Econometric Theory, **20**, 2004, 813–843. Printed in the United States of America.
DOI: 10.1017/S0266466604205011

INSTRUMENTAL VARIABLE ESTIMATION OF A THRESHOLD MODEL

MEHMET CANER
University of Pittsburgh

BRUCE E. HANSEN
University of Wisconsin

Threshold models (sample splitting models) have wide application in economics. Existing estimation methods are confined to regression models, which require that all right-hand-side variables are exogenous. This paper considers a model with endogenous variables but an exogenous threshold variable. We develop a two-stage least squares estimator of the threshold parameter and a generalized method of moments estimator of the slope parameters. We show that these estimators are consistent, and we derive the asymptotic distribution of the estimators. The threshold estimate has the same distribution as for the regression case (Hansen, 2000, *Econometrica* 68, 575–603), with a different scale. The slope parameter estimates are asymptotically normal with conventional covariance matrices. We investigate our distribution theory with a Monte Carlo simulation that indicates the applicability of the methods.

1. INTRODUCTION

Threshold models have some popularity in current applied econometric practice. The model splits the sample into classes based on the value of an observed variable—whether or not it exceeds some threshold. When the threshold is unknown (as is typical in practice) it needs to be estimated, and this increases the complexity of the econometric problem. A theory of estimation and inference is fairly well developed for linear models with exogenous regressors, including Chan (1993), Hansen (1996), Hansen (1999), Hansen (2000), and Caner (2002). These papers explicitly exclude the presence of endogenous variables, and this has been an impediment to empirical application, including dynamic panel models.

This paper develops an estimator and a theory of inference for linear models with endogenous variables and an exogenous threshold variable. We derive a

We thank the two referees and co-editor for constructive comments. Hansen thanks the National Science Foundation for financial support. Caner thanks University of Pittsburgh Central Research Development Fund for financial support. Address correspondence to Mehmet Caner, Department of Economics, WW Posvar Hall 4S01, Pittsburgh, PA 15260, USA; e-mail: caner@pitt.edu; or to Bruce E. Hansen, Department of Economics, 1180 Observatory Drive, Madison, WI 53706, USA; e-mail: bhansen@ssc.wisc.edu.

large sample distribution theory for the parameter estimates and test statistics. The estimator is based on estimation of a reduced form regression for the endogenous variables as a function of exogenous instruments. This requires the development of a model of the conditional mean of the endogenous variables as a function of the exogenous variables. Based on the reduced form, predicted values for the endogenous variables are formed and substituted into the structural equation of interest. Least-squares (LS) minimization yields the estimate of the threshold. Estimation of the slope parameters of this equation occurs in the third step, where the sample is split based on the estimated threshold, and then conventional two-stage least squares (2SLS) or generalized method of moments (GMM) estimation is performed on the subsamples.

Although we demonstrate that our estimator is consistent, we do not know if it is efficient, as other estimators are possible and efficiency is difficult to establish in nonregular models.

We make the important assumption that the threshold variable is exogenous. This is an important feature of the model and limits potential applications. In some cases, it may be desired to treat the threshold variable as endogenous. This would be a substantially different model and would require a distinct estimator. Hence we do not treat this case in this paper, and we leave it to future research.

Our asymptotic theory allows either for a random sample or for weakly dependent time-series data (thereby excluding trends and unit roots).

Our statistical analysis of our threshold estimator follows Hansen (2000) by using a "small threshold" asymptotic framework. Specifically, the difference in the regression slopes between regions is modeled as decreasing as the sample size grows. This device reduces the convergence rate of the threshold estimate and allows the development of a simpler distributional approximation. The analysis is therefore probably most relevant to empirical applications where the threshold effect is "small." The small threshold assumption was first developed in the change-point literature by Picard (1985) and Bai (1997).

An example of a potential application is q theory, which specifies that marginal q (the ratio of a firm's market value to replacement value) should be a sufficient predictor for firm investment. Empirical evidence has suggested that this relation is violated by financially constrained firms, whose investment is also affected by the level of cash flow. This literature, including Fazzari, Hubbard, and Petersen (1988), Barnett and Sakellaris (1998), Hu and Schiantarelli (1998), and Hansen (1999), uses threshold models for the investment equation where the threshold variables are measures of financial constraints. Erickson and Whited (2000) challenge this literature by arguing that these findings are artifacts of measurement error in marginal q. They use GMM estimates with sample splits based on firm size and bond ratings but do not estimate the threshold levels. Our methods would allow these split points (threshold parameters) to be estimated rather than fixed at arbitrary values.

The plan of the paper is as follows. In Section 2, we lay out the model. In Section 3, we describe our proposed estimators for the model parameters. Section 4 presents the asymptotic distribution theory. Section 5 discusses testing for a threshold. A Monte Carlo simulation is reported in Section 6. The conclusion is Section 7. The Appendix contains the proofs of the asymptotic distribution results.

A Gauss program that computes the statistics discussed in the paper is available at http://www.ssc.wisc.edu/~bhansen/.

2. MODEL

The observed sample is $\{y_i, z_i, x_i\}_{i=1}^n$, where y_i is real valued, z_i is an m-vector, and x_i is a k-vector with $k \geq m$. The threshold variable $q_i = q(x_i)$ is an element (or function) of the vector x_i and must have a continuous distribution. The data are either a random sample or a weakly dependent time series (so that unit roots and stochastic trends are excluded).

The structural equation of interest is

$$y_i = \theta_1' z_i + e_i, \qquad q_i \leq \gamma,$$

$$y_i = \theta_2' z_i + e_i, \qquad q_i > \gamma,$$

which also may be written in the form

$$y_i = \theta_1' z_i 1(q_i \leq \gamma) + \theta_2' z_i 1(q_i > \gamma) + e_i. \tag{1}$$

The threshold parameter is $\gamma \in \Gamma$ where Γ is a strict subset of the support of q_i. This parameter is assumed unknown and needs to be estimated.

The model allows the slope parameters θ_1 and θ_2 to differ depending on the value of q_i. The magnitude of the threshold effect is the difference between these parameters. Our statistical analysis of our threshold estimator will utilize a "small threshold" asymptotic framework, where $\delta_n = \theta_2 - \theta_1$ will tend to zero slowly as n diverges. We do not interpret this as a behavioral assumption but rather as a device for the construction of a useful asymptotic approximation.

The equation error is a martingale difference sequence

$$E(e_i | \Im_{i-1}) = 0, \tag{2}$$

where (x_i, z_i) are measurable with respect to $\Im_{i-1}$, the sigma field generated by $\{x_{i-j}, z_{i-j}, e_{i-1-j} : j \geq 0\}$. It is important that the error e_i satisfy this strong assumption, as simple orthogonality assumptions are insufficient to identify nonlinear models (including threshold models).

In the special case where $x_i = z_i$, (2) implies that (1) is a regression, but in general e_i may be correlated with z_i, so z_i is endogenous. It is important for our analysis and methods that the threshold variable q_i is treated as exogenous. Our methods do not generalize to the case of endogenous threshold variable, and different methods will need to be developed for that case.

The reduced form is a model of the conditional expectation of z_i given x_i:

$$z_i = g(x_i, \pi) + u_i, \tag{3}$$

$$E(u_i | x_i) = 0, \tag{4}$$

where π is a $p \times 1$ parameter vector, $g(\cdot,\cdot)$ maps $R^k \times R^p$ to R^m, and u_i is $m \times 1$. The function g is presumed known, whereas the parameter π is unknown. For simplicity, when performing the evaluation at the true value we will write

$g_i = g(x_i, \pi_0)$.

It will turn out to be useful to substitute (3) into (1), yielding

$$y_i = \theta_1' g_i 1(q_i \le \gamma) + \theta_2' g_i 1(q_i > \gamma) + v_i, \tag{5}$$

where

$$v_i = \theta_1' u_i 1(q_i \le \gamma) + \theta_2' u_i 1(q_i > \gamma) + e_i. \tag{6}$$

This will be important, as it turns out that the first-order asymptotic theory for our estimate of γ will behave as if it has been estimated directly from equation (5), i.e., as if the conditional mean g_i were observable. The error v_i (equation (6)) thus plays an important role in this distribution theory.

Our analysis will apply to several reduced form models. We explicitly provide regularity conditions for two examples. One is linear regression:

$$g(x_i, \pi) = \Pi' x_i, \tag{7}$$

where Π is $k \times m$. The second is threshold regression:

$$g(x_i, \pi) = \Pi_1' x_i 1(q_i \le \rho) + \Pi_2' x_i 1(q_i > \rho). \tag{8}$$

In the latter specification, the reduced form threshold parameter ρ may equal the threshold γ in the structural equation, but this is not necessary, and this restriction will not be used in estimation. For this model, our asymptotic analysis will assume that $\Pi_1 \ne \Pi_2$ are fixed parameters (in contrast to θ_1 and θ_2). Under these conditions Chan (1993) shows that the LS estimator $\hat{\rho}$ for ρ is $O(n^{-1})$ consistent. This fast rate of convergence is critical and is exploited in our theory. A reasonable interpretation is that if a threshold regression (8) is used for the reduced form, our theory is most appropriate if the latter is well identified with a large threshold effect.

3. ESTIMATION

We estimate the parameters sequentially. First, we estimate the reduced form parameter π by LS. Second, we estimate the threshold γ using predicted values of the endogenous variables z_i. Third, we estimate the slope parameters θ_1 and θ_2 by 2SLS or GMM on the split samples implied by the estimate of γ.

3.1. Reduced Form

It is helpful to partition $z_i = (z_{1i}, z_{2i})$ where $z_{2i} \in x_i$ are "exogenous" (a function of x_i) and z_{1i} are endogenous. Similarly, partition $g = (g_1, g_2)$, so that the reduced form parameters π enter only g_1.

Because (3) is a regression, the reduced form parameter π is estimated by LS. If there are no cross-equation restrictions (common parameters) in the m equations, this is equation-by-equation LS (for each variable in z_{1i}). If there are cross-equation restrictions, then the multivariate LS estimator solves

$$\hat{\pi} = \underset{\pi}{\operatorname{argmin}} \det\left(\sum_{i=1}^{n} (z_{1i} - g_1(x_i, \pi))(z_{1i} - g_1(x_i, \pi))' \right). \tag{9}$$

Given $\hat{\pi}$, the predicted values for z_i are

$$\hat{z}_i = \hat{g}_i = g(x_i, \hat{\pi}).$$

For example, in the threshold regression model (8) the threshold parameter ρ is common across equations—a cross-equation restriction—so the multivariate estimator (9) is appropriate. The solution is found as follows. For each $\rho \in \Gamma$ define

$$\hat{\Pi}_1(\rho) = \left(\sum_{i=1}^{n} x_i x_i' 1(q_i \le \rho) \right)^{-1} \sum_{i=1}^{n} x_i z_{1i}' 1(q_i \le \rho),$$

$$\hat{\Pi}_2(\rho) = \left(\sum_{i=1}^{n} x_i x_i' 1(q_i > \rho) \right)^{-1} \sum_{i=1}^{n} x_i z_{1i}' 1(q_i > \rho),$$

$$\hat{u}_i(\rho) = z_{1i} - \hat{\Pi}_1(\rho)' x_i 1(q_i \le \rho) - \hat{\Pi}_2(\rho)' x_i 1(q_i > \rho).$$

Then we obtain the LS estimates (9) by minimization of the concentrated LS criterion:

$$\hat{\rho} = \underset{\rho \in \Gamma}{\operatorname{argmin}} \det\left(\sum_{i=1}^{n} \hat{u}_i(\rho) \hat{u}_i(\rho)' \right),$$

$$\hat{\Pi}_1 = \hat{\Pi}_1(\hat{\rho}),$$

$$\hat{\Pi}_2 = \hat{\Pi}_2(\hat{\rho}).$$

For this model of the reduced form, the predicted values are

$$\hat{z}_i = \hat{g}_i = \hat{\Pi}_1' x_i 1(q_i \le \hat{\rho}) + \hat{\Pi}_2' x_i 1(q_i > \hat{\rho}).$$

3.2. Threshold Estimation

We now turn to estimation of the threshold γ in the structural equation. For any γ, let Y, $\hat{Z}_\gamma$, and $\hat{Z}_\perp$ denote the matrices of stacked vectors y_i, $\hat{z}_i' 1(q_i \le \gamma)$, and $\hat{z}_i' 1(q_i > \gamma)$, respectively. Let $S_n(\gamma)$ denote the LS residual sum of squared errors from a regression of Y on $\hat{Z}_\gamma$ and $\hat{Z}_\perp$. Our 2SLS estimator for γ is the minimizer of the sum of squared errors:

$$\hat{\gamma} = \operatorname*{argmin}_{\gamma \in \Gamma} S_n(\gamma).$$

As a by-product of estimation, we obtain natural test statistics for hypotheses on γ, which take the form $H_0: \gamma = \gamma_0$. Following Hansen (2000), we consider the LR-like statistic

$$LR_n(\gamma) = n \frac{S_n(\gamma) - S_n(\hat{\gamma})}{S_n(\hat{\gamma})}.$$

3.3. Slope Estimation

Given the estimate $\hat{\gamma}$ of the threshold γ, the sample can be split into two subsamples, based on the indicators $1(q_i \le \hat{\gamma})$ and $1(q_i > \hat{\gamma})$. The slope parameters θ_1 and θ_2 can then be estimated by 2SLS or GMM separately on each subsample. We focus our discussion on the case where the reduced form is linear in x_i in each subsample.

Let $\hat{X}_1$, $\hat{X}_2$, $\hat{Z}_1$, and $\hat{Z}_2$ denote the matrices of stacked vectors $x_i'1(q_i \le \hat{\gamma})$, $x_i'1(q_i > \hat{\gamma})$, $z_i'1(q_i \le \hat{\gamma})$, and $z_i'1(q_i > \hat{\gamma})$, respectively. The 2SLS estimators for θ_1 and θ_2 are

$$\tilde{\theta}_1 = (\hat{Z}_1'\hat{X}_1(\hat{X}_1'\hat{X}_1)^{-1}\hat{X}_1'\hat{Z}_1)^{-1}(\hat{Z}_1'\hat{X}_1(\hat{X}_1'\hat{X}_1)^{-1}\hat{X}_1'Y),$$

$$\tilde{\theta}_2 = (\hat{Z}_2'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{Z}_2)^{-1}(\hat{Z}_2'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'Y).$$

The residual from this equation is

$$\tilde{e}_i = y_i - z_i'\tilde{\theta}_1 1(q_i \le \hat{\gamma}) - z_i'\tilde{\theta}_2 1(q_i > \hat{\gamma}).$$

Construct the weight matrices

$$\tilde{\Omega}_1 = \sum_{i=1}^{n} x_i x_i' \tilde{e}_i^2 1(q_i \le \hat{\gamma}),$$

$$\tilde{\Omega}_2 = \sum_{i=1}^{n} x_i x_i' \tilde{e}_i^2 1(q_i > \hat{\gamma}).$$

The GMM estimators for θ_1 and θ_2 are

$$\hat{\theta}_1 = (\hat{Z}_1'\hat{X}_1\tilde{\Omega}_1^{-1}\hat{X}_1'\hat{Z}_1)^{-1}(\hat{Z}_1'\hat{X}_1\tilde{\Omega}_1^{-1}\hat{X}_1'Y), \tag{10}$$

$$\hat{\theta}_2 = (\hat{Z}_2'\hat{X}_2\tilde{\Omega}_2^{-1}\hat{X}_2'\hat{Z}_2)^{-1}(\hat{Z}_2'\hat{X}_2\tilde{\Omega}_2^{-1}\hat{X}_2'Y). \tag{11}$$

The estimated covariance matrices for the GMM estimators are

$$\hat{V}_1 = (\hat{Z}_1'\hat{X}_1\tilde{\Omega}_1^{-1}\hat{X}_1'\hat{Z}_1)^{-1}, \tag{12}$$

$$\hat{V}_2 = (\hat{Z}_2'\hat{X}_2\tilde{\Omega}_2^{-1}\hat{X}_2'\hat{Z}_2)^{-1}. \tag{13}$$

We now briefly discuss the asymptotic efficiency of these estimators. As we show in Section 4.3, they are asymptotically equivalent to their ideal counterparts constructed with the unknown true value of γ rather than the estimated value $\hat{\gamma}$, and so for the purposes of asymptotic efficiency, we can examine the case of known γ. The GMM estimators $(\hat{\theta}_1, \hat{\theta}_2)$ are easily seen to be efficient estimators of (θ_1, θ_2) (in the sense of Chamberlain, 1987) under the moment conditions

$$\begin{aligned} &E(x_i e_i 1(q_i \le \gamma)) = 0, \\ &E(x_i e_i 1(q_i > \gamma)) = 0. \end{aligned} \tag{14}$$

These equations are implied by the assumed multidimensional scaling (MDS) assumption (2) but do not in general exhaust its implications, suggesting that our GMM estimator is not fully efficient. However, under the leading assumption of conditional homoskedasticity $E(e_i^2|\Im_{i-1}) = \sigma^2$, both the 2SLS and GMM estimators achieve the semiparametric efficiency bound.

4. DISTRIBUTION THEORY

4.1. Assumptions

Define the moment functionals

$$M(\gamma) = E(g_i g_i' 1(q_i \le \gamma)),$$

$$D_1(\gamma) = E(g_i g_i' | q_i = \gamma),$$

and

$$D_2(\gamma) = E(g_i g_i' v_i^2 | q_i = \gamma).$$

Let $f(q)$ denote the density function of q_i, γ_0 denote the true value of γ, $D_1 = D_1(\gamma_0)$, $D_2 = D_2(\gamma_0)$, $f = f(\gamma_0)$, and $M = E(g_i g_i')$.

Assumption 1.

1. (x_i, g_i, e_i, u_i) is strictly stationary and ergodic with ρ mixing coefficients $\sum_1^\infty \rho_m^{1/2} < \infty$;
2. $E(e_i|\Im_{i-1}) = 0$;
3. $E(u_i|\Im_{i-1}) = 0$;
4. $E|g_i|^4 < \infty$ and $E|g_i v_i|^4 < \infty$;
5. for all $\gamma \in \Gamma$, $E(|g_i|^4 v_i^4 | q_i = \gamma) \le C$ and $E(|g_i|^4 | q_i = \gamma) \le C$ for some $C < \infty$;
6. for all $\gamma \in \Gamma$, $f(\gamma) \le \bar{f} < \infty$;

7. $f(\gamma)$, $D_1(\gamma)$, and $D_2(\gamma)$ are continuous at $\gamma = \gamma_0$;
8. $\delta_n = \theta_1 - \theta_2 = cn^{-\alpha}$ with $c \neq 0$ and $0 < \alpha < \frac{1}{2}$;
9. $c'D_1c > 0$, $c'D_2c > 0$, and $f > 0$;
10. $M > M(\gamma) > 0$ for all $\gamma \in \Gamma$.

Assumption 1.1 is relevant for time series applications and is trivially satisfied for independent observations. The assumption of stationarity excludes time trends and integrated processes. Assumptions 1.2 and 1.3 impose the correct specification of the conditional mean in the structural equation and reduced form. Assumptions 1.4 and 1.5 are unconditional and conditional moment bounds. Assumptions 1.6 and 1.7 require the threshold variable to have a continuous distribution and essentially require the conditional variance $E(v_i^2|q_i = \gamma)$ to be continuous at γ_0, which excludes regime-dependent heteroskedasticity. Assumption 1.8 is the small threshold effect assumption. Assumptions 1.9 and 1.10 are full rank conditions needed to have nondegenerate asymptotic distributions.

We require that the reduced form predicted values be consistent for the true reduced form conditional mean. Let

$$\hat{r}_i = g_i - \hat{g}_i$$

denote the estimation error from the reduced form estimation. The following high-level conditions are sufficient for our theory. Let $a_n = n^{1-2\alpha}$.

Assumption 2. Let $H_i = \{g_i, v_i, \hat{r}_i\}$. First,

$$\sup_{\gamma\in\Gamma}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} H_i \hat{r}_i' 1(q_i \le \gamma)\right| = O_p(1). \tag{15}$$

Second, there exists a $0 < B < \infty$ such that for all $\varepsilon > 0$ and $\delta > 0$, there is a $\bar{v} < \infty$ and $\bar{n} < \infty$ such that for all $n \ge \bar{n}$,

$$P\left(\sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B}\left|\frac{\sum_{i=1}^{n} H_i \hat{r}_i'(1(q_i \le \gamma) - 1(q_i \le \gamma_0))}{n^{1-\alpha}|\gamma - \gamma_0|}\right| > \delta\right) < \varepsilon. \tag{16}$$

Third,

$$\sup_{|\nu|\le\bar{v}} n^{-\alpha}\left|\sum_{i=1}^{n} H_i \hat{r}_i'(1(q_i \le \gamma_0 + \nu/a_n) - 1(q_i \le \gamma_0))\right| \to_p 0. \tag{17}$$

We can show that Assumption 2 holds for important cases of both linear and threshold reduced form models.

Lemma 1. *If $g(x_i, \pi)$ takes the linear form (7), or if $g(x_i, \pi)$ takes the threshold form (8) with $\Pi_2 \neq \Pi_1$, and Assumption 1 holds, then Assumption 2 holds.*

4.2. Threshold Estimate

Let

$$\sigma_v^2 = Ev_i^2,$$

$$\omega = \frac{c'D_2 c}{(c'D_1 c)^2 f},$$

$$\eta^2 = \frac{c'D_2 c}{\sigma_v^2 c'D_1 c}.$$

In the leading case of conditional homoskedasticity $E(v_i^2|x_i) = \sigma_v^2$, then these constants simplify as follows:

$$\omega = \frac{\sigma_v^2}{c'D_1 cf},$$

$$\eta^2 = 1.$$

Let $W(r)$ denote a two-sided Brownian motion on the real line. Define the random variables

$$T = \underset{-\infty<r<\infty}{\text{argmax}}\left(-\frac{1}{2}|r| + W(r)\right)$$

and

$$\xi = \sup_{r\in R}(-|r| + 2W(r)).$$

THEOREM 1. *Under Assumptions 1 and 2,*

$$n^{1-2\alpha}(\hat{\gamma} - \gamma_0) \xrightarrow{d} \omega T, \tag{18}$$

$$LR(\gamma_0) \xrightarrow{d} \eta^2 \xi. \tag{19}$$

The rate of convergence $n^{1-2\alpha}$ and asymptotic distribution T for the threshold estimate shown in (18) are the same as in LS estimation of threshold regression models (see Hansen, 2000, Theorem 1). The main difference is that in the 2SLS case the scale ω (which inversely determines precision) is proportional to the variance of v_i (the variance of the error (6) from equation (5)) rather than that of the equation error e_i and is inversely proportional to the conditional design matrix $E(g_i g_i'|q_i = \gamma_0)$ of g_i, the conditional expectation of regressors z_i given the instruments x_i, rather than the conditional design of the regressors. The distribution function of T is derived by Bhattacharya and Brockwell (1976).

Theorem 1 makes the crucial assumption that $\delta_n = \theta_1 - \theta_2 = cn^{-\alpha} \to 0$ as $n \to \infty$. This is the small threshold effect assumption. In contrast, if we assume $\delta_n \neq 0$ fixed, then the convergence rate for $\hat{\gamma}$ is $O(n^{-1})$ as shown for LS estimation by Chan (1993). However, the asymptotic distribution for $n(\hat{\gamma} - \gamma_0)$ is quite complicated and not useful for inference on γ. By adopting the small threshold effect assumption, we effectively slow down the rate of convergence from n to $n^{1-2\alpha}$, which allows asymptotic averaging to simplify the sampling distribution.

The asymptotic distribution of the LR-like test $LR(\gamma_0)$ in (19) takes the same form as for the LS case. If the error v_i is homoskedastic, then $\eta^2 = 1$ and the asymptotic distribution is free of nuisance parameters, facilitating testing and confidence interval construction. Otherwise, η^2 can be estimated as in Section 3.4 of Hansen (2000). The distribution function of ξ is derived in Theorem 2 of Hansen (2000) and is $P(\xi \le x) = (1 - e^{-x/2})^2$. Some critical values are provided in Table 1 of Hansen (2000).

To form an asymptotic confidence interval for γ we use the test-inversion method advocated by Hansen (2000). Let C be the 95% percentile of the distribution of ξ. The most straightforward method assumes that the errors v_i are homoskedastic and then sets

$$\hat{\Gamma} = \{\gamma : LR_n(\gamma) \le C\},$$

the set of values of γ such that the LR-like statistic is below the 5% asymptotic critical value. For a confidence region robust to heteroskedasticity, set

$$\hat{\Gamma} = \{\gamma : LR_n(\gamma) \le \hat{\eta}^2 C\},$$

where $\hat{\eta}^2$ is an estimate of η^2. Theorem 1 shows that $\hat{\Gamma}$ is an asymptotically valid 95% confidence region for γ_0.

A useful method to visually assess the estimator $\hat{\gamma}$ and its precision is to plot $LR_n(\gamma)$ against γ. The point where $LR_n(\gamma)$ strikes zero is the estimator $\hat{\gamma}$. The points where $LR_n(\gamma) \le C$ are the points in the confidence region. The parameter γ_0 is more precisely estimated the more "peaked" is the graph of $LR_n(\gamma)$. In samples with strong information about γ, $LR_n(\gamma)$ will tend to have a sharp V shape with clearly delineated minimum. In samples with low information about γ, $LR_n(\gamma)$ will tend to have a more irregular shape with less clearly defined minimum.

4.3. Slope Parameters

We first state the asymptotic distribution of the 2SLS slope estimators.

THEOREM 2. *Under Assumptions 1 and 2,*

$$n^{1/2}(\tilde{\theta}_1 - \theta_1) \xrightarrow{d} N(0, V_1^{2SLS}),$$

$$n^{1/2}(\tilde{\theta}_2 - \theta_2) \xrightarrow{d} N(0, V_2^{2SLS}),$$

where

$$V_1^{2SLS} = (R_1' Q_1 R_1)^{-1} R_1' Q_1 \Omega_1 Q_1 R_1 (R_1' Q_1 R_1)^{-1},$$
$$V_2^{2SLS} = (R_2' Q_2 R_2)^{-1} R_2' Q_2 \Omega_2 Q_2 R_2 (R_2' Q_2 R_2)^{-1},$$
$$Q_1 = E(x_i x_i' 1(q_i \leq \gamma_0)),$$
$$Q_2 = E(x_i x_i' 1(q_i > \gamma_0)),$$
$$R_1 = E(x_i z_i' 1(q_i \leq \gamma_0)),$$
$$R_2 = E(x_i z_i' 1(q_i > \gamma_0)),$$
$$\Omega_1 = E(x_i x_i' e_i^2 1(q_i \leq \gamma_0)),$$
$$\Omega_2 = E(x_i x_i' e_i^2 1(q_i > \gamma_0)).$$

Second, we give the asymptotic distribution of the GMM slope estimators.

THEOREM 3. *Under Assumptions 1 and 2*

$$n^{1/2}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} N(0, V_1),$$
$$n^{1/2}(\hat{\theta}_2 - \theta_2) \xrightarrow{d} N(0, V_2),$$

where

$$V_1 = (R_1' \Omega_1^{-1} R_1)^{-1},$$
$$V_2 = (R_2' \Omega_2^{-1} R_2)^{-1}.$$

Furthermore,

$$n\hat{V}_1 \rightarrow_p V_1,$$
$$n\hat{V}_2 \rightarrow_p V_2.$$

Theorems 2 and 3 give the asymptotic distributions of the 2SLS and GMM estimators of the slope coefficients under the small threshold effect assumption. It is not hard to see that if instead we make the assumption that $\theta_1 - \theta_2 = \delta$ is fixed with sample size, then the results are unaltered.

5. TESTING FOR A THRESHOLD

In model (1), the threshold effect disappears under the hypothesis

$$H_0 : \theta_1 = \theta_2.$$

To test H_0 we recommend an extension of the Davies (1977) Sup test to the GMM framework.

The statistic is formed as follows. First, fix $\gamma \in \Gamma$ at any value. Given this fixed threshold, estimate the model (1) by GMM under the moment conditions

(14). These estimators take the form (10) and (11) except that they are evaluated at this fixed value of γ rather than $\hat{\gamma}$. Corresponding to these estimates are their estimated covariance matrices $\hat{V}_1(\gamma)$ and $\hat{V}_2(\gamma)$, which take the form (12) and (13) except that again they are evaluated at γ rather than $\hat{\gamma}$. Then, still for this fixed value of γ, the Wald statistic for H_0 is

$$W_n(\gamma) = (\hat{\theta}_1(\gamma) - \hat{\theta}_2(\gamma))'(\hat{V}_1(\gamma) + \hat{V}_2(\gamma))^{-1}(\hat{\theta}_1(\gamma) - \hat{\theta}_2(\gamma)).$$

This calculation is repeated for all $\gamma \in \Gamma$. The Davies Sup statistic for H_0 is then the largest value of these statistics:

$$\text{SupW} = \sup_{\gamma \in \Gamma} W_n(\gamma).$$

We now present the asymptotic null distribution of this statistic. Define

$$\Omega_1(\gamma) = E(x_i x_i' e_i^2 1(q_i \le \gamma)),$$

$$Q_1(\gamma) = E(x_i z_i' 1(q_i \le \gamma)),$$

$$V_1(\gamma) = (Q_1(\gamma)'\Omega_1(\gamma)^{-1} Q_1(\gamma))^{-1},$$

$$\Omega_2(\gamma) = E(x_i x_i' e_i^2 1(q_i > \gamma)),$$

$$Q_2(\gamma) = E(x_i z_i' 1(q_i > \gamma)),$$

$$V_2(\gamma) = (Q_2(\gamma)'\Omega_2(\gamma)^{-1} Q_2(\gamma))^{-1}.$$

Let $S_1(\gamma)$ be a mean-zero Gaussian process with covariance kernel $E(S_1(\gamma_1)S_1(\gamma_2)') = \Omega_1(\gamma_1 \wedge \gamma_2)$, let $S = \text{plim}_{\gamma \to \infty} S_1(\gamma)$, and let $S_2(\gamma) = S - S_1(\gamma)$. Following the analysis of Davies (1977), Andrews and Ploberger (1994), and Hansen (1996), we have the following theorem. The proof is in the Appendix.

THEOREM 4. *Under Assumption 1 plus the null hypothesis* $\theta_1 = \theta_2$,

$$\begin{aligned}\text{SupW} \to_d \sup_{\gamma \in \Gamma} &(S_1(\gamma)'\Omega_1(\gamma)^{-1} Q_1(\gamma) V_1(\gamma) - S_2(\gamma)'\Omega_2(\gamma)^{-1} Q_2(\gamma) V_2(\gamma)) \\ &\times (V_1(\gamma) + V_2(\gamma))^{-1} \\ &\cdot (V_1(\gamma) Q_1(\gamma)'\Omega_1(\gamma)^{-1} S_1(\gamma) - V_2(\gamma) Q_2(\gamma)'\Omega_2(\gamma)^{-1} S_2(\gamma)).\end{aligned}$$

Because the parameter γ is not identified under the null hypothesis, this asymptotic distribution is not chi-square but can be written as the supremum of a chi-square process. This asymptotic distribution is nonpivotal but easily can be calculated by simulation. The argument presented in Hansen (1996) extends to the present case. Define the pseudodependent variable $y_i^* = \hat{e}_i(\gamma)\eta_i$, where $\hat{e}_i(\gamma)$ is the estimated residual under the unrestricted model for each γ, and η_i is independent and identically distributed (i.i.d.) $N(0,1)$. Then when we repeat the calculation presented previously using this pseudodependent variable in place of y_i, the resulting statistic SupW* has the same asymptotic distribution[1] as SupW. Thus by repeated simulation draws, the asymptotic p-value of the statistic SupW can be calculated with arbitrary accuracy.

6. MONTE CARLO SIMULATION

6.1. Model

The structural and reduced form equations are

$$y_i = \theta_1' z_i 1(q_i \leq \gamma) + \theta_2' z_i 1(q_i > \gamma) + e_i,$$

$$z_i = \begin{pmatrix} 1 \\ z_{1i} \end{pmatrix},$$

$$z_{1i} = (\pi_{11} + \pi_{12} x_i) 1(q_i \leq \rho) + (\pi_{21} + \pi_{22} x_i) 1(q_i > \rho) + u_i.$$

The structural equation has a single endogenous variable z_{1i} and a single excluded exogenous variable x_i.

We generate the exogenous variables as $x_i \sim N(0,1)$ and $q_i \sim N(2,1)$, and we generate the errors as $u_i \sim N(0,1)$ and $e_i = 0.5u_i$. Making e_i perfectly correlated with u_i is an extreme specification for endogeneity and is done to illustrate the robustness of the results to extreme settings.

We set the reduced form parameters as $\rho = 2$, $\pi_{11} = 1$, $\pi_{12} = 2$, $\pi_{21} = 1$, and $\pi_{22} = 1$. In the structural equation, we set $\gamma = 2$. The statistics we report depend on θ_1 and θ_2 only through the difference $\delta = \theta_1 - \theta_2 = (\delta_1, \delta_2)'$. We set $\delta_1 = 1$ and vary δ_2 and the sample size n. All results are based on 1,000 simulation replications.

For each simulated sample, we estimate the threshold reduced form by LS, substitute the predicted values of the endogenous variable z_{1i} into the structural equation, and then estimate the structural equation threshold by LS and finally estimate the slopes by GMM, as described in Sections 3.1–3.3.

6.2. Threshold Estimation

We first assess the performance of the threshold estimator $\hat{\gamma}$. Table 1 reports the 5%, 50%, and 95% quantiles of the simulation distribution of $\hat{\gamma}$, varying δ_2

TABLE 1. Quantiles of $\hat{\gamma}$ distribution, $\gamma = 2$

	$\delta_2 = 0.25$			$\delta_2 = 1$			$\delta_2 = 2$		
Quantiles	5	50	95	5	50	95	5	50	95
$n = 100$	−0.08	1.35	3.78	1.13	1.97	2.16	1.83	1.98	2.06
$n = 250$	−0.03	1.66	3.70	1.82	1.99	2.05	1.94	1.99	2.02
$n = 500$	0.05	1.97	3.21	1.95	2.00	2.04	1.97	2.00	2.01

TABLE 2. Nominal 90% confidence interval coverage for γ

δ_2	0.25	0.5	1.0	1.5	2.0
$n = 50$	76	86	92	95	97
$n = 100$	73	88	96	98	98
$n = 250$	80	94	98	98	99
$n = 500$	86	96	99	98	98
$n = 1{,}000$	92	97	98	99	98

among 0.25, 1, and 2 and n among 100, 250, and 500. Performance improves, as expected, as δ_2 and/or n increases. In particular, we observe that for a small threshold effect ($\delta_2 = 0.25$), the distribution of $\hat{\gamma}$ is quite dispersed, whereas a large threshold effect ($\delta_2 = 2$) yields a tight sampling distribution, even for a sample size as small as $n = 100$.

Second, we assess the performance of our proposed confidence interval $\hat{\Gamma}$ for γ. Table 2 reports simulated coverage probabilities of a nominal 90% interval $\hat{\Gamma}$, constructed without a correction for heteroskedasticity, varying δ_2 among 0.25, 0.5, 1.0, 1.5, and 2.0 and n among 50, 100, 250, 500, and 1,000. We see that for any value of δ_2, the coverage probability increases as n increases, becoming fairly conservative for the large sample sizes. Similarly, for fixed n, the coverage probability increases as δ_2 increases. These findings do not contradict the distribution theory of Theorem 1, as that result requires that the threshold effect δ_2 decreases as n increases, which implies taking a diagonal path in Table 2 roughly from the upper right toward the lower left, where the coverage probabilities indeed fall close to the nominal 90% level. Interestingly, the results in Table 2 are consistent with Theorem 3 of Hansen (2000), which suggests that at least in the leading case of i.i.d. Gaussian errors, the confidence interval $\hat{\Gamma}$ is asymptotically conservative for fixed parameter values as n goes to infinity.

6.3. Slope Parameters

Theorem 3 shows that the GMM slope estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ are asymptotically normal and standard errors can be consistently computed from the covariance estimators $\hat{V}_1$ and $\hat{V}_2$. This implies that conventional asymptotic confidence intervals can be constructed based on the normal approximation. We denote this interval as $\hat{\Theta}_0$, for reasons given later. In the first row of Table 3 we present the finite sample coverage of nominal 95% confidence intervals for δ_2 constructed using this method. We see that if δ_2 is large, the interval $\hat{\Theta}_0$ has about the correct coverage, but coverage rates are quite poor for small values of δ_2. Coverage improves as the sample size n increases, but even for $n = 500$ coverage is quite poor for small values of δ_2.

TABLE 3. Nominal 95% confidence interval coverage for δ_2

	$n = 100$				$n = 250$				$n = 500$			
δ_2	0.25	0.5	1.0	2.0	0.25	0.5	1.0	2.0	0.25	0.5	1.0	2.0
$\hat{\Theta}_0$	54	69	87	93	59	82	93	94	70	87	92	96
$\hat{\Theta}_{0.5}$	86	89	95	95	84	94	96	95	90	96	97	95
$\hat{\Theta}_{0.8}$	94	94	96	97	94	98	98	96	96	98	97	97
$\hat{\Theta}_{0.95}$	98	99	99	98	99	99	98	98	99	99	98	98

To improve the coverage rates, we can use the Bonferroni-type approach advocated in Hansen (2000). The premise is that the poor coverage rates for $\hat{\Theta}_0$ are because it does not take into account uncertainty concerning γ. The solution is to incorporate the confidence interval $\hat{\Gamma}$ for γ developed in Section 4.2.

First, for any fixed value of γ we can calculate the GMM estimators of Section 3.3. Specifically, the sample is split by the indicators $1(q_i \leq \gamma)$ and $1(q_i > \gamma)$, the slope coefficients estimated by GMM on each subsample, and standard errors calculated using the conventional GMM formula. Given these estimates and standard errors, let $\hat{\Theta}(\gamma)$ denote the constructed C-level confidence region for θ (for this fixed value of γ).

Second, for any $0 \leq \kappa < 1$, let $\hat{\Gamma}(\kappa)$ denote the confidence interval of Section 4.2 for γ with asymptotic coverage κ.

Third, construct the union of the intervals $\hat{\Theta}(\gamma)$, where the union is taken over the values of γ in $\hat{\Gamma}(\kappa)$,

$$\hat{\Theta}_\kappa = \bigcup_{\gamma \in \hat{\Gamma}(\kappa)} \hat{\Theta}(\gamma).$$

Theorem 3 shows that $\hat{\Theta}_0 = \hat{\Theta}(\hat{\gamma})$ has asymptotic coverage C. Because $\hat{\Theta}_0 \subset \hat{\Theta}_\kappa$ it follows that

$$P(\theta \in \hat{\Theta}_\kappa) \geq P(\theta \in \hat{\Theta}_0) \to C$$

as $n \to \infty$. Thus the intervals $\hat{\Theta}_\kappa$ should be asymptotically conservative.

We report in the four rows of Table 3 the coverage rates of the interval $\hat{\Theta}_\kappa$ using $\kappa = 0$, 0.5, 0.8, and 0.95. As expected, the coverage rates increase substantially over the $\hat{\Theta}_0$ interval. We see that in this example $\kappa = 0.8$ provides good coverage in each case, which is the same recommendation as for the regression model investigated in Hansen (2000).

7. CONCLUSION

We have developed consistent estimators for the threshold in a model with endogenous variables and an exogenous threshold variable. The estimator for the threshold is a 2SLS estimator, and the estimator of the slope parameters is a

GMM estimator. It may be possible to construct alternative estimators of the parameters based on the GMM principle, and we make no claim to asymptotic efficiency. We specifically focus on the case of an exogenous threshold variable. The case of an endogenous threshold variable would require an alternative estimation approach, and this would be a worthwhile subject for future research.

NOTE

1. A formal argument for this claim follows the same reasoning as Theorem 2 in Hansen (1996). The key condition to verify is equation (13) of Hansen (1996) which is satisfied in our case by (A.65) in the Appendix.

REFERENCES

Andrews, D.W.K. & V. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.

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

Barnett, S.A. & P. Sakellaris (1998) Nonlinear response of firm investment to q: Testing a model of convex and non-convex adjustment costs. *Journal of Monetary Economics* 42, 261–288.

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

Caner, M. (2002) A note on LAD estimation of a threshold model. *Econometric Theory* 18, 800–814.

Chamberlain, G. (1987) Asymptotic efficiency in estimation with conditional Monet restrictions. *Journal of Econometrics* 34, 305–334.

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

Davies, R.B. (1977) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.

Erickson, T. & T.M. Whited (2000) Measurement error and the relationship between investment and q. *Journal of Political Economy* 108, 1027–1057.

Fazzari, S.M., R.G. Hubbard, & B.C. Petersen (1988) Financing constraints and corporate investment. *Brookings Papers on Economic Activity* 1, 141–195.

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

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

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

Hu, X. & F. Schiantarelli (1998) Investment and capital market imperfections: A switching regression approach using U.S. firm panel data. *Review of Economics and Statistics* 80, 466–479.

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

Picard, D. (1985) Testing and estimating change-points in time series. *Advances in Applied Probability* 17, 841–867.

APPENDIX

Proof of Lemma 1. We need to show that Assumptions 1 and 2 imply equations (15)–(17) for both the reduced form linear model (7) and the reduced form threshold model (8). First, we note the form of $\hat{r}_i$ and the rates of convergence of the parameter estimates in these two models. In the linear model (7),

$$\hat{r}_i = (\Pi - \hat{\Pi})x_i,$$

$$\sqrt{n}(\Pi - \hat{\Pi}) = O_p(1). \tag{A.1}$$

In the threshold model (8), we let ρ_0 denote the true value of ρ and write

$$\Delta_i(\gamma) = 1(q_i \leq \gamma) - 1(q_i \leq \gamma_0),$$

$$\Delta_{2i}(\rho) = 1(q_i \leq \rho) - 1(q_i \leq \rho_0).$$

Then

$$\hat{r}_i = (\Pi_1 - \hat{\Pi}_1)x_i 1(q_i \leq \rho_0) + (\Pi_2 - \hat{\Pi}_2)x_i 1(q_i > \rho_0) + (\hat{\Pi}_2 - \hat{\Pi}_1)x_i \Delta_{2i}(\hat{\rho}), \tag{A.2}$$

$$\sqrt{n}(\Pi_1 - \hat{\Pi}_1) = O_p(1), \tag{A.3}$$

$$\sqrt{n}(\Pi_2 - \hat{\Pi}_2) = O_p(1), \tag{A.4}$$

$$(\hat{\Pi}_2 - \hat{\Pi}_1) = O_p(1), \tag{A.5}$$

$$n(\hat{\rho} - \rho_0) = O_p(1). \tag{A.6}$$

Equation (A.2) is simple algebra, and equations (A.3)–(A.6) are shown by Chan (1993).
We now sequentially establish (15)–(17).

Proof of (15). First, take the linear model (7).

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} H_i \hat{r}_i' 1(q_i \leq \gamma) \right| \leq \frac{1}{n} \sum_{i=1}^{n} |H_i x_i'| |\Pi - \hat{\Pi}| \sqrt{n} = O_p(1),$$

establishing (15).

Second, take the threshold model (8). Chan (1993) establishes that

$$\left| \sum_{i=1}^{n} H_i x_i' 1(q_i \leq \gamma) \Delta_{2i}(\hat{\rho}) \right| \leq \sum_{i=1}^{n} |H_i x_i'| |\Delta_{2i}(\hat{\rho})| = O_p(1). \tag{A.7}$$

Combining this with (A.5), we conclude that

$$\begin{aligned} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} H_i \hat{r}_i' 1(q_i \leq \gamma) \right| &\leq \frac{1}{n} \sum_{i=1}^{n} |H_i x_i'| \sqrt{n}(|\Pi_1 - \hat{\Pi}_1| + |\Pi_2 - \hat{\Pi}_2|) \\ &\quad + \frac{1}{n^{1/2}} \sum_{i=1}^{n} |H_i x_i'| |\Delta_{2i}(\hat{\rho})| |\hat{\Pi}_2 - \hat{\Pi}_1| \\ &= O_p(1), \end{aligned}$$

which is (15). ∎

Proof of (16). By Lemma A.7 of Hansen (2000), there exist constants B and k such that for all $\eta > 0$ and $\varepsilon > 0$ there exists a $\bar{v} < \infty$ such that for all n

$$P\left(\sup_{\frac{\bar{v}}{a_n} \le |\gamma - \gamma_0| \le B} \frac{\sum_{i=1}^{n} |H_i x_i'| |\Delta_i(\gamma)|}{n|\gamma - \gamma_0|} > (1+\eta)k \right) \le \frac{\varepsilon}{2}. \tag{A.8}$$

First, take the linear model (7). Using (A.8),

$$\sup_{\frac{\bar{v}}{a_n} \le |\gamma - \gamma_0| \le B} \left| \frac{\sum_{i=1}^{n} H_i \hat{r}_i' \Delta_i(\gamma)}{n^{1-\alpha}|\gamma - \gamma_0|} \right| \le \sup_{\frac{\bar{v}}{a_n} \le |\gamma - \gamma_0| \le B} \frac{\sum_{i=1}^{n} |H_i x_i'| |\Delta_i(\gamma)|}{n|\gamma - \gamma_0|} n^{\alpha} |\Pi - \hat{\Pi}| \to_p 0$$

because $n^{1/2}|\Pi - \hat{\Pi}| = O_p(1)$. Similarly

$$\sup_{\frac{\bar{v}}{a_n} \le |\gamma - \gamma_0| \le B} \left| \frac{\sum_{i=1}^{n} \hat{r}_i \hat{r}_i' \Delta_i(\gamma)}{n^{1-\alpha}|\gamma - \gamma_0|} \right| \le \sup_{\frac{\bar{v}}{a_n} \le |\gamma - \gamma_0| \le B} \frac{\sum_{i=1}^{n} |x_i x_i'| |\Delta_i(\gamma)|}{n|\gamma - \gamma_0|} n^{\alpha} |\Pi - \hat{\Pi}|^2 \to_p 0,$$

which is (16).

Second, take the threshold model (8). By (A.8), (A.3)–(A.6), and (A.7), we can pick $\bar{v}$, $\bar{v}_2$, K_1, K_2, and $\bar{n}$ so that for all $n \ge \bar{n}$, with probability exceeding $1 - \varepsilon$,

$$\sup_{\frac{\bar{v}}{a_n} \le |\gamma - \gamma_0| \le B} \frac{\sum_{i=1}^{n} |H_i x_i'| |\Delta_i(\gamma)|}{n|\gamma - \gamma_0|} \le (1+\eta)k,$$

$$n^{\alpha}(|\Pi_1 - \hat{\Pi}_1| + |\Pi_2 - \hat{\Pi}_2|) \le \frac{\delta}{2(1+\eta)k},$$

$$|\hat{\Pi}_2 - \hat{\Pi}_1| \le K_1,$$

$$n|\hat{\rho} - \rho_0| \le \bar{v}_2, \tag{A.9}$$

$$n^{-\alpha} \sum_{i=1}^{n} |H_i x_i'| |\Delta_{2i}(\hat{\rho})| \le \frac{\bar{v}\delta}{2K_1}. \tag{A.10}$$

Using (A.2)

$$\begin{aligned}\sum_{i=1}^{n} H_i \hat{r}_i' \Delta_i(\gamma) &= \sum_{i=1}^{n} H_i x_i' \Delta_i(\gamma) 1(q_i \le \rho_0)(\Pi_1 - \hat{\Pi}_1)' \\ &\quad + \sum_{i=1}^{n} H_i x_i' \Delta_i(\gamma) 1(q_i > \rho_0)(\Pi_2 - \hat{\Pi}_2)' \\ &\quad + \sum_{i=1}^{n} H_i x_i' \Delta_i(\gamma) \Delta_{2i}(\hat{\rho})(\hat{\Pi}_2 - \hat{\Pi}_1)'.\end{aligned}$$

Observe that if $\bar{v}/a_n \leq |\gamma - \gamma_0| \leq B$, then

$$|\Delta_i(\gamma)| \leq |\Delta_i(\gamma_0 + B)| + |\Delta_i(\gamma_0 - B)|$$

and

$$\frac{1}{n^{1-\alpha}|\gamma - \gamma_0|} \leq \frac{1}{\bar{v}n^{\alpha}}.$$

Hence

$$\sup_{\frac{\bar{v}}{a_n} \leq |\gamma-\gamma_0| \leq B} \left| \frac{\sum_{i=1}^{n} H_i x_i' \Delta_i(\gamma)\Delta_{2i}(\hat{\rho})}{n^{1-\alpha}|\gamma - \gamma_0|} \right| \leq \frac{1}{\bar{v}n^{\alpha}} \sum_{i=1}^{n} |H_i x_i'|(|\Delta_i(\gamma_0 + B)| + |\Delta_i(\gamma_0 - B)|)|\Delta_{2i}(\hat{\rho})| \equiv A_{1n}. \quad \textbf{(A.11)}$$

Thus with probability exceeding $1 - \varepsilon$,

$$\begin{aligned}
&\sup_{\frac{\bar{v}}{a_n} \leq |\gamma-\gamma_0| \leq B} \left| \frac{\sum_{i=1}^{n} H_i \hat{r}_i' \Delta_i(\gamma)}{n^{1-\alpha}|\gamma - \gamma_0|} \right| \\
&\leq \sup_{\frac{\bar{v}}{a_n} \leq |\gamma-\gamma_0| \leq B} \left| \frac{\sum_{i=1}^{n} H_i x_i' \Delta_i(\gamma)}{n|\gamma - \gamma_0|} \right| n^{\alpha}(|\Pi_1 - \hat{\Pi}_1| + |\Pi_2 - \hat{\Pi}_2|) \\
&\quad + \sup_{\frac{\bar{v}}{a_n} \leq |\gamma-\gamma_0| \leq B} \left| \frac{\sum_{i=1}^{n} H_i x_i' \Delta_i(\gamma)\Delta_{2i}(\hat{\rho})}{n^{1-\alpha}|\gamma - \gamma_0|} \right| |\hat{\Pi}_2 - \hat{\Pi}_1| \\
&\leq \frac{\delta}{2} + A_{1n}K_1.
\end{aligned} \quad \textbf{(A.12)}$$

We conclude by showing $|A_{1n}|K_1 \leq \delta/2$, which completes the proof of (16). This is simplest when $\rho_0 \neq \gamma_0$, for then we pick B and $\bar{n}$ so that

$$B + \frac{\bar{v}_2}{\bar{n}} \leq |\gamma_0 - \rho_0|. \quad \textbf{(A.13)}$$

By (A.9) and the triangle inequality,

$$|\gamma_0 - \rho_0| \leq |\gamma_0 - \hat{\rho}| + |\hat{\rho} - \rho_0| \leq |\gamma_0 - \hat{\rho}| + \frac{\bar{v}_2}{\bar{n}}. \quad \textbf{(A.14)}$$

Expressions (A.13) and (A.14) imply $|\hat{\rho} - \gamma_0| > B$. Thus for all i

$$(|\Delta_i(\gamma_0 + B)| + |\Delta_i(\gamma_0 - B)|)|\Delta_{2i}(\hat{\rho})| = 0,$$

and thus $A_{1n} = 0$.

For the case $\rho_0 = \gamma_0$, we pick $\bar{n}$ so that $\bar{n}^{2\alpha} \geq \bar{v}/\bar{v}_2$. (A.9) implies

$$|\hat{\rho} - \rho_0| \leq \bar{v}/a_n \leq \bar{v}_2/n,$$

and thus for all i

$$(|\Delta_i(\gamma_0 + B)| + |\Delta_i(\gamma_0 - B)|)|\Delta_{2i}(\hat{\rho})| \leq |\Delta_{2i}(\hat{\rho})|.$$

Hence using (A.10), $A_{1n}K_1 \leq \delta/2$ for sufficiently large n. ■

Proof of (17). Define

$$\Delta_i^*(v) = 1(q_i \leq \gamma_0 + v/a_n) - 1(q_i \leq \gamma_0).$$

By Lemma A.10 of Hansen (2000),

$$\sup_{|v| \leq \bar{v}} n^{-2\alpha} \sum_{i=1}^{n} |H_i x_i'| \Delta_i^*(v) = O_p(1). \tag{A.15}$$

First, take the linear model (7). Using (A.15) and (A.1),

$$\sup_{|v| \leq \bar{v}} n^{-\alpha} \left| \sum_{i=1}^{n} H_i \hat{r}_i' \Delta_i^*(v) \right| \leq \sup_{|v| \leq \bar{v}} n^{-2\alpha} \left| \sum_{i=1}^{n} H_i x_i' \Delta_i^*(v) \right| n^{\alpha} |\Pi - \hat{\Pi}| = o_p(1)$$

and

$$\sup_{|v| \leq \bar{v}} n^{-\alpha} \left| \sum_{i=1}^{n} \hat{r}_i \hat{r}_i' \Delta_i^*(v) \right| \leq \sup_{|v| \leq \bar{v}} n^{-2\alpha} \left| \sum_{i=1}^{n} x_i x_i' \Delta_i^*(v) \right| n^{\alpha} |\Pi - \hat{\Pi}|^2 = o_p(1)$$

as desired.

Second, take the threshold model (8). Using (A.2) and (A.15),

$$\begin{aligned}
\sup_{|v| \leq \bar{v}} n^{-\alpha} \left| \sum_{i=1}^{n} H_i \hat{r}_i' \Delta_i^*(v) \right| &\leq \sup_{|v| \leq \bar{v}} n^{-2\alpha} \sum_{i=1}^{n} |H_i x_i'| \Delta_i^*(v) n^{\alpha} (|\Pi_1 - \hat{\Pi}_1| + |\Pi_2 - \hat{\Pi}_2|) \\
&\quad + \sup_{|v| \leq \bar{v}} n^{-\alpha} \sum_{i=1}^{n} |H_i x_i'| |\Delta_i^*(v)| |\Delta_{2i}(\hat{\rho})| |\hat{\Pi}_2 - \hat{\Pi}_1| \\
&\leq o_p(1) + O_p(1) \cdot A_{2n},
\end{aligned}$$

where

$$A_{2n} = n^{-\alpha} \sum_{i=1}^{n} |H_i x_i'| (|\Delta_i^*(\bar{v})| + |\Delta_i^*(-\bar{v})|) |\Delta_{2i}(\hat{\rho})|.$$

The proof is completed by showing that $A_{2n} = o_p(1)$. If $\gamma_0 \neq \rho_0$, then for large enough n, $(|\Delta_i^*(\bar{v})| + |\Delta_i^*(-\bar{v})|)|\Delta_{2i}(\hat{\rho})| = 0$ and $A_{2n} = 0$ with probability arbitrarily close to 1. If $\gamma_0 = \rho_0$, then for large enough n, $(|\Delta_i^*(\bar{v})| + |\Delta_i^*(-\bar{v})|)|\Delta_{2i}(\hat{\rho})| = |\Delta_{2i}(\hat{\rho})|$ and

$$A_{2n} = n^{-\alpha} \sum_{i=1}^{n} |H_i x_i'| |\Delta_{2i}(\hat{\rho})| = o_p(1)$$

by (A.7). ■

Let $\hat{v}_i = \hat{r}_i'\theta_2 + v_i$ and $\hat{v} = \hat{r}\theta_2 + v$. Let G_0 be the matrix obtained by stacking the vectors $g_i'1(q_i \le \gamma_0)$.

LEMMA 2. *Uniformly in* $\gamma \in \Gamma$,

$$\frac{1}{n}\hat{Z}_\gamma'\hat{Z}_\gamma = \frac{1}{n}\sum_{i=1}^{n}\hat{z}_i\hat{z}_i'1(q_i \le \gamma) \to_p M(\gamma), \tag{A.16}$$

$$\frac{1}{n}\hat{Z}_0'G_0 = \frac{1}{n}\sum_{i=1}^{n}\hat{z}_i g_i'1(q_i \le \gamma_0) \to_p M_0 = M(\gamma_0), \tag{A.17}$$

$$\frac{1}{\sqrt{n}}\hat{Z}_\gamma'\hat{v} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{z}_i\hat{v}_i'1(q_i \le \gamma) = O_p(1). \tag{A.18}$$

Proof of Lemma 2. To show (A.16), because $\hat{z}_i = \hat{g}_i = g_i - \hat{r}_i$, using (15) and Lemma 1 of Hansen (1996),

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\hat{z}_i\hat{z}_i'1(q_i \le \gamma) &= \frac{1}{n}\sum_{i=1}^{n}g_i g_i'1(q_i \le \gamma) - \frac{1}{n}\sum_{i=1}^{n}g_i\hat{r}_i'1(q_i \le \gamma)\\ &\quad - \frac{1}{n}\sum_{i=1}^{n}\hat{r}_i g_i'1(q_i \le \gamma) + \frac{1}{n}\sum_{i=1}^{n}\hat{r}_i\hat{r}_i'1(q_i \le \gamma)\\ &\to_p M(\gamma)\end{aligned}$$

uniformly in $\gamma \in \Gamma$. Equation (A.17) follows similarly. To show (A.18), using Lemma A.4 of Hansen (2000) and (15),

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{z}_i\hat{v}_i'1(q_i \le \gamma) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_i v_i'1(q_i \le \gamma) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{r}_i v_i'1(q_i \le \gamma)\\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_i\hat{r}_i'\theta_2 1(q_i \le \gamma) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{r}_i\hat{r}_i'\theta_2 1(q_i \le \gamma)'\\ &= O_p(1).\end{aligned}$$

■

LEMMA 3. $\hat{\gamma} \to_p \gamma_0$.

Proof of Lemma 3. It will be convenient to write (5) in the format

$$\begin{aligned}y_i &= g_i'(\delta_n + \theta_2)1(q_i \le \gamma_0) + g_i'\theta_2 1(q_i > \gamma_0) + v_i\\ &= g_i'\theta_2 + g_i'1(q_i \le \gamma_0)\delta_n + v_i.\end{aligned}$$

Define the matrices $\hat{G}$, G, $\hat{r}$, and v by stacking $\hat{g}_i'$, g_i', $\hat{r}_i$, and v_i. This equation can be written in matrix format as

$$Y = G\theta + G_0 cn^{-\alpha} + v, \tag{A.19}$$

where we use $\delta_n = cn^{-\alpha}$.

Let $P_\gamma = \hat{Z}_\gamma(\hat{Z}_\gamma'\hat{Z}_\gamma)^{-1}\hat{Z}_\gamma'$, $P_\perp = \hat{Z}_\perp(\hat{Z}_\perp'\hat{Z}_\perp)^{-1}\hat{Z}_\perp'$, and $P_\gamma^* = \hat{Z}_\gamma^*(\hat{Z}_\gamma^{*\prime}\hat{Z}_\gamma^*)^{-1}\hat{Z}_\gamma^{*\prime}$ where $\hat{Z}_\gamma^* = (\hat{Z}_\gamma, \hat{Z}_\perp)$. Because $\hat{Z}_\gamma'\hat{Z}_\perp = 0$, it can be shown that $P_\gamma^* = P_\gamma + P_\perp$.

Because $G = \hat{G} + \hat{r}$ and $\hat{Z} = \hat{G}$ is in the span of $\hat{Z}_\gamma^*$, then

$(I - P_\gamma^*)G = (I - P_\gamma^*)\hat{r}.$

Thus using (A.19),

$(I - P_\gamma^*)Y = (I - P_\gamma^*)(G_0 c n^{-\alpha} + \hat{v}).$

Hence

$$\begin{aligned} S_n(\gamma) &= Y'(I - P_\gamma^*)Y \\ &= (n^{-\alpha}c'G_0' + \hat{v}')(I - P_\gamma^*)(G_0 c n^{-\alpha} + \hat{v}) \\ &= (n^{-\alpha}c'G_0' + \hat{v}')(G_0 c n^{-\alpha} + \hat{v}) - (n^{-\alpha}c'G_0' + \hat{v}')P_\gamma^*(G_0 c n^{-\alpha} + \hat{v}). \qquad \textbf{(A.20)} \end{aligned}$$

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

$$\begin{aligned} S_n^*(\gamma) &= n^{2\alpha-1}(n^{-\alpha}c'G_0' + \hat{v}')P_\gamma^*(G_0 c n^{-\alpha} + \hat{v}) \\ &= n^{-1}c'G_0'P_\gamma^* G_0 c + n^{\alpha-1}2c'G_0'P_\gamma^*\hat{v} + n^{2\alpha-1}\hat{v}'P_\gamma^*\hat{v}. \end{aligned}$$

From Lemma 2, we see that uniformly in $\gamma \in [\gamma_0, \bar{\gamma}]$,

$$G_0'P_\gamma G_0 = \frac{1}{n} G_0'\hat{Z}_0\left(\frac{1}{n}\hat{Z}_\gamma'\hat{Z}_\gamma\right)^{-1}\frac{1}{n}\hat{Z}_0'G_0 \to_p M_0 M(\gamma)^{-1}M_0,$$

$$n^{\alpha-1}G_0'P_\gamma\hat{v} = n^{\alpha-1/2}\frac{1}{n}G_0'\hat{Z}_0\left(\frac{1}{n}\hat{Z}_\gamma'\hat{Z}_\gamma\right)^{-1}\frac{1}{\sqrt{n}}\hat{Z}_\gamma'\hat{v} \to_p 0,$$

$$n^{2\alpha-1}\hat{v}'P_\gamma\hat{v} = n^{2\alpha-1}\frac{1}{\sqrt{n}}\hat{v}'\hat{Z}_\perp\left(\frac{1}{n}\hat{Z}_\perp'\hat{Z}_\perp\right)^{-1}\frac{1}{\sqrt{n}}\hat{Z}_\perp'\hat{v} \to_p 0,$$

and

$$n^{2\alpha-1}\hat{v}'P_\perp\hat{v} = n^{2\alpha-1}\frac{1}{\sqrt{n}}\hat{v}'\hat{Z}_\gamma\left(\frac{1}{n}\hat{Z}_\gamma'\hat{Z}_\gamma\right)^{-1}\frac{1}{\sqrt{n}}\hat{Z}_\gamma'\hat{v} \to_p 0.$$

When $\gamma > \gamma_0$ then $P_\perp G_0 = 0$, so $P_\gamma^* G_0 = P_\gamma G_0$ and

$$\begin{aligned} S_n^*(\gamma) &= n^{-1}c'G_0'P_\gamma G_0 c + n^{\alpha-1}2c'G_0'P_\gamma\hat{v} + n^{2\alpha-1}\hat{v}'P_\gamma\hat{v} + n^{2\alpha-1}\hat{v}'P_\perp\hat{v} \\ &\to_p c'M_0 M(\gamma)^{-1}M_0 c \end{aligned}$$

uniformly on $\gamma \in [\gamma_0, \bar{\gamma}]$, which is uniquely maximized at γ_0 as shown in the proof of Lemma A.5 of Hansen (2000). Symmetrically, on $\gamma \in [\underline{\gamma}, \gamma_0]$, $S_n^*(\gamma)$ converges uniformly to a limit function uniquely maximized at γ_0. Because $\hat{\gamma}$ maximizes $S_n^*(\gamma)$, it follows that $\hat{\gamma} \to_p \gamma_0$. ■

LEMMA 4. $a_n(\hat{\gamma} - \gamma_0) = O_p(1)$.

Proof of Lemma 4. Let the constants B, d, k be defined as $B > 0$, $0 < d < \infty$, $0 < k < \infty$. Let $\bar{M} = \sup_{|\gamma-\gamma_0|\le B}|M(\gamma)^{-1}|$ and $\bar{D} = \sup_{|\gamma-\gamma_0|\le B}|D_1(\gamma)f(\gamma)|$. Fix $\varepsilon > 0$. Pick κ and reduce B if necessary so that

$$\kappa + 3kM^*(\bar{D}B + 2\kappa)(1 + M^*(M_0 + \kappa)) < d/12, \tag{A.21}$$

$$\kappa(M_0 + \kappa)M^*[1 + 3kM^*] \le d/12, \tag{A.22}$$

$$\kappa^2 M^*[2 + 3kM^*] \le d/12, \tag{A.23}$$

where $M^* = \overline{M} + \overline{M}^2\kappa$ and $M_0 = M(\gamma_0)$. To simplify some inequalities, assume without loss of generality that $\kappa \le k$.

Define

$$\Delta_i(\gamma) = 1(q_i \le \gamma) - 1(q_i \le \gamma_0).$$

Define the joint event

$$\sup_{\bar{v}/a_n \le |\gamma - \gamma_0| \le B} \frac{\sum_{i=1}^n |g_i|^2 \Delta_i(\gamma)}{n(\gamma - \gamma_0)} \le 13k/12, \tag{A.24}$$

$$\inf_{\bar{v}/a_n \le |\gamma - \gamma_0| \le B} \frac{\sum_{i=1}^n (c'g_i)^2 \Delta_i(\gamma)}{n(\gamma - \gamma_0)} \ge 11d/12, \tag{A.25}$$

$$\sup_{\bar{v}/a_n \le |\gamma - \gamma_0| \le B} \left| -\frac{\sum_{i=1}^n g_i \hat{r}_i' \Delta_i(\gamma)}{n(\gamma - \gamma_0)} - \frac{\sum_{i=1}^n \hat{r}_i g_i' \Delta_i(\gamma)}{n(\gamma - \gamma_0)} + \frac{\sum_{i=1}^n \hat{r}_i \hat{r}_i' \Delta_i(\gamma)}{n(\gamma - \gamma_0)} \right| \le \kappa, \tag{A.26}$$

$$\sup_{\bar{v}/a_n \le |\gamma - \gamma_0| \le B} \left| \frac{\sum_{i=1}^n (g_i - \hat{r}_i)\hat{v}_i' \Delta_i(\gamma)}{n^{1-\alpha}(\gamma - \gamma_0)} \right| \le \kappa, \tag{A.27}$$

$$|\hat{\gamma} - \gamma_0| \le B, \tag{A.28}$$

$$\sup_{\gamma \in \Gamma} \left| \frac{1}{n} \hat{Z}_\gamma' \hat{Z}_\gamma - M(\gamma) \right| \le \kappa, \tag{A.29}$$

$$\sup_{\gamma \in \Gamma} \left| \left| \frac{1}{n} \hat{Z}_\gamma' \hat{Z}_\gamma \right| - |M(\gamma)| \right| \le \kappa, \tag{A.30}$$

$$\sup_{\gamma \in \Gamma} \left| \frac{1}{n} \hat{Z}_0' G_0 - M_0 \right| \le \kappa, \tag{A.31}$$

$$\sup_{\gamma \in \Gamma} \left| \frac{1}{n^{1-\alpha}} \hat{Z}_\gamma' \hat{v} \right| \le \kappa. \tag{A.32}$$

By Lemma A.7 of Hansen (2000) and (16), there exist sufficiently large $\bar{v} = \bar{v}(\varepsilon) < \infty$ and $\bar{n} = \bar{n}(\varepsilon) < \infty$ so that for all $n \ge \bar{n}$, equations (A.24)–(A.27) hold jointly with probability exceeding $1 - \varepsilon/2$. By Lemmas 2 and 3, equations (A.28)–(A.32) hold jointly with probability exceeding $1 - \varepsilon/2$ (increasing $\bar{n}$ if necessary). Thus (A.24)–(A.32) hold jointly with probability exceeding $1 - \varepsilon$.

We show subsequently that (A.24)–(A.32) imply

$$\inf_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} c' \left(\frac{G_0'(P_0^* - P_\gamma^*)G_0}{n(\gamma-\gamma_0)} \right) c \ge 5d/6, \tag{A.33}$$

$$\sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} \left| \frac{c' G_0'(P_0^* - P_\gamma^*)\hat{v}}{n^{1-\alpha}(\gamma-\gamma_0)} \right| \le d/12, \tag{A.34}$$

$$\sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} \left| \frac{\hat{v}'(P_0^* - P_\gamma^*)\hat{v}}{n^{1-2\alpha}(\gamma-\gamma_0)} \right| \le d/6. \tag{A.35}$$

Using (A.20) and applying (A.33)–(A.35) we can calculate that for $\bar{v}/a_n \le |\gamma - \gamma_0| \le B$, (A.24)–(A.32) imply

$$\frac{S_n(\gamma) - S_n(\gamma_0)}{n^{1-2\alpha}(\gamma-\gamma_0)} = \frac{\hat{v}'(P_0^* - P_\gamma^*)\hat{v}}{n^{1-2\alpha}(\gamma-\gamma_0)} + 2\,\frac{\hat{v}'(P_0^* - P_\gamma^*)G_0 c}{n^{1-\alpha}(\gamma-\gamma_0)} + \frac{c' G_0'(P_0^* - P_\gamma^*)G_0 c}{n(\gamma-\gamma_0)}$$
$$\ge d/2.$$

Because $S_n(\hat{\gamma}) - S_n(\gamma_0) \le 0$, this establishes that (A.24)–(A.32) imply $|\hat{\gamma} - \gamma_0| \le \bar{v}/a_n$. As discussed previously, (A.24)–(A.32) hold jointly with probability exceeding $1 - \varepsilon$ for all $n \ge \bar{n}$. Thus $P(a_n|\hat{\gamma} - \gamma_0| > \bar{v}) \le \varepsilon$ for $n \ge \bar{n}$ as required.

The proof is completed by showing that (A.24)–(A.32) imply (A.33)–(A.35).

For simplicity, we restrict attention to the region $[\gamma_0 + \bar{v}/a_n \le \gamma \le \gamma_0 + B]$, as the analysis for the case $[\gamma_0 - \bar{v}/a_n \ge \gamma \ge \gamma_0 - B]$ is similar. This restriction implies $G_0'\hat{Z}_\gamma = G_0'\hat{Z}_0$, $\hat{Z}_0'\hat{Z}_\gamma = \hat{Z}_0'\hat{Z}_0$, and $P_\perp G_0 = 0$ so $(P_0^* - P_\gamma^*)G_0 = (P_0 - P_\gamma)G_0$.

We start with some useful bounds. Equation (A.30) implies

$$B_{1n} = \sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} |(n^{-1}\hat{Z}_\gamma'\hat{Z}_\gamma)^{-1}|$$
$$\le \sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} \left| |M(\gamma)|^{-1} - |M(\gamma)|^{-2} \right| \left\| \frac{1}{n}\hat{Z}_\gamma'\hat{Z}_\gamma \right| - |M(\gamma)| \right\|$$
$$\le \bar{M} + \bar{M}^2\kappa \equiv M^*. \tag{A.36}$$

Equations (A.29), (A.31), (A.36), and a Taylor's expansion imply

$$B_{2n} = \sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} |I_m - (\hat{Z}_\gamma'\hat{Z}_\gamma)^{-1}(\hat{Z}_0' G_0)|$$
$$\le B_{1n} \sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} |\hat{Z}_\gamma'\hat{Z}_\gamma - \hat{Z}_0' G_0|$$
$$\le B_{1n}\left(\sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} |M(\gamma) - M(\gamma_0)| + 2\kappa \right)$$
$$\le M^*(\bar{D}B + 2\kappa). \tag{A.37}$$

Expressions (A.31), (A.32), and (A.36) imply

$$B_{3n} = \sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} |(\hat{Z}_\gamma'\hat{Z}_\gamma)^{-1}(\hat{Z}_0' G_0)| \le M^*(M_0 + \kappa) \tag{A.38}$$

and

$$B_{4n} = \sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} |n^{\alpha}(\hat{Z}_\gamma'\hat{Z}_\gamma)^{-1}(\hat{Z}_0'\hat{v})| \le M^*\kappa. \tag{A.39}$$

Next, observe that

$$\hat{Z}_\gamma'\hat{Z}_\gamma - \hat{Z}_0'\hat{Z}_0 = \sum_{i=1}^n g_i g_i' \Delta_i(\gamma) - \sum_{i=1}^n g_i \hat{r}_i' \Delta_i(\gamma) - \sum_{i=1}^n \hat{r}_i g_i' \Delta_i(\gamma) + \sum_{i=1}^n \hat{r}_i \hat{r}_i' \Delta_i(\gamma). \tag{A.40}$$

Equations (A.24)–(A.27) imply

$$C_{1n} = \sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} \left| \frac{\hat{Z}_\gamma'\hat{Z}_\gamma - \hat{Z}_0'\hat{Z}_0}{n(\gamma-\gamma_0)} \right| \le (1+\eta)k + \kappa \le 3k, \tag{A.41}$$

$$C_{2n} = \inf_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} c' \left(\frac{\hat{Z}_\gamma'\hat{Z}_\gamma - \hat{Z}_0'\hat{Z}_0}{n(\gamma-\gamma_0)} \right) c \ge (1-\eta)\, d - \kappa, \tag{A.42}$$

and

$$C_{3n} = \sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} \left| \frac{\hat{Z}_\gamma'\hat{v} - \hat{Z}_0'\hat{v}}{n^{1-\alpha}(\gamma-\gamma_0)} \right| = \sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} \left| \frac{\sum_{i=1}^n (g_i - \hat{r}_i)\hat{v}_i'\Delta_i(\gamma)}{n^{1-\alpha}(\gamma-\gamma_0)} \right| \le \kappa. \tag{A.43}$$

We can now show (A.33). We calculate that

$$\begin{aligned} G_0'(P_0^* - P_\gamma^*)G_0 &= (\hat{Z}_\gamma'\hat{Z}_\gamma - \hat{Z}_0'\hat{Z}_0) - (\hat{Z}_\gamma'\hat{Z}_\gamma - \hat{Z}_0'\hat{Z}_0)(I_m - (\hat{Z}_\gamma'\hat{Z}_\gamma)^{-1}\hat{Z}_0'G_0) \\ &\quad - (I_m - G_0'\hat{Z}_0(\hat{Z}_0'\hat{Z}_0)^{-1})(\hat{Z}_\gamma'\hat{Z}_\gamma - \hat{Z}_0'\hat{Z}_0)(\hat{Z}_\gamma'\hat{Z}_\gamma)^{-1}\hat{Z}_0'G_0. \end{aligned} \tag{A.44}$$

Hence, using (A.37), (A.38), (A.41), and (A.42),

$$\begin{aligned} &\inf_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} c' \left(\frac{G_0'(P_0^* - P_\gamma^*)G_0}{n(\gamma-\gamma_0)} \right) c \\ &\quad \ge C_{2n} - C_{1n}B_{2n} - C_{1n}B_{2n}B_{3n} \\ &\quad \ge (1-\eta)\, d - \kappa - 3kM^*(\bar{D}B + 2\kappa)(1 + M^*(M_0 + \kappa)) \\ &\quad \ge (1-2\eta)\, d, \end{aligned}$$

the last inequality being (A.21). This is (A.33).

Next, we show (A.34). We calculate that

$$\begin{aligned} G_0'(P_0^* - P_\gamma^*)\hat{v} &= G_0'\hat{Z}_0(\hat{Z}_0'\hat{Z}_0)^{-1}(\hat{Z}_\gamma'\hat{Z}_\gamma - \hat{Z}_0'\hat{Z}_0)(\hat{Z}_\gamma'\hat{Z}_\gamma)^{-1}\hat{Z}_0'\hat{v} \\ &\quad - G_0'\hat{Z}_0(\hat{Z}_\gamma'\hat{Z}_\gamma)^{-1}(\hat{Z}_\gamma'\hat{v} - \hat{Z}_0'\hat{v}). \end{aligned} \tag{A.45}$$

Hence, using (A.38), (A.39), (A.41) and (A.43),

$$\sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} \left| \frac{c'G_0'(P_0^* - P_\gamma^*)\hat{v}}{n^{1-\alpha}(\gamma-\gamma_0)} \right| \le C_{1n}B_{3n}B_{4n} + B_{3n}C_{3n}$$

$$\le 3kM^*(M_0+\kappa)M^*\kappa + M^*(M_0+\kappa)\kappa \le \eta d,$$

the last inequality being (A.22). This is (A.34).

Finally, we show (A.35). Observe that

$$\hat{v}'(P_0^* - P_\gamma^*)\hat{v} = \hat{v}'(P_0 - P_\gamma)\hat{v} + \hat{v}'(P_{\perp 0} - P_\perp)\hat{v}. \tag{A.46}$$

We examine the first term of (A.46). We calculate that

$$\hat{v}'(P_0 - P_\gamma)\hat{v} = \hat{v}'\hat{Z}_0(\hat{Z}_0'\hat{Z}_0)^{-1}(\hat{Z}_\gamma'\hat{Z}_\gamma - \hat{Z}_0'\hat{Z}_0)(\hat{Z}_\gamma'\hat{Z}_\gamma)^{-1}\hat{Z}_0'\hat{v}$$

$$- 2\hat{v}'\hat{Z}_0(\hat{Z}_\gamma'\hat{Z}_\gamma)^{-1}(\hat{Z}_\gamma'\hat{v} - \hat{Z}_0'\hat{v}). \tag{A.47}$$

Hence using (A.39), (A.41), and (A.43),

$$\sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} \left| \frac{\hat{v}'(P_0 - P_\gamma)\hat{v}}{n^{1-2\alpha}(\gamma-\gamma_0)} \right| \le B_{4n}^2 C_{1n} + 2B_{4n}C_{3n}$$

$$\le M^{*2}\kappa^2 3k + 2M^*\kappa^2 \le d\eta,$$

the last inequality being (A.23). A similar argument applies to the second term of (A.46). Together, we find

$$\sup_{\bar{v}/a_n \le |\gamma-\gamma_0| \le B} \left| \frac{\hat{v}'(P_0^* - P_\gamma^*)\hat{v}}{n^{1-2\alpha}(\gamma-\gamma_0)} \right| \le 2d\eta.$$

This is (A.35) and completes the proof. ■

LEMMA 5. *On* $[-\bar{v}, \bar{v}]$,

$$Q_n(\nu) = S_n(\gamma_0) - S_n(\gamma_0 + \nu/a_n) \Rightarrow -\mu|\nu| + 2\lambda^{1/2}W(\nu),$$

where $\mu = c'D_1 cf$ *and* $\lambda = c'D_2 cf$.

Proof of Lemma 5. Reparameterize all functions of γ instead as functions of ν. For example, $Z_\nu = Z_{\gamma_0+\nu/a_n}$, $P_\nu = P_{\gamma_0+\nu/a_n}$, $\Delta_i(\nu) = \Delta_i(\gamma_0 + \nu/a_n)$. We show subsequently that uniformly on $\nu \in [-\bar{v}, \bar{v}]$,

$$n^{-2\alpha}c'G_0'(P_0^* - P_\nu^*)G_0c \Rightarrow |\nu|\mu, \tag{A.48}$$

$$n^{-\alpha}c'G_0'(P_0^* - P_\nu^*)\hat{v} \Rightarrow \lambda^{1/2}W(\nu), \tag{A.49}$$

$$\hat{v}'(P_0^* - P_\nu^*)\hat{v} \Rightarrow 0. \tag{A.50}$$

Using (A.20) and (A.48)–(A.50), we find

$$Q_n(\nu) = (n^{-\alpha}c'G_0' + \hat{v}')P_\nu^*(G_0 cn^{-\alpha} + \hat{v}) - (n^{-\alpha}c'G_0' + \hat{v}')P_0^*(G_0 cn^{-\alpha} + \hat{v})$$

$$\Rightarrow 2\lambda^{1/2}W(\nu) - |\nu|\mu$$

as stated. The proof is completed by demonstrating (A.48)–(A.50).

First, observe that using the decomposition (A.40), Lemma A.10 of Hansen (2000), and (17), we have

$$n^{-2\alpha}|\hat{Z}_\nu'\hat{Z}_\nu - \hat{Z}_0'\hat{Z}_0|$$

$$\leq n^{-2\alpha}\sum_{i=1}^{n}|g_i|^2|\Delta_i(\nu)| + 2n^{-2\alpha}\left|\sum_{i=1}^{n} g_i\hat{r}_i'\Delta_i(\nu)\right| + n^{-2\alpha}\left|\sum_{i=1}^{n}\hat{r}_i\hat{r}_i'\Delta_i(\nu)\right|$$

$$\Rightarrow |D_1 f||\nu|,$$

so

$$n^{-2\alpha}\sup_{|\nu|\leq\bar{v}}|\hat{Z}_\nu'\hat{Z}_\nu - \hat{Z}_0'\hat{Z}_0| = O_p(1). \tag{A.51}$$

Second, by Lemma 2

$$n^{-1}\hat{Z}_\nu'\hat{Z}_\nu \Rightarrow M(\gamma_0) = M_0. \tag{A.52}$$

Then, by (A.44), (A.51), (A.52), Lemma 2, (A.40), (17), and Lemma A.10 of Hansen (2000),

$$n^{-2\alpha}c'G_0'(P_0^* - P_\nu^*)G_0 c'$$

$$= n^{-2\alpha}c'(\hat{Z}_\nu'\hat{Z}_\nu - \hat{Z}_0'\hat{Z}_0)c$$

$$- n^{-2\alpha}c'(\hat{Z}_\nu'\hat{Z}_\nu - \hat{Z}_0'\hat{Z}_0)(I_m - (\hat{Z}_\nu'\hat{Z}_\nu)^{-1}\hat{Z}_0'Z_0)c$$

$$- c'(I_m - G_0'\hat{Z}_0(\hat{Z}_0'\hat{Z}_0)^{-1})n^{-2\alpha}(\hat{Z}_\nu'\hat{Z}_\nu - \hat{Z}_0'\hat{Z}_0)(\hat{Z}_\nu'\hat{Z}_\nu)^{-1}\hat{Z}_0'G_0 c$$

$$= n^{-2\alpha}\sum_{i=1}^{n}(c'g_i)^2\Delta_i(\nu) + o_p(1)$$

$$\Rightarrow \mu|\nu|.$$

This is (A.48).

By Lemma 2 and (A.52), uniformly in $\nu \in [-\bar{v},\bar{v}]$,

$$n^{\alpha}(\hat{Z}_\nu'\hat{Z}_\nu)^{-1}\hat{Z}_0'\hat{v} = (n^{-1}\hat{Z}_\nu'\hat{Z}_\nu)^{-1}n^{-(1-\alpha)}\hat{Z}_0'\hat{v} = o_p(1). \tag{A.53}$$

By (17)

$$
\begin{aligned}
n^{-\alpha}(\hat{Z}_\nu'\hat{v} - \hat{Z}_0'\hat{v}) &= n^{-\alpha}\sum_{i=1}^{n}\hat{g}_i\hat{v}_i\Delta_i(\nu) \\
&= n^{-\alpha}\sum_{i=1}^{n}\hat{g}_i\hat{r}_i'\theta_2\Delta_i(\nu) + n^{-\alpha}\sum_{i=1}^{n}g_i v_i\Delta_i(\nu) - n^{-\alpha}\sum_{i=1}^{n}\hat{r}_i v_i\Delta_i(\nu) \\
&= n^{-\alpha}\sum_{i=1}^{n}g_i v_i\Delta_i(\nu) + o_p(1) \\
&\Rightarrow B(\nu), \qquad \text{(A.54)}
\end{aligned}
$$

a vector Brownian motion with covariance matrix $D_2 f$, where the final convergence uses Lemma A.11 of Hansen (2000). Thus by (A.45), (A.51), (A.52), (A.53), and (A.54),

$$
\begin{aligned}
n^{-\alpha}G_0'(P_0^* - P_\nu^*)\hat{v} &= G_0'\hat{Z}_0(\hat{Z}_0'\hat{Z}_0)^{-1}n^{-2\alpha}(\hat{Z}_\gamma'\hat{Z}_\gamma - \hat{Z}_0'\hat{Z}_0)n^{\alpha}(\hat{Z}_\nu'\hat{Z}_\nu)^{-1}\hat{Z}_0'\hat{v} \\
&\quad - G_0'\hat{Z}_0(\hat{Z}_\nu'\hat{Z}_\nu)^{-1}n^{-\alpha}(\hat{Z}_\nu'\hat{v} - \hat{Z}_0'\hat{v}) \\
&= -n^{-\alpha}(\hat{Z}_\nu'\hat{v} - \hat{Z}_0'\hat{v}) + o_p(1) \\
&\Rightarrow B(\nu).
\end{aligned}
$$

This yields (A.49).

Finally, by (A.47), (A.51), (A.53), and (A.54)

$$
\begin{aligned}
\hat{v}'(P_0 - P_\nu)\hat{v} &= n^{\alpha}\hat{v}'\hat{Z}_0(\hat{Z}_0'\hat{Z}_0)^{-1}n^{-2\alpha}(\hat{Z}_\nu'\hat{Z}_\nu - \hat{Z}_0'\hat{Z}_0)n^{\alpha}(\hat{Z}_\nu'\hat{Z}_\nu)^{-1}\hat{Z}_0'\hat{v} \\
&\quad - 2n^{-\alpha}(\hat{v}'\hat{Z}_\nu - \hat{v}'\hat{Z}_0)n^{\alpha}(\hat{Z}_\nu'\hat{Z}_\nu)^{-1}\hat{Z}_0'\hat{v} \\
&= o_p(1)
\end{aligned}
$$

uniformly in $\nu \in [-\bar{v},\bar{v}]$. A similar argument applies to $\hat{v}'(P_{\perp 0} - P_{\perp\nu})\hat{v}$. Combined with (A.46) this establishes (A.50) and completes the proof. ■

Proof of Theorem 1. By Lemma 4, $a_n(\hat{\gamma} - \gamma_0) = \text{argmax}_\nu Q_n(\nu) = O_p(1)$ and by Lemma 5, $Q_n(\nu) \Rightarrow -\mu|\nu| + 2\lambda^{1/2}W(\nu)$, where the limit functional is continuous with a unique maximum almost surely. Appealing to Theorem 2.7 of Kim and Pollard (1990), (18) and (19) follow by the argument in the proofs of Theorem 1 and 2 of Hansen (2000). ■

The 2SLS and GMM estimators of θ_1 and θ_2 introduced in Section 3 are special cases of the class of estimators

$$\bar{\theta}_1 = (\hat{Z}_1'\hat{X}_1\hat{W}_1\hat{X}_1'\hat{Z}_1)^{-1}(\hat{Z}_1'\hat{X}_1\hat{W}_1\hat{X}_1'Y), \qquad \text{(A.55)}$$

$$\bar{\theta}_2 = (\hat{Z}_2'\hat{X}_2\hat{W}_2\hat{X}_2'\hat{Z}_2)^{-1}(\hat{Z}_2'\hat{X}_2\hat{W}_2\hat{X}_2'Y), \qquad \text{(A.56)}$$

where $\hat{W}_1$ and $\hat{W}_2$ are sequences of weight matrices.

LEMMA 6. *If* $\hat{W}_1 \to_p W_1 > 0$ *and* $\hat{W}_2 \to_p W_2 > 0$ *then*

$$\sqrt{n}(\bar{\theta}_1 - \theta_1) \to_d N(0, \bar{V}_1),$$

$$\sqrt{n}(\bar{\theta}_2 - \theta_2) \to_d N(0, \bar{V}_2),$$

where

$$\bar{V}_1 = (R_1' W_1 R_1)^{-1} R_1' W_1 Q_1 W_1 R_1 (R_1' W_1 R_1)^{-1}, \tag{A.57}$$

$$V_2 = (R_2' W_2 R_2)^{-1} R_2' W_2 Q_2 W_2 R_2 (R_2' W_2 R_2)^{-1}. \tag{A.58}$$

Proof of Lemma 6. We provide the details of the proof for $\bar{\theta}_1$. Let Z_v, $Z_\perp$, ΔZ_v, X_v denote the matrices obtained by stacking, respectively,

$$z_i' 1(q_i \le \gamma_0 + n^{-(1-2\alpha)} v),$$

$$z_i' 1(q_i > \gamma_0 + n^{-(1-2\alpha)} v),$$

$$z_i' 1(q_i \le \gamma_0 + n^{-(1-2\alpha)} v) - z_i' 1(q_i \le \gamma_0),$$

$$x_i' 1(q_i \le \gamma_0 + n^{-(1-2\alpha)} v).$$

By Lemma 1 of Hansen (1996), Lemma A.4 of Hansen (2000), and Lemma A.10 of Hansen (2000), uniformly on $\nu \in [-\bar{v}, \bar{v}]$,

$$\frac{1}{n} X_v' Z_v \to_p R_1, \tag{A.59}$$

$$\frac{1}{\sqrt{n}} X_v' e \Rightarrow N_1 \sim N(0, \Omega_1), \tag{A.60}$$

$$\frac{1}{n^{2\alpha}} X_v' \Delta Z_v = O_p(1). \tag{A.61}$$

Let

$$\bar{\theta}_1(v) = (Z_v' X_v \hat{W}_1 X_v' Z_v)^{-1} (Z_v' X_v \hat{W}_1 X_v' Y).$$

A little rewriting of the model shows that

$$Y = Z_v \theta_1 + Z_\perp \theta_2 - \Delta Z_v \delta_n + e.$$

Uniformly on $\nu \in [-\bar{v}, \bar{v}]$, by (A.59)–(A.61),

$$\begin{aligned}\sqrt{n}(\bar{\theta}_1(v) - \theta_1) &= \left(\frac{1}{n} Z_v' X_v \hat{W}_1 \frac{1}{n} X_v' Z_v\right)^{-1} \left(\frac{1}{n} Z_v' X_v \hat{W}_1 \left(\frac{1}{\sqrt{n}} X_v' e - \frac{1}{\sqrt{n}} X_v' \Delta Z_v \delta_n\right)\right) \\ &\Rightarrow (R_1' W_1 R_1)^{-1} (R_1' W_1 N_1).\end{aligned}$$

Because $\hat{v} = n^{1-2\alpha}(\hat{\gamma} - \gamma_0) = O_p(1)$ and $\bar{\theta}_1 = \bar{\theta}_1(\hat{v})$, it follows that

$$\sqrt{n}(\bar{\theta}_1 - \theta_1) = \sqrt{n}(\bar{\theta}_1(\hat{v}) - \theta_1) \Rightarrow (R_1' W_1 R_1)^{-1} (R_1' W_1 N_1) \sim N(0, \bar{V}_1)$$

as stated. ∎

Proof of Theorem 2. The 2SLS estimators $(\tilde{\theta}_1, \tilde{\theta}_2)$ fall in the class (A.55) and (A.56) with

$$\hat{W}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' 1(q_i \le \hat{\gamma}),$$

$$\hat{W}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' 1(q_i > \hat{\gamma}).$$

By Lemma 1 of Hansen (1996) and the consistency of $\hat{\gamma}$, $\hat{W}_1 \to_p Q_1$ and $\hat{W}_2 \to_p Q_2$. Thus $(\tilde{\theta}_1, \tilde{\theta}_2)$ are asymptotically normal, with covariance matrices given by the formula (A.57) and (A.58) with Q_1 and Q_2 replacing W_1 and W_2, yielding the stated result. ■

Proof of Theorem 3. Let

$$\hat{\Omega}_1(\gamma) = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' \tilde{e}_i^2 1(q_i \le \gamma).$$

It will be enough to show that

$$\hat{\Omega}_1(\gamma) \to_p E(x_i x_i' e_i^2 1(q_i \le \gamma)) \tag{A.62}$$

uniformly in $\gamma \in \Gamma$, for then by the consistency of $\hat{\gamma}$, $n^{-1}\hat{\Omega}_1 = \hat{\Omega}_1(\hat{\gamma}) \to_p \Omega_1$, and the theorem follows by Lemma 6. Hence we show (A.62).

Set $z_i^* = (z_i' 1(q_i \le \gamma_0), z_i' 1(q_i > \gamma_0))'$, $\Delta\hat{z}_i = z_i(1(q_i \le \hat{\gamma}) - 1(q_i \le \gamma_0))$, and $\tilde{\delta} = \tilde{\theta}_1 - \tilde{\theta}_2$. Algebraic manipulation shows that

$$\tilde{e}_i = e_i - z_i^{*\prime}(\tilde{\theta} - \theta) - \Delta\tilde{z}_i' \tilde{\delta}.$$

Hence

$$\begin{aligned}
\hat{\Omega}_1(\gamma) - \frac{1}{n}\sum_{i=1}^{n} x_i x_i' e_i^2 1(q_i \le \gamma) = &-\frac{2}{n}\sum_{i=1}^{n} x_i x_i' 1(q_i \le \gamma) e_i z_i^{*\prime}(\tilde{\theta} - \theta) \\
&- \frac{2}{n}\sum_{i=1}^{n} x_i x_i' 1(q_i \le \gamma) e_i \Delta\hat{z}_i' \tilde{\delta} \\
&+ \frac{1}{n}\sum_{i=1}^{n} x_i x_i' 1(q_i \le \gamma)(\tilde{\theta} - \theta)' z_i^* z_i^{*\prime}(\tilde{\theta} - \theta) \\
&+ \frac{2}{n}\sum_{i=1}^{n} x_i x_i' 1(q_i \le \gamma)(\tilde{\theta} - \theta)' z_i^* \Delta\hat{z}_i' \tilde{\delta} \\
&+ \frac{1}{n}\sum_{i=1}^{n} x_i x_i' 1(q_i \le \gamma) \tilde{\delta}' \Delta\hat{z}_i \Delta\hat{z}_i' \tilde{\delta}.
\end{aligned}$$

It is straightforward to show that the terms on the right-hand side converge in probability to zero, uniformly in γ. For example, the first term is bounded by

$$\frac{2}{n}\left|\sum_{i=1}^{n} x_i x_i' 1(q_i \le \gamma) e_i z_i^{*\prime}(\tilde{\theta} - \theta)\right| \le \frac{2}{n}\sum_{i=1}^{n} |x_i|^2 |e_i||z_i||\tilde{\theta} - \theta| \to_p 0$$

because the data have bounded fourth moments and $|\tilde{\theta} - \theta| \to_p 0$. Hence,

$$\hat{\Omega}_1(\gamma) = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' e_i^2 1(q_i \leq \gamma) + o_p(1) \to_p E(x_i x_i' e_i^2 1(q_i \leq \gamma))$$

uniformly in γ, by Lemma 1 of Hansen (1996), which is (A.62). This completes the proof. ■

Proof of Theorem 4. Under the null of $\theta_1 = \theta_2$,

$$\hat{\theta}_1(\gamma) - \theta_1 = (\hat{Z}_1'\hat{X}_1\tilde{\Omega}_1^{-1}\hat{X}_1'\hat{Z}_1)^{-1}(\hat{Z}_1'\hat{X}_1\tilde{\Omega}_1^{-1}\hat{X}_1'e).$$

Then by Lemma 1 of Hansen (1996) and Assumption 1.3, uniformly in γ

$$\frac{\hat{X}_1'\hat{Z}_1}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i z_i' 1(q_i \leq \gamma) \xrightarrow{p} Q_1(\gamma). \tag{A.63}$$

Via Lemma A.4 of Hansen (2000)

$$\frac{\hat{X}_1'e}{n} \Rightarrow S_1(\gamma). \tag{A.64}$$

Use (A.63) and (A.64) with (A.62) to have

$$n^{1/2}(\hat{\theta}_1(\gamma) - \theta_1) \Rightarrow V_1(\gamma)Q_1(\gamma)'\Omega_1(\gamma)^{-1}S_1(\gamma). \tag{A.65}$$

In the same manner we derive

$$n^{1/2}(\hat{\theta}_2(\gamma) - \theta_2) \Rightarrow V_2(\gamma)Q_2(\gamma)'\Omega_2(\gamma)^{-1}S_2(\gamma).$$

A similar argument applies to the covariance matrices. Combining these results completes the proof. ■