Methods for Using Selection on Observed Variables to Address Selection on Unobserved Variables[1] (Work in Progress: Preliminary and Incomplete)

Joseph G. Altonji
Timothy Conley
Todd E. Elder
Christopher R. Taber

8th September 2010

[1]We have received helpful comments from seminar participants at Northwestern, University of Chicago, University of Michigan, University of Pennsylvania, University of Wisconsin at Madison, U.C. Santa Barbara, Georgetown, and Yale. We are grateful for financial support from the National Institute of Child Health and Development grant R01 HD36480-03 (Altonji and Taber) and from the Economic Growth Center, Yale University (Altonji).

Abstract

We develop new estimation methods for estimating causal effects based on the idea that the amount of selection on the observed explanatory variables in a model provides a guide to the amount of selection on the unobservables. We discuss two approaches, one of which involves the use of a factor model as a way to infer properties of unobserved covariates from the observed covariates. We construct an interval estimator that asymptotically covers the true value of the causal effect, and we propose related confidence regions that cover the true value with fixed probability.

1 Introduction

Distinguishing between correlation and causality is the most difficult challenge faced by empirical researchers in the social sciences. Social scientists are rarely in a position to run a well controlled experiment. Consequently, they rely on a priori restrictions about the relationships between the variables that are observed or unobserved. These restrictions are typically in the form of exclusion restrictions or assumptions about the functional form of the model, the distribution of the unobserved variables, or dynamic interactions. Occasionally, the restrictions are derived from a widely accepted theory or are supported by other studies that had access to a richer set of data. However, in most cases, doubt remains about the validity of the identifying assumptions and the inferences that are based on them. This reality has lead a number of researchers to focus on the estimation of bounds under weaker assumptions than those that are conventionally imposed.

In this paper, we develop estimation strategies that may be helpful when strong prior information is unavailable regarding the exogeneity of the variable of interest or instruments for that variable. This is the situation in many applications in economics and the other social sciences, with examples including the effectiveness of private schools, the effects of education on crime, the effects of crime on labor market outcomes, or the effects of obesity on health outcomes.

Our approach uses the degree of selection on observed variables as a guide to the degree of selection on the unobservables. Researchers often informally argue for the exogeneity of an explanatory variable or an instrumental variable by examining the relationship between the instrumental variable and a set of observed characteristics, or by assessing whether point estimates are sensitive to the inclusion of additional control variables.[1] We provide a formal theoretical analysis confirming the intuition that such evidence can be informative in some situations. More importantly, we provide ways to quantitatively assess the degree of selection bias or omitted variables bias and in some situations provide ways to estimate bounds. To fix ideas, let the Y be a continuous outcome of interest determined by:

[1] See for example, Currie and Duncan (1995), Engen et al (1996), Poterba et al (1994), Angrist and Evans (1998), Jacobsen et al. (1999), Bronars and Grogger (1994), Udry (1996),Cameron and Taber (2001), or Angrist and Krueger (1999). Wooldridge's (2000) undergraduate textbook contains a computer exercise (15.14) that instructs students to look for a relationship between an observable (IQ) and an instrumental variable (closeness to college).

$$Y = \alpha T + X'\Gamma_X + W^{c\prime}\Gamma^c \tag{1.1}$$

where T is a treatment variable.[2] The parameter of interest is α, the causal effect of T on Y. X is a vector of observed variables with coefficient vector Γ_X. X contains routinely measured characteristics, like basic demographics, that are not at risk of being unmeasured. W^c is a vector of all additional characteristics that are relevant for determining the outcome. Some elements of W^c are observed and some are unobserved. Using the notation $W'\Gamma$ to refer to the vector of observed components of $W^{c\prime}\Gamma^c$, we can rewrite the model as:

$$Y = \alpha T + X'\Gamma_X + W'\Gamma + \varepsilon \tag{1.2}$$

with the term ε capturing all the unobservable components of $W^{c\prime}\Gamma^c$.

The key idea in our paper is to model the relationship between W and W^c. Our operational definition of "selection on unobservables is like selection on observables" involves thinking about the breakdown of exactly which characteristics are in W (and which are unobserved) as being determined by random chance. In addition, we view both W and W^c as having a large number of elements, none of which dominates in determining Y.[3] Dominant characteristics, like gender or schooling in a wage regression, are assumed always measured and in X. Finally, although the principal source of endogeneity bias here is that T is correlated with ε, an additional source of bias stems from the correlation between W and ε. In the context of a model for the determination of W, the correlations between the elements of W are informative about the nature of the correlation between W and ε.

To illustrate the nature of the restrictions we use, consider the linear projection of T onto X, $W'\Gamma$ and ε :

$$\text{Proj}(T|X, W'\Gamma, \varepsilon) = \phi_0 + X'\phi_X + \phi W'\Gamma + \phi_\varepsilon \varepsilon. \tag{1.3}$$

In the context of this projection, our formalization of the idea that, after controlling for X, "selection on the unobservables is the same as selection on the remaining observables" leads to:

[2] We will also discuss a binary depedent variable model in which the outcome is $1(Y > 0)$.
[3] We will utilize approximations that take the number of regressors in W^c (and W) to be large.

Condition 1.

$$\phi_{\varepsilon} = \phi.$$

One may contrast Condition 1 with the implication of the usual OLS orthogonality conditions:

Condition 2.

$$\phi_{\varepsilon} = 0.$$

Roughly speaking, Condition 1 says that conditional on X, the part of Y that is related to the observables and the part related to the unobservables have the **same** relationship with T. Condition 2 says that the part of Y related to the unobservables has **no** relationship with T. We also present a set of assumptions regarding how W is determined from W^c that imply an intermediate condition between the extremes of Conditions 1 and 2:

Condition 3.

$$0 \leq \phi_{\varepsilon} \leq \phi \textit{ if } \phi \geq 0$$
$$0 \geq \phi_{\varepsilon} \geq \phi \textit{ if } \phi < 0.$$

We propose two alternative estimators that differ in how they model the relationship between W and ε. We refer to the first estimator as OU, which refers to using properties of observed ("O") covariates to infer the properties of unobserved ("U") covariates. OU amounts to estimating equation (1.2) using moment conditions that X and W are orthogonal to ε and the restriction $\phi_{\varepsilon} = \phi$. This estimates a lower (upper) bound on α if ϕ is greater (less) than 0. It requires a high level assumption that implies, roughly speaking, that conditional on X, the coefficient of the regression of T on $(Y - \alpha T)$ has the same sign and is at least as large in absolute value as the coefficient of the regression of the part of T that is orthogonal to W on the corresponding part of $Y - \alpha T$. The high level assumption is required because the estimator does not make direct use of how the observed and unobserved explanatory variables are interrelated to assess the consequences of omitted variables that affect both the treatment and the outcome. Essentially, it treats W as exogenous, in common with the vast IV literature that focusses on endogeneity of T but treats the "controls"as exogenous. Furthermore, it does not provide a way to account for the fact that randomness in which elements of W^c are observed influences the distribution of the estimator. This estimator has been applied in Altonji, Elder and Taber (2005a, 2005b; hereafter, AET) to study the effectiveness of Catholic schools, as well as in a large number of other studies. We complete the theoretical analysis of the estimator that is presented in preliminary form in AET (2002).

We also propose a second estimator that we believe is a more satisfactory approach because it relaxes the assumption that W is essentially exogenous. In this second approach, we develop a method of moments procedure that uses the bounds on selection embodied in Condition 3 and also uses a factor structure to model the covariance between the observable and unobservable covariates. This structure allows us to infer properties of unobserved covariates based on the observed correlation structure of the observed covariates W. We show that this estimator, which we name OU-Factor, consistently identifies a set that contains α. We also provide a general bootstrap procedure that may be used to construct confidence regions for the identified set, as well as a less computationally demanding bootstrap procedure that typically works well in practice.

The paper continues in Section 2, where we provide a formal model of which covariates are observed and which are unobserved. We provide an explicit set of assumptions under which Condition 1, Condition 2, and Condition 3 hold, and we elaborate on why Condition 3 is the most plausible of the three. In Section 3 we present the OU estimator. We also show that in general, Condition 1 is not sufficient to provide point identification of α. As a practical matter, this is not critical, because we focus on the use of Condition 3 to identify a range of admissible values for α rather than on point identification of α. We then turn to the OU-Factor estimator based on specifying a factor structure for W^c. In Section 4 we provide some Monte Carlo evidence on the performance of OU and OU-Factor. We offer brief conclusions in Section 5.

2 Selection Bias and the Link Between the Observed and Unobserved Determinants of the Instrument and Outcome

In this section, we begin with a formal discussion of how the observables W are chosen from the full set W^c. This is the first step in developing a theoretical foundation for using the relationship between a potentially endogenous variable (or an instrument for that variable) and the observables to make inferences about the relationship between such a variable and the unobservables. In doing so, we provide a foundation for quantitatively assessing the importance of the bias from the unobservables. We then provide a set of conditions under which Condition 3 holds, which is central to OU and OU-factor.

2.1 How are Observables Chosen?

We do not know of a formal discussion of how variables are chosen for inclusion in data sets. Here we make a few general comments that apply to many social science data sets. First, most large scale data sets such as the National Longitudinal Survey of Youth 1979, the British Household Panel, the Panel Study of Income Dynamics, and the German Socioeconomic Panel are collected to address many questions. Data set content is a compromise among the interests of multiple research, policy making, and funding constituencies. Burden on the respondents, budget, and access to administrative data sources serve as constraints. Obviously, content is also shaped by what is known about the factors that really matter for particular outcomes and by variation in the feasibility of collecting useful information on particular topics. Major data sets with large samples and extensive questionnaires are designed to serve multiple purposes rather than to address one relatively specific question. As a result, explanatory variables that influence a large set of important outcomes (such as family income, race, education, gender, or geographical information) are more likely to be collected. Because of limits on the number of the factors that we know matter, that we know how to collect, and that we can afford to collect, many elements of W^c are left out. This is reflected in the relatively low explanatory power of most social science models of individual behavior. Furthermore, in many applications, the treatment variable T is correlated with many of the elements of W^c.

These considerations suggest that Condition 2, which underlies single equation methods in econometrics, will rarely hold in practice. The optimal survey design for estimation of α would be to assign the highest priority to variables that are important determinants of *both* T and Y (it would also be to useful to collect potential instrumental variables that determine T but not Y). Condition 2 is based on the extreme assumption that surveys are sufficiently well designed to ensure that $\phi_\varepsilon = 0$.

At the other extreme, one might suspect that the constraints on data collection are sufficiently severe that it is better to think of the elements of W as a more or less random subset of the elements of W^c, rather than a set that has been systematically chosen to eliminate bias. Indeed, a natural way to formalize the idea that "selection on the observables is the same as selection on the unobservables" is to treat observables and unobservables symmetrically by assuming that the observables are a random subset of a large number of underlying variables. More formally, we use the notation S_j to denote whether covariate W_j is observed in the data set. In this notation, "selection on the observables is the same as selection on the unobservables" amounts to assuming that S_j is an iid binary random variable which is equal to one with probability P_S for all covariates in W^c. Of course, there are other ways to capture the idea of equality of selection on observables and unobservables. For example, consider a more general notation $P_{S_j} = Pr(S_j = 1)$. This object may vary across types of variables but have no systematic relationship with the influence of the variables on Y relative to the influence of the variables on T. Also, in many applications a small set of exogenous variables may play a critical role in determining Y and T and are likely to be available in data sets appropriate for the research topic in question. These variables are represented by X.

To the extent that the data set was designed for the study of the effect of T on Y, one might expect that $\phi > \phi_\varepsilon$ if $\phi > 0$ in equation (1.3). Furthermore, in many problems Y is a future outcome and will depend on unobserved factors that are determined after T or Z, a potential instrument for T, are determined. Consider the case of the effect of Catholic high schools on 12th grade test scores studied by AET. In this case, ε will reflect variability in test performance on a particular day, which presumably has nothing to do with the decision to attend Catholic high school. Furthermore, high school outcomes will be influenced by non-anticipated shocks that occur after the beginning of high school, but all of the W used in AET are measured in eighth grade. Given this sequencing, these shocks influence high school outcomes but cannot affect the probability of starting a Catholic high school. Similarly, in health applications, ε may reflect health shocks (such as an accident or exposure to a virus) that occur after the treatment choice T has been made.

With these considerations in mind, we partition W^c into two categories of variables. The first, W^*, consists of K^* variables that affect Y and potentially T (and possibly Z) and may or may not be observed by the econometrician. The subvector W of W^* is observed, while the subvector W^u is not. The second category consists of the vector W^{**}, which represents variables that have a zero probability of being observed and used. In this case,

$$\begin{aligned} W'\Gamma &= \sum_{j=1}^{K^*} S_j W_j \Gamma_j \\ \varepsilon &= \sum_{j=1}^{K^*} (1 - S_j)\, W_j \Gamma_j \ + W^{**\prime}\Gamma^{**} = W^{u\prime}\Gamma^u + \xi \end{aligned}$$

where Γ^u is the subvector of Γ^c that corresponds to W^u, Γ^{**} is the subvector of Γ^c that corresponds to W^{**}, and $\xi = W^{**\prime}\Gamma^{**}$. Given that W^{**} represents unanticipated covariates, we assume that ξ is orthogonal to (W^*, T, Z). This implies Condition 3

$$\begin{aligned} 0 &\leq \phi_\varepsilon \leq \phi \text{ if } \phi > 0 \\ 0 &\geq \phi_\varepsilon \geq \phi \text{ if } \phi < 0 \end{aligned} \tag{2.1}$$

as the basis for the estimation strategies developed below, which focus on estimation of a confidence set for α that contains the true value rather than on point estimation.

Often there is a third category of variables, X, consisting of factors that play an essential role in determining Y and potentially Z and T. These would be included in any serious study of Y and may be different in nature from the other variables and thus not informative about how the properties of $W^{u\prime}\Gamma^u$. In AET's study of Catholic schools, Catholic religion is such a variable.

2.2 Implications of Random Selection of Observables

We are now ready to consider the implications of random selection from W^*. We begin with the general case. We first derive the probability limit of ϕ_ε/ϕ as the number of covariates in W^* becomes large. We then consider several special cases.

For individual i, we define Y_i and Z_i as outcomes for a sequence of models indexed by K^*, where K^* is the number of elements of W^*.[4] A natural part of the thought experiment in which K^* varies across models is the idea that the importance of each individual factor declines with K^*. We take the dimensions of X and W^{**} as fixed.

Define $\mathcal{G}^{K^*}$ as the information set consisting of the realizations of the S_j, the Γ_j, and the joint distribution of W_{ij} conditional on $j = 1, ..., K^*$. That is, $E(W_{ij} \mid \mathcal{G}^{K^*})$ is the mean for

[4] The "local to unity" literature in time series econometrics" (e.g., Stock, 1994) and the "weak instruments" literatures (e.g., Staiger and Stock, 1997) are other examples in econometrics in which the asymptotic approximation is taken over a sequence of models, which in the case of those literatures, depend on sample size.

a given j, where the expectation is only over i, but $E(W_{ij})$ is an unconditional expectation over both i and j. It may be helpful to think of this data generation process as operating in two steps. First the "model" is drawn: for a given K^*, the joint distribution of W_{ij}, T_i, Z_i, ξ_i, and S_j are drawn. We can think of $\mathcal{G}^{K^*}$ as representing this draw. In the second stage of the data generating process, individual data is constructed from these underlying distributions.

The two steps combine to generate Y_i as is represented in Assumption 1.

Assumption 1.

$$Y_i = \alpha T_i + X_i' \Gamma_X + \frac{1}{\sqrt{K^*}} \sum_{j=1}^{K^*} W_{ij} \Gamma_j + \xi_i \tag{2.2}$$

where (W_{ij}, Γ_j) is unconditionally stationary (indexed by j), and X_i includes an intercept.

We use slightly non-standard notation in Assumption 1. Rather than explicitly indexing parameters by K^*, we suppress a K^* index on (W_{ij}, Γ_j) and bring a $\frac{1}{\sqrt{K^*}}$ out in front of the sum. This scaling guarantees that no particular covariate will be any more important *ex ante* than the others. It embodies the idea that a large number of components determine most outcomes in the social sciences. Any variables that play an outsized role in Y and Z are assumed to be observed with probability 1 and are included in the set of special regressors X. The number of elements of X is fixed. Note that Assumption 1 involves unconditional stationarity. Conditional on $\mathcal{G}^{K^*}$, the variance of the W_{ij} and the contribution of the W_{ij} to the variance of Y will differ across j.

Throughout we will project all variables on X and take residuals to remove X from the regression. We will use "tildes" to denote the residuals from these projections, so we define

$$\begin{aligned}
\widetilde{W}_{ij} &\equiv W_{ij} - Proj(W_{ij} \mid X_i; \mathcal{G}^{K^*}) \\
\widetilde{T}_i &\equiv T_i - Proj(T_i \mid X_i; \mathcal{G}^{K^*}) \\
\widetilde{Z}_i &\equiv Z_i - Proj(Z_i \mid X_i; \mathcal{G}^{K^*}) \\
\widetilde{Y}_i &\equiv Y_i - Proj(Y_i \mid X_i; \mathcal{G}^{K^*})
\end{aligned}$$

where $Proj$ denotes a linear projection.[5] Let $\sigma_{j,\ell}^{K^*} = E\left(\widetilde{W}_{ij}\widetilde{W}_{i\ell} \mid \mathcal{G}^{K^*}\right)$. To guarantee that $var(Y_i)$ is bounded as K^* becomes large, we assume that

[5]Formally, the linear projection projection of a generic Y_i on a generic X_i is defined by $X_i'\delta$ where δ satisfies $E[(Y_i - X_i'\delta)X_i \mid \mathcal{G}^{K^*}] = 0$. Hereafter, this projection is meant to be the population projection, i.e., for a very large N, but with K^* fixed.

Assumption 2.

$$0 < \lim_{K^* \to \infty} \frac{1}{K^*} \sum_{j=1}^{K^*} \sum_{\ell=1}^{K^*} E(\sigma_{j,\ell}^{K^*} \Gamma_j \Gamma_\ell) < \infty$$

and

$$\lim_{K^* \to \infty} Var \left(\frac{1}{K^*} \sum_{j=1}^{K^*} \sum_{\ell=1}^{K^*} \sigma_{j,\ell}^{K^*} \Gamma_j \Gamma_\ell) \right) \to 0 \ .$$

The next two assumptions guarantee that $cov(Z_i, Y_i)$ is well behaved as K^* grows.

Assumption 3. *For any* $j = 1, ..., K^*$*,define* $\mu_j^{K^*}$ *so that*

$$E\left(\widetilde{Z}_i \widetilde{W}_{ij} | \mathcal{G}^{K^*}\right) = \frac{\mu_j^{K^*}}{\sqrt{K^*}} \ .$$

Then

$$E(\mu_j^{K^*} \Gamma_j) < \infty$$

and

$$\lim_{K^* \to \infty} Var \left(\frac{1}{K^*} \sum_{j=1}^{K^*} \mu_j^{K^*} \Gamma_j \right) \to 0 \ .$$

Since Assumptions 2 and 3 are quite abstract, it may be helpful to pause to discuss examples of models that satisfy Assumption 1-3 before turning to the rest of our assumptions. The key example is the factor model of the X_i and W_{ij}, which is central to one of our estimation strategies. To avoid repetition, we defer presentation of the factor model until Section 5.

Another example is a case in which the W_{ij} are linked across j through an MA model.[6] The MA example is the most straightforward when one examines Assumptions 1 and 2 given that those assumptions refer to observables as though they have a sequential ordering. To simplify the example, we consider a case in which X_i only contains an intercept term and the W_{ij} are stationary conditional on $\mathcal{G}^{K^*}$. This means that the W_{ij} will have the same marginal distribution for all j. This is not realistic for the types of data sets typically used by economists, and it is not required for Theorem 1,but it simplifies the exposition.[7] Specifically, assume that across individuals i, W_{ij} is generated by independent and identically

[6]The case of a general ARMA structure is conceptually straightforward, but the algebra becomes substantially more complicated.

[7]What is important is that W_{ij} is unconditionally stationary. Conditional on $\mathcal{G}^{K^*}$, the distribution of W_{ij} is not restricted.

distributed stationary $MA(q_w)$ processes

$$W_{ij} = \zeta_{ij} + \sum_{\ell=1}^{q_w} \mu_\ell \zeta_{ij-1} \ ,$$

where ζ_{ij} is i.i.d. with finite variance σ_ζ^2. The W_{ij} processes are also independent of the Γ_j process, and we assume further that Γ_j is generated from a stationary process with finite fourth moments. We think of j as being ordered so that variables that measure related factors appear close to each other in the j sequence.[8] Given our assumptions about the W_{ij} processes and Γ_j, it is almost immediate that Assumption 1 is satisfied by the MA model. In the Appendix we show that the model satisfies Assumption 2.

To consider Assumption 3, we need a model for Z. In the Appendix, we prove that Assumption 3 is satisfied by the MA model if the model for Z takes a form which is similar to the form of Y_i:

Assumption 4.

$$Z_i = X_i'\beta_X + \frac{1}{\sqrt{K^*}} \sum_{j=1}^{K^*} W_{ij}\beta_j + \psi_i, \tag{2.3}$$

where (i) ψ_i is independent of all of the elements of W^c.(ii) β_j is a stationary process with finite second moments. β_j may be correlated with Γ_j.

It is convenient to rewrite the model for Z as

$$Z_i = X_i'\beta_X + \frac{1}{\sqrt{K^*}} \sum_{j=1}^{K} W_{ij}\beta_j + u_i \tag{2.4}$$

where $u_i = \frac{1}{\sqrt{K^*}} \sum_{j=K+1}^{K^*} W_{ij}\beta_j + \psi_i$, and all variables are residuals from linear projections onto the space of X_i. We use the above specification of Z_i in much of the analysis below, but we note that the main results below (Theorem 1 and Corollaries 1-3) do not require assumptions about Z beyond those given in Assumptions 1-3.

Finally, we provide assumptions about the process under which observables are chosen. Consider the case discussed above in which variables are chosen at random:

[8]For example, consider a study of educational attainment in which measures of student behavior (e.g., absenteeism, suspensions, getting into fights, acting out in class) are viewed as potentially important control variables. If these variables appear in sequence, the above model captures the fact that they are dependent and will have Γ_j coefficients that are related. Only a subset of the behavioral variables might actually be observed.

Assumption 5. *For $j = 1, ..., K^*$, S_j is independent and identically distributed with $0 < \Pr(S_j = 1) \equiv P_s \leq 1$. S_j is also independent of all other random variables in the model. If $var(\xi) \equiv \sigma_\xi^2 = 0$, then $P_S < 1$.*

Assumption 6. *ξ is mean zero and uncorrelated with Z and W^*.*

As mentioned above, the assumption that ξ is uncorrelated with Z and W^* is not very restrictive, since for a given value of K^* one can redefine Γ^* and ξ so that ξ is uncorrelated with W^*.[9]

First we consider the relationship between ϕ and ϕ_ε in the general case and then derive three key special cases. Finally, we relax Assumption 5 that $P_{Sj} = P_S \;\forall j$ and instead assume that P_{Sj} is a positive function of the degree to which omitting W_j will lead to bias in the IV estimator.

Note that our asymptotic analysis is nonstandard in two respects. First, we are allowing the number of underlying explanatory variables, K^*, to get large. Second, the random variable W_{ij} is different from the random variables Γ_j and S_j in the following way. For each j we draw one observation on Γ_j and S_j which is the same for every person in the population; however, each individual i draws his own W_{ij}.

Theorem 1. *Define ϕ and ϕ_ε such that*

$$\begin{aligned}
&Proj\left(Z_i \mid X_i, \frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j W_{ij}\Gamma_j, \frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}(1-S_j)W_{ij}\Gamma_j + \xi; \mathcal{G}^K\right) \\
&\quad = X'\phi_X + \phi\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j W_{ij}\Gamma_j\right) + \phi_\varepsilon\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}(1-S_j)W_{ij}\Gamma_j + \xi_i\right).
\end{aligned}$$

Then under assumptions 1-3 and 5-6, if the probability limit of ϕ is nonzero, then

$$\frac{\phi_\varepsilon}{\phi} \xrightarrow[K^*\to\infty]{p} \frac{(1-P_s)A}{(1-P_s)A + \sigma_\xi^2}$$

where

$$A \equiv \lim_{K^* \to\infty} E\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sigma_{j,j}^{K^*}(\Gamma_j)^2\right).$$

If the probability limit of ϕ is zero, then the probability limit of ϕ_ε is also zero.

[9] Assume one can write ξ_i as $\xi_i = \frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} W_{ij}B_{wj} + \tilde{\xi}_i$ where $\tilde{\xi}_i$ is independent of W^* and Z. Replace Γ_j with $\Gamma_j + B_{wj}$ and replace ξ_i with $\tilde{\xi}_i$ in (2.2). The key assumption would then be that $\tilde{\xi}_i$ is uncorrelated with ψ_i.

(Proof in Appendix)

Next we consider three separate cases which we present as corollaries. We omit the proofs of these as they follow immediately from the proof of Theorem 1.

Corollary 1. *When* $\sigma_\xi^2 = 0$,

$$plim(\phi - \phi_\varepsilon) = 0.$$

The case in which $\sigma_\xi^2 = 0$ is the case in which $W^c = W^*$, meaning that W is a random subset of all of elements of W^c. Corollary 1 states that the coefficients of the projection of Z_i onto $\frac{1}{\sqrt{K^*}} \sum_{j=1}^{K^*} S_j W_{ij} \Gamma_j$ and $\frac{1}{\sqrt{K^*}} \sum_{j=1}^{K^*} (1 - S_j) W_{ij} \Gamma_j$ approach each other with probability one as K^* becomes large.

The other extreme is the case in which all the important control variables that affect both Z and Y are included in the model, so the variation in the composite error term ε arises from ξ only:

Corollary 2. *When* $P_s = 1$,

$$plim(\phi_\varepsilon) = 0.$$

What about the case in which selection on observables is stronger than selection on unobservables but there is still some selection on unobservables? This corresponds to the case in which $var(\xi) > 0$ and $P_s < 1$. The next Corollary considers this case:

Corollary 3. *When* $0 < P_s < 1$ *and* $\sigma_\xi^2 > 0$,

either

$$0 < \; plim(\phi_\varepsilon) < plim(\phi),$$

or

$$plim(\phi) < plim(\phi_\varepsilon) < 0,$$

or

$$0 = plim(\phi_\varepsilon) = plim(\phi).$$

This Corollary plays a key role in the estimator below.

2.3 Systematic Variation in P_{s_j}

In this subsection we extend Theorem 1 to the case in which P_{sj} is positively related to the impact of including W_{ij} on the bias in IV estimation of α.

Without loss of generality, assume the correlation between Z_i and $W_i'\Gamma$ is positive as one could multiply Z_i by -1 to change the sign. In general, the impact of including a particular W_j is a complicated function of Γ_j, μ_j, the Γ_ℓ and μ_ℓ of the variables that remain excluded, and the covariances among both the included and excluded variables. Thus, it is not straightforward to characterize the relative impact of the exclusion of particular variables on the bias. Consequently, we do not attempt to formulate a general result, but instead consider a special case in which it is easy to assess the relationship between ϕ_ε and ϕ. It is intuitive that exclusion of W_{ij} variables with a strong positive association with both Z_j and Y will lead to bigger bias, everything else equal. Consequently, we assume that S_j is positively related to $E(Z_j W_j \Gamma_j)$. More specifically, we assume

Assumption 7.

$$E\left(\mu_j \Gamma_j \mid S_j = 1\right) > E\left(\mu_j \Gamma_j \mid S_j = 0\right) > 0.$$

We make additional assumptions that make it very easy to establish the result. First, we assume that S_j is independent of $W_{ij}\Gamma_j$:

Assumption 8. *S_j is independent of $W_j\Gamma_j$.*

This is neither an attractive assumption nor a necessary condition, but it implies that the variation in P_{S_j} will not affect the second moments of

$$\left\{ \frac{1}{\sqrt{K^*}} \sum_{j=1}^{K^*} S_j W_{ij} \Gamma_j, \left(\frac{1}{\sqrt{K^*}} \sum_{j=1}^{K^*} (1 - S_j) W_{ij} \Gamma_j + \xi_i \right) \right\}$$

as K^* gets large.[10] With these assumptions in hand, in the Appendix we establish the following result.

Theorem 2. *Define ϕ and ϕ_ε as in Theorem 1. Then under assumptions 1-3 and 5-8, as K^* gets large,*

$$0 < \phi_\varepsilon < \phi.$$

[10] If we impose (2.4), we might instead assume S_j is positively correlated with $E(\beta_j \Gamma_j)$ but unrelated to the marginal distributions of Γ_j, β_j, and $cov(W_j, W_\ell)$ for all j and ℓ.

(Proof in Appendix)

A key, but perhaps subtle, implication of this theorem is that we get the inequality $\phi_\varepsilon < \phi$ even when $\sigma_\xi^2 = 0$. This model thus gives another explanation for this inequality.

3 Estimators of α

We now discuss ways to estimate α. In Section 4.1 We set the stage by reviewing the OU estimator introduced in AET (2002, 2005). Then we present OU-Factor, beginning with the factor model of W^* that it requires.

3.1 The OU Estimator

The OU Estimator is simple. If were were to estimate the standard treatment effect model embodied in (1.2) by instrumental variables, we can think of this as GMM with the standard moment conditions $E((X,W)\varepsilon) = 0$ and the IV moment equation $E(Z\varepsilon) = 0$. The basic idea of the OU estimator is to simply replace the moment equation $E(Z\varepsilon) = 0$ with condition 3. The problem, however, is that Condition 3 is not operational unless $E(\varepsilon \mid W) = 0$ because Γ is not identified. Mean independence of ε and (X,W) is maintained in virtually all studies of selection problems, because without it, α is not identified even if one has a valid exclusion restriction.[11] Our discussion of how the observables are arrived at makes clear that this is hard to justify in most settings. If the observables are correlated with one another, as in most applications, then the observed and unobserved determinants of Y are also likely to be correlated.

Most applications to date have involved either $T = 1(Z > 0)$ or $T = Z$, so we focus on this case (in which we suppress the norming of individual elements of W by $\sqrt{K^*}$):

$$T = Z = X'\beta_x + W'\beta + u$$

AET address the problem as follows. Assume that $E(\varepsilon \mid X, W)$ is linear, and define G and e to be the slope vector and error term of the "reduced forms":

$$E\left(\widetilde{Y} - \alpha\widetilde{T} \mid \widetilde{W}\right) \equiv \widetilde{W}'G \tag{3.1}$$

$$\widetilde{Y} - E\left(\widetilde{Y} - \alpha\widetilde{T} \mid \widetilde{W}\right) \equiv e. \tag{3.2}$$

[11] The exception is when the instrument is uncorrelated with W (and X) as well as ξ, as when the instrument is randomly assigned in an experimental setting.

Let $\phi_{W'G}$ and ϕ_e be the coefficients of the projection of T on $W'G$ and e (in a regression model that includes X). Sufficient conditions for $0 \leq \phi_e \leq \phi_{W'G}$ when $\phi_{W'G} > 0$ are the assumptions of Theorem 1 and the following condition:

Assumption 9.

$$\frac{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{W}_j \widetilde{W}_{j-\ell}\right) E\left(\beta_j \Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{W}_j \widetilde{W}_{j-\ell}\right) E\left(\Gamma_j \Gamma_{j-\ell}\right)} = \frac{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j \widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\beta_j \Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j \widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\Gamma_j \Gamma_{j-\ell}\right)}, \tag{3.3}$$

for the set of variables W_j in $j = 1, ..., K^*$,

where $\widetilde{\widetilde{W}}_j$ is the component of $\widetilde{W}_j$ that is orthogonal to the observed variables (X, W), for all elements of W^*. Roughly speaking (3.3) says that the regression of T on $\left(\widetilde{Y} - \alpha\widetilde{T} - \xi\right)$ is equal to the regression of the part of $\widetilde{T}$ that is orthogonal to $\widetilde{W}$ on the corresponding part of $\left(\widetilde{Y} - \alpha\widetilde{T} - \xi\right)$. One can show that this condition holds under the standard assumption $E(\varepsilon \mid W) = 0$, in which case G and e equal Γ and ε, respectively. However, $E(\varepsilon \mid W) = 0$ is not necessary for (3.3).[12]

Theorem 3. *Define $\phi_{W'G}$ and ϕ_e such that*

$$\begin{aligned} &Proj\left(\widetilde{Z}_i \mid \frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j \widetilde{W}_{ij} G_j, \frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} \left(1 - S_j\right) \widetilde{W}_{ij}\Gamma_j + \xi; \mathcal{G}^K\right) \\ &\quad = \phi_{W'G}\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j W_{ij}\Gamma_j\right) + \phi_e\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} \left(1 - S_j\right) W_{ij}\Gamma_j + \xi_i\right). \end{aligned}$$

Then under assumptions 1-6 and 9, as K^* gets large, if the probability limit of ϕ is nonzero, then

$$\frac{\phi_e}{\phi_{W'G}} \xrightarrow{p} \frac{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{W}_j \widetilde{W}_{j-\ell}\right) E\left(\Gamma_j \Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{W}_j \widetilde{W}_{j-\ell}\right) E\left(\Gamma_j \Gamma_{j-\ell}\right) + \sigma_\xi^2}.$$

If the probability limit of $\phi_{W'G}$ is zero then the probability limit of ϕ_e is also zero.

(Proof in Appendix)

[12] For example, one can show that (3.3) will also hold if $E\left(\beta_j \Gamma_{j-\ell}\right)$ is proportional to $E\left(\Gamma_j \Gamma_{j-\ell}\right)$ regardless of the correlations among the W_j.

Based on the argument that selection on unobservables is likely to be weaker than selection on observables, one might impose Condition 3 rather than Condition 1. The upshot is that one can work with the system

$$\begin{aligned} Y &= \alpha T + W'G + e. \\ T &= W'\beta + u \\ 0 &\leq \left|\frac{cov(u,e)}{var(e)}\right| \leq \left|\frac{Cov(W'\beta, W'G)}{Var(W'G)}\right|, \end{aligned}$$

and estimate the set of α values that satisfy the above inequality restrictions. In practice, AET find that the lower bound is obtained when the equality of selection condition $\frac{cov(u,e)}{var(e)} = \frac{Cov(W'\beta,W'G)}{Var(W'G)}$ is imposed and the upper bound corresponds to the case in which T is treated as exogenous, with $\frac{cov(u,e)}{var(e)} = 0$.

One can perform statistical inference accounting for variation over i conditional on which W are observed in the usual way. We do not develop this idea here; however, there is no obvious way to account for random variation due to the draws of S_j.

3.1.1 Is Equality of Selection on Observables and Unobservables Enough to Identify α?

We favor using Condition 3 to estimate bounds for α based on a range of the degree of selection on unobservables, but it is interesting to ask whether Condition 1 is sufficient for point identification of α. Perhaps surprisingly, in general the answer is no. To demonstrate this, we assume that Y is determined by $Y = \alpha T + W'\Gamma + \varepsilon$ as above and consider the special case in which $E(\varepsilon \mid W) = 0$, which implies that $G = \Gamma$ and $e = \varepsilon$.

Define π and $\widetilde{\beta}$ so that

$$\text{Proj}\,(Z \mid W) = W'\widetilde{\beta}, \tag{3.1}$$

$$\text{Proj}\,(T \mid W, Z) = W'\pi + \lambda Z, \tag{3.2}$$

and define v and u to be the residual components of Z and T, so that

$$v \equiv Z - W'\widetilde{\beta} \tag{3.3}$$

$$u \equiv T - W'\pi - \lambda Z. \tag{3.4}$$

Theorem 4. *Suppose that ε is independent of W. Under Condition 1, the true value of α is a root of a cubic polynomial. Thus the identified set contains one, two or three values.*

(Proof in Appendix)

This theorem implies that even if $Cov(\varepsilon, W'\Gamma) = 0$, there are typically either three solutions (i.e., three values of α, which we label α^*, that satisfy the moment conditions) or there is a unique solution that equals α.

Theorem 5. *If we impose the same model as above but use T as an instrument for itself, the true value of α is a root of a quadratic polynomial with two roots:*

$$\begin{aligned} \alpha^* &= \alpha \\ \alpha^* &= \alpha + \frac{var(\varepsilon)}{cov(u, \varepsilon)}. \end{aligned}$$

(Proof in Appendix)

Although there are two roots, this result is useful. When an applied researcher is worried about the bias in an IV estimator, including the case when $Z = T$, he or she often has a strong prior about the sign of the bias, which is the sign of $cov(u, \varepsilon)$. Imposing an assumption about the sign of $cov(u, \varepsilon)$ on the data delivers point identification; if one imposes that $cov(u, \varepsilon)$ is positive (negative), then the smaller (larger) of the two solutions is the true value. One should not make too much of this result, because in most applications variables represented by W^{**} will be present, so that $var(\xi)$ will be positive and equality of selection will not hold. Consequently, we focus on the construction of bounds rather than on point estimation.

3.2 OU-Factor: A Bounds Estimator Based on a Factor Model of $\widetilde{W}_{ij}$

3.2.1 A Factor Model of $\widetilde{W}_{ij}$

We now present a factor model of $\widetilde{W}_{ij}$, which is central to the estimator proposed below. The factor model is a convenient way to model the relationship among the covariates. We assume that $\widetilde{W}_{ij}$ has a factor structure

$$\widetilde{W}_{ij} = \frac{1}{\sqrt{K^*}} \widetilde{F}_i' \Lambda_j + v_{ij}, \; j = 1, ..., K^*, \tag{3.5}$$

where F_i is an r dimensional vector. We treat r as finite, so while the dimension of W_{ij} grows, the number of factors remains constant. Keep in mind that $\widetilde{W}_{ij}$ is the component of W_{ij} that is orthogonal to X_i. It may seem arbitrary to assume that (3.5) applies to $\widetilde{W}_{ij}$ rather than W_{ij}. To motivate this assumption, suppose that both W_{ij} and X_i are defined by a factor model:

$$W_{ij} = \alpha_j + \frac{1}{\sqrt{K^*}} F_i' \Lambda_j + \nu_{ij}$$
$$X_i = \begin{bmatrix} 1 \\ \Lambda_x F_i + \omega_i \end{bmatrix},$$

where F_i is the factor, α_j is the mean of W_{ij}, the dimension of X_i is $K_x \times 1$, Λ_x is a $(K_x - 1) \times r$ matrix and ω_i is a $(K_x - 1) \times 1$ vector. Then

$$\begin{aligned}
\widetilde{W}_{ij} &= W_{ij} - Proj(W_{ij} \mid X_i) \\
&= \frac{1}{\sqrt{K^*}} F_i' \Lambda_j + v_{ij} - \frac{1}{\sqrt{K^*}} Proj(F_i \mid X_i)' \Lambda_j \\
&\equiv \frac{1}{\sqrt{K^*}} \widetilde{F}_i' \Lambda_j + v_{ij} \ ,
\end{aligned}$$

where we have defined $\widetilde{F}_i = F_i - proj(F_i \mid X_i)$. In the rest of this section we abstract from X_i and focus on $\widetilde{W}_{ij}$.

We normalize the variance/covariance matrix of $\widetilde{F}_i$ be to the identity matrix. Define $\sigma_j^2 \equiv E(v_{ij}^2 \mid j)$, $j = 1, ..., K^*$. When we refer to the "factor model", we will often mean the model defined by (3.5), the model (2.2) for Y, and the model (2.3) for Z. We continue to assume that ξ_i and ψ_i are independent of all of the W_{ij} and of each other. They may also have factor structures, but the factors are uncorrelated with $\widetilde{F}_i$. The stochastic structure of the model is that Λ_j, Γ_j, β_j and σ_j^2 differ across j, but are identical for all individuals in the population, $i = 1, ..., N$.

We model Z_i according to assumption (2.4), and we define an analogous structure for T_i:

$$T_i = X_i' \delta_X + \frac{1}{\sqrt{K^*}} \sum_{j=1}^{K} W_{ij} \delta_j + \omega_i.$$

We redefine $\mathcal{G}^{K^*}$ to refer to aspects of the model of W_i, T_i, Y_i, and Z_i, that do not vary across individuals:

$$\mathcal{G}^{K^*} = \left\{ (\Gamma_j, \beta_j, \delta_j, \Lambda_j, \sigma_j^2, S_j) \text{ for } j = 1, ..., K^* \right\}.$$

For estimation, we make the following additional assumptions.

Assumption 10. *(i)* $\left(\Gamma_j, \beta_j, \delta_j, \Lambda_j, \sigma_j^2\right)$ *is i.i.d with fourth moments; (ii) the components* ξ_i *and* ψ_i *of* Y_i *and* Z_i *respectively are independent of* W_i^* *and of each other; (iii)* ξ_i *is independent of* X_i.

Assumption 10 (ii) implies that there is a component of Z_i that is independent of the observed and unobserved determinants of Y. Without this there is no hope of identifying α using Z or a component of Z as a source of exogenous variation in T, because there is no exogenous variation. In the Appendix we verify that the factor model of W in conjunction with the model (2.2) for Y and (2.3) for Z satisfies Assumptions 1, 2, and 3 of Theorem 1.

3.2.2 An Estimator of an Admissible Set for α

In contrast to the OU estimator, here we use the factor model to directly address the problem posed by the fact that basic introspection suggests that the elements of W_i (as well as T_i, X_i, and Z_i) are likely correlated with the error term. We study identification under the following assumptions. First, we assume that the econometrician can observe the sequence of models indexed by $K^* = 1, ..., \infty$ and that for each model she observes K,the number of observed covariates in W (but not the number of unobserved covariates), as well as the joint distribution of Y_i, Z_i, T_i, X_i and $\{W_{ij} : S_{ij} = 1\}$. Second, we assume that $K/K^* \to P_{s0}$. Third, we assume that N becomes large faster than K^*, with $\frac{K^*}{N} \to 0$, so that we can take sequential limits. This seems like a good approximation in problems where K and K^* are large, but not for problems in which the number of variables that determine Y_i is small.

In general the model is not point identified, so we provide an estimator of a set that contains the true values. The key subset of the parameter vector of our model is $\theta = \{\alpha, \phi, P_s, \sigma_\xi^2)$, where we abstract from parameters that are point identified given θ. The true value of θ is $\theta_0 = \{\alpha_0, \phi_0, P_{s0}, \sigma_{\xi 0}^2)$ which lies in the compact set $\bar{\Theta}$. Our approach is to estimate a set $\widehat{\Theta}$ that asymptotically will contain the true value θ_0. The key restrictions on the parameter set are that

$$0 < P_{s0} \leq 1, \text{ and} \tag{3.6}$$

$$\sigma_{\xi 0}^2 \geq 0. \tag{3.7}$$

The case in which $P_{s0} = 1$ corresponds to the standard IV case represented by Condition 2, while $\sigma_{\xi 0}^2 = 0$ corresponds to the "unobservables like observables" case represented by Condition 1. We construct an estimate of the set of values of α by estimating the set of

θ that satisfy all of the conditions and then projecting onto the α dimension. We then go on to discuss construction of confidence intervals. While the upper and lower bound of the estimated set does not have to correspond to the cases in which $P_{s0} = 1$ and $\sigma^2_{\xi 0} = 0$, in practice we find that it does.

It will be useful to make use of matrix notation. We assume without loss of generality that the variables are ordered so that $j = 1, .., K$ corresponds to the K observed covariates in W^c. Unless indicated otherwise,

- For a generic variable $B_i, i = 1, ..., N$, B will represent the $N \times 1$ vector.
- For a generic variable $B_j, j = 1, ..., K^*$, B will represent the $K \times 1$ vector of observable characteristics and B^* will represent the full $K^* \times 1$ vector.
- For a generic variable $B_{ij}, i = 1, .., N, j = 1, ..., K^*$, B will represent the $N \times K$ matrix of observable characteristics, $B^*$the full $N \times K^*$matrix of covariates, and B_i represents the $K \times 1$ vector of B_{ij} for a given i.
- We also employ the convention of using capital letters for matrices so, for example, the matrix version of v_{ij} will be written as V.

Given the large amount of notation we concentrate on the 1 factor case ($r = 1$), so $\widetilde{F}_i$ and Λ_j are scalars. We fully expect that the results generalize to the multiple factor case. We now present the estimator, which has two stages.

Stage 1

In the first stage we estimate the $\Lambda_1, .., \Lambda_K$ and $\sigma^2_1, ..., \sigma^2_K$. The moment conditions are the K equations

$$E\left(\widetilde{W}_{ij_1}\widetilde{W}_{ij_2}\right) = \frac{1}{K^*}\Lambda^2_{j_1} + \sigma^2_{j_1};\ j_1 = 1, ..., K,\ j_1 = j_2\ , \tag{3.8}$$

and the $K \cdot (K-1)/2$ equations

$$E\left(\widetilde{W}_{ij_1}\widetilde{W}_{ij_2}\right) = \frac{1}{K^*}\Lambda^2_{j_1}\ ;\ j_1,\ j_2 = 1, ..., K,\ j_1 \neq j_2\ . \tag{3.9}$$

This is a standard GMM problem. As N grows we will obtain $\sqrt{N}$ consistent estimates of $\frac{1}{\sqrt{K^*}}\Lambda_j$ for each j and for $\widehat{\sigma_j^2}$ by using the sample analogues to (3.8) and (3.9). Note that K^* is not known since it depends on the number of unobserved variables. However, the econometrician knows $K = P_{S0}K^*$. To simplify the exposition we define $\widehat{\lambda}_j$ to be the GMM estimate of the parameter $\sqrt{K} \times \frac{1}{\sqrt{K^*}}\Lambda_j = \sqrt{P_{S0}}\Lambda_j$ and λ to be the corresponding vector. In practice we just replace the left side of the equations by $\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{W}_{i_1}\widetilde{W}_{ij_2}\right)$ and choose $\widehat{\lambda}_j$ and $\widehat{\sigma_j^2}$ as the values that minimize the appropriately weighted difference between the values of $\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{W}_{i_1}\widetilde{W}_{ij_2}\right)$ and the predictions summarized in the moment conditions above.

Stage 2

We estimate the rest of the parameters in a second stage. If we knew α_0 we could estimate Γ conditional on α_0 by taking advantage of the moment condition

$$\begin{aligned}
\sqrt{K^*}E\left[\widetilde{W}_{ij}\left(\widetilde{Y}_i - \alpha_0\widetilde{T}_i\right)\right] &= \sqrt{K^*}E\left[\left(\frac{1}{\sqrt{K^*}}\widetilde{F}_i\Lambda_j + v_{ij}\right)\left(\frac{1}{\sqrt{K^*}}\sum_{\ell=1}^{K^*}\frac{1}{\sqrt{K^*}}\widetilde{F}_i\Lambda_\ell\Gamma_\ell + \frac{1}{\sqrt{K^*}}\sum_{\ell=1}^{K^*}v_{ij}\Gamma_\ell\right)\right] \\
&= \Lambda_j\left(\frac{1}{K^*}\sum_{\ell=1}^{K^*}\Lambda_\ell\Gamma_\ell\right) + \sigma_{vj}^2\Gamma_j \\
&\xrightarrow{p} \Lambda_j' E(\Lambda_\ell\Gamma_\ell) + \sigma_{vj}^2\Gamma_j.
\end{aligned}$$

We work with the sample analog of the above expression,

$$\left[\ \sqrt{K^*}\tfrac{1}{N}\widetilde{W}'\left(\widetilde{Y} - \alpha_0\widetilde{T}\right)\ \right] = \left[\ \tfrac{1}{K}\tfrac{1}{P_{s0}}\widehat{\lambda}\widehat{\lambda}'\Gamma + \Sigma\Gamma\ \right],$$

where Σ is the diagonal matrix composed of the σ_j^2 terms. Thus, for the parameter θ we can construct the estimator

$$\widehat{\Gamma}(\theta) \approx \left[\frac{1}{P_s K}\widehat{\lambda}\widehat{\lambda}' + \widehat{\Sigma}\right]^{-1}\frac{1}{N}\widetilde{W}'\left(\widetilde{Y} - \alpha\widetilde{T}\right), \tag{3.10}$$

where we define $\widehat{\Sigma}$ is the diagonal matrix composed of $\widehat{\sigma_j^2}$ which is estimated in the first stage.

One may show that

$$\phi_0 = \frac{\left[E(\Gamma_j\Lambda_j)E(\beta_j\Lambda_j) + E(\Gamma_j\beta_j\sigma_j^2)\right]\left[P_{s0}\left(1-P_{s0}\right)E(\Gamma_j^2\sigma_j^2) + P_{s0}\sigma_{\xi0}^2\right]}{\sigma_{\xi0}^2\left[P_{s0}^2E(\Gamma_j\Lambda_j)^2 + P_{s0}E(\Gamma_j^2\sigma_j^2)\right] + \left[E(\Gamma_j\Lambda_j)^2 + E(\Gamma_j^2\sigma_j^2)\right]\left(1-P_{s0}\right)P_{s0}E(\Gamma_j^2\sigma_j^2)}.$$

Using this fact, we define our estimator based on the following system of equations.

$$
\begin{aligned}
(3.11)\quad q^1_{N,K^*}(\theta) =& \frac{1}{N}\sum_{i=1}^{N} \widetilde{W}_i'\widehat{\Gamma}(\theta) \times \\
& \left[\widetilde{Z}_i - \phi \widetilde{W}_i'\widehat{\Gamma}(\theta) - \phi \frac{(1-P_s)\widehat{\Gamma}(\theta)'\,\widehat{\Sigma}\widehat{\Gamma}(\theta)}{(1-P_s)\widehat{\Gamma}(\theta)\,\widehat{\Sigma}\widehat{\Gamma}(\theta) + P_s\sigma^2_\xi}\left(\widetilde{Y}_i - \alpha\widetilde{T}_i - \widetilde{W}_i'\widehat{\Gamma}(\theta)\right)\right] \\
(3.12)\quad q^2_{N,K^*}(\theta) =& \frac{1}{N}\sum_{i=1}^{N} \left(\left(\widetilde{Y}_i - \alpha\widetilde{T}_i - \widetilde{W}_i'\widehat{\Gamma}(\theta)\right)\right) \times \\
& \left[\widetilde{Z}_i - \phi \widetilde{W}_i'\widehat{\Gamma}(\theta) - \phi \frac{(1-P_s)\widehat{\Gamma}(\theta)'\,\widehat{\Sigma}\widehat{\Gamma}(\theta)}{(1-P_s)\widehat{\Gamma}(\theta)'\,\widehat{\Sigma}\widehat{\Gamma}(\theta) + P_s\sigma^2_\xi}\left(\widetilde{Y}_i - \alpha\widetilde{T}_i - \widetilde{W}_i'\widehat{\Gamma}(\theta)\right)\right] \\
(3.13)\quad q^3_{N,K^*}(\theta) =& \frac{1}{N}\sum_{i=1}^{N} \left(\widetilde{Y}_i - \alpha\widetilde{T}_i\right)^2 - \left(\frac{\widehat{\Gamma}(\theta)'\,\widehat{\lambda}}{P_s}\right)^2 - \frac{\widehat{\Gamma}(\theta)'\,\widehat{\Sigma}\widehat{\Gamma}(\theta)}{P_s} - \sigma^2_\xi
\end{aligned}
$$

subject to $\theta \in \bar{\Theta}$. We will show that when evaluated at θ_0 these equations converge to zero as N and K^* grow.

To understand the first two equations, note that when $\sigma^2_\xi = 0$ they reduce to

$$
\begin{aligned}
q^1_{N,K^*}(\theta) &= \frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{W}_i'\widehat{\Gamma}(\theta)\left[\widetilde{Z}_i - \phi\widetilde{W}_i'\widehat{\Gamma}(\theta) - \phi\left(\widetilde{Y}_i - \alpha\widetilde{T}_i - \widetilde{W}_i'\widehat{\Gamma}(\theta)\right)\right]\right) \\
q^2_{N,K^*}(\theta) &= \frac{1}{N}\sum_{i=1}^{N}\left(\left(\widetilde{Y}_i - \alpha\widetilde{T}_i - \widetilde{W}_i'\widehat{\Gamma}(\theta)\right)\left[\widetilde{Z}_i - \phi\widetilde{W}_i'\widehat{\Gamma}(\theta) - \phi\left(\widetilde{Y}_i - \alpha\widetilde{T}_i - \widetilde{W}_i'\widehat{\Gamma}(\theta)\right)\right]\right).
\end{aligned}
$$

These are the classic moment conditions of a linear regression of $\widetilde{Z}_i$ on $(\widetilde{W}_i'\widehat{\Gamma}(\theta))$ and $(\widetilde{Y}_i - \alpha\widetilde{T}_i - \widetilde{W}_i'\widehat{\Gamma}(\theta))$ when the two regression coefficients are restricted to be the same. They are the empirical analogue of Corollary 1 of Theorem 1. In the general case, the equations are more complicated because the presence of ξ leads to attenuation bias on the regression coefficient on $\left(\widetilde{Y}_i - \alpha\widetilde{T}_i - \widetilde{W}_i'\widehat{\Gamma}(\theta)\right)$.

When $P_S = 1$, the second equation reduces to

$$
q^2_{N,K^*}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(\left(\widetilde{Y}_i - \alpha\widetilde{T}_i - \widetilde{W}_i'\widehat{\Gamma}(\theta)\right)\left[\widetilde{Z}_i - \phi\widetilde{W}_i'\widehat{\Gamma}(\theta)\right]\right).
$$

In this case $\widehat{\Gamma}(\theta)$ could be estimated as the coefficient vector from a linear regression of $\left(\widetilde{Y}_i - \alpha\widetilde{T}_i\right)$ on $\widetilde{W}_i$. (Our estimator is asymptotically equivalent to this with K^* fixed and N getting large.) In that case, $\widetilde{W}_i'\widehat{\Gamma}(\theta)$ would have to be orthogonal to the error term, so this equation would reduce further to

$$q_N^2(\alpha,\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{Y}_i - \alpha\widetilde{T}_i - \widetilde{W}_i'\widehat{\Gamma}(\theta)\right) \times Z_i,$$

which is the standard IV moment equation.

Turning to (3.13), $q^3_{N.K^*}(\theta)$ is the difference between the sample value of $var\left(\widetilde{Y}_i - \alpha\widetilde{T}_i\right)$ for the hypothesized value of α and the variance implied by the model estimate.

We define the estimator $\widehat{\Theta}$ as the set of values of θ that minimize the criterion function

$$Q_{N,K^*}(\theta) = q_{N,K^*}(\theta)'\Omega q_{N,K^*}(\theta),$$

where

$$q_{N,K^*}(\theta) = \left[\begin{array}{ccc} q^1_{N,K^*}(\theta) & q^2_{N,K^*}(\theta) & q^3_{N,K^*}(\theta) \end{array}\right]'$$

and Ω is some predetermined positive definite weighting matrix.

3.3 Consistency of the Estimator

In this section we prove consistently using the standard methods from Chernozhukov, Hong, and Tamer (2007). Define $Q_0(\theta)$ as the probability limit of $Q_{N,K^*}(\theta)$ as N and K^* get large. Specifically we use sequential limits assuming that N grows faster than K^*. The identified set, Θ_I, is defined as the set of values that minimize $Q_0(\theta)$. We verify the conditions in Chernozhukov, Hong, and Tamer (2007) to show that the Hausdorff distance between $\widehat{\Theta}$ and Θ_I converges in probability to zero and that $\theta_0 \in \Theta_i$. Thus as the sample gets large our estimate of $\widehat{\Theta}$ will contain the true value with probability approaching 1.

Assumption 11. *F_i, ξ_i, and ψ_i are all mean 0 and i.i.d. across individuals and are independent of each other with finite second moments. ω_i is i.i.d. across individuals with finite second moments, is independent of F_i, but may be correlated with ξ_i and/or ψ_i. v_{ij} is mean zero and i.i.d. across individuals and covariates with finite variance. The vector $(\Gamma_j, \Lambda_j, \beta_j, \sigma_j^2)$ is i.i.d. across covariates with finite second moments.*

Assumption 12. *$\bar{\Theta}$ is compact with the support of P_s bounded below by $p_s^\ell > 0$.*

Assumption 13. *The dimension of F_i is 1*

Define $d_h(\cdot,\cdot)$to be Hausdorff distance as defined in Chernozhukov, Hong, and Tamer (2007).

Theorem 6. *Under Assumptions 11-13, $d_h(\widehat{\Theta}, \Theta_I)$ converges in probability to zero and $\theta_0 \in \Theta_I$.*

(Proof in Appendix)

One can form a set estimator for α_0 just by taking the projection of $\widehat{\Theta}$ onto α. That is, we can define this set as

$$\widehat{A} \equiv \left\{ \alpha : \text{ there exists some value of } (\phi, P_s, \sigma_\xi^2) \text{ such that } \{\alpha, \phi, P_s, \sigma_\xi^2\} \in \widehat{\Theta} \right\}$$

3.4 Constructing Confidence Intervals

In this section we discuss confidence interval construction. We start with the ideal procedure one would use given unlimited computing resources. We then discuss a more practical approach, which is the parametric bootstrap we use in the Monte Carlos below.

3.4.1 A General Procedure

Before discussing inference it is useful to step back and consider our basic approach. In terms of identification we have four parameters $(\alpha_0, \phi_0, P_S^0, \sigma_\xi^0)$ but only 3 equations: the population and limit of the sequence of models for (q_N^1, q_N^2, q_N^3).[13] However, we also have limits on the parameter space. In particular $0 < P_S \leq 1$ and $\sigma_\xi^0 \geq 0$. In principle, while we cannot get a point estimator for $(\alpha_0, \phi_0, P_S^0, \sigma_\xi^0)$, we construct the set estimator $\widehat{\Theta}$ for this four dimensional parameter. Our set estimate for α_0 is just the set of α that lie within this identified set.

We can construct a confidence region in the analogous manner. That is, we could first construct a confidence set for $(\alpha_0, \phi_0, P_S^0, \sigma_\xi^0)$ and then let our confidence set for α be the values of α that lie within this set. The most natural way to construct the larger confidence set would be to "invert a test statistic." That is, we would first construct a test statistic $T(\theta)$ which has a known distribution under the null hypothesis: $\theta = \theta_0$.[14] For each potential θ, we would construct an acceptance region of the test. When $T(\theta)$ lies within this acceptance region, θ would belong to this confidence set, otherwise it would not. Given the confidence set for the full parameter space, we take the confidence set to be the set of α that lie within

[13] In the definition of the estimator, we have not explicitly defined Λ,Γ,β, or Σ as parameters but the estimates of these objects as functions of the data and θ. The main reason is that the dimension of these objects grows with K^*so in terms of consistency and inference it is easier to focus on the elements of θ.

[14] A natural choice for a test statistic would be the objective function $Q_{N,K^*}(\theta)$.

this set. More formally let $T_{N,K^*}(\theta)$ be the estimated value of the test statistic and let $T^c(\theta)$ the critical value. Assuming we reject when the test statistic is larger than the critical value, the confidence set is defined as

$$\widehat{C}_{N,K^*} = \left\{\theta \in \Theta \mid \widehat{T}(\theta) \leq T^c(\theta)\right\},$$

and our estimated confidence region for α can be written as

$$\widehat{C}_\alpha = \left\{\alpha \in \mathbb{R} \mid (\alpha, \Theta) \cap \widehat{C}_N \neq \emptyset\right\}.$$

There are many test statistics one could use and many ways to calculate the critical value. We consider the following algorithm based on the bootstrap. Consider testing the null hypothesis $\theta = \theta_0$.The most natural test statistic is the normalized criteria function, so that

$$T_{N,K^*}(\theta) = KQ_{N,K^*}(\theta).$$

Such a test statistic would be computed as follows:

1. Estimate parameters to be used in generating data for the bootstrap. This involves using the data generation process for X_i as well. That is, from the joint distribution of (X_i, W_i),
   (a) Estimate (Λ, Λ_X), Σ, and the data generating processes for F_i and v_{ij}.
   (b) Estimate
   $$\begin{aligned}\frac{\widehat{\Gamma}(\theta)}{\sqrt{K^*}} &\equiv \left[\frac{1}{P_s K}\widehat{\lambda}'\widehat{\lambda} + \widehat{\Sigma}\right]^{-1} \frac{1}{N}\widetilde{W}'\left(\widetilde{Y} - \alpha\widetilde{T}\right) \\ \frac{\widehat{\beta}(\theta)}{\sqrt{K^*}} &\equiv \left[\frac{1}{P_s K}\widehat{\lambda}'\widehat{\lambda} + \widehat{\Sigma}\right]^{-1} \frac{1}{N}\widetilde{W}'\widetilde{Z}\end{aligned}$$
   (c) Given knowledge of P_S, estimate the distribution of $(\xi_i, \psi_i, \omega_i)$.

2. Generate N_B bootstrap samples. Where for each sample:
   (a) Draw K observable covariates from the actual set of covariates (with replacement) with appropriate $\left(\widehat{\Gamma}_j, \widehat{\beta}_j, \widehat{\lambda}_j, \widehat{\Sigma}_{jj}\right)$
   (b) Draw $(K^* - K)$ unobservable covariates from the actual set of covariates (with replacement) with appropriate $\left(\widehat{\Gamma}_j, \widehat{\beta}_j, \widehat{\lambda}_j, \widehat{\Sigma}_{jj}\right)$

(c) Now for $i = 1, N$ generate all of the (X_i, W_i^*) using the DGP for f_i and v_{ij}.

(d) Using the DGP for ψ_i and ξ_i generate Z_i and $(Y_i - \alpha_0 T_i)$ *(Note that we do not need to generate data on Y_{ii} and T_i themselves because only $\left(\widetilde{Y}_i - \alpha_0 \widetilde{T}_i\right)$ enters the moment conditions that define the test statistic.)*

(e) Given generated bootstrap data construct the test statistic $Q_{N,K^*}(\theta)$. (This involves the intermediate steps of estimating Σ, λ and Γas well.)

3. From the bootstrap sample we can estimate the distribution of the test statistic and calculate the critical value given the size of the test.

For this critical value to be correct, we need that the bootstrap distribution of $T_{N,K^*}(\theta_0)$ provides a consistent estimate of the actual distribution of $T_{N,K^*}(\theta_0)$.

It will prove useful to define

$$\chi_j = \begin{bmatrix} \Lambda_j \Gamma_j & \Lambda_j \beta_j & \Gamma_j \sigma_j^2 \Gamma_j & \Gamma_j \sigma_j^2 \beta_j & S_j \frac{\Lambda_j^2}{\sigma_j^2} & S_j \Gamma_j \Lambda_j & S_j \Gamma_j \Lambda_j \sigma_j^2 & S_j \beta_j \Lambda_j & S_j \beta_j \Lambda_j \sigma_j^2 & S_j \Gamma_j^2 \sigma_j^2 & S_j \end{bmatrix}'$$

and

$$\chi_0 = E\left(\chi_j\right).$$

Our next goal to show that the limit of $q_{N,K^*}(\theta_0)$ as N gets large is a known function of only θ and the mean of χ_j. This property will the asymptotic distribution straight forward to figure out. The proof is still in progress, so we write it as a congjecture.

Conjecture 7. *Under Assumptions 11-13, the bootstrap distribution of the test statistic is consistent.*

(Proof in progress)

In the appendix we present the algorithm one would use to implement this approach in practice.

3.4.2 A Parametric Bootstrap Procedure

In practice, implementing the procedure above in impractical because testing the null over a four dimensional grid is computationally difficult. Additionally, one often has a strong prior about the sign of the selection bias. We can obtain tighter bounds by imposing this prior (formally defined as "monotone selection" in Manski and Pepper, 2000). While our estimation interval can potentially be much more complicated, for the simulations we have run, we consistently find a compact region with one end of the region occurring at the instrumental variable estimate ($P_S = 1$) and the other occurring at the "observables like unobservables"assumption ($\sigma_\xi = 0$). Without loss of generality we will assume positive selection bias so that the upper bound occurs under the constraint $P_S = 1$. We will also assume that the minimum value occurs at σ_ξ. We propose a parametric bootstrap procedure to construct one-sided confidence interval estimators for the lower and upper bounds of this set, denoted $\alpha_{\min}$ and $\alpha_{\max}$, respectively. We construct these intervals such that the estimator $\hat{\alpha}_{.10,\min}$ has 10% nominal probability of being below $\alpha_{\min}$. The estimator $\hat{\alpha}_{.10,\max}$ has a 10% nominal probability of exceeding $\alpha_{\max}$.

3.4.3 Construction of $\hat{\alpha}_{.10,\min}$

The procedure for estimating $\hat{\alpha}_{.10,\min}$ involves the following steps.

1. Estimate the parameters under the model under the assumption that $\sigma_\xi = 0$. We do this by solving the system of equations

$$0 = q_N^1(\widehat{\alpha}_{min}, \widehat{\phi}, \widehat{P}_S, 0) = q_N^2(\widehat{\alpha}_{min}, \widehat{\phi}, \widehat{P}_S, 0) = q_N^3(\widehat{\alpha}_{min}, \widehat{\phi}, \widehat{P}_S, 0)$$

   for $\widehat{\alpha}, \widehat{\varphi}$, and $\widehat{P}_S$. In doing this we also obtain estimates of Λ, Σ, and γ for the observable covariates.

2. Next we need to estimate some additional parameters that will be used for generating the bootstrap sample.

   (a) Obtain estimates of the distributions for F_i, v_{ij} given the estimates of $[\hat{\Sigma}, \hat{\Lambda}_j]$. This can be done in a number of different ways. One could specify a parametric distribution and estimate the distribution parameters. Alternatively, one could do this completely nonparametrically. A third possibility is to take advantage of the fact that our estimator involves up to second moments of the variables, so only up to 4rth moments of the distributions of these variables matter for the sampling distribution of $\hat{\alpha}_{\min}$. Instead of specifying parametric distributions, one could use a method of moments procedure to estimate up to the fourth moments from sample estimates of $E(\widetilde{W}_{ij}^r \widetilde{W}_{ij'}^s)$ and $\hat{\sigma}_v, \hat{\Lambda}_j$, $j = 1..., K$ for various values of r and s. One could then pick convenient parametric distributions for θ_i and

v_{ij}, $j = 1, ..., K$ and choose parameters of the distributions to match the relevant moments.[15] Call the estimates of the additional parameters of the θ_i distribution $\hat{B}_\theta$ and the additional parameters of the v_{ij} distribution $\hat{B}_{v_j}$.[16]

(b) Next we need to estimate the distribution of $(\xi_i, \psi_i, \omega_i)$. We can use the same three approaches as in the previous case. To use the third we need estimates of fourth moments. To obtain them, one can use the fourth moments of $\widetilde{Y}_i - \hat{\alpha}\widetilde{T}_i$, $\widetilde{Z}_i$ and $\widetilde{T}_i$. Consider

$$E(\xi_i^4) = E(\widetilde{Y}_i - \alpha\widetilde{T}_i)^4 - E(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}\widetilde{W}_{ij}\Gamma_j)^4 - E(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}\widetilde{W}_{ij}\Gamma_j)^2\sigma_\xi^2.$$

We have the estimate of $\hat{\alpha}_{\min}$, so $E(\widetilde{Y}_i - \alpha\widetilde{T}_i)^4$ can be replaced with the corresponding sample moment. We also have estimates of $E(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}\widetilde{W}_{ij}\Gamma_j)^2$ and σ_ξ^2. One can use a similar procedure to estimate $E(\psi_i^4)$. The relevant moment condition is

$$E(\psi_i^4) = E(\widetilde{Z}_i)^4 - E(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}\widetilde{W}_{ij}\beta_j)^4 - E(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}\widetilde{W}_{ij}\beta_j)^2\sigma_\psi^2 \, .$$

Note that this requires an estimate of $\widehat{\beta}$ and σ_ψ^2, but estimating these is completely analogous to estimating $\widehat{\gamma}$ and σ_ξ^2 where the dependent variable is now $\widetilde{Z}_i$ rather than $\widetilde{Y}_i - \alpha\widetilde{T}_i$. Estimation of δ, σ_ω^2 and $E(\omega_i^4)$ is analogous. We would then pick convenient parametric distributions for this joint distribution, and estimate parameters $B_{\xi,\psi,\omega}$. The joint distribution should not constrain the second and fourth moments unless one wishes to impose additional a priori information (such as normality) on it. We leave implicit the fact that $\hat{B}_{\xi,\psi,\omega}$ depends on $\hat{\alpha}_{\min}$.

3. Construct the Bootstrap sample. This involves a few different steps.

[15] Sticking with the one factor case and taking W_{ij} to be mean zero, using independence of θ_i and the v_{ij}, and using the fact that $var(\theta_i) = 1$, the moments are $E(W_{ij}^4) = \Lambda_j^4 E(\theta_i^4) + E(v_{ij}^4) + 4\Lambda^2 1\sigma_{vij}^2$ and $E(W^2 ijW_{ij'}^2) = \Lambda_j^2\Lambda_{j'}^2 E(\theta_i^6) + \sigma_{vj}^2\sigma_{vj'}^2$ for all $j, j' \neq j$ pairs. The idea generalizes to the multiple factor case.

[16] An alternative is to use the K observed W_j, impose the estimates $\hat{\Lambda}_j$ and the estimates of $\hat{\sigma}_{vj}$, choose parametric distributions for θ_i $v_{1i}, ..., v_{Ki}$, and fit the parameters of those distributions. The chosen distributions should not impose constraints on the second and fourth moments. In principle, one could work with nonparametric distributions with the variance constrained to match the σ_{vj}^2. A nonparametric approach is unattractive from a computational point of view – given that our estimators only involve second moments, it does offer any clear advantages.

(a) Using the estimates $[\hat{\beta}_j, \hat{\Gamma}_j, \hat{\sigma}_v, \hat{\Lambda}_j, \hat{B}_j]$, $j = 1, ..., K$, and the estimates $\hat{P}_S$, draw $\hat{K}^*$ values of $[\hat{\beta}_j, \hat{\Gamma}_j, \hat{\sigma}_{vj}, \hat{\Lambda}_j, \hat{B}_j]$ by sampling with replacement from the K estimated values. Let the first K correspond to the "observed" $W's$ for purposes of the bootstrap replication.

(b) Using $(\hat{\sigma}_{vj}, \hat{\Lambda}_j, \hat{B}_j)$ and $\hat{B}_\theta$, generate $(f_i)^{(b)}$, $(v_{ij})^{(b)}$ and then $W_{ij}^{(b)}$, $i = 1...N.$, $j = 1, ..., \hat{K}^*$ where (b) denotes the bth bootstrap replication, $(b) = 1, ..., N_{boot}$.

(c) Using the $\hat{K}^*$ values of $\hat{\beta}_j$, the associated K^* vectors $W_{ij}^{(b)}$, $\hat{\alpha}_{\min}$, and the draws of $\psi_i^{(b)}$, use $\widehat{B}_{\xi,\psi,\omega}$ to generate N values of $(Z_i^{(b)}, T_i^{(b)}, Y_i^{(b)})$.

4. For each bootstrap sample compute $\hat{\alpha}_{\min}^{(b)}$ by solving

$$0 = q^1_{N^{(b)}}(\widehat{\alpha}^{(b)}_{min}, \widehat{\varphi}, \widehat{P}_S, 0) = q^2_{N^{(b)}}(\widehat{\alpha}^{(b)}_{min}, \widehat{\varphi}, \widehat{P}_S, 0) = q^3_{N^{(b)}}(\widehat{\alpha}^{(b)}_{min}, \widehat{\varphi}, \widehat{P}_S, 0)$$

on the bootstrap samples.

5. Calculate the 90^{th} quantile of the bootstrap sample of $\widehat{\alpha}_{min}$ and subtract the different between that and our point estimate from our point estimate of $\widehat{\alpha}_{min}$ to obtain the lower bound of our confidence set.

3.4.4 Construction of $\hat{\alpha}_{.90,\max}$

To obtain $\hat{\alpha}_{.90,\max}$, we assume that the largest value of $\hat{\alpha}$ that satisfies the restrictions of the model is obtained when one imposes the assumption that $\hat{P}_S = 1$ and ignores the possibility that unobserved $\widetilde{W}_j$ that induce positive correlation between $\widetilde{T}_i$ and $\widetilde{Y}_i$. If one sets $\hat{P}_S$ to 1 in the matrix $\left[\frac{1}{\hat{P}_S \cdot K}\widehat{\lambda}'\widehat{\lambda} + \widehat{\Sigma}\right]$ and replaces the matrix with $\widetilde{W}'\widetilde{W}$ in equation 3.10) for $\Gamma(\hat{\theta})$, then the solution for $\hat{\alpha}$ is IV. Under the null, all of the W_j are observed. Thus we do not need to impose a model of how the W_j are related to each other to account for the effects of missing W_j. One can construct the one sided confidence interval estimate using the appropriate robust standard error estimator given assumptions about serial correlation and heteroskedasticity in ξ_i. Alternatively, one can use a conventional bootstrap procedure.

While the simplicity of the above approach is attractive, it has an important shortcoming. We have not been able to prove that OLS is the upper bound when P_S is less than 1 $Cov(W, \varepsilon) \neq 0$. This is because bias in $\hat{\Gamma}$ may lead to a partially offsetting bias in $\hat{\alpha}$.

4 Monte Carlo Evidence (very preliminary)

In this section we present Monte Carlo evidence on the performance of $\hat{\alpha}_{\min}$, which we estimate based on $\hat{\alpha}_{OU-Factor}$, and $\hat{\alpha}_{\max}$, which we estimate based on $\hat{\alpha}_{OLS}$ because in our context $\hat{\alpha}_{\max}$ turns out to be essentially the same as the OLS estimator. We also present evidence on the performance of $\hat{\alpha}_{OU}$.[17]

In discussing the design, we first restate the equations of the model of Y_i, T_i, and W_{ij}:

$$
\begin{aligned}
Y_i &= \alpha_0 T_i + \frac{1}{\sqrt{K^*}} \sum_{j=1}^{K^*} W_{ij}\Gamma_j + \xi_i \\
&= \alpha_0 T_i + \frac{1}{\sqrt{K^*}} \sum_{j=1}^{K_o} W_{ij}\Gamma_j + \frac{1}{\sqrt{K}} \sum_{j=K_o+1}^{K^*} W_{ij}\Gamma_j + \xi_i \\
W_{ij} &= \frac{1}{\sqrt{K^*}} \theta_i' \Lambda_j + v_{ij} \\
T_i &= Z_i = \frac{1}{\sqrt{K^*}} \sum_{j=1}^{K^*} W_{ij}^K \beta_j + \psi_i
\end{aligned}
$$

We focus on the case in which θ is a scalar $(r = 1)$. We vary assumptions about $P_S = K/K^*$, the fraction of the W_{ij} variables that are included in the model.

4.1 W parameters

The distributions of the variables that determine W_{ij} are

$$
\begin{aligned}
&\theta_i \sim N(0,1) \\
&v_{ij} \sim N(0, \sigma_{vj}^2);\ \ \sigma_{vj} \sim U(1.0, 2.0) \\
&\Lambda_j = \bar{\Lambda} + \tilde{\Lambda}_j \\
&\tilde{\Lambda}_j \sim U(-\tilde{\Lambda}_{\max}, \tilde{\Lambda}_{\max})
\end{aligned}
$$

For this specification,

[17] The OLS estimator is essentially the same as the estimate of α based on our moment equations with P_S set to 1. The two differ because we use the moments implied by the estimated factor structure rather than the actual variance covariance matrix of W in the moment condition for $\hat{\Gamma}$. In the designs we consider we found that the maximum value of $\hat{\alpha}$ consistent with $\sigma_\xi^2 > 0$ occured at $P_S = 1$, although we have not proved that this has to be the case for any model with a factor structure.

$$\begin{aligned}E[Cov(W_j, W_{j'})|j &\neq j'] = \frac{1}{K^*}E(\Lambda_j \Lambda_{j'}) = \frac{1}{K^*}\bar{\Lambda}^2 \text{ and} \\ E[Var(W_j)] &= \frac{1}{K^*}\bar{\Lambda}^2 + \frac{1}{3K^*}[\tilde{\Lambda}_{\max}]^2 + E(\sigma_{vj}^2),\end{aligned}$$

where the expectations are defined over j and j'. We report $\frac{E[Cov(W_j,W_{j'})]}{E[Var(W_j)]}$ in the tables below.

4.2 Parameters of the Y_j and T_j Equations

Γ_j and β_j have expected values μ_Γ and μ_β, respectively, and depend on a common component ε_j and the components $\varepsilon_{\Gamma j}$ and $\varepsilon_{\beta j}$ that are specific to Γ_j and β_j. They are determined by

$$\begin{aligned}\Gamma_j &= \mu_\Gamma + \frac{g_\varepsilon}{[g_\varepsilon^2 + (1-g_\varepsilon)^2]^{.5}}\varepsilon_j + \frac{(1-g_\varepsilon)}{[g_\varepsilon^2 + (1-g_\varepsilon)^2]^{.5}}\varepsilon_{\Gamma j} \\ \beta_j &= \mu_\beta + \frac{b_\varepsilon}{[b_\varepsilon^2 + (1-b_\varepsilon)^2]^{.5}}\varepsilon_j + \frac{(1-b_\varepsilon)}{[b_\varepsilon^2 + (1-b_\varepsilon)^2]^{.5}}\varepsilon_{\beta j},\end{aligned}$$

where ε_j, ε_{Γ_j}, and ε_{β_j} are uniform random variables with mean 0 and variance 1. They are mutually independent and independent across j.

The parameters g_ε and b_ε determine relative weights on ε_j and the idiosyncratic terms $\varepsilon_{\Gamma_j}, \varepsilon_{\beta_j}$, thereby determining the covariance between Γ_j and β_j. We have normalized the weights so that $var(\Gamma_j) = var(\beta_j) = 1$ regardless of the choice of g_ε and b_ε. g_ε^2 and b_ε^2 are the shares of the variances accounted for by the common component ε_j, respectively. For the above design,

$$\begin{aligned}E(\Gamma_j \cdot \beta_{j'}) &= \mu_\Gamma\mu_\beta + \frac{g_\varepsilon \cdot b_\varepsilon}{[g_\varepsilon^2 + (1-g_\varepsilon)^2]^{.5} \cdot [b_\varepsilon^2 + (1-b_\varepsilon)^2]^{.5}} \text{ , } j = j' \\ &= \mu_\Gamma\mu_\beta \text{ , } j \neq j'\end{aligned}$$

$$\begin{aligned}cov(\Gamma_j, \beta_{j'}) &= corr(\Gamma_j, \beta_{j'}) = \frac{g_\varepsilon \cdot b_\varepsilon}{[[g_\varepsilon^2 + (1-g_\varepsilon)^2]^{.5} \cdot [b_\varepsilon^2 + (1-b_\varepsilon)^2]^{.5}]}, \, j = j' \\ &= 0, \, j \neq j'.\end{aligned}$$

$$\begin{aligned}E(\Gamma_j \cdot \Gamma_{j'}) &= \mu_\Gamma\mu_\Gamma + 1, \, j = j' \\ &= \mu_\Gamma\mu_\Gamma, \, j \neq j' \\ E(\beta_j \cdot \beta_{j'}) &= \mu_\beta\mu_\beta + 1, \, j = j' \\ &= \mu_\beta\mu_\beta, \, j \neq j'\end{aligned}$$

Below we consider the effects of varying g_ε and b_ε, and we also consider a case in which $\beta_j = 0$ for all j.

4.3 Additional Parameter Values

We also examine the sensitivity of the estimates to the importance of ψ and ξ, the idiosyncractic components of T and Y, respectively. To do this, we vary σ_ξ^2 so as to vary the expected fraction of the variance of the unobservable component of Y that is due to ξ. That is, we choose σ_ξ^2 to manipulate

$$R_\xi^2 \equiv E\left[\sigma_\xi^2/(\frac{1}{K^*}Var(\sum\nolimits_{j=K_0+1}^{K^*} W_j\Gamma_j|\Gamma) + \sigma_\xi^2)\right],$$

where the expectation is defined over the joint distribution of Γ, β, and W. Similarly, we set σ_ξ^2 to control

$$R_\psi^2 \equiv E\left[\sigma_\psi^2/(\frac{1}{K^*}Var(\sum\nolimits_{j=1}^{K^*} W_j\beta_j|\beta) + \sigma_\psi^2)\right].$$

We report R_ψ^2 and R_ξ^2 in the tables below. Note that for a given value of R_ξ^2, the value of σ_ξ^2 will depend on the choice of P_S, but ϕ and ϕ_ε will not. We view this as an attractive parameterization because we are primarily concerned with ensuring that ϕ and ϕ_ε do not depend on P_S.[18] The expected values of ϕ and ϕ_ε at the true α are complicated functions of the parameters of the data generation process, so we simply compute the average values in each design as well as the average estimate of $\hat{\phi}$ at $\hat{\alpha}_{\min}$.

For all experiments, we set $N = 2000$ and report results based on 1000 Monte Carlo replications. The bootstrap estimates of the .10 one sided confidence interval estimate is based on 1000 bootstrap replications for each Monte Carlo replication. We set K^* to 100, R_ψ^2 to 0.5, and α_0 to 1.0 in all the experiments reported, and we vary P_S, R_ξ^2, $\bar{\Lambda}$, $\tilde{\Lambda}_{\max}$, μ_B, μ_Γ, g_ε, and b_ε across experiments. Specifically, we set P_S of 0.2, 0.4, and 0.8 and we set R_ξ^2 to

[18] If we fix $Var(\xi_i)$ at a nonzero value, the ratio ϕ_ε/ϕ approaches 0 (the case in which OLS is unbiased) as P_S approaches 1. In assessing how variation in P_S matters, we wish to hold constant the degree to which selection on observables is similar to selection on unobservables. For each Monte Carlo experiment we set σ_ψ^2 and σ_ξ^2 to the fixed values

$$\begin{aligned}
\sigma_\xi^2 &= E\left[\frac{R_\xi^2}{1-R_\xi^2}\frac{1}{K^*}Var(\sum\nolimits_{j=K_0+1}^{K^*} W_j\Gamma_j|\Gamma)\right] \\
\sigma_\psi^2 &= E\left[\frac{R_\psi^2}{1-R_\psi^2}\frac{1}{K^*}Var(\sum\nolimits_{j=K_0+1}^{K^*} W_j\beta_j|\beta)\right]
\end{aligned}$$

given the values of the other parameters of the experiment.

0, 0.2, and 0.4. We vary μ_B, μ_Γ, g_ε, and b_ε such that $E(\beta_j \Gamma_j) = 0.09$, 0.3, and 0.6. Finally, we vary $\bar{\Lambda}$ and $\tilde{\Lambda}_{\max}$. In one case, we set $\bar{\Lambda} = 0$, which means that $E[Corr(W_{ij}, W_{ij'})] = 0$ if $j \neq j'$. In the other case, $E[Corr(W_{ij}, W_{ij'})] = 0.2$ if $j \neq j'$.

4.4 Monte Carlo Results

We first consider a baseline case in which T_i is randomly assigned. Table MC1 reports results for a design in which $\beta_j = 0$ for all j ($\mu_\beta = 0$, $var(\varepsilon_{\beta j}) = 0$, and $b_\varepsilon = 0$), which means that T does not depend on the W_j. For these designs, $\hat{\alpha}_{OLS}$ is unbiased because $E(\phi) = E(\phi_\varepsilon) = 0$. We use the median as our measure of central tendency but also report the 10th and 90th percentile values. We use the 90th-10th differential as a measure of dispersion. The median values of ϕ, ϕ_ε, and $\widehat{\phi}$ across replications are shown in the three rows of the table.

The estimates of $\hat{\alpha}_{OLS}$ are tightly distributed around 1.0 in all three cases. The dispersion declines with P_S, reflecting a smaller variance of the unobserved components of Y as P_S increases. The values of $\hat{\alpha}_{OU}$ and of $\hat{\alpha}_{\min}$ are also tightly distributed around 1.0, although they are estimated less precisely than the OLS coefficients. When $P_S = 0.2$, the 90th-10th differential of $\hat{\alpha}_{\min}$ is roughly double that of the 90th-10th differential for $\hat{\alpha}_{OLS}$, but when $P_S = 0.8$, the three estimators have similar dispersion. The results are not very sensitive to the value of P_s.

We turn next to designs in which OLS estimates of α_0 are biased. In Table MC2a, we set $\mu_\beta = \mu_\Gamma = 0.3$, which leads to bias OLS estimates for the specification we consider. To see this, note that even if $b_\varepsilon = g_e = 0$, so that the elements of β_j and Γ_j are uncorrelated, OLS will be biased if $P_S < 1$ because $E(\beta_j \Gamma_j) = 0.09$. In the top panel of the table, $\bar{\Lambda} = 0$, so that $E[Corr(W_{ij}, W_{ij'})] = 0 \ \forall \ j \neq j'$. We consider the $b_\varepsilon = g_e = 0$ case in the first three columns of the table. In the first column, with $P_S = 0.2$, ϕ and ϕ_ε are small. (The median of $\phi = 0.043$ and the median of $\phi_\varepsilon = 0.041$. For this design $\phi = \phi_\varepsilon$ and both are positive, so the difference reflects sampling error.) The bias in OLS in this case is small regardless of the value of P_S. The precision of the OLS estimator is also essentially invariant to the value of P_S.[19] In contrast, the performance of the $\hat{\alpha}_{OU}$ and $\hat{\alpha}_{\min}$ estimators improves as P_S increases. $\hat{\alpha}_{OU}$ exhibits some downward bias when $P_S = 0.2$, but $\hat{\alpha}_{\min}$ is approximately unbiased in all cases. $\widehat{\alpha}_{\min}$ and $\widehat{\alpha}_{OU}$ are noisier than OLS but not dramatically so when $P_S = 0.8$.

[19] It is surprising that the bias in OLS is not monotone decreasing in P_S. Sampling error may be the reason, but phenomen shows up in several places in the tables.

In the next three columns of the table, we chose b_ε and g_ε so that $E(\Gamma_j \beta_j) = 0.3$. Not surprisingly, the upward bias in OLS is higher than the corresponding cases in the first three columns of the table, with the median of $\widehat{a}_{OLS}$ rising to 1.256 when $P_S = 0.2$ and 1.101 when $P_S = 0.8$. Again, $\hat{\alpha}_{\min}$ is essentially unbiased in all three cases, with the dispersion declining with P_S. The last three columns increase b_ε and g_ε so that $E(\Gamma_j \beta_j)$ to 0.6 and $Corr(\Gamma_j, \beta_j) = .51$. For each value of P_S, the bias in OLS increases relative to the cases in which $E(\Gamma_j \beta_j) = 0.3$. Interestingly, the $\widehat{\alpha}_{OU}$ and $\widehat{\alpha}_{\min}$ estimators are less noisy as $E(\Gamma_j \beta_j)$ increases. When $E(\Gamma_j \beta_j) = 0.6$ and $P_S = 0.8$ (column 6) shown in the last column, the $\widehat{\alpha}_{OU}$ and $\widehat{\alpha}_{\min}$ estimators have no more sampling error than the OLS estimator.

Table MC2b repeats the calculations found in Table MC2a but introduces a factor structure such that $E[Corr(W_{ij}, W_{ij'})] = 0.2$ if $j \neq j'$. We impose this correlation by setting $\bar{\Lambda}$ to 3.4. In order to keep $E[Var(W_{ij})]$ constant relative to the $\bar{\Lambda} = 0$ case, we reduce $\widetilde{\Lambda}_{\max}$ from 6.2 to 2.0. The bias in OLS tends to be lower for this design, perhaps because the regressors that are included do a better job of controlling for the omitted W_j when the correlation among the W_j is higher. Intuitively, as $E[Corr(W_{ij}, W_{ij'})] \rightarrow 1$, it does not matter which regressors are actually observed and which are not. The increase in the correlation across W_j that comes from θ is associated with an improvement in the performance of $\hat{\alpha}_{\min}$ relative to $\hat{\alpha}_{AET}$. In particular, $\hat{\alpha}_{OU}$ is substantially downward biased unless $E(\Gamma_j \beta_j) = 0.6$ or $P_S = 0.8$. This may be due to fact that the $\hat{\alpha}_{OU}$ estimator is based on the assumption that the restriction $\phi = \phi_\varepsilon$ based on the true Γ_j carries over to the coefficient vector Γ^P of the projection of $Y_i - \alpha_i T$ on the observables W_i. The positive correlation between the observed and unobserved covariates that is present in these designs results in positive omitted variables bias on the observed $\hat{\Gamma}_j$. The bias arises because the unobserved covariates are positively correlated with Y. Since the observed covariates are also positively correlated with T in these designs, the positive bias on the estimates of Γ_j may lead the projection of T on $W_i \Gamma^P$ to overstate the amount of selection bias, inducing a negative bias in the AET estimates of α_0. This negative bias also affects the OLS estimator, partially counteracting the positive bias caused by correlation of T with the unobserved elements of W. This is why the positive bias on the OLS estimates is smaller in Table MC2b than in Table MC2a.

Most importantly, $\hat{\alpha}_{\min}$ performs very well in the presence of a factor structure. It had a median close to 1 in all cases and a 90th-10th differential that is similar to OLS in the cases in which $E(\Gamma_j \beta_j) = 0.3$ or 0.6. This superior performance of $\hat{\alpha}_{\min}$ relative to $\hat{\alpha}_{OU}$ is due to the fact that explicitly accounting for the factor structure eliminates the positive bias on the estimates of Γ_j, which in turn eliminates the negative bias in the estimate of α_0.

In Table MC3a, we relax the assumption that the observables are a random set of all the unobservables by setting $R^2_\xi = 0.2$. In the top panel, $\bar{\Lambda} = 0$ and $\widetilde{\Lambda}_{\max} = 6.2$, as in Table MC2a. Not surprisingly, allowing a positive variance for ξ has no effect on the median of OLS. However, the lower bound estimators $\hat{\alpha}_{OU}$ and $\hat{\alpha}_{\min}$ are now both downward biased for α because the assumption that $\phi = \phi_\varepsilon$ no longer holds. This is easiest to see in the three cases in which P_S equals 0.8; in all three cases ϕ_ε is approximately equal to 0.8ϕ; in other words, selection on unobservables is now only 80 percent as large as selection on observables. When $E(\Gamma_j\beta_j) = 0.3$, the median of $\hat{\alpha}_{OU}$ varies from 0.907 to 0.975 depending on P_S, and the corresponding median values of $\hat{\alpha}_{\min}$ are 0.976, 0.956, and 0.979. However, the sampling variance of the $\hat{\alpha}_{AET}$ and $\hat{\alpha}_{\min}$ estimators is quite wide when P_S is small. When we increase b_ε and g_ε so that $E(\Gamma_j\beta_j) = 0.6$, the positive bias in OLS increases, as in table MC2a, while there is no systematic change for the other estimators. The sampling variances of $\hat{\alpha}_{OU}$ and $\hat{\alpha}_{\min}$ are wider in this case than in the analogous cases in Table MC2a (in which the assumption $\phi = \phi_\varepsilon$ holds.). We do not fully understand this pattern, but in spite of it, the lower bound estimators usefully complement OLS.

Table MC3b again allows for correlation among the elements of W_j by setting $\bar{\Lambda}$ and $\widetilde{\Lambda}_{\max}$ so that $E[Corr(W_{ij}, W_{ij'})] = 0.2$. Relative to the iid case, the performance of $\hat{\alpha}_{\min}$ improves substantially, with median values that are close to 1.0 for all cases. The sampling distribution narrows substantially, perhaps reflecting the fact that when the W_j are correlated, it is easier to "fill in" for the effects of missing covariates using our moment conditions, so that it matters less which elements of W^* are actually observed. Relative to the values in Table MC3a, the negative bias of the $\hat{\alpha}_{OU}$ estimator increases and the positive bias of the $\hat{\alpha}_{OLS}$ declines, again reflecting positive correlation between the observed and unobserved elements of W^*.

Finally, Tables MC4a and MC4b are analogous to tables MC3a and MC3b, except now we set $R^2_\xi = 0.4$, thereby lowering ϕ_ε relative to ϕ. The median of OLS is essentially unchanged relative to the cases in which R^2_ξ is 0 or 0.2, which is not surprising. The performance of $\hat{\alpha}_{OU}$ is poor in all three cases in which $P_S = 0.2$, with large sampling errors and negative bias. The medians of $\widehat{\alpha}_{\min}$ range between 0.786 and 0.982, but this estimator is noisy relative to OLS except when $P_S = 0.8$ and $E(\Gamma_j\beta_j) = 0.6$. As we saw earlier in a comparison of Tables MC3a and MC3b, the performance of $\widehat{\alpha}_{\min}$ improves substantially when $E[Corr(W_{ij}, W_{ij'})] = 0.2$.

There appears to be negative bias in all cases, but this bias is typically small relative to the positive bias in $\widehat{\alpha}_{OLS}$. The negative bias in $\widehat{\alpha}_{OU}$ is substantial in most cases, reflecting the fact that $\phi > \phi_\varepsilon$ as well as the positive correlation between the observed and unobserved elements of W.

The Monte Carlo results may be summarized as follows. First, the median of $\widehat{\alpha}_{\min}$ and $\widehat{\alpha}_{OU}$ are similar when there is no factor structure, although $\widehat{\alpha}_{\min}$ is less dispersed, particularly when $P_S = .2$. $\widehat{\alpha}_{\min}$ performs much better than $\widehat{\alpha}_{OU}$ when there a factor structure, although in some unreported experiments we have found that the estimators perform similarly. Second, both $\widehat{\alpha}_{\min}$ and $\widehat{\alpha}_{OU}$ are biased down when $\phi > \phi_\varepsilon$. This is to be expected, because both estimators are based on the assumption that $\phi = \phi_\varepsilon$ and are to be interpreted as lower bound estimators if $\phi > \phi_\varepsilon > 0$ (in the case $\phi > 0$). Third, the downward bias in $\widehat{\alpha}_{\min}$ when $\phi > \phi_\varepsilon$ is reduced considerably when there is a factor structure, at least in the cases we consider. Fourth, precision is worse than with OLS. The loss of precision depends on the design and is negligible in the case in which T is randomly assigned (Table MC1). However, $\widehat{\alpha}_{\min}$ is sufficiently precise to provide useful information about α in all of the cases that we consider.

5 Conclusion

In many situations, exclusion restrictions, functional form restrictions, or parameter restrictions are not sufficiently well grounded in theory or sufficiently powerful to provide a reliable source of identification. What can one do?

As we noted in the introduction, it is standard procedure to look for patterns in the relationship between an explanatory variable or an instrumental variable and the observed variables in the model when considering exogeneity. We provide a theoretical foundation for thinking about the degree of selection on observed variables relative to unobserved variables, and we propose two estimators that make explicit use of the pattern of selection in the observables to bound the treatment effect. We contrast the standard IV or OLS assumption that the researcher has chosen the control variables so that the instrument (or the treatment itself) are not related to the unobservables with the assumption that the control variables are randomly chosen from the full set variables that influence the outcome, and argue that the truth is likely to lie somewhere in between.

Our estimators build on Theorem 1, which concerns the coefficients of the projection of an outcome on the regression indices of the observables and the unobservables. A number of assumptions are required, but roughly speaking, the theorem says that when the number of observed and unobserved variables that influence the outcome are large, the coefficient on the index of unobservables will lie between 0 and the coefficient on the index of observables. Both OU and the $OU - Factor$ estimators identify bounds by imposing the inequality restriction on the econometric model for the outcome. However, in the likely case that the observed and unobserved variables are related, the coefficients on the control variables will to suffer from omitted variables bias, invalidating the restriction and the case for bounds. The OU estimator combines Theorem 1 with a high level assumption about the link among the observed and unobserved variables. The $OU - Factor$ estimator adds the assumption that the observed and unobserved explanatory variables have a factor structure, which provides additional moment restrictions that permit one to account for the effects of omitted variables. We show that the estimator identifies a set that asymptotially contains the true value of the treatment parameter. We derive the asymptotic distribution of the $OU - Factor$ estimator and present a parametric bootstrap approach to statistical inference. Our Monte Carlo simulations are generally encouraging, particularly for $OU - Factor$.

There is a very long research agenda. More Monte Carlo evidence is needed in the context of real world applications and data sets. Thus far we have not applied the $OU - Factor$ estimator, and we have not performed Monte Carlo studies for designs with multiple factors. The OU estimator has the advantage of simplicity and has already been used in a number of applications. However, a way to account for randomness in which explanatory variables are included in W when constructing confidence intervals is needed. Ultimately, we believe that incorporating a formal model of the relationships among the observed and unobserved variables in W^c is the more promising long-run research path. The linear factor model that we employ in developing the $OU - Factor$ estimator is a natural way to do this, but it is also restrictive. Other models of the joint distribution of the covariates should be explored. We only touch upon the case of heterogeneous treatment effects and so far we have only considered models in which the index that determines the outcome is an additively separable function.

More generally, we think of OU and $OU - Factor$ as a start for an investigation into a broader class of estimators based on the idea that if one has some prior information about how the observed variables were arrived at, then the joint distribution of the outcome, the treatment variable, the instrument, and the observed explanatory variables are informative about the distribution of the unobservables.

In closing, we caution against the potential for misuse of the idea of using observables to draw inferences about selection bias, whether through an informal comparison of means or through the estimators we propose. The conditions required for Theorem 1 imply that it is dangerous to infer too much about selection on the unobservables from selection on the observables if the observables are small in number and explanatory power, or if they are unlikely to be representative of the full range of factors that determine an outcome.

References

Altonji,Joseph G., 1988."The Effects of Family Background and School Characteristics on Education and Labor Market Outcomes," unpublished manuscript, Northwestern University, 1988.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, 2002. "The Effectiveness of Catholic School," unpublished manuscript, Northwestern University.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, 2005(a)."Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy,* 13(1): 151-84.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, 2005(b)."An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling", *Journal of Human Resources*, 40(4): 791-821.

Angrist J., and W. Evans, 1998. "Children and their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size" *American Economic Review*, 88, 450-477.

Angrist, Joshua D., and Alan B. Krueger, 1999."Empirical Strategies in Labor Economics," *Handbook of Labor Economics Vol. 3A,* Ashenfelter and Card (eds.), North Holland.

Bronars, Stephen G. , and Jeff Grogger, "The Economic Consequences of Unwed Motherhood: Using Twins as a Natural Experiment," *American Economic Review* 80(1994), 1141-56.

Cameron, Stephen V. and Heckman, James J., 1998."Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy,* 106, 262-333.

Cameron, Stephen V. and Christopher R. Taber, 2004."Estimating Borrowing Constraints Using the Returns to Schooling," *Journal of Political Economy*, 112(1), 132-182.

Chernozhukov, V., H. Hong, and E. Tamer, 2007. "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243-1284.

Coleman, James S., Thomas Hoffer, and Sally Kilgore, 1982. *High School Achievement: Public, Catholic, and Private Schools Compared* (New York, NY: Basic Books, Inc.).

Coleman, James S., and Thomas Hoffer, 1987. *Public and Private Schools: The Impact of Communities* (New York, NY: Basic Books, Inc., 1987).

Currie, Janet, and Thomas Duncan, 1990. "Does Head Start Make a Difference?" *American Economic Review*, 85, 341-64.

Engen, Eric, William Gale, and John Karl Sholz, 1996. "The Illusory Effects of Saving Incentives on Saving," *Journal of Economic Perspectives,* 10, 113-138.

Evans, William N., and Robert M. Schwab, 1995. "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *Quarterly Journal of Economics,* 110, 947-974.

Goldberger, Arthur S., and Glen C. Cain, 1982. "The Causal Analysis of Cognitive Outcomes in the Coleman, Hoffer and Kilgore Report," *Sociology of Education*, LV, 103-122.

Grogger, Jeff, and Derek Neal, 2000. "Further Evidence on the Benefits of Catholic Secondary Schooling," *Brookings-Wharton Papers on Urban Affairs*, 151-193.

Heckman, James .J., 1990. "Varieties of Selection Bias," *American Economic Review,* 80.

Heckman, J., and Robb, R., 1985. "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer eds.,*Longitudinal Analysis of Labor Market Data*. Cambridge, Cambridge University Press.

Imbens, G., and Angrist, J., 1994. "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-75.

Jacobsen, Joyce P., James W. Pearce III, and Joshua L. Rosenbloom, 1999. "The Effect of Childbearing on Married Women's Labor Supply and Earnings," *Journal of Human Resources* 34(3), pp. 449-474.

Manski, C., 1989. "Anatomy of the Selection Problem," *Journal of Human Resources,* 24, 343-360.

Manski, C., 1994. "The Selection Problem," in C. Sims (ed) *Advances in Econometrics: Sixth World Congress* (Cambridge: Cambridge University Press).

Manski, C., and J. Pepper, 2000."Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica,* 68(4), 997-1010.

McLeish, D. L., 1975. "A Maximal Inequality and Dependent Strong Laws," *The Annals of Probability*, 3, 829-839.

Murnane, Richard J., 1984."A Review Essay—Comparisons of Public and Private Schools: Lessons from the Uproar," *Journal of Human Resources* 19, 263–77.

Murphy, Kevin M., and Robert H. Topel, 1990."Efficiency Wages Reconsidered: Theory and Evidence," in Y. Weiss and R. Topel eds., *Advances in the Theory and Measurement of Unemployment.* New York, St. Martin's Press, 204-40.

Neal, Derek, 1997."The Effects of Catholic Secondary Schooling on Educational Attainment," *Journal of Labor Economics* 15, 98-123.

Poterba, James, Steven Venti, and David Wise, 1994."Targeted Retirement Saving and the Net Worth of Elderly Americans," *American Economic Review,* 84, 180-185.

Rosenbaum, Paul R., 1995. *Observational Studies*, Springer-Verlag, New York, NY.

Staiger, Douglas, and James Stock, 1997."Instrumental Variables Regression with Weak Instruments," *Econometrica,* 65(3), 557-586.

Stock, James, 1994."Unit Roots, Structural Breaks and Trends," *Handbook of Econometrics, Volume 4,* Engle and Mcfaddend eds., Elsevier Science, 2740-2841.

Udry, Christopher, 1996."Gender, Agricultural Production, and the Theory of the Household", *Journal of Political Economy*, 104, 1010-1046.

White, Halbert, 1984. *Asymptotic Theory for Econometricians,* Academic Press, Inc.

Table MC1
Design: Z Randomly Assigned (all β terms=0)

	Factor structure such that $E((Cov(W_j,W_{j'}))=0$		
	P_S=0.2	P_S=0.4	P_S=0.8
Median of φ	0.004	-0.003	-0.002
Median of φ_ε	0.001	-0.001	0.001
Median of estimated φ at α_{min}	0.002	-0.002	-0.002
α_{max}			
10th percentile	0.943	0.986	0.988
Median	1.005	1.002	1.000
90th percentile	1.049	1.017	1.014
α_{OU}			
10th percentile	0.911	0.981	0.985
Median	1.000	1.006	1.000
90th percentile	1.092	1.021	1.013
α_{min}			
10th percentile	0.909	0.981	0.984
Median	1.001	0.994	1.000
90th percentile	1.117	1.010	1.014

Notes: In all specifications, the true value of α=1, E(Γ)=0, all β terms=0, N=2000, and K*=100.

Table MC2a

A: Factor structure such that $E((Corr(W_j,W_{j'}))=0$

	$E(\beta^*\Gamma)$=0.09			$E(\beta^*\Gamma)$=0.3			$E(\beta^*\Gamma)$=0.6		
	P_S=0.2	P_S=0.4	P_S=0.8	P_S=0.2	P_S=0.4	P_S=0.8	P_S=0.2	P_S=0.4	P_S=0.8
Median of φ	0.043	0.058	-0.016	0.432	0.453	0.453	0.848	0.820	0.823
Median of φ_ε	0.041	0.081	-0.019	0.456	0.446	0.452	0.815	0.831	0.817
Median of estimated φ at α_{min}	0.067	0.049	-0.014	0.415	0.465	0.451	0.786	0.816	0.815
α_{max}									
10th percentile	0.959	0.959	0.950	1.167	1.102	1.038	1.376	1.264	1.119
Median	1.040	1.036	1.012	1.256	1.180	1.101	1.477	1.351	1.181
90th percentile	1.128	1.114	1.077	1.345	1.256	1.170	1.550	1.421	1.249
α_{OU}									
10th percentile	0.089	0.694	0.905	0.307	0.749	0.920	0.621	0.797	0.947
Median	0.822	0.986	1.001	0.940	1.007	1.004	0.987	0.997	1.004
90th percentile	1.232	1.248	1.094	1.422	1.269	1.086	1.382	1.156	1.057
α_{min}									
10th percentile	0.624	0.783	0.888	0.624	0.810	0.907	0.739	0.837	0.941
Median	1.017	1.032	0.999	0.993	0.998	1.006	1.001	1.002	1.004
90th percentile	1.384	1.276	1.101	1.371	1.202	1.096	1.195	1.107	1.061

Notes: In all specifications, the true value of α=1, $E(\Gamma)=E(\beta)$=0.3, N=2000, K*=100, R^2_ψ=0.5, and R^2_ξ=0.

Table MC2b

	B: Factor structure such that $E((Corr(W_j,W_{j'}))=0.2$								
		$E(\beta^*\Gamma)=0.09$			$E(\beta^*\Gamma)=0.3$			$E(\beta^*\Gamma)=0.6$	
	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$
Median of φ	0.606	0.587	0.594	0.735	0.772	0.782	0.924	0.929	0.928
Median of φ_ε	0.629	0.575	0.605	0.767	0.727	0.752	0.920	0.932	0.914
Median of estimated φ at α_{min}	0.649	0.620	0.639	0.750	0.785	0.803	0.923	0.943	0.941
α_{max}									
10th percentile	0.956	0.965	0.922	1.084	1.029	0.955	1.224	1.192	1.060
Median	1.038	1.039	1.011	1.137	1.116	1.042	1.294	1.293	1.137
90th percentile	1.102	1.119	1.109	1.228	1.202	1.158	1.448	1.425	1.262
α_{OU}									
10th percentile	0.554	0.539	0.732	0.639	0.616	0.777	0.864	0.872	0.880
Median	0.761	0.771	0.915	0.795	0.866	0.938	0.966	0.979	0.991
90th percentile	0.886	0.948	1.057	0.914	1.057	1.121	1.067	1.066	1.078
α_{min}									
10th percentile	0.851	0.877	0.930	0.864	0.889	0.933	0.911	0.923	0.959
Median	0.998	1.003	1.004	0.989	0.993	1.002	0.983	0.999	1.005
90th percentile	1.156	1.144	1.077	1.105	1.088	1.049	1.068	1.058	1.041

Notes: In all specifications, the true value of $\alpha=1$, $E(\Gamma)=E(\beta)=0.3$, $N=2000$, $K^*=100$, $R^2_\psi=0.5$, and $R^2_\xi=0$.

Table MC3a

A: Factor structure such that $E((Corr(W_j,W_{j'}))=0$

	$E(\beta^*\Gamma)=0.09$			$E(\beta^*\Gamma)=0.3$			$E(\beta^*\Gamma)=0.6$		
	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$
Median of φ	0.041	0.074	0.064	0.438	0.444	0.454	0.850	0.816	0.832
Median of φ_ε	0.084	0.059	0.081	0.353	0.347	0.337	0.632	0.640	0.631
Median of estimated φ at α_{min}	0.026	0.078	0.054	0.412	0.495	0.474	0.670	0.718	0.791
α_{max}									
10th percentile	0.958	0.948	0.949	1.165	1.138	1.039	1.378	1.334	1.120
Median	1.045	1.041	1.019	1.257	1.224	1.101	1.476	1.407	1.181
90th percentile	1.137	1.127	1.089	1.345	1.314	1.171	1.551	1.479	1.248
α_{OU}									
10th percentile	0.030	0.626	0.890	0.213	0.258	0.884	0.033	0.439	0.889
Median	0.844	0.971	0.996	0.868	0.784	0.975	0.783	0.806	0.953
90th percentile	1.347	1.270	1.101	1.388	1.276	1.067	1.346	1.121	1.013
α_{min}									
10th percentile	0.591	0.626	0.881	0.227	0.576	0.865	0.275	0.493	0.882
Median	0.987	0.987	0.997	0.879	0.878	0.979	0.738	0.812	0.952
90th percentile	1.430	1.419	1.112	1.516	1.215	1.077	1.063	0.950	1.017

Notes: In all specifications, the true value of $\alpha=1$, $E(\Gamma)=E(\beta)=0.3$, $N=2000$, $K^*=100$, $R^2_\psi=0.5$, and $R^2_\xi=0.2$.

Table MC3b

	B: Factor structure such that $E((Corr(W_j,W_{j'}))=0.2$								
	$E(\beta^*\Gamma)=0.09$			$E(\beta^*\Gamma)=0.3$			$E(\beta^*\Gamma)=0.6$		
	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$
Median of φ	0.767	0.675	0.615	0.909	0.899	0.806	1.097	1.077	0.956
Median of φ_ε	0.464	0.430	0.446	0.564	0.533	0.580	0.689	0.670	0.697
Median of estimated φ at α_{min}	0.625	0.654	0.639	0.678	0.712	0.802	0.755	0.823	0.915
α_{max}									
10th percentile	1.017	0.938	0.965	1.112	1.091	1.026	1.208	1.202	1.087
Median	1.072	1.036	1.021	1.188	1.157	1.068	1.300	1.284	1.130
90th percentile	1.142	1.136	1.089	1.280	1.231	1.134	1.397	1.379	1.217
α_{OU}									
10th percentile	0.145	0.452	0.753	0.333	0.505	0.725	0.379	0.556	0.852
Median	0.577	0.697	0.910	0.649	0.752	0.922	0.713	0.812	0.948
90th percentile	0.737	0.832	0.958	0.834	0.877	0.970	0.863	0.901	0.982
α_{min}									
10th percentile	0.437	0.738	0.898	0.447	0.690	0.904	0.380	0.641	0.907
Median	0.813	0.929	0.975	0.788	0.871	0.969	0.732	0.840	0.963
90th percentile	1.204	1.198	1.036	1.026	2.202	1.014	0.903	0.922	0.994

Notes: In all specifications, the true value of $\alpha=1$, $E(\Gamma)=E(\beta)=0.3$, $N=2000$, $K^*=100$, $R^2_\psi=0.5$, and $R^2_\xi=0.2$.

Table MC4a

	A: Factor structure such that $E((Corr(W_j,W_{j'}))=0$								
	$E(\beta^*\Gamma)=0.09$			$E(\beta^*\Gamma)=0.3$			$E(\beta^*\Gamma)=0.6$		
	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$
Median of φ	0.060	0.044	0.041	0.440	0.442	0.455	0.841	0.812	0.827
Median of φ_ε	0.047	0.050	0.025	0.259	0.252	0.249	0.461	0.465	0.467
Median of estimated φ at α_{min}	0.031	0.082	0.057	0.402	0.487	0.504	0.512	0.585	0.744
α_{max}									
10th percentile	0.949	0.954	0.947	1.188	1.149	1.028	1.364	1.313	1.125
Median	1.044	1.045	1.014	1.270	1.228	1.093	1.466	1.392	1.165
90th percentile	1.147	1.136	1.078	1.355	1.328	1.166	1.565	1.507	1.250
α_{OU}									
10th percentile	0.002	0.304	0.793	-0.449	-0.280	0.750	-0.656	-0.185	0.712
Median	0.863	0.876	0.989	0.705	0.551	0.887	0.446	0.437	0.825
90th percentile	1.314	1.260	1.141	1.418	1.232	1.028	1.486	0.977	0.911
α_{min}									
10th percentile	0.492	0.608	0.798	-0.505	-0.010	0.762	-0.628	-0.225	0.703
Median	0.979	0.982	0.984	0.657	0.655	0.890	0.288	0.418	0.832
90th percentile	1.378	1.213	1.139	1.702	1.582	1.021	1.080	0.699	0.920

Notes: In all specifications, the true value of $\alpha=1$, $E(\Gamma)=E(\beta)=0.3$, $N=2000$, $K^*=100$, $R^2_\psi=0.5$, and $R^2_\xi=0.4$.

Table MC4b

	B: Factor structure such that $E((Corr(W_j,W_{j'}))=0.2$								
		$E(\beta^*\Gamma)=0.09$			$E(\beta^*\Gamma)=0.3$			$E(\beta^*\Gamma)=0.6$	
	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$	$P_S=0.2$	$P_S=0.4$	$P_S=0.8$
Median of φ	0.878	0.770	0.649	1.076	0.988	0.831	1.311	1.204	0.993
Median of φ_ε	0.323	0.295	0.315	0.385	0.395	0.412	0.489	0.466	0.493
Median of estimated φ at α_{min}	0.500	0.339	0.643	0.537	0.622	0.785	0.565	0.673	0.865
α_{max}									
10th percentile	1.020	0.983	0.965	1.130	1.117	1.023	1.219	1.204	1.084
Median	1.077	1.043	1.016	1.198	1.164	1.084	1.284	1.281	1.140
90th percentile	1.168	1.124	1.075	1.270	1.243	1.134	1.390	1.372	1.186
α_{OU}									
10th percentile	0.455	0.031	0.782	-0.317	0.016	0.789	-0.353	0.034	0.786
Median	0.248	0.482	0.870	0.294	0.513	0.876	0.267	0.537	0.891
90th percentile	0.533	0.697	0.918	0.595	0.720	0.927	0.610	0.736	0.943
α_{min}									
10th percentile	-0.156	0.248	0.838	-0.299	0.175	0.817	-0.555	0.041	0.796
Median	0.516	0.672	0.933	0.377	0.590	0.914	0.297	0.553	0.901
90th percentile	1.643	2.447	1.001	0.932	0.789	0.966	0.662	0.739	0.954

Notes: In all specifications, the true value of $\alpha=1$, $E(\Gamma)=E(\beta)=0.3$, $N=2000$, $K^*=100$, $R^2_\psi=0.5$, and $R^2_\xi=0.4$.

Appendix

Proof of Theorem 1:

To simplify the notation, define

$$\begin{aligned}
M_{11}^{K^*} &= E\left(\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j \widetilde{W}_{ij}\Gamma_j\right)^2 ; \mathcal{G}^{K^*}\right) \\
M_{22}^{K^*} &= E\left(\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} (1-S_j)\, \widetilde{W}_{ij}\Gamma_j + \xi\right)^2 ; \mathcal{G}^{K^*}\right) \\
M_{12}^{K^*} &= E\left(\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j \widetilde{W}_{ij}\Gamma_j\right)\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} (1-S_j)\, \widetilde{W}_{ij}\Gamma_j + \xi\right) ; \mathcal{G}^{K^*}\right) \\
M_{Z1}^{K^*} &= E\left(Z_i\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j \widetilde{W}_{ij}\Gamma_j\right) ; \mathcal{G}^{K^*}\right) \\
M_{Z2}^{K^*} &= E\left(Z_i\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} (1-S_j)\, \widetilde{W}_{ij}\Gamma_j + \xi\right) ; \mathcal{G}^{K^*}\right).
\end{aligned}$$

By definition of the projection,

$$\begin{aligned}
\phi &= \frac{M_{Z1}^{K^*} M_{22}^{K^*} - M_{Z2}^{K^*} M_{12}^{K^*}}{M_{22}^{K^*} M_{11}^{K^*} - M_{12}^{K^*} M_{12}^{K^*}} \\
\phi_\varepsilon &= \frac{M_{Z2}^{K^*} M_{11}^{K^*} - M_{Z1}^{K^*} M_{12}^{K^*}}{M_{22}^{K^*} M_{11}^{K^*} - M_{12}^{K^*} M_{12}^{K^*}}.
\end{aligned}$$

Thus, if $M_{22}^{K^*} M_{11}^{K^*} - M_{12}^{K^*} M_{12}^{K^*}$ is nonzero,

$$\frac{\phi_\varepsilon}{\phi} = \frac{M_{Z2}^{K^*} M_{11}^{K^*} - M_{Z1}^{K^*} M_{12}^{K^*}}{M_{Z1}^{K^*} M_{22}^{K^*} - M_{Z2}^{K^*} M_{12}^{K^*}}. \tag{1}$$

Notice first that

$$\begin{aligned}
E\left(\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} \widetilde{W}_{ij}\Gamma_j\right)^2 ; \mathcal{G}^{K^*}\right) &= E\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*} \widetilde{W}_{ij}\widetilde{W}_{i\ell}\Gamma_j\Gamma_\ell ; \mathcal{G}^{K^*}\right) \\
&= \frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*} \sigma_{j\ell}^{K^*}\Gamma_j\Gamma_\ell.
\end{aligned}$$

Let

$$\begin{aligned}
A &\equiv \lim_{K^* \to\infty} E\left(\frac{1}{K^*}\sum_{j=1}^{K^*} \sigma_{jj}^{K^*}\left(\Gamma_j\right)^2\right) \\
B &\equiv \lim_{K^* \to\infty} E\left(\frac{1}{K^*}\sum_{\substack{\ell=1 \\ \ell\neq j}}^{K^*} \left(\sigma_{j\ell}^{K^*}\Gamma_j\Gamma_\ell\right)\right).
\end{aligned}$$

Assumptions 1 and 2 guarantee that

$$\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}\sigma_{j\ell}^{K^*}\Gamma_j\Gamma_\ell \xrightarrow{p} \lim_{K^*\to\infty} E\left[\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}\sigma_{j\ell}^{K^*}\Gamma_j\Gamma_\ell\right] = A + B$$

and that A and B are finite.

Applying the same steps to $M_{11}^{K^*}$,

$$\begin{aligned}
M_{11}^{K^*} &= E\left(\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j\widetilde{W}_{ij}\Gamma_j\right)^2; \mathcal{G}^{K^*}\right) \\
&= E\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*} S_jS_\ell\widetilde{W}_{ij}\widetilde{W}_{i\ell}\Gamma_j\Gamma_\ell; \mathcal{G}^{K^*}\right) \\
&= \frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*} S_jS_\ell\sigma_{j\ell}^{K^*}\Gamma_j\Gamma_\ell \\
&= \frac{1}{K^*}\sum_{j=1}^{K^*} S_j\sigma_{jj}^{K^*}\Gamma_j\Gamma_j + \frac{1}{K^*}\sum_{j=1}^{K^*-1}\sum_{\substack{\ell=1\\ \ell\neq j}}^{K^*} S_jS_\ell\sigma_{j\ell}^{K^*}\Gamma_j\Gamma_\ell \\
&\xrightarrow{p} E\left(S_j\sigma_{jj}^{K^*}\Gamma_j\Gamma_j\right) + \sum_{j=1}^{\infty}\sum_{\ell=j+1}^{\infty} E\left(S_jS_\ell\sigma_{j\ell}^{K^*}\Gamma_j\Gamma_\ell\right) = P_SA + P_S^2B
\end{aligned}$$

Similarly,

$$\begin{aligned}
M_{22}^{K^*} &= \frac{1}{K^*}\sum_{j=1}^{K^*}(1-S_j)\,\sigma_{jj}^{K^*}\Gamma_j\Gamma_j + \frac{1}{K^*}\sum_{j=1}^{K^*-1}\sum_{\substack{\ell=1\\ \ell\neq j}}^{K^*}(1-S_j)\,(1-S_\ell)\,\sigma_{j\ell}^{K^*}\Gamma_j\Gamma_\ell + \sigma_\xi^2 \\
&\xrightarrow{p} (1-P_S)\,A + (1-P_S)^2\,B + \sigma_\xi^2 \\
M_{12}^{K^*} &= \frac{1}{K^*}\sum_{j=1}^{K^*-1}\sum_{\ell+1=1}^{K^*}(1-S_j)\,S_\ell\sigma_{j\ell}^{K^*}\Gamma_j\Gamma_\ell \\
&\xrightarrow{p} (1-P_S)\,P_SB
\end{aligned}$$

Using Assumption 3, now notice that

$$\begin{aligned}
M_{Z1}^{K^*} &= E\left(Z\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j\widetilde{W}_{ij}\Gamma_j\right); \mathcal{G}^{K^*}\right) \\
&= \frac{1}{K^*}\sum_{j=1}^{K^*} S_j\mu_j^{K^*}\Gamma_j \\
&\xrightarrow{p} P_SE\left(\mu_j^{K^*}\Gamma_j\right) \\
M_{Z2}^{K^*} &= E\left(Z\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}(1-S_j)\,\widetilde{W}_{ij}\Gamma_j\right); \mathcal{G}^{K^*}\right) \\
&= \frac{1}{K}\sum_{j=1}^{K}(1-S_j)\,\mu_j^{K^*}\Gamma_j \\
&\xrightarrow{p} (1-P_S)\,E\left(\mu_j^{K^*}\Gamma_j\right)
\end{aligned}$$

Notice that the probability limit of the denominator of equation (1) is

$$
\begin{aligned}
&plim\ M_{Z1}^{K^*} M_{22}^{K^*} - M_{Z2}^{K^*} M_{12}^{K^*} \xrightarrow{p} P_S E\left(\mu_j^{K^*}\Gamma_j\right)\left[(1-P_S)A + (1-P_S)^2 B + \sigma_\xi^2\right] - \\
&(1-P_S)\,E\left(\mu_j^{K^*}\Gamma_j\right)(1-P_S)\,P_S B \\
&= E\left(\mu_j^{K^*}\Gamma_j\right)\left[P_s(1-P_S)A + \sigma_\xi^2\right]
\end{aligned}
$$

Given that $A > 0$, and since it cannot be the case that $P_s = 1$ and $\sigma_\xi^2 = 0$, this expression can only equal zero if $E\left(\mu_j^{K^*}\Gamma_j\right) = 0$. In that case, ϕ_ε also converges to zero, proving the second part of the theorem (if plim $\phi = 0$, then $\phi_\varepsilon = 0$).

Now assume that ϕ does not converge to zero. Consider the numerator of equation (1):

$$
\begin{aligned}
M_{Z2}^{K^*} M_{11}^{K^*} - M_{Z1}^{K^*} M_{12}^{K^*} \xrightarrow{p} &(1-P_S)E\left(\mu_j^{K^*}\Gamma_j\right)(P_S A + P_S^2 B) - \\
&P_S E\left(\mu_j^{K^*}\Gamma_j\right)(1-P_S)P_S B) \\
&= E\left(\mu_j^{K^*}\Gamma_j\right) P_s(1-P_S)A
\end{aligned}
$$

Thus,

$$
\frac{\phi_\varepsilon}{\phi} \xrightarrow{p} \frac{(1-P_S)A}{(1-P_S)A + \sigma_\xi^2}.
$$

Verification That MA Model Satisfies Assumptions 2 and 3

To see that Assumption 2 is satisfied first note that

$$
\sigma_{j,\ell}^{K^*} = \begin{cases} < \infty & |j-\ell| \le q_q \\ 0 & \text{otherwise} \end{cases}.
$$

Thus

$$
\begin{aligned}
\frac{1}{K}\sum_{j=1}^{K}\sum_{\ell=1}^{K^*} E(\sigma_{j,\ell}^{K^*}\Gamma_j\Gamma_\ell) &= \frac{1}{K}\sum_{j=1}^{K}\sum_{\ell=j-q_w}^{j+q_w} E(\sigma_{j,\ell}^{K}\Gamma_j\Gamma_\ell) \\
&= \sum_{\ell=j-q_w}^{j+q_w} E(\sigma_{j,\ell}^{K}\Gamma_j\Gamma_\ell)
\end{aligned}
$$

Next consider the second part of Assumption 2. The key is the second moment of the variable of interest

$$
\begin{aligned}
\frac{1}{K^{*2}}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}\sum_{r=1}^{K^*}\sum_{s=1}^{K^*} E(\sigma_{j,\ell}^{K}\sigma_{r,s}^{K}\Gamma_j\Gamma_\ell\Gamma_r\Gamma_s) &= \frac{1}{K^{*2}}\sum_{j=1}^{K^*}\sum_{\ell=j-q_w}^{j+q_w}\sum_{r=1}^{K^*}\sum_{s=r-q_w}^{r+q_w} E(\sigma_{j,\ell}^{K}\sigma_{r,s}^{K}\Gamma_j\Gamma_\ell\Gamma_r\Gamma_s) \\
&= \frac{1}{K^{*2}}\sum_{j=1}^{K^*}\sum_{\ell=j-q_w}^{j+q_w}\sum_{r=1,|r-j|>2q_w}^{K^*}\sum_{s=r-q_w}^{r+q_w} E(\sigma_{j,\ell}^{K}\sigma_{r,s}^{K}\Gamma_j\Gamma_\ell\Gamma_r\Gamma_s) \\
&\quad + \frac{1}{K^{*2}}\sum_{j=1}^{K^*}\sum_{\ell=j-q_w}^{j+q_w}\sum_{r=1,|r-j|\le 2q_w}^{K^*}\sum_{s=r-q_w}^{r+q_w} E(\sigma_{j,\ell}^{K}\sigma_{r,s}^{K}\Gamma_j\Gamma_\ell\Gamma_r\Gamma_s) \\
&\to \left[\sum_{\ell=j-q_w}^{j+q_w} E(\sigma_{j,\ell}^{K}\Gamma_j\Gamma_\ell)\right]^2
\end{aligned}
$$

Since this is just the square of the expected value, Assumption 2 is satisfied.

Next consider Assumption 3.

Notice that

$$\begin{aligned}
\mu_j^{K^*} &= \sqrt{K^*} E\left(Z_i W_{ij} | \mathcal{G}^{K^*}\right) \\
&= \left[\sum_{\ell=1}^{K} W_{i\ell}\beta_\ell\right] W_{ij} \\
&= E\left(\sigma_{j,\ell}^{K^*}\beta_\ell\right)
\end{aligned}$$

Verifying Assumption 3 becomes virtually identical to verifying Assumption 2.

$$\begin{aligned}
E(\mu_j^{K^*}\Gamma_j) &= \frac{1}{K^*}\sum_{j=1}^{K^*} E\left(\mu_j^{K^*}\Gamma_j\right) \\
&= \frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*} E\left(\sigma_{j,\ell}^{K^*}\beta_\ell\Gamma_j\right) \\
&= \frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=j-q_w}^{j+q_w} E(\sigma_{j,\ell}^{K^*}\beta_j\Gamma_\ell) \\
&= \sum_{\ell=j-q_w}^{j+q_w} E(\sigma_{j,\ell}^{K^*}\beta_j\Gamma_\ell)
\end{aligned}$$

and

$$E\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\mu_j^{K^*}\Gamma_j\right)^2 = \frac{1}{K^{*2}}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}\sum_{r=1}^{K^*}\sum_{s=1}^{K^*} E(\sigma_{j,\ell}^{K^*}\sigma_{r,s}^{K^*}\beta_j\Gamma_\ell\beta_r\Gamma_s)$$

$$\begin{aligned}
&= \frac{1}{K^{*2}}\sum_{j=1}^{K^*}\sum_{\ell=j-q_w}^{j+q_w}\sum_{r=1,|r-j|>2q_w}^{K^*}\sum_{s=r-q_w}^{r+q_w} E(\sigma_{j,\ell}^{K^*}\sigma_{r,s}^{K^*}\beta_j\Gamma_\ell\beta_r\Gamma_s) \\
&\quad+\frac{1}{K^{*2}}\sum_{j=1}^{K^*}\sum_{\ell=j-q_w}^{j+q_w}\sum_{r=1,|r-j|\leq 2q_w}^{K^*}\sum_{s=r-q_w}^{r+q_w} E(\sigma_{j,\ell}^{K^*}\sigma_{r,s}^{K^*}\beta_j\Gamma_\ell\beta_r\Gamma_s) \\
&\to \left[\sum_{\ell=j-q_w}^{j+q_w} E(\sigma_{j,\ell}^{K^*}\beta_j\Gamma_\ell)\right]^2
\end{aligned}$$

which is again just the square of the mean so this gives the result.

Proof of Theorem 2

The theorem imposes the assumption

$$\begin{aligned}
A &\equiv \lim_{K^*\to\infty} E\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sigma_{jj}^{K^*}\left(\Gamma_j\right)^2\right) = \lim_{K^*\to\infty} E\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\omega_j\sigma_{jj}^{K^*}\left(\Gamma_j\right)^2\right) \\
B &\equiv \lim_{K^*\to\infty} E\left(\frac{1}{K^*}\sum_{\substack{\ell=1\\ \ell\neq j}}^{K^t}\left(\sigma_{j\ell}^{K^*}\Gamma_j\Gamma_\ell\right)\right) = \lim_{K^*\to\infty} E\left(\frac{1}{K^*}\sum_{\substack{\ell=1\\ \ell\neq j}}^{K^t}\left(\omega_j\omega_l\sigma_{j\ell}^{K^*}\Gamma_j\Gamma_\ell\right)\right).
\end{aligned}$$

Following the proof of theorem 1, this assumption implies that $M_{11}^{K^*} \xrightarrow{p} P_S A + P_S^2 B$, $M_{22}^{K^*} \xrightarrow{p} (1-P_S)A + (1-P_S)^2 B + \sigma_\xi^2$, and $M_{12}^{K^*} \xrightarrow{p} (1-P_S) P_S B$, as in the i.i.d. case.

Following that proof, it is straightforward to show that

$$\begin{aligned}
M_{Z1}^{K^*} &= P_S E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right) \\
M_{Z2}^{K^*} &= (1-P_S) E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 0\right)
\end{aligned}$$

Consider the numerator of $\phi_\varepsilon^{K^*}$

$$\begin{aligned}
&M_{Z2}^{K^*} M_{11}^{K^*} - M_{Z1}^{K^*} M_{12}^{K^*} \\
&\xrightarrow{p} \left[(1-P_S) E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 0\right)\right]\left[P_S A + P_S^2 B\right] - \left[P_S E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right)\right]\left[(1-P_S) P_S B\right] \\
&= (1-P_S) P_S A E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 0\right) + (1-P_S) P_S^2 B E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 0\right) \\
&\quad - (1-P_S) P_S^2 B E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right) \\
&= (1-P_S) P_S A E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 0\right) + \left[(1-P_S) P_S^2 B \left[E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 0\right) - E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right)\right]\right]
\end{aligned}$$

The plim of the numerator of ϕ^{K^*} is

$$\begin{aligned}
&M_{Z1}^{K^*} M_{22}^{K^*} - M_{Z2}^{K^*} M_{12}^{K^*} \\
&\xrightarrow{p} \left[P_S E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right)\right]\left[(1-P_S) A + (1-P_S)^2 B + \sigma_\xi^2\right] \\
&\quad - \left[(1-P_S) E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 0\right)\right]\left[(1-P_S) P_S B\right] \\
&= (1-P_S) P_S A E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right) + P_S (1-P_S)^2 B E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right) \\
&\quad - P_S (1-P_S)^2 B E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 0\right) + P_S E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right)\sigma_\xi^2 \\
&= (1-P_S) P_S A E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right) + P_S (1-P_S)^2 B \left[E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right) - E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 0\right)\right] \\
&\quad + P_S E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right)\sigma_\xi^2
\end{aligned}$$

Now notice that we can write the numerator of the expression $\phi - \phi_\varepsilon$ as

$$\begin{aligned}
M_{Z1}^{K^*} M_{22}^{K^*} - M_{Z2}^{K^*} M_{12}^{K^*} - M_{Z2}^{K^*} M_{11}^{K^*} - M_{Z1}^{K^*} M_{12}^{K^*} \\
&= (1-P_S) P_S A E\left[\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right) - E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 0\right)\right] \\
&\quad + \left[P_S (1-P_S)^2 - (1-P_S) P_S^2\right] B \left[E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right) - E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 0\right)\right] \\
&\quad + P_S E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right)\sigma_\xi^2 \\
&= (1-P_S) P_S \left[A + B\right] E\left[\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right) - E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 0\right)\right] \\
&\quad + P_S E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right)\sigma_\xi^2 \\
&> 0
\end{aligned}$$

The result comes from the fact that we know that $A + B > 0$ and that $E\left(\mu_j^{K^*}\Gamma_j \mid S_j = 1\right) > 0$.

Proof of Theorem 3

Following a similar procedure to theorem 1

$$\begin{aligned}
M_{11}^{K^*} &= E\left(\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j\widetilde{W}_{ij}G_j\right)^2; \mathcal{G}^{K^*}\right) \\
M_{22}^{K^*} &= E\left(\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} \widetilde{\widetilde{W}}_{ij}\Gamma_j + \xi\right)^2; \mathcal{G}^{K^*}\right) \\
&\xrightarrow{p} \sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j\widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\Gamma_j\Gamma_{j-\ell}\right) \\
M_{12}^{K^*} &= E\left(\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j\widetilde{W}_{ij}G_j\right)\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} \widetilde{\widetilde{W}}_{ij}\Gamma_j + \xi\right); \mathcal{G}^{K^*}\right) \\
&= 0 \\
M_{Z1}^{K^*} &= E\left(Z_i\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j\widetilde{W}_{ij}G_j\right); \mathcal{G}^{K^*}\right) \\
M_{Z2}^{K^*} &= E\ \left(Z_i\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} (1-S_j)\,\widetilde{\widetilde{W}}_{ij}\Gamma_j + \xi\right); \mathcal{G}^{K^*}\right) \\
&\xrightarrow{p} \sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j\widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\beta_j\Gamma_{j-\ell}\right) + \sigma_\xi^2
\end{aligned}$$

By definition of the projection, and using Assumption ??

$$\begin{aligned}
\phi_{X'G} &= \frac{M_{Z1}^{K^*}}{M_{11}^{K^*}} \\
\phi_e &= \frac{M_{Z2}^{K^*}}{M_{22}^{K^*}} \\
&\xrightarrow{p} \frac{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j\widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\beta_j\Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j\widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\Gamma_j\Gamma_{j-\ell}\right) + \sigma_\xi^2}
\end{aligned}$$

by definition of G_j

$$\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} \widetilde{W}_{ij}\Gamma_j = \frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j\widetilde{W}_{ij}G_j + \frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} \widetilde{\widetilde{W}}_{ij}\Gamma_j$$

so one can write Assumption 9 as

$$\begin{aligned}
&\frac{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{W}_j \widetilde{W}_{j-\ell}\right) E\left(\beta_j \Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{W}_j \widetilde{W}_{j-\ell}\right) E\left(\Gamma_j \Gamma_{j-\ell}\right)} \\
=&\frac{plim M_{Z1}^{K^*} + \sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j \widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\beta_j \Gamma_{j-\ell}\right)}{plim M_{11}^{K^*} - \sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j \widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\Gamma_j \Gamma_{j-\ell}\right)} \\
=&\frac{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j \widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\beta_j \Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j \widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\Gamma_j \Gamma_{j-\ell}\right)}
\end{aligned}$$

which can be rewritten as

$$\frac{plim M_{Z1}^{K^*}}{plim M_{11}^{K^*}} = \frac{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j \widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\beta_j \Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j \widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\Gamma_j \Gamma_{j-\ell}\right)}$$

Then if $\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j \widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\beta_j \Gamma_{j-\ell}\right) = 0$,

$$\phi_e \xrightarrow{p} 0 \text{ and } \phi_{X'G} \xrightarrow{p} 0,$$

otherwise

$$\frac{\phi_e}{\phi_{X'G}} = \frac{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j \widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\Gamma_j \Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j \widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\Gamma_j \Gamma_{j-\ell}\right) + \sigma_\xi^2}. \tag{2}$$

Proof of Theorem 4

Since we have not made distributional assumptions, we need only consider what is identified from the second moments and Condition 1. Clearly π, λ β are identified from (3.1) and (3.2). Since the model is just identified in the case that $cov(v, \varepsilon) = 0$, we need only think about what the additional information in Condition 1 generates. The 2SLS estimator converges to

$$\widehat{\alpha}_{2SLS} = \alpha + \frac{cov\left(v, \varepsilon\right)}{\lambda var(v)}.$$

Is knowledge of Condition 1 sufficient to identify α? When ε is uncorrelated with W, Condition 1 is equivalent to assuming that

$$\frac{cov(W'\beta, W'\Gamma)}{var(W'\Gamma)} = \frac{cov(v, \varepsilon)}{var(\varepsilon)}.$$

Suppose the model is not identified. Then there would be alternative values α^*, Γ^*, and ε^* with $\alpha^* \neq \alpha$ such that

$$\widehat{\alpha} = \alpha^* + \frac{cov\left(v, \varepsilon^*\right)}{\lambda var(v)}.$$

Under these conditions, note that

$$\begin{aligned}
Y - \alpha^* T &= \left(\alpha - \alpha^*\right) T + W'\Gamma + \varepsilon \\
&= \left(\alpha - \alpha^*\right) \left[W'\pi + u + \lambda\left(W'\beta + v\right)\right] + W'\Gamma + \varepsilon,
\end{aligned}$$

and thus

$$\begin{aligned}\gamma^* &= \gamma + (\alpha - \alpha^*)(\beta + \lambda\pi) \\ \varepsilon^* &= \varepsilon + (\alpha - \alpha^*)(u + \lambda v).\end{aligned}$$

But if this model satisfies the assumptions, we know that

$$\frac{cov(W'\beta, W'\Gamma^*)}{var(W'\Gamma^*)} = \frac{cov(v, \varepsilon^*)}{var(\varepsilon^*)},$$

which is equivalent to

$$\begin{aligned}&\frac{cov\,(W'\beta, W'\Gamma) + (\alpha - \alpha^*)\,cov\,(W'\beta, (W'\pi + \lambda W'\beta))}{var\,(W'\Gamma) + 2\,(\alpha - \alpha^*)\,cov(W'\Gamma, (W'\pi + \lambda W'\beta)) + (\alpha - \alpha^*)^2\,var(W'\pi + \lambda W'\beta)} \\ &= \frac{cov\,(v, \varepsilon) + (\alpha - \alpha^*)\,cov\,(v, (u + \lambda v))}{var(\varepsilon) + 2\,(\alpha - \alpha^*)\,cov(\varepsilon, (u + \lambda v)) + (\alpha - \alpha^*)^2\,var(u + \lambda v)}.\end{aligned}$$

Imposing the restriction from the true model,

$$\phi \equiv \frac{cov(W'\beta, W'\Gamma)}{var(W'\Gamma)} = \frac{cov(v, \varepsilon)}{var(\varepsilon)},$$

yields

$$\begin{aligned}&\frac{\phi + (\alpha - \alpha^*)\,\frac{cov\left(W'\beta, \left(W'\pi + \lambda W'\beta\right)\right)}{var(W'\Gamma)}}{1 + 2\,(\alpha - \alpha^*)\,\frac{cov(W'\Gamma, (W'\pi + \lambda W'\beta))}{var(W'\Gamma)} + (\alpha - \alpha^*)^2\,\frac{var(W'\pi + \lambda W'\beta)}{var(W'\Gamma)}} \\ &= \frac{\phi + (\alpha - \alpha^*)\,\frac{cov(v, (u + \lambda v))}{var(\varepsilon)}}{1 + 2\,(\alpha - \alpha^*)\,\frac{cov(\varepsilon, (u + \lambda v))}{var(\varepsilon)} + (\alpha - \alpha^*)^2\,\frac{var(u + \lambda v)}{var(\varepsilon)}}.\end{aligned}$$

Solving out yields

$$\begin{aligned}0 = (\alpha - \alpha^*)^3 &\left[\frac{cov\,(v, (u + \lambda v))}{var(\varepsilon)}\frac{var(W'\pi + \lambda W'\beta)}{var(W'\Gamma)} - \frac{cov\,(W'\beta, (W'\pi + \lambda W'\beta))}{var(W'\Gamma)}\frac{var(u + \lambda v)}{var(\varepsilon)}\right] \\ + (\alpha - \alpha^*)^2 &\left[\phi\frac{var(W'\pi + \lambda W'\beta)}{var(W'\Gamma)} + 2\frac{cov\,(v, (u + \lambda v))}{var(\varepsilon)}\frac{cov(W'\Gamma, (W'\pi + \lambda W'\beta))}{var(W'\Gamma)}\right. \\ &\left. -\phi\frac{var(u + \lambda v)}{var(\varepsilon)} - 2\frac{cov\,(W'\beta, (W'\pi + \lambda W'\beta))}{var(W'\Gamma)}\frac{cov(\varepsilon, (u + \lambda v))}{var(\varepsilon)}\right] \\ + (\alpha - \alpha^*) &\left[\frac{cov\,(v, (u + \lambda v))}{var(\varepsilon)} + 2\phi\frac{cov(W'\Gamma, (W'\pi + \lambda W'\beta))}{var(W'\Gamma)}\right. \\ &\left. -\frac{cov\,(W'\beta, (W'\pi + \lambda W'\beta))}{var(W'\Gamma)} - 2\phi\frac{cov(\varepsilon, (u + \lambda v))}{var(\varepsilon)}\right].\end{aligned}$$

Thus Condition 1 restricts the solutions α^* to be the solutions of a cubic equation, one of which is α.

Proof of Theorem 5

If we use T as the instrument we obtain exactly the cubit from Theorem 4 with $\beta = 0, u = 0$, and $\lambda = 1$. Thus

$$\begin{aligned}&\left[\frac{cov\,(v, (u + \lambda v))}{var(\varepsilon)}\frac{var(W'\pi + \lambda W'\beta)}{var(W'\Gamma)} - \frac{cov\,(W'\beta, (W'\pi + \lambda W'\beta))}{var(W'\Gamma)}\frac{var(u + \lambda v)}{var(\varepsilon)}\right] \\ = &\left[\frac{var\,(v)}{var(\varepsilon)}\frac{var(W'\beta)}{var(W'\Gamma)} - \frac{var\,(W'\beta)}{var(W'\Gamma)}\frac{var(v)}{var(\varepsilon)}\right] \\ = &\ 0.\end{aligned}$$

Using the definition of ϕ from the proof of Theorem 2,

$$
\begin{aligned}
&\left[\phi\frac{var(W'\pi+\lambda W'\beta)}{var(W'\Gamma)}+2\frac{cov\left(v,(u+\lambda v)\right)}{var(\varepsilon)}\frac{cov(W'\Gamma,(W'\pi+\lambda W'\beta))}{var(W'\Gamma)}\right.\\
&\left.-\phi\frac{var(u+\lambda v)}{var(\varepsilon)}-2\frac{cov\left(W'\beta,(W'\pi+\lambda W'\beta)\right)}{var(W'\Gamma)}\frac{cov(\varepsilon,(u+\lambda v))}{var(\varepsilon)}\right]\\
&=\left[\phi\frac{var(W'\beta)}{var(W'\Gamma)}+2\frac{var\left(v\right)}{var(\varepsilon)}\frac{cov(W'\Gamma,W'\beta)}{var(W'\Gamma)}\right.\\
&\left.-\phi\frac{var(v)}{var(\varepsilon)}-2\frac{var\left(W'\beta\right)}{var(W'\Gamma)}\frac{cov(\varepsilon,v)}{var(\varepsilon)}\right]\\
&=\phi\frac{var(W'\beta)}{var(W'\Gamma)}+2\frac{var\left(v\right)}{var(\varepsilon)}\phi-\phi\frac{var(v)}{var(\varepsilon)}-2\frac{var\left(W'\beta\right)}{var(W'\Gamma)}\phi\\
&=\phi\left[\frac{var\left(v\right)}{var(\varepsilon)}-\frac{var\left(W'\beta\right)}{var(W'\Gamma)}\right]\\
&\left[\frac{cov\left(v,(u+\lambda v)\right)}{var(\varepsilon)}+2\phi\frac{cov(W'\Gamma,(W'\pi+\lambda W'\beta))}{var(W'\Gamma)}\right.\\
&\left.-\frac{cov\left(W'\beta,(W'\pi+\lambda W'\beta)\right)}{var(W'\Gamma)}-2\phi\frac{cov(\varepsilon,(u+\lambda v))}{var(\varepsilon)}\right]\\
&=\left[\frac{var\left(v\right)}{var(\varepsilon)}+2\phi\frac{cov(W'\Gamma,W'\beta)}{var(W'\Gamma)}-\frac{var\left(W'\beta\right)}{var(W'\Gamma)}-2\phi\frac{cov(\varepsilon,v)}{var(\varepsilon)}\right]\\
&=\left[\frac{var\left(v\right)}{var(\varepsilon)}-\frac{var\left(W'\beta\right)}{var(W'\Gamma)}\right]
\end{aligned}
$$

Thus, the model is a solution to the quadratic

$$
(\alpha-\alpha^*)^2\,\phi\left[\frac{var\left(v\right)}{var(\varepsilon)}-\frac{var\left(W'\beta\right)}{var(W'\Gamma)}\right]+(\alpha-\alpha^*)\left[\frac{var\left(v\right)}{var(\varepsilon)}-\frac{var\left(W'\beta\right)}{var(W'\Gamma)}\right],
$$

which has two solutions:

$$
\begin{aligned}
1)\ \alpha^* &= \alpha\\
2)\ \alpha^* &= \alpha+\frac{1}{\phi}\\
&= \alpha+\frac{var(\varepsilon)}{cov(u,\varepsilon)}
\end{aligned}
$$

Verification that The Factor Model Satisfies Assumptions 2 and 3

Notice that

$$
\begin{aligned}
\sigma_{j,\ell}^{K^*} &= E\left(\widetilde{W}_{ij}\widetilde{W}_{i\ell}\mid\mathcal{G}^K\right)\\
&= \begin{cases}\frac{1}{K^*}\Lambda_{j_1}'\Lambda_{j_1}+\sigma_{j_1}^2 & j_1=j_2\\ \frac{1}{K^*}\Lambda_{j_1}'\Lambda_{j_2} & j_1\neq j_2\end{cases}
\end{aligned}
$$

Then

$$
\begin{aligned}
\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}E(\sigma_{j,\ell}^{K^*}\Gamma_j\Gamma_\ell) &= \frac{1}{K^{*2}}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}E(\Lambda_j'\Lambda_\ell\Gamma_j\Gamma_\ell)+\frac{1}{K^*}\sum_{j=1}^{K^*}E(\sigma_j^2\Gamma_j^2)\\
&= \frac{1}{K^{*2}}\sum_{j=1}^{K^*}E(\Lambda_j'\Lambda_j\Gamma_j^2)+\frac{1}{K^{*2}}\sum_{j=1}^{K^*}\sum_{\substack{\ell=1\\ \ell\neq j}}^{K^*}E(\Lambda_j'\Gamma_j)E(\Lambda_\ell\Gamma_\ell)+\frac{1}{K^*}\sum_{j=1}^{K^*}E(\sigma_j^2\Gamma_j^2)\\
&\xrightarrow{p} E(\Lambda_j\Gamma_j)E(\Lambda_j\Gamma_j)+E(\sigma_{j_1}^2\Gamma_j^2)
\end{aligned}
$$

thus the first part of Assumption 2 is satisfied.

Now consider the second part,

$$
\begin{aligned}
& E\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*} E(\sigma_{j,\ell}^{K^*}\Gamma_j\Gamma_\ell)\right)^2 \\
= \; & E\left(\frac{1}{K^{*2}}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*} \Lambda_j'\Lambda_\ell\Gamma_j\Gamma_\ell + \frac{1}{K^*}\sum_{j=1}^{K^*}\sigma_j^2\Gamma_j^2\right)^2 \\
= \; & E\left(\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\Lambda_{j_1}'\Gamma_j\right)\left(\frac{1}{K^*}\sum_{\ell=1}^{K^*}\Lambda_{j_2}\Gamma_\ell\right)\right)^2 \\
& +2\left(\left(\frac{1}{K^{*2}}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}\Lambda_{j_1}'\Lambda_{j_2}\Gamma_j\Gamma_\ell\right)\left(\frac{1}{K^*}\sum_{r=1}^{K^*}\sigma_r^2\Gamma_r^2\right)\right) \\
& +E\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sigma_j^2\Gamma_j^2\right)^2 \\
\xrightarrow{p} \; & \left(E\left(\Lambda_j'\Gamma_j\right)E\left(\Gamma_\ell\Lambda_\ell\right)\right)^2 + 2E\left(\Lambda_j'\Gamma_j\right)E(\sigma_{j_1}^2\Gamma_j^2) + E(\sigma_{j_1}^2\Gamma_j^2)^2 \\
= \; & \left(E(\Lambda_j'\Gamma_j)E(\Lambda_j\Gamma_j) + E(\sigma_{j_1}^2\Gamma_j^2)\right)^2.
\end{aligned}
$$

To see where the three pieces come from, if you multiplied everything out all the terms involving fourth moments, third moments and second moments would disappear since we are dividing by K^4.

Thus

$$
\begin{aligned}
& Var\left(\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}E(\sigma_{j,\ell}^{K^*}\Gamma_j\Gamma_\ell)\right)^2\right) \\
= \; & \frac{1}{K^2}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}\sum_{r=1}^{K^*}\sum_{s=1}^{K^*}E(\sigma_{j,\ell}^{K^*}\sigma_{r,s}^{K^*}\Gamma_j\Gamma_\ell\Gamma_r\Gamma_s) \\
& -\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}E\left(\sigma_{j,\ell}^{K^*}\Gamma_j\Gamma_\ell\right)\right)^2 \\
\xrightarrow{p} \; & \left(E(\Lambda_j'\Gamma_j)E(\Lambda_j\Gamma_j) + E(\sigma_{j_1}^2\Gamma_j^2)\right)^2 - \left(E(\Lambda_j'\Gamma_j)E(\Lambda_j\Gamma_j) + E(\sigma_{j_1}^2\Gamma_j^2)\right)^2 \\
= \; & 0
\end{aligned}
$$

We next proceed to verify Assumption 3 in a way almost identical to that for Assumption 2.

$$
\begin{aligned}
\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}E(\sigma_{j,\ell}^{K^*}\Gamma_j\beta_\ell) \; &= \; \frac{1}{K^{*2}}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}E(\Lambda_j'\Lambda_\ell\Gamma_j\beta_\ell) + \frac{1}{K^*}\sum_{j=1}^{K^*}E(\sigma_j^2\Gamma_j\beta_j) \\
&\xrightarrow{p} \; E(\Lambda_j'\Gamma_j)E(\Lambda_j\beta_j) + E(\sigma_{j_1}^2\Gamma_j\beta_j)
\end{aligned}
$$

thus the first part of Assumption 3 is satisfied.

Now consider the second part,

$$
\begin{aligned}
& E\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}E(\sigma_{j,\ell}^{K^*}\Gamma_j\beta_\ell)\right)^2 \\
= \; & E\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}\Lambda_j'\Lambda_\ell\Gamma_j\beta_\ell+\frac{1}{K^*}\sum_{j=1}^{K^*}\sigma_j^2\Gamma_j\beta_j\right)^2 \\
= \; & E\left(\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\Lambda_{j_1}'\Gamma_j\right)\left(\frac{1}{K^*}\sum_{\ell=1}^{K^*}\Lambda_{j_2}\beta_\ell\right)\right)^2 \\
& +2\left(\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}\Lambda_{j_1}'\Lambda_{j_2}\Gamma_j\beta_\ell\right)\left(\frac{1}{K^*}\sum_{r=1}^{K^*}\sigma_r^2\Gamma_r\beta_r\right)\right) \\
& +E\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sigma_j^2\Gamma_j\beta_j\right)^2 \\
\xrightarrow{p} \; & \left(E\left(\Lambda_j'\Gamma_j\right)E\left(\beta_\ell\Lambda_\ell\right)\right)^2+2E\left(\Lambda_j'\Gamma_j\right)E(\sigma_{j_1}^2\Gamma_j\beta_j)+E(\sigma_{j_1}^2\Gamma_j\beta_j)^2 \\
= \; & \left(E(\Lambda_j'\Gamma_j)E(\Lambda_j\beta_j)+E(\sigma_{j_1}^2\Gamma_j\beta_j)\right)^2.
\end{aligned}
$$

Thus

$$
\begin{aligned}
& Var\left(\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}E(\sigma_{j,\ell}^{K^*}\Gamma_j\beta_\ell)\right)^2\right) \\
= \; & \frac{1}{K^{*2}}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}\sum_{r=1}^{K^*}\sum_{s=1}^{K^*}E(\sigma_{j,\ell}^{K^*}\sigma_{r,s}^{K^*}\Gamma_j\beta_\ell\Gamma_r\beta_s) \\
& -\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sum_{\ell=1}^{K^*}E\left(\sigma_{j,\ell}^{K}\Gamma_j\beta_\ell\right)\right)^2 \\
\xrightarrow{p} \; & \left(E(\Lambda_j'\Gamma_j)E(\Lambda_j\beta_j)+E(\sigma_{j_1}^2\Gamma_j\beta_j)\right)^2-\left(E(\Lambda_j'\Gamma_j)E(\Lambda_j\beta_j)+E(\sigma_{j_1}^2\Gamma_j\beta_j)\right)^2 \\
= \; & 0.
\end{aligned}
$$

Lemma 1

In this lemma, we will define and prove 23 intermediate results which will be useful for the proof of Theorem 6. Define

$$
\begin{aligned}
\widetilde{\Gamma}(\alpha) &\equiv \Gamma+(\alpha_0-\alpha)\delta \\
\widetilde{\Gamma}^*(\alpha) &\equiv \Gamma^*+(\alpha_0-\alpha)\delta^* \\
\widetilde{\Gamma}_j(\alpha) &\equiv \Gamma_j+(\alpha_0-\alpha)\delta_j
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\lambda} &\underset{N\to\infty}{\overset{p}{\to}} \sqrt{P_{s0}}\Lambda && (1)\\
\widehat{\Sigma} &\underset{N\to\infty}{\overset{p}{\to}} \Sigma && (2)\\
\frac{1}{N}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &\underset{N\to\infty}{\overset{p}{\to}} \frac{1}{\sqrt{K^*}}\Sigma\widetilde{\Gamma}(\alpha) && (3)\\
\frac{1}{N}V'\widetilde{Z} &\underset{N\to\infty}{\overset{p}{\to}} \frac{1}{\sqrt{K^*}}\Sigma\beta && (4)\\
\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &\underset{N\to\infty}{\overset{p}{\to}} \left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right) && (5)\\
&\overset{p}{\to} E(\Lambda_j\Gamma_j)\\
\frac{1}{N}\widetilde{F}'\widetilde{Z} &\underset{N\to\infty}{\overset{p}{\to}} \left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right) && (6)\\
&\overset{p}{\to} E(\Lambda_j\beta_j)\\
\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{Z} &\underset{N\to\infty}{\overset{p}{\to}} \left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+\frac{1}{K^*}\widetilde{\Gamma}^*(\alpha)'\Sigma\beta^* && (7)\\
&\overset{p}{\to} E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)E\left(\Lambda_j\beta_j\right)+E\left(\widetilde{\Gamma}_j(\alpha)\sigma_j^2\beta_j\right)\\
\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &\underset{N\to\infty}{\overset{p}{\to}} \left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)+\frac{\widetilde{\Gamma}^*(\alpha)^{*\prime}\Sigma\widetilde{\Gamma}^*(\alpha)^*}{K^*} && (8)\\
&\quad +\sigma_{\xi 0}^2+2(\alpha-\alpha_0)\sigma_{\xi\omega 0}+(\alpha-\alpha_0)^2\sigma_\omega^2\\
&\overset{p}{\to} E\left(\Lambda_j\Gamma_j\right)^2+E\left(\Gamma_{j_1}^2\sigma_{j_1}^2\right)+\sigma_{\xi 0}^2\\
\frac{1}{K}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda} &\underset{N\to\infty}{\overset{p}{\to}} \frac{P_{s0}}{K}\Lambda'^{-1}\Lambda && (9)\\
&\overset{p}{\to} P_{s0}E(\Lambda_j^2/\sigma_j^2)\\
\frac{1}{N}\widetilde{F}'\widetilde{F} &\underset{N\to\infty}{\overset{p}{\to}} 1 && (10)\\
\frac{1}{N}\frac{1}{\sqrt{K}}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}\widehat{\lambda} &\underset{N\to\infty}{\overset{p}{\to}} \frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\sum_{j=1}^{K}\widetilde{\Gamma}_j(\alpha)\Lambda_j && (11)\\
&\overset{p}{\to} P_{s0}E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)\\
\frac{1}{N}\frac{1}{\sqrt{K}}\widehat{\lambda}'V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &\underset{N\to\infty}{\overset{p}{\to}} \frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\sum_{j=1}^{K}\widetilde{\Gamma}_j(\alpha)\Lambda_j\sigma_j^2 && (12)\\
&\overset{p}{\to} P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)\\
\frac{1}{N}\frac{1}{\sqrt{K}}\widetilde{F}'V\widehat{\Sigma}^{-1}\widehat{\lambda} &\underset{N\to\infty}{\overset{p}{\to}} 0 && (13)
\end{aligned}
$$

$$\frac{1}{N^2}\widetilde{F}'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \underset{N\to\infty}{\overset{p}{\to}} 0 \tag{14}$$

$$\begin{aligned}\frac{1}{KN}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}\widehat{\lambda} &\underset{N\to\infty}{\overset{p}{\to}} P_{s0}\frac{1}{K}\sum_{j=1}^{K}\frac{\Lambda_j^2}{\sigma_j^2} \\ &\overset{p}{\to} P_{s0}E(\Lambda_j^2/\sigma_j^2)\end{aligned} \tag{15}$$

$$\begin{aligned}\frac{1}{N^2}\frac{1}{\sqrt{K}}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &\underset{N\to\infty}{\overset{p}{\to}} \frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha) \\ &\overset{p}{\to} P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\end{aligned} \tag{16}$$

$$\begin{aligned}\frac{1}{N^2}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &\underset{N\to\infty}{\overset{p}{\to}} \frac{\widetilde{\Gamma}(\alpha)'\Sigma\widetilde{\Gamma}(\alpha)}{K^*} \\ &\overset{p}{\to} P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]\end{aligned} \tag{17}$$

$$\begin{aligned}\frac{1}{N\sqrt{K}}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\widetilde{Z} &\underset{N\to\infty}{\overset{p}{\to}} \frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\beta \\ &\overset{p}{\to} P_{s0}E\left(\Lambda_j\beta_j\right)\end{aligned} \tag{18}$$

$$\begin{aligned}\frac{1}{N}\widetilde{F}'\widetilde{Z} &\underset{N\to\infty}{\overset{p}{\to}} \frac{\Lambda^{*\prime}\Gamma^*}{K^*} \\ &\overset{p}{\to} E(\Lambda_j\beta_j)\end{aligned} \tag{19}$$

$$\begin{aligned}\frac{1}{N^2}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}V'\widetilde{Z} &\underset{N\to\infty}{\overset{p}{\to}} \frac{\widetilde{\Gamma}(\alpha)'\Sigma\beta}{K^*} \\ &\overset{p}{\to} P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)\beta_j\right]\end{aligned} \tag{20}$$

$$\begin{aligned}\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\Sigma^{-1}V'V\Sigma^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &\underset{N\to\infty}{\overset{p}{\to}} \frac{\widetilde{\Gamma}(\alpha)'\Sigma\widetilde{\Gamma}(\alpha)}{K^*} \\ &\overset{p}{\to} P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right)\end{aligned} \tag{21}$$

$$\begin{aligned}\frac{1}{\sqrt{K^*}N}\Lambda'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &\underset{N\to\infty}{\overset{p}{\to}} \frac{1}{K^*}\Lambda'\widetilde{\Gamma}(\alpha) \\ &\overset{p}{\to} P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\end{aligned} \tag{22}$$

$$\begin{aligned}\frac{1}{\sqrt{K^*}N}\Lambda'\widehat{\Sigma}^{-1}V'\widetilde{Z} &\underset{N\to\infty}{\overset{p}{\to}} \frac{1}{K^*}\Lambda'\beta \\ &\overset{p}{\to} P_{s0}E\left(\Lambda_j\beta_j\right)\end{aligned} \tag{23}$$

Proof of Lemma 1

Result L1.1

Clearly this $\sqrt{P_{S0}}\Lambda$ is identified and with K fixed is a standard GMM problem so it is consistent.

Result L1.2

This follows directly from the Law of Large Numbers and Result L1.1.

Result L1.3

$$\begin{aligned}
\frac{1}{N}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &= \frac{1}{N}\sum_{i=1}^{N}V_i\left(\widetilde{F}_i\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)+\frac{V_i^{*\prime}\widetilde{\Gamma}^*(\alpha)^*}{\sqrt{K^*}}+\xi_i+(\alpha-\alpha_0)\,\omega_i\right)\\
&\underset{N\to\infty}{\overset{p}{\to}} \frac{1}{\sqrt{K^*}}\Sigma\widetilde{\Gamma}^*(\alpha)
\end{aligned}$$

Result L1.4

$$\begin{aligned}
\frac{1}{N}V'\widetilde{Z} &= \frac{1}{N}\sum_{i=1}^{N}V_i\left(\widetilde{F}_i\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+\frac{V_i^{*\prime}\beta^*}{\sqrt{K^*}}+u_i\right)\\
&\underset{N\to\infty}{\overset{p}{\to}} \frac{1}{\sqrt{K^*}}\Sigma\beta
\end{aligned}$$

Result L1.5

$$\begin{aligned}
\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &= \frac{1}{N}\widetilde{F}'\left[\widetilde{F}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)+V^*\frac{\widetilde{\Gamma}^*(\alpha)}{\sqrt{K^*}}+\xi+(\alpha-\alpha_0)\,\omega\right]\\
&= \left(\frac{\Lambda^{*\prime}\Gamma^*}{K^*}\right)\frac{1}{N}\sum_{i=1}^{N}\widetilde{F}_i^2+\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}\left[\frac{1}{N}\sum_{i=1}^{N}\widetilde{F}_iV_{ij}\Gamma_j\right]+\frac{1}{N}\sum_{i=1}^{N}\widetilde{F}_i\left[\xi_i+(\alpha-\alpha_0)\,\omega_i\right]\\
&\underset{N\to\infty}{\overset{p}{\to}} \left(\frac{\Lambda^{*\prime}\Gamma^*}{K^*}\right)\\
&\overset{p}{\to} E(\Lambda_j\Gamma_j)
\end{aligned}$$

$_i$ with each piece falling from the law of large numbers

Result L1.6

$$\begin{aligned}
\frac{1}{N}\widetilde{F}'\widetilde{Z} &= \frac{1}{N}\widetilde{F}'\left[\widetilde{F}\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+V^*\frac{\beta^*}{\sqrt{K^*}}+u\right]\\
&= \left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)\frac{1}{N}\sum_{i=1}^{N}\widetilde{F}_i^2+\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}\left[\frac{1}{N}\sum_{i=1}^{N}\widetilde{F}_iV_{ij}\beta_j\right]+\frac{1}{N}\sum_{i=1}^{N}\widetilde{F}_iu_i\\
&\underset{N\to\infty}{\overset{p}{\to}} \left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)\\
&\overset{p}{\to} E(\Lambda_j\beta_j)
\end{aligned}$$

with each piece falling from the law of large numbers

Result L.1.7

$$\begin{aligned}
\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{Z} &= \frac{1}{N}\left[\widetilde{F}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)^{*}}{K^{*}}\right)+V^{*}\frac{\widetilde{\Gamma}^{*}(\alpha)}{\sqrt{K^{*}}}+\xi+(\alpha-\alpha_{0})\,\omega\right]'\left[\widetilde{F}\left(\frac{\Lambda^{*\prime}\beta^{*}}{K^{*}}\right)+V^{*}\frac{\beta^{*}}{\sqrt{K^{*}}}+u\right]\\
&= \left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^{*}}\right)\frac{1}{N}\left[\widetilde{F}'\widetilde{F}\left(\frac{\Lambda^{*\prime}\beta^{*}}{K^{*}}\right)+\widetilde{F}'V^{*}\frac{\beta^{*}}{\sqrt{K^{*}}}+\widetilde{F}'u\right]\\
&\quad+\frac{1}{N}\frac{\widetilde{\Gamma}^{*}(\alpha)'}{\sqrt{K^{*}}}V^{*\prime}\left[\widetilde{F}\left(\frac{\Lambda^{*\prime}\beta^{*}}{K^{*}}\right)+V^{*}\frac{\beta^{*}}{\sqrt{K^{*}}}+u\right]\\
&\quad+\frac{1}{N}\xi'\left[\widetilde{F}\left(\frac{\Lambda^{*\prime}\beta^{*}}{K^{*}}\right)+V^{*}\frac{\beta^{*}}{\sqrt{K^{*}}}+u\right]\\
&\quad+\frac{1}{N}(\alpha-\alpha_{0})\,\omega'\left[\widetilde{F}\left(\frac{\Lambda^{*\prime}\beta^{*}}{K^{*}}\right)+V^{*}\frac{\beta^{*}}{\sqrt{K^{*}}}+u\right]\\
&\underset{N\to\infty}{\overset{p}{\to}} \left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^{*}}\right)\left(\frac{\Lambda^{*\prime}\beta^{*}}{K^{*}}\right)+\frac{1}{K^{*}}\widetilde{\Gamma}^{*}(\alpha)'\Sigma^{*}\beta^{*}+(\alpha-\alpha_{0})\,\sigma_{\omega,u}\\
&\overset{p}{\to} E\left(\Lambda_{j}\widetilde{\Gamma}_{j}(\alpha)\right)E\left(\Lambda_{j}\beta_{j}\right)+E\left(\widetilde{\Gamma}_{j}(\alpha)\sigma_{j}^{2}\beta_{j}\right)+(\alpha-\alpha_{0})\,\sigma_{\omega,u}
\end{aligned}$$

Result L1.8

$$\begin{aligned}
\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &= \frac{1}{N}\left[\widetilde{F}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^{*}}\right)+V^{*}\frac{\widetilde{\Gamma}^{*}(\alpha)}{\sqrt{K^{*}}}+\xi+(\alpha-\alpha_{0})\,\omega\right]'\times\\
&\qquad\left[\widetilde{F}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^{*}}\right)+V^{*}\frac{\widetilde{\Gamma}^{*}(\alpha)}{\sqrt{K^{*}}}+\xi+(\alpha-\alpha_{0})\,\omega\right]\\
&= \left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^{*}}\right)\frac{1}{N}\left[\widetilde{F}'\widetilde{F}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^{*}}\right)+2\widetilde{F}'V^{*}\frac{\widetilde{\Gamma}^{*}(\alpha)}{\sqrt{K^{*}}}+2\widetilde{F}'\xi\right]\\
&\quad+\frac{1}{N}\frac{\widetilde{\Gamma}^{*}(\alpha)'}{\sqrt{K^{*}}}V^{*\prime}\left[V^{*}\frac{\widetilde{\Gamma}^{*}(\alpha)}{\sqrt{K^{*}}}+2\xi\right]+\frac{1}{N}\xi'\xi+(\alpha-\alpha_{0})\frac{2}{N}\xi'\omega+(\alpha-\alpha_{0})^{2}\frac{1}{N}\omega'\omega\\
&\underset{N\to\infty}{\overset{p}{\to}} \left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^{*}}\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^{*}}\right)+\frac{1}{K^{*}}\widetilde{\Gamma}^{*}(\alpha)'\Sigma^{*}\widetilde{\Gamma}^{*}(\alpha)+\sigma_{\xi 0}^{2}+2(\alpha-\alpha_{0})\sigma_{\xi\omega 0}+(\alpha-\alpha_{0})^{2}\sigma_{\omega}^{2}\\
&\overset{p}{\to} E\left(\Lambda_{j}\widetilde{\Gamma}_{j}(\alpha)\right)E\left(\Lambda_{j}\beta_{j}\right)+E\left(\widetilde{\Gamma}_{j}(\alpha)\sigma_{j}^{2}\beta_{j}\right)+\sigma_{\xi 0}^{2}+2(\alpha-\alpha_{0})\sigma_{\xi\omega 0}+(\alpha-\alpha_{0})^{2}\sigma_{\omega}^{2}
\end{aligned}$$

Result L1.9

$$\begin{aligned}
\frac{1}{K}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda} &= \frac{1}{K}\sum_{j=1}^{K}\frac{\widehat{\lambda}_{j}^{2}}{\widehat{\sigma}_{j}^{2}}\\
&\underset{N\to\infty}{\overset{p}{\to}} \frac{P_{s0}}{K}\sum_{j=1}^{K}\frac{\Lambda_{j}^{2}}{\sigma_{j}^{2}}\\
&\overset{p}{\to} P_{s0}E\left(\frac{\Lambda_{j}^{2}}{\sigma_{j}^{2}}\right)
\end{aligned}$$

by the law of large numbers.

Result L1.10

$$\frac{1}{N}\widetilde{F}'\widetilde{F} = \frac{1}{N}\sum_{i=1}^{N}\widetilde{F}_i^2 \xrightarrow{p} 1$$

by the law of large numbers

Result L1.11

$$\begin{aligned}
\frac{1}{N}\frac{1}{\sqrt{K}}\left(\widetilde{Y}-\alpha\widetilde{T}\right)' V\widehat{\Sigma}^{-1}\widehat{\lambda} &= \frac{1}{\sqrt{K}}\sum_{j=1}^{K}\frac{\widehat{\lambda}_j}{\widehat{\sigma}_j^2}\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{Y}_i-\alpha\widetilde{T}_i\right)v_{ij} \\
&= \frac{1}{N}\frac{1}{\sqrt{K}}\sum_{i=1}^{N}\sum_{j=1}^{K}\left(\widetilde{F}_i\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)+\sum_{\ell=1}^{K^*}v_{i\ell}\frac{\widetilde{\Gamma}_\ell(\alpha)}{\sqrt{K^*}}+\xi_i+(\alpha-\alpha_0)\,\omega_i\right)\frac{v_{ij}\widehat{\lambda}_j}{\widehat{\sigma}_j^2} \\
&\underset{N\to\infty}{\xrightarrow{p}} \frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\sum_{j=1}^{K}\Gamma_j\Lambda_j \\
&\xrightarrow{p} P_{s0}E\left(\Gamma_j\Lambda_j\right)
\end{aligned}$$

Result L1.12

This is similar to the previous one.

$$\begin{aligned}
\frac{1}{N}\frac{1}{\sqrt{K}}\left(\widetilde{Y}-\alpha\widetilde{T}\right)' V\widehat{\lambda} &= \frac{1}{\sqrt{K}}\sum_{j=1}^{K}\widehat{\lambda}_j\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)v_{ij} \\
&= \frac{1}{N}\frac{1}{\sqrt{K}}\sum_{i=1}^{N}\sum_{j=1}^{K}\left(\widetilde{F}_i\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)+\sum_{\ell=1}^{K^*}v_{i\ell}\frac{\widetilde{\Gamma}_\ell(\alpha)}{\sqrt{K^*}}+\xi_i+(\alpha-\alpha_0)\,\omega_i\right)v_{ij}\widehat{\lambda}_j \\
&\underset{N\to\infty}{\xrightarrow{p}} \frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\sum_{j=1}^{K}\Gamma_j\Lambda_j\sigma_j^2 \\
&\xrightarrow{p} P_{s0}E\left(\Gamma_j\Lambda_j\sigma_j^2\right)
\end{aligned}$$

Result L1.13

$$\begin{aligned}
\frac{1}{N}\frac{1}{\sqrt{K}}\sum_{i=1}^{N}\sum_{j=1}^{K}\widetilde{F}_i\frac{v_{ij}\widehat{\lambda}_j}{\widehat{\sigma}_j^2} &= \frac{1}{\sqrt{K}}\sum_{j=1}^{K}\frac{\widehat{\lambda}_j}{\widehat{\sigma}_j^2}\left[\frac{1}{N}\sum_{i=1}^{N}\widetilde{F}_i v_{ij}\right] \\
&\underset{N\to\infty}{\xrightarrow{p}} 0
\end{aligned}$$

Result L1.14

$$\begin{aligned}
\frac{1}{N^2}\widetilde{F}'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &= \frac{1}{N^2}\sum_{j=1}^{K}\frac{1}{\widehat{\sigma}_j^2}\left[\sum_{i_1=1}^{N}\widetilde{F}_{i_1}v_{i_1j}\right]\sum_{i_2=1}^{N}v_{i_2j}\left(\widetilde{F}_{i_2}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)+\sum_{\ell=1}^{K^*}v_{i_2\ell}\frac{\widetilde{\Gamma}_\ell(\alpha)}{\sqrt{K^*}}+\xi_{i_2}+(\alpha-\alpha_0)\,\omega_{i_2}\right) \\
&= \left(\frac{\Lambda^{*\prime}\Gamma^*}{K^*}\right)\sum_{j=1}^{K}\left[\frac{1}{N^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\frac{\widetilde{F}_{i_1}v_{i_1j}v_{i_2j}}{\widehat{\sigma}_j^2}\widetilde{F}_{i_2}\right] \\
&\quad+\sum_{j=1}^{K}\left[\sum_{\ell=1}^{K^*}v_{i_2\ell}\frac{\Gamma_\ell}{\sqrt{K^*}}\right]\left[\frac{1}{N^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\frac{\widetilde{F}_{i_1}v_{i_1j}v_{i_2j}}{\widehat{\sigma}_j^2}\right] \\
&\quad+\sum_{j=1}^{K}\left[\frac{1}{N^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\frac{\widetilde{F}_{i_1}v_{i_1j}v_{i_2j}}{\widehat{\sigma}_j^2}\left[\xi_{i_2}+(\alpha-\alpha_0)\,\omega_{i_2}\right]\right] \\
&\underset{N\to\infty}{\overset{p}{\to}} 0
\end{aligned}$$

Result L1.15

$$\begin{aligned}
\frac{1}{KN}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}\widehat{\lambda} &= \frac{1}{K}\sum_{j_1=1}^{K}\sum_{j_2=1}^{K}\frac{\widehat{\lambda}_{j_1}\widehat{\lambda}_{j_2}}{\widehat{\sigma}_{j_1}^2\widehat{\sigma}_{j_2}^2}\left[\frac{1}{N}\sum_{i=1}^{N}v_{ij_1}v_{ij_2}\right] \\
&\underset{N\to\infty}{\overset{p}{\to}} \frac{P_{s0}}{K}\sum_{j=1}^{K}\frac{\Lambda_j^2}{\sigma_j^2} \\
&\overset{p}{\to} P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)
\end{aligned}$$

Result L1.16

$$\begin{aligned}
\frac{1}{N^2}\frac{1}{\sqrt{K}}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &= \frac{1}{\sqrt{K}}\widehat{\lambda}'\widehat{\Sigma}^{-1}\left(\frac{1}{N}V'V\right)\widehat{\Sigma}^{-1}\left(\frac{1}{N}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
&\underset{N\to\infty}{\overset{p}{\to}} \frac{\sqrt{P_{s0}}}{\sqrt{K}}\Lambda'\Sigma^{-1}\Sigma\Sigma^{-1}\left(\frac{1}{\sqrt{K^*}}\Sigma\widetilde{\Gamma}(\alpha)\right) \\
&= \frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha) \\
&\overset{p}{\to} P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)
\end{aligned}$$

Result L1.17

$$\begin{aligned}
\frac{1}{N^2}Y'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &= \left(\frac{1}{N}Y'V\right)\widehat{\Sigma}^{-1}\left(\frac{1}{N}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
&\underset{N\to\infty}{\overset{p}{\to}} \left(\frac{1}{\sqrt{K^*}}\widetilde{\Gamma}(\alpha)'\,\Sigma\right)\Sigma^{-1}\left(\frac{1}{\sqrt{K^*}}\Sigma\widetilde{\Gamma}(\alpha)\right) \\
&= \frac{1}{K^*}\widetilde{\Gamma}(\alpha)'\,\Sigma\widetilde{\Gamma}(\alpha) \\
&\overset{p}{\to} P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}(\alpha)_j^2\right)
\end{aligned}$$

Result L1.18

$$\begin{aligned}
\frac{1}{N\sqrt{K}}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\widetilde{Z} & \underset{N\to\infty}{\overset{p}{\to}} \frac{1}{\sqrt{K}}\sqrt{P_{s0}}\Lambda'\Sigma^{-1}\left(\frac{1}{\sqrt{K^*}}\Sigma\beta\right) \\
&= \frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\beta \\
&= P_{s0}E\left(\Lambda_j\beta_j\right)
\end{aligned}$$

Result L1.19

$$\begin{aligned}
\frac{1}{N}\widetilde{F}'\widetilde{Z} &= \frac{1}{N}\widetilde{F}'\left[\widetilde{F}\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+V^*\frac{\beta^*}{\sqrt{K^*}}+u\right] \\
&= \left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)\frac{1}{N}\sum_{i=1}^{N}\widetilde{F}_i^2+\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K}\beta_j\left[\frac{1}{N}\sum_{i=1}^{N}\widetilde{F}_iV_{ij}\right]+\frac{1}{N}\sum_{i=1}^{N}\widetilde{F}_iu_i \\
&\underset{N\to\infty}{\overset{p}{\to}} \frac{\Lambda^{*\prime}\beta^*}{K^*} \\
&\overset{p}{\to} E\left(\Lambda_j\beta_j\right)
\end{aligned}$$

Result L1.20

$$\begin{aligned}
\frac{1}{N^2}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}V'Z &\underset{N\to\infty}{\overset{p}{\to}} \frac{1}{K^*}\widetilde{\Gamma}\left(\alpha\right)'\Sigma\Sigma^{-1}\Sigma\beta \\
&= \frac{1}{K^*}\widetilde{\Gamma}\left(\alpha\right)'\Sigma\beta \\
&\overset{p}{\to} P_{s0}E\left(\widetilde{\Gamma}_j\left(\alpha\right)\sigma_j^2\beta_j\right)
\end{aligned}$$

Result L1.21

$$\begin{aligned}
\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &= \left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\right)\widehat{\Sigma}^{-1}\left(\frac{1}{N}V'V\right)\widehat{\Sigma}^{-1}\left(\frac{1}{N}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
&\underset{N\to\infty}{\overset{p}{\to}} \frac{1}{K^*}\widetilde{\Gamma}\left(\alpha\right)'\Sigma\Sigma^{-1}\Sigma\Sigma^{-1}\Sigma\widetilde{\Gamma}\left(\alpha\right) \\
&= \frac{1}{K^*}\widetilde{\Gamma}\left(\alpha\right)'\Sigma\widetilde{\Gamma}\left(\alpha\right) \\
&\overset{p}{\to} P_{s0}E\left(\widetilde{\Gamma_j}(\alpha)^2\sigma_j^2\right)
\end{aligned}$$

Result L.1.22

$$\begin{aligned}
\frac{1}{N\sqrt{K^*}}\Lambda'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) &\underset{N\to\infty}{\overset{p}{\to}} \frac{1}{\sqrt{K^*}}\Lambda'\Sigma^{-1}\left(\frac{1}{\sqrt{K^*}}\Sigma\widetilde{\Gamma}\left(\alpha\right)\right) \\
&= \frac{1}{K^*}\Lambda'\widetilde{\Gamma}\left(\alpha\right) \\
&\overset{p}{\to} P_{s0}E\left(\Lambda_j\widetilde{\Gamma_j}(\alpha)\right)
\end{aligned}$$

Result L.1.23

$$\begin{aligned}\frac{1}{N\sqrt{K^*}}\Lambda'\widehat{\Sigma}^{-1}V'\widetilde{Z} &\underset{N\to\infty}{\overset{p}{\to}} \frac{1}{\sqrt{K^*}}\Lambda'\Sigma^{-1}\left(\frac{1}{\sqrt{K^*}}\Sigma\beta\right)\\ &= \frac{1}{K^*}\Lambda'\beta\\ &\overset{p}{\to} P_{s0}E\left(\Lambda_j\beta_j\right)\end{aligned}$$

Lemma 2

Here we define and prove 6 additional results which will be useful in proving Theorem 6.

Let the notation U_p denote uniform convergence in probability.

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{W}_i'\widehat{\Gamma}\left(\theta\right)\widetilde{Z}_i\right) \qquad (1)\\ &\underset{N\to\infty}{\overset{U_p}{\to}}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{1}{K^*}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{1}{K^*}\Lambda'\beta\right)+\left(\frac{1}{K^*}\Lambda'\Gamma\right)\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+\frac{\widetilde{\Gamma}(\alpha)'\Sigma\beta}{K^*}\\ &-\frac{\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left[\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)^2\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\beta\right)\right]}{P_s+\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda}\\ &-\frac{\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Gamma\right)\left[\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\beta\right]}{P_s+\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda}\\ &\overset{U_p}{\to}E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E(\Lambda_j^2/\sigma_j^2)E(\Lambda_j\beta_j)+2E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E\left(\Lambda_j\beta_j\right)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)\beta_j\right)\\ &-\frac{E(\Lambda_j\widetilde{\Gamma}_j(\alpha))E(\Lambda_j\beta_j)P_{s0}^2\left[E(\Lambda_j^2/\sigma_j^2)+1\right]^2}{P_s+P_{s0}E(\Lambda_j^2/\sigma_j^2)}\end{aligned}$$

$$\begin{aligned}&\widehat{\Gamma}_{K^*}\left(\theta\right)'\widehat{\Sigma}\widehat{\Gamma}_{K^*}\left(\theta\right) \qquad (2)\\ &\underset{N\to\infty}{\overset{U_p}{\to}}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\Lambda'\Sigma^{-1}\Lambda\frac{1}{K^*}\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)+2\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\Lambda'\widetilde{\Gamma}(\alpha)}{K^*}\right)+\left(\frac{\widetilde{\Gamma}(\alpha)'\Sigma\widetilde{\Gamma}(\alpha)}{K^*}\right)\\ &-2\left[\frac{1}{1+\frac{P_{s0}}{KP_s}\Lambda'\Sigma^{-1}\Lambda}\right]\frac{1}{P_s}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda+\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right)^2\\ &+\left[\frac{1}{1+\frac{P_{s0}}{KP_s}\Lambda'\Sigma^{-1}\Lambda}\right]^2\frac{1}{P_s^2}\left(\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)^2\left(\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\right.\\ &\quad+2\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right)\\ &\quad\left.+\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Gamma\right)^2\left(P_{s0}\Lambda'\frac{1}{K}\Sigma^{-1}\Lambda\right)\right)\\ &\overset{U_p}{\to}P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]\\ &+\left(E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\right)^2\left[P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+2P_{s0}-2\frac{\left(P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}\right)^2}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)\left[\frac{P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right]^2\right]\end{aligned}$$

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{W}_i'\widehat{\Gamma}\left(\theta\right)\widetilde{W}_i'\widehat{\Gamma}\left(\theta\right)\right) \qquad (3)\\
&\underset{N\to\infty}{\overset{U_p}{\to}}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)\left(\frac{\Lambda'\Sigma^{-1}\Lambda}{K^*}\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)+2\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)\left(\frac{\Lambda'\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)+\frac{\widetilde{\Gamma}(\alpha)'\Sigma\widetilde{\Gamma}(\alpha)}{K^*}\\
&\quad-2\left(\frac{1}{P_s+\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda}\right)\left[\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)\left(\frac{\sqrt{P_{s0}}\Lambda'\Sigma^{-1}\Lambda}{\sqrt{K}\sqrt{K^*}}\right)+\left(\frac{\sqrt{P_{s0}}\Lambda'\widetilde{\Gamma}(\alpha)}{\sqrt{K}\sqrt{K^*}}\right)\right]^2\\
&\quad+\left(\frac{1}{P_s+\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda}\right)^2\left(\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widehat{\Sigma}^{-1}\Lambda\frac{\widetilde{\Gamma}^{*}(\alpha)'\Lambda^*}{K^*}+\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right)^2\\
&\quad+\left[\left(\frac{1}{K^*}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)+\frac{\widetilde{\Gamma}(\alpha)'\Lambda}{K^*}\right.\\
&\quad\left.-\left(\frac{1}{P_s+\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda}\right)\left[\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)^2\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)+\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\widetilde{\Gamma}(\alpha)'\Lambda\right)\right]\right]^2\\
&\overset{U_p}{\to}E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)+2E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)P_{s0}E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right)\\
&\quad-2\frac{\left[P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}\right]^2}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)^2+\left(\frac{E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+1}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right)^2P_{s0}^3E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)^2\\
&\quad+\left[P_{s0}E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)\left[1-\frac{P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right]\right]^2
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{W}_i'\widehat{\Gamma}\left(\theta\right)\left(\widetilde{Y}_i-\alpha\widetilde{T}_i\right)\right)\underset{N\to\infty}{\overset{U_p}{\to}}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)\left(\frac{\Lambda'\Sigma^{-1}\Lambda}{K^*}\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)+2\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)\left(\frac{\widetilde{\Gamma}_j(\alpha)'\Lambda}{K^*}\right)+\frac{\widetilde{\Gamma}_j(\alpha)'\Sigma\widetilde{\Gamma}_j(\alpha)}{K^*} \qquad (4)\\
&\qquad-\frac{\left[\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)+\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}_j(\alpha)\right]^2}{P_s+\frac{1}{K}\Lambda'\Sigma^{-1}\Lambda}\\
&\qquad\overset{U_p}{\to}E(\Lambda_j\widetilde{\Gamma}_j(\alpha))^2P_{s0}E(\Lambda_j^2/\sigma_j^2)+2E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right)\\
&\qquad-E(\Lambda_j\widetilde{\Gamma}_j(\alpha))^2\frac{\left[P_{s0}E(\Lambda_j^2/\sigma_j^2)+P_{s0}\right]^2}{P_s+P_{s0}E(\Lambda_j^2/\sigma_j^2)}
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{\sqrt{K}}\widehat{\lambda}'\widehat{\Gamma}\left(\theta\right)\underset{N\to\infty}{\overset{U_p}{\to}}\left[\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^{*}(\alpha)}{K^*}\right)+\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\widetilde{\Gamma}(\alpha)'\Lambda\right]\left[\frac{P_s}{P_s+\frac{P_{s0}\Lambda'\Sigma^{-1}\Lambda}{K}}\right] \qquad (5)\\
&\qquad\overset{U_p}{\to}P_sE(\Lambda_j\widetilde{\Gamma}_j(\alpha))\left[\frac{P_{s0}E(\Lambda_j^2/\sigma_j^2)+P_{s0}}{P_s+E(\Lambda_j^2/\sigma_j^2)P_{s0}}\right]
\end{aligned}
$$

Proof of Lemma 2

Result L2.1

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{W}_i'\widehat{\Gamma}\left(\theta\right)\widetilde{Z}_i\right) =& \frac{1}{N}\widehat{\Gamma}\left(\theta\right)'\widetilde{W}'\widetilde{Z}\\
=& \frac{1}{N^2}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\widetilde{Z}\\
=& \frac{1}{N^2}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\widehat{\Sigma}^{-1}-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\right]\widetilde{W}'\widetilde{Z}\\
=& \frac{1}{N^2}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\left(\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'+V\right)\left[\widehat{\Sigma}^{-1}-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\right]\\
&\times\left(\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'+V'\right)\widetilde{Z}\\
=& \left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\frac{1}{K^*}\Lambda'\widehat{\Sigma}^{-1}\Lambda\right)\left(\frac{1}{N}\widetilde{F}'\widetilde{Z}\right)+\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)\widetilde{F}\right)\left(\frac{1}{\sqrt{K^*}N}\Lambda'\widehat{\Sigma}^{-1}V'\widetilde{Z}\right)\\
&-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{P_s}\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\frac{1}{\sqrt{K}\sqrt{K^*}}\Lambda'\widehat{\Sigma}^{-1}\widehat{\lambda}\right)^2\left(\frac{1}{N}\widetilde{F}'\widetilde{Z}\right)\\
&-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{P_s}\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\frac{1}{\sqrt{K}\sqrt{K^*}}\Lambda'\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{\sqrt{K}N}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\widetilde{Z}\right)\\
&+\left(\frac{1}{\sqrt{K^*}N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}\Lambda\right)\left(\frac{1}{N}\widetilde{F}'\widetilde{Z}\right)+\frac{1}{N^2}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}V'\widetilde{Z}\\
&-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{P_s}\left(\frac{1}{\sqrt{K}N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{\sqrt{K^*}\sqrt{K}}\widehat{\lambda}'\widehat{\Sigma}^{-1}\Lambda\right)\left(\frac{1}{N}\widetilde{F}'\widetilde{Z}\right)\\
&-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{P_s}\left(\frac{1}{\sqrt{K}N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{\sqrt{K}N}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\widetilde{Z}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\overset{U_p}{\underset{N\to\infty}{\to}} \left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{1}{K^*}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{1}{K^*}\Lambda'\beta\right)\\
&\quad-\frac{1}{1+\frac{P_{s0}}{KP_s}\Lambda'\Sigma^{-1}\Lambda}\frac{1}{P_s}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)^2\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)\\
&\quad-\frac{1}{1+\frac{P_{s0}}{KP_s}\Lambda'\Sigma^{-1}\Lambda}\frac{1}{P_s}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\beta\right)\\
&\quad+\left(\frac{1}{K^*}\Lambda'\widetilde{\Gamma}(\alpha)\right)\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+\frac{\widetilde{\Gamma}^*(\alpha)'\Sigma\beta}{K^*}\\
&\quad-\frac{1}{1+\frac{P_{s0}}{KP_s}\Lambda'\Sigma^{-1}\Lambda}\frac{1}{P_s}\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)\\
&\quad-\frac{1}{1+\frac{P_{s0}}{KP_s}\Lambda'\Sigma^{-1}\Lambda}\frac{1}{P_s}\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\beta\right)\\
&=\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{1}{K^*}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{1}{K^*}\Lambda'\beta\right)+\left(\frac{1}{K^*}\Lambda'\widetilde{\Gamma}(\alpha)\right)\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+\frac{\widetilde{\Gamma}(\alpha)'\Sigma\beta}{K^*}\\
&\quad-\frac{\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left[\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)^2\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\beta\right)\right]}{P_s+\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda}\\
&\quad-\frac{\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right)\left[\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\beta^*}{K^*}\right)+\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\beta\right]}{P_s+\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda}
\end{aligned}
$$

$$
\begin{aligned}
\overset{U_p}{\to}\quad & E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E(\Lambda_j^2/\sigma_j^2)E(\Lambda_j\beta_j)+E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E\left(\Lambda_j\beta_j\right)\\
&+P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)E(\Lambda_j\beta_j)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)\beta_j\right)\\
&-\frac{E(\Lambda_j\widetilde{\Gamma}_j(\alpha))\left[\left(P_{s0}E(\Lambda_j^2/\sigma_j^2)\right)^2E(\Lambda_j\beta_j)+P_{s0}E(\Lambda_j^2/\sigma_j^2)P_{s0}E\left(\Lambda_j\beta_j\right)\right]}{P_s+P_{s0}E(\Lambda_j^2/\sigma_j^2)}\\
&-\frac{\left(P_{s0}E(\Lambda_j\widetilde{\Gamma}_j(\alpha))\right)\left[\left(P_{s0}E(\Lambda_j^2/\sigma_j^2)\right)E(\Lambda_j\beta_j)+P_{s0}E(\Lambda_j\beta_j)\right]}{P_s+P_{s0}E(\Lambda_j^2/\sigma_j^2)}\\
=\quad & E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E(\Lambda_j^2/\sigma_j^2)E(\Lambda_j\beta_j)+2E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E\left(\Lambda_j\beta_j\right)\\
&+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)\beta_j\right)-\frac{E(\Lambda_j\widetilde{\Gamma}_j(\alpha))E(\Lambda_j\beta_j)P_{s0}^2\left[E(\Lambda_j^2/\sigma_j^2)+1\right]^2}{P_s+P_{s0}E(\Lambda_j^2/\sigma_j^2)}
\end{aligned}
$$

The pointwise convergence follows from the algebra and lemma 1. Uniformity comes from the fact that P_s is bounded from below by P_s^{ℓ} (and above by some finite number) in which case the relevant functions are clearly stochastically equicontinuitous. Throughout the rest of the proof we see the same basic structure throughout, so we do not repeat this statement each time.

Result L2.2

$$\begin{aligned}
\widehat{\Gamma}(\theta)'\widehat{\Sigma}\widehat{\Gamma}(\theta) &= \frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widehat{\Sigma}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\frac{1}{N}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
&= \frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\widehat{\Sigma}^{-1}-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\right]\widehat{\Sigma} \\
&\quad\times\left[\widehat{\Sigma}^{-1}-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\right]\frac{1}{N}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
&= \frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\left(\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'+V\right)\widehat{\Sigma}^{-1}\frac{1}{N}\left(\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'+V\right)'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
&\quad -2\left[\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\right]\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\left(\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'+V\right) \\
&\qquad\times\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\frac{1}{N}\left(\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'+V\right)'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
&\quad +\left[\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\right]^2\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\left(\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'+V\right)\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\Sigma} \\
&\qquad\times\left[\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\right]\frac{1}{N}\left(\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'+V\right)'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
&= \left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\Lambda'\widehat{\Sigma}^{-1}\Lambda\frac{1}{K^*}\right)\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
&\quad +2\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\frac{1}{N}\frac{1}{\sqrt{K^*}}\Lambda'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)+\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}\frac{1}{N}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
&\quad -2\left[\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\right]\frac{1}{P_s}\left(\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\frac{1}{\sqrt{K}\sqrt{K^*}}\Lambda'\widehat{\Sigma}^{-1}\widehat{\lambda}\right)^2\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)\right. \\
&\qquad +2\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\Lambda'\widehat{\Sigma}^{-1}\widehat{\lambda}\frac{1}{\sqrt{K}\sqrt{K^*}}\right)\left(\frac{1}{\sqrt{K}}\widehat{\lambda}'\widehat{\Sigma}^{-1}\frac{1}{N}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
&\qquad \left.+\left(\frac{1}{\sqrt{K}}\widehat{\lambda}'\widehat{\Sigma}^{-1}\frac{1}{N}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)^2\right) \\
&\quad +\left[\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\right]^2\frac{\left(\widehat{\lambda}'\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda}\right)}{P_s^2}\left(\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\frac{1}{\sqrt{K}\sqrt{K^*}}\Lambda'\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\right. \\
&\qquad \left.+\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\frac{1}{\sqrt{K}}\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\right)^2
\end{aligned}$$

$$
\begin{aligned}
\overset{U_p}{\underset{N\to\infty}{\to}} \quad & \left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\Lambda'\Sigma^{-1}\Lambda\frac{1}{K^*}\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)+2\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\Lambda'\widetilde{\Gamma}(\alpha)}{K^*}\right)+\left(\frac{\widetilde{\Gamma}(\alpha)'\Sigma\widetilde{\Gamma}(\alpha)}{K^*}\right)\\
& -2\left[\frac{1}{1+\frac{P_{s0}}{KP_s}\Lambda'\Sigma^{-1}\Lambda}\right]\frac{1}{P_s}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda+\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right)^2\\
& +\left[\frac{1}{1+\frac{P_{s0}}{KP_s}\Lambda'\Sigma^{-1}\Lambda}\right]^2\frac{1}{P_s^2}\left(\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)^2\left(\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\right.\\
& \quad +2\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right)\\
& \quad \left.+\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right)^2\left(P_{s0}\Lambda'\frac{1}{K}\Sigma^{-1}\Lambda\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
\overset{U_p}{\to} \quad & E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)^2 P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+2E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)P_{s0}+P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]\\
& -2\left[\frac{1}{1+\frac{P_{s0}}{P_s}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right]\frac{1}{P_s}\left(E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\left(P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)\right)^2E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\right.\\
& \quad \left.+2E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)+\left(P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\right)^2\right)\\
& +\left[\frac{1}{1+\frac{P_{s0}}{P_s}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right]^2\frac{1}{P_s^2}\left(E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\left(P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)\right)^2P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\right.\\
& \quad \left.+2E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\left(P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)\right)\left(P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)\right)P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)+\left(P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\right)^2\left(P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)\right)\right)\\
= \quad & P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]+\left(E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\right)^2\left[P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+2P_{s0}-2\frac{\left(P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}\right)^2}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right.\\
& \quad \left.+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)\left[\frac{P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right]^2\right]
\end{aligned}
$$

Result L.2.3

$$\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{W}_i'\widehat{\Gamma}(\theta)\widetilde{W}_i'\widehat{\Gamma}(\theta)\right)$$
$$=\frac{1}{N}\widehat{\Gamma}(\theta)'\widetilde{W}'\widetilde{W}\widehat{\Gamma}(\theta)$$
$$=\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)$$
$$=\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\left(\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'+V'\right)\left(\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'+V\right)\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)$$
$$=\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\frac{1}{K^*}\Lambda\widetilde{F}'\widetilde{F}\Lambda'\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)$$
$$+2\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'V\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)$$
$$+\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}V'V\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)$$

Lets consider each of the three pieces above.

First

$$\begin{aligned}
&\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_{s0}K}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}V'V\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
=\;&\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\widehat{\Sigma}^{-1}-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda\lambda}'\widehat{\Sigma}^{-1}\right]V'V\\
&\times\left[\widehat{\Sigma}^{-1}-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda\lambda}'\widehat{\Sigma}^{-1}\right]\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
=\;&\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
&-2\left(\frac{1}{P_s+\frac{1}{K}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\right)\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\widehat{\Sigma}^{-1}V'V\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda\lambda}'\widehat{\Sigma}^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
&+\left(\frac{1}{P_s+\frac{1}{K}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\right)^2\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda\lambda}'\widehat{\Sigma}^{-1}V'V\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda\lambda}'\widehat{\Sigma}^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)
\end{aligned}$$

Now deal with each of these three pieces

$$
\begin{aligned}
& \frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)' W\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}W'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
= \quad & \frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\left(\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'+V\right)\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}\left(\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'+V'\right)\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
= \quad & \frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
& +2\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
& +\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
= \quad & \left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\frac{1}{NK^*}\Lambda'\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}\Lambda\right)\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
& +2\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\frac{1}{N^2\sqrt{K^*}}\Lambda'\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
& +\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
\underset{N\to\infty}{\overset{p}{\to}} \quad & \left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\Lambda'\Sigma^{-1}\Lambda}{K^*}\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right) \\
& +2\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\Lambda'\widetilde{\Gamma}(\alpha)}{K^*}\right)+\frac{\widetilde{\Gamma}(\alpha)'\Sigma\widetilde{\Gamma}(\alpha)}{K^*} \\
\overset{p}{\to} \quad & E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E(\widetilde{\Gamma}_j(\alpha)\Lambda_j) \\
& +2E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)P_{s0}E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right)
\end{aligned}
$$

$$
\begin{aligned}
& \frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\widehat{\Sigma}^{-1}V'V\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda\lambda}'\widehat{\Sigma}^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
= \quad & \frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\left(\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'+V\right)\widehat{\Sigma}^{-1}V'V\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda\lambda}'\widehat{\Sigma}^{-1}\left(\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'+V'\right)\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
= \quad & \frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'\widehat{\Sigma}^{-1}V'V\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda\lambda}'\widehat{\Sigma}^{-1}\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
& +\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'\widehat{\Sigma}^{-1}V'V\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda\lambda}'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
& +\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}V'V\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda\lambda}'\widehat{\Sigma}^{-1}\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
& +\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}V'V\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda\lambda}'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
= \quad & \left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)\left(\frac{1}{KN}\Lambda'\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{K^*}\widehat{\lambda}'\widehat{\Sigma}^{-1}\Lambda\right)\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
& +\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)\left(\frac{1}{NK}\Lambda'\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{N\sqrt{K^*}}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
& +\left(\frac{1}{N^2\sqrt{K}}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}V'V\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{\sqrt{K}\sqrt{K^*}}\widehat{\lambda}'\widehat{\Sigma}^{-1}\Lambda\right)\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
& +\left(\frac{1}{N^2\sqrt{K}}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}V'V\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{N\sqrt{K}}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
\xrightarrow[N\to\infty]{p} \quad & \left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\sqrt{P_{s0}}\Lambda'\Sigma^{-1}\Lambda}{K}\right)\left(\frac{\sqrt{P_{s0}}}{K^*}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right) \\
& +\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\sqrt{P_{s0}}\Lambda'\Sigma^{-1}\Lambda}{K}\right)\left(\frac{\sqrt{P_{s0}}\Lambda'\widetilde{\Gamma}(\alpha)}{K^*}\right) \\
& +\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Gamma\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right) \\
& +\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right) \\
= \quad & \left[\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\sqrt{P_{s0}}\Lambda'\Sigma^{-1}\Lambda}{\sqrt{K}\sqrt{K^*}}\right)+\left(\frac{\sqrt{P_{s0}}\Lambda'\widetilde{\Gamma}(\alpha)}{\sqrt{K}\sqrt{K^*}}\right)\right]^2 \\
\xrightarrow{p} \quad & \left[E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)\right]^2 \\
= \quad & E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)^2\left[P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}\right]^2
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'V\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
&\qquad=\left(\frac{1}{N}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'V\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{\sqrt{K}N}\widehat{\lambda}'\widehat{\Sigma}^{-1}\left(\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'+V'\right)\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)^2\\
&\qquad=\left(\frac{1}{N}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'V\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\\
&\qquad\quad\times\left(\left(\widehat{\lambda}'\widehat{\Sigma}^{-1}\Lambda\frac{1}{\sqrt{K}\sqrt{K^*}}\right)\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)+\frac{1}{\sqrt{K}N}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)^2\\
&\qquad\xrightarrow[N\to\infty]{p}\left(\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\frac{\widetilde{\Gamma}^*(\alpha)'\Lambda^*}{K^*}+\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right)^2\\
&\qquad\xrightarrow{p}P_{s0}^3E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)^2\left(E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+1\right)^2
\end{aligned}
$$

Now put these three pieces together

$$
\begin{aligned}
&\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}V'V\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
&\quad\underset{N\to\infty}{\overset{U_p}{\to}}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\Lambda'\Sigma^{-1}\Lambda}{K^*}\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)+2\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\Lambda'\widetilde{\Gamma}(\alpha)}{K^*}\right)+\frac{\widetilde{\Gamma}(\alpha)'\Sigma\widetilde{\Gamma}(\alpha)}{K^*}\\
&\qquad-2\left(\frac{1}{P_s+\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda}\right)\left[\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\sqrt{P_{s0}}\Lambda'\Sigma^{-1}\Lambda}{\sqrt{K}\sqrt{K^*}}\right)+\left(\frac{\sqrt{P_{s0}}\Lambda'\widetilde{\Gamma}(\alpha)}{\sqrt{K}\sqrt{K^*}}\right)\right]^2\\
&\qquad+\left(\frac{1}{P_s+\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda}\right)^2\left(\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda\right)\\
&\qquad\quad\times\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widehat{\Sigma}^{-1}\Lambda\frac{\Gamma^{*\prime}\Lambda^*}{K^*}+\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right)^2\\
&\quad\overset{U_p}{\to}E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)+2E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)P_{s0}E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right)\\
&\qquad-2\frac{\left[P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}\right]^2}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)^2\\
&\qquad+\left(\frac{E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+1}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right)^2P_{s0}^3E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)^2
\end{aligned}
$$

which is the third of the three terms above.

Next consider the first

$$
\begin{aligned}
&\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\frac{1}{K^*}\Lambda\widetilde{F}'\widetilde{F}\Lambda'\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
=&\left(\frac{1}{N}\widetilde{F}'\widetilde{F}\right)\left(\frac{1}{N\sqrt{K}}\Lambda'\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)^2
\end{aligned}
$$

Focus on the term

$$
\begin{aligned}
&\frac{1}{N\sqrt{K^*}}\Lambda'\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
&\quad=\frac{1}{N\sqrt{K^*}}\Lambda'\left[\widehat{\Sigma}^{-1}-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\right]\left(\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'+V'\right)\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
&\quad=\frac{1}{N\sqrt{K^*}}\Lambda'\widehat{\Sigma}^{-1}\frac{1}{\sqrt{K^*}}\Lambda F'\left(\widetilde{Y}-\alpha\widetilde{T}\right)+\frac{1}{N\sqrt{K^*}}\Lambda'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
&\qquad-\frac{1}{N\sqrt{K^*}}\Lambda'\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
&\qquad+\frac{1}{N\sqrt{K^*}}\Lambda'\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
&\quad=\left(\frac{1}{K^*}\Lambda'\widehat{\Sigma}^{-1}\Lambda\right)\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)+\frac{1}{N\sqrt{K^*}}\Lambda'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
&\qquad-\left(\frac{1}{P_s+\frac{1}{K}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\right)\left(\frac{1}{\sqrt{K^*}\sqrt{K}}\Lambda'\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{\sqrt{K^*}\sqrt{K}}\widehat{\lambda}'\widehat{\Sigma}^{-1}\Lambda\right)\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)\\
&\qquad-\left(\frac{1}{P_s+\frac{1}{K}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\right)\left(\frac{1}{\sqrt{K^*}\sqrt{K}}\Lambda'\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{N\sqrt{K}}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)\\
&\underset{N\to\infty}{\overset{U_p}{\to}}\left(\frac{1}{K^*}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)+\frac{\widetilde{\Gamma}(\alpha)'\Lambda}{K^*}\\
&\qquad-\left(\frac{1}{P_s+\frac{P_{s0}}{K}\Lambda'\Sigma^{-1}\Lambda}\right)\left[\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)^2\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)+\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\widetilde{\Gamma}(\alpha)'\Lambda\right)\right]\\
&\overset{U_p}{\to}P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)+P_{s0}E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)\\
&\qquad-\left(\frac{1}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right)\left[\left(P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)\right)^2E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)P_{s0}E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)\right]\\
&\quad=P_{s0}E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)\left[1-\frac{P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right]
\end{aligned}
$$

Now finally the middle piece

$$
\begin{aligned}
&\frac{1}{N^3}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'V\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
=&\left\{\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\frac{1}{\sqrt{K^*}}\Lambda\right\}\left\{\frac{1}{N^2}\widetilde{F}'V\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right\}
\end{aligned}
$$

Note that we have already derived the plim of $\left\{\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\frac{1}{\sqrt{K^*}}\Lambda\right\}$. So the final piece to deal with is

$$
\begin{aligned}
& \frac{1}{N^2} F'V\left[\frac{1}{P_s K}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1} W'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
= \quad & \frac{1}{N^2}\widetilde{F}'V\left[\widehat{\Sigma}^{-1}-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\right]\left(\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'+V'\right)\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
= \quad & \frac{1}{N^2}\widetilde{F}'V\widehat{\Sigma}^{-1}\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)+\frac{1}{N^2}\widetilde{F}'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
& -\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{N^2}\widetilde{F}'V\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
& -\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{N^2}\widetilde{F}'V\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
= \quad & \left(\frac{1}{N\sqrt{K^*}}\widetilde{F}'V\widehat{\Sigma}^{-1}\Lambda\right)\left(\frac{1}{N}\widetilde{F}'Y\right)+\frac{1}{N^2}\widetilde{F}'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
& -\left(\frac{1}{P_s+\frac{1}{K}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\right)\left(\frac{1}{N\sqrt{K^*}}\widetilde{F}'V\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{K}\widehat{\lambda}'\widehat{\Sigma}^{-1}\Lambda\right)\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
& -\left(\frac{1}{P_s+\frac{1}{K}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\right)\left(\frac{1}{N\sqrt{K}}\widetilde{F}'V\frac{1}{K}\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{N\sqrt{K}}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
\underset{N\to\infty}{\overset{U_p}{\to}} \quad & 0
\end{aligned}
$$

Putting the pieces together we get the result.

Result L.2.4

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{W}_i'\widehat{\Gamma}(\theta)\left[\widetilde{Y}_i-\alpha\widetilde{T}_i\right]\right)=&\frac{1}{N^2}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
=&\frac{1}{N^2}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{W}\left[\widehat{\Sigma}^{-1}-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\right]\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
=&\frac{1}{N^2}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\left(\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'+V\right)\left[\widehat{\Sigma}^{-1}-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\right]\\
&\times\left(\frac{1}{\sqrt{K^*}}\widetilde{F}\Lambda'+V\right)'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
=&\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\frac{1}{K^*}\Lambda'\widehat{\Sigma}^{-1}\Lambda\right)\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)\\
&+2\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\frac{1}{\sqrt{K^*}N}\Lambda'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)\\
&-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{P_s}\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\frac{1}{\sqrt{K^*}\sqrt{K}}\Lambda'\widehat{\Sigma}^{-1}\widehat{\lambda}\right)^2\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)\\
&-\frac{2}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{P_s}\left(\frac{1}{N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'\widetilde{F}\right)\left(\frac{1}{\sqrt{K^*}\sqrt{K}}\Lambda'\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{\sqrt{K}N}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)\\
&+\frac{1}{N^2}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\\
&-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{P_s}\left(\frac{1}{\sqrt{K}N}\left(\widetilde{Y}-\alpha\widetilde{T}\right)'V\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{\sqrt{K}N}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
\underset{N\to\infty}{\overset{U_p}{\to}}\quad&\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\Lambda'\Sigma^{-1}\Lambda}{K^*}\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)+2\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\widetilde{\Gamma}(\alpha)'\Lambda}{K^*}\right)\\
&-\frac{1}{1+\frac{P_{s0}}{KP_s}\Lambda'\Sigma^{-1}\Lambda}\frac{1}{P_s}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)^2\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\\
&-\frac{2}{1+\frac{P_{s0}}{KP_s}\Lambda'\Sigma^{-1}\Lambda}\frac{1}{P_s}\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Gamma\right)+\frac{\widetilde{\Gamma}(\alpha)'\Sigma\widetilde{\Gamma}(\alpha)}{K^*}\\
&\frac{1}{1+\frac{P_{s0}}{KP_s}\Lambda'\Sigma^{-1}\Lambda}\frac{1}{P_s}\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right)^2\\
=\quad&\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\Lambda'\Sigma^{-1}\Lambda}{K^*}\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)+2\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)\left(\frac{\widetilde{\Gamma}(\alpha)'\Lambda}{K^*}\right)+\frac{\widetilde{\Gamma}(\alpha)'\Sigma\widetilde{\Gamma}(\alpha)}{K^*}\\
&-\frac{\left[\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)+\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\widetilde{\Gamma}(\alpha)\right]^2}{P_s+\frac{1}{K}\Lambda'\Sigma^{-1}\Lambda}\\
&\overset{U_p}{\to}E(\Lambda_j\widetilde{\Gamma}_j(\alpha))^2P_{s0}E(\Lambda_j^2/\sigma_j^2)+2E(\Lambda_j\Gamma_j)P_{s0}E\left(\Lambda_j\Gamma_j\right)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right)\\
&-E(\Lambda_j\widetilde{\Gamma}_j(\alpha))^2\frac{\left[P_{s0}E(\Lambda_j^2/\sigma_j^2)+P_{s0}\right]^2}{P_s+P_{s0}E(\Lambda_j^2/\sigma_j^2)}
\end{aligned}
$$

Result L2.5

$$
\begin{aligned}
\frac{1}{\sqrt{K}}\widehat{\lambda}'\widehat{\Gamma}(\theta) &= \frac{1}{\sqrt{K}N}\widehat{\lambda}'\left[\frac{1}{P_sK}\widehat{\lambda}'\widehat{\lambda}+\widehat{\Sigma}\right]^{-1}\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
&= \frac{1}{\sqrt{K}N}\widehat{\lambda}'\left[\widehat{\Sigma}^{-1}-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\right]\widetilde{W}'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
&= \frac{1}{\sqrt{K}N}\widehat{\lambda}'\left[\widehat{\Sigma}^{-1}-\frac{1}{1+\frac{1}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{KP_s}\widehat{\Sigma}^{-1}\widehat{\lambda}\widehat{\lambda}'\widehat{\Sigma}^{-1}\right]\left(\frac{1}{\sqrt{K^*}}\Lambda\widetilde{F}'+V'\right)\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
&= \left(\frac{1}{\sqrt{K}\sqrt{K^*}}\widehat{\lambda}'\widehat{\Sigma}^{-1}\Lambda\right)\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
&\quad -\frac{1}{1+\frac{P_{s0}}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{P_s}\left(\frac{1}{K}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{\sqrt{K}\sqrt{K^*}}\widehat{\lambda}'\widehat{\Sigma}^{-1}\Lambda\right)\left(\frac{1}{N}\widetilde{F}'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
&\quad +\frac{1}{\sqrt{K}N}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right) \\
&\quad -\frac{1}{1+\frac{P_{s0}}{KP_s}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}}\frac{1}{P_s}\left(\frac{1}{K}\widehat{\lambda}'\widehat{\Sigma}^{-1}\widehat{\lambda}\right)\left(\frac{1}{\sqrt{K}N}\widehat{\lambda}'\widehat{\Sigma}^{-1}V'\left(\widetilde{Y}-\alpha\widetilde{T}\right)\right) \\
&\underset{N\to\infty}{\overset{U_p}{\to}} \left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right) \\
&\quad -\frac{1}{P_s+\frac{P_{s0}\Lambda'\Sigma^{-1}\Lambda}{K}}\left(\frac{P_{s0}\Lambda'\Sigma^{-1}\Lambda}{K}\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right) \\
&\quad +\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\widetilde{\Gamma}(\alpha)'\Lambda-\frac{1}{P_s+\frac{P_{s0}\Lambda'\Sigma^{-1}\Lambda}{K}}\left(\frac{P_{s0}\Lambda'\Sigma^{-1}\Lambda}{K}\right)\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\widetilde{\Gamma}(\alpha)'\Lambda\right) \\
&= \left[\left(\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\Lambda'\Sigma^{-1}\Lambda\right)\left(\frac{\Lambda^{*\prime}\widetilde{\Gamma}^*(\alpha)}{K^*}\right)+\frac{\sqrt{P_{s0}}}{\sqrt{K}\sqrt{K^*}}\widetilde{\Gamma}(\alpha)'\Lambda\right]\left[\frac{P_s}{P_s+\frac{P_{s0}\Lambda'\Sigma^{-1}\Lambda}{K}}\right] \\
&\overset{U_p}{\to} \left[P_{s0}E(\Lambda_j^2/\sigma_j^2)E(\Lambda_j\widetilde{\Gamma}_j(\alpha))+P_{s0}E(\Lambda_j\widetilde{\Gamma_j}(\alpha))\right]\left[\frac{P_s}{P_s+E(\Lambda_j^2/\sigma_j^2)P_{s0}}\right] \\
&= P_sE(\Lambda_j\widetilde{\Gamma}_j(\alpha))\left[\frac{P_{s0}E(\Lambda_j^2/\sigma_j^2)+P_{s0}}{P_s+E(\Lambda_j^2/\sigma_j^2)P_{s0}}\right]
\end{aligned}
$$

Proof of Theorem 6:

The vast majority of work for this proof has been done in Lemmas 1 and 2.

To simplify the notation, define

$$
a(P_s,\alpha)\equiv\left(E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\right)^2\left[P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+2P_{s0}-2\frac{\left(P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}\right)^2}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)\left[\frac{P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right]^2\right]
$$

and note that

$$
a(P_{s0})=0.
$$

First consider

$$
\begin{aligned}
q^1_{N,K^*}(\theta) \quad &= \quad \frac{1}{N}\sum_{i=1}^{N}\widetilde{W}_i'\widehat{\Gamma}(\theta)\widetilde{Z}_i - \frac{\phi P_s\sigma_\xi^2}{(1-P_s)\widehat{\Gamma}_{K^*}(P_s)'\widehat{\Sigma}\widehat{\Gamma}_{K^*}(P_s)+P_s\sigma_\xi^2}\frac{1}{N}\sum_{i=1}^{N}\widetilde{W}_i'\widehat{\Gamma}(\theta)\widetilde{W}_i'\widehat{\Gamma}(\theta) \\
&\quad -\frac{\phi(1-P_s)\widehat{\Gamma}(\theta)'\widehat{\Sigma}\widehat{\Gamma}(\theta)}{(1-P_s)\widehat{\Gamma}(\theta)'\widehat{\Sigma}\widehat{\Gamma}(\theta)+\sigma_\xi^2}\frac{1}{N}\sum_{i=1}^{N}\widetilde{W}_i'\widehat{\Gamma}(\theta)\left[\widetilde{Y}_i-\alpha\widetilde{T}_i\right]
\end{aligned}
$$

From Lemmas 1 and 2, this converges uniformly to

$$
\begin{aligned}
& E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E(\Lambda_j^2/\sigma_j^2)E(\Lambda_j\beta_j)+2E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E\left(\Lambda_j\beta_j\right)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)\beta_j\right) \\
& -\frac{E(\Lambda_j\widetilde{\Gamma}_j(\alpha))E(\Lambda_j\beta_j)P_{s0}^2\left[E(\Lambda_j^2/\sigma_j^2)+1\right]^2}{P_s+P_{s0}E(\Lambda_j^2/\sigma_j^2)} \\
& -\left(\frac{\phi P_s\sigma_\xi^2}{(1-P_s)\left[P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]+a(P_s)\right]+P_s\sigma_\xi^2}\right)\times \\
& \Bigg[E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)+2E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)P_{s0}E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right) \\
& -2\frac{\left[P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}\right]^2}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}E\left(\Gamma_j\Lambda_j\right)^2+\left(\frac{E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+1}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right)^2P_{s0}^3E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)^2 \\
& +\left[P_{s0}E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)\left[1-\frac{P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right]\right]^2\Bigg] \\
& -\left(\frac{\phi(1-P_s)\left[P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]+a(P_s)\right]}{(1-P_s)\left[P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]+a(P_s)\right]+P_s\sigma_\xi^2}\right)\times \\
& \Bigg[E(\Lambda_j\widetilde{\Gamma}_j(\alpha))^2P_{s0}E(\Lambda_j^2/\sigma_j^2)+2E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right) \\
& -E(\Lambda_j\widetilde{\Gamma}_j(\alpha))^2\frac{\left[P_{s0}E(\Lambda_j^2/\sigma_j^2)+P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\right]^2}{P_s+P_{s0}E(\Lambda_j^2/\sigma_j^2)}\Bigg]
\end{aligned}
$$

Substituting θ_0 for θ yields

$$
\begin{aligned}
& E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)P_{s0}E(\Lambda_j^2/\sigma_j^2)E(\Lambda_j\beta_j)+2E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)P_{s0}E\left(\Lambda_j\beta_j\right)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)\beta_j\right)\\
& -E(\Lambda_j\widetilde{\Gamma}_j(\alpha))E(\Lambda_j\beta_j)P_{s0}\left[E(\Lambda_j^2/\sigma_j^2)+1\right]\\
& -\left(\frac{\phi_0P_{s0}\sigma_{\xi0}^2}{(1-P_{s0})P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]+P_{s0}\sigma_{\xi0}^2}\right)\times\\
& \left[E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)+2E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)P_{s0}E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)+P_{s0}E\left(\sigma_j^2\Gamma_j^2\right)-2P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)^2\right.\\
& \left.-2P_{s0}E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)^2+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)^2+\left[P_{s0}E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)^2\right]^2\right]\\
& -\left(\frac{\phi_0(1-P_s)P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]}{(1-P_{s0})P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]+P_{s0}\sigma_{\xi0}^2}\right)\times\\
& \left[E(\Lambda_j\widetilde{\Gamma}_j(\alpha))^2P_{s0}E(\Lambda_j^2/\sigma_j^2)+2E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right)\right.\\
& \left.-E(\Lambda_j\widetilde{\Gamma}_j(\alpha))^2\left[P_{s0}E(\Lambda_j^2/\sigma_j^2)+P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\right]\right]\\
= & E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right))PE\left(\Lambda_j\beta_j\right)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)\beta_j\right)\\
& -\phi_0\left(\frac{P_{s0}\sigma_{\xi0}^2}{(1-P_{s0})P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]+P_{s0}\sigma_{\xi0}^2}\right)\left[P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right)+\left[P_{s0}E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)\right]^2\right]\\
& -\phi_0\left(\frac{(1-P_{s0})P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]}{(1-P_{s0})P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]+P_{s0}\sigma_{\xi0}^2}\right)\left[E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right)\right]\\
= & P_{s0}\left(E(\Lambda_j\widetilde{\Gamma}_j(\alpha))E\left(\Lambda_j\beta_j\right)+E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)\beta_j\right)\right.\\
& \qquad\qquad\qquad\qquad\frac{-\phi_0}{(1-P_{s0})P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]+P_{s0}\sigma_{\xi0}^2}\times\\
& \left[\sigma_{\xi0}^2\left[P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right)+\left[P_{s0}E(\widetilde{\Gamma}_j(\alpha)\Lambda_j)\right]^2\right]+(1-P_s)E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]\left[E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right)\right]\right]\\
= & 0
\end{aligned}
$$

Next consider

$$
\begin{aligned}
q_{N,K^*}^2(\theta) &= \frac{1}{N}\sum_{i=1}^N\left(\widetilde{Y}_i-\alpha\widetilde{T}_i-\widetilde{W}_i'\widehat{\Gamma}(\theta)\right)\times\\
&\quad\left[\widetilde{Z}_i-\phi\widetilde{W}_i'\widehat{\Gamma}(\theta)-\frac{\phi(1-P_s)\widehat{\Gamma}(\theta)'\widehat{\Sigma}\widehat{\Gamma}(\theta)}{(1-P_s)\widehat{\Gamma}_{K^*}(P_s)'\widehat{\Sigma}\widehat{\Gamma}_{K^*}(P_s)+P_s\sigma_\xi^2}\left(\widetilde{Y}_i-\alpha\widetilde{T}_i-\widetilde{W}_i'\widehat{\Gamma}_{K^*}(P_s)\right)\right]\\
&= \frac{1}{N}\sum_{i=1}^N\left(\widetilde{Y}_i-\alpha\widetilde{T}_i\right)\left[\widetilde{Z}_i-\phi\widetilde{W}_i'\widehat{\Gamma}(\theta)-\frac{\phi(1-P_s)\widehat{\Gamma}_{K^*}(P_s)'\widehat{\Sigma}\widehat{\Gamma}_{K^*}(P_s)}{(1-P_s)\widehat{\Gamma}(\theta)'\widehat{\Sigma}\widehat{\Gamma}(\theta)+P_s\sigma_\xi^2}\left(\widetilde{Y}_i-\alpha\widetilde{T}_i-\widetilde{W}_i'\widehat{\Gamma}_{K^*}(P_s)\right)\right]\\
&\quad -q_{N,K^*}^1(\theta)\\
&= \frac{1}{N}\sum_{i=1}^N\left(\widetilde{Y}_i-\alpha\widetilde{T}_i\right)\widetilde{Z}_i-\frac{\phi P_s\sigma_\xi^2}{(1-P_s)\widehat{\Gamma}(\theta)'\widehat{\Sigma}\widehat{\Gamma}(\theta)+P_s\sigma_\xi^2}\frac{1}{N}\sum_{i=1}^N\left(\widetilde{Y}_i-\alpha\widetilde{T}_i\right)\widetilde{W}_i'\widehat{\Gamma}_{K^*}(P_s)\\
&\quad -\frac{\phi(1-P_s)\widehat{\Gamma}(\theta)'\widehat{\Sigma}\widehat{\Gamma}(\theta)}{(1-P_s)\widehat{\Gamma}(\theta)'\widehat{\Sigma}\widehat{\Gamma}(\theta)+P_s\sigma_\xi^2}\frac{1}{N}\sum_{i=1}^N\left(\widetilde{Y}_i-\alpha\widetilde{T}_i\right)-q_{N,K*}^1(\theta).
\end{aligned}
$$

This expression converges uniformly in probability to

$$
\begin{aligned}
& E(\Lambda_j\widetilde{\Gamma}_j(\alpha))E(\Lambda_j\beta_j)+E\left(\widetilde{\Gamma}_j(\alpha)\beta_j\sigma_j^2\right)-p\lim(q^1_{N,K*}(\theta)) \\
& -\left(\frac{\phi P_s\sigma_\xi^2}{(1-P_s)\left[P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]+a(P_s)\right]+P_s\sigma_\xi^2}\right)\times \\
& \left[E(\Lambda_j\widetilde{\Gamma}_j(\alpha))^2P_{s0}E(\Lambda_j^2/\sigma_j^2)+2E(\Lambda_j\widetilde{\Gamma}_j(\alpha))P_{s0}E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)+P_{s0}E\left(\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right)-E(\Lambda_j\widetilde{\Gamma}_j(\alpha))^2\frac{\left[P_{s0}E(\Lambda_j^2/\sigma_j^2)+P_{s0}\right]^2}{P_s+P_{s0}E(\Lambda_j^2/\sigma_j^2)}\right] \\
& -\left(\frac{\phi(1-P_s)\left[P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]+a(P_s)\right]}{(1-P_s)\left[P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]+a(P_s)\right]+P_s\sigma_\xi^2}\right)\times \\
& \left[E(\Lambda_j\widetilde{\Gamma}_j(\alpha))^2+E\left(\widetilde{\Gamma}_j(\alpha)^2\sigma_j^2\right)+\sigma_{\xi 0}^2\right]
\end{aligned}
$$

Plugging in θ_0 for θ yields

$$
\begin{aligned}
& E(\Lambda_j\Gamma_j)E(\Lambda_j\beta_j)+E\left(\Gamma_j\beta_j\sigma_j^2\right) \\
& -\phi_0\left(\frac{P_{s0}\sigma_\xi^2}{(1-P_{s0})P_{s0}E\left[\sigma_j^2\Gamma_j^2\right]+P_{s0}\sigma_\xi^2}\right)\times \\
& \left[E(\Lambda_j\Gamma_j)^2P_{s0}E(\Lambda_j^2/\sigma_j^2)+2E(\Lambda_j\Gamma_j)P_{s0}E\left(\Lambda_j\Gamma_j\right)+P_{s0}E\left(\sigma_j^2\Gamma_j^2\right)-E(\Lambda_j\Gamma_j)^2\left[P_{s0}E(\Lambda_j^2/\sigma_j^2)+P_{s0}\right]\right] \\
& -\phi_0\left(\frac{(1-P_{s0})P_{s0}E\left[\sigma_j^2\Gamma_j^2\right]}{(1-P_{s0})P_{s0}E\left[\sigma_j^2\Gamma_j^2\right]+P_{s0}\sigma_\xi^2}\right)\left[E(\Lambda_j\Gamma_j)^2+E\left(\Gamma_j^2\sigma_j^2\right)+\sigma_{\xi 0}^2\right] \\
= \; & E(\Lambda_j\Gamma_j)E(\Lambda_j\beta_j)+E\left(\Gamma_j\beta_j\sigma_j^2\right) \\
& -\phi_0\left(\frac{P_{s0}\sigma_{\xi 0}^2}{(1-P_{s0})\left[P_{s0}E\left[\sigma_j^2\Gamma_j^2\right]\right]+P_{s0}\sigma_\xi^2}\right)\left[E(\Lambda_j\Gamma_j)^2P_{s0}+P_{s0}E\left(\sigma_j^2\Gamma_j^2\right)\right] \\
& -\phi_0\left(\frac{(1-P_{s0})P_{s0}E\left[\sigma_j^2\Gamma_j^2\right]}{(1-P_{s0})P_{s0}E\left[\sigma_j^2\Gamma_j^2\right]+P_{s0}\sigma_\xi^2}\right)\left[E(\Lambda_j\Gamma_j)^2+E\left(\Gamma_j^2\sigma_j^2\right)+\sigma_{\xi 0}^2\right] \\
= \; & E(\Lambda_j\Gamma_j)E(\Lambda_j\beta_j)+E\left(\Gamma_j\beta_j\sigma_j^2\right) \\
& -\phi_0\left(\frac{P_{s0}\sigma_{\xi 0}^2\left[E(\Lambda_j\Gamma_j)^2P_{s0}+P_{s0}E\left(\sigma_j^2\Gamma_j^2\right)\right]+(1-P_{s0})P_{s0}E\left[\sigma_j^2\Gamma_j^2\right]\left[E(\Lambda_j\Gamma_j)^2+E\left(\Gamma_j^2\sigma_j^2\right)+\sigma_{\xi 0}^2\right]}{(1-P_{s0})\left[P_{s0}E\left[\sigma_j^2\Gamma_j^2\right]\right]+P_{s0}\sigma_\xi^2}\right) \\
= \; & E(\Lambda_j\Gamma_j)E(\Lambda_j\beta_j)+E\left(\Gamma_j\beta_j\sigma_j^2\right) \\
& -\phi_0\left(\frac{P_{s0}\sigma_{\xi 0}^2\left[E(\Lambda_j\Gamma_j)^2P_{s0}+E\left(\sigma_j^2\Gamma_j^2\right)\right]+(1-P_{s0})P_{s0}E\left[\sigma_j^2\Gamma_j^2\right]\left[E(\Lambda_j\Gamma_j)^2+E\left(\Gamma_j^2\sigma_j^2\right)\right]}{(1-P_{s0})\left[P_{s0}E\left[\sigma_j^2\Gamma_j^2\right]\right]+P_{s0}\sigma_\xi^2}\right) \\
= \; & 0
\end{aligned}
$$

Finally consider

$$
q^3_{N,K^*}(\theta)=\frac{1}{N}\sum_{i=1}^N\left(\widetilde{Y}_i-\alpha\widetilde{T}_i\right)^2-\left(\frac{\widehat{\Gamma}(\theta)'\widehat{\lambda}}{P_s}\right)^2-\frac{\widehat{\Gamma}(\theta)'\widehat{\Sigma}\widehat{\Gamma}(\theta)}{P_s}-\sigma_\xi^2,
$$

which converges uniformly in probability to

$$
\begin{aligned}
& E\left(\widetilde{\Gamma}_j(\alpha)\Lambda_j\right)^2 + E(\widetilde{\Gamma}_j(\alpha)^2\sigma_j^2) + \sigma_{\xi 0}^2 - \left(\frac{P_s E\left(\Lambda_j \widetilde{\Gamma}_j(\alpha)\right)) \left[\frac{P_{s0}E(\Lambda_j^2/\sigma_j^2)+P_{s0}}{P_s+E(\Lambda_j^2/\sigma_j^2)P_{s0}}\right]}{P_s}\right)^2 \\
& -\frac{P_{s0}E\left[\sigma_j^2\widetilde{\Gamma}_j(\alpha)^2\right]}{P_s} \\
& -\frac{\left(E\left(\Lambda_j\widetilde{\Gamma}_j(\alpha)\right)\right)^2\left[P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+2P_{s0}-2\frac{\left(P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}\right)^2}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)\left[\frac{P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)+P_{s0}}{P_s+P_{s0}E\left(\frac{\Lambda_j^2}{\sigma_j^2}\right)}\right]^2\right]}{P_s} \\
& -\sigma_\xi^2
\end{aligned}
$$

Evaluated at $\theta = \theta_0$ this is

$$
\begin{aligned}
& E(\Gamma_j\Lambda_j)^2 + E(\Gamma_j^2\sigma_j^2) + \sigma_{\xi 0}^2 - (E(\Lambda_j\Gamma_j))^2 \\
& -\frac{P_{s0}E\left[\sigma_j^2\Gamma_j^2\right]}{P_{s0}} - \sigma_{\xi 0}^2 \\
= \; & 0.
\end{aligned}
$$

We have thus shown that $Q_0(\theta_0) = 0$, so $\theta_0 \in \Theta_I$. Since the convergence of $q_{N,K^*}(\theta)$ is uniform, the convergence of $Q_{N,K^*}(\theta)$ to $Q_0(\theta)$ is uniform as well. Given that the parameter space of θ is compact, Assumption C.1 of Chernozhokov, Han, and Tamer (2007) is satisisfied. Applying their Theorem 3.1 gives the desired result.