given γ based on the objective functions employed in our CF approaches. This type of estimator of β is not studied in CH or KST. Recall that in the Heckit model the generated regressor affects the asymptotic variance of the estimates of regular parameters such as β. A similar phenomenon occurs here. Interestingly, in CF-I, the extra randomness in the generated regressor and the randomness present in the original error term are correlated, whereas in CF-II they are not. The second type of estimator relies on GMM estimation and is based on two new sets of moment conditions. The first GMM estimator requires more moments than the second and can be treated as an extension of CH's GMM estimator to the endogenous q case for which their GMM estimator is inconsistent. The second GMM estimator is an extension of the first type of estimator (i.e., the CF estimator of β). For all β estimators, the γ estimates do not affect their asymptotic variances even though $\widehat{\gamma}$ may not be $o_p\left(n^{-1/2}\right)$ convergent.

Section 5 reports simulation comparisons of the finite sample performance of all existing estimators and CIs of θ when instruments are available. It turns out that the CF-II estimator and CI for γ and the associated second GMM estimator and CI for β dominate the other methods. Section 6 applies the new methodology of this paper to an empirical application from international trade to illustrate its

usefulness in practice. Section 7 concludes. Proofs of the main results (theorems, propositions, and lemmas) are given in Appendixes A–C. Additional assumptions, lemmas, and theorems as well as further details concerning the various procedures in this paper and the properties of existing procedures in the literature are given in the Supplementary Material which contains five sections labeled SD.1–SD.5.

A word on notation: Throughout this paper, we use the subscript 0 to designate the true value of a parameter when it is not obvious and a hat to indicate its estimate. The subscripts "$\leq \gamma$" and "$> \gamma$" signify use of the indicator functions $1(q_i \leq \gamma)$ and $1(q_i > \gamma)$, so that for instance $\mathbf{x}_{i, \leq \gamma} = \mathbf{x}_i 1(q_i \leq \gamma)$. For three random vectors x, y, z, $x \perp y$ means x is independent of y, and $x \perp y|z$ means x is independent of y conditional on z. For a matrix A, $A > 0$ means A is positive definite. I_m is the $m \times m$ identity matrix. $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$ is the inverse Mills ratio, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf. Throughout the paper, the symbol ℓ is used to indicate the presence of two regimes in (1) and is not written out explicitly as "$\ell = 1, 2$." The slope parameter is sometimes partitioned according to x and q, as well as regime, giving the component representation $\beta_\ell = \left(\beta'_{\ell x}, \beta_{\ell q}\right)'$.

A word on assumptions: Because CF-II is our preferable method, we list only the assumptions for consistency and asymptotic distributions of CF-II in the main paper, and relegate the counterparts for CF-I to Section SD.2 of the Supplementary Material.

2. INCONSISTENCY OF KST'S STR ESTIMATOR

This section begins with a review of the STR estimator in KST, followed by a simple example to show its inconsistency with an accompanying analysis and discussion. To simplify the development, we assume the data are randomly sampled in this section.

2.1. Review of KST's STR Estimator

KST assume in the first-stage regression,

$$
\begin{aligned}
x &= \Pi'_x \mathbf{z} + v_x, \\
q &= \pi' \mathbf{z} + v_q,
\end{aligned}
\tag{6}
$$

where $\mathbb{E}[v_x|\mathbf{z}] = \mathbf{0}$, $\mathbb{E}\left[v_q|\mathbf{z}\right] = 0$, and $\left(u, v_q\right)|\mathbf{z} \sim \mathcal{N}(0, (\sigma^2, \rho\sigma; \rho\sigma, 1))$ with $\rho \neq 0$ due to the endogeneity of q.[1] KST also impose the following key assumption to simplify their discussion.

Assumption K: $\mathbf{x} = x$, and $v_x \perp 1(q \leq \gamma_0)|\mathbf{z}$.

---

[1] In this formulation, it is only required that $E\left[u|v_q\right]$ be linear in v_q, but $v_q \sim \mathcal{N}(0, 1)$ is necessary. Also, $q = \pi'\mathbf{z} + v_q$ with $v_q \perp \mathbf{z}$ and $v_q \sim \mathcal{N}(0, 1)$ implies that q follows a (mean) mixture normal distribution with the density $\int \phi\left(\pi'\mathbf{z}, 1\right) dF_\mathbf{z}(\mathbf{z})$, where $\phi\left(\mu, \sigma^2\right)$ is the normal density with mean μ and variance σ^2, and $F_\mathbf{z}(\mathbf{z})$ is the cdf of $\mathbf{z}$.

Assumption K excludes q as a regressor; otherwise, the analysis is more complicated as detailed below. Under Assumption K,

$$\mathbb{E}[y|\mathbf{z}, q \leq \gamma_0]$$
$$= \mathbb{E}\left[\left(\beta'_{10}g_x + \beta'_{10}v_x\right) 1(q \leq \gamma_0) + \left(\beta'_{20}g_x + \beta'_{20}v_x\right) 1(q > \gamma_0) + u|\mathbf{z}, q \leq \gamma_0\right]$$
$$= \beta'_{10}g_x - \kappa_0\lambda\left(\gamma_0 - \pi'_0\mathbf{z}\right),$$

where $g_x = \Pi'_{x0}\mathbf{z}$, $\mathbb{E}[v_x 1(q \leq \gamma_0)|\mathbf{z}, q \leq \gamma_0] = \mathbb{E}[v_x|\mathbf{z}, q \leq \gamma_0] 1(q \leq \gamma_0) = \mathbb{E}[v_x|\mathbf{z}] 1(q \leq \gamma_0) = \mathbf{0}$, $\mathbb{E}[u|\mathbf{z}, q \leq \gamma_0] = \mathbb{E}\left[u|\mathbf{z}, v_q \leq \gamma_0 - \pi'_0\mathbf{z}\right] = -\kappa_0\lambda\left(\gamma_0 - \pi'_0\mathbf{z}\right)$, and $\kappa_0 = \rho_0\sigma_0$. Similarly,

$$\mathbb{E}[y|\mathbf{z}, q > \gamma_0] = \beta'_{20}g_x + \kappa_0 \cdot \lambda\left(\pi'_0\mathbf{z} - \gamma_0\right).$$

Note here that the conditioning events are $q \leq \gamma_0$ and $q > \gamma_0$ rather than $q \leq \gamma$ and $q > \gamma$ for an arbitrary γ. As shown at the end of Section SD.1 of the Supplementary Material, the formulas above cannot be extended to other γ values. When $\mathbf{x} = (x', q)'$,

$$\mathbb{E}[y|\mathbf{z}, q \leq \gamma_0] = \left(g'_x, g_q\right)\beta_{10} - \kappa_{10} \cdot \lambda\left(\gamma_0 - \pi'_0\mathbf{z}\right),$$
$$\mathbb{E}[y|\mathbf{z}, q > \gamma_0] = \left(g'_x, g_q\right)\beta_{20} + \kappa_{20} \cdot \lambda\left(\pi'_0\mathbf{z} - \gamma_0\right),$$
$$\tag{7}$$

where $g_q = \pi'_0\mathbf{z}$, and $\kappa_{\ell 0} = \rho_0\sigma_0 + \beta_{\ell q0}$. To be consistent with KST, we maintain Assumption K below. Now,

$$y = \left[\beta'_{10}g_x + \kappa_0 \cdot \lambda_1\left(\gamma_0 - \pi'_0\mathbf{z}\right)\right] 1\left(q \leq \gamma_0\right)$$
$$+ \left[\beta'_{20}g_x + \kappa_0 \cdot \lambda_2\left(\gamma_0 - \pi'_0\mathbf{z}\right)\right] 1\left(q > \gamma_0\right) + e^*,$$
$$\tag{8}$$

where we follow the KST notation and define $\lambda_1(\cdot) = -\lambda(\cdot)$ and $\lambda_2(\cdot) = \lambda(-\cdot)$, and

$$e^* = u - \kappa_0 \cdot \lambda_1\left(\gamma_0 - \pi'_0\mathbf{z}\right) 1\left(q \leq \gamma_0\right) - \kappa_0 \cdot \lambda_2\left(\gamma_0 - \pi'_0\mathbf{z}\right) 1\left(q > \gamma_0\right).$$
$$\tag{9}$$

In other words, the model becomes a nonlinear threshold regression, and γ can be estimated by minimizing

$$S_n(\theta, \kappa) = \frac{1}{n}\sum_{i=1}^{n}\left\{\left[y_i - \beta'_1\widehat{g}_{xi} - \kappa \cdot \lambda_1\left(\gamma - \widehat{g}_{qi}\right)\right]^2 1(q_i \leq \gamma)\right.$$
$$\left. + \left[y_i - \beta'_2\widehat{g}_{xi} - \kappa \cdot \lambda_2\left(\gamma - \widehat{g}_{qi}\right)\right]^2 1(q_i > \gamma)\right\},$$
$$\tag{10}$$

where $\widehat{g}_{xi} = \widehat{\Pi}'_x\mathbf{z}_i$ and $\widehat{g}_{qi} = \widehat{\pi}'\mathbf{z}_i$ with $\widehat{\Pi}_x$ and $\widehat{\pi}$ obtained from a first-stage regression. As usual, we can concentrate on parameter γ, and regress y_i on $\widehat{g}_{xi}1(q_i \leq \gamma)$, $\widehat{g}_{xi}1(q_i > \gamma)$ and $\widehat{\Lambda}_i(\gamma) := \lambda_1\left(\gamma - \widehat{g}_{qi}\right) 1(q_i \leq \gamma) + \lambda_2\left(\gamma - \widehat{g}_{qi}\right) 1(q_i > \gamma)$ to obtain $\widehat{\beta}_1(\gamma), \widehat{\beta}_2(\gamma)$ and $\widehat{\kappa}(\gamma)$. Then γ can be estimated by the extremum problem

$$\widehat{\gamma} = \underset{\gamma \in \Gamma}{\mathrm{argmin}}\, S_n(\gamma),$$

where $\Gamma = [\underline{\gamma}, \overline{\gamma}]$ is a proper subset of the support of q, and

$$S_n(\gamma) = S_n\left(\gamma, \widehat{\beta}_1(\gamma), \widehat{\beta}_2(\gamma), \widehat{\kappa}(\gamma)\right).$$

Given $\widehat{\gamma}$, the parameter β can be estimated by 2SLS/GMMs as in CH. In the small-threshold-effect framework of Hansen (2000), KST show that $\widehat{\gamma}$ is $n^{1-2\alpha}$-consistent and its asymptotic distribution is based on a functional of two-sided Brownian motion, under the assumption that both δ_β and κ are $O(n^{-\alpha})$ with $\alpha \in (0, 1/2)$.

The terms $\kappa \cdot \lambda_1(\gamma - \widehat{g}_{qi})$ and $\kappa \cdot \lambda_2(\gamma - \widehat{g}_{qi})$ in the criterion (10) are CF, as in the Heckit model. Unfortunately in the present case, they do not take the correct form as is illustrated in the simple example in the next subsection. When q is exogenous or $\kappa_0 = 0$, KST's STR estimator reduces to the 2SLS estimator of CH which proves that $\widehat{\gamma}$ is consistent.

2.2. Inconsistency of STR Estimator $\widehat{\gamma}$

We consider the case where q is endogenous and to highlight the reason why the STR estimator is inconsistent, we consider the simple form of model (1) in which

$$\begin{aligned}
y_i &= \delta_0 1\left(q_i \leq \gamma_0\right) + u_i, \\
q_i &= z_i \pi_0 + v_{qi},
\end{aligned} \qquad (11)$$

with $u_i = \kappa_0 v_{qi} + e_{ui}$, $\left(v_{qi}, e_{ui}\right)' \sim \mathcal{N}(0, I_2)$, $z \sim \mathcal{N}(0, 1)$ independent of $\left(v_q, e_u\right)$,[2] and $\gamma_0 = 0$. In other words, in model (1), $\mathbf{x} = 1$, $\beta_{20} = 0$ is known, $\delta_\beta = \delta_0$, $\delta_\sigma = 0$, and there is only one instrument. For the analysis of $\widehat{\gamma}$ consistency it is immaterial whether π_0 is known or is consistently estimated. So here, we simply assume that $\pi_0 = -1$ is known.

The extremum criterion (10) is now

$$S_n(\gamma, \delta, \kappa) = \frac{1}{n} \sum_{i=1}^{n} \left[y_i - \delta 1\left(q_i \leq \gamma\right) - \kappa \lambda_{1i}^\gamma 1\left(q_i \leq \gamma\right) - \kappa \lambda_{2i}^\gamma 1\left(q_i > \gamma\right)\right]^2, \qquad (12)$$

where $\lambda_{1i}^\gamma = \lambda_1(\gamma + z_i)$ and $\lambda_{2i}^\gamma = \lambda_2(\gamma + z_i)$. Assume that the unknown parameters $(\gamma, \delta, \kappa)'$ lie in a compact set with their true value in the interior. By a Glivenko–Cantelli theorem, $S_n(\gamma, \delta, \kappa)$ converges uniformly to

$$S(\gamma, \delta, \kappa) = \mathbb{E}\left[\left(y_i - \delta 1\left(q_i \leq \gamma\right) - \kappa \lambda_{1i}^\gamma 1\left(q_i \leq \gamma\right) - \kappa \lambda_{2i}^\gamma 1\left(q_i > \gamma\right)\right)\right]^2.$$

Define centered versions of S_n and S as

$$Q_n(\gamma, \delta, \kappa) = S_n(\gamma, \delta, \kappa) - S_n(\gamma_0, \delta_0, \kappa_0) \qquad (13)$$

---

[2]As emphasized in Yu (2013), although in ELR, we can linearly project $\mathbf{x}_i$ on $\mathbf{z}_i$ in the first step even if $E[\mathbf{x}_i|\mathbf{z}_i]$ is nonlinear, in CH's framework of ETR, we must nonetheless regress $\mathbf{x}_i$ on $\mathbf{z}_i$ (i.e., obtain $E[\mathbf{x}_i|\mathbf{z}_i]$) in the first step to consistently estimate γ. Accordingly, it is critical to assume $E\left[v_q|z\right] = 0$ rather than $E\left[zv_q\right] = 0$; making z independent of v_q is a strengthening of $E\left[v_q|z\right] = 0$.

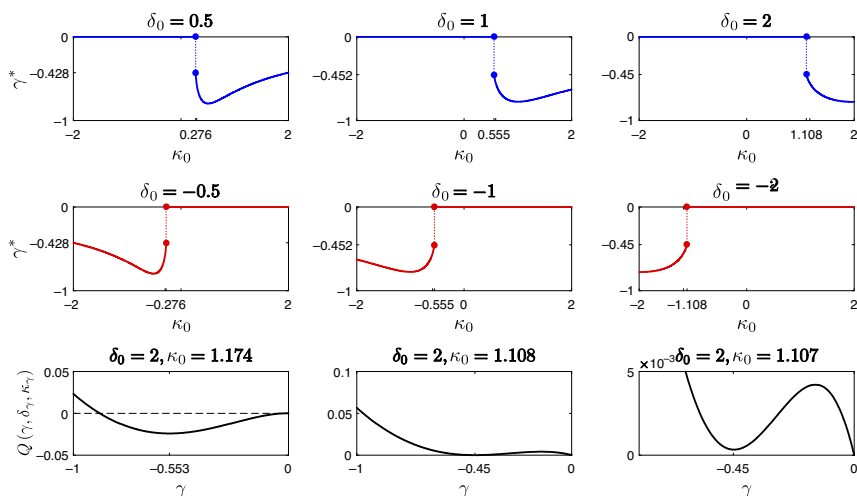**FIGURE 1.** γ^* As a function of $\kappa_0 \in [-2, 2]$ for $\delta_0 = \pm 0.5, \pm 1, \pm 2$

and

$$Q(\gamma, \delta, \kappa) = S(\gamma, \delta, \kappa) - S(\gamma_0, \delta_0, \kappa_0),$$

where recentering S_n and S ensures that $Q_n(\gamma_0, \delta_0, \kappa_0) = 0 = Q(\gamma_0, \delta_0, \kappa_0)$. Then by Theorem 2.1 of Newey and McFadden (1994), we need only show that $\underset{\gamma \in \Gamma}{\arg\min} \underset{\delta, \kappa}{\min} Q(\gamma, \delta, \kappa) = \gamma_0$.

To simplify notation, we denote the pseudo-true value of γ by γ^* so that $\arg\min_\gamma Q(\gamma, \delta_\gamma, \kappa_\gamma) = \gamma^*$, where $(\delta_\gamma, \kappa_\gamma)$ is the concentrated value of (δ, κ) given γ. The concentrated objective function $Q(\gamma, \delta_\gamma, \kappa_\gamma)$ is symmetric about γ_0, so we examine only $\gamma \in [\underline{\gamma}, \gamma_0]$, that is, we let the parameter space be $\Gamma = [\underline{\gamma}, \gamma_0]$. In Section SD.1 of the Supplementary Material, we derive the formula of $Q(\underline{\gamma}, \delta_\gamma, \kappa_\gamma)$ and show that as long as $\kappa_0/\delta_0 > 0.587$, irrespective of whether δ_0 is positive or negative, $\gamma^* < 0 = \gamma_0$; we also discuss the case where κ_0 and δ_0 are known in Section SD.1 of the Supplementary Material.

Figure 1 shows γ^* as a function of $\kappa_0 \in [-2, 2]$ for $\delta_0 = \pm 0.5, \pm 1, \pm 2$. First of all, $\gamma^*(\delta_0, \kappa_0) = \gamma^*(-\delta_0, -\kappa_0)$ because $Q(\gamma, \delta_\gamma, \kappa_\gamma)$ remains the same when (δ_0, κ_0) changes to $(-\delta_0, -\kappa_0)$. Obviously, $\widehat{\gamma}$ is consistent only if $\kappa_0/\delta_0 \leq 0.587$. As a result, there does not exist a κ_0 such that $\widehat{\gamma}$ is consistent for all $\delta_0 \neq 0$, or the κ_0 that makes $\widehat{\gamma}$ consistent uniformly in $\delta_0 \neq 0$ is 0. In other words, only if q is exogenous is the KST estimator consistent for any $\delta_0 \neq 0$. Even when $\kappa_0/\delta_0 < 0.587$, $\widehat{\gamma}$ may still be inconsistent. In the third row of Figure 1, we show $Q(\gamma, \delta_\gamma, \kappa_\gamma)$ as a function of γ when $\delta_0 = 2$, and $\kappa_0 = 1.174, 1.108$, and 1.107. Indeed, when $\kappa_0 = 1.174 = 0.587\delta_0$, $dQ(\gamma, \delta_\gamma, \kappa_\gamma)/d\gamma|_{\gamma=\gamma_0} = 0$, but $\arg\min_\gamma Q(\gamma, \delta_\gamma, \kappa_\gamma) = -0.553 < 0$. The case where $\kappa_0 = 1.108$ is marginal and $Q(\gamma, \delta_\gamma, \kappa_\gamma)|_{\gamma=-0.45} =$

$Q(\gamma, \delta_\gamma, \kappa_\gamma)|_{\gamma=0}$. When $\kappa_0 = 1.107$, $Q(\gamma, \delta_\gamma, \kappa_\gamma)|_{\gamma=-0.45} > Q(\gamma, \delta_\gamma, \kappa_\gamma)|_{\gamma=0} = 0$ and then $\operatorname{argmin}_\gamma Q(\gamma, \delta_\gamma, \kappa_\gamma) = 0$, as is evident in the final panel of Figure 1.

The arguments above provide a possible explanation for why the simulations in KST work well. The setup above is a special case of their data generating processes (DGPs). Specifically, in our setup, their $x_{2i} = 1$, x_{1i} is missing and $q_i = -z_i + v_{qi}$ (rather than $2 + z_i + v_{qi}$). In our notation, they set $\delta_0 = 1, 2, 3$ and $\kappa_0 = 0.053, 0.316, 0.634$. Obviously, their κ_0's are all relatively small compared to δ_0; this parameter setting is the reason why their simulation based estimates of $\widehat{\gamma}$ appear unbiased.

In Section SD.1 of the Supplementary Material, we point out the key problems in KST's consistency proof using the framework of our simple example. Basically, they miss two terms in their calculation of the probability limit of the objective function. When κ is small, these two terms are dominated, so the minimizer of the limit objective function is still γ_0. But when κ is large, these two terms are not neglectable, so the minimizer is different from γ_0. We also show that we cannot modify KST's objective function (pursuing their approach) to obtain a consistent estimator of γ.

CMT suffer a similar problem as KST, where they replace $\lambda_{\ell i}^\gamma$ by some functions derived from copula theory. The problem here is similar to that of KST, that is, they calculate $\mathbb{E}[y|\mathbf{z}, q \leq \gamma]$ and $\mathbb{E}[y|\mathbf{z}, q > \gamma]$ rather than $\mathbb{E}[y|\mathbf{z}, q]$. The good performances of CMT's estimators in their simulations (even in the misspecified scenarios) are also attributed to the large threshold effects compared to the endogeneity levels in their DGPs.

2.3. Discussion

As emphasized in Section 3.3 of Yu (2015), conditional moment restrictions rather than orthogonality conditions must be used to consistently estimate γ. Given that KST's estimation is indeed based on the conditional mean $\mathbb{E}[y|z, q \leq \gamma_0]$ and $\mathbb{E}[y|z, q > \gamma_0]$, a natural question is why KST's estimator is inconsistent. The key point here is that when the objective function takes the form of a sum of squares (like in KST), the conditioning set in the conditional mean on which the objective function is based cannot depend on the unknown parameter to achieve a consistent estimator. This phenomenon emerges regardless of whether the parameter is nonregular (like γ) or not (like a mean). We here use a simple example to illustrate this point.

Suppose y follows the exponential distribution, and we want to estimate its mean $\theta_0 = 1$. It is easy to see that for any $x \in (0, \infty)$,

$$
\begin{aligned}
\mathbb{E}[y|y \leq x] &= \int_0^x y \frac{e^{-y/\theta_0}}{\theta_0(1 - e^{-x/\theta_0})} dy = \theta_0 - x \frac{e^{-x/\theta_0}}{1 - e^{-x/\theta_0}}, \\
\mathbb{E}[y|y > x] &= \int_x^\infty y \frac{e^{-y/\theta_0}}{\theta_0 e^{-x/\theta_0}} dy = \theta_0 + x,
\end{aligned}
\tag{14}
$$
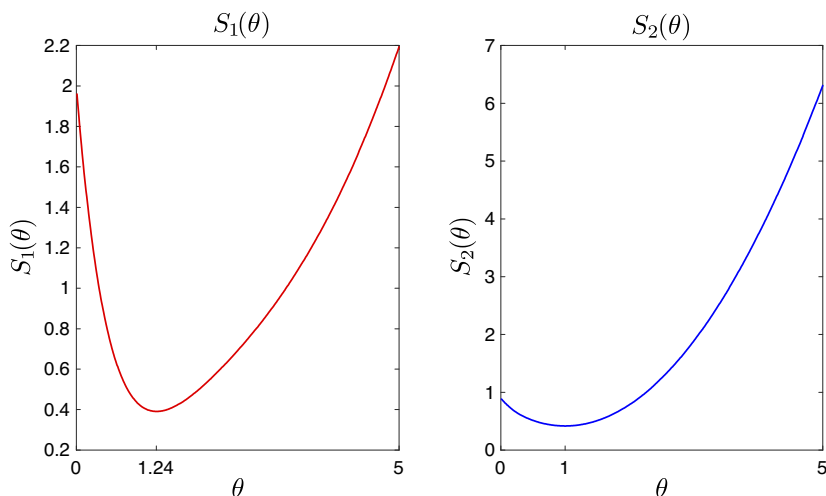
FIGURE 2. $S_1(\theta)$ and $S_2(\theta)$

which certainly hold when $x = \theta_0$, so it is natural to estimate θ by

$$S_{1n}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \theta + \theta \frac{e^{-1}}{1 - e^{-1}} \right)^2 1(y_i \leq \theta) + \frac{1}{n} \sum_{i=1}^{n} (y_i - \theta - \theta)^2 1(y_i > \theta).$$

The probability limit of $S_{1n}(\theta)$ is

$$S_1(\theta) = \mathbb{E}\left[\left(y - \frac{1 - 2e^{-1}}{1 - e^{-1}}\theta \right)^2 1(y \leq \theta) \right] + \mathbb{E}\left[(y - 2\theta)^2 1(y > \theta) \right]$$

with y following the standard exponential distribution. As shown in Figure 2, $S_1(\theta)$ is minimized at around 1.24, different from θ_0.[3] On the other hand, if the conditioning sets do not depend on θ, say, $\{y \leq 1\}$ and $\{y > 1\}$, then

$$\mathbb{E}[y|y \leq 1] = \theta_0 - \frac{e^{-1/\theta_0}}{1 - e^{-1/\theta_0}} \text{ and } \mathbb{E}[y|y > 1] = 1 + \theta_0, \tag{15}$$

and we can set the objective function as

$$S_{2n}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \theta + \frac{e^{-1/\theta}}{1 - e^{-1/\theta}} \right)^2 1(y_i \leq 1) + \frac{1}{n} \sum_{i=1}^{n} (y_i - 1 - \theta)^2 1(y_i > 1).$$

---

[3] Of course, the above estimation approach need not always generate an inconsistent estimator. For example, if we replace the standard exponential distribution by the standard uniform distribution and use $E[y|y \leq \theta] = \frac{\theta}{2}$ and $E[y|y > \theta] = \frac{1+\theta}{2}$, $\theta \in (0,1)$, to identify $\theta_0 = \frac{1}{2}$, it indeed works. Actually, as shown in Section 2.2, when the endogeneity level is relatively low compared to the threshold effect, KST's estimator also converges to the true γ_0. However, because the true DGP is unknown, we cannot use these estimators without discretion or additional information.

From Figure 2, the probability limit of $S_{2n}(\theta)$,

$$S_2(\theta) = \mathbb{E}\left[\left(y - \theta + \frac{e^{-1/\theta}}{1 - e^{-1/\theta}}\right)^2 1(y \leq 1)\right] + \mathbb{E}\left[(y - 1 - \theta)^2 1(y > 1)\right]$$

$$=: S_{21}(\theta) + S_{22}(\theta),$$

is indeed minimized at θ_0 since both $S_{21}(\theta)$ and $S_{22}(\theta)$ are minimized at θ_0.

KST try to calculate $\mathbb{E}[y|\mathbf{z}, q \leq \gamma_0]$ and $\mathbb{E}[y|\mathbf{z}, q > \gamma_0]$ whose conditioning sets depend on γ_0; this is why their estimator of γ may not be consistent. If we make the conditioning set independent of γ_0, for example, replacing $q \leq \gamma_0$ by $q \in \Gamma$ or $q \in \mathbb{R}$, then we can indeed get a consistent estimator. In KST's setup, suppose $x \in \mathbf{z}$ and only q is endogenous; then

$$\mathbb{E}[y|\mathbf{z}, q \in \Gamma] = \beta_{10}'xP\left(q \leq \gamma_0|\mathbf{z}, \underline{\gamma} \leq q \leq \overline{\gamma}\right) + \beta_{20}'xP\left(q > \gamma_0|\mathbf{z}, \underline{\gamma} \leq q \leq \overline{\gamma}\right)$$

$$+ \kappa_0\mathbb{E}\left[v_q|\mathbf{z}, \underline{\gamma} \leq q \leq \overline{\gamma}\right]$$

$$= \beta_{10}'x\frac{\Phi\left(\gamma_0 - \pi_0'\mathbf{z}\right) - \Phi\left(\underline{\gamma} - \pi_0'\mathbf{z}\right)}{\Phi\left(\overline{\gamma} - \pi_0'\mathbf{z}\right) - \Phi\left(\underline{\gamma} - \pi_0'\mathbf{z}\right)} + \beta_{20}'x\frac{\Phi\left(\overline{\gamma} - \pi_0'\mathbf{z}\right) - \Phi\left(\gamma_0 - \pi_0'\mathbf{z}\right)}{\Phi\left(\overline{\gamma} - \pi_0'\mathbf{z}\right) - \Phi\left(\underline{\gamma} - \pi_0'\mathbf{z}\right)}$$

$$+ \kappa_0\frac{\phi\left(\underline{\gamma} - \pi_0'\mathbf{z}\right) - \phi\left(\overline{\gamma} - \pi_0'\mathbf{z}\right)}{\Phi\left(\overline{\gamma} - \pi_0'\mathbf{z}\right) - \Phi\left(\underline{\gamma} - \pi_0'\mathbf{z}\right)},$$

where the last equality is from $q = \pi'\mathbf{z} + v_q$ and $v_q \sim \mathcal{N}(0, 1)$. The resulting least square estimator of γ is $\sqrt{n}$ consistent (rather than n consistent) because the objective function is smooth in γ. Letting $\underline{\gamma} = -\infty$ and $\overline{\gamma} = \infty$, we get

$$\mathbb{E}[y|\mathbf{z}] = \mathbb{E}[y|\mathbf{z}, q \in \mathbb{R}] = \beta_{10}'x\Phi\left(\gamma_0 - \pi_0'\mathbf{z}\right) + \beta_{20}'x\left(1 - \Phi\left(\gamma_0 - \pi_0'\mathbf{z}\right)\right),$$

and a $\sqrt{n}$-consistent estimator can also result from this conditional mean. When $\mathbf{x}$ is also endogenous, the analysis is much messier even if we assume $v_x \perp v_q|\mathbf{z}$ (which strengthens KST's assumption $v_x \perp 1(q \leq \gamma_0)|\mathbf{z}$). This is because $x = \Pi_x'\mathbf{z} + v_x$, so

$$\mathbb{E}[y|\mathbf{z}, \underline{\gamma} \leq q \leq \overline{\gamma}]$$

$$= \beta_{10}'\left(\Pi_{x0}'\mathbf{z}\right)P\left(q \leq \gamma_0|\mathbf{z}, \underline{\gamma} \leq q \leq \overline{\gamma}\right) + \beta_{20}'\left(\Pi_{x0}'\mathbf{z}\right)P\left(q > \gamma_0|\mathbf{z}, \underline{\gamma} \leq q \leq \overline{\gamma}\right)$$

$$+ \mathbb{E}\left[\beta_{10}'v_x 1(q \leq \gamma_0)|\mathbf{z}, \underline{\gamma} \leq q \leq \overline{\gamma}\right] + \mathbb{E}\left[\beta_{20}'v_x 1(q > \gamma_0)|\mathbf{z}, \underline{\gamma} \leq q \leq \overline{\gamma}\right]$$

$$+ \kappa_0\mathbb{E}\left[v_q|\mathbf{z}, \underline{\gamma} \leq q \leq \overline{\gamma}\right],$$

where the terms involving v_x will not disappear.

Note that KST's estimation is inspired by Heckman ([1979](#)). In the Heckit model (5), we note that $d = 1(\mathbf{z}'\gamma + v > 0)$ is observable, whereas the conditioning events $1(q \leq \gamma_0)$ and $1(q > \gamma_0)$ in KST's estimation are unobservable. Estimation is based on $\mathbb{E}[y|d = 1, \mathbf{x}, \mathbf{z}]$, that is, the conditioning set does not depend on any unknown parameter, which results in consistent estimation of β. Similarly, our CF-I is based on $\mathbb{E}[y|\mathbf{z}, q]$ and CF-II is based on $\mathbb{E}[y|\mathbf{z}, \mathbf{x}]$, that is, neither conditioning set depends on γ_0; this is why our two CF estimators are consistent. Compared with CF-I, KST integrate out q on a range depending on γ_0 in $\mathbb{E}[y|\mathbf{z}, q]$ (or project the q dimension onto the two indicator functions $1(q \leq \gamma_0)$ and $1(q > \gamma_0)$).

In Section SD.1 of the Supplementary Material, we provide more discussion on why we cannot include unknown parameters in the conditioning set of M-estimation to generate consistent estimators by comparing M estimators and Z estimators in the general sense.

3. TWO CONTROL FUNCTION APPROACHES FOR γ

In this section, we propose two CF approaches to estimate γ. In both approaches, we allow x and q to be endogenous, treating the exogenous q specification as a special case. First, we introduce some notation for future use. Define $\mathcal{F}_{i-1}$ as the sigma field generated by $\{\mathbf{z}_{i-j}, \mathbf{x}_{i-1-j}, u_{i-1-j} : j \geq 0\}$; then $\mathbb{E}[u_i|\mathcal{F}_{i-1}] = 0$, that is, $\{u_i\}_{i=1}^n$ is a martingale difference sequence.[4] When the sample is random, we can just replace $\mathcal{F}_{i-1}$ by $\mathbf{z}_i$. Suppose the first-stage regression is written in stacked form as

$$\mathbf{x}_i = \Pi'\mathbf{z}_i + \mathbf{v}_i,$$

with components x_i and q_i taking the form (6) so that $\Pi = (\Pi_x, \pi)$ and $\mathbf{v}_i = (v'_{xi}, v_{qi})'$, where $\mathbb{E}[\mathbf{v}_i|\mathcal{F}_{i-1}] = \mathbf{0}$. Denote the estimates of Π and $\mathbf{v}_i$ as $\widehat{\Pi} = (\widehat{\Pi}_x, \widehat{\pi})$ and $\widehat{\mathbf{v}}_i = (\widehat{v}'_{xi}, \widehat{v}_{qi})'$, and that of $\mathbf{g}_i = \Pi'\mathbf{z}_i$ by $\widehat{\mathbf{g}}_i = \widehat{\Pi}'\mathbf{z}_i$. For exogenous covariates in x_i such as 1, the corresponding components in Π are known and the corresponding components of v_{xi} are zero since such covariates are included in $\mathbf{z}_i$. In the two approaches, we use only the first-stage error terms and residuals for endogenous regressors; to avoid introducing further notation, we still use $\mathbf{v}_i$ and $\widehat{\mathbf{v}}_i$ to denote such error terms and residuals. To unify the notations in the two approaches, define $\vartheta = (\theta', \kappa')'$, where $\theta = (\gamma, \beta')'$ is the parameter of interest, and κ collects all other nuisance parameters which may be different in the two CF approaches. We will also use the same notations $\check{\mathbf{x}}_i$ (and $\tilde{\mathbf{x}}_i$) for the regressors and $\overline{\beta}_\ell$ for the slope parameters in the two approaches although they may be different. As mentioned in the Section 1, CF-II is our favorite approach, so CF-I is introduced here only for completeness and all its properties such as consistency, asymptotic distribution and LR inference are provided in Section SD.2 of the Supplementary Material.

---

[4]In CH, there is a typo in the definition of $\mathcal{F}_{i-1}$; they miswrite $\mathbf{x}_{i-1-j}$ as $\mathbf{x}_{i-j}$.

3.1. Definition of the Two Control Function Estimators for γ

In the first CF approach (labeled CF-I), assume $\mathbb{E}[\mathbf{v}_i|\mathcal{F}_{i-1}, q_i] = \mathbb{E}[\mathbf{v}_i|v_{qi}]$ and $\mathbb{E}[u_i|\mathcal{F}_{i-1}, q_i] = \mathbb{E}[u_i|v_{qi}]$. Then

$$
\begin{aligned}
\mathbb{E}[y_i|\mathcal{F}_{i-1}, q_i] &= \left\{\beta_1'\left(\Pi'\mathbf{z}_i\right) + \mathbb{E}\left[\beta_1'\mathbf{v}_i|v_{qi}\right] + \sigma_1\mathbb{E}\left[u_i|v_{qi}\right]\right\}1(q_i \le \gamma_0) \\
&\quad + \left\{\beta_2'\left(\Pi'\mathbf{z}_i\right) + \mathbb{E}\left[\beta_2'\mathbf{v}_i|v_{qi}\right] + \sigma_2\mathbb{E}\left[u_i|v_{qi}\right]\right\}1(q_i > \gamma_0),
\end{aligned}
$$

where conditioning also on q_i rather than only on the exogenous variables is because y_i is nonlinear in q_i, and to lighten notational burden the subscript 0 is only added to γ to indicate the true value, not to other parameters whenever this does not create confusion. If we assume further that $\mathbb{E}[\mathbf{v}_i|v_{qi}] = \mathbf{g}_1(v_{qi}; \varphi)$ and $\mathbb{E}[u_i|v_{qi}] = g_2(v_{qi}; \kappa)$ for some possibly nonlinear functions $\mathbf{g}_1$ and g_2 with φ and κ being unknown parameters, then

$$
\begin{aligned}
\mathbb{E}[y_i|\mathcal{F}_{i-1}, q_i] &= \left[\beta_1'\mathbf{g}_i + \beta_1'\mathbf{g}_1\left(v_{qi}; \varphi\right) + \sigma_1 g_2\left(v_{qi}; \kappa\right)\right]1(q_i \le \gamma_0) \\
&\quad + \left[\beta_2'\mathbf{g}_i + \beta_2'\mathbf{g}_1\left(v_{qi}; \varphi\right) + \sigma_1 g_2\left(v_{qi}; \kappa\right)\right]1(q_i > \gamma_0),
\end{aligned}
$$

and the error term

$$
\begin{aligned}
e_i^0 &= \left[\beta_1'\left(\mathbf{v}_i - \mathbf{g}_1\left(v_{qi}; \varphi\right)\right) + \sigma_1\left(u_i - g_2\left(v_{qi}; \kappa\right)\right)\right]1(q_i \le \gamma_0) \\
&\quad + \left[\beta_2'\left(\mathbf{v}_i - \mathbf{g}_1\left(v_{qi}; \varphi\right)\right) + \sigma_2\left(u_i - g_2\left(v_{qi}; \kappa\right)\right)\right]1(q_i > \gamma_0) \\
&=: e_{1i}1(q_i \le \gamma_0) + e_{2i}1(q_i > \gamma_0),
\end{aligned}
$$

where $\mathbb{E}\left[e_i^0|\mathcal{F}_{i-1}, q_i\right] = 0$, and we use the superscript 0 in e_i^0 to emphasize that it is γ_0 appearing in the definition of e_i^0. We can estimate γ by minimizing

$$
\begin{aligned}
S_n(\theta) = \sum_{i=1}^{n}\Big[&y_i - \left(\beta_1'\widehat{\mathbf{g}}_i + \beta_1'\mathbf{g}_1\left(\widehat{v}_{qi}; \varphi\right) + \sigma_1 g_2\left(\widehat{v}_{qi}; \kappa\right)\right)1(q \le \gamma) \\
&- \left(\beta_2'\widehat{\mathbf{g}}_i + \beta_2'\mathbf{g}_1\left(\widehat{v}_{qi}; \varphi\right) + \sigma_2 g_2\left(\widehat{v}_{qi}; \kappa\right)\right)1(q > \gamma)\Big]^2,
\end{aligned} \tag{16}
$$

where $\kappa = (\varphi', \kappa', \sigma_1, \sigma_2)' =: (\varphi', \kappa', \sigma')'$ collects the nuisance parameters in θ.

In the second CF approach (labeled CF-II), assume that $\mathbb{E}[u_i|\mathcal{F}_{i-1}, \mathbf{x}_i] = \mathbb{E}[u_i|\mathbf{v}_i]$. Then

$$
\mathbb{E}[y_i|\mathcal{F}_{i-1}, \mathbf{x}_i] = \left(\mathbf{x}_i'\beta_1 + \sigma_1\mathbb{E}[u_i|\mathbf{v}_i]\right)1(q_i \le \gamma_0) + \left(\mathbf{x}_i'\beta_2 + \sigma_2\mathbb{E}[u_i|\mathbf{v}_i]\right)1(q_i > \gamma_0),
$$

where we condition also on x_i besides $\mathcal{F}_{i-1}$ and q_i as in CF-I. If assume further that $\mathbb{E}[u_i|\mathbf{v}_i] = g_3(\mathbf{v}_i; \psi)$ for some possibly nonlinear function g_3 with ψ being unknown parameters, then

$$
\begin{aligned}
\mathbb{E}[y_i|\mathcal{F}_{i-1}, \mathbf{x}_i] &= \left(\mathbf{x}_i'\beta_1 + \sigma_1 g_3\left(\mathbf{v}_i; \psi\right)\right)1(q_i \le \gamma_0) + \left(\mathbf{x}_i'\beta_2 + \sigma_2 g_3\left(\mathbf{v}_i; \psi\right)\right)1(q_i > \gamma_0),
\end{aligned}
$$

and the error term

$$e_i^0 = \sigma_1 \left(u_i - g_3 \left(\mathbf{v}_i; \psi\right)\right) 1 \left(q_i \leq \gamma_0\right) + \sigma_2 \left(u_i - g_3 \left(\mathbf{v}_i; \psi\right)\right) 1 \left(q_i > \gamma_0\right)$$
$$=: e_{1i} 1(q_i \leq \gamma_0) + e_{2i} 1(q_i > \gamma_0), \tag{17}$$

where $\mathbb{E}\left[e_i^0 | \mathcal{F}_{i-1}, \mathbf{x}_i\right] = 0$. Then γ can be estimated by minimizing

$$S_n(\theta)$$
$$= \sum_{i=1}^n \left[y_i - \left(\mathbf{x}_i' \beta_1 + \sigma_1 g_3 \left(\widehat{\mathbf{v}}_i; \psi\right)\right) 1 \left(q_i \leq \gamma\right) - \left(\mathbf{x}_i' \beta_2 + \sigma_2 g_3 \left(\widehat{\mathbf{v}}_i; \psi\right)\right) 1 \left(q_i > \gamma\right)\right]^2, \tag{18}$$

where $\kappa = \left(\psi', \sigma_1, \sigma_2\right)' =: \left(\psi', \sigma'\right)'$.

CH and KST use the idea of CF-I, while the classical CF approach in ELR uses the idea of CF-II, that is, conditioning on all endogenous regressors and exogenous instruments. In other words, CF-II is the genuine CF approach in the literature where the endogeneity is controlled by $\mathbb{E}[u_i|\mathbf{v}_i]$ as the regressor $\mathbf{x}_i$ includes two components, $\mathbf{z}_i$ and $\mathbf{v}_i$, and only $\mathbf{v}_i$ is correlated with u_i. We state CF-I here to show the correct form of CFs in the KST approach. In ELR, using CFs and using moment conditions are equivalent. However, in ETR, they are different (see the Sections 1 and 2.3 for the differences between GMM estimators and M-estimators like CF-I and CF-II). CF-I imposes assumptions on both $\mathbb{E}[u_i|\mathcal{F}_{i-1}, q_i]$ and $\mathbb{E}[\mathbf{v}_i|\mathcal{F}_{i-1}, q_i]$, whereas CF-II imposes no assumptions on $\mathbb{E}[\mathbf{v}_i|\mathcal{F}_{i-1}, q_i]$ but only assumptions on $\mathbb{E}[u_i|\mathcal{F}_{i-1}, x_i, q_i]$. Of course, rigorously speaking, these two groups of assumptions do not nest each other.

Because CF-II is based on $\mathbb{E}[y_i|\mathcal{F}_{i-1}, x_i, q_i]$, whereas CF-I is based on $\mathbb{E}[y_i|\mathcal{F}_{i-1}, q_i]$, CF-II controls for more endogeneity effects than CF-I (when $\mathbf{x}_i = x_i$ and only q_i is endogenous, they are the same). So CF-II is indeed more appealing from a theory perspective. Actually, CF-I projects all endogeneity on the v_{qi} dimension, that is, these two approaches are connected. Specifically,

$$\mathbb{E}[y_i|\mathcal{F}_{i-1}, q_i]$$
$$= \mathbb{E}[\mathbb{E}[y_i|\mathcal{F}_{i-1}, x_i, q_i]|\mathcal{F}_{i-1}, q_i]$$
$$= \mathbb{E}\left[\left(\mathbf{x}_i' \beta_1 + \sigma_1 g_3 \left(\mathbf{v}_i; \psi\right)\right) 1 \left(q_i \leq \gamma_0\right)\right.$$
$$\left. + \left(\mathbf{x}_i' \beta_2 + \sigma_2 g_3 \left(\mathbf{v}_i; \psi\right)\right) 1 \left(q_i > \gamma_0\right)|\mathcal{F}_{i-1}, q_i\right]$$
$$= \left[\beta_1' \left(\Pi' \mathbf{z}_i\right) + \beta_1' \mathbb{E}\left[\mathbf{v}_i|v_{qi}\right] + \sigma_1 \mathbb{E}\left[g_3 \left(\mathbf{v}_i; \psi\right)|v_{qi}\right]\right] 1(q_i \leq \gamma_0)$$
$$+ \left[\beta_2' \left(\Pi' \mathbf{z}_i\right) + \beta_2' \mathbb{E}\left[\mathbf{v}_i|v_{qi}\right] + \sigma_2 \mathbb{E}\left[g_3 \left(\mathbf{v}_i; \psi\right)|v_{qi}\right]\right] 1(q_i > \gamma_0)$$
$$= \left[\beta_1' \mathbf{g}_i + \beta_1' \mathbf{g}_1 \left(v_{qi}; \varphi\right) + \sigma_1 g_2 \left(v_{qi}; \kappa\right)\right] 1(q_i \leq \gamma_0)$$
$$+ \left[\beta_2' \mathbf{g}_i + \beta_2' \mathbf{g}_1 \left(v_{qi}; \varphi\right) + \sigma_2 g_2 \left(v_{qi}; \kappa\right)\right] 1(q_i > \gamma_0), \tag{19}$$

where we need strengthen $\mathbb{E}[\mathbf{v}_i|\mathcal{F}_{i-1}, q_i] = \mathbb{E}[\mathbf{v}_i|v_{qi}]$ to $\mathbb{E}[g_3(\mathbf{v}_i; \psi)|\mathcal{F}_{i-1}, q_i] = \mathbb{E}[g_3(\mathbf{v}_i; \psi)|v_{qi}]$, and note that $\mathbb{E}[g_3(\mathbf{v}_i; \psi)|v_{qi}] = \mathbb{E}[\mathbb{E}[u_i|\mathbf{v}_i]|v_{qi}] = \mathbb{E}[u_i|v_{qi}] = g_2(v_{qi}; \kappa)$ by the law of iterated expectation.

We now show the consistency of the CF-II estimator of γ based on (18) to formalize our identification scheme. First, the following assumptions are imposed. Note here that we label our assumptions as "C-II" since the label "C-I" has been used for CF-I in Section SD.2 of the Supplementary Material.

Assumption C-II:

1. $\{w_i\}_{i=1}^n := \{(y_i, x_i', q_i, \mathbf{z}_i')'\}_{i=1}^n$ are strictly stationary and ergodic; $\theta \in \Theta$ with Θ being compact; $(\beta_{10}', \sigma_{10})' \neq (\beta_{20}', \sigma_{20})'$.
2. $\mathbb{E}[\mathbf{v}_i | \mathcal{F}_{i-1}] = \mathbf{0}$, and $\mathbb{E}[u_i | \mathcal{F}_{i-1}, \mathbf{x}_i] = \mathbb{E}[u_i | \mathbf{v}_i] = g_3(\mathbf{v}_i; \psi_0)$.
3. $\mathbb{E}[\mathbf{z}_i \mathbf{z}_i'] > 0$, $\mathbb{E}[\|\mathbf{z}_i\|^2] < \infty$ and $\mathbb{E}[\|\mathbf{v}_i\|^2] < \infty$.
4. $\mathbb{E}[g_3(\mathbf{v}_i; \psi_0)^2] < \infty$ and $\mathbb{E}[(e_i^0)^2] < \infty$.
5. For all $\gamma \in \Gamma$, $f(\gamma) \leq \bar{f} < \infty$, $f(\gamma)$ is continuous at γ_0 and $f := f(\gamma_0) > 0$, $P(q < \underline{\gamma}) > 0$ and $P(q > \overline{\gamma}) > 0$, where $f(\cdot)$ is the density function of q.
6. $g_3(\mathbf{v}; \psi)$ is a Lipschitz function in each of its arguments, that is, there is a positive constant $C < \infty$ such that

   $$\|g_3(\mathbf{v}_1; \psi) - g_3(\mathbf{v}_2; \psi)\| \leq C \|\mathbf{v}_1 - \mathbf{v}_2\|,$$
   $$\|g_3(\mathbf{v}; \psi_1) - g_3(\mathbf{v}; \psi_2)\| \leq C(\mathbf{v}) \|\psi_1 - \psi_2\|,$$

   with $\mathbb{E}[C(\mathbf{v})^2] < C$.
7. $P(\mathbf{x}'\beta_\ell + \sigma_\ell g_3(\mathbf{v}; \psi) \neq \mathbf{x}'\beta_{\ell 0} + \sigma_{\ell 0} g_3(\mathbf{v}; \psi_0) | q) > 0$ for any $\theta_\ell \neq \theta_{\ell 0}$ and any q value in its support, where $\theta_\ell = (\beta_\ell', \sigma_\ell, \psi')'$.

Condition C-II.1 covers weakly dependent data beyond random samples but excludes time trends and integrated processes; it also assumes the existence of threshold effects in the conditional mean and/or conditional variance of y. C-II.2 collects the nonlinear endogeneity assumptions which imply the correct specification of the conditional mean in the structural equation and reduced form. C-II.3 implies the consistency of the first-stage estimator $\widehat{\Pi}$. C-II.4 is implied by $\mathbb{E}[u_i^2] < \infty$ but we express it in this way for convenience. C-II.5 includes some regularity conditions on $f(\cdot)$; especially, it requires the threshold variable to have a continuous distribution and excludes the possibility that γ_0 falls on the boundary of q's support. C-II.6 guarantees that replacing $\mathbf{v}_i$ by $\widehat{\mathbf{v}}_i$ will not affect the consistency of $\widehat{\gamma}$. C-II.7 is the key assumption for identification of γ when $\widehat{\mathbf{v}}_i$ is replaced by $\mathbf{v}_i$, for example, it excludes the case where $g_3(\mathbf{v}; \psi)$ takes a linear form in ψ and implicitly assumes $\delta_\beta' \mathbb{E}[\mathbf{x}_i \mathbf{x}_i' | q] \delta_\beta > 0$. It guarantees the uniqueness of $\arg\min_\theta \text{plim}(S_n(\theta))$ at θ_0; it is actually the form of the identification assumptions in nonlinear least squares (see, e.g., Section 2.2.2 of Newey and McFadden, 1994) required in nonlinear TR models, where linear and nonlinear TR models refer to the models where the conditional mean of y_i in each regime is linear and nonlinear in parameters, respectively.

LEMMA 1. *Under Assumption C-II, $\widehat{\theta} \xrightarrow{p} \theta_0$.*

3.2. Identification of γ and Linear Endogeneity Forms

From the analysis in Section 3.1, we can see three drawbacks of the nonlinear specification of endogeneity forms. First, different from linear TR models, it is hard to find general primitive conditions for consistency of $\widehat{\gamma}$, that is, we need check C-II.7 case by case. Second, there is typically no guideline for the nonlinear specification and the functional form of $g_3(\cdot;\cdot)$ is often arbitrary. On the other hand, linear TR models introduce nonlinearity into the system only through the threshold effect, thereby providing a class of parsimonious models; and, as shown in Petruccelli (1992), linear threshold autoregressive models can approximate a general class of time series processes (e.g., exponential autoregressive and invertible bilinear processes) almost surely. In other words, use of the linear TR model reduces the need to introduce nonlinearity in each regime. Third, and most importantly, optimization when the endogeneity takes nonlinear forms can be troublesome in practice. In particular, we need first concentrate on $(\psi',\gamma)'$ and then on γ to find $\widehat{\gamma}$; and when the dimension of ψ is large, numerical minimization is burdensome. These three drawbacks go some way to explain why the literature seldom considers nonlinear TR models.

To ease discussion and provide the most practical estimation procedures, we assume that endogeneity takes linear forms in both approaches. In CF-I, we further assume that $\mathbb{E}[\mathbf{v}_i|v_{qi}] = \varphi v_{qi}$ and $\mathbb{E}[u_i|v_{qi}] = \kappa v_{qi}$. Then

$$\mathbb{E}[y_i|\mathcal{F}_{i-1}, q_i] = [\beta_1'\mathbf{g}_i + \kappa_1 v_{qi}]\,1(q_i \le \gamma_0) + [\beta_2'\mathbf{g}_i + \kappa_2 v_{qi}]\,1(q_i > \gamma_0), \tag{20}$$

where we note that when $q \in \mathbf{x}$ and is endogenous, the last element of φ is 1, and $\kappa_\ell = \beta_\ell'\varphi + \sigma_\ell\kappa$. Different from the nonlinear endogeneity case, φ and κ cannot be identified separately. In fact, we must exclude the cases where $\mathbf{g}_1(\cdot;\cdot)$ and $g_2(\cdot;\cdot)$ take linear forms in (16) to identify φ and κ. Since our main interest lies in $\theta = (\gamma, \beta')'$, absorbing φ and κ in the nuisance parameter $\kappa = (\kappa_1, \kappa_2)'$ is not critical. Now, γ is estimated by minimizing

$$S_n(\theta) = \sum_{i=1}^{n} \left[y_i - \left(\beta_1'\widehat{\mathbf{g}}_i + \kappa_1\widehat{v}_{qi}\right)1(q \le \gamma) - \left(\beta_2'\widehat{\mathbf{g}}_i + \kappa_2\widehat{v}_{qi}\right)1(q_i > \gamma) \right]^2$$

$$:= \sum_{i=1}^{n} \left[y_i - \widetilde{\widehat{\mathbf{x}}}_i'\overline{\beta}_1 1(q_i \le \gamma) - \widetilde{\widehat{\mathbf{x}}}_i'\overline{\beta}_2 1(q_i > \gamma) \right]^2, \tag{21}$$

where $\theta = (\theta', \kappa')' =: (\gamma, \underline{\theta}')'$, $\underline{\theta}' = (\beta', \kappa') =: (\overline{\beta}_1', \overline{\beta}_2')$ with $\overline{\beta}_\ell' = (\beta_\ell', \kappa_\ell)$, and $\widetilde{\widehat{\mathbf{x}}}_i' = (\widehat{\mathbf{g}}_i', \widehat{v}_{qi})$ is the generated regressor of $\check{\mathbf{x}}_i' = (\mathbf{g}_i', v_{qi})$.

Although the objective function (21) is based on the conditional mean $\mathbb{E}[y_i|\mathcal{F}_{i-1}, q_i]$, (21) carries some threshold effect information in the conditional variance because $\delta_\kappa := \kappa_1 - \kappa_2 = \delta_\beta'\varphi + \delta_\sigma\kappa$ involves δ_σ. In contrast, the usual LS estimator (without endogeneity) and CH's 2SLS estimator (with endogenous x_i but exogenous q_i) explore only the threshold effect information in the conditional mean even though a threshold effect in the conditional variance may be present

(i.e., $\delta_\sigma \neq 0$). Note that, were $\delta_\sigma = 0$, we would still have $\kappa_1 \neq \kappa_2$ unless $\varphi = \mathbf{0}$. But when $q_i \in \mathbf{z}_i$, $\varphi = \mathbf{0}$ and $\kappa = 0$, so we return to the framework of CH. Also, different from KST, κ is assumed to be fixed rather than to shrink to zero (i.e., we allow for strong endogeneity) and v_{qi} need not follow $\mathcal{N}(0,1)$.

In CF-II, we assume that $\mathbb{E}[u_i|\mathbf{v}_i] = \psi'\mathbf{v}_i = \psi_x' v_{xi} + \psi_q v_{qi}$ with $\psi' = (\psi_x', \psi_q)$. Then

$$\mathbb{E}[y_i|\mathcal{F}_{i-1}, \mathbf{x}_i] = (\mathbf{x}_i'\beta_1 + \sigma_1 \psi'\mathbf{v}_i)\,1\,(q_i \leq \gamma_0) + (\mathbf{x}_i'\beta_2 + \sigma_2 \psi'\mathbf{v}_i)\,1\,(q_i > \gamma_0)\,. \quad \textbf{(22)}$$

So γ can be estimated by minimizing

$$\begin{aligned}
S_n(\theta) &= \sum_{i=1}^n \left[y_i - (\mathbf{x}_i'\beta_1 + \sigma_1 \psi'\widehat{\mathbf{v}}_i)\,1\,(q_i \leq \gamma) - (\mathbf{x}_i'\beta_2 + \sigma_2 \psi'\widehat{\mathbf{v}}_i)\,1\,(q_i > \gamma) \right]^2 \\
&=: \sum_{i=1}^n \left[y_i - (\mathbf{x}_i'\beta_1 + \kappa_1'\widehat{\mathbf{v}}_i)\,1\,(q_i \leq \gamma) - (\mathbf{x}_i'\beta_2 + \kappa_2'\widehat{\mathbf{v}}_i)\,1\,(q_i > \gamma) \right]^2 \\
&=: \sum_{i=1}^n \left[y_i - \widehat{\widetilde{\mathbf{x}}}_i'\overline{\beta}_1 1\,(q_i \leq \gamma) - \widehat{\widetilde{\mathbf{x}}}_i'\overline{\beta}_2 1\,(q_i > \gamma) \right]^2,
\end{aligned} \quad \textbf{(23)}$$

where the nuisance parameter $\kappa = (\kappa_1', \kappa_2')'$ with $\kappa_\ell = \sigma_\ell \psi$, $\overline{\beta}_\ell' = (\beta_\ell', \kappa_\ell')$, and $\widehat{\widetilde{\mathbf{x}}}_i' = (\mathbf{x}_i', \widehat{\mathbf{v}}_i')$ is the generated regressor of $\widetilde{\mathbf{x}}_i' = (\mathbf{x}_i', \mathbf{v}_i')$. Note here that we express the objective function of CF-II in the same format as in CF-I by redefining $\widehat{\widetilde{\mathbf{x}}}_i'$ and $\overline{\beta}_\ell'$. Like (19), we can discuss the relationship between the parameters in CF-I and CF-II. Now, $\mathbb{E}[u_i|v_{qi}] = \mathbb{E}[\mathbb{E}[u_i|\mathbf{v}_i]|v_{qi}] = \psi'\varphi v_{qi}$, that is, $\psi'\varphi$ plays the role of κ in CF-I. So the coefficients of v_{qi} in the conditional mean of u_i in these two approaches (κ in CF-I and ψ_q in CF-II) are not equal unless either ψ_x or φ_x is zero, that is, either x_i is exogenous or $x_i \perp q_i|\mathcal{F}_{i-1}$.

As in CF-I, ψ cannot be identified separately, and the identification of ψ excludes the linear specification of $g_3(\cdot; \cdot)$ in (18); the threshold effect information in the conditional variance is also explored through $\delta_\kappa := \kappa_1 - \kappa_2 = \delta_\sigma \psi$; ψ is fixed, that is, CF-II also allows for strong endogeneity. Different from CF-I, even if q is exogenous (i.e., in CH's setup), the threshold effect information in the conditional variance is still explored. When q is exogenous, $\psi'\mathbf{v}_i = \psi_x'\mathbf{v}_{xi}$, and we need only control for the endogeneity of x. If $\delta_\sigma = 0$, then $\kappa_1 = \kappa_2$ and the model is simpler. As shown in Perron and Qu (2006), the asymptotic distribution of the estimator of γ based on (23) is not affected when some constraints on the slope parameters in the two regimes are imposed as long as the true model satisfies these constraints. The intuition is simple: the constraints can affect only the efficiency (i.e., the asymptotic variance) of the slope estimator, while as long as the slope estimator is $\sqrt{n}$-consistent, the asymptotic distribution of the γ estimator is the same, which is in turn due to the asymptotic independence between the slope estimator and the γ estimator. In Section SD.4 of the Supplementary Material, we develop the asymptotic theory for γ and β estimators when the model satisfies some constraints such as part of β_1 and β_2 are the same since such constraints are imposed in our

simulations and empirical application. The CF-II approach is not considered in the ETR literature.

As in CMT, we can extend v_{qi} in (20) and $\mathbf{v}_i$ in (22) to some functions of them and meanwhile maintain the linear TR forms. To save space, we focus on CF-II here. For one example, if $u_i \sim \mathcal{N}(0,1)$, and $(u_i, \mathbf{v}_i')'$ has a Gaussian copula, then by Proposition 2 of CMT, $\mathbb{E}[u_i|\mathbf{v}_i] = \psi^{*'}\mathbf{v}_i^*$ which reduces to $\psi'\mathbf{v}_i$ when $(u_i, \mathbf{v}_i')'$ follows a jointly normal distribution, where $\mathbf{v}_i^* = (v_{1i}^*, \ldots, v_{d-1,i}^*, v_q^*)'$ with $v_{ji}^* = \Phi^{-1}(F_{v_j}(v_{ji}))$, and $\psi^{*'} = P'\Xi$ with $P' = Corr(u_i, \mathbf{v}_i^{*'})$ and Ξ being the correlation matrix of $\mathbf{v}_i^*$. For another example, if $u_i \sim t_{v_u}$, and $(u_i, \mathbf{v}_i')'$ has a t copula, then $\mathbb{E}[u_i|\mathbf{v}_i] = \psi^{*'}\mathbf{v}_i^*$, where $\mathbf{v}_i^* = (v_{1i}^*, \ldots, v_{d-1,i}^*, v_{qui}^*)'$, $v_{ji}^* = T_{v_u}^{-1}(F_{v_j}(v_{ji}))$ with $T_{v_u}^{-1}$ being the quantile function of t_{v_u} will reduce to v_{ji}/σ_j when $v_{ji} \sim \sigma_j t_{v_u}$, $v_{qui}^* = T_{v_u}^{-1}(F_{v_q}(v_{qi}))$, and ψ^* takes the same form as above.[5] In other words, instead of assuming $\mathbb{E}[u_i|\mathbf{v}_i] = g_3(\mathbf{v}_i; \psi)$, we can assume $\mathbb{E}[u_i|\mathbf{v}_i] = \psi^{*'}g_3(\mathbf{v}_i)$ for some nonlinear function g_3 implied from copula theory. Now, the analysis in this paper goes through, just replacing $\mathbf{v}_i$ (or $\widehat{\mathbf{v}}_i$) in CF-II by $g_3(\mathbf{v}_i)$ (or $\widehat{g}_3(\widehat{\mathbf{v}}_i)$) everywhere. In short, the linear endogeneity forms are not as restrictive as they might seem to be.

The above extensions require that all components of $\mathbf{v}$ are continuous. When x_i contains some discrete regressors, it seems to us that the linear endogeneity assumption in CF-II is more robust than in CF-I. In such a case, assuming $\mathbb{E}[\mathbf{v}_i|v_{qi}]$ to be linear in v_{qi} in CF-I seems questionable. For such discrete regressors, assuming $\mathbb{E}[\mathbf{x}_i|\mathcal{F}_{-1}] = \Pi'\mathbf{z}_i$ also seems questionable. This is why CH assumes $\mathbb{E}[\mathbf{x}_i|\mathcal{F}_{-1}] = g(\mathbf{z}_i; \Pi)$ for some possibly nonlinear (e.g., logit or probit) function of $\mathbf{z}_i$. In other words, in such scenarios, we must use an objective function like (16). On the other hand, assuming $\mathbb{E}[u_i|\mathbf{v}_i] = \psi_x'v_{xi} + \psi_q v_{qi}$ seems reasonable even if some components of x_i are discrete. When x_i is discrete, it is better to replace v_{xi} by the generalized error of Gouriéroux, Monfort, Renault and Trognon (1987). For example, if $x_i \in \mathbb{R}$ satisfies the reduced form equation $x_i = 1(\Pi_x'\mathbf{z}_i + v_{xi} > 0)$, then we can replace v_{xi} in $\mathbb{E}[u_i|v_{xi}, v_{qi}] = \psi_x v_{xi} + \psi_q v_{qi}$ by $r_{xi} = x_i \lambda(\Pi_x'\mathbf{z}_i) - (1-x_i)\lambda(-\Pi_x'\mathbf{z}_i)$. However, we cannot replace v_{xi} by r_{xi} in CF-I to calculate $\mathbb{E}[\mathbf{v}_i|v_{qi}]$. As in CH, we can also assume the first-stage reduced form is unstable, $\mathbb{E}[\mathbf{x}_i|\mathcal{F}_{-1}] = \Pi_1'\mathbf{z}_i 1(z_{1i} \leq \rho) + \Pi_2'\mathbf{z}_i 1(z_{1i} > \rho)$, and then $\mathbf{g}_i = \mathbb{E}[\mathbf{x}_i|\mathcal{F}_{-1}]$ and $\mathbf{v}_i = \mathbf{x}_i - \mathbb{E}[\mathbf{x}_i|\mathcal{F}_{-1}]$, where $\Pi_1 - \Pi_2$ is fixed and z_{1i} is a component of $\mathbf{z}_i$. For all these specifications of $\mathbb{E}[\mathbf{x}_i|\mathcal{F}_{-1}]$, as long as $\mathbb{E}[u_i|\mathbf{v}_i]$ (or $\mathbb{E}[u_i|r_{xi}, v_{qi}]$) is linear, our consistency proof of $\widehat{\gamma}$ in CF-II in the next subsection can still go through if $\widehat{\Pi}$ (or the parameters in the first-stage equation) is $\sqrt{n}$-consistent

---

[5]Different from CMT, we calculate $\mathbb{E}[u_i|\mathbf{v}_i]$ rather than $\mathbb{E}[u_i|\mathbf{x}_i]$. If we impose similar assumptions on $(u_i, \mathbf{x}_i')'$ rather than $(u_i, \mathbf{v}_i')'$, then we can similarly have $\mathbb{E}[u_i|\mathbf{x}_i] = \psi^{*'}\mathbf{x}_i^*$. However, $\mathbf{x}_i^* = (x_{1i}^*, \ldots, x_{d-1,i}^*, q_i^*)'$ with $x_{ji}^* = \Phi^{-1}(F_{x_j}(x_{ji}))$ or $T_{v_u}^{-1}(F_{x_j}(x_{ji}))$ and $q_i^* = \Phi^{-1}(F_q(q_i))$ or $T_{v_u}^{-1}(F_q(q_i))$, so when $F_{x_j} = \Phi$ or T_{v_u}, $x_{ji}^* = x_{ji}$, and when $F_q = \Phi$ or T_{v_u}, $q_i^* = q_i$, we have the multicollinearity problem given that x_{ji} or q_i itself is a regressor. Although F_{x_j} and F_q are unknown so that $\mathbf{x}_i^*$ need be estimated, that is, $\mathbf{x}_i^*$ and $\mathbf{x}_i$ are not perfectly multicollinear in finite samples, they are indeed so asymptotically. As a result, the asymptotic distributions of $\widehat{\gamma}$ and $\widehat{\beta}$, which are not stated in CMT, cannot be well developed. On the other hand, an advantage of CMT's specification is that no instruments are required.

and the asymptotic distribution of $\widehat{\gamma}$ takes the same form with the meaning of notations appropriately adjusted.[6] Because CF-II is our favored approach and the linear endogeneity assumption is reasonable in CF-II, we will maintain such an assumption in CF-II in the rest of the paper. As to CF-I, we also impose the linear endogeneity assumption to avoid technical difficulties, that is, it is better to assume the endogenous regressors in x_i to be continuous if practitioners wish to use CF-I. Finally, note that the existing literature such as CH, KST, and CMT does not explicitly consider discrete x_i scenarios.

In the semiparametric framework, we need to replace $\kappa_1 \widehat{v}_{qi}$ and $\kappa_2 \widehat{v}_{qi}$ by $h_1\left(\widehat{v}_{qi}\right)$ and $h_2\left(\widehat{v}_{qi}\right)$ in CF-I and replace $\kappa_1' \widehat{\mathbf{v}}_i$ and $\kappa_2' \widehat{\mathbf{v}}_i$ by $h_1\left(\widehat{\mathbf{v}}_i\right)$ and $h_2\left(\widehat{\mathbf{v}}_i\right)$ (or by $h_{1x}\left(\widehat{v}_{xi}\right) + h_{1q}\left(\widehat{v}_{qi}\right)$ and $h_{2x}\left(\widehat{v}_{xi}\right) + h_{2q}\left(\widehat{v}_{qi}\right)$) in CF-II, where h_1 and h_2 can be any (smooth) functions. It seems that the first approach does not suffer from the curse of dimensionality compared with the second approach. Following this paper, Kourtellos et al. (2022) do the semiparametric extension of CF-I, and their method may be applied to the extension of CF-II. We will not discuss semiparametric approaches in this paper because the key advantage of the CF approaches above is their parametric setup. If the setup does contain nonparametric components (e.g., $\mathbb{E}[\mathbf{x}_i | \mathbf{z}_i]$, $\mathbb{E}\left[\mathbf{v}_i | v_{qi}\right]$, $\mathbb{E}\left[u_i | v_{qi}\right]$ and $\mathbb{E}[u_i | \mathbf{v}_i]$), nonparametric estimators such as the IDKE in YP and YLP seem preferable because they do not require any instruments.[7] In summary, the parametric setup in this paper is not only preferable practically (not requiring series expansion on $h_1\left(\widehat{v}_{qi}\right)$ and $h_2\left(\widehat{v}_{qi}\right)$ as in Kourtellos et al. (2022)), but also preferable theoretically when instruments are employed (given that nonparametric methods do not require instruments).

Because we are interested only in γ in this section, we can concentrate out β and κ in both approaches to get a concentrated objective function $S_n\left(\gamma\right)$. Specifically, for any γ, let Y, $\widehat{X}_\gamma$, and $\widehat{X}_\perp$ denote the matrices of stacked vectors y_i, $\widehat{\widetilde{\mathbf{x}}}_i' 1(q_i \leq \gamma)$, and $\widehat{\widetilde{\mathbf{x}}}_i' 1(q_i > \gamma)$ in both approaches. Then $S_n\left(\gamma\right)$ is the LS residual sum of squared errors from a regression of Y on $\widehat{X}_\gamma$ and $\widehat{X}_\perp$. Our two CF estimators for γ are the minimizers of the sum of squared errors:

$$\widehat{\gamma} = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \; S_n\left(\gamma\right).$$

There is an interval of γ, $[\widehat{\gamma}_-, \widehat{\gamma}_+)$, minimizing $S_n\left(\gamma\right)$. Following Yu (2012; 2015), we therefore take the mid-point of the interval as our estimator of γ because the mid-point $\frac{\widehat{\gamma}_- + \widehat{\gamma}_+}{2}$ is more efficient than the left-endpoint $\widehat{\gamma}_-$ in most cases.

---

[6]Boldea, Hall and Han (2012) and Hall, Han and Boldea (2012) also study the unstable reduced form in the context of structural change models and allow both fixed and shrinking (at potentially different rates from $n^{-\alpha}$) $\Pi_1 - \Pi_2$, but note that in structural change models, the threshold variable is the time index and so it cannot be endogenous (see also Perron and Yamamoto, 2014, for related discussions). Note also that although the form of $\widehat{\gamma}$'s asymptotic distribution will not change, the form of $\widehat{\beta}$'s asymptotic variance matrix critically depends on the formation of the first-stage regression.

[7]Of course, if $\dim\left(\mathbf{v}_i\right) < \dim\left(\mathbf{x}_i\right)$, CF-II suffers less curse of dimensionality because the IDKE needs to estimate $\mathbb{E}[y_i | x_i, q_i = \gamma \pm]$.

In the following two subsections, we will discuss the asymptotic properties of the CF-II estimator of γ above. Compared with usual threshold regression estimation such as that of Chan (1993) and Hansen (2000), the only new component in the objective function (23) is the presence of the generated regressors from the first step.

3.3. Asymptotic Theory for $\widehat{\gamma}$

First, define some further notation to ease the exposition. Let $e_i^0 = e_{1i}1(q_i \leq \gamma_0) + e_{2i}1(q_i > \gamma_0)$ with $e_{\ell i} = \sigma_\ell \left(u_i - \psi' \mathbf{v}_i\right)$. $a_n = n^{1-2\alpha}$, and $\bar{\delta}_n := \overline{\beta}_1 - \overline{\beta}_2 = \left(\delta_\beta', \delta_\kappa'\right)'$. Write $M = \mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}_i'\right]$, $M_\gamma = \mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}_i' 1(q_i \leq \gamma)\right]$, $M_0 = M_{\gamma_0}$, $D_\gamma = \mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}_i' | q_i = \gamma\right]$, $D_0 = D_{\gamma_0}$, $V_\gamma^- = \mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}_i' e_{1i}^2 | q_i = \gamma-\right]$, $V_\gamma^+ = \mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}_i' e_{2i}^2 | q_i = \gamma+\right]$, $V_0^- = V_{\gamma_0}^-$ and $V_0^+ = V_{\gamma_0}^+$.

We impose the following assumptions which somewhat strengthen Assumption C-II so notations such as w_i and $f(\cdot)$ used there apply here. Like the label "C-II," we use the label "II" for our assumptions here since the label "I" has been used for CF-I in Section SD.2 of the Supplementary Material.

Assumption II:

1. $\{w_i\}_{i=1}^n$ are strictly stationary and ergodic with ρ mixing coefficients that satisfy $\sum_{m=1}^\infty \rho_m^{1/2} < \infty$; $\gamma \in \Gamma = [\underline{\gamma}, \overline{\gamma}]$ with Γ being compact and γ_0 is in the interior of Γ.

2. $\mathbb{E}[\mathbf{v}_i | \mathcal{F}_{i-1}] = \mathbf{0}$, and $\mathbb{E}[u_i | \mathcal{F}_{i-1}, \mathbf{x}_i] = \mathbb{E}[u_i | \mathbf{v}_i] = \psi' \mathbf{v}_i$.

3. $\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_i'\right] > 0$, $\mathbb{E}\left[\|\mathbf{z}_i\|^4\right] < \infty$, and $\mathbb{E}\left[\|\mathbf{v}_i\|^4\right] < \infty$.

4. $\mathbb{E}\left[\|\check{\mathbf{x}}_i\|^4\right] < \infty$ and $\mathbb{E}\left[\|\check{\mathbf{x}}_i e_i^0\|^4\right] < \infty$.

5. For all $\gamma \in \Gamma$, $\mathbb{E}\left[\|\check{\mathbf{x}}_i\|^4 | q_i = \gamma\right] \leq C$ and $\mathbb{E}\left[\|\check{\mathbf{x}}_i e_i^0\|^4 | q_i = \gamma\right] \leq C$ for some $0 < C < \infty$.

6. For all $\gamma \in \Gamma, f(\gamma) \leq \overline{f} < \infty$.

7. $f(\gamma), D_\gamma$, and $V_\gamma^\pm$ are continuous at $\gamma = \gamma_0$.

8. $\delta_n = \left(\delta_\beta', \delta_\sigma\right)' = \left(c_\beta', c_\sigma\right)' n^{-\alpha} = cn^{-\alpha}$ with $c \neq \mathbf{0}$ and $0 < \alpha < 1/2$.

9. $\overline{c}' D_0 \overline{c} > 0, \overline{c}' V_0^\pm \overline{c} > 0$ and $f > 0$, where $\overline{c} = \begin{pmatrix} I_d & \mathbf{0} \\ \mathbf{0} & \psi \end{pmatrix} c$.

10. $M > M_\gamma > 0$ for all $\gamma \in \Gamma$.

Conditions II.1,8 strengthen C-II.1, II.2-4 strengthen C-II.2-4 when x_i includes 1, II.6,7,9,10 imply C-II.5 (especially, II.10 implies that Γ cannot include the boundary region of q's support), II.2,3 imply C-II.6 when Θ is compact, and II.9,10 play the role of (but are not limited to) C-II.7 in the linear endogeneity scenario. This is understandable because Assumption II is used to derive the asymptotic distribution of $\widehat{\gamma}$ while Assumption C-II is only used to prove its consistency (although in the nonlinear endogeneity scenario). We next provide more explanation on the conditions in Assumption II beyond those in Assumption C-II. The

assumptions on $\{w_i\}_{i=1}^n$ in II.1 imply the same assumptions on $\left\{\left(\mathbf{z}_i, \check{\mathbf{x}}_i, \mathbf{v}_i, u_i\right)\right\}_{i=1}^n$ because the latter is a function of the former. II.3 implies $\widehat{\Pi}$ is $O_p\left(n^{-1/2}\right)$. II.4-10 parallel Assumption 1.4-1.10 of CH, and we indicate the differences below. Note that II.3 implies $E\left[\left\|\check{\mathbf{x}}_i\right\|^4\right] < \infty$ in II.4, but to make our assumptions comparable with CH, we still state this explicitly in II.4. Note also that II.4,5 imply $\mathbb{E}\left[\left\|\mathbf{z}_i e_i^0\right\|^4\right] < \infty$, and $\sup_{\gamma \in \Gamma} \mathbb{E}\left[\left\|\mathbf{z}_i e_i^0\right\|^4 | q_i = \gamma\right] \le C$, which are used in our proof. Different from CH, II.7 allows regime-dependent heteroskedasticity. II.8 is the small threshold effect assumption in the original TR model (1), where the quantity c represents the threshold effect and among which c_β is the quantity required in CH; it implies small threshold effect assumptions in the augmented TR model defined by (22), where $\bar{c}$ in II.9 represents the threshold effect and is the quantity required in this paper. II.7,9 implicitly assume that D_γ and $V_\gamma^{\pm}$ are well defined in a neighborhood of γ_0. The assumption $\bar{c}'D_0\bar{c} > 0$ in II.9 excludes the continuous threshold regression (CTR) of Chan and Tsay (1998) (see also Hansen, 2017) in the augmented regression. When $c_\sigma = 0$, the CTR in the original regression implies the CTR in the augmented regression asymptotically. Specifically, $\bar{c}'D_0\bar{c} = \mathbb{E}\left[\left(\mathbf{x}'c_\beta\right)^2 | q = \gamma_0\right] = c_\beta'\mathbb{E}\left[\mathbf{xx}' | q = \gamma_0\right]c_\beta$, so the original CTR implies the augmented CTR. When $c_\sigma \neq 0$, however, even if the original regression is a CTR, the augmented regression need not be, given that $\bar{c}'D_0\bar{c} = c_\sigma^2\mathbb{E}\left[\left(\psi'\mathbf{v}\right)^2 | q = \gamma_0\right]$ is positive in general. Since the discontinuous threshold regression has more identification information than the CTR, the threshold effects in conditional variance can actually improve identification power. Also, $D_0 > 0$ implies that $\dim\left(\mathbf{z}_i\right) \ge \dim\left(\mathbf{x}_i\right) = d$. This is because the rank of

$$D_0 = \mathbb{E}\left[\left(\begin{array}{c} \mathbf{x}_i \\ \mathbf{v}_i \end{array}\right)\left(\mathbf{x}_i', \mathbf{v}_i'\right)\bigg| q_i = \gamma_0\right]$$

is the same as the rank of

$$\left(\begin{array}{cc} I_d & -I_d \\ \mathbf{0} & I_d \end{array}\right)D_0\left(\begin{array}{cc} I_d & \mathbf{0} \\ -I_d & I_d \end{array}\right) = \mathbb{E}\left[\left(\begin{array}{c} \mathbf{g}_i \\ \mathbf{v}_i \end{array}\right)\left(\mathbf{g}_i', \mathbf{v}_i'\right)\bigg| q_i = \gamma_0\right],$$

and thus $\mathbb{E}\left[\mathbf{g}_i\mathbf{g}_i' | q_i = \gamma_0\right] = \Pi'\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i' | q_i = \gamma_0\right]\Pi > 0$ implies $\dim(\mathbf{z}_i) \ge d$.

THEOREM 1. *Under Assumption II,*

$$a_n\left(\widehat{\gamma} - \gamma_0\right) \xrightarrow{d} \omega \cdot \zeta(\phi),$$

where

$$\omega = \frac{\bar{c}'V_0^-\bar{c}}{\left(\bar{c}'D_0\bar{c}\right)^2 f} \quad and \quad \zeta(\phi) = \underset{r}{\operatorname{argmax}} \begin{cases} W_1(-r) - \frac{|r|}{2}, & if \ r \le 0, \\ \sqrt{\phi}W_2(r) - \frac{|r|}{2}, & if \ r > 0, \end{cases}$$

with $\phi = \bar{c}'V_0^+\bar{c}/\bar{c}'V_0^-\bar{c}$ and $W_\ell(r)$, $\ell = 1, 2$, being two independent standard Wiener processes on $[0, \infty)$.

The density of $\zeta(\phi)$ is reported in Appendix B of Bai (1997). If $\delta_\sigma = 0$, then $\bar{c} = \begin{pmatrix} I_d \\ \mathbf{0} \end{pmatrix} c_\beta$ is simplified; we can replace the $\check{\mathbf{x}}_i'$ in both D_0 and $V_0^\pm$ by $\mathbf{x}_i'$ and $\bar{c}$ by c_β, so the endogeneity level ψ does not affect the asymptotic distribution of $\widehat{\gamma}$. From Section SD.2 of the Supplementary Material, the asymptotic distribution of the CF-I $\widehat{\gamma}$ takes the same form with ω and ϕ reinterpreted.

It is helpful to compare the efficiency of the two CF estimators with the LS estimator without endogeneity. There is no clear result unless some simplifications are assumed. If $\mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' e_{1i}^2 | q_i = \gamma_0 -\right] = \mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' | q_i = \gamma_0\right] \mathbb{E}\left[e_{1i}^2\right]$ and $\mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' e_{2i}^2 | q_i = \gamma_0 +\right] = \mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' | q_i = \gamma_0\right] \mathbb{E}\left[e_{2i}^2\right]$, then $\phi = \frac{\mathbb{E}\left[e_{2i}^2\right]}{\mathbb{E}\left[e_{1i}^2\right]}$ and $\omega = \frac{\mathbb{E}\left[e_{1i}^2\right]}{\bar{c}' D_0 \bar{c} f}$.

In LS, if $\mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i' u_i^2 | q_i = \gamma_0 \pm\right] = \mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i' | q_i = \gamma_0\right] \mathbb{E}\left[u_i^2\right]$, then $\phi = \frac{\sigma_2^2}{\sigma_1^2}$ and $\omega = \frac{\sigma_1^2 \mathbb{E}\left[u_i^2\right]}{c_\beta' \mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i' | q_i = \gamma_0\right] c_\beta}$. In Section SD.2 of the Supplementary Material, we show that it is hard to compare the efficiency of CF-I with LS. However, in CF-II, its value of $\phi = \frac{\sigma_2^2}{\sigma_1^2}$ is the same as that of LS, and its value of $\omega = \frac{\sigma_1^2 \mathbb{E}\left[(u_i - \psi'\mathbf{v}_i)^2\right]}{\bar{c}' D_0 \bar{c}' f}$ should be less than that of LS because $\mathbb{E}\left[(u_i - \psi'\mathbf{v}_i)^2\right] \leq \mathbb{E}\left[u_i^2\right]$ and

$$\bar{c}' D_0 \bar{c} = (c_\beta', c_\sigma \psi') \begin{pmatrix} \mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i' | q_i = \gamma_0\right] & \mathbb{E}\left[\mathbf{x}_i\mathbf{v}_i' | q_i = \gamma_0\right] \\ \mathbb{E}\left[\mathbf{v}_i\mathbf{x}_i' | q_i = \gamma_0\right] & \mathbb{E}\left[\mathbf{v}_i\mathbf{v}_i' | q_i = \gamma_0\right] \end{pmatrix} \begin{pmatrix} c_\beta \\ \psi c_\sigma \end{pmatrix}$$

should be larger than $c_\beta' \mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i' | q_i = \gamma_0\right] c_\beta$ given that the threshold information on the conditional variance is also explored (when $c_\sigma = 0$, they are the same). It is interesting here to observe that endogeneity actually provides useful information in CF-II (in the sense that when there is no endogeneity, that is, $\psi = \mathbf{0}$, both ϕ and ω of CF-II are the same as those of LS even without the simplifications at the beginning of this paragraph); the extra efficiency beyond LS originates from the extra variables $\mathbf{z}_i$.

Interestingly, extra randomness in the generated regressors would not affect the asymptotic distribution of $\widehat{\gamma}$. This is different from the results for usual regularly estimable parameters (see Theorem 2 or recall the Heckit model). From Yu (2012; 2015), this outcome arises because the extra randomness affects $\widehat{\gamma}$ only locally so the effect does not accumulate sufficiently to influence estimation of γ. From the proof of Theorem 1, when we replace the true regressors $\check{\mathbf{x}}_i'$ by $\widehat{\check{\mathbf{x}}}_i'$, an extra term $\widetilde{e}_i = \widehat{r}_i'\kappa_2$ is added to the original error term e_i^0, and this extra term is $O_p(n^{-1/2})$, where $\widehat{r}_i' = \mathbf{v}_i' - \widehat{\mathbf{v}}_i' = (\widehat{\Pi} - \Pi)' \mathbf{z}_i$. For regular parameters, a positive proportion of $\widetilde{e}_i$ would contribute; since $\sum_{i=1}^n \widetilde{e}_i \approx O_p(1)$, the effect of $\widetilde{e}_i$ would not disappear. For $\widehat{\gamma}$, however, only an order $O(1/a_n)$ portion of $\widetilde{e}_i$ would contribute; then, since $\sum_{i=1}^{n/a_n} \widetilde{e}_i \approx O_p(\sqrt{n/a_n \cdot 1/n}) = O_p(\sqrt{1/a_n}) = o_p(1)$, the effect of $\widetilde{e}_i$ disappears asymptotically.

3.4. Likelihood Ratio Tests

As a by-product of estimation, we obtain natural test statistics for hypotheses concerning γ such as $H_0 : \gamma = \gamma_0$. As emphasized in Hansen (2000), the following LR-like statistic has better finite-sample performance than a typical t-like statistic when the threshold effect is small:

$$LR_n(\gamma) = \frac{S_n(\gamma) - S_n(\widehat{\gamma})}{2\widehat{\eta}^2},$$

where $\widehat{\eta}^2$ is a consistent estimator of $\eta^2 = \overline{c}'V_0^-\overline{c}/\overline{c}'D_0\overline{c}$. Also, $LR_n(\gamma)$ avoids having to estimate f in the asymptotic distribution of $\widehat{\gamma}$, as is apparent in the following result (see also Theorem 2 in KST, noting the correction given in the proof of the corollary in Appendix A).

COROLLARY 1. *Under the assumptions in Theorem 1,*

$$LR_n(\gamma_0) \xrightarrow{d} \xi(\phi),$$

where $\xi(\phi) = \sup_r \begin{cases} W_1(-r) - \frac{|r|}{2}, & \text{if } r \leq 0, \\ \sqrt{\phi}W_2(r) - \frac{|r|}{2}, & \text{if } r > 0, \end{cases}$ *has the distribution* $P(\xi(\phi) \leq x) = (1 - e^{-x})(1 - e^{-x/\phi})$.

From Section SD.2 of the Supplementary Material, the CF-I LR statistic takes the same form and has the same asymptotic distribution with ϕ reinterpreted. If the model is homoskedastic, then $\phi = 1$ and $\widehat{\eta}^2$ in $LR_n(\gamma)$ can be replaced by an estimate of $\mathbb{E}[e_i^{02}]$, such as $S_n(\widehat{\gamma})/n$ as in CH. However, the model is generally heteroskedastic. For example,

$$V_0^- = \sigma_1^2 \mathbb{E}\left[\breve{\mathbf{x}}_i\breve{\mathbf{x}}_i'(u_i - \psi'\mathbf{v}_i)^2 | q_i = \gamma_0-\right],$$
$$V_0^+ = \sigma_2^2 \mathbb{E}\left[\breve{\mathbf{x}}_i\breve{\mathbf{x}}_i'(u_i - \psi'\mathbf{v}_i)^2 | q_i = \gamma_0+\right].$$

Even if $\mathbb{E}\left[\breve{\mathbf{x}}_i\breve{\mathbf{x}}_i'(u_i - \psi'\mathbf{v}_i)^2 | q_i = \gamma\right]$ is continuous at γ_0, $V_0^- \neq V_0^+$ if $\sigma_1 \neq \sigma_2$. In fact, even in CH's setup, V_0^- is likely to be unequal to V_0^+, as we now explain. In their setup, $e_i^0 = \beta_1'\mathbf{v}_i 1(q_i \leq \gamma_0) + \beta_2'\mathbf{v}_i 1(q_i > \gamma_0) + u_i$,

$$V_0^- = \mathbb{E}\left[\mathbf{g}_i\mathbf{g}_i'(e_i^0)^2 | q_i = \gamma_0-\right] = \mathbb{E}\left[\mathbf{g}_i\mathbf{g}_i'(\beta_1'\mathbf{v}_i + u_i)^2 | q_i = \gamma_0-\right],$$
$$V_0^+ = \mathbb{E}\left[\mathbf{g}_i\mathbf{g}_i'(e_i^0)^2 | q_i = \gamma_0+\right] = \mathbb{E}\left[\mathbf{g}_i\mathbf{g}_i'(\beta_2'\mathbf{v}_i + u_i)^2 | q_i = \gamma_0+\right],$$

where $\mathbf{v}$ does not include v_q because their $\mathbf{x}$ is just our x and their q is exogenous so is included in $\mathbf{z}$. Obviously, V_0^- is unlikely to equal V_0^+.

For inference based on $LR_n(\gamma)$, we need to estimate η^2 and ϕ. Following Hansen (2000), we can use the following procedure. Let $r_{1i} = (\overline{\delta}_n'\breve{\mathbf{x}}_i)^2 e_{1i}^2$, $r_{2i} = (\overline{\delta}_n'\breve{\mathbf{x}}_i)^2 e_{2i}^2$

and $r_{3i} = (\bar{\delta}'_n \check{\mathbf{x}}_i)^2$. Then

$$\eta^2 = \frac{\mathbb{E}[r_{1i}|q_i = \gamma_0]}{\mathbb{E}[r_{3i}|q_i = \gamma_0]} \text{ and } \phi = \frac{\mathbb{E}[r_{2i}|q_i = \gamma_0]}{\mathbb{E}[r_{1i}|q_i = \gamma_0]}$$

are ratios of two conditional expectations. Since r_{1i}, r_{2i} and r_{3i} are unobserved, let $\widehat{r}_{1i} = (\widehat{\bar{\delta}}'_n \widehat{\mathbf{x}}_i)^2 \widehat{e}^2_{1i}, \widehat{r}_{2i} = (\widehat{\bar{\delta}}'_n \widehat{\mathbf{x}}_i)^2 \widehat{e}^2_{2i}$ and $\widehat{r}_{3i} = (\widehat{\bar{\delta}}'_n \widehat{\mathbf{x}}_i)^2$ denote their sample counterparts, where $\widehat{e}_{1i} = y_i - \widehat{\mathbf{x}}'_i \widehat{\bar{\beta}}_1$ is defined only for $i \in \mathcal{I}_1 := \{i | q_i \leq \widehat{\gamma}\}$, and $\widehat{e}_{2i} = y_i - \widehat{\mathbf{x}}'_i \widehat{\bar{\beta}}_2$ is defined only for $i \in \mathcal{I}_2 := \{i | q_i > \widehat{\gamma}\}$ with some consistent estimators of $\bar{\beta}_\ell$ in Section 4. We can then estimate η^2 and ϕ by kernel regression, for example,

$$\widehat{\eta}^2 = \frac{\sum K^-_h (q_i - \widehat{\gamma}) \widehat{r}_{1i}}{\sum K_h (q_i - \widehat{\gamma}) \widehat{r}_{3i}} \text{ and } \widehat{\phi} = \frac{\sum K^+_h (q_i - \widehat{\gamma}) \widehat{r}_{2i}}{\sum K^-_h (q_i - \widehat{\gamma}) \widehat{r}_{1i}},$$

where $K^\pm_h(\cdot) = h^{-1} K^\pm(\cdot/h)$ and $K_h(\cdot) = h^{-1} K(\cdot/h)$ for some bandwidth h and (boundary) kernel functions $K^\pm(\cdot)$ and $K(\cdot)$. Alternatively, we can use a polynomial regression, such as a quadratic by fitting OLS regressions of the form

$$\widehat{r}_{1i} = \widehat{\mu}_{10} + \widehat{\mu}_{11} q_i + \widehat{\mu}_{12} q_i^2 + \widehat{\varepsilon}_{1i}, i \in \mathcal{I}_1,$$
$$\widehat{r}_{2i} = \widehat{\mu}_{20} + \widehat{\mu}_{21} q_i + \widehat{\mu}_{22} q_i^2 + \widehat{\varepsilon}_{2i}, i \in \mathcal{I}_2,$$
$$\widehat{r}_{3i} = \widehat{\mu}_{30} + \widehat{\mu}_{31} q_i + \widehat{\mu}_{32} q_i^2 + \widehat{\varepsilon}_{3i}, i \in \mathcal{I}_1 \cup \mathcal{I}_2 = \{1, 2, \ldots, n\},$$

and setting

$$\widehat{\eta}^2 = \frac{\widehat{\mu}_{10} + \widehat{\mu}_{11} \widehat{\gamma} + \widehat{\mu}_{12} \widehat{\gamma}^2}{\widehat{\mu}_{30} + \widehat{\mu}_{31} \widehat{\gamma} + \widehat{\mu}_{32} \widehat{\gamma}^2} \text{ and } \widehat{\phi} = \frac{\widehat{\mu}_{20} + \widehat{\mu}_{21} \widehat{\gamma} + \widehat{\mu}_{22} \widehat{\gamma}^2}{\widehat{\mu}_{10} + \widehat{\mu}_{11} \widehat{\gamma} + \widehat{\mu}_{12} \widehat{\gamma}^2}.$$

If $\mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}'_i (u_i - \psi' \mathbf{v}_i)^2 | q_i = \gamma\right]$ is continuous at γ_0, then $\phi = \sigma_2^2/\sigma_1^2$ can be estimated in a simpler way. For example, if the homoskedasticity assumption $\mathbb{E}\left[(u_i - \psi' \mathbf{v}_i)^2 | q_i\right] = \mathbb{E}\left[(u_i - \psi' \mathbf{v}_i)^2\right]$ holds, then a suitable estimate is $\widehat{\phi} = \widehat{\sigma}_2^2/\widehat{\sigma}_1^2$ where $\widehat{\sigma}_\ell^2 = |\mathcal{I}_\ell|^{-1} \sum_{i \in \mathcal{I}_\ell} \widehat{e}^2_{\ell i}$ is a consistent estimator of $\sigma_\ell^2 \mathbb{E}\left[(u_i - \psi' \mathbf{v}_i)^2\right]$. If $\mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}'_i (u_i - \psi' \mathbf{v}_i)^2 | q_i = \gamma_0 -\right] = \mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}'_i | q_i = \gamma_0\right] \mathbb{E}\left[(u_i - \psi' \mathbf{v}_i)^2\right]$, which holds under the homoskedasticity assumption $Var(u|\mathbf{x}, \mathbf{z}) = Var(u|\mathbf{v}) = \mathbb{E}[Var(u|\mathbf{v})]$ because

$$\mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}'_i (u_i - \psi' \mathbf{v}_i)^2 | q_i\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}'_i (u_i - \psi' \mathbf{v}_i)^2 | \mathbf{x}_i, \mathbf{z}_i\right] | q_i\right] = \mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}'_i Var(u_i | \mathbf{v}_i) | q_i\right]$$
$$= \mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}'_i | q_i\right] \mathbb{E}[Var(u_i | \mathbf{v}_i)] = \mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}'_i | q_i\right] \mathbb{E}\left[(u_i - \psi' \mathbf{v}_i)^2\right],$$

then $\eta^2 = \sigma_1^2 \mathbb{E}[(u_i - \psi' \mathbf{v}_i)^2]$ can be estimated by $\widehat{\sigma}_1^2$ under the above homoskedasticity assumption.

Given these components, the $(1 - \alpha)$ LR-CI for γ follows by inversion of the statistic from

$$\{\gamma : LR_n(\gamma) \leq \widehat{\mathrm{crit}}\},$$

where $\widehat{\mathrm{crit}}$ is the $(1 - \alpha)$ quantile of ξ with ϕ being replaced by $\widehat{\phi}$.

4. ESTIMATION OF SLOPE PARAMETERS

Having obtained $\widehat{\gamma}$, we propose two approaches for estimating β. The first is to estimate β as a by-product of the two CF approaches above, that is, we estimate β by $\widehat{\beta} = (\widehat{\beta}_1(\widehat{\gamma})', \widehat{\beta}_2(\widehat{\gamma})')'$, where $\widehat{\beta}(\gamma) = (\widehat{\beta}_1(\gamma)', \widehat{\beta}_2(\gamma)')'$ is the concentrated estimator of β given γ. We label these two estimation methods of β (with either the CF-I estimator of γ or the CF-II estimator of γ as $\widehat{\gamma}$) as CF-I and CF-II, and jointly as CF. The second approach is to estimate β by setting up moment conditions as in CH and KST. However, we show in Section SD.3 of the Supplementary Material that the estimates based on the moment conditions in CH and KST are in general inconsistent when q is endogenous. We therefore provide two new sets of moment conditions that generate consistent estimates of β. The resulting estimation methods of β are labeled GMM-1 and GMM-2, respectively; when the CF-I estimator of γ is used as $\widehat{\gamma}$, the corresponding estimation methods of β using these two sets of moment conditions are labeled as GMM-I1 and GMM-I2, and when the CF-II estimator of γ is used as $\widehat{\gamma}$, labeled as GMM-II1 and GMM-II2, respectively. Since GMM-II2 is the most efficient among all methods in our simulation studies, we will discuss CF/GMM-1 and GMM-II1 only briefly in the main text and leave the detailed discussions concerning asymptotic properties to Sections SD.3 and SD.2 of the Supplementary Material, respectively.

4.1. CF and GMM-1 Estimators

In the CF estimators, we estimate $\overline{\beta}_\ell$ by

$$\widehat{\overline{\beta}}_\ell = (\widehat{X}'_\ell \widehat{X}_\ell)^{-1} \widehat{X}'_\ell Y, \tag{24}$$

and estimate $\overline{\delta}_n$ by $\widehat{\overline{\delta}}_n = \widehat{\overline{\beta}}_1 - \widehat{\overline{\beta}}_2$, where $\widehat{X}_1 = \widehat{X}_{\widehat{\gamma}}$, and $\widehat{X}_2$ is $\widehat{X}_\perp$ with γ replaced by $\widehat{\gamma}$. Given $\widehat{\overline{\beta}}_\ell$, we can estimate β_ℓ by extracting the corresponding components of $\widehat{\overline{\beta}}_\ell$, say, $\widehat{\beta}_\ell$, and estimate δ_β by $\widehat{\beta}_1 - \widehat{\beta}_2$. This type of estimate essentially employs the moment conditions

$$\mathbb{E}\left[\check{\mathbf{x}}_i e_{1i} 1(q_i \leq \gamma_0)\right] = \mathbf{0} \text{ and } \mathbb{E}\left[\check{\mathbf{x}}_i e_{2i} 1(q_i > \gamma_0)\right] = \mathbf{0} \tag{25}$$

to estimate β_1 and β_2 separately.

Before introducing the GMM-1 estimators, we first show in Section SD.3 of the Supplementary Material that the moment conditions in CH (the resulting GMM

is labeled as CH's GMM) and KST will not generate consistent estimates of β in general when q is endogenous. Although the moment conditions used in CH,

$$\mathbb{E}\left[\mathbf{z}_i u_i 1(q_i \leq \gamma_0)\right] = \mathbf{0} \text{ and } \mathbb{E}\left[\mathbf{z}_i u_i 1(q_i > \gamma_0)\right] = \mathbf{0}, \tag{26}$$

may not hold, $\mathbb{E}\left[\mathbf{z}_i u_i\right] = \mathbf{0}$ seems reasonable, where we assume $\delta_\sigma = 0$ and absorb σ_2 in u as mentioned in the Section 1. As a result, we can estimate β by the following minimization

$$\widehat{\beta} = \operatorname*{argmin}_{\beta} n \cdot \overline{m}_n (\beta)' W_n \overline{m}_n (\beta),$$

where $W_n \xrightarrow{p} W > 0$, and

$$\overline{m}_n (\beta) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i \left(y_i - \mathbf{x}'_{i, \leq \widehat{\gamma}} \beta_1 - \mathbf{x}'_{i, > \widehat{\gamma}} \beta_2\right),$$

in which $\widehat{\gamma}$ is the CF-I/CF-II estimator of γ in GMM-I1/GMM-II1. In other words, the GMM-1 estimator of β is

$$\widehat{\beta} = \left[\left(X^{*\prime} Z\right) W_n \left(Z' X^*\right)\right]^{-1} \left[\left(X^{*\prime} Z\right) W_n \left(Z' Y\right)\right],$$

where X^* stacks $(\mathbf{x}'_{i, \leq \widehat{\gamma}}, \mathbf{x}'_{i, > \widehat{\gamma}})$, and Z stacks $\mathbf{z}'_i$.

Compared with the CF estimators of β, the GMM-1 estimates of β suffer two drawbacks. First, the CF estimates require only d instruments, while the GMM-1 estimates require $2d$ instruments; $2d$ instruments are required because we estimate β_1 and β_2 jointly. When q is exogenous, we can estimate β_1 and β_2 separately as in CH's GMM, in which case only d instruments are required. It is well known that good instruments are often hard to find and justify in practical work, making this a relevant consideration. Second, the GMM-1 estimates are hard to extend to the $\delta_\sigma \neq 0$ case although not impossible (see Section SD.3 of the Supplementary Material for detailed analysis).

4.2. GMM-2 Estimators

The two drawbacks of GMM-1 estimators of β above originate from the fact that they estimate β_1 and β_2 jointly. To avoid these drawbacks, we can estimate β_1 and β_2 separately as in CH's GMM. Specifically, we use the following moment conditions to estimate β_1 and β_2:

$$\mathbb{E}\left[\check{\mathbf{z}}_i e_{1i} 1(q_i \leq \gamma_0)\right] = \mathbf{0} \text{ and } \mathbb{E}\left[\check{\mathbf{z}}_i e_{2i} 1(q_i > \gamma_0)\right] = \mathbf{0}, \tag{27}$$

where

$$\check{\mathbf{z}}_i = \left(\mathbf{z}'_i, q_i\right)' \text{ in CF-I and } \check{\mathbf{z}}_i = \left(\mathbf{z}'_i, \mathbf{x}'_i\right)' \text{ in CF-II.}$$

When some elements of $\mathbf{x}_i$ are exogenous, then they can be included in $\mathbf{z}_i$ and $\mathbf{x}_i$ is correspondingly shortened. The moment conditions (27) hold because of the analysis in Section 3, especially (20) and (22). Although we can use any function of $\check{\mathbf{z}}_i$ as instruments, we do not pursue this particular moment selection problem

in the current paper. Interestingly, although q_i and $\mathbf{x}_i$ may be endogenous, they are used as instruments because the error terms here are $e_{\ell i}$ rather than u_i. Actually, the CF estimators of β are special cases of this kind of GMM estimator as is evident by noting that

$$\check{\mathbf{x}}_i = \begin{pmatrix} \Pi' & 0 \\ -\pi' & 1 \end{pmatrix} \check{\mathbf{z}}_i \text{ in CF-I and } \check{\mathbf{x}}_i = \begin{pmatrix} \mathbf{0} & I_d \\ -\Pi' & I_d \end{pmatrix} \check{\mathbf{z}}_i \text{ in CF-II.}$$

We list the CF estimators of β separately because they are natural byproducts of CF estimation of γ.

Solving the minimization problem

$$\min_{\overline{\beta}_\ell} n \cdot \overline{m}_{\ell n} \left(\overline{\beta}_\ell \right)' W_{\ell n} \overline{m}_{\ell n} \left(\overline{\beta}_\ell \right),$$

we have

$$\widehat{\overline{\beta}}_\ell = \left[\left(\widehat{X}'_\ell \widehat{Z}_\ell \right) W_{\ell n} \left(\widehat{Z}'_\ell \widehat{X}_\ell \right) \right]^{-1} \left[\left(\widehat{X}'_\ell \widehat{Z}_\ell \right) W_{\ell n} \left(\widehat{Z}'_\ell Y \right) \right],$$

where $W_{\ell n} \xrightarrow{p} W_\ell > 0$, $\widehat{X}_\ell$ is defined in (24), $\widehat{Z}_1$ and $\widehat{Z}_2$ are the matrices obtained by stacking $\check{\mathbf{z}}'_i 1(q_i \leq \widehat{\gamma})$ and $\check{\mathbf{z}}'_i 1(q_i > \widehat{\gamma})$, and

$$\overline{m}_{1n} \left(\overline{\beta}_1 \right) = \frac{1}{n} \sum_{i=1}^n \check{\mathbf{z}}_i \left(y_i - \widehat{\check{\mathbf{x}}}'_i \overline{\beta}_1 \right) 1(q_i \leq \widehat{\gamma}),$$

$$\overline{m}_{2n} \left(\overline{\beta}_2 \right) = \frac{1}{n} \sum_{i=1}^n \check{\mathbf{z}}_i \left(y_i - \widehat{\check{\mathbf{x}}}'_i \overline{\beta}_2 \right) 1(q_i > \widehat{\gamma}).$$

It might seem that GMM-I2 is an extension of CH's GMM method. However, even if q is exogenous, they are different because $e_{\ell i}$ rather than u_i are used as the error terms in (27). The limit theory of GMM-II2 is given in the following result.

THEOREM 2. *Under Assumption II,*

$$n^{1/2}(\widehat{\overline{\beta}}_\ell - \overline{\beta}_\ell) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma_\ell),$$

where

$$\Sigma_\ell = (G'_\ell W_\ell G_\ell)^{-1} (G'_\ell W_\ell \Omega_\ell W_\ell G_\ell)(G'_\ell W_\ell G_\ell)^{-1}.$$

In Σ_ℓ, $G_1 = \mathbb{E}[\check{\mathbf{z}}_i \check{\mathbf{x}}'_{i, \leq \gamma_0}]$, $G_2 = \mathbb{E}[\check{\mathbf{z}}_i \check{\mathbf{x}}'_{i, > \gamma_0}]$, $G'_\ell W_\ell G_\ell > 0$, and

$$\Omega_\ell = \Omega^\ell_1 + \Omega^\ell_2$$

with

$$\Omega^1_1 = \mathbb{E}[\check{\mathbf{z}}_i \check{\mathbf{z}}'_{i, \leq \gamma_0} e^2_{1i}] > 0, \Omega^2_1 = \mathbb{E}\left[\check{\mathbf{z}}_i \check{\mathbf{z}}'_{i, > \gamma_0} e^2_{2i} \right] > 0,$$

$$\Omega^1_2 = \mathbb{E}\left[\check{\mathbf{z}}_i \mathbf{z}'_{i, \leq \gamma_0} \right] \mathbb{E}\left[\mathbf{z}_i \mathbf{z}'_i \right]^{-1} \mathbb{E}\left[\mathbf{z}_i \mathbf{z}'_i \left(\kappa'_1 \mathbf{v}_i \mathbf{v}'_i \kappa_1 \right) \right] \mathbb{E}\left[\mathbf{z}_i \mathbf{z}'_i \right]^{-1} \mathbb{E}\left[\mathbf{z}_i \check{\mathbf{z}}'_{i, \leq \gamma_0} \right],$$

$$\Omega^2_2 = \mathbb{E}\left[\check{\mathbf{z}}_i \mathbf{z}'_{i, > \gamma_0} \right] \mathbb{E}\left[\mathbf{z}_i \mathbf{z}'_i \right]^{-1} \mathbb{E}\left[\mathbf{z}_i \mathbf{z}'_i \left(\kappa'_2 \mathbf{v}_i \mathbf{v}'_i \kappa_2 \right) \right] \mathbb{E}\left[\mathbf{z}_i \mathbf{z}'_i \right]^{-1} \mathbb{E}\left[\mathbf{z}_i \check{\mathbf{z}}'_{i, > \gamma_0} \right].$$

When $W_\ell = \Omega_\ell^{-1}$, Σ_ℓ reduces to $(G_\ell'\Omega_\ell^{-1}G_\ell)^{-1}$. The asymptotic covariance matrix between $n^{1/2}(\widehat{\overline{\beta}}_1 - \overline{\beta}_1)$ and $n^{1/2}(\widehat{\overline{\beta}}_2 - \overline{\beta}_2)$ is

$$C_{GMM} = (G_1'W_1G_1)^{-1}G_1'W_1\Omega_{12}W_2G_2(G_2'W_2G_2)^{-1}$$

with $\Omega_{12} = \mathbb{E}\left[\check{z}_i z_{i,\leq\gamma_0}'\right]\mathbb{E}\left[z_iz_i'\right]^{-1}\mathbb{E}\left[z_iz_i'(\kappa_1'v_iv_i'\kappa_2)\right]\mathbb{E}\left[z_iz_i'\right]^{-1}\mathbb{E}\left[z_i\check{z}_{i,>\gamma_0}'\right]$. The $\mathcal{N}(\mathbf{0}, \Sigma_\ell)$ limit distribution is independent of $\zeta(\phi)$ in Theorem 1.

Note that the components Ω_2^ℓ stem from contributions by the generated regressors. If the first-stage regression is homoskedastic, that is, $Var(v_i|z_i) = Q$, then the Ω_2^ℓ simplify as follows:

$$\Omega_2^1 = \kappa_1'Q\kappa_1 \cdot \mathbb{E}\left[\check{z}_i z_{i,\leq\gamma_0}'\right]\mathbb{E}\left[z_iz_i'\right]^{-1}\mathbb{E}\left[z_i\check{z}_{i,\leq\gamma_0}'\right],$$

$$\Omega_2^2 = \kappa_2'Q\kappa_2 \cdot \mathbb{E}\left[\check{z}_i z_{i,>\gamma_0}'\right]\mathbb{E}\left[z_iz_i'\right]^{-1}\mathbb{E}\left[z_i\check{z}_{i,>\gamma_0}'\right];$$

and, similarly, the limit covariance matrix structure is

$$C_{GMM} = \kappa_1'Q\kappa_2 \cdot (G_1'W_1G_1)^{-1}G_1'W_1\mathbb{E}\left[\check{z}_i z_{i,\leq\gamma_0}'\right]\mathbb{E}\left[z_iz_i'\right]^{-1}\mathbb{E}\left[z_i\check{z}_{i,>\gamma_0}'\right]$$
$$\times W_2G_2(G_2'W_2G_2)^{-1}.$$

Note that the extra randomness from the generated regressors is not correlated with the original error term. Take $\widehat{\overline{\beta}}_1$ as an example: the random component of the influence function from the original error term is $\check{z}_{i,\leq\gamma_0}e_{1i}$ and that from the generated regressors is $z_iv_i'\kappa_1$. Obviously,

$$\mathbb{E}\left[\check{z}_{i,\leq\gamma_0}z_i'e_{1i}v_i'\kappa_1\right] = \mathbb{E}\left[\mathbb{E}\left[\check{z}_{i,\leq\gamma_0}z_i'e_{1i}v_i'\kappa_1 \,\middle|\, x_i, z_i\right]\right]$$
$$= \sigma_1\mathbb{E}\left[\check{z}_{i,\leq\gamma_0}z_i'v_i'\kappa_1\mathbb{E}\left[u_i - \psi'v_i \,\middle|\, v_i\right]\right] = \mathbf{0}.$$

In Section SD.2 of the Supplementary Material, we show that the asymptotic variance matrix of GMM-I2 takes a similar form as in GMM-II2 but there are two extra terms in Ω_ℓ beyond Ω_1^ℓ and Ω_2^ℓ; this is because the two random components in the influence functions are correlated now. Although $\Omega_1^\ell > 0$ is assumed, Ω_2^ℓ is not positive definite because $\dim(z_i) < \dim(\check{z}_i)$. Of course, $\Omega_\ell = \Omega_1^\ell + \Omega_2^\ell > 0$ and so $\Sigma_\ell > 0$; this is because $G_\ell'W_\ell G_\ell > 0$ and $W_\ell > 0$ imply G_ℓ is full column rank (necessarily, $\dim(\check{z}_i) \geq \dim(\check{x}_i)$ or $\dim(z_i) \geq d$). Unlike LS estimation, $\widehat{\overline{\beta}}_1$ and $\widehat{\overline{\beta}}_2$ are not asymptotically independent because the generated regressors $\widehat{v}_i$ involve all data points; this is why $C_{GMM} \neq \mathbf{0}$. As a result, the asymptotic variance matrix of $\widehat{\overline{\delta}}_n = \widehat{\overline{\beta}}_1 - \widehat{\overline{\beta}}_2$ is $\Sigma_1 + \Sigma_2 - C_{GMM} - C_{GMM}'$ rather than $\Sigma_1 + \Sigma_2$ as in LS estimation.

As for estimation of Σ_ℓ, we just replace the population mean by the sample mean, population random vectors by their sample analogs (i.e., $\check{x}_i$ by $\widehat{\check{x}}_i$, v_i by $\widehat{v}_i$, and $e_{\ell i}$ by $\widehat{e}_{\ell i}$), and the population parameters by their estimates (i.e., θ_0 by $\widehat{\theta}$). In practice, we can set $W_{\ell n} = \widehat{Z}_\ell'\widehat{Z}_\ell$ to get an initial estimator of $\overline{\beta}_\ell$, and then set $W_{\ell n}$ as a consistent estimator of Ω_ℓ^{-1} to obtain the optimal estimator. Given $\widehat{\Sigma}_\ell$, we can construct CIs for each component of $\underline{\theta}$ by inverting the t-statistic; we can

also test some interesting hypotheses, for example, by testing whether $\kappa_1 = \kappa_2 = \mathbf{0}$ in CF-II to check whether there is endogeneity (such a test cannot be applied to CF-I because whether or not $\kappa_\ell = \beta_\ell' \varphi + \sigma_\ell \kappa = 0$ does not indicate the existence of endogeneity when $\mathbf{z}_i$ does not contain $\mathbf{x}_i$ in the first-stage equation). Also, the usual nonparametric bootstrap should be valid for inference on $\underline{\theta}$ although it is questionable for γ (see Yu, 2014, for further discussion on this point).

As distinct from GMM-1, only d instruments (both included and excluded) are needed in GMM-2 estimation irrespective of whether q is endogenous. On the other hand, the asymptotic distributions of GMM-2 rely on first-stage specification such as stable or unstable reduced forms, which is similar to CF but different from GMM-1. To be specific, both CF and GMM-2 require $\widehat{\mathbf{v}}_i$ which is generated from the first-stage estimation, while GMM-1 relies on moment conditions $\mathbb{E}[\mathbf{z}_i u_i] = \mathbf{0}$ that do not involve $\mathbf{v}_i$; this is an advantage of GMM-1. Of course, we can use the nonparametric bootstrap to conduct inferences on $\underline{\theta}$ to avoid explicit formulas of the asymptotic variances of $\widehat{\underline{\theta}}$.

Because GMM-2 employs more moment conditions than CF, the resulting estimators are expected to be more efficient than the CF estimators.[8] However, it is hard to compare the efficiency of CF and GMM-2 with GMM-1. This is because, as distinct from GMM-1, CF and GMM-2 take advantage of the triangular structure of the simultaneous equations and also restrictions such as $\mathbb{E}[\mathbf{v}|v_q] = \varphi v_q$ and $\mathbb{E}[u|v_q] = \kappa v_q$ in CF-I and GMM-I2 and $\mathbb{E}[u|\mathbf{x}, \mathbf{z}, q] = \psi' \mathbf{v}$ in CF-II and GMM-II2. Roughly speaking, GMM-1 explores a different part of model information from that of CF and GMM-2. In summary, it is fair to claim that GMM-2 is the most preferable method to practitioners because GMM-1 suffers from the two drawbacks discussed above and CF is less efficient. Between GMM-I2 and GMM-II2, GMM-II2 seems more attractive because it usually uses a more efficient $\widehat{\gamma}$ (i.e., CF-II) and more instruments than GMM-I2 (i.e., $\dim(\mathbf{x}_i) \geq \dim(q_i)$). Of course, GMM-II2 usually estimates more parameters.

Finally, the asymptotic independence between $n^{1/2}\left(\widehat{\underline{\theta}} - \underline{\theta}\right)$ and $a_n(\widehat{\gamma} - \gamma_0)$ occurs because the global information explored in the former and the local information explored in the latter are independent (see Yu (2012; 2015) for more discussion on this point). This information independence also implies that $\widehat{\gamma}$ will not affect the asymptotic variance of $\widehat{\underline{\theta}}$ (as verified in the theorem) although $\widehat{\gamma}$ need not be $o_p\left(n^{-1/2}\right)$.

5. SIMULATIONS

This section reports the results of three simulation experiments conducted to assess the performance of the various estimates of γ and β and their corresponding CIs when q is endogenous and instruments are employed. Simulation 1 seeks to reveal the bias of the STR estimator of KST in finite samples when endogeneity

---

[8]Because generated regressors are involved, we can add the same group of moment conditions to (25) and (27), as in Newey (1984) and Murphy and Topel (1985), to see that GMM-2 indeed employs more moment conditions than CF.

is severe, and to compare the 2SLS estimator of YLP and the two CF estimators in this paper for γ. Under the DGP of this simulation, the two CF estimators are the same; also, the efficient GMM estimator of γ in YLP is the 2SLS estimator. Simulation 2 compares the two CF estimators of γ under a DGP for which they differ. Simulation 3 compares the performances of seven estimators of $\underline{\theta}$, viz., the 2SLS estimator of YLP, the two CF estimators in Section 4.1, and the four GMM estimators in Sections 4.1 and 4.2. Similar to Simulation 1, the efficient GMM estimator of $\underline{\theta}$ in YLP is the 2SLS estimator. In Section SD.5 of the Supplementary Material, we also report the risk of the integrated difference kernel estimator (IDKE) of γ in YP where no instruments are employed and compare the performance of the two CF approaches when q is exogenous.

Simulation 1 uses the following simple setup which generates the same DGP as in (11),

$$y_i = \delta 1 (q_i \leq \gamma_0) + u_i,$$
$$q_i = \pi' \mathbf{z}_i + v_{qi},$$

where $\mathbf{z}_i = (1, z_i)'$, $u_i = \kappa v_{qi} + e_{ui}$, $(z_i, v_{qi}, e_{ui})' \sim \mathcal{N}(\mathbf{0}, I_3)$, $\beta_2 = 0$ is known, $\gamma_0 = 0$, and $\pi = (0, -1)'$. Since $\mathbf{x} = 1$ is exogenous, the two CF approaches are exactly the same. In Simulations 2 and 3,

$$y_i = x_i \delta 1 (q_i \leq \gamma_0) + u_i,$$
$$x_i = \Pi'_x \mathbf{z}_i + v_{xi},$$
$$q_i = \pi' \mathbf{z}_i + v_{qi},$$

where $\mathbf{z}_i = (1, z_i)'$, $u_i = \psi_x v_{xi} + \psi_q v_{qi} + e_{ui} =: \psi' \mathbf{v}_i + e_{ui}$, $v_{xi} = \varphi v_{qi} + e_{xi}$, $(z_i, v_{qi}, e_{ui}, e_{xi})' \sim \mathcal{N}(\mathbf{0}, I_4)$, $\beta_2 = 0$ is known, $\gamma_0 = 0$, $\psi = (\psi_x, \psi_q)' = \kappa(1, 1)'$, $\varphi = 1$ and $\pi = \Pi_x = (0, -1)'$. We label these two DGPs as DGP1 and DGP2, respectively; in both DGPs, the data are randomly sampled. In DGP1, $\check{\mathbf{x}}_i = (1, v_{qi})'$, $\kappa_1 = \kappa_2 = \kappa$, and $e_{1i} = e_{2i} = e_{ui} =: e_i^0$ for both CF-I and CF-II. In DGP2, $\check{\mathbf{x}}_i = (\Pi'_x \mathbf{z}_i, v_{qi})'$, $\kappa_1 = (\delta + \psi_x) \varphi + \psi_q$, $\kappa_2 = \psi_x \varphi + \psi_q$, $\delta_\kappa = \delta \varphi$, $e_{1i} = (\delta + \psi_x) e_{xi} + e_{ui}$, and $e_{2i} = \psi_x e_{xi} + e_{ui}$ if CF-I is employed, and $\check{\mathbf{x}}_i = (x_i, v_{xi}, v_{qi})'$, $\kappa_1 = \kappa_2 = \psi$, $\delta_\kappa = 0$, and $e_{1i} = e_{2i} = e_{ui} =: e_i^0$ if CF-II is employed; $\check{\mathbf{z}}_i = (\mathbf{z}'_i, q_i)'$ in GMM-I2 and $\check{\mathbf{z}}_i = (\mathbf{z}'_i, x_i, q_i)'$ in GMM-II2.

In both DGPs, we consider three δ values, 0.5, 1, and 2, indicating small, medium, and large threshold effects, and two κ values, 0.2δ and δ, indicating weak and strong endogeneity, respectively. The parameter space Γ is taken as the interval between the 5% and 95% quantiles of q. We consider $N = 1,000$ replications with sample size $n = 200$ and 800 for all simulations. All CIs are constructed using the asymptotic critical values except YLP's 2SLS where the nonparametric bootstrap is applied since the asymptotic variance matrix involves density as in quantile regression. When the nonparametric bootstrap is used, the resampling times are $B = 399$ when $n = 200$, and $B = 199$ when $n = 800$.

5.1. Comparison of Three Estimators of γ

For the 2SLS estimator of YLP, we use the moment conditions $\mathbb{E}\left[\mathbf{z}_i\left(y_i - \delta 1\left(q_i \leq \gamma_0\right)\right)\right] = \mathbf{0}$ to estimate δ and γ_0 jointly. From YLP,

$$\begin{pmatrix} n^{1/2} & 0 \\ 0 & n^{1/2-\alpha} \end{pmatrix} \begin{pmatrix} \widehat{\delta} - \delta \\ \widehat{\gamma} - \gamma_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbb{E}\left[u^2\right] G^{-1} W^{-1} G'^{-1}\right), \tag{28}$$

where $\mathbb{E}\left[u^2\right] = \kappa^2 + 1$,

$$G = \left(\mathbb{E}\left[\mathbf{z}_{i,\leq\gamma_0}\right], \mathbb{E}\left[\mathbf{z}_i|q_i = \gamma_0\right] c_\beta f\right) \text{ with } c_\beta = n^\alpha \delta \text{ and } W = \mathbb{E}\left[\mathbf{z}_i \mathbf{z}_i'\right]^{-1}.$$

In KST,

$$\begin{aligned} y_i &= \left[\delta + \kappa \cdot \lambda_1\left(\gamma_0 - \pi'\mathbf{z}_i\right)\right] 1\left(q_i \leq \gamma_0\right) + \kappa \cdot \lambda_2\left(\gamma_0 - \pi'\mathbf{z}_i\right) 1\left(q_i > \gamma_0\right) + e_i^* \\ &= \delta \cdot 1\left(q_i \leq \gamma_0\right) + \kappa \Lambda_i\left(\gamma_0\right) + e^*, \end{aligned}$$

where $\Lambda_i\left(\gamma\right) := \lambda_1\left(\gamma - \pi'\mathbf{z}_i\right) 1\left(q_i \leq \gamma\right) + \lambda_2\left(\gamma - \pi'\mathbf{z}_i\right) 1\left(q_i > \gamma\right)$, and in our CF approaches,

$$\begin{aligned} y_i &= \left[\delta + \kappa v_{qi}\right] 1\left(q_i \leq \gamma_0\right) + \kappa v_{qi} 1\left(q_i > \gamma_0\right) + e_{ui} \\ &= \delta \cdot 1\left(q_i \leq \gamma_0\right) + \kappa v_{qi} + e_i^0. \end{aligned}$$

From Section 2, the limit STR objective function is symmetric and its minimizer is not unique, so we use the mean absolute deviation (MAD) as a measure of risk. In our two CF estimators, ϕ and η^2 can be simplified as follows:

$$\phi = 1 \text{ and } \eta^2 = \mathbb{E}\left[e_{ui}^2\right] = 1.$$

In the STR estimator of KST, we set $\phi = 1$ and $\eta^2 = \mathbb{E}\left[e^{*2}\right]$ for comparison.

Table 1 summarizes the MAD outcomes for the three estimates of γ. Some conclusions can be drawn from the results. First, when endogeneity is weak, that is, $\kappa = 0.2\delta$, the MAD of the STR estimator lies between those of our CF estimators and the 2SLS estimator; when endogeneity is strong, that is, $\kappa = \delta$, the MAD of the STR estimator is the largest. This is because, as is apparent from Figure 1, the STR estimator is consistent in the former case but inconsistent in the latter case. Actually, in the latter case, when $\delta = 0.5$, 1, and 2, the absolute asymptotic biases are 0.797, 0.775, and 0.778, respectively. Correspondingly, the MAD increases according to this scale. Second, the 2SLS estimator is the most inefficient among these three estimators (for the STR estimator, consider only the $\kappa = 0.2\delta$ case). This is because its convergence rate is $n^{1/2-\alpha}$ from YLP, which is much slower than the convergence rate $n^{1-2\alpha}$ of the other two estimators. Third, our CF estimators are the most efficient among these three estimators in all cases. Fourth, for all estimators, endogeneity is evidently harmful, that is, the $\kappa = \delta$ case has a larger MAD than the $\kappa = 0.2\delta$ case. Fifth, comparing the cases where $n = 200$ and $n = 800$, we see that the convergence rate for the 2SLS estimator is roughly $\sqrt{n}$,

TABLE 1. MAD for three estimators of γ

$n \to$		200							800			
$\delta \to$	0.5		1		2		0.5		1		2	
$\kappa \to$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
2SLS	0.900	0.928	0.458	0.617	0.246	0.494	0.417	0.471	0.216	0.290	0.118	0.232
STR	0.511	1.104	0.093	1.006	0.019	0.966	0.074	0.987	0.015	0.944	0.005	0.925
CFs	0.258	0.277	0.066	0.086	0.019	0.021	0.058	0.066	0.015	0.015	0.005	0.005

TABLE 2. Coverage of nominal 95% confidence intervals for γ

$n \to$		200							800			
$\delta \to$	0.5		1		2		0.5		1		2	
$\kappa \to$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
2SLS	0.674	0.672	0.918	0.856	0.957	0.919	0.965	0.934	0.962	0.976	0.952	0.963
STR	0.959	0.576	0.995	0.369	0.996	0.220	0.997	0.334	0.993	0.190	1.00	0.097
CFs	0.974	0.959	0.983	0.974	0.988	0.982	0.977	0.969	0.972	0.980	0.991	0.991

whereas for our CF estimators, it is roughly n.[9] Sixth, when the STR estimator is consistent, it may be exactly the same as our CF estimators. For example, when $\kappa = 0.2\delta, n = 200, \delta = 2$ or $\kappa = 0.2\delta, n = 800, \delta = 1$ and 2, their MADs are exactly the same; this is because the estimators in almost all simulations are the same. In Table 12 in Section SD.5 of the Supplementary Material, we also report the MAD of the IDKE in YP (which reduces to the DKE since q is the only covariate). From Table 12 in Section SD.5 of the Supplementary Material, we can see that the risk of the IDKE is between that of the CF estimators and the 2SLS estimator in all cases. This result is expected because the CF estimators have the same convergence rate as the IDKE but use more data (i.e., z_i's) while the 2SLS estimator has a slower convergence rate than the IDKE.

Tables 2 and 3 summarize performance of the CIs. Table 2 reports coverage and Table 3 reports average length. From Table 2, we draw the following conclusions. First, when the STR estimator is consistent, the corresponding CI has good coverage (mostly over-coverage); but when it is inconsistent, the CI undercovers and the coverage can be lower than 10%. Second, the 2SLS-CI coverage is good when $n\delta$ is large; when $n\delta$ is small (especially when endogeneity is strong), the CI suffers from under-coverage. Third, our CF-CI coverage is excellent in all cases and is unaffected by the severity of endogeneity.

From Table 3, we draw these conclusions. First, our CF-CI has the shortest length in all cases, followed by the STR-CI, and the 2SLS-CI which is longest. This

---

[9]As emphasized in CH, letting $0 < \alpha < 1/2$ is only a device for the construction of a useful asymptotic approximation rather than a behavioral assumption. In this simulation, $\alpha = 0$ gives a better approximation for the MAD.

TABLE 3. Length of nominal 95% confidence intervals for γ

$n \to$			200						800			
$\delta \to$	0.5		1		2		0.5		1		2	
$\kappa \to$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
2SLS	3.656	3.847	2.449	3.023	1.372	2.591	2.285	2.522	1.096	1.542	0.599	1.195
STR	2.639	2.143	0.690	1.318	0.147	0.889	0.657	1.138	0.124	0.637	0.036	0.398
CFs	1.748	1.879	0.362	0.387	0.097	0.100	0.346	0.355	0.079	0.081	0.025	0.025

outcome matches the efficiency results in Table 1. Combined with the results in Table 2, the findings show that CF-CI is shortest and has the best coverage. Second, the STR-CI is much longer than the CF-CI when endogeneity is strong but, even so, the coverage is very low. This outcome is of course due to the inconsistency of the STR estimator. Third, comparing the cases where $n = 200$ and $n = 800$, it is apparent that the length of the 2SLS-CI when $n = 800$ is roughly $1/2$ of that when $n = 200$, whereas for the CF-CI it is roughly $1/4$. These findings match the convergence rate findings in Table 1. Finally, from Tables 1 and 3, it is clear that larger δ tends to be more beneficial, that is, larger δ implies lower risk and shorter CIs.

5.2. Comparison between Two CF Estimators of γ

From Section SD.2 of the Supplementary Material, ϕ and η^2 in CF-I can be simplified as follows:

$$\phi = \frac{\mathbb{E}\left[e_{2i}^2\right]}{\mathbb{E}\left[e_{1i}^2\right]} = \frac{\mathbb{E}\left[(\psi_x e_{xi} + e_{ui})^2\right]}{\mathbb{E}\left[((\delta + \psi_x)\,e_{xi} + e_{ui})^2\right]} = \frac{\psi_x^2 + 1}{(\delta + \psi_x)^2 + 1},$$

and

$$\eta^2 = \mathbb{E}\left[e_{1i}^2\right] = \mathbb{E}\left[((\delta + \psi_x)\,e_{xi} + e_{ui})^2\right] = (\delta + \psi_x)^2 + 1.$$

In CF-II, ϕ and η^2 can again be simplified as

$$\phi = 1 \text{ and } \eta^2 = \mathbb{E}\left[e_{ui}^2\right] = 1.$$

Tables 4–6 summarize performance of the two CF estimators and corresponding CIs. For completeness, we also report performance of the 2SLS estimator based on the moment conditions $\mathbb{E}[\mathbf{z}_i\,(y_i - x_i\delta\mathbf{1}\,(q_i \le \gamma_0))] = \mathbf{0}$. From YLP, the asymptotic distribution of this 2SLS estimator takes the same form as in (28) except that G is redefined as $G = \left(\mathbb{E}\left[\mathbf{z}_i x_{i,\le\gamma_0}\right], \mathbb{E}[\mathbf{z}_i x_i | q_i = \gamma_0]\,c_\beta f\right)$. From these three tables, the following conclusions are drawn. First, CF-II dominates the other two methods with lowest risk, highest coverage and shortest length. Second, and different from Simulation 1, the 2SLS estimator can be more efficient than the CF-I estimator,

TABLE 4. MAD for three estimators of γ

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
2SLS	0.670	0.736	0.511	0.685	0.458	0.665	0.424	0.508	0.343	0.509	0.283	0.488
CF-I	0.756	0.830	0.527	0.688	0.410	0.625	0.453	0.515	0.331	0.421	0.277	0.391
CF-II	0.269	0.280	0.070	0.075	0.030	0.032	0.059	0.063	0.019	0.019	0.008	0.008

TABLE 5. Coverage of nominal 95% confidence intervals for γ

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
2SLS	0.707	0.698	0.770	0.712	0.747	0.723	0.783	0.763	0.776	0.736	0.741	0.732
CF-I	0.960	0.961	0.952	0.952	0.953	0.955	0.967	0.959	0.963	0.960	0.952	0.960
CF-II	0.976	0.970	0.978	0.979	0.988	0.981	0.972	0.982	0.981	0.981	0.990	0.986

TABLE 6. Length of nominal 95% confidence intervals for γ

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
2SLS	3.814	4.052	3.058	3.922	2.560	3.831	2.844	3.320	2.060	3.092	1.602	3.015
CF-I	2.669	2.996	1.802	2.432	1.564	2.233	1.526	1.744	1.099	1.433	0.951	1.338
CF-II	1.307	1.387	0.370	0.377	0.130	0.136	0.320	0.336	0.092	0.093	0.033	0.033

especially when $n\delta$ is small; also, the respective CIs always undercover. Third, for all methods, endogeneity is harmful and larger δ is beneficial (stronger endogeneity and smaller δ implies higher risk and longer CIs). Fourth, comparing the risks and CI lengths when $n = 200$ and $n = 800$, we see that the convergence rates of the 2SLS and CF-I estimators are slower than $\sqrt{n}$, while the convergence rate of the CF-II estimator is roughly of order n. In summary, the performance of CF-II is extremely good and the other two methods are not recommended. In Table 13 in Section SD.5 of the Supplementary Material, we also report the MAD of the IDKE in YP. From Table 13 in Section SD.5 of the Supplementary Material, we can see that the risk of the IDKE is between that of the CF-II estimator and the CF-I estimator (also the 2SLS estimator) in all cases. Although it is expected that the IDKE is less efficient than the CF-II estimator since the latter uses more data (i.e.,

z_i's), it is surprising to see that the CF-I estimator is worse than the IDKE given that it also uses the same set of additional data (z_i's). This result corroborates the discussion in Section 3 where we argue that CF-II is the preferable CF approach compared with CF-I.

5.3. Comparison among Seven Estimators of δ

For the estimation of δ, we can use the moment conditions $\mathbb{E}[\mathbf{z}_i(y_i - x_i\delta 1(q_i \leq \gamma_0))] = \mathbf{0}$ to estimate δ and γ_0 jointly as in the 2SLS procedure of YLP. Note that under our DGP

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{z}_i(y_i - x_i\delta)1(q_i \leq \gamma_0)\right] &= \mathbb{E}[\mathbf{z}_i u_i 1(q_i \leq \gamma_0)] \\
&= \mathbb{E}[\mathbf{z}_i\left((\psi_x\varphi + \psi_q)v_{qi} + \psi_x e_{xi} + e_{ui}\right)1(\pi'\mathbf{z}_i + v_{qi} \leq \gamma_0)] \\
&= \left(\psi_x\varphi + \psi_q\right)\mathbb{E}\left[\mathbf{z}_i v_{qi}1(v_{qi} \leq \gamma_0 - \pi'\mathbf{z}_i)\right] \\
&= -\left(\psi_x\varphi + \psi_q\right)\mathbb{E}\left[\mathbf{z}_i\lambda\left(\gamma_0 - \pi'\mathbf{z}_i\right)\right] \neq \mathbf{0}.
\end{aligned}
$$

So the moment $\mathbb{E}[\mathbf{z}_i(y_i - x_i\delta)1(q_i \leq \gamma_0)]$ cannot be used to estimate δ in our case. Instead, we will use the six estimators in Section 4, that is, CF-I, GMM-I1, GMM-II1, CF-II, GMM-I2, and GMM-II2, to estimate δ. Since δ is a regular parameter, we use the usual root-mean-square error (RMSE) criterion to evaluate its risk. Except for YLP's 2SLS, where the nonparametric bootstrap is used to construct CIs, we invert the t-statistics using the asymptotic critical values to construct CIs for the other six estimators.

In CF-II and GMM-II2, we can estimate δ and ψ jointly to improve the efficiency of δ estimation. The details of the asymptotic variances can be found at the end of Section SD.4 of the Supplementary Material. Due to the special setup of our DGP2, the asymptotic variances of the CF-I, GMM-1, and GMM-I2 estimators can also be simplified; we detail these simplifications in Section SD.3 of the Supplementary Material after the discussion on the general results.

The performance of these seven estimators of δ are summarized in Tables 7–9. From these tables, the following conclusions are drawn. First, the rank of the risk from lowest to highest among these seven estimators in all cases is GMM-II2, CF-II, GMM-II1, GMM-I1, CF-I, GMM-I2, and 2SLS. So the methods based on CF-II $\widehat{\gamma}$ are generally better than those based on CF-I $\widehat{\gamma}$. The superiority of GMM-II1 compared with GMM-I1 can be explained by the more precise estimation of γ in CF-II. But the superiority of CF-II and GMM-II2 compared with CF-I and GMM-I2 cannot be explained solely by this factor. It seems that the former two methods themselves have some intrinsic advantages given that their risks can be less than one half of the latter two methods. GMM-I2 is a little worse than CF-I, while GMM-II2 is better than CF-II especially when endogeneity is severe. Second, although the 2SLS estimator is the most inefficient, the corresponding CIs have the best coverage uniformly for all cases. The superiority of the 2SLS coverage may be explained by the fact that δ is estimated jointly with γ so, different from the other methods, the uncertainty of $\widehat{\gamma}$ is incorporated. On the other hand, the

TABLE 7. RMSE for four estimators of δ

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
2SLS	0.187	0.224	0.286	0.388	0.419	0.782	0.094	0.145	0.090	0.251	0.108	0.452
CF-I	0.166	0.242	0.183	0.366	0.275	0.615	0.063	0.092	0.076	0.142	0.110	0.242
GMM-I1	0.119	0.168	0.125	0.282	0.157	0.483	0.053	0.075	0.054	0.118	0.068	0.208
GMM-I2	0.167	0.244	0.184	0.369	0.277	0.625	0.063	0.093	0.076	0.144	0.111	0.245
CF-II	0.073	0.092	0.078	0.139	0.087	0.242	0.036	0.047	0.037	0.070	0.042	0.119
GMM-II1	0.098	0.142	0.106	0.240	0.134	0.439	0.049	0.072	0.051	0.118	0.062	0.207
GMM-II2	0.073	0.088	0.078	0.098	0.083	0.104	0.036	0.044	0.037	0.049	0.040	0.052

TABLE 8. Coverage of nominal 95% confidence intervals for δ

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
2SLS	0.965	0.965	0.969	0.966	0.978	0.957	0.970	0.951	0.976	0.966	0.982	0.977
CF-I	0.895	0.875	0.905	0.874	0.922	0.895	0.923	0.908	0.917	0.911	0.939	0.942
GMM-I1	0.942	0.939	0.939	0.920	0.951	0.949	0.940	0.948	0.939	0.952	0.951	0.965
GMM-I2	0.895	0.873	0.896	0.864	0.911	0.889	0.925	0.905	0.913	0.906	0.934	0.938
CF-II	0.948	0.946	0.938	0.940	0.952	0.952	0.953	0.941	0.951	0.942	0.954	0.964
GMM-II1	0.949	0.950	0.957	0.944	0.949	0.959	0.947	0.955	0.947	0.946	0.959	0.959
GMM-II2	0.947	0.947	0.933	0.951	0.949	0.948	0.950	0.935	0.949	0.952	0.954	0.956

2SLS CI is much longer than the other methods. Third, CF-I and GMM-I2 suffer from some under-coverage especially when $n\delta$ is small,[10] whereas the coverage of GMM-1, CF-II and GMM-II2 is excellent or near perfect in all cases. Fourth, as expected, the rank of the CI length matches the rank of risk. Fifth, similar to γ estimation, endogeneity is harmful to all methods; again stronger endogeneity implies more risk and longer CIs (but in larger magnitudes than for γ estimation). Sixth, comparing the risks and CI lengths when $n = 200$ and $n = 800$, we see that the convergence rates of all estimators are roughly $\sqrt{n}$. In summary, the GMM-II2 estimator performs best, and after balancing coverage and length, GMM-II2 delivers the best CI.

---

[10] As in CH, we could use Bonferroni-type methods to improve coverage for these CIs. Because there is as yet no rigorous procedure to select the coverage level of γ in this Bonferroni-type CI construction, we will not investigate it in this paper.

TABLE 9. Length of nominal 95% confidence intervals for δ

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
2SLS	0.969	1.146	1.401	2.062	2.101	3.946	0.512	0.680	0.479	1.188	0.518	2.271
CF-I	0.488	0.660	0.582	1.063	0.870	1.961	0.220	0.302	0.266	0.486	0.399	0.907
GMM-I1	0.435	0.643	0.458	1.032	0.583	1.891	0.198	0.293	0.214	0.472	0.275	0.880
GMM-I2	0.488	0.659	0.580	1.061	0.864	1.957	0.220	0.302	0.265	0.486	0.395	0.906
CF-II	0.284	0.362	0.294	0.544	0.336	0.978	0.140	0.181	0.146	0.272	0.168	0.488
GMM-II1	0.392	0.573	0.412	0.923	0.503	1.710	0.189	0.278	0.202	0.452	0.248	0.848
GMM-II2	0.284	0.334	0.293	0.386	0.322	0.428	0.140	0.168	0.146	0.192	0.161	0.207

In Section SD.5 of the Supplementary Material, we also compare the performance of the two CF approaches when q is exogenous. All the results in Simulations 2 and 3 follow as above (especially the CF-II (GMM-II2) estimators and associated CIs of γ (δ) are found to perform best). When q is exogenous, we can also estimate δ based on the moment conditions $\mathbb{E}\left[\mathbf{z}_i\left(y_i - x_i\delta\right)1\left(q_i \leq \gamma_0\right)\right] = \mathbf{0}$, as in CH. It turns out that the corresponding estimator is more efficient than that based on the moment conditions $\mathbb{E}\left[\mathbf{z}_i\left(y_i - x_i\delta 1\left(q_i \leq \gamma_0\right)\right)\right] = \mathbf{0}$ both asymptotically and in finite samples. This is because δ is only related to the data points with $q_i \leq \gamma_0$, whereas the moment conditions $\mathbb{E}\left[\mathbf{z}_i\left(y_i - x_i\delta 1\left(q_i \leq \gamma_0\right)\right)\right] = \mathbf{0}$ also use data points with $q_i > \gamma_0$, which introduces some redundancy in the estimation of δ and thereby blunts the sharper information in the former moment conditions. Note further that CH's GMM has the lowest risk among the four estimators of δ that are associated with CF-I $\widehat{\gamma}$, that is, CF-I, CH's GMM, GMM-I1, and GMM-I2. In other words, if we use CH's $\widehat{\gamma}$, their GMM estimator of δ is indeed the best.

6. EMPIRICAL APPLICATION

As in CMT, we estimate a threshold model of the foreign-trade multiplier relationship, quantifying the effects of exports on real output growth, where the level of the real exchange rate volatility is used as the threshold variable. Specifically, we employ the following TR model:

$$y_i = \left(\mathbf{x}_i'\beta_1 + \sigma_1 u_i\right)1\left(q_i \leq \gamma\right) + \left(\mathbf{x}_i'\beta_2 + \sigma_2 u_i\right)1\left(q_i > \gamma\right) + \rho y_{i-1},$$

where y_i is the real GDP growth rate, $\mathbf{x}_i = (1, x_i)'$ with x_i being the export growth rate of goods and services, and q_i is a measure of the real effective exchange rage volatility (REER). Adding y_{i-1} into the system is to control for the effects of persistent shocks in y_i, but we assume such effects do not exhibit any threshold effect. As noted in Tenreyro (2007), the two regimes defined by REER may reflect differences between flexible and stable (or pegging) exchange rate regimes, and

the threshold variable REER q_i is correlated with both x_i and u_i. Following CMT, we do not consider the endogeneity of x_i because their test cannot reject that x_i is exogenous. In this case, our CF-I and CF-II estimators of γ are the same. The method of CMT does not require any instruments. For our methods, we use $\mathbf{z}_i = (1, x_i, y_{i-1}, q_{i-1})'$ as instruments, that is, we use q_{i-1} as the only excluded instrument, and specify the first-stage equation as

$$q_i = \pi' \mathbf{z}_i + v_{qi},$$

and assume $\mathbb{E}\left[u_i | v_{qi}\right] = \psi v_{qi}$. The linear form of $\mathbb{E}\left[u_i | v_{qi}\right]$ can be justified as follows. The determination of GDP growth y_i and exchange rate volatility q_i involves many factors beyond $\mathbf{x}_i$ and $\mathbf{z}_i$, and all these factors combine to influence u_i and v_{qi}. By a commonly used central limit argument the resulting variables u_i and v_{qi} in aggregating these influences would approximately follow normal distributions, which in turn implies a linear form of $\mathbb{E}\left[u_i | v_{qi}\right]$. The primary purpose of this empirical application is to show the difference between KST's STR estimator and our two CF estimators in this setting.

The data of CMT contain seven OECD countries: Canada (CA), France (FR), Italy (IT) Japan (JP), United Kingdom (UK), United States (US), and Netherlands (NL), from 1966 to 2014.[11] Since the tests in CMT cannot reject that CA and FR do not exhibit any threshold effects, we consider the remaining five countries. Table 10 reports the estimates of model parameters γ and $\underline{\theta}$ based on CF-II and GMM-II2 (which are the same as CF-I and GMM-I2 in this application) since they are the most efficient in our simulations. From Table 10, we can draw the following conclusions. First, the percentages of observations in the high volatility regime "2" are bounded away from 0 and 1, which indicates that there are sufficient observations in each regime for our estimation. Second, ρ is bounded away from 1, that is, there is no unit root problem, which indicates that our assumption of stationarity is reasonable in this application. Third, for IT, JP, and US, we can reject $\delta_\kappa = 0$ (i.e., $\kappa_1 = \sigma_1 \psi = \sigma_2 \psi = \kappa_2$), which implies $\sigma_1 \neq \sigma_2$ when $\psi \neq 0$ (this is the case at least for IT and JP from the CIs for κ_ℓ). In other words, our specification of a threshold effect in the variance of error term is necessary in this application. Fourth, all CIs (whether for γ or β) are quite wide, which is due to the small sample size $n = 49$. As a result, the CIs of γ in our CF approach cover the KST and CMT estimators. Fifth, our $\widehat{\gamma}$'s for IT and JP are very different from the KST and CMT estimators, while for UK, US, and NL, the three estimators are similar or even the same. This can be explained from the testing results for $\kappa_1 = \kappa_2 = 0$. For IT and JP, this endogeneity test rejects the null at the 5% level; in other words, the endogeneity level is comparable to the jump size. In this case, the KST and CMT estimators are inconsistent, as explained in Section 2. On the other hand, for the remaining three countries, the endogeneity test cannot reject or only marginally rejects the null at the 5% level, that is, the endogeneity level is negligible compared to the jump size. In this case, all three estimators are consistent. Note that our

---

[11] Our data start from 1966 rather than 1965 as CMT because we use q_{i-1} as an instrument.

endogeneity test results differ from those in CMT where the null is rejected for all five countries. Sixth, we check the effects of x_i on y_i via the coefficients β_{12} and β_{22}. It turns out that our results are qualitatively similar to CMT. On the one hand, β_{12} is significantly positive for IT, US, and NL, which supports the view that low real exchange rate volatility favors exports and economic growth (e.g., Grier and Smallwood, 2013). On the other hand, β_{22} is significant and greater than 1 for JP and the UK, that is, high real exchange rate volatility favors exports and growth (e.g., Egert and Morales-Zumaquero, 2008); see CMT for more references on this phenomenon.

7. CONCLUSION

We conclude this paper by providing a summary of the strengths and weaknesses of the available estimators of γ and β in ETR. Table 11 lists the number of instruments required by each estimator. YLP's 2SLS estimator requires the most instruments in estimating both γ and β and has a slower convergence rate in estimating γ. Although we can use transformations of $\mathbf{z}$ as instruments when more than $\dim(\mathbf{z})$ instruments are required, such instruments provide information only through nonlinearity and seem unattractive for practical use. The CI for β (but not γ) has good coverage (but not length) because this method estimates γ and β jointly so it takes into account the randomness in estimating γ. The joint property of this estimation method differs from all the other methods. CH's 2SLS cannot be used to estimate γ and CH's GMM cannot be used to estimate β when q is endogenous. In such cases, we can use CF-I or GMM-I1 to substitute for CH's 2SLS or CH's GMM. In this sense, our CF-I estimate of γ combined with the GMM-I1 estimate of β may be viewed as an extension of CH's 2SLS combined with CH's GMM. However, GMM-I1 requires more instruments than CH's GMM and cannot handle the $\delta_\sigma \neq 0$ case easily. GMM-I2 is an extension of CF-I; both differ from GMM-I1 and do not suffer the problems of GMM-I1. The CF-II estimator of γ combined with the corresponding estimators of β are parallel developments of the CF-I estimators.

Compared with CF-I, CF-II estimation of γ and β has advantages in both its theory foundation and its finite sample performance. Among all estimators and CIs for β, the GMM-II2 estimator and CI have the best finite sample performance. After balancing all dimensions of each method, we suggest the use of CF-II to estimate and construct CIs for γ and use of GMM-II2 for estimation of β when instruments are available. When instruments are absent, we may use YP's IDKE method to estimate γ. Of course, we can use YP's IDKE even when instruments are available. However, the method is nonparametric so may suffer from the curse of dimensionality and requires bandwidth selection in practice which may be troublesome when $\dim(\mathbf{x})$ is large. For β, YP's IDKE can estimate δ_β when $\delta_\sigma = 0$ and $\mathbb{E}[u|\mathbf{x}]$ is smooth, but in general instruments are required to estimate β. The latter is not unexpected in view of endogeneity in the regression.

TABLE 10. Estimates of model parameters of γ and $\underline{\theta}$

	IT	JP	UK	US	NL
KST	0.040	0.030	0.051	0.045	0.022
CMT	0.024	0.086	0.054	0.045	0.027
CF	0.055	0.047	0.051	0.045	0.022
	$(0.019, 0.081)$	$(0.014, 0.080)$	$(0.021, 0.087)$	$(0.043, 0.068)$	$(0.012, 0.026)$
β_{12}	1.296	-1.699	0.079	1.492	0.841
	$(0.920, 1.672)$	$(-4.422, 1.024)$	$(-0.490, 0.648)$	$(0.339, 2.645)$	$(0.699, 0.983)$
β_{22}	0.312	1.227	1.587	0.544	0.431
	$(-0.149, 0.774)$	$(0.646, 1.808)$	$(1.082, 2.093)$	$(-1.903, 2.990)$	$(0.266, 0.596)$
ρ	0.467	0.423	0.183	0.400	0.455
	$(0.237, 0.697)$	$(0.167, 0.679)$	$(-0.106, 0.471)$	$(0.118, 0.571)$	$(0.322, 0.587)$
κ_1	1.127	-3.464	-0.213	0.813	0.858
	$(0.330, 1.923)$	$(-4.900, -2.027)$	$(-0.754, 0.328)$	$(-0.023, 1.649)$	$(-0.078, 1.795)$
κ_2	-0.458	0.059	-0.080	-0.590	0.000
	$(-0.827, -0.088)$	$(-0.102, 0.220)$	$(-0.626, 0.466)$	$(-1.348, 0.167)$	$(-0.524, 0.525)$
$\delta_\kappa = 0$	0.000	0.000	0.740	0.016	0.118
$\kappa_1 = \kappa_2 = 0$	0.006	0.000	0.955	0.047	0.259
%obs"2"	24.49	85.71	57.14	28.57	77.55

Note: The first three rows report different γ estimators, $\beta_{\ell 2}$ is the coefficient of x_i in each regime, 95% CIs are reported in the parentheses, $\delta_\kappa = 0$ and $\kappa_1 = \kappa_2 = 0$ report the p-values for the two tests, and %obs"2" reports the percentages of observations in regime "2."

TABLE 11. Number of instruments required in estimation of γ and β by each method

Parameters →	γ			β	
Endogeneity of q →	$q \in \mathbf{z}$	$q \notin \mathbf{z}$		$q \in \mathbf{z}$	$q \notin \mathbf{z}$
CH's 2SLS	d	NA	CH's GMM	d	NA
CF-I	d	d	CF-I	d	d
			GMM-I1 ($\delta_\sigma = 0$)	$2d$	$2d$
			GMM-I2	d	d
CF-II	d	d	CF-II	d	d
			GMM-II1 ($\delta_\sigma = 0$)	$2d$	$2d$
			GMM-II2	d	d
YLP's 2SLS	$2d+1$	$2d+1$	YLP's 2SLS	$2d+1$	$2d+1$
YP's IDKE	0	0	YP's IDKE	NA	NA

Appendix A. Proofs

Notation is collected together here for convenient reference in the proofs. The letter C is used as a generic positive constant, which need not be the same in each occurrence, $a_n = n^{1-2\alpha}$, and $\rightsquigarrow$ signifies weak convergence over a compact metric space. The matrices $X, X_\gamma, X_\perp, \breve{X},$ $\breve{X}_\gamma, \breve{X}_\perp, V, \widehat{X}, \widehat{X}_\gamma, \widehat{X}_\perp, \widehat{V}, \widehat{V}_\gamma, Z, \mathbf{e}^0$ and $\mathbf{u}$ are defined by stacking $\mathbf{x}'_i, \mathbf{x}'_{i,\leq\gamma}, \mathbf{x}'_{i,>\gamma}, \breve{\mathbf{x}}'_i, \breve{\mathbf{x}}'_{i,\leq\gamma},$ $\breve{\mathbf{x}}'_{i,>\gamma}, \mathbf{v}'_i, \widehat{\mathbf{x}}'_i, \widehat{\mathbf{x}}'_{i,\leq\gamma}, \widehat{\mathbf{x}}'_{i,>\gamma}, \widehat{\mathbf{v}}'_i, \widehat{\mathbf{v}}'_{i,\leq\gamma}, \mathbf{z}'_i, e^0_i$ and $u_i, \breve{X}_0 = \breve{X}_{\gamma 0}, \widehat{X}_0 = \widehat{X}_{\gamma 0}, \widehat{X}^*_\gamma = (\widehat{X}_\gamma, \widehat{X}_\perp),$ $\theta = (\gamma, \underline{\theta}')' = (\gamma, \overline{\beta}'_1, \overline{\beta}'_2)'$ or equivalently $(\gamma, \overline{\beta}'_2, \overline{\delta}'_n)'$ with $\overline{\beta}_\ell = (\beta'_\ell, \kappa'_\ell)', \ell = 1, 2,$ and $\overline{\delta}_n = (\delta'_\beta, \delta'_\kappa)',$ where the upper bar on the parameters and the breve over the variables mean "augmented." $\overline{c} = (n^\alpha \delta'_\beta, n^\alpha \delta_\sigma \psi')' = \begin{pmatrix} I_d & \mathbf{0} \\ \mathbf{0} & \psi \end{pmatrix} c. \widehat{e}_i = \widehat{r}'_i \kappa_2 + e^0_i,$ and $\widehat{\mathbf{e}} = \widehat{\mathbf{r}}\kappa_2 + \mathbf{e}^0$ stacks $\widehat{e}_i,$ where $\widehat{r}'_i = \mathbf{v}'_i - \widehat{\mathbf{v}}'_i = \mathbf{z}'_i(\widehat{\Pi} - \Pi),$ and $\widehat{\mathbf{r}} = V - \widehat{V}$ stacks $\widehat{r}'_i.$

Proof of Lemma 1. First, we show that $S_n(\theta)$ and $\widetilde{S}_n(\theta)$ have the same probability limit, where $\widetilde{S}_n(\theta)$ is the same as $S_n(\theta)$ except replacing $\widehat{\mathbf{v}}_i$ by $\mathbf{v}_i$. Since

$$
\begin{aligned}
n^{-1}\left(S_n(\theta) - \widetilde{S}_n(\theta)\right) = &-2n^{-1}\sum_{i=1}^n \left[y_i - \left(\mathbf{x}'_i\beta_1 + \sigma_1 g_{3,i,\psi} + \sigma_1 \Delta\widehat{g}_{3,i,\psi}/2\right)1\left(q_i \leq \gamma\right)\right. \\
&\left. - \left(\mathbf{x}'_i\beta_2 + \sigma_2 g_{3,i,\psi} + \sigma_2 \Delta\widehat{g}_{3,i,\psi}/2\right)1\left(q_i > \gamma\right)\right] \\
&\cdot \left[\sigma_1 \Delta\widehat{g}_{3,i,\psi}1\left(q_i \leq \gamma\right) + \sigma_2 \Delta\widehat{g}_{3,i,\psi}1\left(q_i > \gamma\right)\right] \\
= &-2n^{-1}\sum_{i=1}^n \left[y_i - \left(\mathbf{x}'_i\beta_1 + \sigma_1 g_{3,i,\psi}\right)1\left(q_i \leq \gamma\right) - \left(\mathbf{x}'_i\beta_2 + \sigma_2 g_{3,i,\psi}\right)1\left(q_i > \gamma\right)\right] \\
&\times \left[\sigma_1 \Delta\widehat{g}_{3,i,\psi}1\left(q_i \leq \gamma\right) + \sigma_2 \Delta\widehat{g}_{3,i,\psi}1\left(q_i > \gamma\right)\right] \\
&+ n^{-1}\sum_{i=1}^n \left[\sigma_1 \Delta\widehat{g}_{3,i,\psi}1\left(q_i \leq \gamma\right) + \sigma_2 \Delta\widehat{g}_{3,i,\psi}1\left(q_i > \gamma\right)\right]^2,
\end{aligned}
$$

where $g_{3,i,\psi} = g_3(\mathbf{v}_i; \psi)$, and

$$\Delta\widehat{g}_{3,i,\psi} = g_3(\widehat{\mathbf{v}}_i; \psi) - g_3(\mathbf{v}_i; \psi) \le C \|\widehat{\mathbf{v}}_i - \mathbf{v}_i\| \le C \|\widehat{\Pi} - \Pi\| \|\mathbf{z}_i\| = o_p(1) \|\mathbf{z}_i\|,$$

with the first inequality from C-II.6, the second inequality from $\widehat{\mathbf{v}}_i - \mathbf{v}_i = (\Pi - \widehat{\Pi})' \mathbf{z}_i$, and the last equality from C-II.2-3, we need only show that

$$n^{-1} \sum_{i=1}^{n} \left| y_i - (\mathbf{x}_i'\beta_1 + \sigma_1 g_{3,i,\psi}) \mathbf{1}(q_i \le \gamma) - (\mathbf{x}_i'\beta_2 + \sigma_2 g_{3,i,\psi}) \mathbf{1}(q_i > \gamma) \right| \|\mathbf{z}_i\| = O_p(1), \tag{A.1}$$

uniformly in θ, and

$$n^{-1} \sum_{i=1}^{n} \|\mathbf{z}_i\|^2 = O_p(1). \tag{A.2}$$

(A.2) is straightforward by the ergodic theorem and C-II.3. As for (A.1), we can apply Lemma 2.4 of Newey and McFadden (1994). Although Lemma 2.4 of Newey and McFadden (1994) is stated for i.i.d. observations, the result requires only application of a weak law of large numbers, which holds under C-II.1 by the ergodic theorem. The almost sure continuity condition in Lemma 2.4 of Newey and McFadden (1994) is implied by the continuity of $g_{3,i,\psi}$ in ψ (C-II.6) and the distribution of q (C-II.5). To check the summand is dominated by a function of w_i with finite first moment, we need only show $\mathbb{E}\left[y_i^2\right] < \infty$, $\mathbb{E}\left[\|\mathbf{x}_i\|^2\right] < \infty$ and $\mathbb{E}\left[g_{3,i,\psi_0}^2\right] < \infty$ by the Cauchy–Schwarz inequality and

$$\mathbb{E}\left[g_{3,i,\psi}^2\right] \le C \left(\mathbb{E}\left[g_{3,i,\psi_0}^2\right] + \|\psi - \psi_0\|^2\right)$$

for any ψ (which is implied by C-II.6). $\mathbb{E}\left[\|\mathbf{x}_i\|^2\right] < \infty$ is implied by C-II.3, $\mathbb{E}\left[g_{3,i,\psi_0}^2\right] < \infty$ is assumed in C-II.4, and $\mathbb{E}\left[y_i^2\right] < \infty$ is implied by these two results and $\mathbb{E}\left[(e_i^0)^2\right] < \infty$ (C-II.4).

Second, we prove the consistency of $\widehat{\theta}$ by applying Theorem 2.1 of Newey and McFadden (1994). For this purpose, we need only show that $\widetilde{S}_n(\theta)$ converges uniformly in probability to $\overline{S}(\theta)$ which is continuous and minimized uniquely at θ_0. By Lemma 2.4 of Newey and McFadden (1994) and the analysis above, $\widetilde{S}_n(\theta)$ converges uniformly in probability to

$$\overline{S}(\theta) = \mathbb{E}\left[\left\{y_i - (\mathbf{x}_i'\beta_1 + \sigma_1 g_{3,i,\psi}) \mathbf{1}(q_i \le \gamma) - (\mathbf{x}_i'\beta_2 + \sigma_2 g_{3,i,\psi}) \mathbf{1}(q_i > \gamma)\right\}^2\right],$$

which is continuous in θ. From Section 2.2.2 of Newey and McFadden (1994), $\overline{S}(\theta)$ is minimized uniquely at θ_0 if $g(w; \theta) = g(w; \theta_0)$ implies $\theta = \theta_0$, where $g(w; \theta) := (\mathbf{x}'\beta_1 + \sigma_1 g_3(\mathbf{v}; \psi)) \mathbf{1}(q \le \gamma) - (\mathbf{x}'\beta_2 + \sigma_2 g_3(\mathbf{v}; \psi)) \mathbf{1}(q > \gamma)$. If not, then there are two cases: $\gamma \ne \gamma_0$ and $\gamma = \gamma_0$. For $\gamma \ne \gamma_0$, there are three subcases:

> **Case (i):** $\gamma \ne \gamma_0$, $\underline{\theta} = \underline{\theta}_0$, where $\theta := (\gamma, \underline{\theta}')'$. By C-II.5, $P(\gamma < q \le \gamma_0) > 0$, where we assume without loss of generality that $\gamma < \gamma_0$. So C-II.1 and C-II.7 imply $P(g(w; \theta) \ne g(w; \theta_0)) > 0$, a contradiction, and thus $\gamma = \gamma_0$.

Case (ii): $\gamma > \gamma_0, \underline{\theta} \neq \underline{\theta}_0, \theta_1 = \theta_{20}$ or $\gamma < \gamma_0, \underline{\theta} \neq \underline{\theta}_0, \theta_2 = \theta_{10}$. From C-II.5, $P(q < \gamma) > 0$ and $P(q > \gamma) > 0$. Then by C-II.1 and C-II.7, $P(g(w; \theta) \neq g(w; \theta_0)) > 0$, so $\gamma = \gamma_0$;

Case (iii): $\gamma > \gamma_0, \theta_1 \neq \theta_{20}$ or $\gamma < \gamma_0, \theta_2 \neq \theta_{10}$. Similar arguments as in Case (i) lead to $\gamma = \gamma_0$.

So $\gamma = \gamma_0$. With $\gamma = \gamma_0$, by C-II.7, $\theta_\ell = \theta_{\ell 0}$. In summary, $g(w; \theta) = g(w; \theta_0)$ implies $\theta = \theta_0$. □

Proof of Theorem 1. By Proposition B.2 in Appendix B, $a_n(\widehat{\gamma} - \gamma_0) = \text{argmax}_\nu Q_n(\nu) = O_p(1)$ and by Proposition B.3 in Appendix B,

$$Q_n(\nu) \rightsquigarrow \begin{cases} 2\sqrt{\overline{c}'V_0^-\overline{c}f}\,W_1(-\nu) - \overline{c}'D_0\overline{c}f\,|\nu|, & \text{if } \nu \leq 0, \\ 2\sqrt{\overline{c}'V_0^+\overline{c}f}\,W_2(\nu) - \overline{c}'D_0\overline{c}f\,|\nu|, & \text{if } \nu > 0, \end{cases}$$

$$=: \begin{cases} 2\sqrt{\lambda^-}\,W_1(-\nu) - \mu\,|\nu|, & \text{if } \nu \leq 0, \\ 2\sqrt{\lambda^+}\,W_2(\nu) - \mu\,|\nu|, & \text{if } \nu > 0, \end{cases}$$

where the limit functional is continuous with a unique maximum almost surely. Appealing to Theorem 2.7 of Kim and Pollard (1990), we have

$$a_n(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \text{argmax}_\nu \begin{cases} 2\sqrt{\lambda^-}\,W_1(-\nu) - \mu\,|\nu|, & \text{if } \nu \leq 0, \\ 2\sqrt{\lambda^+}\,W_2(\nu) - \mu\,|\nu|, & \text{if } \nu > 0. \end{cases}$$

Making the change-of-variables $\nu = \frac{\lambda^-}{\mu^2}r$, noting the distributional equality $W_\ell(a^2 r) =_d a W_\ell(r)$, and setting $\omega = \lambda^-/\mu^2$, we can rewrite the asymptotic distribution as

$$\text{argmax}_\nu \begin{cases} 2\sqrt{\lambda^-}\,W_1(-\nu) - \mu\,|\nu|, & \text{if } \nu \leq 0, \\ 2\sqrt{\lambda^+}\,W_2(\nu) - \mu\,|\nu|, & \text{if } \nu > 0, \end{cases}$$

$$= \frac{\lambda^-}{\mu^2}\,\text{argmax}_r \begin{cases} 2\sqrt{\lambda^-}\,W_1\left(-\frac{\lambda^-}{\mu^2}r\right) - \frac{\lambda^-}{\mu}\,|r|, & \text{if } r \leq 0, \\ 2\sqrt{\lambda^+}\,W_2\left(\frac{\lambda^-}{\mu^2}r\right) - \frac{\lambda^-}{\mu}\,|r|, & \text{if } r > 0, \end{cases}$$

$$= \omega\,\text{argmax}_r \begin{cases} 2\frac{\lambda^-}{\mu}\,W_1(-r) - \frac{\lambda^-}{\mu}\,|r|, & \text{if } r \leq 0, \\ 2\sqrt{\frac{\lambda^+}{\lambda^-}}\frac{\lambda^-}{\mu}\,W_2(r) - \frac{\lambda^-}{\mu}\,|r|, & \text{if } r > 0, \end{cases}$$

$$= \omega\,\text{argmax}_r \begin{cases} W_1(-r) - \frac{|r|}{2}, & \text{if } r \leq 0, \\ \sqrt{\phi}\,W_2(r) - \frac{|r|}{2}, & \text{if } r > 0. \end{cases}$$

□

Proof of Corollary 1. By the CMT,

$$S_n(\gamma_0) - S_n(\widehat{\gamma}) \xrightarrow{d} \sup_\nu \begin{cases} 2\sqrt{\lambda^-}\,W_1(-\nu) - \mu\,|\nu|, & \text{if } \nu \leq 0, \\ 2\sqrt{\lambda^+}\,W_2(\nu) - \mu\,|\nu|, & \text{if } \nu > 0, \end{cases}$$

$$= \sup_r \begin{cases} 2\frac{\lambda^-}{\mu}\,W_1(-r) - \frac{\lambda^-}{\mu}\,|r|, & \text{if } r \leq 0, \\ 2\sqrt{\frac{\lambda^+}{\lambda^-}}\frac{\lambda^-}{\mu}\,W_2(r) - \frac{\lambda^-}{\mu}\,|r|, & \text{if } r > 0, \end{cases}$$

$$
= \frac{2\lambda^-}{\mu} \sup_r \begin{cases} W_1(-r) - \frac{|r|}{2}, & \text{if } r \leq 0, \\ \sqrt{\phi} W_2(r) - \frac{|r|}{2}, & \text{if } r > 0, \end{cases}
$$

$$
= \frac{2\lambda^-}{\mu} \max\{\xi_1, \xi_2(\phi)\} = \frac{2\lambda^-}{\mu} \xi(\phi),
$$

where $\xi_1 = \sup_{r \leq 0} \left\{ W_1(-r) - \frac{1}{2}|r| \right\}$, $\xi_2(\phi) = \sup_{r > 0} \left\{ \sqrt{\phi} W_2(r) - \frac{1}{2}|r| \right\}$, and ξ_1 and ξ_2 are independent. From Bhattacharya and Brockwell (1976), ξ_1 has a standard exponential distribution, and $\xi_2(\phi)$ has an exponential distribution with mean ϕ. We deduce that

$$
P(\xi(\phi) \leq x) = P(\xi_1 \leq x, \xi_2(\phi) \leq x) = P(\xi_1 \leq x) P(\xi_2(\phi) \leq x) = (1 - e^{-x})(1 - e^{-x/\phi}),
$$

and[12] the required result follows by Slutsky's theorem. □

Proof of Theorem 2. Take $\widehat{\overline{\beta}}_1$ as an example since $\widehat{\overline{\beta}}_2$ can be similarly analyzed. Note that

$$
n^{1/2}(\widehat{\overline{\beta}}_1 - \overline{\beta}_1)
$$

$$
= n^{1/2} \left[(\widehat{X}_1' \widehat{Z}_1) W_{1n} (\widehat{Z}_1' \widehat{X}_1) \right]^{-1} \left[(\widehat{X}_1' \widehat{Z}_1) W_{1n} \widehat{Z}_1' \left(\check{X}\overline{\beta}_2 + \check{X}_0 \bar{c} n^{-\alpha} - \widehat{X}_1 \overline{\beta}_1 + \mathbf{e}^0 \right) \right]
$$

$$
= \left[\left(\frac{1}{n} \widehat{X}_1' \widehat{Z}_1 \right) W_{1n} \left(\frac{1}{n} \widehat{Z}_1' \widehat{X}_1 \right) \right]^{-1}
$$

$$
\times \left[\left(\frac{1}{n} \widehat{X}_1' \widehat{Z}_1 \right) W_{1n} \frac{1}{\sqrt{n}} \widehat{Z}_1' \left(\left(\check{X}_1 - \widehat{X}_1 \right) \overline{\beta}_1 - \check{X}_1 \overline{\beta}_1 + \check{X}_1 \overline{\beta}_2 + \check{X}_0 \bar{c} n^{-\alpha} + \mathbf{e}^0 \right) \right]
$$

$$
= \left[\left(\frac{1}{n} \widehat{X}_1' \widehat{Z}_1 \right) W_{1n} \left(\frac{1}{n} \widehat{Z}_1' \widehat{X}_1 \right) \right]^{-1}
$$

$$
\times \left[\left(\frac{1}{n} \widehat{X}_1' \widehat{Z}_1 \right) W_{1n} \frac{1}{\sqrt{n}} \widehat{Z}_1' \left(\left(\check{X}_1 - \widehat{X}_1 \right) \overline{\beta}_1 - \check{X}_b \bar{c} n^{-\alpha} + \mathbf{e}^0 \right) \right]
$$

$$
= \left[\left(\frac{1}{n} \widehat{X}_1' \widehat{Z}_1 \right) W_{1n} \left(\frac{1}{n} \widehat{Z}_1' \widehat{X}_1 \right) \right]^{-1} \left[\left(\frac{1}{n} \widehat{X}_1' \widehat{Z}_1 \right) W_{1n} \frac{1}{\sqrt{n}} \widehat{Z}_1' \left(\widehat{\mathbf{r}} \kappa_1 - \check{X}_b \bar{c} n^{-\alpha} + \mathbf{e}^0 \right) \right],
$$

where $\check{X}_1$ and $\check{X}_b$ are matrices stacking $\check{\mathbf{x}}_i' 1(q_i \leq \widehat{\gamma})$ and $\check{\mathbf{x}}_i' (1(q_i \leq \widehat{\gamma}) - 1(q_i \leq \gamma_0))$, respectively. By the proof of Lemma C.1 in Appendix C, the continuity of $\mathbb{E} \left[\check{\mathbf{x}}_i \check{\mathbf{z}}_{i, \leq \gamma}' \right]$ in γ, and the consistency of $\widehat{\gamma}$, $\frac{1}{n} \widehat{X}_1' \widehat{Z}_1 \xrightarrow{p} \mathbb{E} \left[\check{\mathbf{x}}_i \check{\mathbf{z}}_{i, \leq \gamma_0}' \right] = G_1'$. So we concentrate on the asymptotic distributions of $\frac{1}{\sqrt{n}} \widehat{Z}_1' \widehat{\mathbf{r}}_1 \kappa_1$, $\frac{n^{-\alpha}}{\sqrt{n}} \sum_{i=1}^n \check{\mathbf{z}}_i \check{\mathbf{x}}_i' \bar{c} (1(q_i \leq \widehat{\gamma}) - 1(q_i \leq \gamma_0))$ and $\frac{1}{\sqrt{n}} \widehat{Z}_1' \mathbf{e}^0$. First,

$$
\frac{1}{\sqrt{n}} \widehat{Z}_1' \widehat{\mathbf{r}} \kappa_1 = \frac{1}{\sqrt{n}} \widehat{Z}_1' Z (\widehat{\Pi} - \Pi) \kappa_1 = \frac{1}{\sqrt{n}} \widehat{Z}_1' Z (Z'Z)^{-1} Z' V \kappa_1
$$

$$
= \left(\frac{1}{n} \widehat{Z}_1' Z \right) \left(\frac{1}{n} Z'Z \right)^{-1} \frac{1}{\sqrt{n}} Z' V \kappa_1,
$$

---

[12] The distribution in Theorem 2 of KST is incorrect: $\sqrt{\varphi}$ should be replaced by $1/\varphi$.

where $\frac{1}{n}\widehat{Z}_1'Z \xrightarrow{p} \mathbb{E}[\check{\mathbf{z}}_i\mathbf{z}_{i,\,\leq\gamma_0}']$, $\frac{1}{n}Z'Z \xrightarrow{p} \mathbb{E}[\mathbf{z}_i\mathbf{z}_i']$, and $\frac{1}{\sqrt{n}}Z'V\kappa_1 \xrightarrow{d} \mathcal{N}(\mathbf{0},\mathbb{E}[\mathbf{z}_i\mathbf{z}_i'(\mathbf{v}_i'\kappa_1)^2])$ with the influence function $\mathbf{z}_i\mathbf{v}_i'\kappa_1$. Second, by Lemma A.10 of Hansen (2000), uniformly on $\nu \in [-\bar{\nu},\bar{\nu}]$,

$$n^{-2\alpha}\sum_{i=1}^{n}\check{\mathbf{z}}_i\check{\mathbf{x}}_i'1\,(q_i \leq \gamma_0 + \nu/a_n) - 1\,(q_i \leq \gamma_0) = O_p(1).$$

Since $a_n(\widehat{\gamma} - \gamma_0) = O_p(1)$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\check{\mathbf{z}}_i\check{\mathbf{x}}_i'\bar{c}n^{-\alpha}\,(1\,(q_i \leq \widehat{\gamma}) - 1\,(q_i \leq \gamma_0)) = n^{-1/2}O_p\left(n^{2\alpha}\right)n^{-\alpha} = O_p(a_n^{-1/2}) = o_p(1).$$

Third, by Lemma A.4 of Hansen (2000),

$$\frac{1}{\sqrt{n}}\widehat{Z}_1'\mathbf{e}^0 \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\mathbb{E}\left[\check{\mathbf{z}}_i\check{\mathbf{z}}_{i,\,\leq\gamma_0}'e_{1i}^2\right]\right)$$

with the influence function $\check{\mathbf{z}}_{i,\,\leq\gamma_0}e_{1i}$. Because the two influence functions $\mathbf{z}_i\mathbf{v}_i'\kappa_1$ and $\check{\mathbf{z}}_{i,\,\leq\gamma_0}e_{1i}$ are uncorrelated, the asymptotic variance of $n^{1/2}\left(\widehat{\overline{\beta}}_1 - \overline{\beta}_1\right)$ is Σ_1.

To study the asymptotic covariance matrix between $\widehat{\overline{\beta}}_1$ and $\widehat{\overline{\beta}}_2$, note that the influence function for $\widehat{\overline{\beta}}_1$ is

$$\left(G_1'W_1G_1\right)^{-1}G_1'W_1\left(\mathbb{E}\left[\check{\mathbf{z}}_i\mathbf{z}_{i,\,\leq\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i\mathbf{v}_i'\kappa_1 + \check{\mathbf{z}}_{i,\,\leq\gamma_0}e_{1i}\right)$$

and for $\widehat{\overline{\beta}}_2$

$$\left(G_2'W_2G_2\right)^{-1}G_2'W_2\left(\mathbb{E}\left[\check{\mathbf{z}}_i\mathbf{z}_{i,\,>\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i\mathbf{v}_i'\kappa_2 + \check{\mathbf{z}}_{i,\,>\gamma_0}e_{2i}\right).$$

The second parts of the two influence functions are uncorrelated, so we need only consider the correlation between the first parts. Specifically, the covariance matrix is

$$\left(G_1'W_1G_1\right)^{-1}G_1'W_1\mathbb{E}\left[\check{\mathbf{z}}_i\mathbf{z}_{i,\,\leq\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\left(\mathbf{v}_i'\kappa_1\right)\left(\mathbf{v}_i'\kappa_2\right)\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\check{\mathbf{z}}_{i,\,>\gamma_0}'\right]$$
$$\times\, W_2G_2\left(G_2'W_2G_2\right)^{-1}.$$

The asymptotic independence between $\mathcal{N}(\mathbf{0},\Sigma_\ell)$ and $\zeta(\phi)$ follows from Proposition B.3 in Appendix B. $\square$

Appendix B. Propositions

PROPOSITION B.1. $\widehat{\gamma} \xrightarrow{p} \gamma_0$, $n^\alpha\left(\widehat{\overline{\beta}}_2 - \overline{\beta}_2\right) = o_p(1)$ and $n^\alpha\left(\widehat{\overline{\delta}}_n - \overline{\delta}_n\right) = o_p(1)$.

Proof. It is convenient to write the model as

$$y_i = \check{\mathbf{x}}_i'\overline{\beta}_2 + \check{\mathbf{x}}_{i,\,\leq\gamma_0}'\overline{\delta}_n + e_i^0,$$

and in matrix notation

$$Y = \check{X}\overline{\beta}_2 + \check{X}_0\bar{c}n^{-\alpha} + \mathbf{e}^0.$$

Define $P_\gamma^* = \widehat{X}_\gamma^* \left(\widehat{X}_\gamma^{*\prime}\widehat{X}_\gamma^*\right)^{-1}\widehat{X}_\gamma^{*\prime}$. Note that $P_\gamma^* = P_\gamma + P_\perp$ with $P_\gamma = \widehat{X}_\gamma \left(\widehat{X}_\gamma'\widehat{X}_\gamma\right)^{-1}\widehat{X}_\gamma'$ and $P_\perp = \widehat{X}_\perp \left(\widehat{X}_\perp'\widehat{X}_\perp\right)^{-1}\widehat{X}_\perp'$ since $\widehat{X}_\gamma'\widehat{X}_\perp = \mathbf{0}$. Because $\check{X} = \widehat{X} + (\mathbf{0},\widehat{\mathbf{r}})$, and $\widehat{X}$ is in the span of $\widehat{X}_\gamma^*$,

$$\left(I - P_\gamma^*\right)Y = \left(I - P_\gamma^*\right)\left(\check{X}_0\bar{c}n^{-\alpha} + \widehat{\mathbf{r}}\kappa_2 + \mathbf{e}^0\right) =: \left(I - P_\gamma^*\right)\left(\check{X}_0\bar{c}n^{-\alpha} + \widehat{\mathbf{e}}\right),$$

where $\widehat{\mathbf{e}}$ stacks $\widehat{e}_i = \widehat{r}_i'\kappa_2 + e_i^0$. The concentrated objective function is

$$\begin{aligned}
S_n(\gamma) &= Y'\left(I - P_\gamma^*\right)Y = \left(n^{-\alpha}\bar{c}'\check{X}_0' + \widehat{\mathbf{e}}'\right)\left(I - P_\gamma^*\right)\left(\check{X}_0\bar{c}n^{-\alpha} + \widehat{\mathbf{e}}\right) \\
&= \left(n^{-\alpha}\bar{c}'\check{X}_0' + \widehat{\mathbf{e}}'\right)\left(\check{X}_0\bar{c}n^{-\alpha} + \widehat{\mathbf{e}}\right) - \left(n^{-\alpha}\bar{c}'\check{X}_0' + \widehat{\mathbf{e}}'\right)P_\gamma^*\left(\check{X}_0\bar{c}n^{-\alpha} + \widehat{\mathbf{e}}\right).
\end{aligned}$$

Because the first term in the last expression does not depend on γ, and $\widehat{\gamma}$ minimizes $S_n(\gamma)$, we see that $\widehat{\gamma}$ maximizes

$$\begin{aligned}
S_n^*(\gamma) &= n^{2\alpha-1}\left(n^{-\alpha}\bar{c}'\check{X}_0' + \widehat{\mathbf{e}}'\right)P_\gamma^*\left(\check{X}_0\bar{c}n^{-\alpha} + \widehat{\mathbf{e}}\right) \\
&= n^{-1}\bar{c}'\check{X}_0'P_\gamma^*\check{X}_0\bar{c} + n^{\alpha-1}2\bar{c}'\check{X}_0'P_\gamma^*\widehat{\mathbf{e}} + n^{2\alpha-1}\widehat{\mathbf{e}}'P_\gamma^*\widehat{\mathbf{e}}.
\end{aligned}$$

For $\gamma \in \left[\gamma_0,\overline{\gamma}\right]$, $P_\perp\check{X}_0 = \mathbf{0}$, so

$$S_n^*(\gamma) = n^{-1}\bar{c}'\check{X}_0'P_\gamma\check{X}_0\bar{c} + n^{\alpha-1}2\bar{c}'\check{X}_0'P_\gamma\widehat{\mathbf{e}} + n^{2\alpha-1}\widehat{\mathbf{e}}'\left(P_\gamma + P_\perp\right)\widehat{\mathbf{e}}.$$

By Lemma C.1 in Appendix C, uniformly in $\gamma \in \left[\gamma_0,\overline{\gamma}\right]$,

$$\begin{aligned}
\frac{1}{n}\check{X}_0'P_\gamma\check{X}_0 &= \frac{1}{n}\check{X}_0'\widehat{X}_\gamma\left(\frac{1}{n}\widehat{X}_\gamma'\widehat{X}_\gamma\right)^{-1}\frac{1}{n}\widehat{X}_\gamma'\check{X}_0 \\
&= \frac{1}{n}\check{X}_0'\widehat{X}_0\left(\frac{1}{n}\widehat{X}_\gamma'\widehat{X}_\gamma\right)^{-1}\frac{1}{n}\widehat{X}_0'\check{X}_0 \xrightarrow{p} M_0M_\gamma^{-1}M_0,
\end{aligned}$$

$$\begin{aligned}
n^{\alpha-1}\check{X}_0'P_\gamma\widehat{\mathbf{e}} &= n^{\alpha-1/2}\frac{1}{n}X_0'\widehat{X}_0\left(\frac{1}{n}\widehat{X}_\gamma'\widehat{X}_\gamma\right)^{-1}\frac{1}{\sqrt{n}}\widehat{X}_\gamma'\widehat{\mathbf{e}} \\
&= n^{\alpha-1/2}O_p(1)O_p(1)O_p(1) = o_p(1),
\end{aligned}$$

$$\begin{aligned}
n^{2\alpha-1}\widehat{\mathbf{e}}'P_\gamma\widehat{\mathbf{e}} &= n^{2\alpha-1}\frac{1}{\sqrt{n}}\widehat{\mathbf{e}}'\widehat{X}_\gamma\left(\frac{1}{n}\widehat{X}_\gamma'\widehat{X}_\gamma\right)^{-1}\frac{1}{\sqrt{n}}\widehat{X}_\gamma'\widehat{\mathbf{e}} \\
&= n^{2\alpha-1}O_p(1)O_p(1)O_p(1) = o_p(1),
\end{aligned}$$

$$\begin{aligned}
n^{2\alpha-1}\widehat{\mathbf{e}}'P_\perp\widehat{\mathbf{e}} &= n^{2\alpha-1}\frac{1}{\sqrt{n}}\widehat{\mathbf{e}}'\widehat{X}_\perp\left(\frac{1}{n}\widehat{X}_\perp'\widehat{X}_\perp\right)^{-1}\frac{1}{\sqrt{n}}\widehat{X}_\perp'\widehat{\mathbf{e}} \\
&= n^{2\alpha-1}O_p(1)O_p(1)O_p(1) = o_p(1).
\end{aligned}$$

In summary, uniformly in $\gamma \in \left[\gamma_0,\overline{\gamma}\right]$, we have

$$S_n^*(\gamma) \xrightarrow{p} \bar{c}'M_0M_\gamma^{-1}M_0\bar{c},$$

which is uniquely maximized at γ_0, as shown in the proof of Lemma A.5 of Hansen (2000). Symmetrically, on $\gamma \in \left[\underline{\gamma},\gamma_0\right]$, $S_n^*(\gamma)$ converges uniformly to a limit function uniquely

maximized at γ_0. Because $\widehat{\gamma}$ maximizes $S_n^*(\gamma)$, it follows that $\widehat{\gamma} \xrightarrow{p} \gamma_0$ by Theorem 2.1 of Newey and McFadden (1994).

Now,

$$
n^\alpha \left[\left(\widetilde{\overline{\beta}}_2', \widetilde{\overline{\delta}}_n' \right)' - \left(\overline{\beta}_2', \overline{\delta}_n' \right)' \right] = n^\alpha \left(\widehat{X}^{*\prime} \widehat{X}^* \right)^{-1} \widehat{X}^{*\prime} (\check{X}\overline{\beta}_2 + \check{X}_0 \overline{c} n^{-\alpha} - \widehat{X}^* (\overline{\beta}_2', \overline{c}' n^{-\alpha})' + \mathbf{e}^0)
$$

$$
= \left(\frac{1}{n} \widehat{X}^{*\prime} \widehat{X}^* \right)^{-1} \frac{1}{n} \widehat{X}^{*\prime} ((\check{X} - \widehat{X}) n^\alpha \overline{\beta}_2 + (\check{X}_0 - \widehat{X}_{\widehat{\gamma}}) \overline{c} + n^\alpha \mathbf{e}^0)
$$

$$
= \left(\frac{1}{n} \widehat{X}^{*\prime} \widehat{X}^* \right)^{-1} \frac{1}{n} \widehat{X}^{*\prime} \left[n^\alpha \widehat{\mathbf{r}} \kappa_2 + (\check{X}_0 - \widehat{X}_{\widehat{\gamma}}) \overline{c} + n^\alpha \mathbf{e}^0 \right],
$$

where $\widehat{X}^* = \left(\widehat{X}, \widehat{X}_{\widehat{\gamma}} \right)$. From Lemma C.1 in Appendix C,

$$
\frac{n^\alpha}{n} \widehat{X}^{*\prime} \widehat{\mathbf{r}} \kappa_2 = n^{\alpha - 1/2} \frac{1}{\sqrt{n}} \widehat{X}^{*\prime} \widehat{\mathbf{r}} \kappa_2 = n^{\alpha - 1/2} O_p(1) = o_p(1),
$$

$$
\frac{n^\alpha}{n} \widehat{X}^{*\prime} \mathbf{e}^0 = n^{\alpha - 1/2} \frac{1}{\sqrt{n}} \widehat{X}^{*\prime} \mathbf{e}^0 = n^{\alpha - 1/2} O_p(1) = o_p(1).
$$

Also, by Lemma C.1 in Appendix C, the continuity of M_γ and the consistency of $\widehat{\gamma}$,

$$
\frac{1}{n} \widehat{X}^{*\prime} \left(\check{X}_0 - \widehat{X}_{\widehat{\gamma}} \right) \xrightarrow{p} \begin{pmatrix} M_0 \\ M_0 \end{pmatrix} - \begin{pmatrix} M_0 \\ M_0 \end{pmatrix} = \mathbf{0},
$$

and

$$
\frac{1}{n} \widehat{X}^{*\prime} \widehat{X}^* \xrightarrow{p} \begin{pmatrix} M & M_0 \\ M_0 & M_0 \end{pmatrix}.
$$

The determinant of the limit matrix of is $|M_0| |M - M_0| > 0$ by Assumption II.9. So by the CMT, $\left(\frac{1}{n} \widehat{X}^{*\prime} \widehat{X}^* \right)^{-1} \xrightarrow{p} \begin{pmatrix} M & M_0 \\ M_0 & M_0 \end{pmatrix}^{-1} > 0$. In summary, $n^\alpha \left[\left(\widetilde{\overline{\beta}}_2', \widetilde{\overline{\delta}}_n' \right)' - \left(\overline{\beta}_2', \overline{\delta}_n' \right)' \right] = O_p(1) o_p(1) = o_p(1)$. $\qquad\square$

PROPOSITION B.2. $a_n (\widehat{\gamma} - \gamma_0) = O_p(1)$.

Proof. First, note that

$$
\frac{S_n(\gamma) - S_n(\gamma_0)}{n^{1-2\alpha}(\gamma - \gamma_0)} = \frac{\widehat{\mathbf{e}}' \left(P_0^* - P_\gamma^* \right) \widehat{\mathbf{e}}}{n^{1-2\alpha}(\gamma - \gamma_0)} + 2 \frac{\widehat{\mathbf{e}}' \left(P_0^* - P_\gamma^* \right) \check{X}_0 \overline{c}}{n^{1-\alpha}(\gamma - \gamma_0)} + \frac{\overline{c}' \check{X}_0' \left(P_0^* - P_\gamma^* \right) \check{X}_0 \overline{c}}{n(\gamma - \gamma_0)}.
$$

From Lemma C.2 in Appendix C, uniformly on $\overline{v}/a_n \leq |\gamma - \gamma_0| \leq B$ for a sufficiently large $\overline{v}$, $\frac{\overline{c}' \check{X}_0' \left(P_0^* - P_\gamma^* \right) \check{X}_0 \overline{c}}{n(\gamma - \gamma_0)} \geq \frac{5}{6} d$, $\left| \frac{\widehat{\mathbf{e}}' \left(P_0^* - P_\gamma^* \right) \check{X}_0 \overline{c}}{n^{1-\alpha}(\gamma - \gamma_0)} \right| \leq \frac{d}{12}$, and $\left| \frac{\widehat{\mathbf{e}}' \left(P_0^* - P_\gamma^* \right) \widehat{\mathbf{e}}}{n^{1-2\alpha}(\gamma - \gamma_0)} \right| \leq \frac{d}{6}$ for some positive constant d with probability approaching 1, where B is a suitably chosen small positive constant from Proposition B.1 above, so

$$
\frac{S_n(\gamma) - S_n(\gamma_0)}{n^{1-2\alpha}(\gamma - \gamma_0)} \geq \frac{d}{2}
$$

on $\overline{v} \leq |\gamma - \gamma_0| \leq B$. Because $S_n(\widehat{\gamma}) - S_n(\gamma_0) \leq 0$, this implies that $P(a_n |\widehat{\gamma} - \gamma_0| > \overline{v})$ can be made arbitrarily small for n large enough. $\qquad\square$

PROPOSITION B.3. *On any compact set* $[-\bar{v}, \bar{v}]$,

$$Q_n(v) = S_n(\gamma_0) - S_n(\gamma_0 + v/a_n) \rightsquigarrow \begin{cases} 2\sqrt{\bar{c}'V_0^-\bar{c}f}W_1(-v) - \bar{c}'D_0\bar{c}f\,|v|, & \text{if } v \le 0, \\ 2\sqrt{\bar{c}'V_0^+\bar{c}f}W_2(v) - \bar{c}'D_0\bar{c}f\,|v|, & \text{if } v > 0, \end{cases}$$

where the $W_\ell(v)$ *are two independent standard Wiener processes on* $[0,\infty)$ *which are independent of* $\mathcal{N}(\mathbf{0}, \Omega_\ell)$.

Proof. Note that

$$Q_n(v) = \widehat{\mathbf{e}}'\left(P_v^* - P_0^*\right)\widehat{\mathbf{e}} + 2n^{-\alpha}\bar{c}'\check{X}_0'\left(P_v^* - P_0^*\right)\widehat{\mathbf{e}} + n^{-2\alpha}\bar{c}'\check{X}_0'\left(P_v^* - P_0^*\right)\check{X}_0\bar{c},$$

where $P_v^* = P_{\gamma_0+v/a_n}^*$ is the localized P_γ^*. Similarly, define $\check{X}_v = \check{X}_{\gamma_0+v/a_n}$ and $\Delta_i(v) = \Delta_i(\gamma_0 + v/a_n) = 1(q_i \le \gamma_0 + v/a_n) - 1(q_i \le \gamma_0)$. By Lemma C.3 in Appendix C,

$$Q_n(v) = -n^{-2\alpha}\sum_{i=1}^n \left(\bar{c}'\check{\mathbf{x}}_i\right)^2 |\Delta_i(v)| + 2n^{-\alpha}\bar{c}'\sum_{i=1}^n \check{\mathbf{x}}_i e_i^0 \Delta_i(v) + o_p(1),$$

uniformly on $v \in [-\bar{v}, \bar{v}]$. First, by Lemma A.10 of Hansen (2000),

$$\sup_{v\in[-\bar{v},\bar{v}]}\left| n^{-2\alpha}\sum_{i=1}^n \left(\bar{c}'\check{\mathbf{x}}_i\right)^2 |\Delta_i(v)| - \text{sign}(v)\bar{c}'\int_0^v D_{\gamma_0+v/a_n}f(\gamma_0+v/a_n)\,dv\bar{c}\right| \xrightarrow{p} 0.$$

So by the continuity of D_γ and $f(\gamma)$ at γ_0, we have

$$\sup_{v\in[-\bar{v},\bar{v}]}\left| n^{-2\alpha}\sum_{i=1}^n \left(\bar{c}'\check{\mathbf{x}}_i\right)^2 |\Delta_i(v)| - \bar{c}'D_0\bar{c}f\,|v|\right| \xrightarrow{p} 0.$$

We now show $n^{-\alpha}\bar{c}'\sum_{i=1}^n \check{\mathbf{x}}_i e_i^0 \Delta_i(v) \rightsquigarrow \begin{cases} \sqrt{\bar{c}'V_0^-\bar{c}f}W_1(-v), & \text{if } v \le 0, \\ \sqrt{\bar{c}'V_0^+\bar{c}f}W_2(v), & \text{if } v > 0, \end{cases}$ and is asymp-

totically independent of $\widehat{\theta}$. The weak convergence is shown in Lemma A.11 of Hansen (2000), so we focus on the asymptotic independence below.

For $v_1 < 0$ and $v_2 > 0$, define

$$S_i = \left(S_{1i}, S_{2i}, S_{3i}'\right)',$$

with

$$S_{1i} = n^{-\alpha}\bar{c}'\check{\mathbf{x}}_i e_i^0 \Delta_i(v_1), \quad S_{2i} = n^{-\alpha}\bar{c}'\check{\mathbf{x}}_i e_i^0 \Delta_i(v_2),$$

$$S_{3i} = \frac{1}{\sqrt{n}}\left(\kappa_1'\mathbf{v}_i\mathbf{z}_i', e_{1i}\check{\mathbf{z}}_{i,\le\gamma_0}', \kappa_2'\mathbf{v}_i\mathbf{z}_i', e_{2i}\check{\mathbf{z}}_{i,>\gamma_0}'\right)' =: \frac{1}{\sqrt{n}}s_{3i},$$

where S_{3i} is the asymptotic random component in $\widehat{\theta}$. By Taylor expansion, for x small, $\log(1+x) \sim x - x^2/2$. Taking exponentials and rearranging, we obtain the approximation, $\exp(x) \sim (1+x)\exp\left(x^2/2\right)$. Fix $t := (t_1, t_2, t_3')'$. Define

$$T_n = \prod_{i=1}^n \left[1 + \sqrt{-1}t'S_i\right], \quad V_n = \sum_{i=1}^n S_i S_i' \text{ and } V_{jn} = \sum_{i=1}^n S_{ji}S_{ji}', j = 1, 2, 3,$$

where by the ergodic theorem,

$$V_n - \mathbb{E}[V_n] \xrightarrow{p} \mathbf{0}.$$

Note that

$$
\begin{aligned}
\mathbb{E}[T_n] &= \mathbb{E}\big[\mathbb{E}[T_n|\mathcal{F}_{i-1}, \mathbf{x}_i]\big] \\
&= \mathbb{E}\Big[T_{n-1}\mathbb{E}\big[1 + \sqrt{-1}t'S_n|\mathcal{F}_{i-1}, \mathbf{x}_i\big]\Big] = \mathbb{E}[T_{n-1}] = \cdots = \mathbb{E}[T_1] = 1.
\end{aligned}
$$

Now,

$$
\begin{aligned}
\mathbb{E}\Big[\exp\Big\{\sqrt{-1}t'\sum_{i=1}^n S_i\Big\}\Big] &= \mathbb{E}\Big[\prod_{i=1}^n \exp\big\{\sqrt{-1}t'S_i\big\}\Big] \\
&\approx \mathbb{E}\Big[\prod_{i=1}^n\Big(1 + \sqrt{-1}t'S_i\Big)\exp\Big\{-\frac{1}{2}t'S_iS_i't\Big\}\Big] \\
&= E\Big[T_n\exp\Big\{-\frac{1}{2}t'V_nt\Big\}\Big] \approx \mathbb{E}\Big[T_n\exp\Big\{-\frac{1}{2}t'\mathbb{E}[V_n]t\Big\}\Big] \\
&= \mathbb{E}\Big[\exp\Big\{-\frac{1}{2}t'\mathbb{E}[V_n]t\Big\}\Big],
\end{aligned}
$$

where for a rigorous justification of the second approximation, see the proof of Theorem 3.2 of Hall and Heyde (1980). Note that

$$
\begin{aligned}
-\frac{1}{2}t'\mathbb{E}[V_n]t = &-\frac{t_1^2}{2}\overline{c}'\mathbb{E}\big[\check{\mathbf{x}}_i\check{\mathbf{x}}_i'e_{1i}^2|q_i = \gamma_0-\big]\overline{c}f|v_1| \\
&-\frac{t_2^2}{2}\overline{c}'\mathbb{E}\big[\check{\mathbf{x}}_i\check{\mathbf{x}}_i'e_{2i}^2|q_i = \gamma_0+\big]\overline{c}fv_2 - \frac{1}{2}t_3'\mathbb{E}\big[s_{3i}s_{3i}'\big]t_3' + o(1),
\end{aligned}
$$

where the cross term between S_{3i} and $S_{1i}(S_{2i})$ is $n \times O(n^{-1/2}n^{-\alpha}a_n^{-1}) = o(1)$. As a result, $n^{-\alpha}\overline{c}'\sum_{i=1}^n \check{\mathbf{x}}_ie_i^0\Delta_i(v_1)$, $n^{-\alpha}\overline{c}'\sum_{i=1}^n \check{\mathbf{x}}_ie_i^0\Delta_i(v_2)$, and $\widehat{\theta}$ are asymptotically independent. □

Appendix C. Lemmas

LEMMA C.1. *Uniformly in $\gamma \in \Gamma$,*

$$
\frac{1}{n}\sum_{i=1}^n \mathbf{v}_i\widehat{r}_i'1(q_i \le \gamma) = O_p(n^{-1/2}), \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\widehat{r}_i'1(q_i \le \gamma) = O_p(n^{-1/2}),
$$

$$
\frac{1}{n}\sum_{i=1}^n \widehat{r}_i\widehat{r}_i'1(q_i \le \gamma) = O_p(n^{-1}), \frac{1}{n}\sum_{i=1}^n \widehat{r}_ie_i^0 1(q_i \le \gamma) = O_p(n^{-1}),
$$

which implies that, uniformly in $\gamma \in \Gamma$,

$$
\frac{1}{n}\widehat{X}_\gamma'\widehat{X}_\gamma \xrightarrow{p} M_\gamma, \frac{1}{n}\widehat{X}_0'\check{X}_0 \xrightarrow{p} M_0 \text{ and } \frac{1}{\sqrt{n}}\widehat{X}_\gamma'\widehat{\mathbf{e}} = O_p(1). \tag{C.1}
$$

Proof. First, note that $\widehat{\mathbf{v}}_i = \mathbf{x}_i - \widehat{\Pi}'\mathbf{z}_i = \mathbf{x}_i - \Pi'\mathbf{z}_i - (\widehat{\Pi} - \Pi)'\mathbf{z}_i = \mathbf{v}_i - (\widehat{\Pi} - \Pi)'\mathbf{z}_i$, that is, $\widehat{r}_i = (\widehat{\Pi} - \Pi)'\mathbf{z}_i$. So

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{v}_i\widehat{r}_i'1\,(q_i \le \gamma) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{v}_i\mathbf{z}_i'1\,(q_i \le \gamma)\,(\widehat{\Pi} - \Pi)$$

$$= \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{v}_i\mathbf{z}_i'1\,(q_i \le \gamma)\right]O_p(n^{-1/2}) = O_p(n^{-1/2}),$$

where $\widehat{\Pi} - \Pi = O_p(n^{-1/2})$ if $\mathbb{E}[\mathbf{z}\mathbf{z}'] > 0$, $\mathbb{E}[\|\mathbf{z}\|^4] < \infty$, and $\mathbb{E}[\|\mathbf{v}\|^4] < \infty$ by the martingale difference (MD) CLT (see, e.g., Billingsley, 1961), and $\frac{1}{n}\sum_{i=1}^{n}\mathbf{v}_i\mathbf{z}_i'1\,(q_i \le \gamma) = O_p(1)$ from Lemma 1 of Hansen (1996). Similarly,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\widehat{r}_i'1\,(q_i \le \gamma) = O_p\left(n^{-1/2}\right) \text{ and } \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\widehat{r}_i'1\,(q_i \le \gamma) = O_p\left(n^{-1/2}\right).$$

Next,

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{r}_i\widehat{r}_i'1\,(q_i \le \gamma) = (\widehat{\Pi} - \Pi)'\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i'1\,(q_i \le \gamma)\,(\widehat{\Pi} - \Pi)$$

$$= O_p(n^{-1/2})\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i'1\,(q_i \le \gamma)\right]O_p(n^{-1/2}) = O_p\left(n^{-1}\right).$$

Finally,

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{r}_i e_i^0 1\,(q_i \le \gamma) = (\widehat{\Pi} - \Pi)'\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i e_i^0 1\,(q_i \le \gamma) = O_p(n^{-1/2})O_p(n^{-1/2}) = O_p(n^{-1}),$$

where $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i e_i^0 1\,(q_i \le \gamma) = O_p(1)$ uniformly in $\gamma \in \Gamma$ is from Lemma A.4 of Hansen (2000). The statements in (C.1) are implied by

$$\frac{1}{n}\widehat{V}_\gamma'\widehat{V}_\gamma \xrightarrow{p} \mathbb{E}[\mathbf{v}\mathbf{v}'1\,(q \le \gamma)],$$

$$\frac{1}{n}X_\gamma'\widehat{V}_\gamma \xrightarrow{p} \mathbb{E}[\mathbf{x}\mathbf{v}'1\,(q \le \gamma)], \frac{1}{n}X_\gamma'X_\gamma \xrightarrow{p} \mathbb{E}[\mathbf{x}\mathbf{x}'1\,(q \le \gamma)],$$

and

$$\frac{1}{\sqrt{n}}\widehat{V}_\gamma'\widehat{\mathbf{e}} = O_p(1), \frac{1}{\sqrt{n}}X_\gamma'\widehat{\mathbf{e}} = O_p(1).$$

Now,

$$
\begin{aligned}
\frac{1}{n}\widehat{V}'_\gamma\widehat{V}_\gamma &= \frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{v}}_i\widehat{\mathbf{v}}'_i 1\,(q_i \le \gamma) \\
&= \frac{1}{n}\sum_{i=1}^n \mathbf{v}_i\mathbf{v}'_i 1\,(q_i \le \gamma) - \frac{1}{n}\sum_{i=1}^n \left(\widehat{r}_i\mathbf{v}'_i + \mathbf{v}_i\widehat{r}'_i\right) 1\,(q_i \le \gamma) + \frac{1}{n}\sum_{i=1}^n \widehat{r}_i\widehat{r}'_i 1\,(q_i \le \gamma) \\
&= \frac{1}{n}\sum_{i=1}^n \mathbf{v}_i\mathbf{v}'_i 1\,(q_i \le \gamma) - O_p(n^{-1/2}) + O_p(n^{-1}) \xrightarrow{p} \mathbb{E}\left[\mathbf{v}_i\mathbf{v}'_i 1\,(q \le \gamma)\right],
\end{aligned}
$$

where the convergence is from Lemma 1 of Hansen (1996). By a similar argument, we can show

$$
\begin{aligned}
\frac{1}{n}X'_\gamma\widehat{V}_\gamma &= \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\widehat{\mathbf{v}}'_i 1\,(q_i \le \gamma) \\
&= \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{v}'_i 1\,(q_i \le \gamma) - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\widehat{r}'_i 1\,(q_i \le \gamma) \xrightarrow{p} \mathbb{E}\left[\mathbf{x}_i\mathbf{v}'_i 1\,(q_i \le \gamma)\right].
\end{aligned}
$$

The convergence $\frac{1}{n}X'_\gamma X_\gamma \xrightarrow{p} \mathbb{E}\left[\mathbf{xx}' 1\,(q \le \gamma)\right]$ is implied by Lemma 1 of Hansen (1996).
   Finally,

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\widehat{V}'_\gamma\widehat{\mathbf{e}} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (\mathbf{v}_i - \widehat{r}_i)\left(\widehat{r}'_i\kappa_2 + e_i^0\right) 1\,(q_i \le \gamma) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{v}_i e_i^0 1\,(q_i \le \gamma) - \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n \mathbf{v}_i\widehat{r}'_i 1\,(q_i \le \gamma)\right]\kappa_2 \\
&\quad - \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n \widehat{r}_i\widehat{r}'_i 1\,(q_i \le \gamma)\right]\kappa_2 - \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n \widehat{r}_i e_i^0 1\,(q_i \le \gamma)\right] \\
&= O_p(1) - O_p(1) - O_p\left(n^{-1/2}\right) - O_p\left(n^{-1/2}\right) = O_p(1),
\end{aligned}
$$

where the result $\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{v}_i e_i^0 1\,(q_i \le \gamma) = O_p(1)$ uniformly in $\gamma \in \Gamma$ in the last equality is from Lemma A.4 of Hansen (2000). Similarly,

$$
\begin{aligned}
\frac{1}{\sqrt{n}}X'_\gamma\widehat{\mathbf{e}} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{x}_i\left(\widehat{r}'_i\kappa_2 + e_i^0\right) 1\,(q_i \le \gamma) \\
&= \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\widehat{r}'_i 1\,(q_i \le \gamma)\right]\kappa_2 + \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{x}_i e_i^0 1\,(q_i \le \gamma) \\
&= O_p(1) + O_p(1) = O_p(1).
\end{aligned}
$$

$\square$

LEMMA C.2. *With probability approaching 1,* $\displaystyle\inf_{\bar{v}/a_n \leq |\gamma - \gamma_0| \leq B} \frac{\bar{c}'\check{X}_0'\left(P_0^* - P_\gamma^*\right)\check{X}_0\bar{c}}{n(\gamma - \gamma_0)} \geq \frac{5}{6}d,$

$\displaystyle\sup_{\bar{v}/a_n \leq |\gamma - \gamma_0| \leq B} \left| \frac{\widehat{\mathbf{e}}'\left(P_0^* - P_\gamma^*\right)\check{X}_0\bar{c}}{n^{1-\alpha}(\gamma - \gamma_0)} \right| \leq \frac{d}{12}$ *and* $\displaystyle\sup_{\bar{v}/a_n \leq |\gamma - \gamma_0| \leq B} \left| \frac{\widehat{\mathbf{e}}'\left(P_0^* - P_\gamma^*\right)\widehat{\mathbf{e}}}{n^{1-2\alpha}(\gamma - \gamma_0)} \right| \leq \frac{d}{6},$ *where d is a positive constant.*

Proof. The proof is exactly the same as the proof of (A.33), (A.34), and (A.35) of CH; just note that $\check{X}_0', \widehat{X}_\gamma, \widehat{X}_\perp$, and $\bar{c}$ play the roles of their $G_0, \widehat{Z}_\gamma, \widehat{Z}_\perp$, and c, respectively. Note also that the proof of CH requires $n^\alpha\left(\widehat{\bar{\beta}}_2 - \bar{\beta}_2\right) = o_p(1)$ and $n^\alpha\left(\widehat{\bar{\delta}}_n - \bar{\delta}_n\right) = o_p(1)$ which are shown in Proposition B.1 of Appendix B. $\square$

LEMMA C.3. *Uniformly for $v \in [-\bar{v}, \bar{v}]$,*

$$n^{-2\alpha}\bar{c}'\check{X}_0'\left(P_0^* - P_v^*\right)\check{X}_0\bar{c} = n^{-2\alpha}\sum_{i=1}^n \left(\bar{c}'\check{\mathbf{x}}_i\right)^2 |\Delta_i(v)| + o_p(1),$$

$$n^{-\alpha}\bar{c}'\check{X}_0'\left(P_0^* - P_v^*\right)\widehat{\mathbf{e}} = -n^{-\alpha}\bar{c}'\sum_{i=1}^n \check{\mathbf{x}}_i e_i^0 \Delta_i(v) + o_p(1),$$

$$\widehat{\mathbf{e}}'\left(P_0^* - P_v^*\right)\widehat{\mathbf{e}} = o_p(1).$$

Proof. The proof is the same as that of Lemma 5 of CH; just note that $\check{\mathbf{x}}_i, \widehat{e}_i$ and e_i^0 play the roles of their $\mathbf{g}_i, \widehat{v}_i$ and v_i, respectively. Be careful that v is assumed to be positive in Lemma 5 of CH; for $v < 0$, $n^{-2\alpha}\bar{c}'\check{X}_0'\left(P_0^* - P_v^*\right)\check{X}_0\bar{c} = -n^{-2\alpha}\sum_{i=1}^n \left(\bar{c}'\check{\mathbf{x}}_i\right)^2 \Delta_i(v)$, so here we use $|\Delta_i(v)|$ instead of $\Delta_i(v)$. $\square$

Supplementary Material

REFERENCES

Acemoglu, D., S. Johnson, & J.A. Robinson (2001) The colonial origins of comparative development: An empirical investigation. *American Economic Review* 91, 1369–1401.

Bai, J. (1997) Estimation of a change point in multiple regression models. *Review of Economics and Statistics* 79, 551–563.

Bhattacharya, P.K. & P.J. Brockwell (1976) The minimum of an additive process with applications to signal estimation and storage theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 37, 51–75.

Billingsley, P. (1961) The Lindeberg–Levy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.

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

Caner, M. & B.E. Hansen (2004) Instrumental variable estimation of a threshold model. *Econometric Theory* 20, 813–843.

Chan, K.S. (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics* 21, 520–533.

Chan, K.S. & R.S. Tsay (1998) Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika* 85, 413–426.

Christopoulos, D., P. McAdam, & E. Tzavalis (2021) Dealing with endogeneity in threshold models using copulas. *Journal of Business & Economic Statistics* 39, 166–178.

Egert, B. & A. Morales-Zumaquero (2008) Exchange rate regimes, foreign exchange volatility and export performance in central and eastern Europe: Just another blur project? *Review of Development Economics* 12, 577–593.

Frankel, J. & D. Romer (1999) Does trade cause growth? *American Economic Review* 89, 379–399.

Gouriéroux, C., A. Monfort, E. Renault, & A. Trognon (1987) Generalized residuals. *Journal of Econometrics* 34, 5–32.

Grier, K.B. & A.D. Smallwood (2013) Exchange rate shocks and trade: A multivariate GARCH-M approach. *Journal of International Money and Finance* 37, 282–305.

Hall, A.R., S. Han, & O. Boldea (2012) Inference regarding multiple structural changes in linear models with endogenous regressors. *Journal of Econometrics* 170, 281–302.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and its Application*. Academic Press.

Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.

Hansen, B.E. (2000) Sample splitting and threshold estimation. *Econometrica* 68, 575–603.

Hansen, B.E. (2011) Threshold autoregression in economics. *Statistics and Its Interface* 4, 123–127.

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

Heckman, J.J. (1979) Sample selection bias as a specification error. *Econometrica* 47, 153–161.

Kim, J. & D. Pollard (1990) Cube root asymptotics. *Annals of Statistics* 18, 191–219.

Kourtellos, A., T. Stengos, & Y. Sun (2022) Endogeneity in semiparametric threshold regression. *Econometric Theory* 38, 562–595.

Kourtellos, A., T. Stengos, & C.-M. Tan (2016) Structural threshold regression. *Econometric Theory* 32, 827–860.

Murphy, K.M. & R.H. Topel (1985) Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics* 3, 370–379.

Navarro, S. (2008) Control functions. In S.N. Durlauf & L.E. Blume (eds.), *The New Palgrave Dictionary of Economics*, *2nd* ed., pp. 20–28. Palgrave Macmillan.

Newey, W.K. (1984) A method of moments interpretation of sequential estimators. *Economics Letters* 14, 201–206.

Newey, W.K. & D.L. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Eagle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2111–2245. Elsevier Science B.V.

Papageorgiou, C. (2002) Trade as a threshold variable for multiple regimes. *Economics Letters* 77, 85–91.

Perron, P. & Z. Qu (2006) Estimating restricted structural change models. *Journal of Econometrics* 134, 373–399.

Perron, P. & Y. Yamamoto (2014) A note on estimating and testing for multiple structural changes in models with endogenous regressors via 2SLS. *Econometric Theory* 30, 491–507.

Petruccelli, J.D. (1992) On the approximation of time series by threshold autoregressive models. *Sankhya Series B* 54, 54–61.

Seo, M.H. & Y. Shin (2016) Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics* 195, 169–186.

Tan, C.M. (2010) No one true path: Uncovering the interplay between geography, institutions, and fractionalization in economic development. *Journal of Applied Econometrics* 25(7), 1100–1127.

Tenreyro, S. (2007) On the trade impact of nominal exchange rate volatility. *Journal of Development Economics* 82, 485–508.

Wooldridge, J.M. (2015) Control function methods in applied econometrics. *Journal of Human Resources* 50, 420–445.

Yu, P. (2012) Likelihood estimation and inference in threshold regression. *Journal of Econometrics* 167, 274–294.

Yu, P. (2013) Inconsistency of 2SLS estimators in threshold regression with endogeneity. *Economics Letters* 120, 532–536.

Yu, P. (2014) The bootstrap in threshold regression. *Econometric Theory* 30, 676–714.

Yu, P. (2015) Adaptive estimation of the threshold point in threshold regression. *Journal of Econometrics* 189, 83–100.

Yu, P., Q. Liao, & P.C.B. Phillips (2018) Inference and Specification Testing in Threshold Regression with Endogeneity. Cowles Foundation Discussion Paper No. 2209, Yale University.

Yu, P. & P.C.B. Phillips (2018) Threshold regression with endogeneity. *Journal of Econometrics* 203, 50–68.