Controlling for ability using test scores

Benjamin Williams*
George Washington University

July 23, 2018

Abstract

This paper proposes a semiparametric method to control for ability using standardized test scores, or other item response assessments, in a regression model. The proposed method is based on a model in which the parameter of interest is invariant to monotonic transformations of ability. I show that the estimator is consistent as both the number of observations and the number of items on the test grow to infinity. I also derive conditions under which this estimator is root-n consistent and asymptotically normal. The proposed method is easy to implement, does not impose a parametric item response model, and does not require item level data. I demonstrate the finite sample performance in a Monte Carlo study and implement the procedure for a wage regression using data from the NLSY1979.

JEL codes: C14, C38, C39, C55, I21, I26, J24

Keywords: test scores, scale invariance, measurement error, item response, ability bias

*Department of Economics, George Washington University, 2115 G St. NW, Washington DC 20052, phone: (202)994-6685, email: bdwilliams@gwu.edu. This paper was previously circulated under the title "Bias correction for the use of educational and psychological assessments as covariates in linear regression". The paper has benefited greatly from comments from Don Parsons and Robert Phillips, invaluable suggestions from the editor and anonymous referees, and correspondence with Dan Black and Lynne Schofield. I also thank Eyan Fu for excellent research assistance.

1 Introduction

In classical test theory (CTT), a test score, typically measured as the number or percent of items answered correctly, is viewed as an error-laden measurement of ability (Lord and Novick, 1968). In item response theory (IRT), performance on a test is modeled at the level of each individual question, or item, and such models are used to develop more sophisticated scoring of tests (van der Linden and Hambleton, 1997). The common practice of controlling for ability by conditioning on test scores, and the interpretation of the test score as a "proxy" (in the sense of Wooldridge, 2009), is not consistent with standard CTT or IRT models.[1] In this paper, I use item response theory to argue that there are two important distinctions between latent ability and measured test scores. First, to the extent that adding items to a test will change test scores, due to idiosyncratic factors influencing responses to each item, but will not change ability, the test score is contaminated with measurement error. Second, since ability is measured on an ordinal scale (Lord, 1975; Zwick, 1992), any test score imposes an arbitrary scale restriction and regression estimates are dependent on this scale.

This paper studies these problems in the context of a regression model with binary item responses. I propose estimating a partially linear regression model using the double residual method of Robinson (1988). The paper makes three main theoretical contributions. First, I argue that the partially linear regression model is invariant to monotonic transformations of ability. Second, I show that the estimator is consistent as both the number of observations (n) and the number of items on the test (J) grow to infinity. This type of asymptotic analysis has been used previously to study IRT models with a large number of items (Haberman, 1977; Douglas, 1997; Williams, 2018a) and alternative asymptotic sequences have been used similarly in other contexts as well (Chesher, 1991; Bekker, 1994; Hahn and Kuersteiner, 2002). Third, I derive conditions under which this estimator is root-n consistent and asymptotically normal. These conditions include the requirement that $\sqrt{n}/J \to 0$, which ensures that the measurement error bias is asymptotically negligible. The method proposed in this paper is easy to implement, does not impose a parametric item response model, and does not require item level data. In a companion paper, Williams (2018b), I discuss the possibility of bias-correction when $\sqrt{n}/J \to \gamma > 0$ with item level data.

This paper is related to two separate lines of research. Mislevy (1991) argues for the publication of institutional plausible values rather than test scores to address the measurement error. Schofield et al. (2014) raise some concerns with this approach and suggest correcting for measurement error by jointly modeling the economic outcome and the test items (see also Junker et al., 2012; Schofield, 2015). Lockwood and McCaffrey (2014) evaluate the effect of test score measurement error in teacher value-added estimates and propose various methods for bias correction. These papers rely on

[1] An exception is if the outcome is determined in part by the test score itself, not latent ability. This is the case, for example, if the outcome is college enrollment and the test score is the individual's SAT score since scores on the SAT are used to determine acceptance into college.

parametric item response models to correct for measurement error and assume that ability enters the outcome equation linearly. Ballou (2009), Ho (2009), Bond and Lang (2013), Nielsen (2015), and others have studied the scale-dependence of test scores, focusing primarily on difficulties this poses for measuring differences in ability across subpopulations or over time. This paper complements this important work.

The methodological issues studied in this paper are also closely related to problems studied in the measurement error literature (Hu and Schennach, 2008; Cunha et al., 2010; Hu and Sasaki, 2015, 2017; Hu, 2017; Agostinelli and Wiswall, 2017). Hu and Sasaki (2015) and Hu and Sasaki (2017) emphasize that only one measurement needs to be measured to scale. Agostinelli and Wiswall (2017) show that the scale problem can be partially resolved by using economic restrictions in a model for the dynamic evolution of latent ability. The analysis in this paper differs in that ability is allowed to be continuously distributed, while the test score is not continuously distributed as it is constructed from binary responses to individual items. Latent variables models with this structure have received less attention in econometrics. One exception is Spady (2007) who jointly models economic behaviors and individual responses to questionnaires meant to elicit cultural and economic attitudes using a flexible parametric item response model to estimate the effect of these attitudes on behaviors. See also Williams (2018a).

Controlling for ability is particularly important in estimating the returns to schooling (Becker, 1967). Various methods have been used to control for ability bias while addressing the measurement error problem. Bollinger (2003) uses the Klepper and Leamer (1984) bounds to study the black-white wage gap, controlling for ability. Goldberger (1972), Chamberlain and Griliches (1975), and others have employed structural equation modeling in this setting and, more recently, more sophisticated methods for using additional measurements in a factor analytic approach in linear and nonlinear models have been developed (Cunha et al., 2010; Heckman et al., 2013). Many empirical studies of the returns to education use a test score to control for ability without addressing the measurement error in the statistical analysis. The results of this paper provide a formal foundation for this approach.

In an empirical application, I demonstrate the method developed in this paper in a regression model for the returns to education that controls for ability using a sample from the NLSY1979. I show that regression estimates of the returns to schooling are sensitive to the scale of the test score. Compared to estimates from regression models where the test score enters the outcome equation linearly, estimates from the partially linear regression model suggest a 10% larger effect of education on wages in the subsample of white males. Similar results are found for other subgroups though sometimes the correction is in the opposite direction.

The use of a test score in place of latent ability is common in many other applications where conditioning on ability (or other traits) is necessary. See Junker et al. (2012) for a discussion of examples in labor economics. Teacher value-added models that are used to identify the role of individual teachers in student learning are often estimated with test scores without accounting for measurement error (Chetty et al., 2014; Lockwood and McCaffrey, 2014).[2] More generally, the use of a scalar "score" that aggregates individual items in place of a latent variable is widespread, not only for test scores but also in personality assessments, political opinion polls, happiness scales, mental health diagnostics, and many other areas. This approach has also been used to control for hard to measure economic primitives. Bloom and Van Reenen (2007), for example, aggregate discrete items from a survey on managerial practices to control for managerial productivity in estimation of a production function.

The rest of the paper is structured as follows. In Section 2, I describe an item response model and outline the main results. In Section 3, I describe the proposed semiparametric estimator, provide the main theoretical results regarding this estimator, and provide some Monte Carlo evidence regarding its behavior in finite samples. In Section 4, the proposed method is used to estimate the return to education controlling for various dimensions of ability. Section 5 discusses the lack of a need for a conditioning model and discusses some related work. Section 6 concludes. In Appendix A I state additional technical conditions for the theoretical results and all proofs are contained in the supplementary appendix.

2 An item response model and the bias of OLS

Let $\mathbf{M}_i = (M_{i1}, \ldots, M_{iJ})$ where each M_{ij} denotes individual i's binary response to the j^{th} item on the test. The theoretical results in this paper pertain to the percent correct measure of ability, $\bar{M}_i = J^{-1}\sum_{j=1}^{J} M_{ij}$. Item response theory provides microfoundations for this nonparametric estimate of ability (Lord and Novick, 1968). As a starting point, consider the following restrictions on the joint distribution of the item responses, $\mathbf{M}_i$, and the latent ability, θ_i, which are typical in the item response literature (Sijtsma and Junker, 2006).

Assumption 2.1.

LI *(Local independence) For any* $\mathbf{m} \in \{0,1\}^J$,

$$Pr(\mathbf{M}_i = \mathbf{m} \mid \theta_i) = \prod_{j=1}^{J} Pr(M_{ij} = m_j \mid \theta).$$

U *(Unidimensionality)* θ_i *is scalar.*

[2]Teacher value-added in some cases is estimated as the coefficients on teacher dummies in a regression of test scores on the previous year's test scores.

M *(Monotonicity) The item response functions, $p_j(\theta_i) := Pr(M_{ij} = 1 \mid \theta_i)$, are nondecreasing in θ_i, for each j.*

Most parametric IRT models for binary responses can be derived from a linear latent index structure. For example, consider the model $M_{ij} = \mathbf{1}(\delta_j(\theta_i - \alpha_j) \geq \varepsilon_{ij})$ where ε_{ij} is independent of θ_i and is assumed to follow a standard normal distribution or a logistic distribution. Under this parameterization, δ_j is called the discrimination parameter and α_j is called the difficulty parameter. If ε_{ij} is normally distributed, this is commonly referred to as the normal ogive model (Lord, 1952). The Rasch model (Rasch, 1961) imposes $\delta_j = 1$ and uses the logistic distribution for ε_{ij}. The three parameter logistic (3PL) model is a popular extension of the Rasch model that allows for a nonzero probability of a positive response when $\theta_i \to -\infty$ (Birnbaum, 1968). In the 3PL model the item response functions take the form

$$\gamma_j + \frac{1 - \gamma_j}{1 + \exp(-\delta_j(\theta_i - \alpha_j))}.$$

While these models have been extended to models with multinomial M_{ij}, models with multiple dimensions of θ_i, ordered response models, models of adaptive tests, and models with partial credit scoring, among many others (van der Linden and Hambleton, 1997), it is still common practice to score tests using the 3PL model. In this paper, however, I do not specify any functional form for the item response functions.

The score $\bar{M}_i$ is a noisy measure of a monotonic transformation of ability. To see this, define $\bar{p}(\theta_i) := J^{-1} \sum_{j=1}^{J} p_j(\theta_i)$. Then $\bar{M}_i = \bar{p}(\theta_i) + \eta_i$ where $E(\eta_i \mid \theta_i) = 0$ because $\bar{p}(\theta_i) = E(\bar{M}_i \mid \theta_i)$. So $\bar{M}_i$ is a noisy measure of $\bar{p}(\theta_i)$ with mean zero measurement error η_i. If $E(\bar{M}_i \mid \theta_i, X_i) = E(\bar{M}_i \mid \theta_i)$ then estimation of a regression of $\bar{M}_i$ (as the *dependent* variable) on explanatory variables, X_i, is not biased due to the presence of measurement error (i.e., regression slope estimates are not biased). The same conclusion holds if $E(\bar{M}_i \mid \theta_i, Z_i) = E(\bar{M}_i \mid \theta_i)$ for a vector of variables Z_i that includes X_i (Schofield et al., 2014).

Nevertheless, use of $\bar{M}_i$ imposes a particular scale on the distribution of ability through the function $\bar{p}$. Ho (2009), Bond and Lang (2013), and Nielsen (2015), among others, have demonstrated how taking different monotonic transformations can change estimates in models with a test score as a dependent variable, sometimes dramatically.[3] To address this problem, Ho (2009) proposes methods for estimating these objects that are scale-invariant.[4] If condition **M** of Assumption 2.1 is strengthened so that $\bar{p}$ is a strictly monotonic function (not just a nondecreasing function), then $\bar{p}(\theta_i)$ is a valid measure of ability when using scale-invariant statistics such as those proposed by Ho

[3] Bond and Lang (2013), for example, demonstrate how this affects estimates of the black-white gap and trends over time in this gap in early childhood. They find that under some monotonic transformations the gap widens substantially between Kindergarten and 3rd grade but under other monotonic transformations it does not grow at all.

[4] Altonji et al. (2012) propose a related method for converting two tests to the same scale.

(2009).

While this paper studies the use of $\bar{M}_i$ as a measure of ability, similar results could be derived for the case where the score consists of an estimate from a parametric IRT model. Parametric IRT models, such as the 3PL model, are typically estimated via maximum likelihood or other likelihood-based methods. The item parameters, α_j, δ_j, and γ_j, as well as the individual ability parameters, θ_i, are estimated. In large-scale education surveys, reported test scores often take the form of estimates of θ_i from such a model. See, for example, NCES (2009) and Ing et al. (2012). One common methodology in such studies involves first estimating the item parameters using a marginal maximum likelihood method with a discrete distribution for θ_i, and second estimating θ_i by separately maximizing the likelihood of each individual i at the estimated item parameter values (Mislevy and Bock, 1990; Muraki and Bock, 1997).

These estimates, $\hat{\theta}_i$, similarly exhibit measurement error and are scale-dependent. Even if the item parameters were known, the score $\hat{\theta}_i$ is an estimate based on a sample of size J. Therefore, for each i, $\hat{\theta}_i = \theta_i + \nu_i$ where ν_i represents finite sample estimation error. Typically ν_i is only mean zero in the limit as $J \to \infty$. Furthermore, as ability is understood by most psychologists to be ordinal, not interval-scaled (Lord, 1975; Zwick, 1992; Ballou, 2009), the criticisms of Ho (2009), Bond and Lang (2013), and Nielsen (2015) regarding scale-dependence apply to the IRT estimate of ability too.

While the methods proposed here are meant to complement, rather than replace, other methods based on $\hat{\theta}_i$, there are several advantages to using the percent correct score, $\bar{M}_i$. First, it does not rely on the correct specification of an item response model.[5] Second, it is much simpler to calculate than $\hat{\theta}_i$ and, for this reason, it is commonly used in situations where only raw test results are available. Third, use of $\bar{M}_i$ obviates the need for consistency between the econometrician's model and the model of the primary analyst who reports the test score (see Mislevy, 1991; Schofield et al., 2014, for a discussion of this problem). Finally, $\bar{M}_i$, is a sufficient statistic in the Rasch model, and it has also played a central role in nonparametric item response models (Mokken, 1971; Stout, 1987, 1990; Junker, 2001). One apparent disadvantage of the use of $\bar{M}_i$, rather than the IRT score, $\hat{\theta}_i$, is that it can be less efficient as it, for example, provides equal weight to items of varying discrimination.

What if, as is common practice among economists, and researchers in other fields, scores such as $\hat{\theta}_i$ or $\bar{M}_i$ are used as covariates in a regression to control for ability? The measurement error and scale-dependence of these test scores translate to biased estimates of causal parameters.[6] Again, I

[5] On the other hand, Junker et al. (2012) argue that it is natural to rely on the parametric IRT model used to develop the test.

[6] In principle, the fact that ability is ordinal should not bias estimates of causal parameters. Suppose, for example, that $X_i = (D_i, W_i)$ where D_i is a treatment indicator or binary choice and suppose that $E(Y_i \mid D_i = d, W_i, \theta_i) = E(Y_{di} \mid W_i, \theta_i)$, using the potential outcomes notation. This is the unconfoundedness or selection on observables assumption. Then, if $\theta_i' = \tau(\theta_i)$, as long as τ is a strictly monotonic transformation, $E(Y_i \mid D_i = d, W_i, \theta_i') = E(Y_i \mid W_i, \theta_i')$. Thus treatment effect parameters are identified using either θ_i or θ_i'. However, the linear model $Y_i = \beta_0 + \beta_1 D_i + \beta_2' W_i + \beta_3 \theta_i + u_i$ is clearly not scale-invariant in this sense. Indeed, this is the principle underlying

focus on the use of $\bar{M}_i$ but similar results can be shown for $\hat{\theta}_i$. Let $\hat{\beta}^{OLS}$ denote the OLS estimator obtained by regressing Y_i on X_i and $\bar{M}_i$,

$$\hat{\beta}^{OLS} := \left(\sum_{i=1}^{n} W_i W_i' \right)^{-1} \sum_{i=1}^{n} W_i Y_i,$$

where $W_i = (1, X_i', \bar{M}_i)'$. Also, let $\tilde{\beta}^{OLS}$ denote the infeasible version of this estimator that replaces W_i with $W_i^* = (1, X_i', \bar{p}(\theta_i))'$. First, consider the following conditions, which can be used to relate the feasible estimator $\hat{\beta}^{OLS}$ to the infeasible estimator $\tilde{\beta}^{OLS}$.

Assumption 2.2.

LIx *(Local independence) For any* $\mathbf{m} \in \{0,1\}^J$,

$$Pr(\mathbf{M}_i = \mathbf{m} \mid X_i, \theta_i) = \prod_{j=1}^{J} Pr(M_{ij} = m_j \mid X_i, \theta_i).$$

U *(Unidimensionality)* θ_i *is scalar.*

ER *(Exclusion restriction)* $Pr(M_{ij} = 1 \mid X_i, \theta_i) = Pr(M_{ij} = 1 \mid \theta_i)$ *for each* j.

M' *(Monotonicity) The average item response functions,* $\bar{p}(\theta_i)$, *is strictly increasing in* θ_i.

ND *(Non-differential measurement)* $Y_i \perp\!\!\!\perp \mathbf{M}_i \mid \theta_i, X_i$.

The first four conditions of Assumption 2.2 extend the IRT assumptions in Assumption 2.1 to restrict the distribution of $\mathbf{M}_i \mid X_i, \theta_i$, rather than the distribution of $\mathbf{M}_i \mid \theta_i$. Also, the monotonicity assumption has been modified because, as will be demonstrated below, condition **M** of Assumption 2.1 is not sufficient for the use of $\bar{M}_i$ in place of θ_i in a regression analysis. Condition **ND** requires the test score to be uninformative about the outcome, Y_i, conditional on latent ability and X_i. Assumption 2.2 will be discussed further below.

The feasible estimator $\hat{\beta}^{OLS}$ is asymptotically biased relative to the infeasible estimator $\tilde{\beta}^{OLS}$ due to the presence of measurement error. Under condition **ND** of Assumption 2.2, as $n \to \infty$, for a fixed J,

$$\hat{\beta}^{OLS} - \tilde{\beta}^{OLS} \to_p -E(W_i W_i')^{-1} E(W_i (W_i - W_i^*)') \beta^{OLS}, \tag{2.1}$$

where $\beta^{OLS} = plim_{n \to \infty} \tilde{\beta}^{OLS}$ denotes the regression estimand. Because η_i is mean zero conditional on X_i and θ_i, the factor $E(W_i(W_i - W_i^*)')$ can be written as $Var(\eta_i) e_{K+2} e_{K+2}'$ where $e_{K+2} = (0, \ldots, 0, 1)'$. Further, under condition **LIx** of Assumption 2.2, $Var(\eta_i) = O(J^{-1})$.

the semiparametric model proposed in this paper.

In Williams (2018b), I derive the asymptotic distribution of $\hat{\beta}^{OLS}$ under a double asymptotic sequence with $n, J \to \infty$. I show that if $\sqrt{n}/J \to \gamma < \infty$ then $\sqrt{n}(\hat{\beta}^{OLS} - \beta^{OLS}) \to N(\gamma\bar{B}, \bar{V})$. I then propose a bias-corrected estimator for β^{OLS} based on a nonparametric estimate of the asymptotic bias, $\bar{B}$.

However, suppose that

$$Y_i = \beta_1' X_i + h(\theta_i) + e_i \tag{2.2}$$

for some function h. The components of β^{OLS} corresponding to coefficients on the vector X_i coincide with β_1 in this model only if $h(\theta_i)$ is a linear transformation of $\bar{p}(\theta_i)$.[7] However, linearity in $\bar{p}(\theta_i)$ is no more (or less) plausible an assumption than linearity in any other monotonic transformation of θ_i given that ability is not interval-scaled. Moreover, using different monotonic transformations of θ_i as a covariate in a regression can produce substantially different estimates of the coefficient on X_i, as these provide different approximations of the conditional expectation function. This is demonstrated in the empirical application in Section 4.

Instead, β_1 can be estimated using semiparametric methods for the partially linear regression model. Indeed the model of equation (2.2) is scale-invariant in the sense that β_1 could be estimated in this model using θ_i or $\bar{p}(\theta_i)$ or any other monotonic transformation of θ_i. However, the validity of an estimator of β_1 that uses $\bar{M}_i$ in place of θ_i does not follow immediately because $\bar{M}_i$ is not a monotonic transformation of θ_i. However, because $\max_{1 \leq i \leq n} |\bar{M}_i - \bar{p}(\theta_i)|$ converges to 0 as $J \to \infty$, consistency of such an estimator as $n, J \to \infty$ should follow under sufficient smoothness conditions. In Section 3, I describe a double-residual method (Robinson, 1988) to estimate the semiparametric regression model using $\bar{M}_i$. I provide sufficient regularity conditions for consistency in Theorem 3.1. Asymptotic normality at the root-n rate of semiparametric estimators generally requires the first stage nonparametric estimator to converge sufficiently quickly (Newey, 1994). In this case, this requires a restriction on the rate at which J grows with n. The asymptotic distribution is unbiased if $\sqrt{n}/J \to 0$. The size of this measurement error bias is perhaps surprising given that $\max_{1 \leq i \leq n} |\bar{M}_i - \bar{p}(\theta_i)|$ converges to 0, not at the rate $1/J$, but at the slower rate of $\sqrt{\log(J)/J}$. See Theorem 3.2.

Beginning in Section 3 below, I modify the notation to index by J to make explicit the dependence on J in the $n, J \to \infty$ asymptotic analysis. So $\bar{M}_i$ becomes $\bar{M}_{iJ}$, $\bar{p}(\theta_i)$ becomes $\bar{p}_J(\theta_i)$, and so on.

Discussion of Assumptions 2.2 As discussed in Sijtsma and Junker (2006), condition **LI** of Assumption 2.1, and similarly condition **LIx** of Assumption 2.2, is stronger than necessary for many purposes. Indeed this is one of the advantages of nonparametric IRT models over parametric IRT models given that the latter are typically estimated using a likelihood derived under this assumption.

[7] Or, alternatively if $E(X_i \mid \theta_i)$ is a linear transformation of $\bar{p}(\theta_i)$.

As discussed in Williams (2018b), $\hat{\beta}^{OLS}$ is a consistent estimator of β^{OLS} as $n, J \to \infty$ as long as $Var(\eta_i) \to 0$. The semiparametric estimator of Section 3 requires stronger conditions than this but it is apparent that condition **LIx** could be weakened, though at the cost of a more complex bias correction formula and slower convergence rates. Williams (2018a), for example, uses a conditional mixing condition in place of full conditional independence in a similar context.

One case where these weaker versions of **LIx** might still be violated is if there is strong temporal dependence among items. Jannarone (1997)'s model of learning during the test, for example, implies such dependence. Another implication of **LIx** is that it limits the type of measurement error allowed for. It does not, for example, allow for error induced by physical test conditions or the physical or mental state of the test-taker.

Assumption **ER** is crucial to the results obtained in this paper. Versions of this restriction are also imposed in the models studied in Junker et al. (2012), Schofield et al. (2014), and Lockwood and McCaffrey (2014). Without this assumption the methods studied in this paper fail to disentangle the effect of X_i on Y_i from the effect X_i has on the test score. In some contexts, however, this assumption is clearly violated. For example, several studies have pointed out that the AFQT test scores in the NLSY may be affected by an individual's education level at the time of the test. This problem has been addressed by Hansen et al. (2004), among others. Williams (2018a) suggests a general solution to this problem in a nonseparable model when $p_j(X_i, \theta_i) = p_j(\theta_i)$ for a single item j but not for the rest. Also, see Remark 1 following Theorem 3.2 regarding another way to relax this assumption.

If θ_i is not scalar or $\bar{p}$ is not a monotone function then conditioning on $\bar{p}(\theta_i)$ will not fully control for ability, particularly if one dimension of ability plays a substantially different role in determining the outcome, Y_i, than it does in performance on the test. In fact, conditioning on the scalar $\bar{p}(\theta_i)$ when θ_i is multidimensional can even lead to a larger bias in an estimate of the effect of X_i than the regression of Y_i on X_i (Heckman and Navarro, 2004). When θ_i is multidimensional a better approach would be to use scores from multiple tests or to divide the items on a single test into multiple separate subscores.

Condition **ND** of Assumption 2.2 requires the test score to be uninformative about the outcome, Y_i, conditional on latent ability and X_i. This condition is violated if the outcome is directly affected by the test score which might be the case, for example, if the observed test score were used for differentiated instruction or to determine acceptance into college. This condition is required in Junker et al. (2012) and Lockwood and McCaffrey (2014) as well.

Lastly, note that Assumption 2.2 allows for the possibility that M_{ij} represents a misreported performance, under certain assumptions regarding the nature of the misreporting. Measures of child development sometimes consist of parent or teacher reports of behavior. These reports may not be accurate, whether intentionally or unintentionally. Let M^*_{ij} denote an accurate report of the skill or behavior. First, conditions **LIx** and **ND** could be violated if there is a latent propensity to misreport. However, if misreporting is independent across items and independent of the outcome, then **LIx** and **ND** will still hold. In that case, there would still be a concern about conditions **ER** and **M**'. The item response functions can be written as

$$Pr(M_{ij} = 1 \mid \theta_i, X_i) = \delta_{0j}(\theta_i, X_i) + (\delta_{1j}(\theta_i, X_i) - \delta_{0j}(\theta_i, X_i))\, Pr(M^*_{ij} = 1 \mid \theta_i, X_i),$$

where $\delta_{sj}(\theta_i, X_i) = Pr(M_{ij} = 1 \mid M^*_{ij} = s, \theta_i, X_i)$ for $s = 0, 1$. If, for example, $\delta_{sj}(\theta_i, X_i) = \delta_{sj}$ and the true item response functions, $Pr(M^*_{ij} = 1 \mid \theta_i, X_i)$, satisfy conditions **ER** and **M**' then Assumption 2.2 would still hold in the presence of misreporting.[8]

3 Semiparametric estimation

Because θ_i is ordinal, not interval-scaled, a specification for equation (2.2) that restricts the functional form of h – e.g., by assuming that it is linear – is an arbitrary specification choice.[9] As demonstrated in Monte Carlo simulations and the empirical example below, this specification choice can affect estimates of β_1. In this section, I propose a solution to this misspecification problem and develop its asymptotic properties as $n, J \to \infty$. The method does not require a parametric specification of the item response model.

Suppose that equation (2.2) holds for some $\beta_1, h(\cdot)$, and e_i where $E(e_i \mid X_i, \theta_i) = 0$. If $\theta^*_i = \tau(\theta_i)$ for some strictly monotonic function τ, then there is a function $h^*(\cdot)$ such that $Y_i = \beta_1' X_i + h^*(\theta^*_i) + e_i$. Thus the parameter β_1 is invariant to monotonic transformations of θ_i in this model. The parameter β_1 can be estimated using θ_i or $\bar{p}_J(\theta_i)$ or any monotonic transformation of θ_i. Thus, if $\bar{p}_J(\theta_i)$ were observed directly, β_1 could be estimated consistently as $n \to \infty$ for a fixed J using any of the many well-known semiparametric estimators for the partially linear regression model.

The question remains whether $\bar{p}_J(\theta_i)$ can be replaced by $\bar{M}_{iJ}$. The analysis in Williams (2018a) shows that $E(Y_i \mid X_i = x, \bar{M}_{iJ} = m) \approx \beta_1' x + h(\bar{p}_J^{-1}(m))$ as $J \to \infty$.[10] This suggests that β_1 can be consistently estimated as $n, J \to \infty$ using semiparametric regression techniques applied to the estimating equation

$$Y_i = \beta_1' X_i + \tilde{g}(\bar{M}_{iJ}) + \tilde{u}_{iJ} \tag{3.1}$$

where $\tilde{g}(m) = h(\bar{p}_J^{-1}(m))$ and $\tilde{u}_{iJ} = e_i + h(\theta_i) - \tilde{g}(\bar{M}_{iJ})$.

One common estimator for the partially linear model is the double-residual method of Robinson

[8] This analysis regarding misreporting was suggested to me by an anonymous referee.

[9] See Braun and von Davier (2017) for a different perspective on this issue in the context of large-scale assessments.

[10] That is, $\lim_{J\to\infty} \left\{E(Y_i \mid X_i = x, \bar{M}_{iJ} = m) - \beta_1' x - h(\bar{p}_J^{-1}(m))\right\} = 0$.

(1988). Equation (2.2) implies that $E(Y_i \mid \theta_i) = \beta_1' E(X_i \mid \theta_i) + h(\theta_i)$. Therefore,

$$Y_i - E(Y_i \mid \theta_i) = \beta_1'(X_i - E(X_i \mid \theta_i)) + e_i$$

The double residual method is based on the idea that if the conditional expectations $E(Y_i \mid \theta_i)$ and $E(X_i \mid \theta_i)$ were known then β_1 could be estimated by a least squares regression of the residual $Y_i - E(Y_i \mid \theta_i)$ on the residual $X_i - E(X_i \mid \theta_i)$.

While the conditional expectation functions $h_y(\theta) := E(Y_i \mid \theta_i = \theta)$ and $h_x(\theta) := E(X_i \mid \theta_i = \theta)$ cannot be estimated, it is possible to consistently estimate $E(Y_i \mid \bar{p}_J(\theta_i) = m)$ and $E(X_i \mid \bar{p}_J(\theta_i) = m)$ as $n, J \to \infty$ by a nonparametric regression on $\bar{M}_{iJ}$. Thus, I propose the following estimator for β_1 in equation (2.2).

$$\hat{\beta}_{1J} = \left(\sum_{i=1}^{n} \hat{w}(\bar{M}_{iJ}) \hat{\tilde{X}}_i \hat{\tilde{X}}_i' \right)^{-1} \sum_{i=1}^{n} \hat{w}(\bar{M}_{iJ}) \hat{\tilde{X}}_i \hat{\tilde{Y}}_i \tag{3.2}$$

where $\hat{\tilde{X}}_i = X_i - \hat{g}_x(\bar{M}_{iJ})$, $\hat{\tilde{Y}}_i = Y_i - \hat{g}_y(\bar{M}_{iJ})$, and $\hat{g}_y(m)$ and $\hat{g}_x(m) = (\hat{g}_{x_1}(m), \ldots, \hat{g}_{x_K}(m))'$ are kernel regression estimators,

$$\hat{g}_y(m) = \frac{\sum_{i=1}^{n} h_y^{*-1} K_y \left(\frac{\bar{M}_{iJ} - m}{h_y^*} \right) Y_i}{\sum_{i=1}^{n} h_y^{*-1} K_y \left(\frac{\bar{M}_{iJ} - m}{h_y^*} \right)}$$

$$\hat{g}_{x_k}(m) = \frac{\sum_{i=1}^{n} h_{x_k}^{*-1} K_{x_k} \left(\frac{\bar{M}_{iJ} - m}{h_{x_k}^*} \right) X_{ik}}{\sum_{i=1}^{n} h_{x_k}^{*-1} K_{x_k} \left(\frac{\bar{M}_{iJ} - m}{h_{x_k}^*} \right)}$$

with bandwidth parameters h_y^* and $h_{x_1}^*, \ldots, h_{x_K}^*$, and $\hat{w}(m)$ is a weighting function. This weighting function must be smooth enough and must be 0 outside a region where $\hat{g}_y$ and $\hat{g}_x$ are uniformly well-behaved. The weighting solves two problems. First, the estimators $\hat{g}_y$ and $\hat{g}_x$ perform poorly where the distribution of $\bar{M}_{iJ}$ becomes thin. Second, the derivatives of $\tilde{g}_y(m) := h_y(\bar{p}_J^{-1}(m))$ and $\tilde{g}_x(m) := h_x(\bar{p}_J^{-1}(m))$ may be unbounded and this would lead to bias even if $\tilde{\mathbf{g}} = (\tilde{g}_y, \tilde{g}_x')'$ were known. The function $\hat{w}(m) = \psi \left(\frac{m - (\hat{q}_\delta(\bar{M}_{iJ}) + \hat{q}_{1-\delta}(\bar{M}_{iJ}))/2}{\frac{1}{2}(\hat{q}_{1-\delta}(\bar{M}_{iJ}) - \hat{q}_\delta(\bar{M}_{iJ}))} \right)$ where $\hat{q}_\tau(\bar{M}_{iJ})$ is the empirical quantile function and

$$\psi(u) = \begin{cases} \exp\left(\frac{1}{u^2 - 1}\right) & |u| \leq 1 \\ 0 & |u| > 1 \end{cases} \tag{3.3}$$

satisfies the required conditions stated in the appendix.

Robinson (1988) and Andrews (1994a), among others, have derived properties of the double-residual estimator that could be applied if $\bar{p}_J(\theta_i)$ were observed without error. However, these results do not apply here for several reasons. First, the analysis requires a double asymptotic sequence where $n, J \to \infty$ and previous results do not immediately apply in this case. Further, because $\hat{g}_y(\bar{M}_{iJ}) - h_y(\theta_i)$ is equal to the sum $\big(\hat{g}_y(\bar{M}_{iJ}) - \tilde{g}_y(\bar{M}_{iJ})\big) + \big(\tilde{g}_y(\bar{M}_{iJ}) - h_y(\theta_i)\big)$, the proof requires results on $\sup_m |\hat{g}_y(m) - \tilde{g}_y(m)|$ *and* an argument bounding $\tilde{g}_y(\bar{M}_{iJ}) - h_y(\theta_i)$. Results in the literature on $\sup_m |\hat{g}_y(m) - \tilde{g}_y(m)|$ require the regressor, in this case $\bar{M}_{iJ}$, to be continuous or discrete with a fixed support. And there are no results in the literature pertaining to $\tilde{g}_y(\bar{M}_{iJ}) - h_y(\theta_i)$. An additional difficulty arises because it is not possible to restrict the support of θ_i. We are only able to do this indirectly by restricting the support of $\bar{M}_{iJ}$.

Under Assumption 2.2 and additional regularity conditions stated in Appendix A, $\hat{\beta}_{1J}$ is consistent as $n, J \to \infty$, as stated in the following theorem.

Theorem 3.1. *If equation (2.2) holds with $E(e_i \mid X_i, \theta_i) = 0$, Assumptions 2.2, A.1, and A.2 are satisfied, and $(Y_i, X_i, \theta_i, \mathbf{M}_i)$, $i = 1, \ldots, n$ is an i.i.d. sequence of random variables then $\hat{\beta}_{1J} \to_p \beta_1$ as $n, J \to \infty$.*

Remark 1: *This result fails if X_i is not independent of $\mathbf{M}_i$ conditional on θ_i (which implies that condition **ER** of Assumption 2.2 does not hold) because in this case the partially linear model of equation (3.1) is misspecified. Suppose, however, that $X_i = (X_{1i}', X_{2i}')'$ and, correspondingly, β_1 is split up into β_{11} and β_{12}. Further, suppose that $Pr(M_{ij} = 1 \mid X_i, \theta_i) = Pr(M_{ij} = 1 \mid X_{2i}, \theta_i)$. This approach could be important, for example, if performance on a test differs between a focal group for which the test was designed and a reference group that is under study.*[11]

In this case, estimation of β_{11} can be based off of the partially linear regression model $Y_i = \beta_{11}' X_{1i} + \tilde{g}(X_{2i}, \bar{M}_{iJ}) + \tilde{u}_{iJ}$. If X_{2i} is discrete then the asymptotic analysis is nearly identical to that in the proof of Theorem 3.1 (and Theorem 3.2 below). If X_{2i} is continuous then the $\sqrt{n}$-asymptotic normality in particular only holds under more restrictive conditions and possibly requires the use of bias-reducing kernel functions, depending on the dimension of X_{2i} (c.f. Robinson, 1988).

Remark 2: *The conditions imposed by Assumptions A.1 and A.2 are fairly straightforward. One condition that is not standard requires the tails of the distribution of θ_i to be thin relative to the tails of the derivatives of the functions $\bar{p}_J$, depending also on how quickly the functions h_x and h_y increase in the tails. In the supplementary appendix I show when these conditions are satisfied in standard parametric IRT models.*

Consistency of $\hat{\beta}_{1J}$ relies primarily on the uniform convergence of $\bar{M}_{iJ}$ to $\bar{p}_J(\theta_i)$ as $J \to \infty$ and the uniform convergence of the kernel regression estimators $\hat{g}_y(m)$ and $\hat{g}_x(m)$ to the functions $\tilde{g}_y(m)$ and $\tilde{g}_x(m)$, uniformly over $m = \bar{p}_J(\theta)$ as θ varies within a compact subset of its support. The latter is sufficient because the trimming parameter δ does not have to be taken to 0 in order for $\hat{\beta}_1$ to be consistent. The uniform convergence holds under Assumption A.1. Uniform convergence plus

[11] I thank an anonymous referee for providing this example.

Assumption 2.2 and the additional regularity conditions in Assumption A.2 are then used to show consistency of $\hat{\beta}_{1J}$. Assumptions A.1 and A.2 do not impose any additional restrictions on the rate of growth of J beyond the assumption that J grows as n^r for some $r > 0$.

It is possible to derive conditions that are sufficient to ensure $\sqrt{n}$-asymptotic normality of $\hat{\beta}_{1J}$ as well. There is an extensive literature that provides sufficient conditions for $\sqrt{n}$-asymptotic normality for general classes of semiparametric estimators (Andrews, 1994a; Newey, 1994; Chen et al., 2003, among other). Mammen et al. (2016) derive such a result for a class of models where the covariates used in the first stage nonparametric estimation are generated. However, none of these results apply here because they do not allow for a double asymptotic sequence where $n, J \to \infty$. While $\bar{M}_{iJ}$ can be viewed as a generated covariate – an estimate of the covariate $\bar{p}_J(\theta_i)$ – this does not satisfy the conditions of Mammen et al. (2016), who consider a setup where a "true" covariate $r(Z_i)$ is replaced by a generated covariate $\hat{r}(Z_i)$ for some finite-dimensional Z_i and a consistent estimate $\hat{r}$ of the function r. The following theorem is thus a novel contribution to the literature on the asymptotic normality of semiparametric estimators.

Theorem 3.2. *If equation (2.2) holds with $E(e_i \mid X_i, \theta_i) = 0$, Assumptions 2.2, A.3, and A.4 are satisfied, and $(Y_i, X_i, \theta_i, \mathbf{M}_i)$, $i = 1, \ldots, n$ is an i.i.d. sequence of random variables then*

$$\sqrt{n} V_{1J}^{-1/2} Q_{0,J}^{*} (\hat{\beta}_{1J} - \beta_1 - B_J) \to_d N(0, I)$$

and $B_J = O(J^{-1})$ where $B_J = Q_{0,J}^{*-1} B_{1J}$,

$$\begin{aligned} B_{1J} &= E\left(w(\bar{p}_J(\theta_i)) \eta_i^2 \frac{\partial}{\partial \theta} h(\theta_i) \frac{\partial}{\partial \theta} h_x(\theta_i) \right) \\ V_{1J} &= E(w(\bar{p}_J(\theta_i))^2 e_i^2 (X_i - h_x(\theta_i))(X_i - h_x(\theta_i))') \\ Q_{0,J}^{*} &= E\left(w(\bar{p}_J(\theta_i))(X_i - h_x(\theta_i))(X_i - h_x(\theta_i))' \right), \end{aligned}$$

and $w(m) := plim_{n,J \to \infty} \hat{w}(m)$ for each m. In addition, if $V_{1J} \to \bar{V}_1$ and $Q_{0,J}^{*} \to \bar{Q}_0^{*}$ then $\sqrt{n}(\hat{\beta}_{1J} - \beta_1 - B_J) \to_d N(0, \bar{V})$ where $\bar{V} = \bar{Q}_0^{*-1} \bar{V}_1 \bar{Q}_0^{*-1}$, and if $\sqrt{n} B_J \to \gamma \bar{B}$ then $\sqrt{n}(\hat{\beta}_{1J} - \beta_1) \to_d N(\gamma \bar{B}, \bar{V})$.

Remark 3: *Consistency requires that $\hat{\mathbf{g}}(\bar{M}_{iJ})$ converges uniformly to $\tilde{\mathbf{g}}(\bar{p}_J(\theta_i))$. Asymptotic normality at the $\sqrt{n}$ rate, on the other hand, requires $\sqrt{n}||\hat{\mathbf{g}}(\bar{M}_{iJ}) - \tilde{\mathbf{g}}(\bar{p}_J(\theta_i))||^2 = o_p(1)$. If $\sqrt{n}/J \to \gamma > 0$ then this result does not hold, resulting in an asymptotic bias. After subtracting off the bias term B_J, the remainder is $o_p(1)$ provided that $\hat{\mathbf{g}}$ converges uniformly to $\tilde{\mathbf{g}}$ at a rate that is faster than $n^{-1/4}$ and also faster than $\sqrt{\log(J)/J}$. Assumption A.3 states sufficient conditions for this uniform convergence to hold.*

Remark 4: *The conditions of Assumption A.3 differ from those of Assumption A.1 in two ways. First,* *the various smoothness conditions and moment conditions are stronger. Second, restrictions on the rate of convergence of the bandwidth parameters and on the rate of growth of J are more stringent. In particular, it is assumed that J grows as n^r for some $r > \frac{1}{3}$. In other words, the result holds provided that $\sqrt{n}/J^{1+\alpha} \to \gamma < \infty$ for some $\alpha < 1/2$. Similarly the conditions of Assumption A.4 are stronger than those of Assumption A.2, in part to ensure that a sufficient stochastic equicontinuity property is satisfied and in part to ensure that the conditions of a central limit theorem hold.*

Remark 5: *It follows from the first conclusion of the theorem that $\hat{\beta}_{1J} - \beta_1 - B_J = O_p(n^{-1/2})$ because $||V_{1J}^{-1/2} Q_{0,J}^*||$ is bounded from above and bounded away from 0 by assumption.*

Remark 6: *Both $Q_{0,J}^*$ and V_{1J} can be estimated by their sample analogs, replacing $h_x(\theta_i)$ with $\hat{g}_x(\bar{M}_i)$, $\tau_{0J}(\theta_i)$ with $\hat{w}(\bar{M}_{iJ})$, and e_i with $\hat{e}_i = Y_i - \hat{\beta}_{1J}' X_i - \hat{g}(\bar{M}_{iJ})$. It is a straightforward extension of the results proved in this paper that the resulting asymptotic variance estimator is consistent as $n, J \to \infty$. This consistent estimator can then be used in practice because $||V_{1J}^{-1/2} Q_{0,J}^*||$ is bounded.*

Remark 7: *If $\sqrt{n}/J \to \gamma > 0$ then valid inference would require an estimate of the bias term, B_J. In Williams (2018b) I develop an estimator for $E(\eta_i^2)$. Given estimates of the derivatives, $\frac{\partial}{\partial \theta} h(\theta_i)$ and $\frac{\partial}{\partial \theta} h_x(\theta_i)$, this estimator could be adapted to estimate B_J.*

Implementing the estimator $\hat{\beta}_{1J}$ requires specifying the kernel functions, bandwidths, and trimming parameter. The conditions on the kernel functions stated in the appendix require them to each be compactly supported and smooth. One choice that satisfies the conditions is the function ψ defined above in (3.3). This kernel was suggested by Andrews (1994a) and is found to work well in the simulations and empirical application in this paper.

The bandwidth parameters can be chosen using leave-one-out cross validation to choose each bandwidth, as this method is known to perform well for kernel regression estimators (Hardle and Marron, 1985) and works well in the simulations below. It can be shown that the bandwidth that minimizes the MSE of the kernel estimate satisfies the required rate conditions for asymptotical normality of $\hat{\beta}_{1J}$. In the empirical application below results are not sensitive to moderate variation in the chosen bandwidths.

The theoretical results hold for any choice of δ. I find in the simulations that varying this trimming parameter between 0.05 and 0.2 has little effect on the results. If equation (2.2) is misspecified because of the additive separability between X_i and θ_i then the choice of δ could affect results. In the empirical application, for example, this is a concern if the effect of education varies substantially across the distribution of ability. As more observations are trimmed the estimated coefficient becomes closer to the effect of education for those at the median.

3.1 Monte Carlo

The proposed method is consistent as $n, J \to \infty$. To study the finite sample properties of this proposed estimation strategy, in this section I conduct a Monte Carlo study. I first simulate the following model

$$Y_i = X_i + \theta_i + e_i$$
$$M_{ij} = \mathbf{1}(\delta_j(\theta_i^* - \alpha_j) \geq \eta_{ij})$$

where $\theta_i^* = \tau(\theta_i)$ for $\tau(\theta) = \theta$ (model 1), $\tau(\theta) = sign(\theta)\theta^2$ (model 2) and $\tau(\theta) = sign(\theta)|\theta|^{1/2}$ (model 3). In addition, I simulate $X_i = \mathbf{1}(a\theta_i \geq e_i^x)$ where $\theta_i, e_i, e_i^x, \eta_{i1}, \ldots, \eta_{iJ}$ are drawn independently from either the standard normal distribution (for the first three variables) or the logistic distribution (for the remaining J variables).

In each case the 2PL model is correctly specified but in models 2 and 3 the latent ability enters the outcome equation and the 2PL item response model via different scales. I simulate all three models for $J = 50, 100$, and 500 and $n = 1000$ and 2000. The "difficulty" parameters, α_j, are equally spaced between -1 and 1, and the "discrimination" parameters, δ_j, are equally spaced between 5 and 10. Simulations are run for three different values of the parameter a, which determines the strength of the dependence between X_i and θ_i. The bias and standard deviation of the estimates of $\beta_1 = 1$ from each simulation are reported in Table 1.

I report results of three different estimators for these three models. The first two estimators are infeasible OLS estimators. The estimator labeled "OLSi" is obtained by a regression of Y_i on X_i and $\bar{p}_J(\theta_i^*)$. The estimator labeled "IRTi" is obtained by a regression of Y_i on X_i and θ_i^*. I use two infeasible estimators to illustrate the bias due to misspecification of the scale of the latent variable. The third estimator, labeled "PLR", is the semiparametric estimator introduced in this paper – the double residual regression estimator of the partially linear model based on $\bar{M}_{iJ}$.

The OLS estimator is inconsistent in all three models, and the IRT estimator is inconsistent in models 2 and 3. The IRT estimator is consistent as $n \to \infty$ in model 1, for any J. The PLR estimator is consistent in all three models as $n, J \to \infty$. The finite sample results in Table 1 demonstrate that the proposed method eliminates the misspecification bias fairly well when $J \geq 50$.

First we see that in model 1 the IRT method is nearly unbiased for large enough J but OLS is biased even for $J = 500$. In model 2 we find the opposite – the IRT method exhibits a substantial bias but the bias of OLS is negligible as J increases. The OLS method performs well apparently because $\bar{p}_J(\theta_i^2)$ is approximately linear in θ_i. In model 3 both OLS and IRT exhibit a bias that does not diminish as J increases.

In contrast, the PLR method exhibits a bias that decreases with J in each of the three models. It is apparent, however, that how large J needs to be in order for the bias to be negligible depends on the model and the amount of dependence between X_i and θ_i (determined by the parameter a). Overall these results demonstrate the superior performance of a method based on the partially linear regression model when J is sufficiently large.

Table B.1 in the supplementary material reports results for two different feasible estimators for the same simulations. In this table, OLS is the OLS estimator from a regression of Y_i on X_i and $\bar{M}_{iJ}$, and IRT is the OLS estimator from a regression of Y_i on X_i and $\hat{\theta}_i$, where $\hat{\theta}_1, \ldots, \hat{\theta}_n$ are maximum likelihood estimates from the 2PL model. It is clear that these estimators are affected by J, in contrast with the feasible estimators in Table 1. However, general patterns are hard to infer as the small J bias can be in the opposite direction of the misspecification bias so that in some cases the overall bias becomes worse as J increases.

4 Returns to schooling

In this section I conduct an empirical exercise to demonstrate the use of the methods proposed in this paper. I employ a sample of data from the National Longitudinal Study of Youth 1979 to investigate the effect of education on earnings.

It is well known in the extensive literature on the returns to education that failing to account for individual ability in a wage regression causes a positive ability bias in estimates of the return to education. Becker (1967), for example, showed how this bias would arise using a model of investment in human capital where the marginal benefit of education is increasing in ability.[12] Several different approaches to overcoming the endogeneity of schooling have been considered. One approach is to use earnings data on identical twins to control for genetic and environmental components of ability that are common between twins who make different education choices (Behrman et al., 1977; Ashenfelter and Krueger, 1994). Another approach uses instrumental variables. Card (1999) provides an extensive review of IV estimates of the effect of education on earnings.

A third approach aims to directly control for the unobserved ability that plagues OLS estimates. While some economists have used IQ scores and performance on achievement tests as proxies for unobserved ability it has long been recognized that doing so produces methods that suffer from measurement error bias that tends to bias the effect of education upwards. Thus factor models and structural equations models have been employed to account for the fact that these tests are noisy measures of ability. This approach is typified by studies by Griliches and Mason (1972); Chamberlain (1977); Blackburn and Neumark (1993). Some more recent work (Carneiro et al., 2003; Heckman et al., 2006b) incorporates measures of ability while also using exclusion restrictions to aid in identification. One common criticism of this approach is that these models require normal-

[12] Heckman et al. (2006a) point out that the problem of ability bias in this context had been recognized by economists as far back as Noyes (1945).

izations for identification which are sometimes perceived as arbitrary, in the sense that they are not motivated by an economic model. My analysis in this section avoids these normalizations using the methods developed in this paper.

4.1 Data

The data used in this section is from the National Longitudinal Survey of Youth (NLSY79). The NLSY79 includes information on demographics, educational outcomes, labor market outcomes, health, and criminal behavior for a panel of individuals over thirty years. The respondents were first interviewed in 1979 when they were between 14 and 22 years old. The respondents were reinterviewed each subsequent year until 1994 after which point they were interviewed on a biennial basis. As discussed below, my analysis is restricted to 30-year olds who reported working at least 35 hours per week on average. Table 2 reports summary statistics for this subsample.

As part of the survey $11,914$ respondents (94% of the respondents) were administered the Armed Services Vocational Aptitude Battery (ASVAB) which consists of ten subtests. Both raw scores ($\bar{M}_{iJ}$) and scale scores (IRT estimates, $\hat{\theta}_i$, from a 3PL model) for each subtest are reported in the data. The Armed Forces Qualifying Test (AFQT), a composite of the mathematical knowledge, arithmetic reasoning, paragraph comprehension, and word knowledge subtests, is also reported. These data are summarized in the second panel of Table 2. The number of items on each test ranges from 15 for paragraph comprehension to 35 for word knowledge. Recently the individual item responses for the AFQT subtests have been offered in a new data release.[13] However, in this paper I do not use this item level data.

4.2 Results

I estimate an earnings regression that controls for mathematical ability.[14] In the NLSY79 documentation (Ing et al., 2012) it is suggested, based on an analysis of the AFQT data from the NLSY79, that the mathematical knowledge and arithmetic reasoning items are all explained by a single "mathematical acuity" factor. Based on these findings they propose pooling the items from these two subtests of the AFQT. I start by estimating the following regression model

$$Y_i = \beta_0 + \beta_1 S_i + \beta_2' X_i + \beta_3 \bar{M}_i + u_i$$

[13] See Schofield (2014) and Williams (2018b) for analysis of measurement error in these scores based on this newly released data.

[14] Using the same sample, in Williams (2018b) I find that verbal ability does not have a statistical significant effect. Therefore, I exclude it from the analysis here.

where Y_i is the log of individual i's average weekly wages, S_i represents education level, X_i is a vector of additional controls, including year dummies, and $\bar{M}_i$ is the combined score ($\%$ answered correctly) on the mathematical knowledge and arithmetic reasoning items of the AFQT.

Because the return to education varies over the life cycle (see, e.g. Heckman et al., 2006a), I restrict the sample to individuals who are 30 years old. I estimate the earnings regression separately for four groups – white males, white females, non-white males, and non-white females. This was done partly due to a concern that the return to education varies with sex and race but also to address potential concerns with the validity of condition **ER** of Assumption 2.2, which states that $Pr(M_{ij} = 1 \mid \theta_i, X_i) = Pr(M_{ij} = 1 \mid \theta_i)$. If the data is pooled then the vector X_i should include sex and race dummies. In that case, we would be required to assume that sex and race do not influence test scores conditional on ability or, in other words, that the measurement error does not vary with sex or race. By separating the analysis based on sex and race I avoid this possibly problematic assumption. See also Remark 1 following Theorem 3.1.

Results are reported in Table 3. Column (1) in each panel reports the coefficient on education when the test score is omitted completely from the regression. The regression reported in column (2) uses $\bar{M}_{iJ}$ as a control and column (3) uses $\hat{\theta}_i$. Generally, the coefficient estimates in columns (2) and (3) are similar because $\hat{\theta}_i$ is roughly linear in $\bar{M}_{iJ}$ except in the tails of the distribution. Both suggest a lower return to education compared to column (1) once the ability bias is mitigated.

Next I explore the possibility of misspecification bias in these results. To demonstrate the problem I first searched across various transformations of the composite math score.[15] Figure 1 demonstrates the range of coefficient estimates obtained for the two specifications. The results suggest that the misspecification bias could potentially be fairly severe. For reference, Figure 2 illustrates the distribution of the the transformed scores under the transformations that lead to the smallest and largest coefficient estimates.

I then estimate the partially linear regression model of equation (3.1) as described in Section 3. The results are reported in column (4) of Table 3. The coefficient on highest grade completed suggests that each additional grade completed increases earnings by 5.2% for white males. This estimate is almost 10% higher than the estimate that controls for $\bar{M}_{iJ}$ linearly. On the other hand, the coefficient on the college dummy in the second specification for white males is overestimated (by over 10%) when $\bar{M}_{iJ}$ enters linearly compared to the partially linear model. The latter suggests that college graduation increases earnings by roughly 24% for white males, compared to 21% in the linear model. Similar results are found for women and non-whites. It is interesting to note that the direction of the misspecification bias is positive for some specifications for some subsamples but negative for others. The most dramatic result is that the coefficient on the college dummy for

[15] I considered transformations that were piecewise linear where the slope was allowed to change at a single "knot" but the function was restricted to be continuous at this knot and strictly increasing over the support of the test score. I further restricted the transformations to those which send 0 to 0 and 1 to 1.

non-white males is negative, though statistically insignificant. In comparison, when using $\bar{M}_{iJ}$ in a linear regression I find that college increases earnings by 10%, which is statistically significant at a 10% level.

5 Conditioning model

The method proposed in this paper does not require the specification of the distribution of $\theta_i \mid X_i$, often referred to as a conditioning model. Mislevy (1991) and Schofield et al. (2014), among others, emphasize the importance of the conditioning model. Schofield et al. (2014), e.g., discuss the role of the conditioning model used to produce institutional plausible values. Plausible values are random draws from $\theta_i \mid M_i, Z_i$, for some vector of variables Z_i, that are sometimes reported rather than, or in addition to, an estimate, $\hat{\theta}_i$, or a posterior mean, $E(\theta_i \mid M_i, Z_i)$. Schofield et al. (2014) argue that plausible values can be used when ability plays the role of a covariate only if the model used to produce the plausible values coincides with the econometrician's model. When it does not, they suggest using item level data to estimate a model of the form

$$
\begin{aligned}
Y_i \mid \theta_i, X_i &\sim N(\beta_0 + \beta_1' X_i + \beta_2 \theta_i, \sigma^2) \\
Pr(M_{ij} = 1 \mid X_i, \theta_i) &= \gamma_j + \frac{1 - \gamma_j}{1 + \exp(-\delta_j(\theta_i - \alpha_j))} \\
\theta_i \mid X_i &\sim N(\alpha' X_i, \tau^2)
\end{aligned}
$$

where $Y_i, M_{i1}, \ldots, M_{iJ}$ are assumed mutually independent conditional on X_i, θ_i so that the distribution of $Y_i, M_{i1}, \ldots, M_{iJ} \mid X_i$ is fully specified up to a finite-dimensional parameter (Schofield, 2015). This model can be estimated using maximum likelihood or MCMC methods.

Schofield et al. (2014) also argue that the bias due to the use of the wrong conditioning model vanishes as $J \to \infty$. What I have shown in this paper is that, due to similar logic, the conditioning model is not needed in order to estimate β_1 consistently as $J \to \infty$. An important advantage of the methods due to Mislevy (1991) and Schofield (2015) is their validity when J is small, if the conditioning model is correctly specified. As seen in the Monte Carlo exercises, when J is small the performance of the method proposed in this paper deteriorates as the dependence between θ_i and X_i grows. The advantages of this method, however, are that it does not rely on any functional form assumptions and it does not require item level data. The model of this paper is also more general in that it does not require that θ_i enters the outcome equation linearly.

While the work of Mislevy (1991) and Schofield et al. (2014) demonstrates the value of the conditioning model when J is small, other approaches to the measurement error problem do not require a conditioning model. See Lockwood and McCaffrey (2014) for a comparison of some of these approaches. The MOM method, for example, uses a parametric IRT model to implement a bias correction to improve estimates when J is small. Similarly, Williams (2018b) discusses the possibility of bias corrections without specifying a parametric IRT model.

6 Conclusion

The use of test scores and other item response assessments as controls for a latent ability or trait is common. The use of a percent correct test score introduces both misspecification bias and measurement error bias in regression analysis. In some cases the use of institutional plausible values or the modeling jointly can mitigate these issues. This paper proposes, as an alternative approach, the estimation of a partially linear model based on the percent correct test score using the double residual method of Robinson (1988) when the number of items is large.

This partially linear model is invariant to monotonic transformations of latent ability. I show that the proposed estimator is consistent as $n, J \to \infty$. I also provide conditions under which the estimator is $\sqrt{n}$-consistent and asymptotically normal. If $\sqrt{n}/J \to 0$ then there is no asymptotic bias and the estimator provides the basis for valid asymptotic inference. I show through Monte Carlo simulations that, in cases where misspecification is the source of substantial, the proposed method performs well provided that J is sufficiently large.

The method is easy to implement, relies on weak assumptions, and does not require item level data. The theoretical results, a novel contribution to the literature on the asymptotic normality of semiparametric estimators, are of independent interest. Finally, this paper, along with Williams (2018a) and Williams (2018b), develops a new framework, building on the nonparametric IRT framework in psychometrics (Douglas, 1997; Junker and Sijtsma, 2001), that can be applied in many different areas of economics where similar measurement problems arise.

References

AGOSTINELLI, F. AND M. WISWALL (2017): "Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development," WP #22441, National Bureau of Economic Research.

ALTONJI, J. G., P. BHARADWAJ, AND F. LANGE (2012): "Changes in the characteristics of American youth: Implications for adult outcomes," *Journal of Labor Economics*, 30, 783–828.

ANDREWS, D. W. (1992): "Generic uniform convergence," *Econometric theory*, 8, 241–257.

——— (1994a): "Asymptotics for semiparametric econometric models via stochastic equicontinuity," *Econometrica: Journal of the Econometric Society*, 43–72.

——— (1994b): "Empirical process methods in econometrics," *Handbook of econometrics*, 4, 2247–2294.

ASHENFELTER, O. C. AND A. B. KRUEGER (1994): "Estimates of the economic returns to schooling from a new sample of twins," *American Economic Review*, 84, 1157–73.

BALLOU, D. (2009): "Test scaling and value-added measurement," *Education*, 4, 351–383.

BECKER, G. (1967): "Human Capital and the Personal Distribution of Income: An Analytical Approach," Woytinsky lecture no. 1, University of Michigan.

BEHRMAN, J., P. TAUBMAN, AND T. WALES (1977): "Controlling for and measuring the effects of genetics and family environment in equations for schooling and labor market success," *Kinometrics: Determinants of socioeconomic success within and between families*, 35–96.

BEKKER, P. A. (1994): "Alternative approximations to the distributions of instrumental variable estimators," *Econometrica: Journal of the Econometric Society*, 657–681.

BIRNBAUM, A. (1968): "Some latent train models and their use in inferring an examinee's ability," *Statistical theories of mental test scores*, 395–479.

BLACKBURN, M. L. AND D. NEUMARK (1993): "Omitted-ability bias and the increase in the return to schooling," *Journal of Labor Economics*, 11, 521.

BLOOM, N. AND J. VAN REENEN (2007): "Measuring and Explaining Management Practices across Firms and Countries," *The Quarterly Journal of Economics*, 1351–1408.

BOLLINGER, C. R. (2003): "Measurement error in human capital and the black-white wage gap," *Review of Economics and Statistics*, 85, 578–585.

BOND, T. N. AND K. LANG (2013): “The evolution of the Black-White test score gap in Grades K–3: The fragility of results,” *Review of Economics and Statistics*, 95, 1468–1479.

BRAUN, H. AND M. VON DAVIER (2017): “The use of test scores from large-scale assessment surveys: psychometric and statistical considerations,” *Large-scale Assessments in Education*, 5, 17.

CARD, D. (1999): “The causal effect of education on earnings,” *Handbook of labor economics*, 3, 1801–1863.

CARNEIRO, P., K. T. HANSEN, AND J. J. HECKMAN (2003): “2001 Lawrence R. Klein Lecture: Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice,” *International Economic Review*, 44, 361–422.

CHAMBERLAIN, G. (1977): “Education, income, and ability revisited,” *Journal of Econometrics*, 5, 241 – 257.

CHAMBERLAIN, G. AND Z. GRILICHES (1975): “Unobservables with a variance-components structure: Ability, schooling, and the economic success of brothers,” *International Economic Review*, 16, 422–449.

CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2003): “Estimation of semiparametric models when the criterion function is not smooth,” *Econometrica*, 71, 1591–1608.

CHESHER, A. (1991): “The effect of measurement error,” *Biometrika*, 78, 451–462.

CHETTY, R., J. N. FRIEDMAN, AND J. E. ROCKOFF (2014): “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates,” *The American Economic Review*, 104, 2593–2632.

CUNHA, F., J. J. HECKMAN, AND S. M. SCHENNACH (2010): “Estimating the technology of cognitive and noncognitive skill formation,” *Econometrica*, 78, 883–931.

DOUGLAS, J. (1997): “Joint consistency of nonparametric item characteristic curve and ability estimation,” *Psychometrika*, 62, 7–28.

——— (2001): “Asymptotic identifiability of nonparametric item response models,” *Psychometrika*, 66, 531–540.

GOLDBERGER, A. S. (1972): “Structural equation methods in the social sciences,” *Econometrica*, 40, 979–1001.

GRILICHES, Z. AND W. M. MASON (1972): "Education, income, and ability," *Journal of Political Economy*, 80, S74–S103.

HABERMAN, S. J. (1977): "Maximum Likelihood Estimates in Exponential Response Models," *Annals of Statistics*, 5, 815–841.

HAHN, J. AND G. KUERSTEINER (2002): "Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large," *Econometrica*, 70, 1639–1657.

HANSEN, K. T., J. J. HECKMAN, AND K. J. MULLEN (2004): "The effect of schooling and ability on achievement test scores," *Journal of Econometrics*, 121, 39 – 98.

HARDLE, W. AND J. S. MARRON (1985): "Optimal bandwidth selection in nonparametric regression function estimation," *The Annals of Statistics*, 1465–1481.

HECKMAN, J. AND S. NAVARRO (2004): "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models," *Review of Economics and Statistics*, 86, 30–57.

HECKMAN, J., R. PINTO, AND P. SAVELYEV (2013): "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes," *The American Economic Review*, 103, 1–35.

HECKMAN, J. J., L. J. LOCHNER, AND P. E. TODD (2006a): "Earnings functions, rates of return and treatment effects: The Mincer equation and beyond," *Handbook of the Economics of Education*, 1, 307–458.

HECKMAN, J. J., J. STIXRUD, AND S. URZUA (2006b): "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior," *Journal of Labor Economics*, 24, 411–482.

HO, A. D. (2009): "A nonparametric framework for comparing trends and gaps across tests," *Journal of Educational and Behavioral Statistics*, 34, 201–228.

HU, Y. (2017): "The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics," *Journal of Econometrics*, 200, 154–168.

HU, Y. AND Y. SASAKI (2015): "Closed-form estimation of nonparametric models with non-classical measurement errors," *Journal of Econometrics*, 185, 392–408.

——— (2017): "Closed-Form Identification of Dynamic Discrete Choice Models with Proxies for Unobserved State Variables," *Econometric Theory*, 1–20.

HU, Y. AND S. M. SCHENNACH (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76, 195–216.

ING, P., C. A. LUMNEY, AND R. J. OLSEN (2012): "Reanalysis of the 1980 AFQT Data from the NLSY79," working paper, Center for Human Resource Research, Ohio State University.

JANNARONE, R. J. (1997): "Models for locally dependent responses: Conjunctive item response theory," in *Handbook of modern item response theory*, Springer, 465–479.

JUNKER, B. (2001): "On the interplay between nonparametric and parametric IRT, with some thoughts about the future," in *Essays on item response theory*, Springer, 247–276.

JUNKER, B., L. S. SCHOFIELD, AND L. J. TAYLOR (2012): "The use of cognitive ability measures as explanatory variables in regression analysis," *IZA Journal of Labor Economics*, 1, 1–19.

JUNKER, B. W. AND K. SIJTSMA (2001): "Nonparametric item response theory in action: An overview of the special issue," *Applied Psychological Measurement*, 25, 211–220.

KLEPPER, S. AND E. E. LEAMER (1984): "Consistent sets of estimates for regressions with errors in all variables," *Econometrica*, 52, 163–183.

LOCKWOOD, J. AND D. F. MCCAFFREY (2014): "Correcting for test score measurement error in ANCOVA models for estimating treatment effects," *Journal of Educational and Behavioral Statistics*, 39, 22–52.

LORD, F. (1952): "A theory of test scores." *Psychometric monographs*.

LORD, F. AND M. NOVICK (1968): *Statistical theories of mental test scores*, Addison-Wesley.

LORD, F. M. (1975): "The 'ability' scale in item characteristic curve theory," *Psychometrika*, 40, 205–217.

MAMMEN, E., C. ROTHE, AND M. SCHIENLE (2016): "Semiparametric estimation with generated covariates," *Econometric Theory*, 32, 1140–1177.

MISLEVY, R. (1991): "Randomization based inference about examinees in the estimation of item parameters," *Psychometrika*, 56, 177–193.

MISLEVY, R. J. AND R. D. BOCK (1990): *BILOG 3: Item analysis and test scoring with binary logistic models*, Scientific Software International.

MOKKEN, R. J. (1971): *A theory and procedure of scale analysis: With applications in political research*, vol. 1, Walter de Gruyter.

MURAKI, E. AND R. D. BOCK (1997): *PARSCALE: IRT item analysis and test scoring for rating-scale data*, Scientific Software International.

NCES (2009): "NAEP technical documentation," Tech. rep., U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

NEWEY, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

NIELSEN, E. (2015): "The Income-Achievement Gap and Adult Outcome Inequality," Finance and Economics Discussion Series 041, Board of Governors of the Federal Reserve.

NOYES, C. (1945): "Director's comment," in *Income from Independent Professional Practice*, ed. by M. Friedman and S. S. Kuznets, National Bureau of Economic Research, New York, 405–410.

PAKES, A. AND D. POLLARD (1989): "Simulation and the asymptotics of optimization estimators," *Econometrica: Journal of the Econometric Society*, 1027–1057.

RASCH, G. (1961): "On general laws and the meaning of measurement in psychology," in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, University of California Press Berkeley, CA, vol. 4, 321–333.

ROBINSON, P. M. (1988): "Root-N-consistent semiparametric regression," *Econometrica: Journal of the Econometric Society*, 931–954.

SCHOFIELD, L. S. (2014): "Measurement error in the AFQT in the NLSY79," *Economics letters*, 123, 262–265.

——— (2015): "Correcting for measurement error in latent variables used as predictors," *The annals of applied statistics*, 9, 2133.

SCHOFIELD, L. S., B. JUNKER, L. J. TAYLOR, AND D. A. BLACK (2014): "Predictive inference using latent variables with covariates," *Psychometrika*, 1–21.

SIJTSMA, K. AND B. W. JUNKER (2006): "Item response theory: Past performance, present developments, and future expectations," *Behaviormetrika*, 33, 75–102.

SPADY, R. H. (2007): "Semiparametric methods for the measurement of latent attitudes and the estimation of their behavioral consequences," Working Paper CWP26/07, CEMMAP.

STOUT, W. (1987): "A nonparametric approach for assessing latent trait unidimensionality," *Psychometrika*, 52, 589–617.

STOUT, W. F. (1990): "A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation," *Psychometrika*, 55, 293–325.

VAN DER LINDEN, W. J. AND R. K. HAMBLETON (1997): *Handbook of modern item response theory*, New York, NY: Springer.

VAN DER VAART, A. W. (2000): *Asymptotic statistics*, vol. 3, Cambridge university press.

WILLIAMS, B. (2018a): "Identification of a nonseparable model under endogeneity using binary proxies for unobserved heterogeneity," working paper, George Washington University.

——— (2018b): "A many item analysis of a linear regression with sum scores as regressors," working paper, George Washington University.

WOOLDRIDGE, J. M. (2009): "On estimating firm-level production functions using proxy variables to control for unobservables," *Economics Letters*, 104, 112–114.

ZWICK, R. (1992): "Chapter 7: Statistical and psychometric issues in the measurement of educational achievement trends: Examples from the National Assessment of Educational Progress," *Journal of Educational Statistics*, 17, 205–218.

A Appendix

This appendix states the additional assumptions used to prove consistency and asymptotic normality of the semiparametric estimator defined in equation (3.2). Proofs of Theorems 3.1 and 3.2 are provided in Appendix B in the supplementary material.

A.1 Consistency

Let $V_{iJ} = (Y_i, X_i', \bar{M}_{iJ})'$. For a function $\mathbf{g} = (g_y, g_x')'$, where $g_y : [0,1] \to \mathbb{R}$ and $g_x : [0,1] \to \mathbb{R}^K$ and a function $w : [0,1] \to \mathbb{R}$, let $m(V_{iJ}, \beta_1, \mathbf{g}, w) = w(\bar{M}_{iJ})(Y_i - g_y(\bar{M}_{iJ}) - \beta_1'(X_i - g_x(\bar{M}_{iJ})))(X_i - g_x(\bar{M}_{iJ}))$. The estimator defined in (3.2) can equivalently be defined as the value of the vector β_1 that solves the equation

$$\sum_{i=1}^{n} m(V_{iJ}, \beta_1, \hat{\mathbf{g}}, \hat{w}) = 0$$

where $\hat{\mathbf{g}} = (\hat{g}_y, \hat{g}_x')'$.

Next, let $V_i^* = (Y_i, X_i', \theta_i)$ and for a function $\mathbf{h} = (h_y, h_x')'$, where $h_y : \mathbb{R} \to \mathbb{R}$ and $h_x : \mathbb{R} \to \mathbb{R}^K$, and a function $\tau : \mathbb{R} \to \mathbb{R}^K$, let $m^*(V_i^*, \beta_1, \mathbf{h}, \tau) = \tau(\theta_i)(Y_i - h_y(\theta_i) - \beta_1'(X_i - h_x(\theta_i)))(X_i - h_x(\theta_i))$. Then define

$$\begin{aligned} M^*(\beta_1, \mathbf{h}, \tau) &= E(m^*(V_i^*, \beta_1, \mathbf{h}, \tau)) \\ M(\beta_1, \mathbf{g}, w) &= E(m(V_{iJ}, \beta_1, \mathbf{g}, w)) \\ \hat{M}_n(\beta_1, \mathbf{g}, w) &= n^{-1} \sum_{i=1}^{n} m(V_{iJ}, \beta_1, \mathbf{g}, w) \end{aligned}$$

Define $h_{y,0}(t) = E(Y_i \mid \theta_i = t)$ and $h_{x,0}(t) = E(X_i \mid \theta_i = t)$ and let β_{10} denote the true value of the parameter β_1. Let $\tilde{g}_y(m) = h_{y,0}(\bar{p}_J^{-1}(m))$ and $\tilde{g}_x(m) = h_{x,0}(\bar{p}_J^{-1}(m))$. Then let $\mathbf{h}_0 = (h_{y,0}, h_{x,0}')'$ and $\tilde{\mathbf{g}} = (\tilde{g}_y, \tilde{g}_x')$. Also define $\tau_{0,J}(t) = w_{0,J}(\bar{p}_J(t))$. Then let $\tilde{X}_i = X_i - h_{x,0}(\theta_i)$ and $Q_{0,J}^* = Q^*(\beta_{10}, \mathbf{h}_0, \tau_{0,J})$ where

$$Q^*(\beta_1, h, \tau) = E\left(\tau(\theta_i)\tilde{X}_i\tilde{X}_i'\right).$$

Last, for any $\pi \in [0,1]$, and any random variable Z with distribution function F_Z, let $q_\pi(Z) := inf\{q : F_Z(q) \geq \pi\}$ denote the π^{th} quantile. Then, for a fixed $0 < \delta < 1/2$, let Θ_δ denote the interval $[q_\delta(\theta_i), q_{1-\delta}(\theta_i)]$ and define $\mathcal{M}_\delta = \bar{p}_J(\Theta_\delta) = \{m \in [0,1] : m = \bar{p}_J(\theta) \text{ for some } \theta \in \Theta_\delta\}$. Though it is suppressed in the notation, $\mathcal{M}_\delta$ varies with J.

The following conditions imply uniform convergence of $\hat{g}_y$ and $\hat{g}_x$ by Theorem C.1.

Assumption A.1.

(a) $\mathbf{h}_0(t)$ *is continuous for all* $t \in \mathbb{R}$, $\mathbf{h}_0(t)$ *is differentiable at all* $t \in \mathbb{R}$ *with derivative* $D\mathbf{h}_0(t)$ *that is also continuous at all* $t \in \mathbb{R}$

(b) $\exists J_0$ *such that, for each* $J \geq J_0$, $\bar{p}_J(t)$ *is strictly increasing, continuous and differentiable at all* $t \in \mathbb{R}$ *with derivative* $D\bar{p}_J(t)$ *such that for each* $t \in \mathbb{R}$, *the family of functions* $\{D\bar{p}_J : J \geq J_0\}$ *is equicontinuous at* t. *Moreover, for each* $t \in \mathbb{R}$, $\inf_{J \geq J_0} D\bar{p}_J(t) > 0$.

(c) $\{J_n : n \geq 1\}$ *is a sequence such that* $J_n = O(n^r)$ *and* $J_n^{-1} = O(n^{-r})$ *for some* $r > 0$, $h_n \to 0$, $nh_n^3 \to \infty$, *and* $(J_n^{-1} \log(J_n))^{1/2} h_n^{-1} = o(1)$

(d) *The kernel functions* $K_y, K_{x_1}, \ldots, K_{x_K}$ *each satisfy condition (d) of Assumption C.2.*

(e) θ_i *has absolutely continuous distribution function* F_θ *and density* f_θ *such that* $0 < f_\theta(t) \leq \bar{f}_\theta$.

(f) $E|\tilde{X}_i|^3 < \infty$, $E|e_i|^3 < \infty$, *and for any* $\delta > 0$, $\sup_{\theta \in \Theta_\delta} E(|\tilde{X}_i|^3 \mid \theta_i = \theta) < \infty$ *and* $\sup_{\theta \in \Theta_\delta} E(|e_i|^3 \mid \theta_i = \theta) < \infty$.

Consistency of $\hat{\beta}_{1J}$ is proved using arguments similar to others in the literature on semiparametric estimators (Chen et al., 2003; Pakes and Pollard, 1989; Andrews, 1994a). In addition to uniform convergence of $\hat{g}_y$ and $\hat{g}_x$, the following additional conditions are used.

Assumption A.2.

(a) $|\hat{M}_n(\hat{\beta}_{1J}, \hat{g}, \hat{w})| = \inf_{\beta_1 \in \mathcal{B}} |\hat{M}_n(\beta_1, \hat{g}, \hat{w})|$ *where* $\mathcal{B} \subset \mathbb{R}^K$ *is compact and* $\beta_{10} \in \mathcal{B}$.

(b) $\exists c, J_0 > 0$ *such that* $\lambda_{min}(Q^*_{0,J}) \geq c$ *for all* $J \geq J_0$

(c) $E|Y_i| < \infty$, $E|X_i| < \infty$, $E|X_iY_i| < \infty$ *and* $E||X_iX_i'|| < \infty$

(d) *(i) The function* $\hat{w}(m)$ *is differentiable in* m *on* $[0, 1]$ *(with probability* 1*) with derivative denoted* $\hat{w}'(m)$ *and has support* $\hat{\mathcal{M}}$ *such that* $Pr(\hat{\mathcal{M}} \subset \mathcal{M}_{\delta_1}) \to 1$ *for some* $0 < \delta_1 < 1/2$. *(ii) There exists a function* $w_{0,J}(m)$ *such that* $\sup_{m \in [0,1]} |\hat{w}(m) - w_{0,J}(m)| \to 0$. *(iii)* $\exists J_0 > 0$ *such that for* $J \geq J_0$, $w_{0,J}$ *is differentiable in* m *on* $[0, 1]$ *with derivative denoted* $w'_{0,J}(m)$ *and* $w_{0,J}$ *has support in* $\mathcal{M}_{\delta_2}$ *for some* $\delta_2 \geq \delta_1$. *(iv) There exists a constant* $B < \infty$ *such that* $\sup_{m \in [0,1]} |\hat{w}(m)| < B$, $\sup_{m \in [0,1]} |\hat{w}'(m)| < B$, $\sup_{m \in [0,1]} |w_{0,J}(m)| < B$ *and* $\sup_{m \in [0,1]} |w'_{0,J}(m)| < B$.

(e) *There is a positive function* $\bar{D}(t)$ *and a constant* $J_0 > 0$ *such that* $|D\mathbf{h}_0(t)| \leq |D\bar{p}_J(t)|\bar{D}(t)$ *for all* $t \in \mathbb{R}$ *and all* $J \geq J_0$, $\bar{D}(t)$ *is nonincreasing for* $t \in (-\infty, q_\delta(\theta_i)]$ *and nondecreasing for* $t \in [q_{1-\delta}(\theta_i), \infty)$, *and* $E(\bar{D}(\theta_i)) < \infty$, $E(\bar{D}(\theta_i)|X_i|) < \infty$, *and* $E(\bar{D}(\theta_i)|Y_i|) < \infty$.

Conditions (a)-(c) are standard regularity conditions. The conditions on $\hat{w}$ in (d) are satisfied by the function in equation (3.3).

Condition (e) can be viewed as a restriction on the relative thickness of the tails of the distribution of θ. To see this, suppose that $|D\mathbf{h}_0(t)| \leq C_h(1+|t|^{q_h})$. Then we can consider when the restriction holds in common parametric models for $\bar{p}_J$. First, in the 3PL model $p_j(t) = \gamma_j + (1-\gamma_j)/(1+\exp(-\delta_j(t-\alpha_j)))$ it can be shown that $|D\bar{p}_J(t)| \geq C_p \exp(-\delta_{max} t)$ where $\delta_{max} = \max_j \delta_j$. Then let $\bar{D}(t) = C_h C_p^{-1}(1+|t|^{q_h})\exp(\delta_{max} t)$. Therefore, if $f_\theta(t) \leq C_\theta \exp(-\delta_\theta t)$ for $\delta_\theta > \delta_{max}$ then $E(\bar{D}(\theta)) = \int \bar{D}(t) f_\theta(t) dt \leq \int C_h C_p^{-1} C_\theta (1+|t|^{q_h}) \exp(-(\delta_\theta - \delta_{max})t) dt < \infty$. The other two bounds in (e) also hold if $E(|X_i| \mid \theta_i = t)$ and $E(|Y_i| \mid \theta_i = t)$ are both bounded by a power of t.

Thus if θ_i is normally distributed then the 3PL model satisfied condition (e). Suppose however that $p_j(t) = \gamma_j + (1-\gamma_j)\Phi(\delta_j(t-\alpha_j))$ where $\Phi(\cdot)$ is the standard normal cdf. This is another common parametric specification. Then if $\theta_i \sim N(\mu_\theta, \sigma_\theta^2)$ it can be shown that condition (e) is only satisfied if $\sigma_\theta^2 \cdot \max_j \beta_j^2 < 1$.

It is apparent that, more generally, condition (e) requires the tails of the distribution of θ_i to be relatively thin. If $D\bar{p}_j$ flattens out more quickly or $D\mathbf{h}_0$ increases more quickly in the tails then the tails of the distribution of θ_i have to be thinner.

A.2 Asymptotic normality

Theorem 3.2 requires additional restrictions on the model. The first set of conditions are sufficient for the uniform convergence of $\hat{g}_y$ and $\hat{g}_x$ at the rates given by Theorem C.2.

Assumption A.3.

(a) *$\mathbf{h}_0(t)$ is continuous for all $t \in \mathbb{R}$ and is twice differentiable at all $t \in \mathbb{R}$ with first and second derivatives $D\mathbf{h}_0(t)$ and $D^2\mathbf{h}_0(t)$ that are both continuous at all $t \in \mathbb{R}$*

(b) *$\exists J_0$ such that, for each $J \geq J_0$, $\bar{p}_J(t)$ is strictly increasing, continuous and twice differentiable at all $t \in \mathbb{R}$ with first and second derivatives $D\bar{p}_J(t)$ and $D^2\bar{p}_J(t)$ such that for each $t \in \mathbb{R}$, the family of functions $\{D\bar{p}_J : J \geq J_0\}$ is equicontinuous at t and the family of functions $\{D^2\bar{p}_J : J \geq J_0\}$ is equicontinuous at t. Moreover, for each $t \in \mathbb{R}$, $\inf_{J \geq J_0} D\bar{p}_J(t) > 0$.*

(c) *θ_i has absolutely continuous distribution function F_θ and density f_θ such that $0 < f_\theta(t) \leq \bar{f}_\theta$. The density function f_θ is differentiable with derivative Df_θ that is continuous at all $t \in \mathbb{R}$.*

(d) *$E|Y_i|^3 < \infty$, $E|X_i|^3 < \infty$, $E|\tilde{X}_i|^3 < \infty$, $E|e_i|^3 < \infty$, and for any $\delta > 0$, $\sup_{\theta \in \Theta_\delta} E(|Y_i|^3 \mid \theta_i = \theta) < \infty$, $\sup_{\theta \in \Theta_\delta} E(|X_i|^3 \mid \theta_i = \theta) < \infty$, $\sup_{\theta \in \Theta_\delta} E(|\tilde{X}_i|^3 \mid \theta_i = \theta) < \infty$ and $\sup_{\theta \in \Theta_\delta} E(|e_i|^3 \mid \theta_i = \theta) < \infty$.*

(e) *$\{J_n : n \geq 1\}$ is a sequence such that $J_n = O(n^r)$ and $J_n^{-1} = O(n^{-r})$ for some $r > \frac{1}{3}$.*

(f) *For each $s \in \mathbb{N}$, $2 \leq s < p$, for some $p > \max\{2, \frac{1-r}{3r-1}\}$, the function $\omega_{sJ}(t) = J^{s/2}E(\eta_i^s \mid \theta_i = t)$ is differentiable with derivative $D\omega_{sJ}(t)$ such that for each $t \in \mathbb{R}$, the family of functions $\{\omega_{sJ} : J \geq J_0\}$ is equicontinuous at t and the family of functions $\{D\omega_{sJ} : J \geq J_0\}$ is equicontinuous at t.*

(g) *If $r \leq 1/2$, $nh_n^{\frac{4}{1-r}-\epsilon} \to 0$ for some $\epsilon > 0$ and $nh_n^{2\frac{p-1}{r(p+1)-1}+\epsilon} \to \infty$ for some $\epsilon' > 0$; if $r > 1/2$, $nh_n^{8-\epsilon} \to 0$ for some $\epsilon > 0$ and $nh_n^{\max\{3, 2\frac{p-1}{rp-1/2}\}+\epsilon} \to \infty$ for some $\epsilon' > 0$*

(h) *The kernel functions $K_y, K_{x_1}, \ldots, K_{x_K}$ each satisfy condition (g) of Assumption C.3.*

Let $Z_i = (Y_i, X_i')'$. Recall that $h_{x,0}(\theta_i) := E(X_i \mid \theta_i)$, $h_{y,0}(\theta_i) := E(Y_i \mid \theta_i)$ and $\mathbf{h}_0(\theta_i) = (h_{y,0}(\theta_i), h_{x,0}(\theta_i)')' = E(Z_i \mid \theta_i)$ and $\tilde{X}_i = X_i - h_{x,0}(\theta_i)$. Also, define $\tilde{Z}_i = Z_i - \mathbf{h}_0(\theta_i)$ and $V_{1J} = E(\tau_{0J}(\theta_i)^2 e_i^2 \tilde{X}_i \tilde{X}_i')$.

Asymptotic normality of $\hat{\beta}_{1J}$ is implied by uniform convergence of $\hat{g}_y$ and $\hat{g}_x$ under the following additional conditions.

Assumption A.4.

(a) *$|\hat{M}_n(\hat{\beta}_{1J}, \hat{g}, \hat{w})| = \inf_{\beta_1 \in \mathcal{B}} |\hat{M}_n(\beta_1, \hat{g}, \hat{w})|$ where $\mathcal{B} \subset \mathbb{R}^K$ is compact and $\beta_{10} \in int(\mathcal{B})$.*

(b) *$\exists c, J_0 > 0$ such that $\lambda_{min}(Q_{0,J}^*) \geq c$, $\lambda_{min}(V_{1J}) \geq c$, and $\inf_{1 \leq j,k \leq K} \lambda_{min}(\Sigma_{jk,J}) \geq c$ for all $J \geq J_0$ where $\Sigma_{jk,J} = E(\tau_{0,J}(\theta_i)^2 \tilde{X}_{ij}^2 \tilde{X}_{ik}^2)$.*

(c) *$E|Z_i|^{2+\delta} < \infty$, $E||X_i Z_i'||^{2+\delta} < \infty$, and $E||\tilde{X}_i \tilde{Z}_i'||^{2+\delta} < \infty$ for some $\delta > 0$ and for some $J_0 \in \mathbb{N}$, $\sup_{J \geq J_0} ||V_{1J}|| < \infty$.*

(d) *(i) The function $\hat{w}(m)$ is differentiable in m on $[0,1]$ (with probability 1), with derivative denoted $\hat{w}'(m)$ that is continuous everywhere, and $\hat{w}(m)$ has support $\hat{\mathcal{M}}$ such that $Pr(\hat{\mathcal{M}} \subset \mathcal{M}_{\delta_1}) \to 1$ for some $0 < \delta_1 < 1/2$. (ii) There exists a function $w_{0,J}(m)$ such that $\sup_{m \in [0,1]} |\hat{w}(m) - w_{0,J}(m)| = o_p(n^{-1/6}) + O_p((J^{-1}\log(J))^{1/2}$. (iii) $\exists J_0 > 0$ such that for $J \geq J_0$, $w_{0,J}$ is differentiable in m on $[0,1]$ with derivative denoted $w_{0,J}'(m)$ that is continuous everywhere, and $w_{0,J}$ has support in $\mathcal{M}_{\delta_2}$ for some $\delta_2 \geq \delta_1$. (iv) There exists a constant $B < \infty$ such that $\sup_{m \in [0,1]} |\hat{w}(m)| < B$, $\sup_{m \in [0,1]} |\hat{w}'(m)| < B$, $\sup_{m \in [0,1]} |w_{0,J}(m)| < B$ and $\sup_{m \in [0,1]} |w_{0,J}'(m)| < B$.*

(e) *There is a positive function $\bar{D}(t)$ and a constant $J_0 > 0$ such that $|D\mathbf{h}_0(t)| \leq |D\bar{p}_J(t)|\bar{D}(t)$, $|D^2\mathbf{h}_0(t)| \leq |D\bar{p}_J(t)|\bar{D}(t)$, and $|D^2\bar{p}_J(t)| \leq |D\bar{p}_J(t)|\bar{D}(t)$ for all $t \in \mathbb{R}$ and all $J \geq J_0$, $\bar{D}(t)$ is nonincreasing for $t \in (-\infty, q_\delta(\theta_i)$ and nondecreasing for $t \in [q_{1-\delta}(\theta_i), \infty)$, and $E(\bar{D}(\theta_i)) < \infty$, $E(\bar{D}(\theta_i)^2) < \infty$, $E(\bar{D}(\theta_i)|X_i|) < \infty$, and $E(\bar{D}(\theta_i)|Y_i|) < \infty$.*

(f) *There exists a continuous function $\bar{p}_\infty(t)$ such that $\bar{p}_J(t) \to \bar{p}_\infty(t)$ for each $t \in \mathbb{R}$.*

Table 1. Monte Carlo results for the partially linear regression model

n	J	model 1, a=1						model 1, a=2						model 1, a=4					
		OLSi		IRTi		PLR		OLSi		IRTi		PLR		OLSi		IRTi		PLR	
		bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
1000	50	0.04	0.08	0.00	0.08	-0.01	0.09	-0.07	0.10	0.00	0.09	0.01	0.11	-0.24	0.11	0.00	0.10	0.03	0.13
	100	0.05	0.08	0.01	0.08	0.00	0.09	-0.07	0.10	0.00	0.09	0.00	0.11	-0.25	0.12	0.00	0.10	0.00	0.13
	500	0.04	0.08	0.00	0.08	-0.02	0.10	-0.06	0.10	0.01	0.09	-0.01	0.10	-0.24	0.12	0.00	0.10	-0.01	0.14
2000	50	0.04	0.06	0.00	0.06	0.00	0.07	-0.08	0.07	0.00	0.07	0.01	0.07	-0.25	0.08	0.00	0.07	0.03	0.09
	100	0.05	0.05	0.00	0.05	0.00	0.06	-0.07	0.07	0.00	0.07	0.00	0.08	-0.25	0.09	0.00	0.07	0.00	0.10
	500	0.04	0.06	0.00	0.06	-0.01	0.07	-0.07	0.07	0.00	0.06	-0.01	0.07	-0.25	0.09	0.00	0.07	-0.02	0.10

n	J	model 2, a=1						model 2, a=2						model 2, a=4					
		OLSi		IRTi		PLR		OLSi		IRTi		PLR		OLSi		IRTi		PLR	
		bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
1000	50	0.22	0.07	0.01	0.07	-0.02	0.09	0.15	0.08	-0.10	0.08	-0.02	0.09	-0.01	0.09	-0.29	0.08	-0.02	0.09
	100	0.22	0.07	0.01	0.07	-0.03	0.09	0.16	0.08	-0.10	0.08	-0.03	0.09	0.00	0.10	-0.29	0.09	-0.03	0.09
	500	0.23	0.07	0.01	0.07	-0.04	0.09	0.16	0.08	-0.11	0.08	-0.04	0.09	0.01	0.09	-0.28	0.09	-0.04	0.10
2000	50	0.22	0.05	0.01	0.05	-0.01	0.06	0.15	0.06	-0.10	0.06	-0.01	0.06	0.00	0.07	-0.28	0.07	-0.01	0.07
	100	0.23	0.05	0.01	0.05	-0.01	0.06	0.16	0.06	-0.10	0.06	-0.01	0.06	0.00	0.07	-0.28	0.06	-0.02	0.06
	500	0.23	0.05	0.01	0.05	-0.04	0.06	0.16	0.05	-0.10	0.06	-0.03	0.06	0.00	0.07	-0.29	0.07	-0.04	0.07

n	J	model 3, a=1						model 3, a=2						model 3, a=4					
		OLSi		IRTi		PLR		OLSi		IRTi		PLR		OLSi		IRTi		PLR	
		bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
1000	50	0.05	0.08	0.20	0.08	0.03	0.10	0.17	0.11	0.54	0.10	0.09	0.13	0.36	0.14	0.78	0.11	0.28	0.22
	100	0.05	0.08	0.20	0.08	0.02	0.09	0.19	0.11	0.55	0.10	0.06	0.13	0.39	0.13	0.79	0.10	0.22	0.22
	500	0.05	0.08	0.19	0.09	0.00	0.10	0.18	0.11	0.55	0.10	0.02	0.13	0.38	0.13	0.78	0.11	0.07	0.25
2000	50	0.06	0.06	0.20	0.06	0.04	0.07	0.19	0.09	0.55	0.08	0.10	0.10	0.38	0.09	0.78	0.07	0.29	0.14
	100	0.06	0.06	0.21	0.06	0.02	0.07	0.19	0.08	0.54	0.07	0.04	0.10	0.39	0.09	0.79	0.08	0.21	0.17
	500	0.05	0.06	0.20	0.06	0.00	0.07	0.19	0.08	0.55	0.07	0.02	0.09	0.39	0.09	0.78	0.08	0.07	0.17

Notes: This table reports results of the Monte Carlo exercise described in Section 3.3. All entries are expressed as a fraction of the true parameter value. This table reports results for the coefficient on the observed regressor. The IRT scores were obtained using the known values for the item response parameters rather than estimated values.

Table 2. Descriptive Statistics

	mean	std. dev.	min.	max
Highest grade completed	13.61	2.32	4	20
Avg. weekly wage	816.31	569.52	2.98	14700.00
Hours worked per wk.	45.16	8.76	35	126
Male	0.55	0.50	0	1
White	0.67	0.47	0	1
Urban residence	0.79	0.41	0	1
Married	0.57	0.49	0	1
Children in household	0.94	1.10	0	6
Urban residence, at age 14	0.77	0.42	0	1
Father's highest grade completed	11.50	3.84	0	20
Mother's highest grade completed	11.42	3.00	0	20
AFQT subtests	mean	std. dev.	items	
Math Knowledge	0.56	0.25	25	
Paragraph Comp.	0.74	0.20	30	
Word Knowledge	0.76	0.19	15	
Arith. Reasoning	0.61	0.24	35	

Notes: Statistics calculated on the sample of 2,983 30-year-olds who worked at least 35 hours per week on average. The parents' education is reported at the initial interview in 1979. The scores reported for the ASVAB subtests are the percent correct scores. Average weekly wages are adjusted to 2010 dollars.

Table 3.

		male				female			
		(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
white		0.067	0.048	0.047	0.052	0.099	0.063	0.063	0.055
		(0.009)	(0.011)	(0.011)	(0.014)	(0.011)	(0.012)	(0.012)	(0.013)
		0.318	0.238	0.236	0.209	0.292	0.134	0.134	0.116
		(0.043)	(0.049)	(0.049)	(0.053)	(0.046)	(0.049)	(0.049)	(0.054)
non-white		0.063	0.039	0.035	0.028	0.081	0.071	0.066	0.051
		(0.012)	(0.013)	(0.014)	(0.017)	(0.013)	(0.015)	(0.015)	(0.016)
		0.203	0.100	0.087	-0.026	0.280	0.230	0.223	0.188
		(0.058)	(0.058)	(0.058)	(0.062)	(0.052)	(0.054)	(0.052)	(0.054)

Notes: All regressions were estimated on the subsample of 30 year olds working at least 35 hours per week on average. Each regression includes the following controls: for urban residence, regional dummies, mother's and father's educational level, urban residence at age 14, and year dummies. Column (1) does not control for ability. Column (2) controls for ability with the percent correct score. Column (3) uses the published IRT score. Column (4) controls nonparametrically for ability as described in Section 3. Robust std. errors are reported in parentheses.

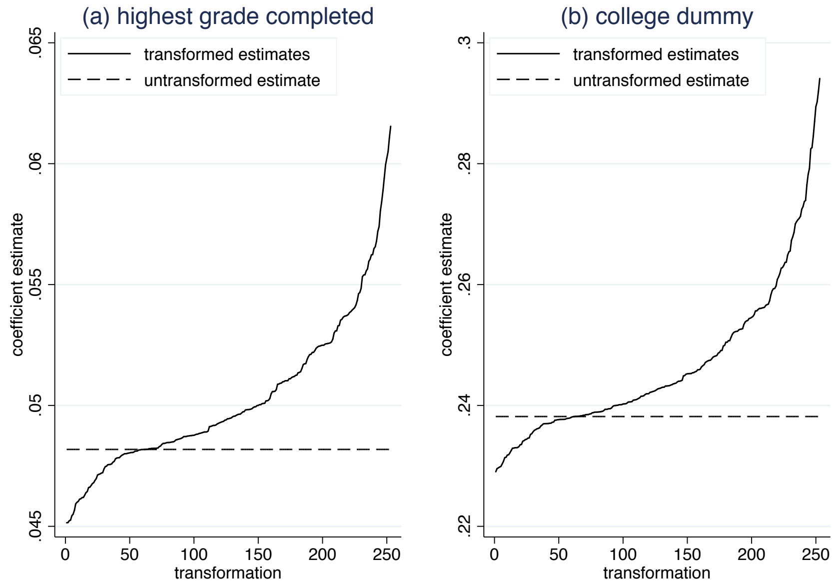

Figure 1: Estimates of education coefficients controlling for transformations of the test score.

Notes: This figure reports the coefficient on the education variable in a regression of log wages on education, a transformation of the test score, urban residence, regional dummies, mother's and father's education, urban residence at age 14, and year dummies, where the transformations are varied as described in the text.

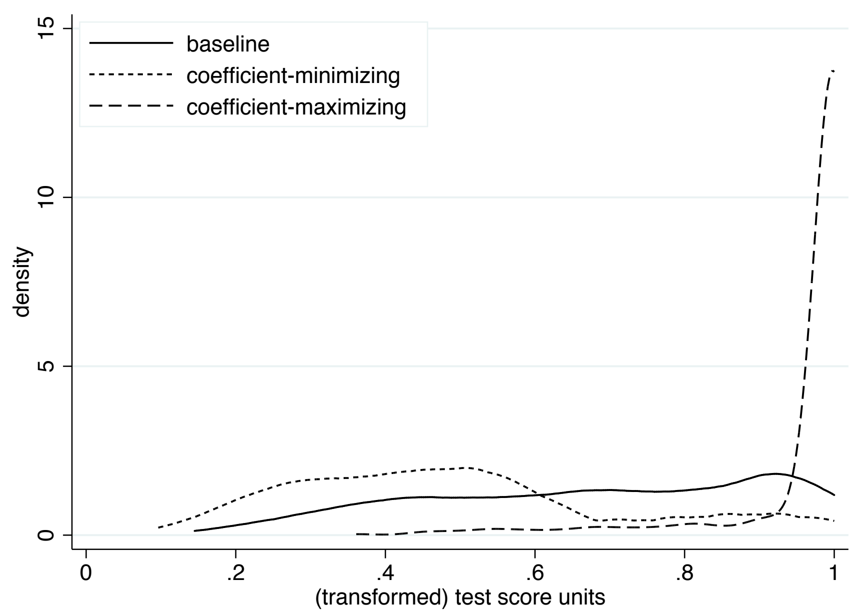

Figure 2: Distribution of transformed test scores

Notes: The solid line is a kernel density estimate of the density of $\bar{M}_i$. The dotted and dashed lines show kernel density estimates of the density of the transformations of $\bar{M}_i$ which maximize or minimize the coefficient on education. The coefficient-maximizing and minimizing transformations were the same for the two regression specifications.

B Proofs

Proof of Theorem 3.1. I consider a sequence $\{J_n : n \geq 1\}$ and show that $plim_{n\to\infty}\hat{\beta}_{1J_n} = \beta_{10}$. First, $m^*(V_i^*, \beta_1, \mathbf{h})$ is linear in β_1 so

$$\begin{aligned}
&M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n}) - M^*(\beta_1, \mathbf{h}_0, \tau_{0,J_n}) \\
&= E(\tau_{0,J_n}(\theta_i)(Y_i - h_{y,0}(\theta_i) - \hat{\beta}'_{1J_n}(X_i - h_{x,0}(\theta_i)))(X_i - h_{x,0}(\theta_i))) \\
&- E(\tau_{0,J_n}(\theta_i)(Y_i - h_{y,0}(\theta_i) - \beta'_1(X_i - h_{x,0}(\theta_i)))(X_i - h_{x,0}(\theta_i))) \\
&= -Q^*_{0,J_n}\left(\hat{\beta}_{1J_n} - \beta_1\right)
\end{aligned}$$

Since, by assumption, equation (2.2) holds for β_{10} with $E(e_i \mid X_i, \theta_i) = 0$, $Y_i - h_{y,0}(\theta_i) - \beta'_{10}(X_i - h_{x,0}(\theta_i)) = e_i$ and

$$\begin{aligned}
M^*(\beta_{10}, \mathbf{h}_0, \tau_{0,J}) &= E(\tau_{0,J}(\theta_i)e_i(X_i - h_{x,0}(\theta_i))) \\
&= E(\tau_{0,J}(\theta_i)E(e_i \mid X_i, \theta_i)(X_i - h_{x,0}(\theta_i))) = 0
\end{aligned}$$

Then, since $Q^*_{0,J}$ is invertible by Assumption A.2(b), $\hat{\beta}_{1J_n} - \beta_{10} = -Q^{*-1}_{0,J_n}M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n})$ and hence for any $\epsilon > 0$

$$Pr(|\hat{\beta}_{1J_n} - \beta_{10}| > \epsilon) \leq Pr(||Q^{*-1}_{0,J_n}|| \cdot |M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n})| \geq \epsilon) \leq Pr(|M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n})| \geq c\epsilon)$$

where the second inequality follows from the bound on $||Q^{*-1}_{0,J}||$ provided by Assumption A.2(b). It will thus be sufficient to show that $Pr(|M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n})| \geq \varepsilon) \to 0$ as $n \to \infty$ for all $\varepsilon > 0$.

By the triangle inequality, $|M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n})| \leq |\hat{M}_n(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w}) - M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n})| + |\hat{M}_n(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w})|$. I will first show that

$$|\hat{M}_n(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w}) - M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n})| = o_p(1) \tag{B.1}$$

using the following decomposition

$$\begin{aligned}
&|\hat{M}_n(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w}) - M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n})| \qquad \text{(B.2)} \\
&\leq |M(\hat{\beta}_{1J_n}, \tilde{\mathbf{g}}, w_{0,J_n}) - M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n})| + |M(\hat{\beta}_{1J_n}, \tilde{\mathbf{g}}, \hat{w}) - M(\hat{\beta}_{1J_n}, \tilde{\mathbf{g}}, w_{0,J_n})| \\
&+ |M(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w}) - M(\hat{\beta}_{1J_n}, \tilde{\mathbf{g}}, \hat{w})| + |\hat{M}_n(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w}) - M(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w})|
\end{aligned}$$

Define $\xi_i(\beta_1) = Y_i - \beta'_1 X_i$ and $\gamma(\beta_1) = (1, \beta'_1)'$ and let $\hat{\xi}_i = \xi_i(\hat{\beta}_{1J})$ and $\hat{\gamma} = \gamma(\hat{\beta}_{1J})$. Then

$M(\beta_1, \mathbf{g}, w) = E(w(\bar{M}_{iJ})(\xi_i(\beta_1) - \gamma(\beta_1)'\mathbf{g}(\bar{M}_{iJ}))(X_i - g_x(\bar{M}_{iJ})))$ so

$$\begin{aligned}|M(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w}) - M(\hat{\beta}_{1J_n}, \tilde{\mathbf{g}}, \hat{w})| &\leq |E(\hat{w}(\bar{M}_{iJ_n})\hat{\xi}_i(\hat{g}_x(\bar{M}_{iJ_n}) - \tilde{g}_x(\bar{M}_{iJ_n})))| \\ &\quad + |E(\hat{w}(\bar{M}_{iJ_n})\gamma'(\hat{\mathbf{g}}(\bar{M}_{iJ_n}) - \tilde{\mathbf{g}}(\bar{M}_{iJ_n}))X_i)| \\ &\quad + |E(\hat{w}(\bar{M}_{iJ_n})\hat{\gamma}'(\hat{\mathbf{g}}(\bar{M}_{iJ_n})\hat{g}_x(\bar{M}_{iJ_n}) - \tilde{\mathbf{g}}(\bar{M}_{iJ_n})\tilde{g}_x(\bar{M}_{iJ_n})))| \\ &\leq C\left(1 + \sup_{m\in\hat{\mathcal{M}}}|\hat{\mathbf{g}}(m)| + \sup_{m\in\hat{\mathcal{M}}}|\tilde{\mathbf{g}}(m)|\right)\sup_{m\in\hat{\mathcal{M}}}|\hat{\mathbf{g}}(m) - \tilde{\mathbf{g}}(m)|\end{aligned} \tag{B.3}$$

where the second inequality follows for some constant $C > 0$ by (a), (c), and (d) of Assumption A.2. Then, using Theorem C.1 and Assumption A.2(d), for any $\varepsilon > 0$,

$$\begin{aligned}&Pr(|M(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w}) - M(\hat{\beta}_{1J_n}, \tilde{\mathbf{g}}, \hat{w})| \geq \varepsilon) \\ &\leq Pr\left(C\left(1 + \sup_{m\in\mathcal{M}_{\delta_1}}|\hat{\mathbf{g}}(m)| + \sup_{m\in\mathcal{M}_{\delta_1}}|\tilde{\mathbf{g}}(m)|\right)\sup_{m\in\mathcal{M}_{\delta_1}}|\hat{\mathbf{g}}(m) - \tilde{\mathbf{g}}(m)| > \varepsilon\right) \\ &+ \left(1 - Pr(\hat{\mathcal{M}} \subset \mathcal{M}_{\delta_1})\right) \to 0\end{aligned}$$

Next,

$$\begin{aligned}&|M(\hat{\beta}_{1J_n}, \tilde{\mathbf{g}}, w_{0,J_n}) - M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n})| \\ &\quad\leq E\left(|w_{0,J_n}(\bar{M}_{iJ_n})|\left|(\hat{\xi}_i - \hat{\gamma}'\tilde{\mathbf{g}}(\bar{M}_{iJ_n}))(X_i - \tilde{g}_x(\bar{M}_{iJ_n})) - (\hat{\xi}_i - \hat{\gamma}'\tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i)))(X_i - \tilde{g}_x(\bar{p}_{J_n}(\theta_i)))\right|\right) \\ &\quad+ E(|w_{0,J_n}(\bar{M}_{iJ_n}) - w_{0,J_n}(\bar{p}_{J_n}(\theta_i))||(\hat{\xi}_i - \hat{\gamma}'\tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i)))(X_i - \tilde{g}_x(\bar{p}_{J_n}(\theta_i)))|)\end{aligned} \tag{B.4}$$

By Assumption A.2(a), the first term in equation (B.4) is bounded by a positive constant times

$$\begin{aligned}&E(|w_{0,J_n}(\bar{M}_{iJ_n})(|X_i| + |Y_i|)|\tilde{\mathbf{g}}(\bar{M}_{iJ_n}) - \tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))|) \\ &+ E(|w_{0,J_n}(\bar{M}_{iJ_n})|\tilde{\mathbf{g}}(\bar{M}_{iJ_n})||\tilde{\mathbf{g}}(\bar{M}_{iJ_n}) - \tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))|) \\ &+ E(|w_{0,J_n}(\bar{M}_{iJ_n})|\tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))||\tilde{\mathbf{g}}(\bar{M}_{iJ_n}) - \tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))|)\end{aligned}$$

It can be shown that each of these three terms can be bounded by an $o(1)$ sequence using essentially the same argument. First, by conditions (a) and (b) of Assumption A.1, $\tilde{\mathbf{g}}$ is continuously differentiable so $|\tilde{\mathbf{g}}(\bar{M}_{iJ_n}) - \tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))| \leq |D\tilde{\mathbf{g}}(p_i^*)||\eta_i|$ for some p_i^* between $\bar{p}_{J_n}(\theta_i)$ and $\bar{M}_{iJ_n}$. Next, by Assumption A.2(e), $|D\tilde{\mathbf{g}}(m)| \leq \bar{D}(\bar{p}_{J_n}^{-1}(m))$ where $\bar{D}(\cdot)$ is nonincreasing on the interval $(-\infty, q_{\delta_2}(\theta_i)]$ and nondecreasing on $[q_{1-\delta_2}(\theta_i), \infty)$. If $w_{0,J_n}(\bar{M}_{iJ_n}) > 0$ and $\theta_i \in \Theta_{\delta_2}$ then

$p_i^* \in \mathcal{M}_{\delta_2}$. If $w_{0,J_n}(\bar{M}_{iJ_n}) > 0$ and $\theta_i \notin \Theta_{\delta_2}$ then $\bar{D}(\bar{p}_{J_n}^{-1}(p_i^*)) \leq \bar{D}(\theta_i)$. Then

$$
\begin{aligned}
&E(|w_{0,J_n}(\bar{M}_{iJ_n})(|X_i| + |Y_i|)|\tilde{\mathbf{g}}(\bar{M}_{iJ_n}) - \tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))|) \\
&\leq \int_{-\infty}^{q_{\delta_2}(\theta_i)} B\bar{D}(t)E(|X_i| + |Y_i| \mid \theta_i = t)f_\theta(t)E(|\eta_i| \mid \theta_i = t)dt \\
&+ \int_{\Theta_{\delta_2}} B\left(\sup_{m\in\mathcal{M}_{\delta_2}} |D\tilde{\mathbf{g}}(m)|\right) E(|X_i| + |Y_i| \mid \theta_i = t)E(|\eta_i| \mid \theta_i = t)f_\theta(t)dt \\
&+ \int_{q_{1-\delta_2}(\theta_i)}^{\infty} B\bar{D}(t)E(|X_i| + |Y_i| \mid \theta_i = t)E(|\eta_i| \mid \theta_i = t)f_\theta(t)dt \\
&\leq \frac{B}{J_n^{1/2}}\left\{E\left(\bar{D}(\theta_i)(|X_i| + |Y_i|)\right) + \left(\sup_{m\in\mathcal{M}_{\delta_2}} |D\tilde{\mathbf{g}}(m)|\right) E(|X_i| + |Y_i|)\right\} = o(1)
\end{aligned}
$$

where the second inequality follows since $\sup_{t\in\mathbb{R}} E(|\eta_i| \mid \theta_i = t) \leq (\sup_{t\in\mathbb{R}} E(\eta_i^2 \mid \theta_i = t))^{1/2} \leq J_n^{-1/2}$ and the final equality follows because $\sup_{m\in\mathcal{M}_{\delta_2}} |D\tilde{\mathbf{g}}(m)| = O(1)$ by Theorem C.1 and $E\left(\bar{D}(\theta_i)(|X_i| + |Y_i|)\right)$ and $E(|X_i| + |Y_i|)$ are both bounded by (c) and (e) of Assumption A.2.

The second term in equation (B.4) is bounded by

$$
\begin{aligned}
&BE(\mathbf{1}(\theta_i \in \Theta_{\delta_2})|\eta_i||(\hat{\xi}_i - \hat{\gamma}'\tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i)))(X_i - \tilde{g}_x(\bar{p}_{J_n}(\theta_i)))|) \\
&\leq \frac{B}{J_n^{1/2}} E(\mathbf{1}(\theta_i \in \Theta_{\delta_2})|(\hat{\xi}_i - \hat{\gamma}'\tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i)))(X_i - \tilde{g}_x(\bar{p}_{J_n}(\theta_i)))|) \\
&\leq \frac{B}{J_n^{1/2}} \sup_{\beta_1\in\mathcal{B}, m\in\mathcal{M}_{\delta_2}} E(|(\xi_i(\beta_1) - \gamma(\beta_1)'\tilde{\mathbf{g}}(m))(X_i - \tilde{g}_x(m))|) = o(1),
\end{aligned}
$$

where convergence follows by Assumption A.2(c) and Theorem C.1.

Next, the second term in equation (B.2) can be bounded as follows.

$$
\begin{aligned}
&|M(\hat{\beta}_{1J_n}, \tilde{\mathbf{g}}, \hat{w}) - M(\hat{\beta}_{1J_n}, \tilde{\mathbf{g}}, w_{0,J_n})| \\
&\leq E(|\hat{w}(\bar{M}_{iJ_n}) - w_{0,J_n}(\bar{M}_{iJ_n})||(\hat{\xi}_i - \hat{\gamma}'\tilde{\mathbf{g}}(\bar{M}_{iJ_n}))(X_i - \tilde{g}_x(\bar{M}_{iJ_n}))|) \\
&\leq C\left(1 + \left(\sup_{m\in\hat{\mathcal{M}}\cup\mathcal{M}_{\delta_2}} |\hat{\mathbf{g}}(m)|\right)^2\right) \sup_{m\in\hat{\mathcal{M}}\cup\mathcal{M}_{\delta_2}} |\hat{w}(m) - w_{0,J_n}(m)| = o_p(1)
\end{aligned}
$$

where the second inequality follows for some constant $C > 0$ by conditions (a) and (c) of Assumption A.2 and the final equality follows by using Theorem C.1 and Assumption A.2(d) as above since $\delta_2 \geq \delta_1$ implies that $\mathcal{M}_{\delta_2} \subseteq \mathcal{M}_{\delta_1}$ and hence $Pr(\hat{\mathcal{M}} \cup \mathcal{M}_{\delta_2} \subset \mathcal{M}_{\delta_1}) = Pr(\hat{\mathcal{M}} \subset \mathcal{M}_{\delta_1}) \to 0$.

Next, $|\hat{M}_n(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w}) - M(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w})| = o_p(1)$ by applying Theorem B.2. Let $\Gamma_n = \mathcal{B} \times \{(w, \mathbf{g}) : w(m) = 0 \forall m \notin \mathcal{M}_{\delta_1}, \sup_{m\in\mathcal{M}_{\delta_1}} |\mathbf{g}(m)| < B, \sup_{m\in\mathcal{M}_{\delta_1}} |D\mathbf{g}(m)| < B, \sup_{m\in\mathcal{M}_{\delta_1}} |w(m)| < B, \sup_{m\in\mathcal{M}_{\delta_1}} |Dw(m)| < B\}$. Define the metric $d_n((\beta_1', \mathbf{g}', w'), (\beta_1, \mathbf{g}, w)) = |\beta_1' - \beta_1| + \sup_{m\in\mathcal{M}_{\delta_1}} |\mathbf{g}'(m) -$

$\mathbf{g}(m)| + \sup_{m\in\mathcal{M}_{\delta_1}} |w'(m) - w(m)|$. Both Γ_n and d_n vary with n because $\mathcal{M}_{\delta_1}$ varies with J_n.

The space Γ_n is uniformly totally bounded because Θ_{δ_1} is compact and because of the conditions in Assumption A.1(b) controlling $\{\bar{p}_J : J \geq J_0\}$. Condition (c) in Theorem B.2 is satisfied under conditions (a) and (c) of Assumption A.1. Condition (b) in the theorem follows from Theorem B.1 since the random variable $|w(\bar{M}_{iJ_n})(\xi_i(\beta_1) - \gamma(\beta_1)'\mathbf{g}(\bar{M}_{iJ_n}))(X_i - g_x(\bar{M}_{iJ_n}))|$ is bounded by a random variable that has finite absolute mean when $(\beta_1, \mathbf{g}, w) \in \Gamma_n$ by condition (c) of Assumption A.2. Lastly, $Pr((\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w}) \in \Gamma_n) \leq Pr(\sup_{m\in\mathcal{M}_{\delta_1}} |\hat{\mathbf{g}}(m)| < B) + Pr(\sup_{m\in\mathcal{M}_{\delta_1}} |D\hat{\mathbf{g}}(m)| < B) + Pr(\sup_{m\notin\mathcal{M}_{\delta_1}} |\hat{w}(m)| = 0, |\hat{w}(m)| \leq B, |D\hat{w}(m)| \leq B)$ and each of the first two terms converges to 1 by Theorem C.1 and the third converges to 1 by Assumption A.2(d).

Thus, I have shown that

$$|M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n})| \leq o_p(1) + |\hat{M}_n(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w})|$$

But, $|\hat{M}_n(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w})| = \inf_{\beta_1\in\mathcal{B}} |\hat{M}_n(\beta_1, \hat{\mathbf{g}}, \hat{w})|$ and

$$\begin{aligned}
\inf_{\beta_1\in\mathcal{B}} |\hat{M}_n(\beta_1, \hat{\mathbf{g}}, \hat{w})| &\leq \inf_{\beta_1\in\mathcal{B}} \left\{ \left|\hat{M}_n(\beta_1, \hat{\mathbf{g}}, \hat{w}) - M^*(\beta_1, \mathbf{h}_0, \tau_{0,J_n})\right| + M^*(\beta_1, \mathbf{h}_0, \tau_{0,J_n}) \right\} \\
&\leq \inf_{\beta_1\in\mathcal{B}} \left|\hat{M}_n(\beta_1, \hat{\mathbf{g}}, \hat{w}) - M^*(\beta_1, \mathbf{h}_0, \tau_{0,J_n})\right| + \inf_{\beta_1\in\mathcal{B}} |M^*(\beta_1, \mathbf{h}_0, \tau_{0,J_n})| \\
&\leq \left|\hat{M}_n(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w}) - M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n})\right| + \inf_{\beta_1\in\mathcal{B}} |M^*(\beta_1, \mathbf{h}_0, \tau_{0,J_n})| \\
&= o_p(1) + \inf_{\beta_1\in\mathcal{B}} |M^*(\beta_1, \mathbf{h}_0, \tau_{0,J_n})| = o_p(1),
\end{aligned}$$

where the third inequality follows since $\hat{\beta}_{1J_n} \in \mathcal{B}$, the first equality follows from (B.1), and the second equality follows because $\beta_{10} \in \mathcal{B}$ and $M^*(\beta_{10}, \mathbf{h}_0, \tau_{0,J_n}) = 0$.

Therefore,

$$|M^*(\hat{\beta}_{1J_n}, \mathbf{h}_0, \tau_{0,J_n})| \leq o_p(1) + |\hat{M}_n(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w})| = o_p(1)$$

□

Theorem 3.2 can be proved through a few lemmas. Let $\hat{Z}_n(\mathbf{g}, w) = n^{-1}\sum_{i=1}^n w(\bar{M}_i)(X_i - g_x(\bar{M}_i))(Z_i - \mathbf{g}(\bar{M}_i))'$ and $\hat{Z}_n^*(\mathbf{h}, \tau) = n^{-1}\sum_{i=1}^n \tau(\theta_i)(X_i - h_x(\theta_i))(Z_i - \mathbf{h}(\theta_i))'$. Then $\hat{M}_n(\beta_1, \mathbf{g}, w) = \hat{Z}_n(\mathbf{g}, w)\gamma(\beta_1)'$ where $\gamma(\beta_1) = (1, \beta_1')'$. I can also define $\hat{M}_n^*(\beta_1, \mathbf{h}, \tau) = \hat{Z}_n^*(\mathbf{h}, \tau)\gamma(\beta_1)'$ and

$$\hat{Q}_n(\mathbf{g}, w) := n^{-1}\sum_{i=1}^n w(\bar{M}_i)(X_i - g_x(\bar{M}_i))(X_i - g_x(\bar{M}_i))' = \hat{Z}_n(\mathbf{g}, w)A'$$

where A is the $K\times K{+}1$ matrix $[0_{K\times 1} \quad I_K]$. Then let $\hat{Q}_n := \hat{Q}_n(\hat{\mathbf{g}}, \hat{w})$ and $\hat{Q}_n^*(\mathbf{h}, \tau) := \hat{Z}_n^*(\mathbf{h}, \tau)A'$.

As in the proof of Theorem 3.1, I consider a sequence $\{J_n : n \geq 1\}$ and then derive the stated results as $n \to \infty$. I first state the following lemmas, which will then be used to prove Theorem 3.2.

Lemma B.1. *Under the assumptions of Theorem 3.2,*

$$\textit{(a)}\ \sqrt{n}\left\{\hat{Z}_n(\hat{\mathbf{g}},\hat{w}) - \hat{Z}_n(\tilde{\mathbf{g}},w_{0,J_n}) - E\left(\hat{Z}_n(\hat{\mathbf{g}},\hat{w}) - \hat{Z}_n(\tilde{\mathbf{g}},w_{0,J_n})\right)\right\} \to_p 0 \quad \textit{and}$$
$$\textit{(b)}\ \sqrt{n}\left\{\hat{Z}_n(\tilde{\mathbf{g}},w_{0,J_n}) - \hat{Z}_n^*(\mathbf{h}_0,\tau_{0,J_n}) - E\left(\hat{Z}_n(\tilde{\mathbf{g}},w_{0,J_n}) - \hat{Z}_n^*(\mathbf{h}_0,\tau_{0,J_n})\right)\right\} \to_p 0$$

Lemma B.2. *Under the assumptions of Theorem 3.2,* $\sqrt{n}(E\left(\hat{M}_n(\beta_{10},\hat{\mathbf{g}},\hat{w})\right) - B_{1J_n}) = o_p(1)$ *and* $B_{1J_n} = O(J_n^{-1})$ *where* $B_{1J} = E\left(\tau_{0J}(\theta_i)\eta_i^2 Dh(\theta_i)Dh_x(\theta_i)\right)$*.*

Lemma B.3. *Under the assumptions of Theorem 3.2,* $\hat{Q}_n - Q^*_{0,J_n} = O_p(r_n) + O\left((J_n^{-1}\log(J_n))^{1/2}\right)$ *where* $r_n = h_n^2 + \frac{\log(n)}{\sqrt{n h_n}} + \frac{\log(J_n)^{p/2}}{h_n^{(p-1)} J_n^{p/2}}$*.*

Proof of Theorem 3.2. First,

$$\begin{aligned}
&\sqrt{n}\left(\hat{M}_n(\beta_{10},\hat{\mathbf{g}},\hat{w}) - B_{1J_n}\right) \qquad\qquad \text{(B.5)}\\
&= \sqrt{n}\left\{\hat{M}_n(\beta_{10},\hat{\mathbf{g}},\hat{w}) - \hat{M}_n^*(\beta_{10},\mathbf{h}_0,\tau_{0,J_n}) - E\left(\hat{M}_n(\beta_{10},\hat{\mathbf{g}},\hat{w}) - \hat{M}_n^*(\beta_{10},\mathbf{h}_0,\tau_{0,J_n})\right)\right\}\\
&+ \sqrt{n}\left(E\left(\hat{M}_n(\beta_{10},\hat{\mathbf{g}},\hat{w})\right) - B_{1J_n}\right) + \sqrt{n}\left(\hat{M}_n^*(\beta_{10},\mathbf{h}_0,\tau_{0,J_n}) - E(\hat{M}_n^*(\beta_{10},\mathbf{h}_0,\tau_{0,J_n}))\right)
\end{aligned}$$

By Lemma B.1, since $\hat{M}_n(\beta_{10},\mathbf{g},w) = \hat{Z}_n(\mathbf{g},w)\gamma_0'$ and $\hat{M}_n^*(\beta_{10},\mathbf{h},\tau) = \hat{Z}_n^*(\mathbf{h},\tau)\gamma_0'$, the first term is $o_p(1)$. By Lemma B.2 the second term is also $o_p(1)$. Therefore,

$$\begin{aligned}
&\sqrt{n}\left(\hat{M}_n(\beta_{10},\hat{\mathbf{g}},\hat{w}) - B_{1J_n}\right) \qquad\qquad \text{(B.6)}\\
&= \sqrt{n}\left(\hat{M}_n^*(\beta_{10},\mathbf{h}_0,\tau_{0,J_n}) - E(\hat{M}_n^*(\beta_{10},\mathbf{h}_0,\tau_{0,J_n}))\right) + o_p(1)
\end{aligned}$$

Since condition (b) of Assumption A.4 implies that $\sup_n ||V_{1J_n}^{-1/2}|| < \infty$,

$$\begin{aligned}
&\sqrt{n}V_{1J_n}^{-1/2}(\hat{M}_n(\beta_{10},\hat{\mathbf{g}},\hat{w}) - B_{1J_n}) \qquad\qquad \text{(B.7)}\\
&= \sqrt{n}V_{1J_n}^{-1/2}\left(\hat{M}_n^*(\beta_{10},\mathbf{h}_0,\tau_{0,J_n}) - E(\hat{M}_n^*(\beta_{10},\mathbf{h}_0,\tau_{0,J_n}))\right) + o_p(1)\\
&\to_d N(0,I)
\end{aligned}$$

where the last line follows from the Lindeberg-Feller central limit theorem for triangular arrays since condition (c) of Assumption A.4 implies the Lyapounov condition and condition (b) implies that $\sup_n ||V_{1J_n}^{-1/2}|| < \infty$ where $V_{1J_n} = E(\tau_{0J_n}(\theta_i)^2 e_i^2 (X_i - h_{0x}(\theta_i))(X_i - h_{0x}(\theta_i))') = Var(\sqrt{n}\hat{M}_n^*(\beta_{10},\mathbf{h}_0,\tau_{0,J_n}))$.

Next, for any β_1, $\hat{M}_n(\beta_1,\hat{\mathbf{g}},\hat{w}) = \hat{M}_n(\beta_{10},\hat{\mathbf{g}},\hat{w}) - \hat{Q}_n(\beta_1 - \beta_{10})$. Therefore, $\sqrt{n}\hat{Q}_n(\hat{\beta}_{1J_n} - \beta_{10}) = \sqrt{n}\hat{M}_n(\beta_{10},\hat{\mathbf{g}},\hat{w}) - \sqrt{n}\hat{M}_n(\hat{\beta}_{1J_n},\hat{\mathbf{g}},\hat{w})$. Rather than assuming that $\hat{M}_n(\hat{\beta}_{1J_n},\hat{\mathbf{g}},\hat{w}) = 0$, the following argument shows that $\sqrt{n}\hat{M}_n(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w}) = o_p(1)$.

The result in (B.7) and condition (b) of Assumption A.4 together imply that $\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w}) - B_{1J_n} = O_p(n^{-1/2})$. Further, since $B_{1J_n} = o(1)$ by Lemma B.2, we have that $\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w}) = o_p(1)$. The, for each β_1, define $\tilde{M}_n(\beta_1, \hat{\mathbf{g}}, \hat{w}) := \hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w}) - Q^*_{0,J_n}(\beta_1 - \beta_{10})$. Let $\bar{\beta}_1 = \beta_{10} + Q^{*-1}_{0,J_n}\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w})$ so that $\tilde{M}_n(\bar{\beta}_1, \hat{\mathbf{g}}, \hat{w}) = 0$. Then $\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w}) = o_p(1)$, so condition (b) of Assumption A.4 implies that $\bar{\beta}_1 - \beta_{10} = Q^{*-1}_{0,J_n}\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w}) \to_p 0$. By condition (a) of Assumption A.4, $\beta_{10} \in int(\mathcal{B})$, so I can assume that $\bar{\beta}_1 \in \mathcal{B}$. Therefore, condition (a) of Assumption A.4 also implies that $|\hat{M}_n(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w})| = \inf_{\beta_1 \in \mathcal{B}} |\hat{M}_n(\beta_1, \hat{\mathbf{g}}, \hat{w})| \leq |\hat{M}_n(\bar{\beta}_1, \hat{\mathbf{g}}, \hat{w})| + o_p(n^{-1/2})$. So it remains to show that $\sqrt{n}\hat{M}_n(\bar{\beta}_1, \hat{\mathbf{g}}, \hat{w}) = o_p(1)$.

But since $\tilde{M}_n(\bar{\beta}_1, \hat{\mathbf{g}}, \hat{w}) = 0$, $|\hat{M}_n(\bar{\beta}_1, \hat{\mathbf{g}}, \hat{w})| \leq |\tilde{M}_n(\bar{\beta}_1, \hat{\mathbf{g}}, \hat{w})| + |(\hat{Q}_n - Q^*_{0,J_n})(\bar{\beta}_1 - \beta_{10})| = |(\hat{Q}_n - Q^*_{0,J_n})(\bar{\beta}_1 - \beta_{10})|$. And $0 = \tilde{M}_n(\bar{\beta}_1, \hat{\mathbf{g}}, \hat{w}) = \hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w}) - Q^*_{0,J_n}(\bar{\beta}_1 - \beta_{10})$ so that

$$
\begin{aligned}
(\hat{Q}_n - Q^*_{0,J_n})(\bar{\beta}_1 - \beta_{10}) &= (\hat{Q}_n - Q^*_{0,J_n})(\bar{\beta}_1 - \beta_{10} - B_{J_n}) + (\hat{Q}_n - Q^*_{0,J})B_{J_n} \\
&= (\hat{Q}_n - Q^*_{0,J_n})Q^{*-1}_{0,J_n}(\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w}) - B_{1J_n}) + (\hat{Q}_n - Q^*_{0,J})B_{J_n}
\end{aligned}
$$

I have already shown that $(\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w}) - B_{1J_n}) = O_p(n^{-1/2})$ so the first term here is $o_p(n^{-1/2})$ by Lemma B.3. Applying both Lemmas B.2 and B.3, the second term is

$$
\left(O_p\left(h_n^2 + \frac{\log(n)}{\sqrt{nh_n}} + \frac{\log(J_n)^{p/2}}{h_n^{(p-1)} J_n^{p/2}} \right) + O((J_n^{-1}\log(J_n))^{1/2}) \right) O(J_n^{-1}),
$$

which is $o_p(n^{-1/2})$ by conditions (e) and (g) of Assumption A.3. Thus I have shown that $\sqrt{n}\hat{M}_n(\hat{\beta}_{1J_n}, \hat{\mathbf{g}}, \hat{w}) = o_p(1)$ and therefore $\sqrt{n}\hat{Q}_n(\hat{\beta}_{1J_n} - \beta_{10}) = \sqrt{n}\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w}) + o_p(1)$.

Next, since $B_J = Q^{*-1}_{0,J}B_{1J}$, $\sup_n ||V^{-1/2}_{1J_n}|| < \infty$, and, as just shown, $(\hat{Q}_n - Q^*_{0,J_n})B_{J_n} = o_p(n^{-1/2})$,

$$
\begin{aligned}
&\sqrt{n}V^{-1/2}_{1J_n}Q^*_{0,J_n}(\hat{\beta}_{1J_n} - \beta_{10} - B_{J_n}) \\
&= \sqrt{n}V^{-1/2}_{1J_n}\hat{Q}_n(\hat{\beta}_{1J_n} - \beta_{10} - B_{J_n}) + \sqrt{n}V^{-1/2}_{1J_n}(Q^*_{0,J_n} - \hat{Q}_n)(\hat{\beta}_{1J_n} - \beta_{10} - B_{J_n}) \\
&= \sqrt{n}V^{-1/2}_{1J_n}\left(\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w}) - \hat{Q}_n B_{J_n} \right) + \sqrt{n}V^{-1/2}_{1J_n}(Q^*_{0,J_n} - \hat{Q}_n)(\hat{\beta}_{1J_n} - \beta_{10} - B_{J_n}) \\
&= \sqrt{n}V^{-1/2}_{1J_n}(\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w}) - B_{1J_n}) \\
&+ \left(V^{-1/2}_{1J_n}(Q^*_{0,J_n} - \hat{Q}_n)Q^{*-1}_{0,J_n}V^{1/2}_{1J_n} \right)\left(\sqrt{n}V^{-1/2}_{1J_n}Q^*_{0,J_n}(\hat{\beta}_{1J_n} - \beta_{10} - B_{J_n}) \right) + o_p(1)
\end{aligned}
$$

Then conditions (b) and (c) of Assumption A.4 and Lemma B.3 imply that

$$
\left(V^{-1/2}_{1J_n}(Q^*_{0,J_n} - \hat{Q}_n)Q^{*-1}_{0,J_n}V^{1/2}_{1J_n} \right) = o_p(1)
$$

so that

$$\sqrt{n}V_{1J_n}^{-1/2}Q_{0,J_n}^*(\hat{\beta}_{1J_n}-\beta_{10}-B_{J_n})=\frac{\sqrt{n}V_{1J_n}^{-1/2}(\hat{M}_n(\beta_{10},\hat{\mathbf{g}},\hat{w})-B_{1J_n})}{1-o_p(1)}+o_p(1)\to_d N(0,I)$$

Since $B_{J_n}=O(J_n^{-1})$ it follows from conditions (b) and (c) of Assumption A.4 that $\hat{\beta}_{1J_n}-\beta_{10}=O_p(n^{-1/2})+O(J_n^{-1})$.

If $V_{1J_n}\to\bar{V}_1$ and $Q_{0,J_n}^*\to\bar{Q}_0^*$ then $\sqrt{n}(\hat{\beta}_{1J_n}-\beta_{10}-B_{J_n})=\left(Q_{0,J_n}^{*-1}V_{1_{J_n}}^{1/2}\right)\sqrt{n}V_{1_{J_n}}^{-1/2}Q_{0,J_n}^*(\hat{\beta}_{1_{J_n}}-\beta_{10}-B_{J_n})\to_d N(0,\bar{Q}_0^{*-1}\bar{V}_1\bar{Q}_0^{*-1})$. If, in addition, $\sqrt{n}B_{J_n}\to\gamma\bar{B}$ then $\sqrt{n}(\hat{\beta}_{1_{J_n}}-\beta_{10})=\sqrt{n}(\hat{\beta}_{1_{J_n}}-\beta_{10}-B_{J_n})+\sqrt{n}B_{J_n}\to_d N(\gamma\bar{B},\bar{V})$. □

Proof of Lemma B.1. Proof of (a): Let $\bar{\mathcal{M}}^*=\bar{p}_\infty(\Theta_\delta)$ for some $0<\delta<\delta_1$. By conditions (d) and (f) of Assumption A.4, $Pr(\hat{M}\subset\bar{\mathcal{M}}^*)\to 1$.

Next, $\hat{Z}_n(\mathbf{g},w)=\sum_{s=1}^4\hat{Z}_{ns}(\mathbf{g},w)$ where $\hat{Z}_{n1}(\mathbf{g},w)=n^{-1}\sum_{i=1}^n w(\bar{M}_{iJ_n})X_iZ_i'$, $\hat{Z}_{n2}(\mathbf{g},w)=-n^{-1}\sum_{i=1}^n w(\bar{M}_{iJ_n})X_i\mathbf{g}(\bar{M}_{iJ_n})'$, $\hat{Z}_{n3}(\mathbf{g},w)=-n^{-1}\sum_{i=1}^n w(\bar{M}_{iJ_n})g_x(\bar{M}_{iJ_n})Z_i'$, and $\hat{Z}_{n4}(\mathbf{g},w)=n^{-1}\sum_{i=1}^n w(\bar{M}_{iJ_n})g_x(\bar{M}_{iJ_n})\mathbf{g}(\bar{M}_{iJ_n})'$.

Then stochastic equicontinuity results of Andrews (1994b) can be applied to each of these four terms separately. For positive integers r,s let $\Gamma_{0,r,s}$ be the space of $r\times s$ matrix-valued functions, $\{\mathbf{f}:\mathbf{f}(m)=0\forall m\notin\bar{\mathcal{M}}^*,\sup_{m\in\bar{\mathcal{M}}^*}|\mathbf{f}(m)|<B,\sup_{m\in\bar{\mathcal{M}}^*}|D\mathbf{f}(m)|<B\}$. Then let $\Gamma_1=\{xz'\}\times\Gamma_{0,1,1}$ and let $\rho_1(f^*,f)=\sup_n E\left(|(f^*(\bar{M}_{iJ_n})-f(\bar{M}_{iJ_n}))X_iZ_i|^2\right)^{1/2}$. Then, by Theorems 1-3 of Andrews (1994b) and condition **LIx** of Assumption 2.2 and condition (c) of Assumption A.4, for any sequence $\delta_n\to 0$,

$$\sup_{f,f^*\in\Gamma_1,\rho_1(f^*,f)<\delta_n}||v_{n1}(f^*)-v_{n1}(f))||\to_p 0$$

where $v_{n1}(f)=n^{-1/2}\sum_{i=1}^n f(\bar{M}_{iJ_n})X_iZ_i'=\sqrt{n}\hat{Z}_{n1}(\mathbf{g},w)$.

Similarly, let $\Gamma_2=\{x\}\times\Gamma_{0,K+1,1}$ and let $\rho_2(\mathbf{f}^*,\mathbf{f})=\sup_n E\left(|X_i(\mathbf{f}^*(\bar{M}_{iJ_n})-\mathbf{f}(\bar{M}_{iJ_n}))'|^2\right)^{1/2}$. Then, by Theorems 1-3 of Andrews (1994b) and condition **LIx** of Assumption 2.2 and condition (c) of Assumption A.4, for any sequence $\delta_n\to 0$,

$$\sup_{\mathbf{f},\mathbf{f}^*\in\Gamma_2,\rho_2(\mathbf{f}^*,\mathbf{f})<\delta_n}||v_{n2}(\mathbf{f}^*)-v_{n2}(\mathbf{f}))||\to_p 0$$

where $v_{n2}(\mathbf{f})=n^{-1/2}\sum_{i=1}^n X_i\mathbf{f}(\bar{M}_{iJ_n})'$ and $v_{n2}(w\mathbf{g})=\sqrt{n}\hat{Z}_{n2}(\mathbf{g},f)$.

Third, let $\Gamma_3=\{z\}\times\Gamma_{0,K,1}$ and let $\rho_3(\mathbf{f}^*,\mathbf{f})=\sup_n E\left(|(\mathbf{f}^*(\bar{M}_{iJ_n})-\mathbf{f}(\bar{M}_{iJ_n}))Z_i'|^2\right)^{1/2}$. Then, by Theorems 1-3 of Andrews (1994b) and condition **LIx** of Assumption 2.2 and condition (c) of

Assumption A.4, for any sequence $\delta_n \to 0$,

$$\sup_{\mathbf{f},\mathbf{f}^* \in \Gamma_3, \rho_3(\mathbf{f}^*,\mathbf{f})<\delta_n} ||v_{n3}(\mathbf{f}^*) - v_{n3}(\mathbf{f})|| \to_p 0$$

where $v_{n3}(\mathbf{f}) = n^{-1/2}\sum_{i=1}^n \mathbf{f}(\bar{M}_{iJ_n})Z_i'$ and $v_{n3}(wg_x) = \sqrt{n}\hat{Z}_{n3}(\mathbf{g}, f)$.

Lastly, let $\Gamma_4 = \Gamma_{0,K,K+1}$ and let $\rho_4(\mathbf{f}^*,\mathbf{f}) = \sup_n E\left(||\mathbf{f}^*(\bar{M}_{iJ_n}) - \mathbf{f}(\bar{M}_{iJ_n})||^2\right)^{1/2}$. Then, by Theorems 1-3 of Andrews (1994b) and condition **LIx** of Assumption 2.2 and condition (c) of Assumption A.4, for any sequence $\delta_n \to 0$,

$$\sup_{\mathbf{f},\mathbf{f}^* \in \Gamma_4, \rho_4(\mathbf{f}^*,\mathbf{f})<\delta_n} ||v_{n4}(\mathbf{f}^*) - v_{n4}(\mathbf{f})|| \to_p 0$$

where $v_{n4}(\mathbf{f}) = n^{-1/2}\sum_{i=1}^n \mathbf{f}(\bar{M}_{iJ_n})$ and $v_{n4}(wg_x\mathbf{g}') = \sqrt{n}\hat{Z}_{n4}(\mathbf{g}, f)$.

Then (a) follows since Theorem C.2 implies (1) that $Pr(\hat{w} \in \Gamma_{0,1,1})$, $Pr(\hat{w}\hat{\mathbf{g}} \in \Gamma_{0,K+1,1})$, $Pr(\hat{w}\hat{g}_x \in \Gamma_{0,K,1})$ and $Pr(\hat{w}\hat{g}_x\hat{\mathbf{g}}' \in \Gamma_{0,K,K+1})$ each converge to 1 and (2) that $\rho_1(\hat{w}, w_{0,J_n}) \to_p 0$, $\rho_2(\hat{w}\hat{\mathbf{g}}, w_{0,J_n}\tilde{\mathbf{g}}) \to_p 0$, $\rho_3(\hat{w}\hat{g_x}, w_{0,J_n}\tilde{\mathbf{g}}) \to_p 0$, and $\rho_4(\hat{w}\hat{g}_x\hat{\mathbf{g}}', w_{0,J_n}\tilde{g}_x\tilde{\mathbf{g}}') \to_p 0$.

Proof of (b): Let

$$\begin{aligned}
\hat{m}_1 = &\frac{1}{n}\sum_{i=1}^n w(\bar{M}_{iJ_n})(X_i - \tilde{g}_x(\bar{M}_{iJ_n}))(Z_i - \tilde{\mathbf{g}}(\bar{M}_{iJ_n}))' - \frac{1}{n}\sum_{i=1}^n w(\bar{M}_{iJ_n})(X_i - h_{x,0}(\theta_i))(Z_i - \mathbf{h}_0(\theta_i))' \\
&- E\left(w(\bar{M}_{iJ_n})(X_i - \tilde{g}_x(\bar{M}_{iJ_n}))(Z_i - \tilde{\mathbf{g}}(\bar{M}_{iJ_n}))' - w(\bar{M}_{iJ_n})(X_i - h_{x,0}(\theta_i))(Z_i - \mathbf{h}_0(\theta_i))'\right)
\end{aligned}$$

and

$$\begin{aligned}
\hat{m}_2 = &\frac{1}{n}\sum_{i=1}^n w(\bar{M}_{iJ_n})(X_i - h_{x,0}(\theta_i))(Z_i - \mathbf{h}_0(\theta_i))' - \frac{1}{n}\sum_{i=1}^n w(\bar{p}_{J_n}(\theta_i))(X_i - h_{x,0}(\theta_i))(Z_i - \mathbf{h}_0(\theta_i))' \\
&- E\left(w(\bar{M}_{iJ_n})(X_i - h_{x,0}(\theta_i))(Z_i - \mathbf{h}_0(\theta_i))' - w(\bar{p}_{J_n}(\theta_i))(X_i - h_{x,0}(\theta_i))(Z_i - \mathbf{h}_0(\theta_i))'\right)
\end{aligned}$$

Then (b) follows if $\sqrt{n}\hat{m}_1 = o_p(1)$ and $\sqrt{n}\hat{m}_2 = o_p(1)$.

First, consider $Var(w(\bar{M}_{iJ_n})(\tilde{g}_s(\bar{M}_{iJ_n}) - h_{0,s}(\theta_i))V_i)$ for V_i equal to Y_i, a component of the vector X_i, or a component of the vector $\mathbf{h}_0(\theta_i)$ where $\tilde{g}_s$ and $h_{0,s}$ represent any component of the vectors $\tilde{\mathbf{g}}$ and $\mathbf{h}_0$, respectively. By a Taylor expansion, $\tilde{g}_s(\bar{M}_{iJ_n}) - h_{0,s}(\theta_i) = D\tilde{g}_s(p_i^*)\eta_i$ for some p_i^* between $\bar{p}_{J_n}(\theta_i)$ and $\bar{M}_{iJ_n}$ so

$$Var(w(\bar{M}_{iJ_n})(\tilde{g}_s(\bar{M}_{iJ_n}) - h_{0,s}(\theta_i))V_i) \leq E(w(\bar{M}_{iJ_n})^2 D\tilde{g}_s(p_i^*)^2\eta_i^2 V_i^2)$$

Next, $E(w(\bar{M}_{iJ_n})^2 D\tilde{g}_s(p_i^*)^2\eta_i^2 V_i^2) = \int E(w(\bar{M}_{iJ_n})^2 D\tilde{g}_s(p_i^*)^2\eta_i^2 V_i^2 \mid \theta_i = \theta)f_\theta(\theta)d\theta$. For $\theta \in \Theta_{\delta_2}$, both $\bar{p}_{J_n}(\theta_i)$ and $\bar{M}_{iJ_n}$ are in $\mathcal{M}_{\delta_2}$ (unless $w(\bar{M}_{iJ_n}) = 0$) so $p_i^* \in \mathcal{M}_{\delta_2}$ and, therefore,

$$
\begin{aligned}
&\int_{\Theta_{\delta_2}} E(w(\bar{M}_{iJ_n})^2 D\tilde{g}_s(p_i^*)^2 \eta_i^2 V_i^2 \mid \theta_i = \theta) f_\theta(\theta) d\theta \\
&\leq B^2 \left(\sup_{m \in \mathcal{M}_{\delta_2}} |D\tilde{g}_s(m)| \right)^2 \int_{\Theta_{\delta_2}} E(V_i^2 \mid \theta_i = t) E(\eta_i^2 \mid \theta_i = t) f_\theta(\theta) d\theta \\
&\leq \frac{1}{J_n} B^2 \left(\sup_{m \in \mathcal{M}_{\delta_2}} |D\tilde{g}_s(m)| \right)^2 \int_{\Theta_{\delta_2}} E(V_i^2) f_\theta(\theta) d\theta
\end{aligned}
$$

where I have used (a) the fact that $E(V_i^2 \eta_i^2 \mid \theta_i) = E(V_i^2 \mid \theta_i) E(\eta_i^2 \mid \theta_i)$ for V_i equal to Y_i, a component of the vector X_i, or a component of the vector $\mathbf{h}_0(\theta_i)$, by Assumption 2.2, (b) the fact that $\sup_\theta E(\eta_i^2 \mid \theta_i = \theta) \leq J_n^{-1}$, by Lemma C.1, and (c) condition (d) of Assumption A.4 which implies that the function $w(m)$ is bounded uniformly by B.

Next, by condition (e) of Assumption A.4, $|D\tilde{g}_s(m)| \leq \bar{D}(\bar{p}_{J_n}^{-1}(m))$ where $\bar{D}(\cdot)$ is nonincreasing on the interval $(-\infty, q_{\delta_2}(\theta_i)]$ and nondecreasing on $[q_{1-\delta_2}(\theta_i), \infty)$. Then

$$
\begin{aligned}
&\int_{\Theta \backslash \Theta_{\delta_2}} E(w(\bar{M}_{iJ_n})^2 D\tilde{g}_s(p_i^*)^2 \eta_i^2 V_i^2 \mid \theta_i = \theta) f_\theta(\theta) d\theta \\
&\leq B^2 \int_{\Theta \backslash \Theta_{\delta_2}} \bar{D}(t)^2 E(V_i^2 \mid \theta_i = \theta) E(\eta_i^2 \mid \theta_i = \theta) f_\theta(\theta) d\theta \\
&\leq \frac{1}{J_n} B^2 E(\bar{D}(\theta_i)^2 V_i^2)
\end{aligned}
$$

so $Var(w(\bar{M}_{iJ_n})(\tilde{g}_s(\bar{M}_{iJ_n}) - h_{0,s}(\theta_i))V_i) = O(J_n^{-1})$ by condition (e) of Assumption A.4 and Theorem C.2, and by Chebyschev's inequality,

$$
n^{-1/2} \sum_{i=1}^{n} \left\{ w(\bar{M}_{iJ_n})(\tilde{g}_s(\bar{M}_{iJ_n}) - h_{0,s}(\theta_i))V_i - E(w(\bar{M}_{iJ_n})(\tilde{g}_s(\bar{M}_{iJ_n}) - h_{0,s}(\theta_i))V_i) \right\} \to_p 0
$$

Since $\sqrt{n}\hat{m}_1$ can be expanded into a sum of (a finite, fixed number of) terms of this form, the desired result follows.

Next, using the Taylor series approximation $w(\bar{M}_{iJ_n}) - w(\bar{p}_{J_n}(\theta_i)) = w'_{0,J_n}(p_i^*)\eta_i$,

$$
\begin{aligned}
&Var((w(\bar{M}_{iJ_n}) - w(\bar{p}_{J_n}(\theta_i))(X_{ik} - h_{x_k,0}(\theta_i))(Z_{il} - \mathbf{h}_{0,l}(\theta_i))) \\
&\leq B^2 E(\eta_i^2 (X_{ik} - h_{x_k,0}(\theta_i))^2 (Z_{il} - \mathbf{h}_{0,l}(\theta_i))^2) \\
&\leq \frac{1}{J_n} E((X_{ik} - h_{x_k,0}(\theta_i))^2 (Z_{il} - \mathbf{h}_{0,l}(\theta_i))^2) = O(J_n^{-1})
\end{aligned}
$$

by condition (f) of Assumption A.3 and conditions (c) and (d) of Assumption A.4 so that by

Chebyschev's inequality, $\sqrt{n}\hat{m}_2 = o_p(1)$.

$\square$

Proof of Lemma B.2. First, by Assumption 2.2, $E(Y_i \mid X_i, \theta_i, \mathbf{M}_i) = \beta_{10}' X_i + h_0(\theta_i)$, $E(X_i \mid \theta_i, \mathbf{M}_i) = h_{x,0}(\theta_i)$, and $h_{y,0}(\theta_i) := E(Y_i \mid \theta_i) = \beta_{10}' h_{x,0}(\theta_i) + h_0(\theta_i)$, and therefore,

$$
\begin{aligned}
E\left(\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w})\right) &= E\left(\hat{w}(\bar{M}_{iJ_n})(X_i - \hat{g}_x(\bar{M}_{iJ_n}))(\beta_{10}' X_i + h_0(\theta_i) - \hat{g}_y(\bar{M}_{iJ_n}) - \beta_{10}'(X_i - \hat{g}_x(\bar{M}_{iJ_n})))\right) \\
&= E\left(\hat{w}(\bar{M}_{iJ_n})(h_{x,0}(\theta_i) - \hat{g}_x(\bar{M}_{iJ_n}))(\mathbf{h}_0(\theta_i) - \hat{\mathbf{g}}(\bar{M}_{iJ_n}))'\right)\gamma_0 \\
&= E\left(\hat{w}(\bar{M}_{iJ_n})\left\{(h_{x,0}(\theta_i) - \tilde{g}_x(\bar{M}_{iJ_n})) + (\tilde{g}_x(\bar{M}_{iJ_n}) - \hat{g}_x(\bar{M}_{iJ_n}))\right\}\right. \\
&\qquad \left.\cdot \left\{(\mathbf{h}_0(\theta_i) - \tilde{\mathbf{g}}(\bar{M}_{iJ_n})) + (\tilde{\mathbf{g}}(\bar{M}_{iJ_n}) - \hat{\mathbf{g}}(\bar{M}_{iJ_n}))\right\}'\right)\gamma_0
\end{aligned}
$$

By condition (d) of Assumption A.4, I can assume that $\hat{\mathcal{M}} \subset \mathcal{M}_{\delta_1}$ since

$$
\begin{aligned}
Pr(|\sqrt{n}E\left(\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w})\right) - B_{1J_n}| \geq \varepsilon) &\leq Pr(|\sqrt{n}E\left(\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w})\right) - B_{1J_n}| \geq \varepsilon, \hat{\mathcal{M}} \subset \mathcal{M}_{\delta_1}) \\
&\quad + (1 - Pr(\hat{\mathcal{M}} \subset \mathcal{M}_{\delta_1})) \\
&= Pr(|\sqrt{n}E\left(\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w})\right) - B_{1J_n}| \geq \varepsilon, \hat{\mathcal{M}} \subset \mathcal{M}_{\delta_1}) + o(1)
\end{aligned}
$$

Let $\hat{m}_1 = E\left(\hat{w}(\bar{M}_{iJ_n})(h_{x,0}(\theta_i) - \tilde{g}_x(\bar{M}_{iJ_n}))(\mathbf{h}_0(\theta_i) - \tilde{\mathbf{g}}(\bar{M}_{iJ_n}))'\right)\gamma_0$. Then, using a second order taylor expansion of $\tilde{\mathbf{g}}$ and $\tilde{g}_x$,

$$
\begin{aligned}
\hat{m}_1 &= E\left(\hat{w}(\bar{M}_{iJ_n})D\tilde{g}_x(\bar{p}_{J_n}(\theta_i))D\tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))'\eta_i^2\right)\gamma_0 + E\left(\hat{w}(\bar{M}_{iJ_n})D\tilde{g}_x(\bar{p}_{J_n}(\theta_i))D^2\tilde{\mathbf{g}}(p_i^*)'\eta_i^3\right)\gamma_0 \\
&\quad + E\left(\hat{w}(\bar{M}_{iJ_n})D^2\tilde{g}_x(p_i^{**})D\tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))'\eta_i^3\right)\gamma_0 + E\left(\hat{w}(\bar{M}_{iJ_n})D^2\tilde{g}_x(p_i^{**})D^2\tilde{\mathbf{g}}(p_i^*)'\eta_i^4\right)\gamma_0 \\
&= E\left(\hat{w}(\bar{M}_{iJ_n})D\tilde{g}_x(\bar{p}_{J_n}(\theta_i))D\tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))'\eta_i^2\right)\gamma_0 + O_p(J_n^{-3/2})
\end{aligned}
$$

where the second equality follows from conditions (d) and (e) of Assumption A.4 by applying the same argument used above in the proof of Lemma B.1. In addition,

$$
\begin{aligned}
&E\left(\hat{w}(\bar{M}_{iJ_n})D\tilde{g}_x(\bar{p}_{J_n}(\theta_i))D\tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))'\eta_i^2\right)\gamma_0 \\
&= E\left(w_{0,J_n}(\bar{p}_{J_n}(\theta_i))D\tilde{g}_x(\bar{p}_{J_n}(\theta_i))D\tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))'\eta_i^2\right)\gamma_0 \\
&+ E\left((w_{0,J_n}(\bar{M}_{iJ_n}) - w_{0,J_n}(\bar{p}_{J_n}(\theta_i)))D\tilde{g}_x(\bar{p}_{J_n}(\theta_i))D\tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))'\eta_i^2\right)\gamma_0 \\
&+ E\left((\hat{w}(\bar{M}_{iJ_n}) - w_{0,J_n}(\bar{M}_{iJ_n}))D\tilde{g}_x(\bar{p}_{J_n}(\theta_i))D\tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))'\eta_i^2\right)\gamma_0 \\
&= E\left(w_{0,J_n}(\bar{p}_{J_n}(\theta_i))D\tilde{g}_x(\bar{p}_{J_n}(\theta_i))D\tilde{\mathbf{g}}(\bar{p}_{J_n}(\theta_i))'\eta_i^2\right)\gamma_0 + O_p(J_n^{-3/2}) + o_p(n^{-1/2})
\end{aligned}
$$

by (e) and (f) of Assumption A.3 and (d) of Assumption A.4. Thus $\sqrt{n}(\hat{m}_1 - B_{1J_n}) = o_p(1)$.

Next, let $\hat{m}_2 = E\left(\hat{w}(\bar{M}_{iJ_n})(\tilde{g}_x(\bar{M}_{iJ_n}) - \hat{g}_x(\bar{M}_{iJ_n}))(\mathbf{h}_0(\theta_i) - \tilde{\mathbf{g}}(\bar{M}_{iJ_n}))'\right)\gamma_0$. Then using a first order Taylor approximation of $\tilde{\mathbf{g}}$,

$$\begin{aligned}
\sqrt{n}|\hat{m}_2| &\le \sqrt{n} \sup_{m\in\mathcal{M}_{\delta_1}} |\hat{g}_x(m) - \tilde{g}_x(m)| E\left(\hat{w}(\bar{M}_{iJ_n})|D\tilde{\mathbf{g}}(p_i^*)\eta_i|\right)|\gamma_0| \\
&= \sqrt{n} O_p\left((J_n^{-1}\log(J_n))^{1/2}\right) \sup_{m\in\mathcal{M}_{\delta_1}} |\hat{g}_x(m) - \tilde{g}_x(m)||\gamma_0| \\
&= o_p(1)
\end{aligned}$$

where the first equality is due to Lemma C.1 and the second because Theorem C.2 and conditions (e) and (g) of Assumption A.3 imply that $\sup_{m\in\mathcal{M}_{\delta_1}} |\hat{g}_x(m) - \tilde{g}_x(m)| = O_p(r_n) = o_p\left((J_n n^{-1}\log(J_n))^{1/2}\right)$ where $r_n = h_n^2 + \frac{\log(n)}{\sqrt{n h_n}} + \frac{\log(J_n)^{p/2}}{h_n^{(p-1)} J_n^{p/2}}$. By essentially the same argument,

$$\sqrt{n}|\hat{m}_3| \le \sqrt{n} O_p\left((J_n^{-1}\log(J_n))^{1/2}\right) \sup_{m\in\mathcal{M}_{\delta_1}} |\hat{g}_x(m) - \tilde{g}_x(m)||\gamma_0| = o_p(1)$$

and $\sqrt{n}|\hat{m}_4| \le \sqrt{n}\sup_{m\in\mathcal{M}_{\delta_1}} |\hat{g}_x(m) - \tilde{g}_x(m)|^2|\gamma_0| = o_p(1)$ where

$$\begin{aligned}
\hat{m}_3 &= E\left(\hat{w}(\bar{M}_{iJ_n})(h_{x,0}(\theta_i) - \tilde{g}_x(\bar{M}_{iJ_n}))(\tilde{\mathbf{g}}(\bar{M}_{iJ_n}) - \hat{\mathbf{g}}(\bar{M}_{iJ_n}))'\right)\gamma_0 \\
\hat{m}_4 &= E\left(\hat{w}(\bar{M}_{iJ_n})(\tilde{g}_x(\bar{M}_{iJ_n}) - \hat{g}_x(\bar{M}_{iJ_n}))(\tilde{\mathbf{g}}(\bar{M}_{iJ_n}) - \hat{\mathbf{g}}(\bar{M}_{iJ_n}))'\right)\gamma_0
\end{aligned}$$

noting that

$$\begin{aligned}
\sqrt{n} \sup_{m\in\mathcal{M}_{\delta_1}} |\hat{g}_x(m) - \tilde{g}_x(m)|^2 &= \left(n^{1/4} \sup_{m\in\mathcal{M}_{\delta_1}} |\hat{g}_x(m) - \tilde{g}_x(m)|\right)^2 \\
&= \left(O\left(\sqrt{n}(J_n^{-1}\log(J_n))^{1/2}\right) \sup_{m\in\mathcal{M}_{\delta_1}} |\hat{g}_x(m) - \tilde{g}_x(m)|\right)^2
\end{aligned}$$

Therefore

$$\begin{aligned}
\sqrt{n}\left(E\left(\hat{M}_n(\beta_{10}, \hat{\mathbf{g}}, \hat{w})\right) - B_{1J_n}\right) &= \sqrt{n}(\hat{m}_1 - B_{1J_n}) + \sqrt{n}\hat{m}_2 + \sqrt{n}\hat{m}_3 + \sqrt{n}\hat{m}_4 \\
&= o_p(1)
\end{aligned}$$

□

Proof of Lemma B.3. Let $A = [0_{K\times 1} \quad I_K]$ and recall that $\hat{Q}_n(\mathbf{g}, w) = \hat{Z}_n(\mathbf{g}, w)A'$ and $\hat{Q}_n^*(\mathbf{h}, \tau) =$

$\hat{Z}_n^*(\mathbf{h}, \tau)A'$. The desired result follows from the following expansion,

$$\begin{aligned}\hat{Q}_n - Q^*_{0,J_n} = &\left\{\hat{Z}_n(\hat{\mathbf{g}}, \hat{w}) - \hat{Z}_n(\tilde{\mathbf{g}}, w_{0,J_n}) - E\left(\hat{Z}_n(\hat{\mathbf{g}}, \hat{w}) - \hat{Z}_n(\tilde{\mathbf{g}}, w_{0,J_n})\right)\right\} A' + \\ &\left\{\hat{Z}_n(\tilde{\mathbf{g}}, w_{0,J_n}) - \hat{Z}_n^*(\mathbf{h}_0, \tau_{0,J_n}) - E\left(\hat{Z}_n(\tilde{\mathbf{g}}, w_{0,J_n}) - \hat{Z}_n^*(\mathbf{h}_0, \tau_{0,J_n})\right)\right\} A' \\ &+ \hat{Q}_n^*(\mathbf{h}_0, \tau_{0,J_n}) - E\left(\hat{Q}_n^*(\mathbf{h}_0, \tau_{0,J_n})\right) + E\left(\hat{Q}_n(\hat{\mathbf{g}}, \hat{w}) - \hat{Q}_n^*(\mathbf{h}_0, \tau_{0,J_n})\right)\end{aligned}$$

The first two terms are $o_p(n^{-1/2})$ by Lemma B.1. The third term is $O_p(n^{-1/2})$ by application of the Lindeberg-Feller central limit theorem for triangular arrays since condition (b) of Assumption A.4 implies the relevant Lyapounov conditions.

Lastly,

$$\begin{aligned}E\left(\hat{Q}_n(\hat{\mathbf{g}}, \hat{w}) - \hat{Q}_n^*(\mathbf{h}_0, \tau_{0,J_n})\right) = &E\left(\hat{Q}_n(\hat{\mathbf{g}}, \hat{w}) - \hat{Q}_n(\tilde{\mathbf{g}}, w_{0,J_n})\right) \\ &+ E\left(\hat{Q}_n(\tilde{\mathbf{g}}, w_{0,J_n}) - \hat{Q}_n^*(\mathbf{h}_0, \tau_{0,J_n})\right)\end{aligned}$$

The first term is $O_p(h_n^2 + \frac{\log(n)}{\sqrt{nh_n}} + \frac{\log(\tilde{J}_n)^{p/2}}{h_n^{(p-1)}\tilde{J}_n^{p/2}})$ by Theorem C.2 and conditions (c) and (d) of Assumption A.4. The second term is $O\left((J_n^{-1}\log(J_n))^{1/2}\right)$ under conditions (d) and (e) of Assumption A.4, the proof of which is nearly identical to the proof of Lemma B.2. □

B.0.1 Some useful weak laws of large numbers

The following is an extension of Khintchin's WLLN that can be proved using the same methods employed to prove the well-known Kolmogorov-Feller WLLN.

Theorem B.1. *Suppose that for each n, the random variables $V_{1n,}, \ldots, V_{nn}$ are i.i.d. Moreover, suppose that there exists an i.i.d. sequence of random variables $V_{1\infty}, \ldots, V_{i\infty}, \ldots$ such that $Pr(|V_{in}| > |V_{i\infty}|) = 0$ and $E|V_{i\infty}| < \infty$. Then $n^{-1}\sum_{i=1}^n (V_{in} - E(V_{in})) \to_p 0$.*

Next, I provide a uniform WLLN. Let $V_{in}, 1 \leq i \leq n$ be a triangular array of random variables where each V_{in} takes values in a (measurable) space $\mathcal{V}_n$ and for each $n \geq 1$ and each $\gamma \in \Gamma_n$, $h(v, \gamma)$ is a measurable function from $\mathcal{V}_n$ to $\mathbb{R}$. The following theorem extends Theorem 3(a) in Andrews (1992) by explicitly allowing for a triangular array and by allowing the parameter space to vary with n. Each parameter space Γ_n is assigned a metric $d_n(\cdot, \cdot)$. Moreover, a uniform version of the totally bounded assumption in Andrews (1992) is required. The family of parameter spaces, $\{\Gamma_n : n \geq 1\}$, is said to be uniformly totally bounded if for all $\varepsilon > 0$ there exists an integer K such that each space Γ_n can be covered by no more than K balls of radius ε.

Theorem B.2. *If (a) $\{\Gamma_n : n \geq 1\}$ is a uniformly totally bounded family of parameter spaces, (b) for any sequence $\gamma_n \in \Gamma_n$, $n^{-1}\sum_{i=1}^n(h(V_{in}, \gamma_n) - E(h(V_{in}, \gamma_n))) \to_p 0$, and (c) $|h(V_{in}, \gamma') -$*

$h(V_{in},\gamma)| \leq |f_1(V_{in})| f_2(d_n(\gamma',\gamma))$ for all $\gamma',\gamma \in \Gamma_n$ almost surely, for a function $f_2(d)$ that converges to 0 as $d \to 0$ and a function f_1 such that $\sup_{n\geq 1} n^{-1}\sum_{i=1}^n E(|f_1(V_{in})|) < \infty$ *then*

$$\sup_{\gamma\in\Gamma_n}\left|\frac{1}{n}\sum_{i=1}^n h(V_{in},\gamma) - E(h(V_{in},\gamma))\right| \to_p 0$$

The proof of this follows as a variation in the proofs in Andrews (1992).

C Uniform convergence of kernel regression estimators

Consistent estimation of β_1 in the partially linear model in Section 3 requires uniform convergence of estimators of $E(W_i \mid \theta_i)$ for a random variable W_i. In this section, I provide three such results for the kernel regression estimator

$$\hat{g}_w(m) = \frac{\sum_{i=1}^n W_i K\left(\frac{\bar{M}_i - m}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\bar{M}_i - m}{h_n}\right)}$$

where $\bar{M}_i = \tilde{J}^{-1}\sum_{j=1}^{\tilde{J}} \tilde{M}_{ij}$. Dependence on $\tilde{J}$ is left implicit in the notation for $\bar{M}_i$ for convenience. Let $\bar{p}_{\tilde{J}}(\theta) = E(\bar{M}_i \mid \theta_i = \theta)$. The results in this section will be applicable for the case where $\tilde{J} = J$ and $\tilde{M}_{ij} = M_{ij}$ for each j but also cases where $\tilde{\mathbf{M}}_i := (\tilde{M}_{i1}, \ldots, \tilde{M}_{i\tilde{J}})$ is some subset of the full vector of J items, $\mathbf{M}_i$. A statement of the main results and the sufficient conditions are collected in the first subsection and proofs are all in a separate section below.

C.1 Assumptions and statement of convergence results

Before stating the main uniform convergence results for $\hat{g}_w(m)$ I first state two important results regarding the convergence of $\bar{M}_i$ to $\bar{p}_{\tilde{J}}(\theta_i)$ under the following assumption.

Assumption C.1.

(a) The binary random variables, $\tilde{M}_{i1},\ldots,\tilde{M}_{i\tilde{J}}$ are mutually independent conditional on θ_i

(b) $\exists J_0$ such that, for each $\tilde{J} \geq J_0$, $\bar{p}_{\tilde{J}}(t)$ is strictly increasing, continuous and differentiable at all $t \in \mathbb{R}$ with derivative $D\bar{p}_{\tilde{J}}(t)$ such that for each $t \in \mathbb{R}$, the family of functions $\{D\bar{p}_{\tilde{J}} : \tilde{J} \geq J_0\}$ is equicontinuous at t. Moreover, for each $t \in \mathbb{R}$, $\inf_{\tilde{J}\geq J_0} D\bar{p}_{\tilde{J}}(t) > 0$.

(c) θ has absolutely continuous distribution function F_θ and density f_θ that is continuous and satisfies $0 < f_\theta(t) \leq \bar{f}_\theta$ for all $t \in \Theta := support(\theta_i)$.

Lemma C.1. *Under Assumption C.1(a), if the sequence of random vectors* $\tilde{\mathbf{M}}_i = (\tilde{M}_{i1}, \ldots, \tilde{M}_{i\tilde{J}})$*,* $i = 1, \ldots, n$ *is i.i.d. for each* $\tilde{J}$ *then*

(a) for any $\varepsilon > 0$*,* $Pr(|\bar{M}_i - \bar{p}_{\tilde{J}}(\theta_i)| > \varepsilon) \leq 2\exp(-2\tilde{J}\varepsilon^2)$

(b) for any $\varepsilon > 0$*,* $Pr(\max_{1\leq i\leq n} |\bar{M}_i - \bar{p}_{\tilde{J}}(\theta_i)| > \varepsilon) \leq 2n\exp(-2\tilde{J}\varepsilon^2)$

(c) for any $s > 0$*,* $\sup_{\theta\in\Theta} E(|\bar{M}_i - \bar{p}_{\tilde{J}}(\theta_i)|^s \mid \theta_i = \theta) = O\left(\left(\tilde{J}^{-1}\log\tilde{J}\right)^{s/2}\right)$

The first two conclusions of this lemma are due to Douglas (2001). Theorem A.2 in Williams (2017) provides a similar result under a more general mixing condition in place of C.1(a). The proof is short but instructive.

Proof of Lemma C.1. First, (a) follows from Hoeffding's inequality since

$$\begin{aligned} Pr(|\bar{M}_i - \bar{p}_{\tilde{J}}(\theta_i)| > \varepsilon) &= \int Pr(|\bar{M}_i - \bar{p}_{\tilde{J}}(\theta_i)| > \varepsilon \mid \theta_i = \theta) f_\theta(\theta)d\theta \\ &\leq \int 2\exp(-2\tilde{J}\varepsilon^2) f_\theta(\theta)d\theta \end{aligned}$$

This then implies (b) since

$$\begin{aligned} Pr(\max_{1\leq i\leq n} |\bar{M}_i - \bar{p}_{\tilde{J}}(\theta_i)| > \varepsilon) &\leq \sum_{i=1}^n Pr(|\bar{M}_i - \bar{p}_{\tilde{J}}(\theta_i)| > \varepsilon) \\ &\leq 2n\exp(-2\tilde{J}\varepsilon^2) \end{aligned}$$

Let $\eta_i = \bar{M}_i - \bar{p}_{\tilde{J}}(\theta_i)$ and define a sequence $\rho_{\tilde{J}} = \left(\frac{s}{4}\tilde{J}^{-1}\log(\tilde{J})\right)^{1/2}$. Then

$$\begin{aligned} &\sup_\theta E(|\bar{M}_i - \bar{p}_{\tilde{J}}(\theta_i)|^s \mid \theta_i = \theta) \\ &\leq \sup_\theta E(|\eta_i|^s \mathbf{1}(|\eta_i|^s \leq \rho_{\tilde{J}}^s) \mid \theta_i = \theta) + E(|\eta_i|^s \mathbf{1}(|\eta_i|^s > \rho_{\tilde{J}}^s) \mid \theta_i = \theta) \\ &\leq \rho_{\tilde{J}}^s + Pr(|\eta_i|^s > \rho_{\tilde{J}}^s \mid \theta_i = \theta) \\ &= \rho_{\tilde{J}}^s + Pr(|\eta_i| > \rho_{\tilde{J}} \mid \theta_i = \theta) \\ &\leq \rho_{\tilde{J}}^s + 2\tilde{J}^{-s/2} \end{aligned}$$

where the final line follows from an application of Hoeffding's inequality and (c) then follows from the definition of $\rho_{\tilde{J}}$. □

In addition to Assumption C.1, I will impose additional regularity conditions and assumptions on the rate of convergence of the bandwidth sequence h_n and impose properties for the kernel function, K to derive asymptotic convergence of $\hat{g}_w$. The following conditions are used for the first result.

Assumption C.2.

(a) $W_i \perp\!\!\!\perp \tilde{\mathbf{M}}_i \mid \theta_i$

(b) *The function* $h_{w,0}(t) := E(W_i \mid \theta_i = t)$ *is continuous for all* $t \in \mathbb{R}$ *and is differentiable at all* $t \in \mathbb{R}$ *with derivative* $Dh_{w,0}(t)$ *that is also continuous at all* $t \in \mathbb{R}$.

(c) $E|e_i|^3 < \infty$ *and for any* $\delta > 0$, $\sup_{\theta \in \Theta_\delta} E(|e_i|^3 \mid \theta_i = \theta) < \infty$ *where* $e_i = W_i - h_{w,0}(\theta_i)$.

(d) K *is nonnegative and twice differentiable with continuous first and second derivatives* K' *and* K''. *All three functions* $K(u)$, $K'(u)$, *and* $K''(u)$ *are bounded by* $\bar{K}\mathbf{1}(|u| \leq 1)$, *and* $K(u) \geq \underline{K}\mathbf{1}(|u| \leq 1/2)$ *for all* $u \in \mathbb{R}$, *for constants* $0 < \underline{K} < \bar{K} < \infty$.

By Assumption C.1(b), for each $\tilde{J}$, the function $\bar{p}_{\tilde{J}}$ has an inverse which is well-defined on its range, which is an interval in $[0,1]$. The inverse $\bar{p}_{\tilde{J}}^{-1}(m)$ can be extended to $[0,1]$ by assigning the values $\inf \Theta$ and $\sup \Theta$ for values of m below and above this interval, respectively. Then define $\tilde{g}_w(m) = h_{w,0}(\bar{p}_{\tilde{J}}^{-1}(m))$.

Also, for a fixed $0 < \delta < 1/2$, let Θ_δ denote the interval $[q_\delta(\theta_i), q_{1-\delta}(\theta_i)]$ and define $\mathcal{M}_\delta = \bar{p}_J(\Theta_\delta) = \{m \in [0,1] : m = \bar{p}_{\tilde{J}}(\theta) \text{ for some } \theta \in \Theta_\delta\}$. Though it is suppressed in the notation, $\mathcal{M}_\delta$ varies with $\tilde{J}$.

Theorem C.1. *Under Assumptions C.1 and C.2, if* $\tilde{J}_n$ *is a sequence such that* $\tilde{J}_n = O(n^r)$ *and* $\tilde{J}_n^{-1} = O(n^{-r})$ *for some* $r > 0$, $h_n \to 0$, $nh_n^3 \to \infty$, *and* $(\tilde{J}_n^{-1}\log(\tilde{J}_n))^{1/2}h_n^{-1} = o(1)$ *then there exists a constant* $0 < B < \infty$ *such that*

(a) $\lim_{J\to\infty} \sup_{m\in\mathcal{M}_\delta} |\tilde{g}_w(m)| \leq B$ *and* $\lim_{J\to\infty} \sup_{m\in\mathcal{M}_\delta} |D\tilde{g}_w(m)| \leq B$

(b) $\lim_{n\to\infty} Pr(\sup_{m\in\mathcal{M}_\delta} \max_{i:|\bar{M}_i - m|\leq h} \sup_{t\in[0,1]} |D\tilde{g}_w(t\bar{p}_{\tilde{J}_n}(\theta_i) + (1-t)m)| > B) = 0$ *and* $\lim_{n\to\infty} Pr(\inf_{m\in\mathcal{M}_\delta} \sup_{t\in[0,1]} |D\bar{p}_{\tilde{J}_n}(\bar{p}_{\tilde{J}_n}^{-1}(t2h_n + (1-t)m))| > B^{-1}) = 0$

(c) $\lim_{n\to\infty} Pr(\sup_{m\in\mathcal{M}} \left|(nh_n)^{-1}\sum_{i=1}^n K\left(\frac{\bar{M}_i - m}{h_n}\right)\right| \leq B^{-1}) = 0$

(d) $\sup_{m\in\mathcal{M}_\delta} |\hat{g}_w(m) - \tilde{g}_w(m)| = O_p(h_n) + O_p(\log(n)(nh_n)^{-1/2})$

(e) $\lim_{n\to\infty} Pr(\sup_{m\in\mathcal{M}_\delta} |\hat{g}_w(m)| \leq B) = 1$

(f) $\lim_{n\to\infty} Pr(\sup_{m\in\mathcal{M}_\delta} |D\hat{g}_w(m)| \leq B) = 1$

The convergence rate in conclusion (d) of Theorem C.1 is not sufficient for $\sqrt{n}-$convergence of semiparametric estimators based on $\hat{g}$ because the convergence rate is not faster than $n^{-1/4}$ when $r \leq 1/2$, which is the case if $\sqrt{n}/\tilde{J}_n \to \gamma > 0$. This is because if $\tilde{J}_n = O(n^r)$ for $r \leq 1/2$ then the restriction $(\tilde{J}_n^{-1}\log(\tilde{J}_n))^{1/2}h_n^{-1} = o(1)$ implies that $h_n^{-1} = O(n^{r/2}) = O(n^{1/4})$ which implies that $O_p(h_n)$ is not $o_p(n^{-1/4})$. Fortunately, this convergence rate can be improved under the following assumption, which implies the conditions of Assumption C.2 but imposes several additional smoothness restrictions.

Assumption C.3.

(a) $W_i \perp\!\!\!\perp \tilde{\mathbf{M}}_i \mid \theta_i$.

(b) *The function* $h_{w,0}(t) := E(W_i \mid \theta_i = t)$ *is continuous for all* $t \in \mathbb{R}$ *and is twice differentiable at all* $t \in \mathbb{R}$ *with first and second derivatives* $Dh_{w,0}(t)$ *and* $D^2 h_{w,0}(t)$ *that are both continuous at all* $t \in \mathbb{R}$.

(c) $\exists J_0$ *such that, for each* $\tilde{J} \geq J_0$, $\bar{p}_{\tilde{J}}(t)$ *is twice differentiable at all* $t \in \mathbb{R}$ *with second derivative* $D^2\bar{p}_{\tilde{J}}(t)$ *such that for each* $t \in \mathbb{R}$, *the family of functions* $\{D^2\bar{p}_{\tilde{J}} : \tilde{J} \geq J_0\}$ *is equicontinuous at* t.

(d) *the density function* f_θ *is differentiable with derivative* $Df_\theta(t)$ *that is continuous at all* $t \in \mathbb{R}$.

(e) *For each* $s \in \mathbb{N}$, $2 \leq s < p$, *the function* $\omega_{s\tilde{J}}(t) = \tilde{J}^{s/2} E(\eta_i^s \mid \theta_i = t)$ *is differentiable with derivative* $D\omega_{s\tilde{J}}(t)$ *such that for each* $t \in \mathbb{R}$, *the family of functions* $\{\omega_{s\tilde{J}} : \tilde{J} \geq J_0\}$ *is equicontinuous at* t *and the family of functions* $\{D\omega_{s\tilde{J}} : \tilde{J} \geq J_0\}$ *is equicontinuous at* t.

(f) $E|e_i|^q < \infty$, $E|W_i|^q < \infty$ *and for any* $\delta > 0$, $\sup_{\theta \in \Theta_\delta} E(|e_i|^q \mid \theta_i = \theta) < \infty$ *and* $\sup_{\theta \in \Theta_\delta} E(|W_i|^q \mid \theta_i = \theta) < \infty$ *for some* $q \geq 3$, *where* $e_i = W_i - h_w(\theta_i)$.

(g) K *is nonnegative and* $p+1$*-times differentiable and, for* $0 \leq s \leq p+1$, $K^{(s)}(u)$ *is continuous for all* $u \in \mathbb{R}$, *where* $K^{(s)}(u) := \frac{d^s}{du^s}K(u)$. *Also, for each* $0 \leq s \leq p+1$, $|K^{(s)}(u)| \leq \bar{K}\mathbf{1}(|u| \leq 1)$ *and* $K(u) \geq \underline{K}\mathbf{1}(|u| \leq 1/2)$ *for all* $u \in \mathbb{R}$, *for constants* $0 < \underline{K} < \bar{K} < \infty$.

Theorem C.2. *Under Assumptions C.1 and C.3, if* $\tilde{J}_n$ *is a sequence such that* $\tilde{J}_n = O(n^r)$ *and* $\tilde{J}_n^{-1} = O(n^{-r})$ *for some* $r > 0$, $h_n \to 0$, $nh_n^3 \to \infty$, $(\tilde{J}_n^{-1}\log(\tilde{J}_n))^{1/2} h_n^{-1} = o(1)$ *then*

$$\sup_{m \in \mathcal{M}_\delta} |\hat{g}_w(m) - \tilde{g}_w(m)| = O_p\left(h_n^2 + \frac{\log(n)}{\sqrt{nh_n}} + \frac{\log(\tilde{J}_n)^{p/2}}{h_n^{(p-1)}\tilde{J}_n^{p/2}}\right).$$

C.2 Proofs

The proof of Theorems C.1 and C.2 both rely on the following lemma. This result is proved below using arguments that are standard in the literature (see, e.g., Hansen (2008)).

Lemma C.2. *Let* $\Delta_n^{Vsa}(m) = (nh_n)^{-1}\sum_{i=1}^n V_i \eta_i^s \kappa\left(\frac{\bar{M}_{i,J}^a \bar{p}_J(\theta_i)^{(1-a)} - m}{h_n}\right)$ *for an i.i.d random vector* $\{V_i\}_{i=1}^n$, *nonnegative integer* s, *and* $a \in \{0, 1\}$. *If* $(V_i, \theta_i, \tilde{\mathbf{M}}_i), i = 1, \ldots, n$ *is an i.i.d. random sequence, Assumption C.1 holds and, in addition,* *(a)* $V_i \perp\!\!\!\perp \tilde{\mathbf{M}}_i \mid \theta_i$, $E|V_i|^q < \infty$ *for some* $q > 2$, *and for any* $\delta > 0$, $\sup_{\theta \in \Theta_\delta} E(|V_i|^q \mid \theta_i = \theta) < \infty$.

(b) $|\kappa(u)| \leq B\mathbf{1}(|u| \leq 1)$ *and* κ *has a derivative,* κ' *which is continuous and is also bounded by* B.

(c) $\tilde{J}_n$ *is a sequence such that* $\tilde{J}_n = O(n^r)$ *and* $\tilde{J}_n^{-1} = O(n^{-r})$ *for some* $r > 0$, $h_n \to 0$, $h_n^{-1} = O(n^\alpha)$ *for some* $\alpha > 0$ *such that* $q(1-\alpha) > 2$, *and* $(\tilde{J}_n^{-1} \log(\tilde{J}_n))^{1/2} h_n^{-1} = o(1)$.

then

$$\sup_{m \in \mathcal{M}_\delta} |\Delta_n^{Vsa}(m) - E(\Delta_n^{Vsa}(m))| = O_p\left(\log(n)(nh_n)^{-1/2}\left(\tilde{J}_n^{-1}\log(\tilde{J}_n)\right)^{s/2}\right)$$

Moreover, if $\Delta_n^{Vs}(m) = (nh_n)^{-1}\sum_{i=1}^n V_i \int_{\bar{p}_J(\theta_i)}^{\bar{M}_i} (\bar{M}_i - t)^{s-1}\kappa\left(\frac{t-m}{h_n}\right) dt$ *then*

$$\sup_{m \in \mathcal{M}_\delta} |\Delta_n^{Vs}(m) - E(\Delta_n^{Vs}(m))| = O_p\left(\log(n)(nh_n)^{-1/2}\left(\tilde{J}_n^{-1}\log(\tilde{J}_n)\right)^{s/2}\right)$$

I now provide the proofs of the three main uniform convergence results. Where it is not necessary for understanding the notation is simplified by omitting the n subscript on $\tilde{J}_n$ and the $\tilde{J}$ subscript on $\bar{p}_{\tilde{J}}$.

Proof of Theorem C.1. (a) $sup_{m \in \mathcal{M}_\delta} |\tilde{g}_w(m)| \leq \sup_{\theta \in \Theta_\delta} |\mathbf{h}_0(\theta)|$, which is bounded since Θ_δ is compact and h_0 is continuous, by Assumption C.2(a). The function $\tilde{g}_w(m) = \mathbf{h}_0(\bar{p}^{-1}(m))$ is differentiable with $D\tilde{g}_w(m) = D\mathbf{h}_0(\bar{p}^{-1}(m))\frac{1}{D\bar{p}(\bar{p}^{-1}(m))}$ since $D\bar{p} > 0$ by Assumption C.2(b). Then

$$\sup_{m \in \mathcal{M}_\delta} |D\tilde{g}_w(m)| \leq \frac{\sup_{\theta \in \Theta_\delta} |D\mathbf{h}_0(\theta)|}{\inf_{\theta \in \Theta_\delta} |D\bar{p}(\theta)|}$$

By Assumption C.2(a), the function $\mathbf{h}_0$ is continuous and hence bounded on the compact set Θ_δ and by Assumption C.2(b) $\inf_{\theta \in \Theta_\delta} D\bar{p}(\theta)$ is bounded away from 0 as $\tilde{J} \to \infty$.

(b) Let the bound found in the proof of (a) above be $B/2$. If $\sup_{m \in \mathcal{M}_\delta} \max_{i:|\bar{M}_i - m| \leq h_n} \sup_{t \in [0,1]} D\tilde{g}_w(t\bar{p}(\theta_i) + (1-t)m) > B$ then there must be $m^* \in \mathcal{M}_\delta$ such that $\bar{M}_i - m^*| \leq h_n$ and $|D\tilde{g}_w(t\bar{p}(\theta_i) + (1-t)m^*) - D\tilde{g}_w(m^*)| > B/2$. By (a) and (b) of Assumption C.2, this implies that there is a constant $\epsilon > 0$ such that $|m^* - \bar{p}(\theta_i)| > \epsilon$. The result follows by Lemma C.1 and Assumption C.2(c) and Assumption C.2(f) since

$$Pr(\sup_{m \in \mathcal{M}_\delta} \max_{i:|\bar{M}_i - m| \leq h_n} |m - \bar{p}(\theta_i)| > \epsilon) \leq Pr(\max_{1 \leq i \leq n} |\bar{M}_i - \bar{p}(\theta_i)| \geq \epsilon - h_n) = o(1)$$

The second part follows from Assumption C.2(b) by a similar argument.

(c) First, define $\hat{f}_1(m) = (nh_n)^{-1} \sum_{i=1}^n K\left(\frac{\bar{M}_i - m}{h_n}\right)$. Then

$$
\begin{aligned}
|\hat{f}_1(m)| &\geq \underline{K}(nh_n)^{-1} \sum_{i=1}^n \mathbf{1}(|\bar{M}_i - m| \leq h_n/2) \\
&\geq \underline{K}(nh_n)^{-1} \sum_{i=1}^n \mathbf{1}(|\bar{p}(\theta_i) - m| \leq h_n/4) - \underline{K}(nh_n)^{-1} \sum_{i=1}^n \mathbf{1}(|\bar{M}_i - \bar{p}(\theta_i)| \leq h_n/4) \\
&\geq \underline{K}(nh_n)^{-1} \sum_{i=1}^n \mathbf{1}(|\bar{p}(\theta_i) - m| \leq h_n/4) - o(1)
\end{aligned}
$$

where the second inequality follows from Assumption C.2(d) and the last line follows from Lemma C.1 and Assumptions C.2(c) and (f) since $\tilde{J}h_n^2 = \tilde{J}\rho_n^2(h_n/\rho_n)^2 \to \infty$.

Then, $\mathbf{1}(|\bar{p}(\theta_i) - m| \leq h_n/4) = \mathbf{1}(\bar{p}(\theta_i) \leq m + h_n/4) - \mathbf{1}(\bar{p}(\theta_i) \leq m - h_n/4)$ so

$$
\begin{aligned}
&\inf_{m \in \mathcal{M}_\delta} \underline{K}(nh_n)^{-1} \sum_{i=1}^n \mathbf{1}(|\bar{p}(\theta_i) - m| \leq h_n/4) \\
&\geq \underline{K}(h_n)^{-1} \inf_{m \in \mathcal{M}_\delta} Pr(|\bar{p}(\theta_i) - m| \leq h_n/4) \\
&- 2\underline{K}(h_n)^{-1} \sup_{s \in [0,1]} \left(n^{-1} \sum_{i=1}^n \mathbf{1}(\bar{p}(\theta_i) \leq s) - Pr(\bar{p}(\theta_i) \leq s) \right)
\end{aligned}
$$

The second term is $O_p(h_n^{-1} n^{-1/2})$ by the DKW inequality (see, e.g., p. 268 of Van der Vaart, 2000) applied to $\sup_{s^* \in \mathbb{R}} \left(n^{-1} \sum_{i=1}^n \mathbf{1}(\theta_i \leq s^*) - Pr(\theta_i \leq s^*)\right)$

Finally, for n large enough, either $m + h_n/4 \in \mathcal{M}_\delta$ or $m - h_n/4 \in \mathcal{M}_\delta$, or both, so I will assume wlog that $m + h_n/4 \in \mathcal{M}_\delta$. Then $Pr(\bar{p}(\theta_i) \leq m + h_n/4) = F_\theta(\bar{p}^{-1}(m + h_n/4))$ and $Pr(\bar{p}(\theta_i) \leq m) = F_\theta(\bar{p}^{-1}(m))$, so

$$
\begin{aligned}
\underline{K}(h_n)^{-1} \inf_{m \in \mathcal{M}_\delta} Pr(|\bar{p}(\theta_i) - m| \leq h_n/4) &\geq \underline{K}(h_n)^{-1} \inf_{m \in \mathcal{M}_\delta} Pr(\bar{p}(\theta_i) \leq m + h_n/4) - Pr(\bar{p}(\theta_i) \leq m) \\
&= \underline{K}(h_n)^{-1} \inf_{m \in \mathcal{M}_\delta} \left(F_\theta(\bar{p}^{-1}(m + h_n/4)) - F_\theta(\bar{p}^{-1}(m))\right) \\
&\geq \frac{\underline{K} \inf_{\theta \in \Theta_\delta} f_\theta(\theta)}{\sup_{m \in \mathcal{M}_\delta} D\bar{p}(\bar{p}^{-1}(m))}
\end{aligned}
$$

which is bounded away from 0 by Assumptions C.2(b) and (e).

(d) First, $W_i = h_{w,0}(\theta_i) + e_i = \tilde{g}_w(\bar{p}(\theta_i)) + e_i = \tilde{g}_w(m) + \tilde{g}_w(\bar{p}(\theta_i)) - \tilde{g}_w(m) + e_i$ where

Assumption C.1(c) implies that $E(e_i \mid \theta_i, \mathbf{M}_i) = 0$. Then

$$|\hat{g}_w(m) - \tilde{g}_w(m)| \leq \frac{(nh_n)^{-1} \left| \sum_{i=1}^n (W_i - \tilde{g}_w(m)) K\left(\frac{\bar{M}_i - m}{h_n}\right) \right|}{|\hat{f}_1(m)|}$$

$$\leq |\hat{f}_1(m)|^{-1} \left| (nh_n)^{-1} \sum_{i=1}^n \left(\tilde{g}_w(\bar{p}(\theta_i)) - \tilde{g}_w(m)\right) K\left(\frac{\bar{M}_i - m}{h_n}\right) + (nh_n)^{-1} \sum_{i=1}^n e_i K\left(\frac{\bar{M}_i - m}{h_n}\right) \right|$$

Next,

$$\left| (nh_n)^{-1} \sum_{i=1}^n \left(\tilde{g}_w(\bar{p}(\theta_i)) - \tilde{g}_w(m)\right) K\left(\frac{\bar{M}_i - m}{h_n}\right) \right|$$
$$\leq \left(\sup_{m \in \hat{\mathcal{M}}_\delta} |D\tilde{g}_w(m)| \right) (nh_n)^{-1} \sum_{i=1}^n \left(|\bar{M}_i - \bar{p}(\theta_i)| + |\bar{M}_i - m| \right) K\left(\frac{\bar{M}_i - m}{h_n}\right)$$
$$\leq \left(\sup_{m \in \hat{\mathcal{M}}_\delta} |D\tilde{g}_w(m)| \right) \left\{ (nh_n)^{-1} \bar{K} \sum_{i=1}^n \mathbf{1}(|\bar{M}_i - \bar{p}(\theta_i)| > \rho_n) + (1 + h_n^{-1}\rho_n) n^{-1} \sum_{i=1}^n \mathbf{1}(|\bar{M}_i - m| \leq h_n) \right\}$$

where $\hat{\mathcal{M}}_\delta = \{t\bar{p}(\theta_i) + (1-t)m : m \in \mathcal{M}_\delta, t \in [0,1], |\bar{M}_i - m| < h_n\}$. The probability that the first term in braces is nonzero is bounded by $Pr(\max_{1 \leq i \leq n} |\bar{M}_i - \bar{p}(\theta_i)| > \rho_n)$ which is $o(n^{-1/2})$ by Assumption C.2(c), Assumption C.2(f), and Lemma C.1. Next, $(1+h_n^{-1}\rho_n) = 1 + o(1)$, $n^{-1}\sum_{i=1}^n \mathbf{1}(|\bar{M}_i - m| \leq h_n) = n^{-1}\sum_{i=1}^n \mathbf{1}(\bar{M}_i \leq m + h_n) - n^{-1}\sum_{i=1}^n \mathbf{1}(\bar{M}_i \geq m - h_n)$, and

$$n^{-1} \sum_{i=1}^n \mathbf{1}(\bar{M}_i \leq m + h_n) \leq n^{-1} \sum_{i=1}^n \mathbf{1}(\bar{p}(\theta_i) \leq m + 2h_n) + n^{-1} \sum_{i=1}^n \mathbf{1}(|\bar{M}_i - \bar{p}(\theta_i)| > h_n)$$

The second term is $o_p(n^{-1/2})$, again by Lemma C.1, since $h_n \geq \rho_n$, at least for n sufficiently large. Therefore,

$$\sup_{m \in \mathcal{M}_\delta} n^{-1} \sum_{i=1}^n \mathbf{1}(|\bar{M}_i - m| \leq h_n) \leq \sup_{m \in \mathcal{M}_\delta} |Pr(\bar{p}(\theta_i) \leq m + 2h_n) - Pr(\bar{p}(\theta_i) \leq m - 2h_n)|$$
$$+2 \sup_{s \in [0,1]} \left| n^{-1} \sum_{i=1}^n \mathbf{1}(\bar{p}(\theta_i) \leq s) - E(\mathbf{1}(\bar{p}(\theta_i) \leq s) \right| + o_p(n^{-1/2})$$

Here, the first term is bounded by $8h\bar{f}_\theta / (\inf_{m \in \hat{\mathcal{M}}_\delta} \inf_{t \in [0,1]} D\bar{p}(\bar{p}^{-1}(tm + (1-t)2h_n))$ by (b) and (e) of Assumption C.2. The second term is $o_p(n^{-1/2})$ (see proof of (c) above).

Next, since $E\left(e_i K\left(\frac{\bar{M}_i - m}{h_n}\right)\right) = 0$, it remains to show that

$$|\Delta_n(m) - E(\Delta_n(m))| = O_p(r_n)$$

where $\Delta_n(m) = (nh_n)^{-1}\sum_{i=1}^n e_i K\left(\frac{\bar{M}_i - m}{h_n}\right)$ and $r_n = \log(n)(nh_n)^{-1/2}$. This follows by applying Lemma C.2 with $\kappa = K$, $V_i = e_i$, $s = 0$, and $a = 1$. Conditions (a)-(c) of the lemma are implied by Assumption C.2.

(e) This follows from parts (a) and (d) of the lemma, which have already been proved.

(f) First,

$$D\hat{g}_w(m) = \frac{D\hat{f}_w(m)}{\hat{f}_1(m)} - \frac{D\hat{f}_1(m)\hat{f}_w(m)}{(\hat{f}_1(m))^2}$$

where $\hat{f}_w(m) = (nh_n)^{-1}\sum_{i=1}^n W_i K\left(\frac{\bar{M}_i - m}{h_n}\right)$. Using $W_i = \tilde{g}_w(m) + \tilde{g}_w(\bar{p}(\theta_i)) - \tilde{g}_w(m) + e_i$, and abbreviating $K_i = K\left(\frac{\bar{M}_i - m}{h_n}\right)$ and $K_i' = K'\left(\frac{\bar{M}_i - m}{h_n}\right)$, the same arguments used in the proof of (d) can be used to show that

$$|D\hat{g}_w(m)| \leq (|\hat{f}_1(m)|)^{-2}\left\{O_p(1)\left|(nh_n^2)^{-1}\sum_{i=1}^n e_i K_i'\right| + O_p(1)\left|(nh_n^2)^{-1}\sum_{i=1}^n e_i K_i\right| + O_p(1)\right\}$$

Then, since $E(e_i K_i) = E(e_i K_i') = 0$, Lemma C.2 can be applied to conclude that

$$\sup_m\left|(nh_n^2)^{-1}\sum_{i=1}^n e_i K_i\right| = h_n^{-1}\sup_m\left|(nh_n)^{-1}\sum_{i=1}^n e_i K_i\right| = O_p((nh_n^3)^{-1/2})$$
$$\sup_m\left|(nh_n^2)^{-1}\sum_{i=1}^n e_i K_i'\right| = h_n^{-1}\sup_m\left|(nh_n)^{-1}\sum_{i=1}^n e_i K_i'\right| = O_p((nh_n^3)^{-1/2})$$

Conditions (a)-(c) of the lemma are implied Assumption C.2. This then implies by part(c) that $\sup_m |D\hat{g}_w(m)| = O_p(1) + O_p((nh_n^3)^{-1/2})$. The desired result follows since $nh_n^3 \to \infty$.

□

Proof of Theorem C.2. First, define $\tilde{e}_i = W_i - \tilde{g}_w(m) = \tilde{g}_w(\bar{p}(\theta_i)) - \tilde{g}_w(m) + e_i$. Then

$$\begin{aligned}|\hat{g}_w(m) - \tilde{g}_w(m)| &\leq |\hat{f}_1(m)|^{-1} \left|(nh_n)^{-1}\sum_{i=1}^{n} \tilde{e}_i K\left(\frac{\bar{M}_i - m}{h}\right)\right| \\ &\leq |\hat{f}_1(m)|^{-1} \left|(nh_n)^{-1}\sum_{i=1}^{n} \tilde{e}_i K\left(\frac{\bar{p}(\theta_i) - m}{h}\right)\right| \\ &+ |\hat{f}_1(m)|^{-1} \left|(nh_n)^{-1}\sum_{i=1}^{n} \tilde{e}_i \left\{K\left(\frac{\bar{M}_i - m}{h}\right) - K\left(\frac{\bar{p}(\theta_i) - m}{h}\right)\right\}\right|\end{aligned}$$

where $\hat{f}_1(m) = (nh_n)^{-1}\sum_{i=1}^{n} K\left(\frac{\bar{M}_i - m}{h_n}\right)$. Since Assumption C.3 implies Assumption C.2, $\sup_{m\in\mathcal{M}_\delta} |\hat{f}_1(m)|^{-1} = O_p(1)$ follows from conclusion (c) of Theorem C.1.

Under condition (f) of Assumption C.3, by a p^{th} order Taylor series expansion, $K(u') - K(u) = \sum_{s=1}^{p-1} \frac{(u'-u)^s}{s!} K^{(s)}(u) + \int_u^{u'} \frac{(u'-t)^{p-1}}{p!} K^{(p)}(t)dt$. Therefore,

$$\begin{aligned}|\hat{g}_w(m) - \tilde{g}_w(m)| \leq |\hat{f}_1(m)|^{-1} \sum_{s=0}^{p-1} (nh_n)^{-1} \left|\sum_{i=1}^{n} \tilde{e}_i \frac{\eta_i^s}{h_n^s s!} K^{(s)}\left(\frac{\bar{p}(\theta_i) - m}{h}\right)\right| \\ + |\hat{f}_1(m)|^{-1}(nh_n)^{-1} \left|\sum_{i=1}^{n} \tilde{e}_i \int_{\bar{p}(\theta_i)}^{\bar{M}_i} \frac{(\bar{M}_i - t)^{p-1}}{p! h_n^p} K^{(p)}\left(\frac{t - m}{h}\right) dt\right|\end{aligned}$$

Since $\tilde{e}_i = W_i - \tilde{g}_w(m)$, for each $0 \leq s < p$,

$$\begin{aligned}&\sup_{m\in\mathcal{M}_\delta} (nh_n)^{-1} \left|\sum_{i=1}^{n} \tilde{e}_i \frac{\eta_i^s}{h_n^s s!} K^{(s)}\left(\frac{\bar{p}(\theta_i) - m}{h}\right)\right| \qquad \text{(C.1)} \\ &\leq \sup_{m\in\mathcal{M}_\delta} (nh_n)^{-1} \left|\sum_{i=1}^{n} \left\{W_i \frac{\eta_i^s}{h_n^s s!} K^{(s)}\left(\frac{\bar{p}(\theta_i) - m}{h}\right) - E\left(W_i \frac{\eta_i^s}{h_n^s s!} K^{(s)}\left(\frac{\bar{p}(\theta_i) - m}{h}\right)\right)\right\}\right| \\ &+ \left(\sup_{m\in\mathcal{M}_\delta} |\tilde{g}_w(m)|\right) \sup_{m\in\mathcal{M}_\delta} (nh_n)^{-1} \left|\sum_{i=1}^{n} \left\{\frac{\eta_i^s}{h_n^s s!} K^{(s)}\left(\frac{\bar{p}(\theta_i) - m}{h}\right) - E\left(\frac{\eta_i^s}{h_n^s s!} K^{(s)}\left(\frac{\bar{p}(\theta_i) - m}{h}\right)\right)\right\}\right| \\ &+ (nh_n)^{-1} \left|\sum_{i=1}^{n} E\left(\tilde{e}_i \frac{\eta_i^s}{h_n^s s!} K^{(s)}\left(\frac{\bar{p}(\theta_i) - m}{h}\right)\right)\right|\end{aligned}$$

By application of Lemma C.2, first with $V_i = W_i$ and second with $V_i = 1$, each of the first two terms is $O_p\left(\log(\tilde{J}_n)^{s/2} h_n^{-s} \tilde{J}_n^{-s/2} \frac{\log(n)}{\sqrt{nh_n}}\right)$. In addition, I show below that

$$\sup_{m\in\mathcal{M}_\delta} (nh_n)^{-1} \left|\sum_{i=1}^{n} E\left(\tilde{e}_i \frac{\eta_i^s}{h_n^s s!} K^{(s)}\left(\frac{\bar{p}(\theta_i) - m}{h}\right)\right)\right| = O\left(h_n^{-(s-2)} \tilde{J}_n^{-s/2}\right) \qquad \text{(C.2)}$$

Next, since $E(e_i \mid \bar{M}_i, \theta_i) = E(e_i \mid \theta_i) = 0$,

$$
\begin{aligned}
&(nh_n)^{-1}\left|\sum_{i=1}^{n} \tilde{e}_i \int_{\bar{p}(\theta_i)}^{\bar{M}_i} \frac{(\bar{M}_i - t)^{p-1}}{p!h_n^p} K^{(p)}\left(\frac{t-m}{h}\right) dt\right| \\
&\leq (nh_n)^{-1}\left|\sum_{i=1}^{n}\left\{e_i \int_{\bar{p}(\theta_i)}^{\bar{M}_i} \frac{(\bar{M}_i - t)^{p-1}}{p!h_n^p} K^{(p)}\left(\frac{t-m}{h}\right) dt - E\left(e_i \int_{\bar{p}(\theta_i)}^{\bar{M}_i} \frac{(\bar{M}_i - t)^{p-1}}{p!h_n^p} K^{(p)}\left(\frac{t-m}{h}\right) dt\right)\right\}\right| \\
&+ (nh_n)^{-1}\left|\sum_{i=1}^{n} (\tilde{g}_w(\bar{p}(\theta_i)) - \tilde{g}_w(m)) \int_{\bar{p}(\theta_i)}^{\bar{M}_i} \frac{(\bar{M}_i - t)^{p-1}}{p!h_n^p} K^{(p)}\left(\frac{t-m}{h}\right) dt\right|
\end{aligned}
$$

Another application of Lemma C.2, this time with $V_i = e_i$, implies that the first term is $O_p\left(\log(\tilde{J})^{p/2}h_n^{-p}\tilde{J}^{-p/2}\frac{\mathrm{lo}}{\sqrt{}}\right.$
Lastly, I will show that

$$
\sup_{m\in\mathcal{M}_\delta} (nh_n)^{-1}\left|\sum_{i=1}^{n} (\tilde{g}_w(\bar{p}(\theta_i)) - \tilde{g}_w(m)) \int_{\bar{p}(\theta_i)}^{\bar{M}_i} \frac{(\bar{M}_i - t)^{p-1}}{p!h_n^p} K^{(p)}\left(\frac{t-m}{h_n}\right) dt\right| = O_p\left(\log(\tilde{J}_n)^{p/2}h_n^{-p}\tilde{J}_n^{-p/2}h_n\right) \tag{C.3}
$$

Then, since $h_n^{-s}\tilde{J}_n^{-s/2} = o(\log(\tilde{J}_n)^{s/2}h_n^{-s}\tilde{J}_n^{-s/2})$ and $\log(\tilde{J}_n)^{s/2}h_n^{-s}\tilde{J}_n^{-s/2} = O(1)$ for any $s \geq 0$,

$$
\begin{aligned}
\sup_{m\in\mathcal{M}_\delta} |\hat{g}_w(m) - \tilde{g}_w(m)| &= O_p(1)O_p\left(\log(\tilde{J}_n)^{p/2}h_n^{-(p-1)}\tilde{J}_n^{-p/2} + \log(\tilde{J}_n)^{p/2}h_n^{-p}\tilde{J}_n^{-p/2}\frac{\log(n)}{\sqrt{nh_n}}\right. \\
&\qquad \left. + \sum_{s=0}^{p-1}\left\{h_n^{-(s-2)}\tilde{J}_n^{-s/2} + \log(\tilde{J}_n)^{s/2}h_n^{-s}\tilde{J}_n^{-s/2}\frac{\log(n)}{\sqrt{nh_n}}\right\}\right) \\
&= O_p\left(h_n^2 + \frac{\log(n)}{\sqrt{nh_n}} + \log(\tilde{J}_n)^{p/2}h_n^{-(p-1)}\tilde{J}_n^{-p/2}\right)
\end{aligned}
$$

Thus, it remains to prove (C.2) and (C.3). First, $E\left(e_i\eta_i^s K^{(s)}\left(\frac{\bar{p}(\theta_i)-m}{h_n}\right)\right) = 0$ so

$$
\begin{aligned}
&E\left(\tilde{e}_i\eta_i^s K^{(s)}\left(\frac{\bar{p}(\theta_i)-m}{h_n}\right)\right) = E\left((\tilde{g}_w(\bar{p}(\theta_i)) - \tilde{g}_w(m))\eta_i^s K^{(s)}\left(\frac{\bar{p}(\theta_i)-m}{h_n}\right)\right) \\
&= \tilde{J}_n^{-s/2}\int (\tilde{g}_w(\bar{p}(\theta)) - \tilde{g}_w(m))\tilde{\omega}_{s\tilde{J}}(\bar{p}(\theta))K^{(s)}\left(\frac{\bar{p}(\theta)-m}{h_n}\right)\tilde{f}_{\theta_i}(\bar{p}(\theta))d\theta \\
&= \tilde{J}_n^{-s/2}h_n\int (\tilde{g}_w(m+uh_n) - \tilde{g}_w(m))\tilde{\omega}_{s\tilde{J}}(m+uh_n)K^{(s)}(u)\,\tilde{f}_{\theta_i}(m+uh_n)d\theta
\end{aligned}
$$

where $\omega_{s\tilde{J}}(\theta) = \tilde{J}^{s/2}E(\eta_i^s \mid \theta_i = \theta)$, $\tilde{\omega}_{s\tilde{J}}(m) = \omega_{s\tilde{J}}(\bar{p}^{-1}(m))$, and $\tilde{f}_{\theta_i}(m) = \frac{f_\theta(\bar{p}^{-1}(m))}{D\bar{p}(\bar{p}^{-1}(m)}$ and the last line follows from the substitution $u = (\bar{p}(\theta) - m)/h_n$.

Next, use three Taylor approximations: $\tilde{g}_w(m^*) - \tilde{g}_w(m) = D\tilde{g}_w(m)(m^*-m) + \frac{1}{2}D^2\tilde{g}_w(m_a)(m^*-m)^2$, $\omega_{s\tilde{J}}(\bar{p}^{-1}(m^*)) = \omega_{s\tilde{J}}(\bar{p}^{-1}(m)) + D\tilde{\omega}_{s\tilde{J}}(m_b)(m^*-m)$, and $\tilde{f}_{\theta_i}(m^*) = \tilde{f}_{\theta_i}(m) + D\tilde{f}_{\theta_i}(m_c)(m^*-$

m) where m_a, m_b and m_c are all between m and m^*. Take n sufficiently large so that m and $m^* = m + uh_n$ are both contained in $\mathcal{M}_{\delta/2}$. Then, by the previous equation

$$\begin{aligned}
&\tilde{J}_n^{s/2} \sup_{m\in\mathcal{M}_\delta} \left| E\left(\tilde{e}_i \eta_i^s K^{(s)}\left(\frac{\bar{p}(\theta_i) - m}{h}\right)\right)\right| \\
&= h_n^2 \sup_{m\in\mathcal{M}_\delta} \left| D\tilde{g}_w(m)\tilde{\omega}_{s\tilde{J}}(m)\frac{\tilde{f}_{\theta_i}(m)}{D\bar{p}(\bar{p}^{-1}(m))}\right| \int uK^{(s)}(u)\,du + O(h_n^3)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sup_{m\in\mathcal{M}_\delta} (nh_n)^{-1} \left| \sum_{i=1}^n E\left(\tilde{e}_i \frac{\eta_i^s}{h_n^s s!} K^{(s)}\left(\frac{\bar{p}(\theta_i) - m}{h}\right)\right)\right| \\
&\le \tilde{J}_n^{-s/2} \frac{1}{h_n^{s+1} s!} O(h_n^3) = O\left(h_n^{-(s-2)} \tilde{J}_n^{-s/2}\right)
\end{aligned}$$

To prove (C.2), first observe that I can take $|\eta_i| \le \delta_n := \left(c_0 \tilde{J}_n^{-1} \log(\tilde{J}_n)\right)^{1/2}$ for each i, where $2rc_0 > 1$, by Lemma C.1. Also, take n sufficiently large so that $\delta_n \le h_n$. Then

$$\begin{aligned}
&\sup_{m\in\mathcal{M}_\delta} (nh_n)^{-1} \left| \sum_{i=1}^n \left(\tilde{g}_w(\bar{p}(\theta_i)) - \tilde{g}_w(m)\right) \int_{\bar{p}(\theta_i)}^{\bar{M}_i} \frac{(\bar{M}_i - t)^{p-1}}{p! h_n^p} K^{(p)}\left(\frac{t-m}{h_n}\right) dt \right| \\
&\le \left(\sup_{m\in\mathcal{M}_{\delta/2}} |D\tilde{g}_w(m)|\right) \sup_{m\in\mathcal{M}_\delta} \frac{\bar{K}\delta_n^p}{p! h_n^p} n^{-1} \sum_{i=1}^n \mathbf{1}(|\bar{p}(\theta_i) - m| \le h_n) \\
&= O_p\left(\frac{\delta_n^p}{h_n^{p-1}}\right)
\end{aligned}$$

The final line follows because, as argued in the proof of Theorem C.1 using the DKW inequality (see, e.g., p. 268 of Van der Vaart, 2000) and the fact that $Pr(|\bar{p}(\theta_i) - m| \le h_n) = O(h_n)$,

$$n^{-1} \sum_{i=1}^n \mathbf{1}(|\bar{p}(\theta_i) - m| \le h_n) = O(h_n) + O_p(n^{-1/2}) = O_p(h_n)$$

□

Now a proof of Lemma C.2 is now provided.

Proof of Lemma C.2. Let $r_n := \log(n)(nh_n)^{-1/2}\left(\tilde{J}_n^{-1}\log(\tilde{J}_n)\right)^{s/2}$. Define b_n such that $b_n^q = n\log(n)$ and let $\bar{V}_{in} = V_i \mathbf{1}(|V_i| \le b_n)$ and $\bar{\eta}_{in} = \eta_i \mathbf{1}(|\eta_i| \le \rho_n)$ where $\rho_n = (r^{-1}\tilde{J}_n^{-1}\log(\tilde{J}_n))^{1/2}$ for

r given by condition (c) of the lemma. Let

$$\begin{aligned}
\Delta_n^{Vsa,r}(m) &= (nh_n)^{-1}\sum_{i=1}^n V_i\eta_i^s\kappa\left(\frac{\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)}-m}{h_n}\right) - (nh_n)^{-1}\sum_{i=1}^n \bar{V}_{in}\bar{\eta}_{in}^s\kappa\left(\frac{\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)}-m}{h_n}\right)\\
&= (nh_n)^{-1}\sum_{i=1}^n V_i\mathbf{1}(|V_i|>b_n)\eta_i^s\kappa\left(\frac{\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)}-m}{h_n}\right)\\
&+ (nh_n)^{-1}\sum_{i=1}^n \bar{V}_{in}\eta_i^s\mathbf{1}(|\eta_i|>\rho_n)\kappa\left(\frac{\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)}-m}{h_n}\right)\\
&= \Delta_n^{Vsa,r1}(m)+\Delta_n^{Vsa,r2}(m)
\end{aligned}$$

Then $|\Delta_n^{Vsa}(m)-E(\Delta_n^{Vsa}(m))|\leq|\bar{\Delta}_n^{Vsa}(m)|+|\Delta_n^{Vsa,r}(m)|+|E(\Delta_n^{Vsa,r}(m))|$ where

$$\bar{\Delta}_n^{Vsa}(m) = (nh_n)^{-1}\sum_{i=1}^n\left(\bar{V}_{in}\bar{\eta}_{in}^s\kappa\left(\frac{\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)}-m}{h_n}\right) - E\left(\bar{V}_{in}\bar{\eta}_{in}^s\kappa\left(\frac{\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)}-m}{h_n}\right)\right)\right)$$

First, for any $t>0$,

$$Pr(\sup_{m\in\mathcal{M}}|\Delta_n^{Vsa,r}(m)|>tr_n)\leq Pr(\max_{1\leq i\leq n}|V_i|>b_n)+Pr(\max_{1\leq i\leq n}|\eta_i|>\rho_n)$$

Then $Pr(\max_{1\leq i\leq n}|V_i|>b_n)\leq nPr(|V_i|>b_n)\leq\frac{E(|V_i|^q)}{\log(n)}\to 0$, where the last inequality follows from Markov's inequality since $\frac{n}{b_n^q}=\frac{1}{\log(n)}$, and the limit holds by condition (a). And $Pr(\max_{1\leq i\leq n}|\eta_i|>\rho_n)=o(n^{-1})$ by Lemma C.1 and condition (c).

Second, by condition (b)

$$\begin{aligned}
\sup_{m\in\mathcal{M}_\delta}|E(\Delta_n^{Vsa,r1}(m))| &\leq h_n^{-1}\sup_{m\in\mathcal{M}_\delta}E\left(|V_i|\mathbf{1}(|V_i|>b_n)|\eta_i|^s\mathbf{1}(|\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)}-m|\leq h_n|)\right)\\
&\leq h_n^{-1}\sup_{m\in\mathcal{M}_\delta}E\left(|V_i|\mathbf{1}(|V_i|>b_n)|\eta_i|^s\mathbf{1}(|\bar{p}(\theta_i)-m|\leq 2h_n|)\right) \qquad \text{(C.4)}\\
&+\left\{h_n^{-1}E\left(|V_i|\mathbf{1}(|V_i|>b_n)|\eta_i|^s\mathbf{1}(|\bar{M}_i-\bar{p}(\theta_i)|>h_n|)\right)\right\}^a
\end{aligned}$$

For n sufficiently large, the first term satisfies

$$
\begin{aligned}
& h_n^{-1} \sup_{m\in\mathcal{M}_\delta} E\left(|V_i|\mathbf{1}(|V_i|>b_n)|\eta_i|^s\mathbf{1}(|\bar{p}(\theta_i)-m|\le 2h_n|)\right) \\
& \le h_n^{-1}\left(\sup_{\theta\in\Theta_{\delta/2}} |E(|\eta_i|^s \mid \theta_i=\theta)|\right) E\left(|V_i|\mathbf{1}(|V_i|>b_n)\mathbf{1}(|\bar{p}(\theta_i)-m|\le 2h_n)\right) \\
& \le h_n^{-1}\frac{1}{b_n^{q-1}}\left(\sup_{\theta\in\Theta_{\delta/2}} |E(|\eta_i|^s \mid \theta_i=\theta)|\right) E\left(|V_i|^q\mathbf{1}(|\bar{p}(\theta_i)-m|\le 2h_n)\right) \\
& \le h_n^{-1}\frac{1}{b_n^{q-1}}\left(\sup_{\theta\in\Theta_{\delta/2}} E(|\eta_i|^s \mid \theta_i=\theta)\right)\left(\sup_{\theta\in\Theta_{\delta/2}} E(|V_i|^q \mid \theta_i=\theta)\right) Pr(|\bar{p}(\theta_i)-m|\le 2h_n)
\end{aligned}
$$

where the first inequality is by the conditional independence between V_i and $\bar{M}_i$ conditional on θ_i under condition (a), the second is because $|V_i| > b_n$ implies that $|V_i| \le b_n^{q-1}|V_i|^q$, and the third is valid under condition (a). This term is $O_p(b_n^{-(q-1)}J_n^{-s/2})$ since $Pr(|\bar{p}(\theta_i)-m|\le 2h_n) \le Pr\left(|\theta_i - \bar{p}^{-1}(m)| \le 2h_n/\inf_{\theta\in\Theta_{\delta/2}} D\bar{p}(\theta)\right) \le 4\sup_{\theta\in\Theta} f_\theta(\theta)h_n/\inf_{\theta\in\Theta_{\delta/2}} D\bar{p}(\theta)$ and because $\sup_{\theta\in\Theta_{\delta/2}} E(|\eta_i|^s \mid \theta_i=\theta) = O\left(\left(\tilde{J}_n^{-1}\log\tilde{J}_n\right)^{s/2}\right)$ by Lemma C.1. Lastly, it is easy to verify that $q>2$ implies that $O_p\left(b_n^{-(q-1)}\left(\tilde{J}^{-1}\log\tilde{J}\right)^{s/2}\right) = o_p(r_n)$ because $b_n^{q-1} > n^{(q-1)/q} > n^{1/2}$.

For any $t>0$, the second term in (C.4), for sufficiently large n, satisfies

$$
\begin{aligned}
& h_n^{-1}E\left(|V_i|\mathbf{1}(|V_i|>b_n)|\eta_i|^s\mathbf{1}(|\bar{M}_i-\bar{p}(\theta_i)|>h_n|)\right) \\
& \le h_n^{-1}E\left(|V_i|\mathbf{1}(|V_i|>b_n)|\eta_i|^s\mathbf{1}(|\bar{M}_i-\bar{p}(\theta_i)|>t\rho_n|)\right) \\
& \le h_n^{-1}E\left(|V_i|^2\right)^{1/2} Pr(|\bar{M}_i-\bar{p}(\theta_i)|>t\rho_n|)^{1/2} \\
& \le 2h_n^{-1}E\left(|V_i|^2\right)^{1/2}\exp(-J_nt^2\rho_n^2)
\end{aligned}
$$

where the first inequality follows from condition (c) in the statement of the lemma, the second follows from the Cauchy-Schwarz inequality and the fact that $|\eta_i| \le 1$, the third follows from Hoeffding's inequality. This term is $o_p(r_n)$ if $t^2 > \frac{1+\alpha+sr}{2}$ because $E\left(|V_i|^2\right) < \infty$ by condition (a), and condition (d) of Assumption C.1 and condition (c) imply that $h_n^{-1}r_n^{-1} = O(n^{\frac{1}{2}(1+\alpha+sr)}\log(n)^{-1})$ for some $\alpha>0$ and $\exp(-2J_nt^2\rho_n^2) \le J_n^{-a^{-1}t^2} = o(n^{-t^2})$.

By Lemma C.1, condition (d) of Assumption C.1 and condition (c), and applying the same argument based on Hoeffding's inequality,

$$
\begin{aligned}
\sup_{m\in\mathcal{M}_\delta}|E(\Delta_n^{Vsa,r2}(m))| &\le h_n^{-1}\sup_{m\in\mathcal{M}_\delta} E\left(|\bar{V}_{in}|\eta_i^s\mathbf{1}(|\eta_i|>\rho_n)|\mathbf{1}(|\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)}-m|\le h_n|)\right) \\
&\le h_n^{-1}b_nPr(|\eta_i|>\rho_n) = o(r_n)
\end{aligned}
$$

Third, since $|\bar{V}_{in}\bar{\eta}_{in}^{s}K\left(\frac{\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)}-m}{h_n}\right)| \leq b_n\rho_n^s B$, I can apply Bernstein's inequality:

$$
\begin{aligned}
Pr(|\bar{\Delta}_n^{Vsa}(m)| > tr_n) &\leq \exp\left(-\frac{(tr_n nh_n)^2}{2nVar\left(\bar{V}_{in}\bar{\eta}_{in}^{s}K\left(\frac{\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)}-m}{h_n}\right)\right) + \frac{4}{3}tBb_n\rho_n^s r_n nh_n}\right) \\
&\leq \exp\left(-\frac{t^2\log(n)}{c_1 + c_2 tb_n\log(n)(nh_n)^{-1/2}}\right) \qquad \text{(C.5)}
\end{aligned}
$$

where the second inequality follows for some positive constants c_1, c_2 because $(r_n nh_n)^2 = O\left(\log(n)nh_n\left(\tilde{J}_n^{-1}\log\right.\right.$ and

$$
\begin{aligned}
&Var\left(\bar{V}_{in}\bar{\eta}_{in}^{s}K\left(\frac{\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)}-m}{h_n}\right)\right) \\
&\leq E\left(\bar{V}_{in}^2\bar{\eta}_{in}^{2s}\mathbf{1}(|\bar{p}(\theta_i)-m| \leq 2h_n)\right) + E\left(\bar{V}_{in}^2\bar{\eta}_{in}^{2s}\mathbf{1}(|\bar{M}_i - \bar{p}(\theta_i)| \geq h_n)\right) \\
&\leq \rho_n^{2s}\left(\sup_{\theta\in\Theta_{\delta/2}} E(V_i^2 \mid \theta_i = \theta)\right) Pr(|\bar{p}(\theta_i)-m| \leq 2h_n) + b_n^2 Pr\left(|\bar{M}_i - \bar{p}(\theta_i)| \geq h_n\right) \\
&= O\left(\left(\tilde{J}_n^{-1}\log(\tilde{J}_n)\right)^s h_n\right)
\end{aligned}
$$

where the last line follows because, using the same argument based on Hoeffding's inequality $b_n^2 Pr\left(|\bar{M}_i - \bar{p}(\theta_i)| \geq h_n\right) = o(n^{-C})$ for any $C > 0$.

Next, partition $\mathcal{M}_\delta$ into $N \leq \frac{1}{r_n h_n}$ intervals of width $r_n h_n$ centered at $\{m_j\}_{j=1}^N$. Since $\left|K\left(\frac{\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)}-m_j}{h_n}\right) - K\right.$ $\frac{|m_j-m|}{h_n}\mathbf{1}(|\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)} - m_j| \leq 2h_n)$, for n sufficiently large, following an argument due to Hansen (2008),

$$
Pr(\sup_{m\in\mathcal{M}_\delta}|\bar{\Delta}_n^{Vsa}(m)| > 3tr_n) \leq NPr(|\bar{\Delta}_n^{Vsa}(m)| > tr_n) + NPr(|\bar{\Delta}_n^{*Vsa}(m)| > tr_n)
$$

provided that $E|\bar{\Delta}_n^{*Vsa}(m)|$ is bounded, where

$$
\begin{aligned}
\bar{\Delta}_n^{*Vsa}(m) = (nh_n)^{-1}\sum_{i=1}^{n}&\left(\bar{V}_{in}\bar{\eta}_{in}^{s}\mathbf{1}(|\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)} - m| \leq h_n)\right. \\
&\left.-E\left(\bar{V}_{in}\bar{\eta}_{in}^{s}\mathbf{1}(|\bar{M}_i^a\bar{p}(\theta_i)^{(1-a)} - m| \leq h_n)\right)\right)
\end{aligned}
$$

The same arguments used above can be repeated to show that $E|\bar{\Delta}_n^{*Vsa}(m)|$ is bounded uniformly over $m \in \mathcal{M}_\delta$ and that the bound on $Pr(|\bar{\Delta}_n^{Vsa}(m)| > tr_n)$ derived in equation (C.5) applies to $NPr(|\bar{\Delta}_n^{*Vsa}(m)| > tr_n)$ as well. Therefore,for t large enough

$$
Pr(\sup_{m\in\mathcal{M}_\delta}|\bar{\Delta}_n^{*Vsa}(m)| > 3tr_n) \leq 2N\exp\left(-\frac{t^2\log(n)}{c_1 + c_2 tb_n\log(n)(nh_n)^{-1/2}}\right) \to 0
$$

where convergence follows because $n^{q^{-1}}(nh_n)^{-1/2} = O(n^{-(\frac{1}{2}-\frac{1}{q}-\frac{\alpha}{2})})$, which implies that $b_n \log(n)(nh_n)^{-1/2} = o(1)$ if $\frac{1}{2}-\frac{1}{q}-\frac{\alpha}{2} > 0$. The latter follows from the restriction in condition (c) that $q(1-\alpha) > 2$.

The result for $\Delta_n^{Vs}(m)$ follows by essentially the same argument. Let

$$\begin{aligned}
\Delta_n^{Vs,r}(m) &= \Delta_n^{Vs}(m) - (nh_n)^{-1}\sum_{i=1}^n \bar{V}_{in}\mathbf{1}(|\eta_i| \leq \rho_n)\int_{\bar{p}(\theta_i)}^{\bar{M}_i}(\bar{M}_i - t)^{s-1}\kappa\left(\frac{t-m}{h_n}\right)dt \\
&= (nh_n)^{-1}\sum_{i=1}^n V_i\mathbf{1}(|V_i| > b_n)\int_{\bar{p}(\theta_i)}^{\bar{M}_i}(\bar{M}_i - t)^{s-1}\kappa\left(\frac{t-m}{h_n}\right)dt \\
&+ (nh_n)^{-1}\sum_{i=1}^n \bar{V}_{in}\mathbf{1}(|\eta_i| > \rho_n)\int_{\bar{p}(\theta_i)}^{\bar{M}_i}(\bar{M}_i - t)^{s-1}\kappa\left(\frac{t-m}{h_n}\right)dt \\
&= \Delta_n^{Vs,r1}(m) + \Delta_n^{Vs,r2}(m)
\end{aligned}$$

Then $|\Delta_n^{Vs}(m) - E(\Delta_n^{Vs}(m))| \leq |\bar{\Delta}_n^{Vs}(m)| + |\Delta_n^{Vs,r}(m)| + |E(\Delta_n^{Vs,r}(m))|$ where

$$\begin{aligned}
\bar{\Delta}_n^{Vs}(m) = (nh_n)^{-1}\sum_{i=1}^n \Bigg(&\bar{V}_{in}\mathbf{1}(|\eta_i| \leq \rho_n)\int_{\bar{p}(\theta_i)}^{\bar{M}_i}(\bar{M}_i - t)^{s-1}\kappa\left(\frac{t-m}{h_n}\right)dt \\
&-E\left(\bar{V}_{in}\mathbf{1}(|\eta_i| \leq \rho_n)\int_{\bar{p}(\theta_i)}^{\bar{M}_i}(\bar{M}_i - t)^{s-1}\kappa\left(\frac{t-m}{h_n}\right)dt\right)\Bigg)
\end{aligned}$$

First, for any $t > 0$,

$$Pr(\sup_{m\in\mathcal{M}_\delta}|\Delta_n^{Vs,r}(m)| > tr_n) \leq Pr(\max_{1\leq i\leq n}|V_i| > b_n) + Pr(\max_{1\leq i\leq n}|\eta_i| > \rho_n) \to 0$$

Second, by condition (b),

$$\left|\int_{\bar{p}(\theta_i)}^{\bar{M}_i}(\bar{M}_i - t)^{s-1}\kappa\left(\frac{t-m}{h_n}\right)dt\right| \leq |\eta_i|^{s-1}B\int_{\bar{p}(\theta_i)}^{\bar{M}_i}\mathbf{1}(|t-m| \leq h_n)dt$$

which implies that

$$\begin{aligned}
\sup_{m\in\mathcal{M}_\delta}|E(\Delta_n^{Vsa,r1}(m))| \leq h_n^{-1}&\sup_{m\in\mathcal{M}_\delta}E\left(|V_i|\mathbf{1}(|V_i| > b_n)|\eta_i|^s\mathbf{1}(|\bar{p}(\theta_i) - m| \leq 2h_n|)\right) \\
&+ h_n^{-1}E\left(|V_i|\mathbf{1}(|V_i| > b_n)|\eta_i|^{s-1}\mathbf{1}(|\bar{M}_i - \bar{p}(\theta_i)| > h_n|)\right)
\end{aligned}$$

Both terms are $o(r_n)$, as argued above. And by Lemma C.1 and conditions (b) and (c)

$$\sup_{m\in\mathcal{M}_\delta}|E(\Delta_n^{Vsa,r2}(m))| \leq h_n^{-1}b_nPr(|\eta_i| > \rho_n) = o(r_n)$$

Third,

$$\left|\bar{V}_{in}\mathbf{1}(|\eta_i| \leq \rho_n)\int_{\bar{p}(\theta_i)}^{\bar{M}_i}(\bar{M}_i - t)^{s-1}\kappa\left(\frac{t-m}{h_n}\right)dt\right| \leq b_n\rho_n^s B$$

and

$$\begin{aligned}
&Var\left(\bar{V}_{in}\mathbf{1}(|\eta_i| \leq \rho_n)\int_{\bar{p}(\theta_i)}^{\bar{M}_i}(\bar{M}_i - t)^{s-1}\kappa\left(\frac{t-m}{h_n}\right)dt\right) \\
&\leq E\left(\bar{V}_{in}^2\bar{\eta}_{in}^{2s}B^2\mathbf{1}(|\bar{p}(\theta_i) - m| \leq 2h_n)\right) + E\left(\bar{V}_{in}^2\bar{\eta}_{in}^{2s}B^2\mathbf{1}(|\bar{M}_i - \bar{p}(\theta_i)| \geq h_n)\right) \\
&= O\left(\left(\tilde{J}_n^{-1}\log(\tilde{J}_n)\right)^s h_n\right)
\end{aligned}$$

so Bernstein's inequality can be applied as above to obtain

$$Pr(|\bar{\Delta}_n^{Vs}(m)| > tr_n) \leq \exp\left(-\frac{t^2\log(n)}{c_1 + c_2 t b_n \log(n)(nh_n)^{-1/2}}\right)$$

The desired result follows by partitioning $\mathcal{M}_\delta$ into $N \leq \frac{1}{r_n h_n}$ intervals of width $r_n h_n$ centered at $\{m_j\}_{j=1}^N$, as above, and combining results since, for n large enough that $r_n < 1$,

$$\begin{aligned}
&\left|\int_{\bar{p}(\theta_i)}^{\bar{M}_i}(\bar{M}_i - t)^{s-1}\kappa\left(\frac{t-m}{h_n}\right)dt - \int_{\bar{p}(\theta_i)}^{\bar{M}_i}(\bar{M}_i - t)^{s-1}\kappa\left(\frac{t-m_j}{h_n}\right)dt\right| \\
&\leq \int_{\bar{p}(\theta_i)}^{\bar{M}_i}|\bar{M}_i - t|^{s-1}\left|\kappa\left(\frac{t-m}{h_n}\right) - \kappa\left(\frac{t-m_j}{h_n}\right)\right|dt \\
&\leq \int_{\bar{p}(\theta_i)}^{\bar{M}_i}|\bar{M}_i - t|^{s-1}\frac{|m-m_j|}{h_n}\mathbf{1}(|t-m_j| \leq 2h_n)dt
\end{aligned}$$

$\square$

Table B.1. Monte Carlo results for the partially linear regression model

n	J	model 1, a=1						model 1, a=2						model 1, a=4					
		OLS		IRT		PLR		OLS		IRT		PLR		OLS		IRT		PLR	
		bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
	50	0.05	0.08	0.22	0.08	-0.01	0.09	-0.04	0.10	0.44	0.09	0.01	0.11	-0.19	0.11	0.54	0.09	0.03	0.13
1000	100	0.06	0.08	0.22	0.08	0.00	0.09	-0.05	0.10	0.40	0.08	0.00	0.11	-0.23	0.12	0.53	0.10	0.00	0.13
	500	0.04	0.08	0.00	0.08	-0.02	0.10	-0.06	0.10	0.00	0.10	-0.01	0.10	-0.24	0.12	-0.01	0.11	-0.01	0.14
	50	0.05	0.06	0.21	0.06	0.00	0.07	-0.05	0.07	0.40	0.06	0.01	0.07	-0.20	0.08	0.55	0.06	0.03	0.09
2000	100	0.05	0.05	0.21	0.05	0.00	0.06	-0.06	0.07	0.44	0.06	0.00	0.08	-0.22	0.09	0.00	0.08	0.00	0.10
	500	0.05	0.06	0.00	0.06	-0.01	0.07	-0.07	0.07	0.00	0.07	-0.01	0.07	-0.24	0.09	-0.01	0.08	-0.02	0.10
		model 2, a=1						model 2, a=2						model 2, a=4					
		OLS		IRT		PLR		OLS		IRT		PLR		OLS		IRT		PLR	
		bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
	50	0.23	0.07	0.09	0.07	-0.02	0.09	0.17	0.08	0.05	0.08	-0.02	0.09	0.01	0.09	0.04	0.08	-0.02	0.09
1000	100	0.23	0.07	0.05	0.07	-0.03	0.09	0.17	0.08	0.03	0.08	-0.03	0.09	0.01	0.10	-0.01	0.09	-0.03	0.09
	500	0.23	0.07	0.03	0.07	-0.04	0.09	0.16	0.08	0.02	0.08	-0.04	0.09	0.01	0.09	-0.03	0.09	-0.04	0.10
	50	0.23	0.05	0.09	0.05	-0.01	0.06	0.16	0.06	0.07	0.06	-0.01	0.06	0.02	0.07	0.05	0.06	-0.01	0.07
2000	100	0.23	0.05	0.05	0.05	-0.01	0.06	0.16	0.06	0.05	0.05	-0.01	0.06	0.01	0.07	0.04	0.06	-0.02	0.06
	500	0.23	0.05	0.03	0.05	-0.04	0.06	0.16	0.05	0.02	0.05	-0.03	0.06	0.01	0.07	-0.04	0.07	-0.04	0.07
		model 3, a=1						model 3, a=2						model 3, a=4					
		OLS		IRT		PLR		OLS		IRT		PLR		OLS		IRT		PLR	
		bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
	50	0.07	0.08	0.48	0.09	0.03	0.10	0.21	0.11	0.99	0.09	0.09	0.13	0.40	0.13	0.59	0.13	0.28	0.22
1000	100	0.06	0.08	0.53	0.08	0.02	0.09	0.21	0.11	0.96	0.09	0.06	0.13	0.41	0.13	0.60	0.13	0.22	0.22
	500	0.05	0.08	0.11	0.09	0.00	0.10	0.19	0.11	0.38	0.11	0.02	0.13	0.38	0.13	0.64	0.12	0.07	0.25
	50	0.08	0.06	0.51	0.06	0.04	0.07	0.23	0.09	0.98	0.07	0.10	0.10	0.42	0.09	1.20	0.06	0.29	0.14
2000	100	0.07	0.06	0.10	0.06	0.02	0.07	0.20	0.08	0.93	0.07	0.04	0.10	0.41	0.09	0.60	0.09	0.21	0.17
	500	0.05	0.06	0.11	0.06	0.00	0.07	0.19	0.08	0.38	0.08	0.02	0.09	0.39	0.09	0.63	0.09	0.07	0.17

Notes: This table reports results of the Monte Carlo exercise described in Section 3.3. All entries are expressed as a fraction of the true parameter value. This table reports results for the coefficient on the observed regressor. The IRT scores were obtained using the known values for the item response parameters rather than estimated values.