A LIKELIHOOD-BASED APPROACH TO THE ANALYSIS OF A CLASS OF NESTED AND NON-NESTED MODELS

YUICHI KITAMURA*

Preliminary Draft

Abstract. It is a common practice in applied econometrics to write a model conditional on some variables, leaving the probability law of the conditioning variables unspecified. Assessing different specifications of conditional models is obviously an important task, for which many well-established methods are available. There are, however, cases to which these conventional methods do not apply easily. This paper proposes a likelihood-based measure of model fit that enables the researcher to evaluate a broad range of conditional models in a unified and coherent manner. The key idea is to introduce "likelihood" for semiparametric models such as conditional mean restriction models and conditional quantile restriction models. Some practical issues for implementing the method are addressed.

1. Introduction

An empirical economic model usually takes a conditional form. For example, an applied researcher routinely writes down a conditional likelihood function and uses the conditional ML estimator. If the researcher uses a parametric conditional model (with tight parameterization), it is often argued that the model should be regarded as an approximation of the underlying true conditional probability law. This notion has led to the extensive literature on the ML estimation for misspecified models. The well-known work by White (1982) discusses the asymptotic properties of ML estimators under misspecification. Vuong (1983) considers conditional ML estimators. One of the basic results in this literature is that the ML estimator converges to a pseudo-true value, which corresponds to the parametric probability measure that is closest to the true measure in terms of expected log likelihood

Date: This Version: April 2003.

Keywords: Conditional Moment Restrictions, Empirical Likelihood, Misspecification, Quantile Regression, Semiparametric Estimators.

JEL Classification Number: C14.

*I acknowledge financial support from the Alfred P. Sloan Foundation Research Fellowship and from the National Science Foundation via grant SES-9905247. I thank Chris Flinn for discussions that motivated materials reported in Section 6.4. I also appreciate helpful comments from Yongmiao Hong and several seminar audiences.

ratio. Vuong (1989) uses this property to proposes a model comparison test based on a likelihood ratio statistic.

Another way to deal with the potential of misspecification of conditional model is using more flexible models. For example, some semiparametric models such as conditional mean restriction models and conditional quantile restriction models avoid distributional assumptions. Obvious examples of such models include mean (quantile) regressions. Also, a standard nonlinear simultaneous equation model is identified by specifying a vector of mean zero residuals conditional on "instrumental variables." These are special cases of a conditional moment restriction model as considered by Chamberlain (1987) and Newey (1990). We shall give a general definition of a conditional moment restriction model shortly.

Using these semiparametric restrictions would certainly reduce the potential impact of misspecification. Even in a semiparametric model, however, it is often more realistic to consider its specification as an approximation. While some asymptotic results for these models under misspecification are available (e.g., White (1981), Kim and White (2002)), these are specific to estimation methods (e.g. the least square method) and are rather hard to reconcile with the elegant results obtained for parametric models as noted above.

This paper is concerned with these issues. It shows that the theory of misspecified ML directly extends to a class of semiparametric models. It then develops a likelihood-based approach that allows the researcher to evaluate the fitness of a broad range of models, including semiparametric models such as conditional mean/quantile restriction models. It applies to many situations that are not covered by conventional methods. For example, the new method enables us to compare a conditional mean restriction model and a conditional median restriction model, or conditional quantile restriction models for two different quantiles. It is also possible to compare a parametric likelihood model and a conditional mean/median restriction model with the method.

In sum, the main goal of this paper is to explore a unified, likelihood-based methodology for analyzing possibly misspecified and non-nested conditional models of various types. It makes three main contributions. First, it develops a unified asymptotic theory for a broad range of conditional models, thereby extending the White-Vuong results to models that are semiparametrically defined. Second, it proposes a new likelihood-based model fit measure which applies to both parametric models and semiparametric models. Third, it shows that a general form of likelihood-ratio test that can be used for model comparison.

2. Theory

2.1. **Framework.** Suppose the econometrician observes random samples $(x_i, z_i), i = 1, 2, ..., n$, taking values in a space $\mathcal{X} \times \mathcal{Z} \subset \mathbb{R}^s \times \mathbb{R}^d$. Denote the joint probability law of (x_i, z_i), the marginal law of x_i and the family of conditional law of z_i given $x \in \mathcal{X}$ by $\mu_{X,Z}$, μ_X and $(\mu_{Z|x})_{x\in\mathcal{X}}$, respectively. Note that the μ's are the true probability measures and not known *a priori* to the econometrician. The econometrician is interested in the unknown $(\mu_{Z|x})_{x\in\mathcal{X}}$, and specifies a set of candidate conditional probability measure families (so each element of this set consists of a family of conditional measures $(P_{Z|X=x})_{x\in\mathcal{X}}$, say), denoted by $\mathcal{P}_{Z|X}$. In other words, $\mathcal{P}_{Z|X}$ is the model specified by the econometrician, who wishes to learn about the true conditional measure $(\mu_{Z|x})_{x\in\mathcal{X}}$ from the data $\{(x_i, z_i)\}_{i=1}^n$. $\mathcal{P}_{Z|X}$ can be a parametric family or a semiparametric family. The marginal measure of x_i is left unspecified.

For a given $\mathcal{P}_{Z|X}$, consider a minimization problem as follows:

$$\text{minimize } I(\mu_{Z|X} \| P_{Z|X}) \text{ subject to } (P_{Z|x})_{x\in\mathcal{X}} \in \mathcal{P}_{Z|X}, \tag{2.1}$$

with respect to a divergence measure $I(\cdot\|\cdot)$ for conditional probability measures. In other words, we approximate the true $(\mu_{Z|x})_{x\in\mathcal{X}}$ using the set of conditional measures $\mathcal{P}_{Z|X}$ with respect to the criterion function $I(\cdot\|\cdot)$. A particularly convenient and important choice of $I(\cdot\|\cdot)$ is the expected log-likelihood ratio, also known as conditional relative entropy, defined as follows. Consider a random variables (U, W), say, with $w \in \mathcal{W}$. Suppose we consider two different sets of conditional measures of U given $w \in \mathcal{W}$, denoted by $(m^1_{U|w})_{w\in\mathcal{W}}$ and $(m^2_{U|w})_{w\in\mathcal{W}}$. Let m_W denote the marginal probability measure of W. Define

$$I(m^1_{U|W} \| m^2_{U|W}) := \int\int \log(dm^1_{U|w}/dm^2_{U|w}) dm^1_{U|w} dm_W \tag{2.2}$$

where the inner integral $\int \log(dm^1_{U|w}/dm^2_{U|w})dm^1_{U|w}$ is infinity if $m^1_{U|w}$ is not absolutely continuous with respect to $m^2_{U|w}$. This is essentially the expected value of the log likelihood ratio $\log(dm^1_{U|w}/dm^2_{U|w})$. $I(m^1_{U|W} \| m^2_{U|W})$ is also said to be the conditional relative entropy of $(m^1_{U|w})_{w\in\mathcal{W}}$ with respect to $(m^2_{U|w})_{w\in\mathcal{W}}$; see, for example, Cover and Thomas (1991) for details. The use of relative entropy is standard in the literature of model comparison. It has been applied to specification tests in econometrics (see Robinson (1991)). It is used in the recent literature of robust control in macroeconomics (Hansen and Sargent (2001)) as well as in the asset pricing literature (see Kitamura and Stutzer (2002) and papers cited therein).

Once the minimization problem (2.1) is solved, it is natural to use the attained minimum value

$$I^* = \inf_{(P_{Z|x})_{x\in\mathcal{X}} \in \mathcal{P}_{Z|X}} I(\mu_{Z|X} \| P_{Z|X}) \tag{2.3}$$

to evaluate the fitness of the model $\mathcal{P}_{Z|X}$ to the true $(\mu_{Z|x})_{x\in\mathcal{X}}$. This is similar in spirit to Hansen and Sargent (1993), who use a counterpart of relative entropy in terms of spectrum to study model approximation errors in rational expectation models. Equivalently, one may use its exponential transformation

$$\phi^* = e^{-I^*}. \tag{2.4}$$

This, by definition, belongs to the unit interval $(0, 1]$. A value of ϕ^* near one (zero) indicates that $\mathcal{P}_{Z|X}$ fits well (poorly) to $(\mu_{Z|x})_{x\in\mathcal{X}}$. In practice, of course, I^* has to be estimated. This is the topic of Section 4, which proposes a consistent estimator $\hat{I}^*$ for I^* and uses it (or $\hat{\phi} = \exp(-\hat{I}^*)$) as a measure of model fit.

The next two subsections discuss specific examples of $\mathcal{P}_{Z|X}$.

2.2. **Parametric $\mathcal{P}_{Z|X}$.** If the set of probability measures $\mathcal{P}_{Z|X}$ is parametrically defined, results obtained by White (1982) and Vuong (1983, 1989) directly apply. Suppose a finite dimensional parameter $\gamma \in \Gamma$ fully determines each conditional measure $(P^{\gamma}_{Z|X})_{x\in\mathcal{X}}$. Then

$$\mathcal{P}_{Z|X} = \cup_{\gamma\in\Gamma}(P^{\gamma}_{Z|x})_{x\in\mathcal{X}}. \tag{2.5}$$

The conditional ML estimator $\hat{\gamma}$ minimizes $\sum_{i=1}^n \log \frac{dP_{Z|x_i}}{dv}(z_i)$ (where v is an appropriate underlying measure). It converges to a pseudo-true value γ^* under mild regularity conditions given in Vuong (1983). This in turn gives rise to a pseudo-true conditional measure $(P^{\gamma^*}_{Z|X})_{x\in\mathcal{X}}$, which solves (2.1). In this sense $(P^{\gamma^*}_{Z|X})_{x\in\mathcal{X}}$ provides a best approximation for the unknown true measure $(\mu^{\gamma^*}_{Z|X})_{x\in\mathcal{X}}$. The quantity

$$I^* = I(\mu_{Z|X} \| P^{\gamma^*}_{Z|X}) \tag{2.6}$$

captures the goodness of approximation of the parametric model (2.5). Section 4 discusses how to estimate I^* (or ϕ^* as in (2.4)).

2.3. **Semiparametric $\mathcal{P}_{Z|X}$.** Next consider the case where $\mathcal{P}_{Z|X}$ is defined by a conditional moment restriction and thus semiparametric. This covers conditional mean restriction models and conditional quantile restriction models as special cases. Let $g(z,\theta) : \mathcal{Z} \times \Theta \to \mathbb{R}^q$ be a known function. Assume that the support of $g(z,\theta)$ is bounded for all $\theta \in \Theta$. This assumption is certainly restrictive, but it is useful to solve a variational problem considered below. This may be relaxed by a sequential truncation approach used by Kitamura (2001); this will be investigated in a later version of the paper. Suppose the identifying assumption made by the econometrician is that

$$\int g(Z,\theta_0)d\mu_{Z|X} = 0, \theta_0 \in \Theta \quad (a.s., \mu_X) \tag{2.7}$$

For an arbitrary value $\bar{\theta}$ in Θ, define

$$\mathcal{P}^{\bar{\theta}}_{Z|X} = \{(P_{Z|X=x})_{x\in\mathcal{X}} : \int g(z,\bar{\theta})dP_{Z|X} = 0 \quad (a.s., \mu_X)\},$$

then let

$$\mathcal{P}_{Z|X} = \cup_{\bar{\theta}\in\Theta}\mathcal{P}^{\bar{\theta}}_{Z|X}. \tag{2.8}$$

By construction $\mathcal{P}_{Z|X}$ is the set of all conditional probability measures that are compatible with the restriction (2.7).

Solving (2.1) with $\mathcal{P}_{Z|X}$ in (2.8), which is semiparametric, is a variational problem. To solve it, it is easier to break up the problem into two parts. The first part consists of the following problem defined for arbitrarily chosen $\bar{x} \in \mathcal{X}$ and $\bar{\theta} \in \Theta$:

$$I(\bar{x},\bar{\theta}) := \inf_{P^{\bar{\theta}}_{Z|\bar{x}}} \int \log(d\mu_{Z|\bar{x}}/dP^{\bar{\theta}}_{Z|\bar{x}})d\mu_{Z|\bar{x}}(z) \tag{2.9}$$

$$\text{subject to} \int g(z,\bar{\theta})dP^{\bar{\theta}}_{Z|\bar{x}}(z) = 0.$$

The second part minimizes the function

$$I(\bar{\theta}) := \int I(x,\bar{\theta})dP_X(x). \tag{2.10}$$

The quantity

$$\inf_{\bar{\theta}\in\Theta} I(\bar{\theta}) \tag{2.11}$$

gives the constrained minimum relative entropy for (2.8).

(2.9) is known as a partially finite convex programming problem (see Borwein and Lewis (1993)). Instead of solving this constrained minimization problem, we can solve the following unconstrained dual problem:

$$\text{maximize} \int \log(1+\lambda' g(z,\theta))d\mu_{Z|\bar{x}} \tag{2.12}$$

$$\text{subject to } \lambda \in \mathbb{R}^q.$$

Let $\lambda(\bar{x}, \bar{\theta})$ be the minimizer of (2.12). Then the solution to (2.9) is given by a Radon-Nikodym derivative

$$\frac{dP^{\bar{\theta}}_{Z|\bar{x}}}{d\mu_{Z|\bar{x}}} = \frac{1}{1 + \lambda(\bar{x}, \bar{\theta})' g(z, \bar{\theta})}. \tag{2.13}$$

See Borwein and Lewis (1993) for details. Using (2.9) and (2.13), write

$$\begin{aligned} I(\bar{x}, \bar{\theta}) &= \int \log(1 + \lambda(\bar{x}, \bar{\theta})' g(z, \bar{\theta})) d\mu_{Z|\bar{x}}(z) \\ &= \max_{\lambda \in \mathbb{R}^q} \int \log(1 + \lambda' g(z, \bar{\theta})) d\mu_{Z|\bar{x}}(z). \end{aligned} \tag{2.14}$$

We can rewrite the second stage objective function $I : \Theta \to \mathbb{R}_+$ in (2.10) using (2.14):

$$I(\bar{\theta}) = \int \max_{\lambda \in \mathbb{R}^q} \int \log(1 + \lambda' g(z, \bar{\theta})) d\mu_{Z|x}(z) d\mu_X(x). \tag{2.15}$$

Suppose θ^* in Θ minimizes the function $I(\cdot)$. Call θ^* a pseudo-true value in the sense used by White (1982), Vuong (1989) and others. Also define a conditional measure $(P^*_{Z|x})_{x \in \mathcal{X}}$ by (2.13) evaluated at θ^*:

$$\frac{dP^*_{Z|x}}{d\mu_{Z|x}} = \frac{1}{1 + \lambda(x, \theta^*)' g(z, \theta^*)}, x \in \mathcal{X}.$$

Analogous to θ^*, call this family of measures a pseudo-true conditional measure. This solves the original problem (2.1) for the set (2.8). It provides the best approximation of the true conditional measure $(\mu_{Z|x})_{x \in \mathcal{X}}$ in the candidate set $\mathcal{P}_{Z|X}$, in terms of the expected log likelihood ratio (2.2). The above result is a modest extension of entropic projection theorems for unconditional measures as considered by Csiszár (1995) to conditional measures.

The minimum value of the expected log-likelihood ratio I^*, attained at $(P^*_{Z|x})_{x \in \mathcal{X}}$, is

$$\begin{aligned} I^* &= I(\theta^*) \\ &= -\int \log \frac{dP^*_{Z|x}}{d\mu_{Z|x}} d\mu_{X,Z} \\ &= \int \max_{\lambda \in \mathbb{R}^q} \int \log(1 + \lambda' g(z, \theta^*)) d\mu_{Z|x}(z) d\mu_X(x). \end{aligned} \tag{2.16}$$

I^* (or its exponential transformation $\phi^* = e^{-I^*}$) measures the goodness of fit of the conditional moment restriction model (2.7). Section 4 develops a consistent estimator for I^* in (2.16).

3. Estimation

The asymptotic theory for the (conditional) MLE for the parametric model (2.5) is well-known. A treatment of the MLE for a correct model can be found in standard textbooks. White (1982) discusses asymptotic properties of the MLE in a misspecified fully parametric model. Vuong (1989) notes that this property extends to a fully parametric conditional model in a straightforward manner. This section develops a semiparametric version of this theory, using the conditional moment restriction model (2.8). It demonstrates that "conditional empirical likelihood" procedure (Kitamura, Tripathi, and Ahn (2000)) has many properties that are analogous to the conventional parametric ML procedure. In large samples, the maximum conditional empirical likelihood estimator finds a conditional measure within $\mathcal{P}_{Z|X}$ such that it maximizes the expected log likelihood, or equivalently, minimizes the conditional relative entropy. This is related, but different from the results on misspecified least square methods obtained by White (1980, 1981), as discussed in Sections 3 and 6.

Consider the conditional moment model defined by (2.8). Suppose $\mathcal{P}_{Z|X}$ does not include the true measure $(P_{Z|x})_{x\in\mathcal{X}}$ as its member. Then the existing theory for the estimation of conditional moment restriction models, as presented by Newey (1990, 1993) and Kitamura, Tripathi, and Ahn (2000) no longer holds. On the other hand, the estimator proposed by Kitamura, Tripathi, and Ahn (2000) has a property that it asymptotically provides a measure that minimizes the conditional relative entropy between the true conditional measure,[1] as we demonstrate below.

The following method operationalizes the theoretical result for the semiparametric model (2.8) obtained in Section 2, by constructing appropriate sample counterparts. That is, the method first calculates a natural sample counterpart of the objective function (2.15), then maximizes the sample objective function with respect to the unknown parameter $\theta \in \Theta$.

The outer integral of the double integral (2.15) can be estimated simply by the sample average over realizations of x's. The inner integral is a conditional expected value, so an appropriate nonparametric regression technique can be used to replace it. Throughout this paper we use the standard Nadaraya-Watson kernel regression method. Let $k(\cdot)$ be a univariate kernel and $K(x) = \Pi_{j=1}^{s} k(x^j)$ be the s-dimensional product kernel. Using bandwidth b that shrinks with n, Nadaraya-Watson

---

[1]Note that efficient IV methods as suggested by Newey does not have this property for two reasons. First, the criterion it uses is, loosely put, akin to the L^2 metric (in the sense used in the maximum entropy literature). This fact is important, especially in developing a model comparison test against a parametric model, since the parametric ML procedure minimizes relative conditional entropy, not the L^2 metric. Second, it is a "plug-in" procedure, which makes it difficult to develop a theory for misspecified models.

weights are defined as $w_{ij} = \frac{K(\frac{x_i - x_j}{b})}{\sum_{j=1}^s K(\frac{x_i - x_j}{b})}$. Note that the denominator of w_{ij} divided by nb^s is the Nadaraya-Watson density estimator $\hat{h}(x_i) := \frac{1}{nb^s}\sum_{j=1}^n K(\frac{x_i - x_j}{b})$. Finally, define a trimming function $\tau_i = 1\{\hat{h}(x_i) \geq b^\delta\}$. Some condition on the constant $\delta > 0$ will be imposed later. Modifying Nadaraya-Watson estimates by τ_i avoids the well known theoretical problem associated with the denominator $\hat{h}$ at points where the density of h_X is too low. This is a standard method in the literature and facilitates our proofs.

We are now ready to compute a sample analogue of (2.15). Define

$$\hat{I}(\theta) = n^{-1} \sum_{i=1}^n \max_{\lambda \in \mathbb{R}^q} \tau_i \sum_{j=1}^n w_{ij} \log(1 + \lambda' g(z_j, \theta)). \tag{3.1}$$

Let $\hat{\theta}$ denote the value of $\theta \in \Theta$ that minimizes $\hat{I}(\cdot)$. Note that this is exactly the form of the estimator developed by Kitamura, Tripathi, and Ahn (2000) and Zhang and Gijbels (2001) (see Owen (2001), Section 9.10). See also LeBlanc and Crowley (1995) for a closely related proposal.

Kitamura, Tripathi, and Ahn (2000) call the technique "smoothed empirical likelihood" and Zhang and Gijbels (2001) call it "sieve empirical likelihood." In this paper we use the terminology "conditional empirical likelihood" to emphasize the conditional nature of the method.

Put loosely, (3.1) corresponds to the negative of the log likelihood of the conditional moment restriction model at θ. The conditional empirical likelihood generalizes the empirical likelihood method developed by Owen (1988, 1990, 1991); see Owen (2001) for a superb introduction to the subject. If we let the bandwidth go to $+\infty$, it coincides with the conventional empirical likelihood. Since $\hat{\theta}$ is a sample counterpart of θ^*, the "analogue principle" (Manski (1988)) suggests that $\hat{\theta}$ approaches to θ^* as the sample size grows. The following assumption lists sufficient conditions for this result. Throughout the paper, assume that x_i is continuously distributed with density $h_X(x)$ defined with respect to the Lebesgue measure.

Assumption 3.1. *(i) $\Theta \in \mathbb{R}^p$ is compact;*

(ii) $\int \|x\|^{1+\varrho} d\mu_X < \infty$ for some $\varrho > 0$;

(iii) h_X is bounded above for all $\mathcal{X}$ and the first and second derivatives of $\|h_X(x)\|$ are uniformly bounded over $\mathcal{X}$;

(iv) $g(z, \cdot)$ is continuous at each θ in Θ with probability one;

(v) $\int \sup_{\theta \in \Theta} \|\log(1 + \lambda(x, \theta)' g(z, \theta))\| d\mu_{X,Z} < \infty$;

(vi) $\|\nabla_{xx} \int \log(1 + \lambda(x, \theta)' g(z, \theta)) h_X(x) d\mu_{Z|x}\|$ is uniformly bounded over $\Theta \times \mathcal{X}$;

(*vii*) $k(\cdot)$ *is zero outside* $[-1,1]$*, continuously differentiable on* $[-1,1]$*, and satisfies* $\int k(u)du = 1$*;*

(*viii*) *For some* $\delta > 0$, $n^{1-\delta} b^{2(s+\delta)} \to \infty$*, and* $b \to 0$*;*

(*ix*) $I(\cdot)$ *has a unique minimum at* θ^* *in* Θ.

The following theorem shows the consistency of $\hat{\theta}$ with respect to the pseudo-true value θ^*.

Theorem 3.1. *Suppose Assumption 3.1 holds. Then* $\hat{\theta}$ *converges to* θ^* *in probability.*

Remark 3.1. The above procedure provides a consistent estimator for the pseudo-true conditional measure $P^*_{Z|x}$. Let A denote a Borel set in the Borel field generated by z. Note that

$$P^*_{Z|x}(A) = \int_A \frac{d\mu_{Z|x}}{1 + \lambda(x, \theta^*)' g(z, \theta^*)}.$$

A natural estimator of $P^*_{Z|x}(A)$ can be constructed using the conditional empirical likelihood outlined above. Let $\hat{\lambda} : \mathcal{X} \times \Theta \to \mathbb{R}^q$ be the solution to the "inner" optimization problem in (3.1):

$$\hat{\lambda}(x, \theta) \in \mathbb{R}^q \text{ maximizes } \sum_{j=1}^{n} \frac{K(\frac{x - x_j}{h})}{\sum_{j=k}^{n} K(\frac{x - x_k}{h})} \log(1 + \lambda' g(z_j, \theta)).$$

Our estimator for $P^*_{Z|x}(A)$ is

$$\hat{P}_{Z|x}(A) := \sum_j \frac{K(\frac{x - x_j}{h})}{\sum_{k=1}^{n} K(\frac{x - x_k}{h})} \frac{I\{z_j \in A\}}{1 + \hat{\lambda}(x, \hat{\theta})' g(z_j, \hat{\theta})}. \tag{3.2}$$

If the conditional moment is correctly specified, the extra weighting factor $1/(1 + \lambda(x, \hat{\theta})' g(z_j, \hat{\theta}))$ provides an efficiency gain, thereby yielding an semiparametrically efficient estimator of the true conditional measure of z_j given x; see Brown and Newey (1998) and Kitamura, Tripathi, and Ahn (2000) for details. Here, the role of the factor is to adjust the conventional estimator $\tilde{P}_{Z|x}(A) := \sum_j \frac{K(\frac{x - x_j}{h})}{\sum_{k=1}^{n} K(\frac{x - x_k}{h})} I\{z_j \in A\}$ so that it converges to the pseudo-true measure $P^*_{Z|x}(A)$.

Remark 3.2. As noted before, this result parallels the theory of misspecified MLE by White (1982). There are some other studies related to the above result. Vuong (1983) provides a formal treatment of the behavior of misspecified conditional MLE in a general parametric setup. Kitamura (1998) develops an asymptotic theory of misspecified unconditional moment restriction models, using so-called exponential tilting as used by Kitamura and Stutzer (1997) and Imbens, Spady, and Johnson (1998). Note that the exponential tilting method has a close connection with empirical likelihood (see Newey and Smith (2000), for example, for this and other types of nonparametric likelihood). See also Hong, Preston, and Shum (2001) for an updated discussion on misspecified unconditional moment restriction models.

Remark 3.3. The conditional moment restriction framework in Section 1 includes (nonlinear) regressions. White (1980, 1981) studies the behavior of the least square estimator for a misspecified nonlinear regression model. The current paper addresses a related, but different question. To fix ideas, suppose the econometrician observe IID samples $\{x_i, y_i\}_{i=1}^n$ drawn from $\mu_{X,Y}$. Let $\mu_{Y|X}$ be the true conditional measure of y_i given x_i. The econometrician uses a parametric specification $m(x,\theta)$, $\theta \in \Theta$, motivated by economic theory, for the conditional expectation $\int y d\mu_{Y|X}$. Suppose $m(x,\theta), \theta \in \Theta$ is misspecified. White shows (among other things) that the misspecified *least squares estimator* converges to the minimizer of MSE, i.e., $\arg\min_\theta \int (y - m(x,\theta))^2 d\mu_{X,Y} (= \theta^{**}$, say). That is, his θ^{**} is defined by imposing the moment condition $\int (y - m(x,\theta^{**}))\nabla_\theta m(x,\theta^{**}) d\mu_{X,Y} = 0$ implied by the least square method. Our method is different: it imposes the conditional moment condition (that the regression function of Y given X takes the form $m(x,\cdot)$) on our estimator. Naturally, our θ^* (that minimizes $I(\cdot)$ in(2.15)) generally differs from θ^{**}.

The next task is to derive the asymptotic distribution of $\hat{\theta}$. This is useful in deriving model comparison tests discussed later, though the asymptotic distribution is interesting in its own right.

Define:

$$D(x,\theta) := \nabla_\theta \int \frac{g(z,\theta)'}{1 + \lambda(x,\theta)' g(z,\theta)} d\mu_{Z|x},$$

$$s(x,z,\theta) := \frac{g(z,\theta)}{1 + \lambda(x,\theta)' g(z,\theta)},$$

$$V(x,\theta) = \int s(x,z,\theta) s(x,z,\theta)' d\mu_{Z|x},$$

and

$$\Omega := \left(\int D(x,\theta^*)' V^{-1}(x,\theta^*) D(x,\theta^*) d\mu_X \right)^{-1}.$$

Let A^{ij} denote the ijth element of a matrix A. We assume some extra conditions:

Assumption 3.2. *There exists a neighborhood N of θ^* such that*

(*i*) $N \subset \Theta$;

(*ii*) $\left(\int \sup_{|\theta-\theta^*| \le \delta} \|s(x,z,\theta) - s(x,z,\theta^*)\|^p d\mu_{X,Z} \right)^{1/p} \le C\psi^\delta$ *for some* $p \ge 2$, $C > 0$, $\psi > 0$, *all* $\delta >$ *in a neighborhood of 0.*

(*iii*) $V(x,\theta)$ *is nonsingular, bounded away from zero and from above, and continuous with probability one at each* $\theta \in N$;

(*iv*) $D(x,\theta)$ *is continuous and* $\sup_{\theta \in N} \|D(x,\theta)\| < \infty$ *with probability one at each* $\theta \in N$.

(*v*) $\|\nabla_{xx} D^{(ij)}(x,\theta) h_X(x)\|$ *is uniformly bounded over* $N \times \mathcal{X}$ *for all* (i,j);

(vi) $\|\nabla_{xx}V^{(ij)}(x,\theta)h_X(x)\|$ *is uniformly bounded over* $N \times \mathcal{X}$ *for all* (i,j)*;*

We will show the following asymptotic normality result in Section 8:

Theorem 3.2. *If Assumptions 3.1 and 3.2 hold, then* $\sqrt{n}(\hat{\theta}-\theta^*) \rightsquigarrow N(0,\Omega)$.

Remark 3.4. This result generalizes Theorem 3.2 by Kitamura, Tripathi, and Ahn (2000), who consider correctly specified conditional moment restriction models. They show that for the true parameter θ_0, $V(x) := \int g(z,\theta_0)g(z,\theta_0)' d\mu_{Z|x}$, $D(x) := \int \nabla_\theta g(z,\theta_0)' d\mu_{Z|x}$, and $\Omega_0 := (\int D(x)'V^{-1}(x)D(x)d\mu_x)^{-1}$,

$$\sqrt{n}(\hat{\theta}-\theta) \rightsquigarrow N(0,\Omega_0).$$

Note that when the model is correct, a pseudo-true parameter and the true parameter coincide ($\theta^* = \theta_0$), and the function $\lambda(x,\theta_0)$ takes the constant value zero over $x \in \mathcal{X}$, since it is the Lagrange multiplier for the conditional moment condition. It is then immediate to see that the asymptotic distribution result for $\hat{\theta}$ in Kitamura, Tripathi, and Ahn (2000) is indeed a special case of Theorem 3.2.

4. A Likelihood-Based Measure of Model Fit

The results in Section 2 suggest that we may use the negative expected log-likelihood ratio (2.3) to assess model fit. The purpose of this section is to operationalize this idea and formulate a goodness-of-fit measure.

First consider the case where $\mathcal{P}_{Z|X}$ is defined parametrically by (2.5). That is, $\mathcal{P}_{Z|X} := \{(P^{\gamma}_{Z|x})_{x\in\mathcal{X}} : \gamma \in \Gamma\}$. Suppose $\mu_{X,Z}$ is absolutely continuous with respect to Lebesgue measure and let $f(x,z)$ be the joint density for X and Z. The conditional MLE $\hat{\gamma}$ converges to a pseudo-true parameter γ^* (Vuong (1983)). Let $f(z|x) = \frac{d\mu_{Z|X}}{dz}$ and $f^{\gamma}(z|x) = \frac{dP^{\gamma}_{Z|X}}{dz}$, then we have

$$\begin{aligned} I^* &= \int\int \log \frac{d\mu_{Z|X}}{dP^{\gamma^*}_{Z|X}} d\mu_{Z|X} d\mu_X \\ &= \int \log\left(\frac{f(z|x)}{f^{\gamma^*}(z|x)}\right) d\mu_{X,Z}. \end{aligned}$$

In practice, of course, $I^* = \int \log f(z|x) d\mu_{X,Z} - \int \log f^{\gamma^*}(z|x) d\mu_{X,Z}$ needs to be estimated. The second integral $\int \log f^{\gamma^*}(z|x) d\mu_{X,Z}$ is consistently estimated by $n^{-1}\sum_{i=1}^n \log f^{\hat{\gamma}}(z_i|x_i)$, as noted by Vuong (1989). Though $f(z|x)$ in the first integral is unknown, this can be estimated using a kernel method. For example, using $K_{X,Z}(x,z) := \Pi_{j=1}^s k(x^j)\Pi_{h=1}^d k(z^h)$, define a kernel density estimator $\hat{g}(x,z) = \frac{1}{nb^{s+d}}\sum_{i=1}^n K_{X,Z}(\frac{x-x_i}{b},\frac{z-z_i}{b})$. Let $\hat{h}$ denote the kernel estimator for the marginal density of

x_i as defined in Section 3. A simple kernel estimator for $f(z|x)$ is given by $\hat{f}(z|x) := \hat{f}(x,z)/\hat{h}(x)$, leading to the following estimator for I^*

$$\hat{I}^* := n^{-1} \sum_{i=1}^{n} \tau_i \left[\log \hat{f}(z_i|x_i) - \log f^{\hat{\gamma}}(z_i|x_i) \right]. \tag{4.1}$$

Note that the term $n^{-1}\sum_{i=1}^{n} \tau_i \log \hat{f}(z_i|x_i)$ in the above definition extends the nonparametric entropy estimator proposed by Ahmad and Lin (1976) (which, in turn, used by Chen, Hong, and Shum (2001)) to the conditional entropy functional $H(f(\cdot|\cdot)) = -\int\int f(z|x)\log f(z|x) dz \mu_X(dx)$; see Beirlant, Dudewicz, Gyölfi, and Meulen (1997) for a review of the vast literature on nonparametric entropy estimation and its use.

The above argument shows that we can use $\hat{I}^*$ in (4.1) or $\hat{\phi}^* := \exp(\hat{I}^*)$ as a model fit measure of the parametric model (2.5). If the researcher wishes to compare model fit of two parametric models, however, our procedure simplifies. The likelihood ratio statistic between the two models is then the difference between the corresponding $\hat{I}^*$'s from the models. This is exactly the statistic proposed by Vuong (1989). Obviously this statistic does not require the nonparametric estimation of $\hat{f}(z_i|x_i)$, as the first term in (4.1) is differenced out. This cancellation, however, does not happen if a parametric model is compared with a conditional moment restriction model. See Section 6.4 for details on this point.

The estimation of I^* for the semiparametric model (2.8) is straightforward. The expression for I^* in (2.16) suggests the following statistic:

$$\begin{aligned} \hat{I}^* &= \hat{I}(\hat{\theta}) \\ &= n^{-1} \sum_{i=1}^{n} \max_{\lambda \in \mathbb{R}^q} \tau_i \sum_{j=1}^{n} w_{ij} \log(1 + \lambda' g(z_j, \hat{\theta})). \end{aligned} \tag{4.2}$$

This is our model fit measure for the conditional moment restriction model (2.8).

A notable feature of $\hat{I}^*$ in (4.1) or (4.2) is that it provides a coherent model fit measure for a wide variety of models. Both (4.1) and (4.2) can be regarded as (negative) log-likelihood values. It enables the researcher to evaluate model fit in terms of likelihood, even if the models is semiparametrically defined.

It may be convenient for practical use to transform $\hat{I}^*$ appropriately. For example, define $\hat{\phi}^* = \exp(-\hat{I}^*)$ as in the expression (2.4). It belongs to the interval $(0, 1)$ and assesses empirical model fit.

5. Quantile Regression

The methodology described in Section 3 remains valid for a conditional quantile restriction model. This section considers quantile regression models as developed by Koenker and Bassett (1978). Quantile regression is now widely used in applied research; see, for instance, Chamberlain (1994), Buchinsky (1994), and Johnson, Kitamura, and Neal (2000). It should be noted that there are some earlier papers that discuss misspecification issues for quantile regressions. For example, Chamberlain (1994) discusses the possibility of misspecification and derives the limiting distribution of a minimum chi-square estimator under such a scenario. Also, a recent paper by Kim and White (2002) develops an asymptotic theory for the LAD-type estimator for quantile regression. Our concern here is different, however, as discussed in Remark 3.3. Our method imposes a conditional quantile restriction on our estimator.

To fix ideas, consider a linear model for the q-th quantile of a random variable y_i conditional on "covariates" x_i, denoted by $Quant_q(y_i|x_i)$. That is,

$$Quant_q(y_i|x_i) = x_i'\theta. \tag{5.1}$$

It is well-known that this model can be put in the conditional moment restriction framework (2.7), with $g(z, \theta) = q - 1 + 1\{y - x'\theta > 0\}$. As noted by LeBlanc and Crowley (1995), there is no need for carrying out numerical optimization over λ in this model, since an analytical solution of the following form is available:

$$\hat{\lambda}(x_i, \theta) = \frac{1 - q - W_i(\theta)}{q(q-1)}, \quad W_i(\theta) = \sum_j w_{ij} 1\{y_j - x_j'\theta > 0\}.$$

Plugging this into (3.1) yields a simple form

$$\hat{I}(\theta) = \frac{1}{n}\sum_i^n \left[W_i(\theta) \log \frac{W_i(\theta)}{1-q} + (1 - W_i(\theta)) \log \left(\frac{1 - W_i(\theta)}{q} \right) \right]. \tag{5.2}$$

$\hat{\theta}$ that minimizes (5.2) is our (conditional empirical likelihood) estimator for the quantile regression model. A care should be taken in computing $\hat{\theta}$, as (5.2) is a discontinuous function. This case nevertheless has the above advantage of not requiring optimizing over λ. Preliminary experiments suggest that obtaining $\hat{\theta}$ for a quantile regression model is indeed computationally inexpensive.

If the model is correctly specified, $\hat{\theta}$ is a consistent and asymptotically efficient estimator of the true quantile regression parameter. This is interesting in its own right, as it is a one-step estimator (as the conventional LAD-type estimator) that attains the semiparametric efficiency (unlike conventional estimators). The focus of the current paper, however, is the misspecified case.

Computing the model fit measure $\hat{I}^*$ for the quantile regression model (5.1) is now straightforward:

$$\hat{I}^* = \frac{1}{n}\sum_{i}^{n}\left[W_i(\hat{\theta})\log\frac{W_i(\hat{\theta})}{1-q} + (1-W_i(\hat{\theta}))\log\left(\frac{1-W_i(\hat{\theta})}{q}\right)\right].$$

This formula shows how to calculate the negative log-likelihood values for evaluating various quantile regression models. It is a potentially useful tool in situations where conventional methods fail to work. Consider, for instance, a situation where the researcher wishes to compare a mean regression specification and a median regression specification. Here conventional fitness criteria cannot be used for model comparison. Comparing a least square criterion and a least absolute deviation criterion does not make sense. The new likelihood-based measure can be applied to both models, providing a coherent measure of model performance. Alternatively, suppose the researcher considers linear quantile regression models at two different quantiles. Quantile regression is usually implemented by minimizing the sample average of "check functions." The criterion functions based on two different check functions are incomparable. Using (5.2) overcomes these problems.

6. Inference

6.1. **Likelihood Ratio Tests.** Vuong (1989) points out that the log-likelihood ratio for a parametric conditional model, divided by the number of samples, consistently estimates the relative entropy of the true conditional measure with respect to the best approximating conditional measure. He proposes to use the estimate to carry out a model comparison test. The results in subsequent sections suggest that this can be done for the conditional moment restrictions model (2.8) as well.

6.2. **Testing between Conditional Moment Restriction Models.** In what follows we develop a likelihood-ratio test for conditional moment restriction models. Consider two functions $g^1 : \mathcal{Z}\times\Theta^1 \to \mathbb{R}^{q^1}$ and $g^2 : \mathcal{Z}\times\Theta^2 \to \mathbb{R}^{q^2}$. Replace (g,Θ) in (2.8) by (g^1,Θ^1) and (g^2,Θ^2) to define $\mathcal{P}^1_{Z|X}$ and $\mathcal{P}^2_{Z|X}$. If either of $\mathcal{P}^1_{Z|X}$ and $\mathcal{P}^2_{Z|X}$ is a subset of the other, the two models are said to be nested. If the two models are nested and at least one of the models include the true conditional measure $(\mu_{Z|x})_{x\in\mathcal{X}}$, the asymptotic theory developed by Kitamura, Tripathi, and Ahn (2000) (or some other inference scheme with or without conditional empirical likelihood) applies. This section focuses on the case where the conditional models $\mathcal{P}^1_{Z|X}$ and $\mathcal{P}^2_{Z|X}$ are possibly non-nested and misspecified.

Suppose $\mathcal{P}^1_{Z|X}$ and $\mathcal{P}^2_{Z|X}$ are indeed non-nested and $\mathcal{P}^1_{Z|X} \not\ni (\mu_{Z|x})_{x\in\mathcal{X}}$ and $\mathcal{P}^2_{Z|X} \not\ni (\mu_{Z|x})_{x\in\mathcal{X}}$. Solve (2.1) using $\mathcal{P}^1_{Z|X}$ and $\mathcal{P}^2_{Z|X}$ as the constraint set for $(P_{Z|x})_{x\in\mathcal{X}}$, following the steps outlined in (2.9)-(2.15). Define $(\lambda^1(\cdot,\cdot), I^1(\cdot), \theta^{*1})$ and $(\lambda^2(\cdot,\cdot), I^2(\cdot), \theta^{*2})$ for $\mathcal{P}^1_{Z|X}$ and $\mathcal{P}^2_{Z|X}$ as we defined

$(\lambda(\cdot,\cdot), I(\cdot), \theta^*)$ for $\mathcal{P}_{Z|X}$ in Section 2. Note that $\lambda^1(\cdot,\cdot)$ and $\lambda^2(\cdot,\cdot)$ are $\mathbb{R}^{q^1}$-valued and $\mathbb{R}^{q^2}$-valued, respectively. The minimum expected log-likelihood values are given by $I^{*1} := I^1(\theta^{*1})$ and $I^{*2} := I^2(\theta^{*2})$.

If $I^{*1} < (>) I^{*2}$, that means that Model $\mathcal{P}^1_{Z|X}$ (Model $\mathcal{P}^2_{Z|X}$) is closer to the true conditional measure $(\mu_{Z|x})_{x\in\mathcal{X}}$ than Model $\mathcal{P}^2_{Z|X}$ (Model $\mathcal{P}^1_{Z|X}$). We then say that Model $\mathcal{P}^1_{Z|X}$ is better (worse) than Model $\mathcal{P}^2_{Z|X}$. If $I^{*1} = I^{*2}$, then Model $\mathcal{P}^1_{Z|X}$ and Model $\mathcal{P}^2_{Z|X}$ are equally good.

In practice, of course, I^{*1} and I^{*2} are not observable. Section 3, however, shows how to estimate them. More specifically, construct the conditional empirical likelihood functions for $\mathcal{P}^1_{Z|X}$ and $\mathcal{P}^2_{Z|X}$, by replacing g in (3.1) with g^1 and g^2. Define $(\hat{\lambda}^1(\cdot,\cdot), \hat{I}^1(\cdot), \hat{\theta}^1)$ and $(\hat{\lambda}^2(\cdot,\cdot), \hat{I}^2(\cdot), \hat{\theta}^2)$ for $\mathcal{P}^1_{Z|X}$ and $\mathcal{P}^2_{Z|X}$ as we defined $(\hat{\lambda}(\cdot,\cdot), \hat{I}(\cdot), \hat{\theta})$ for $\mathcal{P}_{Z|X}$ in Sections 3 and 4. $\hat{I}^{*1} := \hat{I}^1(\hat{\theta}^1)$ and $\hat{I}^{*2} := \hat{I}^2(\hat{\theta}^2)$ are our estimates for I^{*1} and I^{*2}. If $\hat{I}^{*1} < (>) \hat{I}^{*2}$, it gives some empirical evidence in favor of Model $\mathcal{P}^1_{Z|X}$ ($\mathcal{P}^2_{Z|X}$). On the other hand, it is necessary to take account of sampling variations in interpreting these kinds of inequalities. (See Vuong (1989) for the same issue in fully parametric models.) The following addresses this issue by setting up an appropriate null hypothesis and testing it.

The null hypothesis we wish to test is that the two conditional moment models are equally good, or equivalently,

$$\textbf{(H)} \qquad I^{*1} = I^{*2}.$$

To test (**H**), calculate a statistic $D = \sqrt{n}(\hat{I}^{*1} - \hat{I}^{*2})$. Our discussion in Section 3 suggests that $\hat{I}^{*1}$ and $\hat{I}^{*2}$ can be regarded as the log likelihood values (multiplied by $-\frac{1}{n}$) of Model $\mathcal{P}^1_{Z|X}$ and Model $\mathcal{P}^2_{Z|X}$. Therefore $n(\hat{I}^{*1} - \hat{I}^{*2})$ is the log likelihood ratio of the two models. The next task is to evaluate the sampling variation in the normalized likelihood ratio statistic D under (**H**). Let

$$r_1(x,z) = \log(1 + \lambda^1(x,\theta^{*1})' g^1(z,\theta^{*1})) \tag{6.1}$$

for Model $\mathcal{P}^1_{Z|X}$, and similarly define $r_2(x,z)$ for $\mathcal{P}^2_{Z|X}$. Define

$$\sigma^2 = \int (r_1(x,z) - r_2(x,z))^2 d\mu_{X,Z}.$$

The next theorem gives an asymptotic approximation of the distribution of D.

Theorem 6.1. *Suppose Assumptions 3.1 and 3.2 hold. Then $D \rightsquigarrow N(0,\sigma^2)$ under* (**H**).

This result is useful in constructing a studentized test statistic. For studentization, we need a consistent estimator for σ^2. To this end, let

$$\hat{r_{1i}} = \tau_i \log(1 + \hat{\lambda}^1(x_i,\hat{\theta}^1)' g^1(z_i,\hat{\theta}^1)), i = 1, ..., n \tag{6.2}$$

for Model $\mathcal{P}^1_{Z|X}$, and similarly define $\hat{r_{2i}}, i = 1, ..., n$ for $\mathcal{P}^2_{Z|X}$. Then

$$\hat{\sigma}^2 := n^{-1} \sum_{i=1}^{n} (\hat{r_{1i}} - \hat{r_{2i}})^2$$

works for our purpose. Now let

$$d = D/\hat{\sigma}. \tag{6.3}$$

This is our model comparison statistic for conditional moment restrictions. If $\sigma^2 > 0$,

$$d \rightsquigarrow N(0, 1)$$

under (**H**).

Remark 6.1. The studentized likelihood ratio statistic d is asymptotically distributed according to the standard normal distribution under the null, which is convenient for practical use. It extends a nonparametric likelihood ratio for unconditional moment restriction models proposed by Kitamura (1998), whose methodology is further extended and applied to financial models by Christoffersen, Hahn, and Inoue (2001). Kitamura (1998) uses an exponential tilting type method, which is a nonparametric likelihood method closely related to empirical likelihood. Hong, Preston, and Shum (2001) also propose a model comparison test as in Kitamura (1998), though they use empirical likelihood.

While we formulate a Vuong-type test here, other procedures for non-nested testing could be used. An example is the Cox procedure, originally suggested by Cox (1961). The Cox test involves integration of a log likelihood function under the alternative hypothesis. In our case, this could be achieved by using the conditional CDF of the form (3.2), obtained under the alternative model. This means replacing the kernel weight w_{ij} in (3.1) with $\frac{w_{ij}}{1+\hat{\lambda}(x_i,\hat{\theta})'g(z_i,\hat{\theta})}$ calculated under the alternative model. Rigorous analysis of such a procedure is beyond the scope the paper. For unconditional moment restriction models, Smith (1997) discusses such a strategy using unconditional empirical likelihood.

6.3. Degeneracy.

The Problem

To use the normal approximation for the d-statistic (6.3), we need to have the condition $\sigma^2 > 0$ satisfied. If σ^2 is zero, the asymptotic law of D degenerates at the origin. Vuong (1989) recognizes this phenomenon in his investigation of parametric model comparison tests. In the case where this degeneracy can be a problem, he suggests using a pretest for the hypothesis that $\sigma^2 = 0$. He also finds the asymptotic distribution of the pretest. The degeneracy is potentially a problem in Cox tests too, though it appears that many investigators simply 'assume away' the problem in practice.

The degeneracy occurs if the Radon-Nikodym derivatives of the two pseudo-true conditional probability measures with respect to the true conditional measure

$$\frac{dP^{*1}_{Z|x}}{d\mu_{Z|x}} = \frac{1}{(1+\lambda^1(x,\theta^{*1})'g^1(z,\theta^{*1}))} \text{ and } \frac{dP^{*2}_{Z|x}}{d\mu_{Z|x}} = \frac{1}{(1+\lambda^1(x,\theta^{*2})'g^2(z,\theta^{*2}))}$$

happen to coincide (see Vuong (1989) and Kitamura (1998) for related discussions). Kitamura (1998) develops a nonparametric likelihood version of Vuong's pretest for competing unconditional moment restriction models, using exponential tilting. The test has power against parametric (i.e., $n^{-1/2}$) local alternatives. Chen, Hong, and Shum (2001) also propose two pretesting procedures for degeneracy: one is a kernel-based test, the other a Kolmogorov-Smirnov type test. Both of them are very interesting, though extending them to conditional moment models may not be straightforward. They also note that the first test is consistent at a nonparametric rate and that the implementation of the second test could be difficult in practice even with unconditional models.

The pretests by Kitamura (1998) and Chen, Hong, and Shum (2001) are valid for unconditional models. Extending them to conditional moment models is, however, not straightforward either theoretically or practically. Take the Vuong-type degeneracy test by Kitamura (1998), for example. Kitamura (1998) notes an interesting fact about the large sample behavior of the Vuong-type degeneracy test when applied to unconditional moment conditions. The number of the degrees of freedom of the asymptotic (weighted chi-square) distribution includes the number of moment conditions from the models under comparison, in addition to the number of parameters. Since the effective number of moment conditions in a conditional moment model is (asymptotically) infinity, the nature of asymptotics would be quite different. While such a situation is not uncommon in nonparametric specification testing (see Tripathi and Kitamura (2001), for example), global misspecification as considered here complicates asymptotics considerably.

In some cases, one can rule out the possibility of degeneracy *a priori*. If $\mathcal{P}^1_{Z|X}$ and $\mathcal{P}^2_{Z|X}$ are strictly non-nested, i.e. $\mathcal{P}^1_{Z|X} \cap \mathcal{P}^2_{Z|X}$ is an empty set, $\sigma^2 > 0$; see related discussions in Vuong (1989), Kitamura (1998) and Chen, Hong, and Shum (2001). Also, as noted above, the possibility of degeneracy is ignored in many applications of the Cox test, though it is certainly a theoretical possibility (see Kent (1986) for degeneracy in the Cox/Mizon-Richard tests).

Sample Splitting

In applications where the possibility of degeneracy of the normalized likelihood ratio test d-test is an issue, one solution is to use sample splitting, as used by Yatchew (1992) in a different context of nonparametric specification testing. Yatchew's test is based on the difference between sums of squares of residuals with and without a restriction. He recognizes that his test statistics are generally degenerate as he tests nested hypotheses. To avoid this,[2] Yatchew splits the sample into two independent subsamples and calculates the restricted sum of squared residuals from one and the unrestricted sum of squared residuals from the other. Whang and Andrews (1993) also use sample splitting to deal with the same problem. Robinson (1991) uses a sample weighting method which is similar to sample splitting to deal with a degeneracy problem arising in his entropy-based test.

We now apply the sample splitting technique to modify D. Without loss of generality, write $n = 2\bar{n}$. Calculate $\hat{I}^{*1}$ and $\hat{I}^{*2}$ from $\{x_i, z_i\}_{i=1}^{\bar{n}}$ and $\{x_i, z_i\}_{i=\bar{n}+1}^{n}$. Then $D_{\text{split}} := \sqrt{\bar{n}}(\hat{I}^{*1} - \hat{I}^{*2})$ does not degenerate. To see this, note that the degeneracy occurs as the two terms in the second line of (8.4) in the proof of Theorem 6.1 cancel each other if the pseudo true measure coincide. Each of the terms is, however, asymptotically normal with variance

$$\eta_1^2 := \int \left[r_1(x,z) - I^{*1}\right]^2 d\mu_{X,Z}$$

and

$$\eta_2^2 := \int \left[r_2(x,z) - I^{*2}\right]^2 d\mu_{X,Z},$$

respectively (see (6.1) for the definition of r_1 and r_2). With sample splitting, the two terms become independent. Consequently the next result holds regardless of the nature of σ^2.

Theorem 6.2. *Suppose Assumptions 3.1 and 3.2 hold. Then* $D_{\text{split}} \rightsquigarrow N(0, (\eta_1^2 + \eta_2^2))$ *under* (**H**).

Remark 6.2. If $(P^{*1}_{Z|x})_{x\in\mathcal{X}}$ and $(P^{*2}_{Z|x})_{x\in\mathcal{X}}$ coincide, then $\eta_1^2 = \eta_2^2 = \eta^2$, say, so the limiting distribution of D_{split} is $N(0, 2\eta^2)$. In general, of course, it is unknown *a priori* if the two measures coincide or not. Therefore we should not use $2\eta^2$ as an expression of the asymptotic variance for D_{split}. η_1^1 and η_2^2 can be consistently estimated using appropriate sample counterparts. For example, using $\hat{r_1}_i$ and $\hat{r_2}_i$ (see (6.2)), define

$$\hat{\eta}_1^2 = \bar{n}^{-1}\sum_{i=1}^{\bar{n}}\left[\hat{r_1}_i - \hat{I}^1(\hat{\theta}^1)\right]^2$$

[2] If the both models are correctly specified in our d-test, it is possible to analyze the degenerate distribution following approaches as in Hong and White (1995) and Tripathi and Kitamura (2001).

and

$$\hat{\eta}_2^2 = \bar{n}^{-1} \sum_{i=\bar{n}+1}^{n} \left[\hat{r_{2i}} - \hat{I}^2(\hat{\theta}^2)\right]^2.$$

An appropriate model comparison test statistic is then given by

$$d_{\text{split}} := \frac{D_{\text{split}}}{\sqrt{\hat{\eta}_1^2 + \hat{\eta}_2^2}}.$$

This statistic is asymptotically standard normal-distributed under the (**H**), whether or not the degeneracy of D occurs.

6.4. **Testing between a Parametric Conditional Model and a Conditional Moment Restriction Model.** Sections 2 and 3 focused on conditional moment restriction models. As noted in Section 1, however, our basic framework remains valid for other models that are defined conditionally. In this subsection we explore a method that can be used to compare a conditional moment restriction model with a fully parametric conditional model. Such a comparison can be interesting in practice. That is, using statistical evidence to support the decision between a parametric MLE and a method of moment is potentially useful.

Chen, Hong, and Shum (2001) propose comparing an unconditional moment restriction model with an unconditional parametric model. Their idea is to use the empirical likelihood and the MLE to construct an appropriate Vuong-type test. This is an interesting proposal. The procedure in this subsection can be regarded as a generalization of the Chen-Hong-Shum method. The use of conditional models, on the other hand, is likely to be crucially important in many applications. First, a large majority of models in econometrics are defined conditionally. This led Vuong (1989) to consider conditional MLE in construction of the original Vuong test. A conditional parametric model minimizes conditional relative entropy, not unconditional relative entropy. Therefore a fair comparison would be made using the minimum conditional entropy. Second, notice that both a conditional parametric model and a conditional moment restriction model impose an infinite number of moment conditions, whereas an unconditional moment restriction model by definition imposes only a finite number of moment conditions. In most applications the number of moment conditions for an unconditional moment restriction model is somewhat arbitrary and often small. In some cases a model is just identified (as in, for example, regression models based on standard orthogonality conditions), in which case the minimized entropy is zero both in population and in finite samples. This makes results from tests between an unconditional parametric model and an unconditional moment condition model rather difficult to interpret — it is a tricky matter to make a fair comparison in such a situation. We, therefore, focus on comparing conditional models.

Suppose we wish to compare the parametric model in Section 2.2 with the conditional moment restriction model in Section 2.3. Recall our approximation criterion I^* for the conditional moment restriction model and its normalized log-likelihood are given by (2.6) and (2.16), respectively. To avoid confusion, let us call the former $I^*_{\text{parametric}}$ and the latter I^*_{cm}. Similarly, call the goodness-of-fit statistics in (4.1) and (4.2) $\hat{I}^*_{\text{parametric}}$ and $\hat{I}^*_{\text{cm}}$, respectively.

Our null hypothesis is

$$(\mathbf{H}^{\#}) \qquad I^*_{\text{parametric}} = I^*_{\text{cm}},$$

i.e., the parametric model and the conditional moment restriction model approximate the true conditional probability measure equally well. To test this, form a likelihood ratio type statistic as before. Let $D^{\#} = \sqrt{n}(\hat{I}^*_{\text{parametric}} - \hat{I}^*_{\text{cm}})$. Define

$$r(x,z) = \log(1 + \lambda(x,\theta^*)' g(z,\theta^*))$$

as in (6.1). The next theorem shows that that $D^{\#}$ is asymptotically normal with mean 0 and variance

$$\sigma^{\#^2} := \int \left\{ r(x,z) - \left[\log f(z|x) - \log f^{\gamma^*}(z|x) \right] \right\}^2 d\mu_{X,Z}.$$

Theorem 6.3. *Suppose Assumptions 3.1 and 3.2 and assumptions given by Vuong (1989) hold. Then* $D^{\#} \rightsquigarrow N(0, \sigma^{\#^2})$ *under* $(\mathbf{H}^{\#})$.

Remark 6.3. Once we establish the asymptotic normality of $D^{\#}$, we can follow our approach in Sections 6.2 and 6.3 to operationalize our model comparison test. Here is a brief outline of our procedure. If $\sigma^{\#^2}$ is known to be strictly positive, the approach in Section 6.2 works. Define $\hat{r}_i, i = 1, ..., n$ following the formulation in (6.2) and

$$\hat{\sigma}^{\#^2} := n^{-1} \sum_{i=1}^{n} \left\{ \hat{r}_i - \left[\log \hat{f}(z_i|x_i) - \log f^{\hat{\gamma}}(z_i|x_i) \right] \right\}^2.$$

Let $d^{\#} = \frac{D^{\#}}{\hat{\sigma}^{\#}}$. By Theorem 6.3, the studentized likelihood ratio statistic has the limiting distribution as

$$d^{\#} \rightsquigarrow N(0,1)$$

under $(\mathbf{H}^{\#})$.

If the possibility of degeneracy is a concern, we can follow Section 6.3 to employ sample-splitting. Divide the whole sample as before, and calculate $\hat{I}^*_{\text{parametric}}$ and $\hat{I}^*_{\text{cm}}$ using the first subsample and the second subsample, respectively. Call the resulting versions $\hat{I}^*_{\text{parametric}}$ and $\hat{I}^*_{\text{cm}}$.

The likelihood ratio statistic $D^{\#}_{\text{split}} := \sqrt{\bar{n}}(\hat{I}^*_{\text{parametric}} - \hat{I}^*_{\text{cm}})$ is asymptotically normal under the null even with degeneracy. To estimate its variance, take the sum of $\hat{\eta}_1^2$ (see Remark 6.2) and

$$\hat{\eta}_2^2 = \bar{n}^{-1} \sum_{i=\bar{n}+1}^{n} \tau_i \left[\log \hat{f}(z_i|x_i) - \log f^{\hat{\gamma}}(z_i|x_i) - \hat{I}^2_{\text{parametric}}(\hat{\gamma}^2)\right]^2 .$$

A studentized version of $D^{\#}_{\text{split}}$, e.g.,

$$d^{\#}_{\text{split}} := \frac{D^{\#}_{\text{split}}}{\sqrt{\hat{\eta}_1^2 + \hat{\eta}_2^2}}$$

is then distributed according to the standard normal under $(\mathbf{H}^{\#})$ asymptotically.

7. Conclusion

A general framework for analyzing conditional models is presented. Consequences of model misspecification in the application of the "conditional empirical likelihood" (Kitamura, Tripathi, and Ahn (2000)) method are discussed. This investigation leads to a proposal of a likelihood-based goodness fit of measure for a broad range of models. Also, likelihood ratio tests are developed for hypotheses involving conditional moment restriction models. For example, a test between a parametric model and a conditional moment restriction model is proposed. The asymptotic distribution of the proposed statistics is shown to be the standard normal. These procedures are easy to implement in practice.

Though the theory developed in the paper assumes random sampling, the methodology is potentially applicable to dependent series. As it has been shown that empirical likelihood can handle dependent data with appropriate modifications (see Kitamura (1997)), it would be interesting to extend our methodology to dynamic conditional models.

8. Appendix

In what follows we provide some sketches of proofs. Details are omitted in the current version. Let $c_n = (\frac{1}{n^{1-\delta}b^{s+\delta}})^{1/2}$.

Proof of Theorem 3.1. Note that the optimality of $\hat\lambda(x_i,\theta), i = 1,...,n$ implies that

$$\hat I(\theta) = n^{-1}\sum_{i=1}^n \tau_i \sum_{j=1}^n w_{ij}\log(1+\hat\lambda(x_i,\theta)'g(z_j,\theta))$$
$$\geq n^{-1}\sum_{i=1}^n \tau_i \sum_{j=1}^n w_{ij}\log(1+\lambda(x_j,\theta)'g(z_j,\theta))$$

for all $\theta\in\Theta$. Under the stated assumptions, $\tau_i\sum_{j=1}^n w_{ij}\log(1+\lambda(x_j,\theta)'g(z_j,\theta))$ converges to $\int\log(1+\lambda(x_i,\theta)'g(z,\theta))d\mu_{Z|x_i}$ uniformly in i and $\theta\in\Theta$. Therefore the second line of the above inequality has the following approximation,

$$n^{-1}\sum_{i=1}^n \tau_i \int \log(1+\lambda(x_i,\theta)'g(z,\theta))d\mu_{Z|x_i} + o_p(1),\ \text{uniformly in } \theta,$$

which converges to

$$I(\theta) = \int \log(1+\lambda(x_i,\theta)'g(z,\theta))d\mu_{X,Z}$$

in probability uniformly in θ by the UWLLN. Consequently, we have the one-sided uniform convergence result, i.e. $\inf_{\theta\in\Theta}(\hat I(\theta) - I(\theta)) \geq 0$ with probability approaching to one. Moreover, under the assumptions $\max_i \tau_i\|\hat\lambda(x_i,\theta^*) - \lambda(x_i,\theta^*)\| = o_p(c_n)$ uniformly in i. Consequently, $\hat I(\theta^*) \xrightarrow{p} I(\theta^*)$. As $I(\cdot)$ is continuous, by Assumptions 3.1(i),(ix), the result follows. □

Proof of Theorem 6.1. First, we approximate $\hat I^{*1}$ and $\hat I^1(\hat\theta^1)$ using $\tilde I^1(\theta^{1*})$ and $\tilde I^2(\theta^{2*})$ as defined below. Since the results are symmetric between the two models, in what follows we drop the suffix "1" and "2" until it becomes necessary. Let $\tilde I:\Theta\to\mathbb{R}_+$ such that

$$\tilde I(\theta) = n^{-1}\sum_{i=1}^n \tau_i \sum_{j=1}^n w_{ij}\log(1+\lambda(x_i,\theta^*)'g(z_j,\theta^*)).$$

Our approximation proceeds in two steps. The first step is similar to Vuong (1989)'s approximation for a parametric log-likelihood function. The second step is necessary to handle the nonparametric part of the model.

Express $\hat{I}(\theta^*)$ by expanding $\hat{I}(\cdot)$ around $\hat{\theta}$ to the second order using an appropriate mean value $\bar{\theta}$ (the case where the objective function is non-smooth will be included in a later version):

$$\hat{I}(\theta^*) = \hat{I}(\hat{\theta}) + (\theta^* - \hat{\theta})\nabla_\theta \hat{I}(\hat{\theta}) + \frac{1}{2}(\theta^* - \hat{\theta})'\nabla_{\theta\theta}\hat{I}(\bar{\theta})(\theta^* - \hat{\theta}).$$

Note $\nabla_\theta \hat{I}(\hat{\theta}) = 0$ (with probability approaching to one). Moreover, we have shown that $\|\theta^* - \hat{\theta}\| = O_p(n^{-1/2})$ and $\nabla_{\theta\theta}\hat{I}(\bar{\theta}) \xrightarrow{p} H$ as the proof of Theorem 3.2 implies. Therefore the quadratic term is $O_p(n^{-1})$. As a consequence

$$\hat{I}(\theta) = \hat{I}(\theta^*) + O_p(n^{-1}). \tag{8.1}$$

Next, recall the definition

$$\hat{I}(\theta^*) = n^{-1}\sum_{i=1}^n \tau_i \sum_{j=1}^n w_{ij}\log(1 + \hat{\lambda}(x_i,\theta^*)'g(z_j,\theta^*)). \tag{8.2}$$

Notice that the i-th term of the "inner summation" in (8.2) can be rewritten by applying another second order Taylor expansion around $\lambda(x_i,\theta^*)$ with a mean value $\bar{\lambda}(x_i,\theta^*)$:

$$\begin{aligned}
&\sum_{j=1}^n w_{ij}\log(1 + \hat{\lambda}(x_i,\theta^*)'g(z_j,\theta^*)) \\
&= \sum_{j=1}^n w_{ij}\log(1 + \lambda(x_i,\theta^*)'g(z_j,\theta^*)) - (\lambda(x_i,\theta^*) - \hat{\lambda}(x_i,\theta^*))'\sum_{j=1}^n w_{ij}\frac{g(z_j,\theta^*)}{1 + \hat{\lambda}(x_i,\theta^*)'g(z_j,\theta^*)} \\
&+ \frac{1}{2}(\lambda(x_i,\theta^*) - \hat{\lambda}(x_i,\theta^*))'\sum_{j=1}^n w_{ij}\frac{g(z_j,\theta^*)g(z_j,\theta^*)'}{(1 + \bar{\lambda}(x_i,\theta^*)'g(z_j,\theta^*))^2}(\lambda(x_i,\theta^*) - \hat{\lambda}(x_i,\theta^*)).
\end{aligned} \tag{8.3}$$

Again, the linear term drops out by the definition of $\hat{\lambda}(\cdot,\cdot)$. Under the rate condition in Assumption 3.1(viii), $\max_i \|\lambda(x_i,\theta^*) - \hat{\lambda}(x_i,\theta^*)\|^2 = o_p(n^{-1/2})$. Also, under the stated assumptions

$$\max_i \tau_i \|\sum_{j=1}^n w_{ij}\frac{g(z_j,\theta^*)g(z_j,\theta^*)'}{(1 + \bar{\lambda}(x_i,\theta^*)'g(z_j,\theta^*))^2}\| = O_p(1).$$

In sum, the quadratic term in (8.3) is of order $o_p(n^{-1/2})O_p(1) = o_p(n^{-1/2})$ uniformly in i. Then by combining (8.1), (8.2) and (8.3),

$$\hat{I}(\hat{\theta}) = \tilde{I}(\theta^*) + o_p(n^{-1/2}).$$

Applying the above approximation to $\hat{I}^{*1}$ and $\hat{I}^{*2}$, we get

$$\begin{aligned}
D &= \sqrt{n}(\hat{I}^{*1} - \hat{I}^{*2}) \\
&= \sqrt{n}(\tilde{I}^{*1} - \tilde{I}^{*2}) + o_p(1)
\end{aligned} \tag{8.4}$$

By standard U-statistic theory (see, for example, Powell, Stock, and Stoker (1989) or Härdle and Stoker (1989)), we obtain the following asymptotic normality property:

$$\sqrt{n}(\tilde{I}^{*1} - \tilde{I}^{*2}) \rightsquigarrow N(0, \sigma^2). \tag{8.5}$$

(8.4) and (8.5) imply the desired result. □

References

Ahmad, I., and P. Lin (1976): "A nonparametric estimation of the entropy for absolutely continuous distributions," *IEEE Transactions on Information Theory*, pp. 372–375.

Beirlant, J., E. Dudewicz, L. Györfi, and E. Meulen (1997): "Nonparametric Entropy Estimation: An Overview," *International Journal of Mathematical and Statistical Sciences*, 6, 17–39.

Borwein, J., and A. Lewis (1993): "Partially-finite Programming in L_1 and the Existence of Maximum Entropy Estimates," *SIAM Journal of Optimization*, 3, 248–267.

Brown, B. W., and W. K. Newey (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica*, 66(2), 453–464.

Buchinsky, M. (1994): "Changes in the U.S. Wage Structure 1963–1987: Application of Quantile Regression," *Econometrica*, 62, 405–458.

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

——— (1994): "Quantile Regression, Censoring, and the Structure of Wages," in *Advances in Econometrics: Sixth World Congress*, vol. 1, pp. 171–209. Cambridge University Press.

Chen, X., H. Hong, and M. Shum (2001): "Nomparametric Likelihood Selection Tests for Parametric versus Moment Condition Models," Manuscript.

Christoffersen, P., J. Hahn, and A. Inoue (2001): "Testing and Comparing Value at Risk Measures," *Journal of Empirical Finance*, 8, 325–342.

Cover, T. M., and J. A. Thomas (1991): *Elements of Information Theory*. John Wiley and Sons, New York.

Cox, D. (1961): "Tests of Separate Families of hypotheses," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics*, 1, 406–424.

Csiszár, I. (1995): "Generalized Projections for Non-Negative Functions," *Acta Mathematica Hungarica*, 68, 161–185.

Hansen, L. P., and T. J. Sargent (1993): "Seasonality and Approximation Errors in Rational Expectation Models," *Journal of Econometrics*, 55, 21–55.

——— (2001): "Robust Control and Model Uncertainty," *American Economic Review*, 91, 60–66.

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

Hong, H., B. Preston, and M. Shum (2001): "Empirical Likelihood-Based Model Selection Criteria for Moment Condition Models," Manuscript.

Hong, Y., and H. White (1995): "Consistent Specification Testing via Nonparametric Series Regressions," *Econometrica*, 63, 1133–1160.

Imbens, G. W., R. H. Spady, and P. Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333–357.

Johnson, W., Y. Kitamura, and D. Neal (2000): "Evaluating a Simple Method for Estimating Black-White Gaps in Median Wages," *American Economic Review*, 90, 339–343.

Kent, J. T. (1986): "The Underlying Structure of Nonnested Hypothesis Tests," *Biometrika*, 73(2), 333–343.

Kim, T.-H., and H. White (2002): "Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regression," *Advances in Econometrics*, forthcoming.

Kitamura, Y. (1997): "Empirical likelihood methods with weakly dependent processes," *Annals of Statistics*, 25(5), 2084–2102.

——— (1998): "Comparing Misspecified Dynamic Econometric Models Using Nonparametric Likelihood," Working Paper, Department of Economics, University of Wisconsin.

——— (2001): "Asymptotic Optimality of Emptrical Likelihood for Testing Moment Restrictions," *Econometrica*, 69(6), 1661–1672.

Kitamura, Y., and M. Stutzer (1997): "An Information Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65(4), 861–874.

——— (2002): "Connections between Entropic and Linear Projections in Asset Pricing Estimation," *Journal of Econometrics*, 107(1-2), 159–174.

Kitamura, Y., G. Tripathi, and H. Ahn (2000): "Empirical Likelihood-Based Inference for Conditional Moment Restriction Models," Manuscript, Department of Economics, University of Wisconsin-Madison.

Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

LeBlanc, M., and J. Crowley (1995): "Semiparametric Regression Functionals," *Journal of the American Statistical Association*, 90(429), 95–105.

Manski, C. F. (1988): *Analog Estimation Methods in Econometrics.* Chapman and Hall.

Newey, W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

——— (1993): "Efficient Estimation of Models with Conditional Moment Restrictions," in *Handbook of Statistics, vol. 11*, ed. by G. Maddala, C. Rao, and H. Vinod. Elsevier Science B.V., pp. 2111–2245.

Newey, W. K., and R. J. Smith (2000): "Asymptotic Bias and Equivalence of GMM and GEL," Manuscript.

Owen, A. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75(2), 237–249.

——— (1990): "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, 18(1), 90–120.

——— (1991): "Empirical Likelihood for Linear Models," *The Annals of Statistics*, 19(4), 1725–1747.

——— (2001): *Empirical Likelihood.* Chapman and Hall.

Pakes, A., and S. Olley (1995): "A Limit Theorem for a Smooth Class of Semiparametric Estimators," *Journal of Econometrics*, 65(1), 295–332.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57(6), 1403–1430.

Robinson, P. (1991): "Consistent Nonparametric Entropy-Based Testing," *Review of Economic Studies*, 58, 437–453.

Smith, R. J. (1997): "Alternative Semi-parametric Likelihood Approaches to Generalized Method of Moments Estimation," *Economic Journal*, 107, 508–519.

Tripathi, G., and Y. Kitamura (2001): "Testing Conditional Moment Restrictions," *Annals of Statistics*, forthcoming.

Vuong, Q. (1983): "Misspecification and Conditional Maximum Likelihood Estiamtion," Social Science Working Paper 503, California Institute of Technology.

——— (1989): "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," *Econometrica*, 57, 307–333.

WHANG, Y., AND D. ANDREWS (1993): "Tests of Specification for Parametric and Semiparametric Models," *Journal of Econometrics*, 57, 277–318.

WHITE, H. (1980): "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review*, 21, 149–170.

——— (1981): "Consequences and Detection of Misspecified Nonlinear Regression Models," *Journal of the American Statistical Society*, 76, 419–433.

——— (1982): "Maximum Likelihood Estimation of Missoecified Models," *Econometrica*, 50, 1–25.

YATCHEW, A. (1992): "Nonparametric Regression Tests Based on Least Squares," *Econometric Theory*, 8, 435–451.

ZHANG, J., AND I. GIJBELS (2001): "Sieve Empirical Likelihood and Extensions of the Generalized Least Squares," *Scandinavian Journal of Statistics*, Forthcoming.

DEPARTMENT OF ECONOMICS, UNIVERSITY OF PENNSYLVANIA, PHILADELPHIA, PA-19104.

E-mail address: ykitamur@ssc.upenn.edu