Land Economics 94:3, August 2018
"District Subdivision and the Location of Smallholder Forest Conversion in Sumatra"
by Marc N. Conte and Philip Shaw

APPENDIX A

Nonparametric methods

This section provides additional detail about the nonparametric approach employed in our empirical analysis. The most important consideration when estimating a nonparametric model is the choice of bandwidth parameters γ_1 and γ_2. The likelihood cross-validation method is used to calculate the bandwidths. This approach minimizes the following expression:

$$L = -\sum_{j=1}^{n} \ln \hat{\phi}_{-j}(R_j \mid X_j) \qquad [1]$$

where $\hat{\phi}_{-j}$ is the leave-one-out estimator of the conditional density. As noted by Hall (1987a) and Hall (1987b), using likelihood cross-validation can produce bandwidths that oversmooth the data in the presence of fat-tailed distributions. This outcome could potentially lead to inconsistent estimates for $\phi(k,x)$.[1] An alternative to likelihood cross-validation is least squares cross-validation, which is shown to asymptotically smooth out irrelevant regressors in Hall et al. (2004). We are modeling conditional probability functions, which have a bounded range, ruling out fat-tails, so we employ likelihood cross-validation.

Racine et al. (2004) shows that the estimator presented in equation [1] performs very well in Monte Carlo simulations under a variety of data-generating processes, including a probit and multinomial probit specification. The authors also show that their estimator dominates mis-specified parametric estimators and only exhibits a slight loss of efficiency when compared to correctly specified parametric models. Shaw et al. (2015) shows that the nonparametric estimator is also robust to endogeneity generated under a standard bivariate or simultaneous probit population model. Given the similarity between probit and logit models, we expect similarly strong performance from this estimator when applied to logit models as well.

Although the nonparametric estimation of the conditional density is informative, we are more interested in presenting the average derivatives for the continuous variables and the average

[1] Given that our nonparametric estimates substantially outperform all of the alternative parametric models under consideration, loss of consistency is not a valid concern in our case, which differs meaningfully from the conditions explored in Hall (1987a) and Hall (1987b). See figures 4-6 for a graphical representation of relative performance.

differences for the discrete variables. If X_1^c is the continuous variable of interest, then the population average derivative can be expressed as:

$$\theta = E\left[\frac{\partial P(R_j = k \mid X_1^c, X_2^c, \ldots, X_{r_1}^c, X^d)}{\partial X_1^c}\right] = \int_{X_1 \in S_1^c} \int_{X_2 \in S_2^c} \cdots \int_{X_{r_1} \in S_{r_1}^c} \sum_{X^d \in S_{r_2}^d} \frac{\partial \phi(X_1^c, X_2^c, \ldots, X_{r_1}^c, X^d)}{\partial X_1^c}$$

$$\times f(X_1^c, X_2^c, \ldots, X_{r_1}^c, X^d) dX_1 dX_2 \ldots dX_{r_1} \qquad [2]$$

where $f(X_1^c, X_2^c, \ldots, X_{r_1}^c, X^d)$ is the joint density function for all the continuous variables X^c and the discrete variables X^d with X_1^c being the variable of interest. The main focus on estimating the average derivative and the average difference is due to the fact that these estimators are easy to compare to existing parametric estimates and they have a much faster rate of convergence as compared to pointwise nonparametric estimates. Härdle and Stoker (1989) and Coppejans and Sieg (2005) show that common average derivative and difference estimators exhibit root-N convergence even when the pointwise derivatives converge at a much slower rate.

To estimate θ, the conditional probability is first estimated and then a central-difference formula is employed for continuous variables. The estimator can be expressed as follows:

$$\hat{\theta} = \frac{1}{n}\sum_{j=1}^{n}\left[\frac{\hat{\phi}(k \mid X_{j1}^c + \tau, X_{j2}^c, \ldots, X_{jr_1}^c, X_j^d) - \hat{\phi}(k \mid X_{j1}^c - \tau, X_{j2}^c, \ldots, X_{jr_1}^c, X_j^d)}{2\tau}\right] \qquad [3]$$

where τ is the chosen approximation error.[2] For discrete variables, the average difference estimator can be constructed as follows:

$$\hat{\theta} = \frac{1}{n}\sum_{j=1}^{n}\left[\hat{\phi}(k \mid X_{j1}^c, X_{j2}^c, \ldots, X_{jr_1}^c, X_j^d = c_1) - \hat{\phi}(k \mid X_{j1}^c, X_{j2}^c, \ldots, X_{jr_1}^c, X_j^d = c_2)\right] \qquad [4]$$

where c_1 and c_2 are discrete values of interest. If X_d is a zero-one variable then we might have $c_1 = 0$ and $c_2 = 1$.

---

[2] It is well known that truncation error for the central-difference approximation is $O(\tau^2)$. τ is chosen to be $1\text{x}10^{-12}$, which is smaller than the relative error due to rounding in floating point arithmetic.

References

Coppejans, Mark, and Holger Sieg. April 2005. "Kernel Estimation of Average Derivatives and Differences." *Journal of Business & Economic Statistics*, 23(2):211–225.

Hall, Peter. 1987a. "On Kullback-Leibler Loss and Density Estimation." *The Annals of Statistics*, 15(4): 1491–1519.

Hall, Peter. 1987b. "On the Use of Compactly Supported Density Estimates in Problems of Discrimination." *Journal of Multivariate Analysis*, 23:131–158.

Hall, Peter, Jeffrey S. Racine, and Qi Li. 2004. "Cross-validation and the Estimation of Conditional Probability Densities." *Journal of the American Statistical Association*, 99(468):1015–1026.

Härdle, W., and T. Stoker. December 1989. "Investigating Smooth Multiple Regression by the Method of Average Derivatives." *Journal of the American Statistical Association*, 84(408):986–995.

Imbens, G.W., and J.W. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, 47(1):5–86.

Racine, Jeffrey S., Qi Li, and Xi Zhu. 2004. "Kernel Estimation of Multivariate Conditional Distributions. *Annals of Economics and Finance*, 5:211–235.

Rosenbaum, P.R., and D.B. Rubin. 1985."Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *The American Statistician*, 39 (1):33–38.

Rubin, Donald B. 2001. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services and Outcomes Research Methodology*, 2(3):169–188.

Shaw, Philip, Marina-Selini Katsaiti, and Brandon Pecoraro. 2015. "On the Determinants of Educational Corruption: The Case of Ukraine." *Contemporary Economic Policy*, 33(4):698–715.

Stuart, Elizabeth A. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1):1–21.