

USING MIXTURES IN ECONOMETRIC MODELS: A BRIEF REVIEW AND SOME NEW RESULTS

GIOVANNI COMPIANI AND YUICHI KITAMURA*

Very Preliminary

ABSTRACT. This paper is concerned with applications of mixture models in econometrics. Focused attention is given to semiparametric and nonparametric models that incorporate mixture distributions, where important issues about model specifications arise. For example, whether to use a finite mixture or a continuous mixture makes a significant impact on model identifiability as well as estimation. Likewise, the dimension of the latent mixing variables is a critical issue, in particular when a continuous mixture is used. Applications of mixture models to address various problems in econometrics, such as unobserved heterogeneity and multiple equilibria, are presented. New nonparametric identification results are developed for finite mixture models with testable exclusion restrictions — but no large support assumptions — on covariates. The results apply to mixtures with both continuous and discrete covariates, delivering point identification under weak conditions.

1. INTRODUCTION

Mixture models are widely used in economic applications. First of all, mixtures are commonly employed as a modelling device to account for unobserved heterogeneity. Applications range from the labour to the industrial organization literature. Key references include Berry, Carnall, and Spiller (2006), Keane and Wolpin (1997), and Cameron and Heckman (1998). Mixtures also arise when the unobserved element that varies across individuals is the form of their utility functions, as in random utility models (RUM). Another application in microeconometrics is the treatment of multiple equilibria in discrete games (e.g. Cooper (2002); Berry and Tamer (2006); Ciliberto and Tamer (2009)). Measurement error models can be reformulated in terms of mixtures, as in Horowitz and Manski (1995) and Manski (2003). Heckman and Singer (1984a), Heckman and Singer (1984b) and

Date: This Version: November 29, 2013.

Keywords:

JEL Classification Number:

*Kitamura acknowledges financial support from the National Science Foundation. The authors thank Tim Armstrong, Phil Haile, Arthur Lewbel, Lorenzo Magnolfi, Camilla Roncoroni, and in particular Lars Nesheim for very helpful comments.

Abbring and van den Berg (2003) apply mixtures to the analysis of duration models. Finite mixtures have also been employed in regime switching models (e.g. Cho and White (2007)).

The above is an incomplete list of the many applied settings in which mixture models arise. Each of these applications will be analysed in subsequent sections.

The tremendous importance of mixtures in economics as well as other fields has motivated a large body of theoretical research both in the statistics and in the econometrics literature. Given the vastness of this research area, a comprehensive account is far beyond the scope of this paper. Attention will be focussed mainly on the most recent advances and, notably, on the semiparametric and nonparametric treatment of mixtures. These developments have significantly increased the flexibility of the models and reduced reliance on restrictive parametric assumptions. This is important because consistency of estimators for the parameter of interest typically hinges on the correct specification of the unobserved heterogeneity distribution. Cases where this occurs include proportional hazard models (as in Heckman and Singer (1984b)), discrete choice models (e.g., Matzkin (1992)) and switching regression models (e.g., Kitamura (2003)).

The greater flexibility allowed by the semiparametric and nonparametric approaches often comes at the cost of reduced identified power. In other words, relaxing the traditional parametric assumptions often does not allow the researcher to uniquely identify the parameters of interest and only partial identification results can be obtained. A variety of examples in the paper illustrate this trade-off.

We now lay out the general model that will be used throughout the paper. Consider the following:

$$(1.1) \quad F(y) = \int F(y|\alpha) dG(\alpha).$$

In equation (1.1), F is a cumulative distribution function (cdf), $F(\cdot|\alpha)_{\alpha \in \mathcal{A}}$ is a collection of cdf's indexed by a random variable $\alpha \in \mathcal{A}$, where \mathcal{A} is a (possibly infinite) set, and G is a cdf on the space \mathcal{A} . Throughout the paper, we will employ the following terminology. The cdf F will be referred to as the mixture distribution, the cdf's $F(\cdot|\alpha)_{\alpha \in \mathcal{A}}$ will be called the component distributions, G the mixing distribution and α the latent (or mixing) variable. In words, equation (1.1) says that the distribution F is a mixture with components $F(\cdot|\alpha)_{\alpha \in \mathcal{A}}$ and mixing distribution G . In the case where α takes on a finite number of values, the probability mass attached by the distribution G to each α_i is often called the weight of component i .

In a typical setting, the econometrician observes repeated draws $(y_i)_{i=1}^n$ from the mixture distribution F and wishes to estimate the unknown parameters on the right-hand side of (1.1), namely

the component distributions and the mixing distribution. Note that if the mixing variable α were observed, i.e. if the sample were of the form $(y_i, \alpha_i)_{i=1}^n$, then identification and estimation of the unknown parameters would become trivial tasks. More interesting questions arise when α is not observed, which explains why α is often labelled the latent variable and shows that mixtures may be interpreted as missing variable models.

The basic framework of equation (1.1) can be specialized and extended in several directions. First of all, the latent variable α can take a finite number of values or an infinite (typically uncountable) number of values. The case of a finite mixture is dealt with in section 2, while general mixtures are addressed in section 3. Further, the econometrician needs to choose whether to specify a parametric form for the component distributions or to pursue a nonparametric approach. A similar decision applies to the mixing distribution; in particular, in the case of finite mixtures, the question is whether or not to assume that the number of components is known a priori. Thirdly, the dimension of the latent variable α can play an important role, particularly in the case of mixtures with infinitely many components. Finally, one may want to introduce covariates in the basic model of equation (1.1). As will be discussed in the following sections, the assumptions about the way in which the component distributions and/or the mixing distribution depend on the covariates are key to obtaining identification of the parameters of interest.

This paper is organized as follows. Section 2 discusses finite mixtures and presents some novel identification results. Section 3 deals with possibly infinite mixtures; in particular, panel data models, random coefficient models and random utility models are considered. Section 4 concludes.

2. FINITE MIXTURES

Finite mixtures have been analysed extensively in the statistical literature. The survey by Frühwirth-Schnatter (2006) discusses identification and estimation of finite mixtures, with a focus on both the frequentist approach (specifically, likelihood-based methods) and the Bayesian approach. Model specification - and notably the question of how to determine the number of component distributions - is also covered. Another exhaustive account of the statistical literature on mixtures (including infinite mixtures) is provided by Lindsay (1988).

Most of the literature has taken a parametric approach to mixture models, which means that the component distributions are assumed to be known up to finite dimensional parameters. Under such assumptions, identification of the parameters of interest is usually straightforward. However, these identification results are fully driven by the chosen functional form and model misspecification

may lead to inconsistent estimates and invalid inference. Therefore, a growing body of research both in statistics and in econometrics investigates how semiparametric and nonparametric methods can be applied to mixture models. In what follows, we will consider some of these recent developments. Four main cases will be discussed: multiple outcomes with independence property, the case where covariates enter the mixing weights, the case where covariates enter the component cdf's, and a fourth setting where covariates enter both the mixing weights and the component cdf's. For the latter model, new identification results are presented.

2.1. Multiple outcomes with independence property. An important contribution to the non-parametric analysis of mixtures in the statistical literature is provided by Hall and Zhou (2003). They consider two-component mixtures with independence property. The model takes the form:

$$(2.1) \quad F(\mathbf{y}) = \lambda \prod_{j=1}^k F_{j1}(y_j) + (1 - \lambda) \prod_{j=1}^k F_{j2}(y_j),$$

where $\mathbf{y} = (y_1, \dots, y_k)'$ and the cdf of each component factorizes by the independence assumption. The motivation for considering this model comes from the clinical trial literature. In particular, one can think of \mathbf{y} as a vector of outcomes for k clinical tests, whose distribution is allowed to vary depending on whether a patient is affected by a given disease or not. The researcher observes \mathbf{y} , but does not observe the patients' disease status nor the proportion of people affected. Therefore, the problem is to nonparametrically identify and estimate the cdf's F_{ji} , for $j = 1, \dots, k$ and $i = 1, 2$, and the mixture weight λ based on random draws from F . Clearly, if $k = 1$, identification is a hopeless task. For instance, it is easy to see that the same cdf F can be obtained as two different linear combinations of uniform cdf's (see Frühwirth-Schnatter (2006), p.21). For the case $k = 2$, Hall and Zhou (2003) show that the parameters of interest are not identified. In particular, given the estimable cdf F , equation (2.1) has a continuum of solutions indexed by two scalar parameters. On the other hand, when $k = 3$, the authors show that identification (up to switching of the two products on the right-hand side of (2.1)) is achieved under an irreducibility condition¹ on the density f of the cdf F . The proof exploits the fact that every lower-dimensional submodel derived from (2.1) imposes a different restriction on the parameters of interest. The total number of restrictions is $2^k - 1$ while the number of parameters is $2k + 1$. Note that for $k = 3$ these two numbers coincide. Therefore, when

¹The density f is said to be irreducible if none of its bivariate marginal densities factorises into the product of the two corresponding univariate marginals.

$k < 3$ the model is not identified, when $k = 3$ we have just-identification and when $k > 3$ there are over-identifying restrictions.

Given identification (i.e. for $k = 3$), Hall and Zhou (2003) propose a fully nonparametric estimator for the component distributions and the mixing weight. The estimator is obtained by minimising an integrated distance between an estimator of F that does not take into account the structure of the model as specified in (2.1) and an expression (involving the unknowns with respect to which the minimisation is carried out) which is instead based on equation (2.1). This estimation procedure is close in spirit to nonparametric maximum likelihood and empirical likelihood; see Laird (1978), Qin (1998), Qin (1999) and Zou, Fine, and Yandell (2002) for how these frameworks can be applied to mixture models.

The estimators proposed by Hall and Zhou (2003) for the component distributions and the mixing weight are strongly, uniformly consistent. Moreover, under regularity conditions², the estimator for λ achieves the root- n rate of convergence and the estimators for the component are root- n consistent in L^2 .

An attractive feature of the results above is that they rely on relatively mild regularity conditions and only require the number of observations for each individual in the sample to be larger or equal to 3. In particular, this means that the results can be applied to very short panel datasets.

The seminal work by Hall and Zhou (2003) has been extended and complemented by several papers in both the statistics and the econometrics literature. Among the latest developments, Bonhomme, Jochmans, and Robin (2012) provide sufficient conditions for nonparametric identification of the component distributions, the mixing weights and the number of components³. The conditions impose some restrictions on the component densities, namely absolute continuity, square integrability and lack of multicollinearity among the Fourier coefficients of an expansion of the component densities. The identification proof is quite different from that of Hall and Zhou (2003), which allows Bonhomme, Jochmans, and Robin (2012) to obtain two novel results. First, in their setup the number of components is identified from the data (as opposed to being assumed). They provide a consistent estimator for this parameter, although the effect of pretesting the number of components on estimation of the mixing weights and component distributions is not discussed. Secondly, even if one were to take the number of components as given, their approach is more flexible as it allows for arbitrary mixtures,

²A sufficient condition for root- n convergence is that the component distributions be compactly supported. This condition can be relaxed by imposing constraints on the tail behavior of the component distributions.

³As in Hall and Zhou (2003), the dimension of the outcome variable is assumed to be at least 3.

while Hall and Zhou (2003) only deal with the two-component case. Finally, the estimators proposed by Bonhomme, Jochmans, and Robin (2012) are computationally much less costly than Hall and Zhou (2003)'s nonparametric ML estimators.

Another relevant extension encompassing a variety of economics applications can be found in Kasahara and Shimotsu (2009). The focus is on dynamic discrete choice models with unobserved heterogeneity, where mixtures represent the different (latent) types that the agents belong to. The assumption of independence across measurements from a given mixture component is maintained. However, a key difference with respect to equation (2.1) is the introduction of covariates. Indeed, the main source of identifying power in this setting is exactly the fact that the outcome variable Y responds differently to changes in the covariates for different types. In other words, identification requires the covariates to affect the conditional choice probabilities of the agent in a sufficiently heterogeneous way across types. Importantly, the presence of covariates makes identification possible even with relatively short panels, which is the type of datasets that are often available in applications. The previous literature, instead, focused mostly on the long-panel case (see e.g. the estimation exercise in Browning and Carro (2007), where the length of the panel exceeds 100 time periods). The paper also provides sufficient conditions for the nonparametric identification of the number of components, i.e. the number of types in the dynamic discrete choice setting. Interestingly, the number of types can be nonparametrically identified even with just two-period panels. The proof relies on rank assumptions requiring the change in the covariates to have a sufficiently strong effect on the conditional choice probabilities. However, such conditions (which are also needed for identification of the other parameters of interest) cannot be tested using the data.

A computationally attractive procedure for estimation of dynamic discrete models with unobserved heterogeneity is proposed by Kasahara and Shimotsu (2011).

2.2. Covariates in the mixing weights. Another setup considered in the literature is the following:

$$(2.2) \quad F(y|x, w) = \sum_{j=1}^J \lambda_j(x, w) F_j(y|x).$$

Henry, Kitamura, and Salanié (2014) show that, in general, this model is only partially identified. However, the characterization of the identified set is constructive and can be used to extract useful information from the data, such as the nonparametric identification of the number of mixture components.

The main assumption on which the model relies is embedded in equation (2.2) and takes the form

of an excluded covariate restriction. The assumption states that the covariate W affects the mixing weights, but does not change the component distributions. The entire analysis is conditional on values of X for which the excluded covariate assumption holds. Since no other property is imposed on X , conditioning on this variable will be left implicit throughout. In accordance with instrumental variable models, it is also required that the dependence of the mixing weights on the covariate W be strong enough.

The exclusion restriction on W can be justified in a number of applications. First, consider Markov switching models. Assume that the outcome variable Y is a finite-order autoregression conditionally on an unobserved Markov chain of order m . The hidden Markov chain determines the distribution of the outcome variable; for instance, it could determine the expectation (as in mean switching models) or the variance (as in stochastic volatility models). This type of models can be expressed in the form of equation (2.2) by setting $X = (Y_{t-1}, \dots, Y_{t-m})$ and $W = (Y_{t-m-1}, \dots, Y_1)$.

Another example is data contamination. In this setup, the researcher observes an outcome variable Y and potentially flawed measurements T of an underlying categorical regressor T^* . The case of a discrete mismeasured regressor (often referred to as misclassification) is especially relevant, given that in this setting the classical assumption of independence between measurement error and true value is often untenable (see, e.g., Bound, Brown, and Mathiowetz (2001)). This implies that traditional methods based on deconvolution (e.g. Li (2002)) cannot be applied.

A common assumption in the misclassification literature is that, conditional on the true value of the regressor T^* , the outcome Y and the observed regressor T are independent. This assumption is sometimes referred to as nondifferential measurement error. Under this restriction, we can write

$$(2.3) \quad F_{y|T}(Y|T) = \sum_{j=1}^J F_{Y|T,T^*}(y|T, T^* = t_j) \mathbb{P}\{T^* = t_j|T\} = \sum_{j=1}^J F_{Y|T^*}(y|T^* = t_j) \mathbb{P}\{T^* = t_j|T\}.$$

This equation shows that misclassification models may be written in the form of model (2.2) by setting the excluded variable W equal to the observed regressor T and $\lambda_j = \mathbb{P}\{T^* = t_j|T\}$. Here the role of the latent variable giving rise to the mixture is played by the unobserved correctly-classified regressor T .

The nondifferential measurement error assumption is not innocuous. For instance, it may not be satisfied if misclassification is correlated with non-compliance in an experimental setting, or if misclassification stems from misreporting which is correlated with unobserved heterogeneity. Therefore, whether or not using model (2.2) is appropriate depends on the specific application that is being

considered. Further references in the misclassification literature include Molinari (2008) and Chen, Hong, and Nekipelov (2011).

A third setting where model (2.2) can be applied is microeconomic models with unobserved heterogeneity. For example, if the outcome variable is demand for a good, the researcher may be interested in allowing both for observed heterogeneity across buyers (given by the covariates for which data are available) and for unobserved heterogeneity (given by a finite number of types and, possibly, an idiosyncratic shock). In this case, the excluded instrument W could be a set of geographical variables that do not enter preferences nor covariates - and thus do not affect demand for a given type - but do change the distribution of buyer types. Another example is an oligopoly model, where the outcome variable is again demand, the observed covariates are prices and (possibly mismeasured) costs, and the excluded instrument is (possibly mismeasured) profits. As long as profits do not belong to the buyer's information set, they do not affect demand given the covariates and the buyer type. Further, if profits do have an influence on the composition of demand, then the instrument relevance condition is satisfied as well. Additional details on this model can be found in Henry, Kitamura, and Salanié (2014).

Finally, equation (2.2) may be used to account for multiple equilibria in economic models. In this context, each element of the mixture represents a different equilibrium and the mixing weights correspond to the probability distribution over equilibria given by some selection mechanism. The restrictions on the instrument require W not to affect the distribution of the outcome variable in any given equilibrium, but to have an impact on the equilibrium selection mechanism. Policy interventions are often argued to satisfy these conditions. For instance, in the analysis of the airline market by Ciliberto and Tamer (2009), one may claim that anti-collusion policies do not affect a firm's entry decision on a given market, but do influence the equilibrium selection differentially across regional markets. Henry, Kitamura, and Salanié (2014) provide further examples of excluded instruments used in the development, macroeconomics and international finance literatures.

The previous paragraphs listed some of the applications in which the exclusion restriction on W may be plausibly justified. We now discuss the results that Henry, Kitamura, and Salanié (2014) obtain based on model (2.2). The exclusion and relevance restrictions are in general not sufficient to obtain nonparametric point identification of the mixing weights and component distributions. In particular, Henry, Kitamura, and Salanié (2014) show that the parameters of interest can be expressed as functions of $J(J-1)$ scalar parameters, where J is the number of components and is for now assumed to be known. The identified set then is given by all the values in $\mathbb{R}^{J(J-1)}$ which imply values of the

mixing weights and of the component distributions that satisfy obvious restrictions. In particular, the implied mixing weights must be between 0 and 1 and the implied component cdf's must be non-decreasing, right-continuous and have the correct limits as their argument goes to $\pm\infty$. Moreover, the characterization of the identified set for the case of two-component mixtures provides insight into the role played by variation in the outcome and the instrument. Intuitively, the identified set shrinks as variation in the outcome conditional on the instrument increases and as the effect of the instrument on the distribution of the outcome becomes stronger.

While the model is in general not identified, it is possible to achieve point-identification of certain quantities of interest. In the case of two-component mixtures, it is shown that any linear functional of $F_1(y|x) - F_0(y|x)$ is identified up to scale. This implies, for example, that the ratio $\frac{\mathbb{P}_1\{y>a|x\} - \mathbb{P}_0\{y>a|x\}}{\mathbb{E}_1(y|x) - \mathbb{E}_0(y|x)}$ ⁴ is point-identified for all values of a , provided the denominator is nonzero. In the context of randomized experiments with misclassified treatment, this quantity may be interpreted as a quantile treatment effect relative to the (nonzero) average treatment effect.

The results obtained by Henry, Kitamura, and Salanié (2014) can be applied to a variety of situations thanks to the flexibility allowed by model (2.2). In particular, both the outcome Y and the covariates X can be discrete or continuous. This is in contrast with other studies, where more structure is imposed on the model. For example, in the misclassification literature, Molinari (2008) assumes that the outcome takes a finite number of values, and Bollinger (2006) restricts attention to mismeasured binary regressors to derive bounds on $\mathbb{E}(Y|X)$. Moreover, no constraints are put on the mixing weights, whereas Horowitz and Manski (1995) model contaminated data as a two-component mixture and impose an upper bound on the probability of contamination. Another attractive feature of the results in Henry, Kitamura, and Salanié (2014) is that they rely on testable assumptions. In particular, for the $J = 2$ case, the identified set is defined in terms of estimable bounds. Therefore, the data can be used to estimate these bounds and, in case of crossing, provide evidence to reject the model. More precisely, the researcher can jointly test the exclusion restriction and the assumption that the mixture has two components. The same idea may be used to determine the number of components, to which end Henry, Kitamura, and Salanié (2014) propose a simple iterative procedure. See Henry, Jochmans, and Salanié (2013) for the use of additional restrictions on relative tail behavior of component distributions to achieve point identification and nonparametric estimation under such restrictions. A recent paper by Hohmann and Holtzmann (2013) discusses related results, and in

⁴Here, \mathbb{P}_i and \mathbb{E}_i denote probability and expectation under F_i , respectively.

particular, shows that the Hall-Zhou model in Section 2.1 can be studied within the framework of the Henry-Kitamura-Salanié model in this section.

2.3. Covariates in the component distributions. Section 2.2 focused on the case where identification stems from an exclusion restriction, i.e. from the fact that there is a random variable affecting the mixing weights, but not the component distributions. Now we consider the opposite setup, along the lines of Kitamura (2003). The key identifying assumptions will concern the way in which the component distributions depend on covariates, whereas the mixing weights will be assumed to be independent of covariates. The model is given by

$$(2.4) \quad F(y|x) = \sum_{j=1}^J \lambda_j F_j(y - m_j(x)),$$

where we define

$$(2.5) \quad m_j(x) = \mathbb{E}_j(y|x),$$

for $j = 1, \dots, J$. In equation (2.5), \mathbb{E}_j denotes expectation with respect to F_j .

Further, letting

$$\epsilon_j = y - m_j(x),$$

for $j = 1, \dots, J$, we can interpret model (2.4) as a switching regression model, as follows

$$(2.6) \quad y = m_j(x) + \epsilon_j, \quad \epsilon_j \sim F_{\epsilon|X}^j, \quad \text{with probability } \lambda_j$$

The conditional distribution $F_{\epsilon|X}^j$ in (2.6) is defined by

$$(2.7) \quad F_j(y|x) = F_{\epsilon|X}^j(y - m_j(x)).$$

Models of the form (2.6) have traditionally been tackled in a parametric framework. Identification is achieved by assuming specific functional forms for the m_j functions and the conditional error distributions. Once the problem is reduced to a finite dimension, estimation can usually be carried out via maximum likelihood.

However, such parametric assumptions are very strong and, if incorrect, may lead to inconsistent ML estimates. In particular, potential misspecification of the conditional error distributions is an issue. This is in contrast with standard regression models (i.e. without switching) for which the researcher can specify the regression function parametrically and estimate it without making functional form assumptions on the error distribution.

The risk of misspecification inherent in the parametric approach motivates investigating nonparametric identification of model (2.6). Kitamura (2003) undertakes this task.

First, the paper shows that the standard conditional mean restriction on the errors is not sufficient to achieve identification. This is because mean independence allows multiple - in fact, infinitely many - ways to split the mixture distribution into distinct components. Thus, a stronger condition is required for point-identification. Local independence between the errors and the covariates X turns out to be enough. This condition can be interpreted as a “shift restriction”, in the sense that the entire distribution of the ϵ_j needs to remain invariant in a neighborhood of a given covariates value x_0 . In addition, identification of the regression functions m_j requires these functions not to be parallel in a neighborhood of x_0 . Regularity conditions are also imposed on the tail behavior of the moment generating functions of the error terms.⁵

The assumptions listed above are sufficient to point-identify all of the unknown parameters, namely the regression functions m_j , the conditional error distributions $F_{\epsilon_j|X}$ and the mixing weights λ_j . The result is fully nonparametric as it does not rely on any functional form assumptions. Note, however, that the number of components is assumed to be known and finite. Assuming the researcher knows the number of components is not innocuous. Indeed, misspecification of this feature of the model may invalidate inference, which is why several recent contributions focus on testing for the true number of components. See, for instance, Kasahara and Shimotsu (2012) and the paper by Henry, Kitamura and Salanié discussed in section 2.2.

On the other hand, assuming the mixture is finite appears to be less of a concern. In fact, when the distribution of the latent variable is modelled nonparametrically - such as in Heckman and Singer (1984b)⁶ - the resulting estimates are typically discrete distributions with only a few support points. This suggests that assuming ex ante that the number of components is finite should not be too restrictive in practice.

Further, Kitamura (2003) uses the nonparametric identifiability results above to show identification of finite dimensional parameters defined by semiparametric restrictions. This model qualifies as semiparametric because the conditional distribution of the error terms are not specified and thus

⁵Alternatively, Kitamura (2003) provides a set of assumptions on the tail behavior of the characteristic functions of the error terms’ distributions. These alternative assumptions include relevant cases, such as normal distributions with different variances and distributions for which the moment generating function does not exist.

⁶In these papers, the distribution of the latent variable is estimated via nonparametric maximum likelihood, whereas the component distributions are treated parametrically.

are treated as nuisance parameters. The setting is as in equation (2.4) except that Y is now multi-dimensional and some of the Y_i are allowed to be endogenous. In this framework, the variable X plays the role of the instrument. To our knowledge, a fully nonparametric treatment of endogeneity in mixture models is a largely unexplored area of research.

Finally, Kitamura (2003) proposes an algorithm based on an EM-type method that can be used to estimate semiparametric mixture models and is computationally attractive.

2.4. New results on point-identification of finite mixtures. The previous three sections showed that in general strong conditions are needed to ensure identification of finite mixture models. In section 2.1, the key restrictions were the independence property of each component's multivariate distribution and the fact that attention was focused on two-component mixtures only. In section 2.2, the identifying power came from an exclusion restriction on some of the covariates and in general one could obtain only partial identification in that setup. Finally, the results in Section 2.3 did not rely on exclusion restrictions, but more structure was imposed on the model and full independence between the regressors and the error terms was required.

We now present some new results which apply to general finite mixtures. In particular, we show that under appropriate conditions on the way in which the covariates enter the model, we can obtain point-identification of all unknown parameters. The setup is similar to that of Section 2.2, which was concerned with partial identification. For clarity of exposition, we first focus on the case of two-component mixtures; we then extend the results to finite mixtures with an arbitrary - but known - number of components. The first model we consider is

$$(2.8) \quad K(y|\mathbf{z}, w, x) = \lambda(w, x) F_1(y|\mathbf{z}, x) + [1 - \lambda(w, x)] F_2(y|\mathbf{z}, x),$$

where Y is a scalar or a vector outcome variable, and X , W and $\mathbf{Z} = (Z_1, Z_2)'$ are covariates. Note that the mixing weights are assumed to be a function of W , but not of \mathbf{Z} , a restriction to be relaxed in Section 2.4.4. Conversely, the component cdf's are assumed to depend on \mathbf{Z} , but not on W . Both the component cdf's and the mixing weights are allowed to be a function of X . Further, for simplicity we focus on the case where W , Z_1 and Z_2 are all scalar-valued; If not, any element of (W, Z_1, Z_2) not used for identification can be subsumed into X , whose dimension is left unspecified. Suppose Y takes its values in \mathbb{R}^p . The entire analysis that follows is conditional on $X = x$ and $(Z_1, Z_2) = (z_1, z_2)$. We also make the following assumptions.

Assumption 2.1. *The random variables Y , X , \mathbf{Z} and W have support \mathcal{Y} , \mathcal{X} , \mathcal{Z} and \mathcal{W} , respectively, and \mathbf{Z} and W are continuously distributed.*

Note that \mathcal{Y} can be a discrete set, so that the result here applies to a mixture of discrete choice models, or models where the outcome vector has both continuous and discrete elements. For the rest of the paper we use the convention that $\mathcal{X}\mathcal{Y}$ denotes the joint support of (X, Y) , $\mathcal{X}\mathcal{Y}\mathcal{Z}$ the joint support of (Y, X, \mathbf{Z}) , and so on.

Assumption 2.2. *The functions $\lambda(\cdot, x)$ and $F_i(y|\cdot, x)$ for $i = 1, 2$, are differentiable for every $(y, x) \in \mathbb{R}^p \times \mathcal{X}$.*

Assumptions 2.1 and 2.2 are not crucial and will be relaxed later. On the contrary, the following conditions are key to our identification strategy.

Assumption 2.3. *For each $(x, \mathbf{z}) \in \mathcal{X}\mathcal{Z}$ there exists $y^* \in \mathbb{R}^p$ such that*

$$\begin{aligned} \frac{\partial}{\partial z_1} F_1(y^*|\mathbf{z}, x) &\neq 0 & \frac{\partial}{\partial z_2} F_1(y^*|\mathbf{z}, x) &= 0 \\ \frac{\partial}{\partial z_1} F_2(y^*|\mathbf{z}, x) &= 0 & \frac{\partial}{\partial z_2} F_2(y^*|\mathbf{z}, x) &\neq 0. \end{aligned}$$

Let $\mathcal{W}|(x, \mathbf{z})$ be the support of the conditional distribution of W given $(X, \mathbf{z}) = (x, \mathbf{z})$.

Assumption 2.4. *For each $(x, \mathbf{z}) \in \mathcal{X}\mathcal{Z}$ there exists $w^* \in \mathcal{W}|(x, \mathbf{z})$ such that*

$$\frac{\partial}{\partial w} \lambda(w^*, x) \neq 0.$$

Assumption 2.3 requires that, for some value in the support of Y , each component conditional cdf be affected by one of the covariates in the \mathbf{Z} vector, but not by the other one. In section 2.4.3, we relax this condition and consider the case where it holds for only one of the component cdf's.

Assumption 2.4 is simply requiring the covariate W to enter model (2.8). In fact, if Assumption 2.4 did not hold, then W would not affect neither the mixing weights nor the component cdf's and thus it would not even enter the conditioning on the left-hand side of (2.8). Further comments on the identifying assumptions and how they can be substantiated in economic applications are provided in Remarks (2.2) and (2.3) below.

We are now ready to state the first new result.

Theorem 2.1. *Consider model (2.8) and let Assumptions 2.1-2.4 hold. Then, the component cdf's $F_i(y|\mathbf{z}, x)$, for $i = 1, 2$, are nonparametrically identified for every $(y, x, \mathbf{z}) \in \mathbb{R}^p \times \mathcal{X}\mathcal{Z}$ and $\lambda(w, x)$ is nonparametrically identified for every $(x, w) \in \mathcal{X}\mathcal{W}_x^*$, where $\mathcal{X}\mathcal{W}_x^* \equiv \{(x, w) \in \mathcal{X}\mathcal{W} \text{ such that } \frac{\partial}{\partial w} \lambda(w, x) \neq 0\}$.*

Moreover, the results are constructive. Specifically, for every pair (z, x) at which Assumptions 2.3 and 2.4 are satisfied,

$$F_1(y|\mathbf{z}, x) = K(y|\mathbf{z}, w^*, x) - \frac{\frac{\partial}{\partial z_2} K(y^*|\mathbf{z}, w^*, x)}{\frac{\partial^2}{\partial w \partial z_2} K(y^*|\mathbf{z}, w^*, x)} \frac{\partial}{\partial w} K(y|\mathbf{z}, w^*, x)$$

$$F_2(y|\mathbf{z}, x) = K(y|\mathbf{z}, w^*, x) - \frac{\frac{\partial}{\partial z_1} K(y^*|\mathbf{z}, w^*, x)}{\frac{\partial^2}{\partial w \partial z_1} K(y^*|\mathbf{z}, w^*, x)} \frac{\partial}{\partial w} K(y|\mathbf{z}, w^*, x)$$

Further, for every pair (\mathbf{z}, x) for which Assumptions 2.3 and 2.4 are satisfied and for every $w \in \mathcal{W}_x^*$,

$$\lambda(w, x) = \frac{\zeta(w, x)}{1 + \zeta(w, x)}$$

$$\text{where } \zeta(w, x) \equiv -\frac{\frac{\partial}{\partial z_1} K(y^*|\mathbf{z}, w, x) \frac{\partial^2}{\partial w \partial z_2} K(y^*|\mathbf{z}, w, x)}{\frac{\partial}{\partial z_2} K(y^*|\mathbf{z}, w, x) \frac{\partial^2}{\partial w \partial z_1} K(y^*|\mathbf{z}, w, x)}.$$

Proof. Since the entire proof is conditional on $X = x$, we drop this conditioning for notational convenience.

Equation (2.8) can be re-written as

$$(2.9) \quad K(y|\mathbf{z}, w) = F_2(y|\mathbf{z}) + \lambda(w) [F_1(y|\mathbf{z}) - F_2(y|\mathbf{z})]$$

Deriving Equation (2.9) with respect to w , we obtain

$$(2.10) \quad \frac{\partial}{\partial w} K(y|\mathbf{z}, w) = \frac{\partial}{\partial w} \lambda(w) [F_1(y|\mathbf{z}) - F_2(y|\mathbf{z})]$$

for every $y \in \mathbb{R}^p$.

Moreover, if $y = y^*$, we can use Assumption 2.3 to get

$$(2.11) \quad \frac{\partial}{\partial z_1} K(y^*|\mathbf{z}, w) = \lambda(w) \frac{\partial}{\partial z_1} F_1(y^*|\mathbf{z})$$

and

$$(2.12) \quad \frac{\partial}{\partial z_2} K(y^*|\mathbf{z}, w) = [1 - \lambda(w)] \frac{\partial}{\partial z_2} F_2(y^*|\mathbf{z}).$$

Deriving (2.11) and (2.12) with respect to w , we obtain, respectively,

$$(2.13) \quad \frac{\partial^2}{\partial w \partial z_1} K(y^*|\mathbf{z}, w) = \frac{\partial}{\partial w} \lambda(w) \frac{\partial}{\partial z_1} F_1(y^*|\mathbf{z})$$

and

$$(2.14) \quad \frac{\partial^2}{\partial w \partial z_2} K(y^*|\mathbf{z}, w) = -\frac{\partial}{\partial w} \lambda(w) \frac{\partial}{\partial z_2} F_2(y^*|\mathbf{z}).$$

By (2.11) and (2.12),

$$(2.15) \quad \frac{\frac{\partial}{\partial z_1} K(y^*|\mathbf{z}, w)}{\frac{\partial}{\partial z_2} K(y^*|\mathbf{z}, w)} = \frac{\lambda(w)}{1 - \lambda(w)} \cdot \frac{\frac{\partial}{\partial z_1} F_1(y^*|\mathbf{z})}{\frac{\partial}{\partial z_2} F_2(y^*|\mathbf{z})}$$

and by (2.13) and (2.14), for every $w \in \mathcal{W}_x^*$,

$$(2.16) \quad \frac{\frac{\partial^2}{\partial w \partial z_1} K(y^*|\mathbf{z}, w)}{\frac{\partial^2}{\partial w \partial z_2} K(y^*|\mathbf{z}, w)} = - \frac{\frac{\partial}{\partial z_1} F_1(y^*|\mathbf{z})}{\frac{\partial}{\partial z_2} F_2(y^*|\mathbf{z})}.$$

Combining (2.15) and (2.16) yields

$$(2.17) \quad \frac{\lambda(w)}{1 - \lambda(w)} = - \frac{\frac{\partial}{\partial z_1} K(y^*|\mathbf{z}, w)}{\frac{\partial}{\partial z_2} K(y^*|\mathbf{z}, w)} \cdot \frac{\frac{\partial^2}{\partial w \partial z_2} K(y^*|\mathbf{z}, w)}{\frac{\partial^2}{\partial w \partial z_1} K(y^*|\mathbf{z}, w)} \equiv \zeta(w),$$

for every $w \in \mathcal{W}_x^*$. Therefore,

$$(2.18) \quad \lambda(w) = \frac{\zeta(w)}{1 + \zeta(w)}$$

for all $w \in \mathcal{W}_x^*$. Since ζ is only a function of observables, Equation (2.18) shows that the mixing weight $\lambda(w)$ is identified for all $w \in \mathcal{W}_x^*$.

We now turn to showing identification of the component cdf's. The two following equations will be useful:

$$(2.19) \quad \frac{\lambda(w)}{\frac{\partial}{\partial w} \lambda(w)} = \frac{\frac{\partial}{\partial z_1} K(y^*|\mathbf{z}, w)}{\frac{\partial^2}{\partial w \partial z_1} K(y^*|\mathbf{z}, w)}$$

and

$$(2.20) \quad \frac{1 - \lambda(w)}{\frac{\partial}{\partial w} \lambda(w)} = \frac{\frac{\partial}{\partial z_2} K(y^*|\mathbf{z}, w)}{\frac{\partial^2}{\partial w \partial z_2} K(y^*|\mathbf{z}, w)}.$$

Equation (2.19) follows from (2.11) and (2.13), while Equation (2.20) follows from (2.12) and (2.14).

Now let $w^* \in \mathcal{W}_x^*$ and let y be an arbitrary vector in \mathbb{R}^p . By (2.9) and (2.10),

$$(2.21) \quad K(y|\mathbf{z}, w^*) = F_2(y|\mathbf{z}) + \frac{\lambda(w^*)}{\frac{\partial}{\partial w} \lambda(w^*)} \cdot \frac{\partial}{\partial w} F(y|\mathbf{z}, w^*),$$

which, by (2.19), implies

$$(2.22) \quad F_2(y|\mathbf{z}) = K(y|\mathbf{z}, w^*) - \frac{\frac{\partial}{\partial z_1} K(y^*|\mathbf{z}, w^*)}{\frac{\partial^2}{\partial w \partial z_1} K(y^*|\mathbf{z}, w^*)} \cdot \frac{\partial}{\partial w} K(y|\mathbf{z}, w^*).$$

Since the value y was arbitrary, Equation (2.22) shows that F_2 is identified for every $y \in \mathbb{R}^p$, given $Z = z$.

To recover F_1 as a function of observables, note that

$$\begin{aligned}
 F_1(y|\mathbf{z}) &= \frac{\frac{\partial}{\partial w} K(y|\mathbf{z}, w^*)}{\frac{\partial}{\partial w} \lambda(w^*)} + F_2(y|\mathbf{z}) \\
 (2.23) \quad &= \frac{\frac{\partial}{\partial w} K(y|\mathbf{z}, w^*)}{\frac{\partial}{\partial w} \lambda(w^*)} + K(y|\mathbf{z}, w^*) - \frac{\lambda(w^*)}{\frac{\partial}{\partial w} \lambda(w^*)} \cdot \frac{\partial}{\partial w} K(y|\mathbf{z}, w^*) \\
 &= K(y|\mathbf{z}, w^*) + \frac{1 - \lambda(w^*)}{\frac{\partial}{\partial w} \lambda(w^*)} \cdot \frac{\partial}{\partial w} K(y|\mathbf{z}, w^*) \\
 &= K(y|\mathbf{z}, w^*) - \frac{\frac{\partial}{\partial z_2} K(y^*|\mathbf{z}, w^*)}{\frac{\partial^2}{\partial w \partial z_2} K(y^*|\mathbf{z}, w^*)} \cdot \frac{\partial}{\partial w} K(y|\mathbf{z}, w^*),
 \end{aligned}$$

where the first equality follows from (2.10), the second from (2.9) and (2.10), the third equality is trivial and the last follows from (2.20). \square

A few remarks are in order.

Remark 2.1. The identification strategy used to prove Theorem 2.1 does not impose any support assumptions on the covariates. This is in contrast with several contributions in the literature, which rely on “identification at infinity”-type restrictions (see, e.g., Tamer (2003), for the case of multiple equilibria in discrete games). It is a well-documented fact that identification at infinity of a parameter tends to be associated with slow rates of convergence of its estimators. See, for instance, Chamberlain (1986) and Andrews and Schafgans (1998). Therefore, our constructive identification results may be used to obtain estimators with better finite-sample properties than existing estimators.

Remark 2.2. The variable W is required to affect the mixing weights (Assumption 2.4), but not the component distributions. This is the same exclusion restriction that was imposed in Henry, Kitamura, and Salanié (2014); see section 2.2 for a discussion of a range of economic applications where this condition may be plausibly justified. Note that, Henry, Kitamura, and Salanié (2014) only relied on the exclusion restriction on W and obtained a partial identification result. Here, we achieve point-identification thanks to the additional assumptions on the Z covariates.

Moreover, Assumption 2.4 is testable. Using, e.g., Equation (2.13) evaluated at a candidate value of W , one can test the hypothesis that $\frac{\partial}{\partial w} \lambda(w)$ is zero by looking at $\frac{\partial^2}{\partial w \partial z_1} K(y^*|\mathbf{z}, w)$, which is estimable from the data.

Remark 2.3. A sufficient condition for Assumption 2.3 is that, for a given value of X , $F_1(y|\mathbf{z}, x) = F_1(y|\mathbf{z}_1, x)$ and $F_2(y|\mathbf{z}, x) = F_2(y|\mathbf{z}_2, x)$. This constraint is similar in spirit to exclusion restrictions

that are often imposed in discrete games of complete information in presence of multiple equilibria. For example, Tamer (2003) and Bajari, Hong, and Ryan (2010) consider entry games where the exclusion restriction may be satisfied by variables that enter a firm's profit function, but not those of its competitors.

Remark 2.4. Assumption 2.3 imposes restrictions on the cdf's of the component (conditional) distributions. These conditions may be reformulated using any other linear functional of the cdf's, such as the characteristic functions evaluated at a given point or the moments of a given order.

Remark 2.5. Assumption 2.3 is essential for identification of the model. However, one can relax this restriction and still identify some of the parameters of interest. In section 2.4.3, we consider the case where Assumption 2.3 holds for only one of the component cdf's and we show that the corresponding cdf is point-identified under an additional assumption.

Remark 2.6. Assumption 2.1 may be relaxed with minor changes in the statement of the theorem. Specifically, section 2.4.2 discusses the case where \mathbf{Z} and W have a discrete distribution.

2.4.1. *Mixtures with arbitrary number of components.* The results in Theorem 2.1 extend to the case of mixtures with an arbitrary, known number of components. Consider the model:

$$(2.24) \quad K(y|\mathbf{z}, w, x) = \sum_{j=1}^J \lambda_j(w, x) F_j(y|\mathbf{z}, x),$$

which is useful to re-write as

$$(2.25) \quad K(y|\mathbf{z}, w, x) = F_J(y|\mathbf{z}, x) + \sum_{j=1}^{J-1} \lambda_j(w, x) [F_j(y|\mathbf{z}, x) - F_J(y|\mathbf{z}, x)].$$

In the two-component case, we assumed that \mathbf{Z} was a bivariate random variable. In order to achieve identification, we now need \mathbf{Z} to be J -valued.

As in the bivariate mixture case, we make some assumptions that are not essential, but simplify the analysis.

Assumption 2.5. *The random variables Y , X , \mathbf{Z} and W are continuously distributed, with support \mathcal{Y} , \mathcal{X} , \mathcal{Z} and \mathcal{W} , respectively.*

Assumption 2.6. *The functions $\lambda_j(\cdot, x)$ and $F_j(y|\cdot, x)$ for $j = 1, \dots, J$, are differentiable for every $y \in \mathbb{R}^p$.*

As in the two-component case, the key identifying assumptions concern the way in which \mathbf{Z} and W affect the component cdf's and the mixing weights, respectively.

Concerning the component distributions, the generalization of Assumption 2.3 is given by the following.

Assumption 2.7. *There exists $y^* \in \mathbb{R}^p$ such that, for $j = 1, \dots, J$,*

$$\frac{\partial}{\partial z_j} F_j(y^* | \mathbf{z}, x) \neq 0$$

and

$$\frac{\partial}{\partial z_i} F_j(y^* | \mathbf{z}, x) = 0$$

for all $i \neq j$.

As for the restrictions imposed on the mixing weights, we first need some definitions.

For a vector $\mathbf{w} = (w_1, \dots, w_{J-1})$ in \mathcal{W}^{J-1} , define $\Lambda(\mathbf{w})$ as the $(J-1)$ -by- $(J-1)$ matrix with ij -th element $\lambda_i(w_j, x)$. In other words, each column of $\Lambda(\mathbf{w})$ contains the $J-1$ independent mixing weights corresponding to a given value of W .

Similarly, define $D_w \Lambda(\mathbf{w})$ as the $(J-1)$ -by- $(J-1)$ matrix with ij -th element $\frac{\partial}{\partial w} \lambda_i(w_j, x)$.

Given the definitions above, the generalization of Assumption 2.4 takes the following form.

Assumption 2.8. *There exists a vector $\mathbf{w}^* = (w_1^*, \dots, w_{J-1}^*)$ in \mathcal{W}^{J-1} such that $\Lambda(\mathbf{w}^*)$ and $D_w \Lambda(\mathbf{w}^*)$ are invertible.*

In this setup, the model is point-identified, as formalized below.

Theorem 2.2. *Consider model (2.24) and let Assumptions 2.5-2.8 hold. Then, the component cdf's $F_j(y | \mathbf{z}, x)$, for $j = 1, \dots, J$, are identified for every $y \in \mathbb{R}^p$. Further, the matrix $\Lambda(\mathbf{w})$ is identified for every $\mathbf{w} \in \mathcal{W}_x^{J-1}$, where $\mathcal{W}_x^{J-1} \equiv \{w \in \mathcal{W}^{J-1} \text{ such that } \Lambda(\mathbf{w}) \text{ and } D_w \Lambda(\mathbf{w}) \text{ are invertible}\}$.*

Proof. See the Appendix. □

2.4.2. Discrete covariates. A second extension of the basic results covers the case of discrete covariates.

For a function $g : \mathbb{R}^d \rightarrow \mathbb{R}$, let

$$(2.26) \quad \Delta_{x'_i, x_i} g(x_1, \dots, x_d) \equiv g(x_1, \dots, x'_i, \dots, x_d) - g(x_1, \dots, x_i, \dots, x_d)$$

denote the difference with respect to the i -th argument. We work under the following assumptions.

Assumption 2.9. *The variables Z and W have discrete distributions with support \mathcal{Z} and \mathcal{W} , respectively.*

Assumption 2.10. *Given $\mathbf{z} \in \mathcal{Z}$, there exists $y^* \in \mathbb{R}^p$ and $\mathbf{z}' \in \mathcal{Z}$ such that*

$$\begin{aligned} \Delta_{z'_1, z_1} F_1(y^* | \mathbf{z}, x) &\neq 0 & \Delta_{z'_2, z_2} F_1(y^* | \mathbf{z}, x) &= 0 \\ \Delta_{z'_1, z_1} F_2(y^* | \mathbf{z}, x) &= 0 & \Delta_{z'_2, z_2} F_2(y^* | \mathbf{z}, x) &\neq 0 \end{aligned}$$

Assumption 2.11. *There exist $w^*, w^{*'} \in \mathcal{W}$ such that*

$$\Delta_{w^{*'}, w^*} \lambda(w^*, x) \neq 0$$

Theorem 2.3. *Consider model (2.8) and let Assumptions 2.9, 2.10 and 2.11 hold. Then, the component cdf's $F_i(y | \mathbf{z}, x)$, for $i = 1, 2$, are nonparametrically identified for every $y \in \mathbb{R}^p$ and $\lambda(w, x)$ is nonparametrically identified for every $w \in \mathcal{W}_{x,d}^*$, where $\mathcal{W}_{x,d}^* \equiv \{w \in \mathcal{W} \text{ such that } \Delta_{w', w} \lambda(w^*, x) \neq 0 \text{ for some } w' \in \mathcal{W}\}$. Moreover, the results are constructive. Specifically, for every pair (z, x) at which Assumptions 2.10 and 2.11 are satisfied,*

$$\begin{aligned} F_1(y | \mathbf{z}, x) &= K(y | \mathbf{z}, w^*, x) - \frac{\Delta_{z'_2, z_2} K(y^* | \mathbf{z}, w^*, x)}{\Delta_{w^{*'}, w^*} \Delta_{z'_2, z_2} K(y^* | \mathbf{z}, w^*, x)} \Delta_{w^{*'}, w^*} K(y | \mathbf{z}, w^*, x) \\ F_2(y | \mathbf{z}, x) &= K(y | \mathbf{z}, w^*, x) - \frac{\Delta_{z'_1, z_1} K(y^* | \mathbf{z}, w^*, x)}{\Delta_{w^{*'}, w^*} \Delta_{z'_1, z_1} K(y^* | \mathbf{z}, w^*, x)} \Delta_{w^{*'}, w^*} K(y | \mathbf{z}, w^*, x) \end{aligned}$$

Further, for every pair (z, x) for which Assumptions 2.10 and 2.11 are satisfied and for every $w \in \mathcal{W}_{x,d}^*$,

$$\lambda(w, x) = \frac{\zeta(w, x)}{1 + \zeta(w, x)}$$

$$\text{where } \zeta(w, x) \equiv -\frac{\Delta_{z'_1, z_1} K(y^* | \mathbf{z}, w, x) \Delta_{w^{*'}, w^*} \Delta_{z'_2, z_2} K(y^* | \mathbf{z}, w, x)}{\Delta_{z'_2, z_2} K(y^* | \mathbf{z}, w, x) \Delta_{w^{*'}, w^*} \Delta_{z'_1, z_1} K(y^* | \mathbf{z}, w, x)}.$$

Proof. The proof proceeds along the same lines as the proof of Theorem 2.1 with the difference operator defined in (2.26) replacing the differential operator. \square

Remark 2.7. Even when \mathbf{Z} and W are continuously distributed, the above result is potentially useful in implementing nonparametric estimation, as it allows the econometrician to avoid nonparametric estimation of derivatives, which slows down convergence rate.

2.4.3. Relaxing Assumption 2.3. The identification strategy for the baseline model crucially relies on Assumption 2.3. We now try to relax this restriction. In particular, we focus on the case in which Assumption 2.3 holds for one of the component cdf's, but may be violated for the other one. For fixed values of X and \mathbf{Z} , the model is again

$$K(y|\mathbf{z}, w, x) = \lambda(w, x) F_1(y|\mathbf{z}, x) + [1 - \lambda(w, x)] F_2(y|\mathbf{z}, x).$$

and we impose the following milder restriction:

Assumption 2.12. *There exists $y^* \in \mathbb{R}^p$ such that*

$$\frac{\partial}{\partial z_1} F_1(y^*|\mathbf{z}, x) \neq 0 \quad \frac{\partial}{\partial z_2} F_1(y^*|\mathbf{z}, x) = 0$$

No conditions are imposed on F_2 except for the standard requirements that define a cdf. In this setting, the model is not identified in general. However, under an additional restriction, it is possible to point-identify the component cdf on which the conditions are imposed. The following formalizes this idea.

Assumption 2.13. *The cross-derivative $\frac{\partial^2}{\partial z_1 \partial z_2} F_1(y^*|\mathbf{z}, x)$ exists and is nonzero.*

Theorem 2.4. *Consider model (2.8) and let Assumptions 2.1, 2.2, 2.4, 2.12 and 2.13 hold. Then, $F_1(y|\mathbf{z}, x)$ is nonparametrically identified for every $y \in \mathbb{R}^p$.*

Proof. See the Appendix. □

2.4.4. Generalizing the mixture weight function λ . In this subsection we consider a model of the form

$$(2.27) \quad K(y|\mathbf{z}, w, x) = \lambda(\mathbf{z}, w, x) F_1(y|\mathbf{z}, x) + [1 - \lambda(\mathbf{z}, w, x)] F_2(y|\mathbf{z}, x).$$

As before, $K(\cdot|\cdot, \cdot, \cdot)$ is the conditional distribution function of Y given $(\mathbf{Z}, W, X) = (\mathbf{z}, w, x)$, and available from the joint distribution of the vector valued random variable (Y, \mathbf{Z}, W, X) , which is observable. The component distributions (CDFs) $F_1(\cdot|\cdot, \cdot)$, $F_2(\cdot|\cdot, \cdot)$ and the mixing weight $\lambda(\cdot, \cdot, \cdot)$ are unknown function to be identified from the knowledge of $K(\cdot|\cdot, \cdot, \cdot)$. We maintain Assumption 2.1, thus \mathbf{Z} and w are continuously distributed.

Assumption 2.14. *The functions $\lambda(\cdot, \cdot, x)$ and $F_i(y|\cdot, \cdot, x)$ for $i = 1, 2$, are differentiable for every $y \in \mathbb{R}^p$ and $x \in \mathcal{X}$.*

As before, it is possible to drop the continuity/differentiability assumptions; see the discussion in Remark 2.9. The following additional assumptions are made. Recall \mathcal{Y} denotes the support of Y .

Assumption 2.15. *For each $(x, w, \mathbf{z}) \in \mathcal{XWZ}$, there exist non-empty, non-singleton sets $\mathcal{Y}_1(x, w, \mathbf{z})$ and $\mathcal{Y}_2(x, w, \mathbf{z})$ in \mathcal{Y} such that*

- (i) $\frac{\partial}{\partial z_2} F_1(\cdot | \mathbf{z}, x) = 0$ and $\frac{\partial}{\partial z_1} F_1(\cdot | \mathbf{z}, x) \neq 0$ on $\mathcal{Y}_1(x, w, \mathbf{z})$;
- (ii) $\frac{\partial}{\partial z_1} F_2(\cdot | \mathbf{z}, x) = 0$ and $\frac{\partial}{\partial z_2} F_2(\cdot | \mathbf{z}, x) \neq 0$ on $\mathcal{Y}_2(x, w, \mathbf{z})$;
- (iii) the functions $F_1(\cdot | \mathbf{z}, x) - F_2(\cdot, \mathbf{z}, x)$ and $\frac{\partial}{\partial z_1} F_1(\cdot | \mathbf{z}, x)$ are linearly independent on $\mathcal{Y}_1(x, w, \mathbf{z})$;
- (iv) the functions $F_1(\cdot | \mathbf{z}, x) - F_2(\cdot, \mathbf{z}, x)$ and $\frac{\partial}{\partial z_2} F_2(\cdot | \mathbf{z}, x)$ are linearly independent on $\mathcal{Y}_2(x, w, \mathbf{z})$;

□

As before, let $\mathcal{W}|(x, \mathbf{z})$ be the support of the conditional distribution of W given $(X, \mathbf{z}) = (x, \mathbf{z})$.

Assumption 2.16. *For each $(x, \mathbf{z}) \in \mathcal{XZ}$ there exists $w^* \in \mathcal{W}|(x, \mathbf{z})$ such that $\frac{\partial}{\partial w} \lambda(\mathbf{z}, w^*, x) \neq 0$.*

□

The new model (2.27) differs from the original model (2.8) in that it allows the covariate \mathbf{Z} to enter into the mixture weight λ in a completely free manner. Thus it can be views as a generalization of the model (2.8). Strictly speaking, however, the assumptions made for the model (2.8) and the current assumption (i.e. Assumptions 2.15 and 2.16) are non-nested, reflecting a major difference in the identification strategies for the two models. For the current theorem we assume that there are more than one value of y for which Assumptions 2.15(i)(ii) hold. On the other hand Assumption 2.3 holding at one value of y suffices for Theorem 2.1.

Once again, the main application would be a model with unobserved heterogeneity or a model with multiple equilibria. For example, take the generic demand model described in HKS Section 1.2.3. If consumer preferences have two types, say a “high” type and a “low” type, each corresponding to F_1 and F_2 , then as far as there exists a variable (e.g. geographical variable) that is correlated with the types but does not enter the utility functions, then such a variable works as w . Moreover, if there are covariates Z_1 and Z_2 such that each only affects each type of consumers, then they satisfy Assumption 2.15. The previous result in Theorem 2.1 was proved under the formulation (2.8), which, applied to the current example, would mean that individual types and the covariates $\mathbf{Z} = (Z_1, Z_2)$ are uncorrelated; this seems a strong restriction. The current theorem removes this restriction, by allowing for a fully unrestricted and nonparametric dependence of the mixing weights λ on \mathbf{Z} . Alternatively, suppose we have a model with two equilibria at hand. If there exists covariates Z_1 and Z_2 such that Z_1 only

affects one of the two equilibrium distributions and Z_2 the other, the model falls into the category discussed here. In this case the function $\lambda(\cdot, \cdot, \cdot)$ determines the equilibrium selection mechanism, and the specification in this section allows the equilibrium selection mechanism to depend on the covariate \mathbf{Z} in a fully general, nonparametric manner.

The following theorem shows that the component distribution functions $F_i(\cdot|\cdot, \cdot)$ and the mixture weight function $\lambda(\cdot, \cdot, \cdot)$ are nonparametrically identified. Define

$$\begin{aligned} & \mathcal{K}_1(x, w, \mathbf{z}) \\ := & \left\{ (y_a, y_b) \in \mathbb{R}^{2p} : \frac{\partial}{\partial w} K(y_a|\mathbf{z}, w, x) \frac{\partial^2}{\partial w \partial z_1} K(y_b|\mathbf{z}, w, x) - \frac{\partial^2}{\partial w \partial z_1} K(y_a|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_b|\mathbf{z}, w, x) \neq 0 \right\} \end{aligned}$$

and

$$\begin{aligned} & \mathcal{K}_2(x, w, \mathbf{z}) \\ := & \left\{ (y_c, y_d) \in \mathbb{R}^{2p} : -\frac{\partial}{\partial w} K(y_c|\mathbf{z}, w, x) \frac{\partial^2}{\partial w \partial z_2} K(y_d|\mathbf{z}, w, x) + \frac{\partial^2}{\partial w \partial z_2} K(y_c|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_d|\mathbf{z}, w, x) \neq 0 \right\}. \end{aligned}$$

It is straightforward to see that the sets $\mathcal{K}_1(x, w, \mathbf{z}) \cap (\mathcal{Y}_1(x, w, \mathbf{z}) \times \mathcal{Y}_1(x, w, \mathbf{z}))$ and $\mathcal{K}_2(x, w, \mathbf{z}) \cap (\mathcal{Y}_2(x, w, \mathbf{z}) \times \mathcal{Y}_2(x, w, \mathbf{z}))$ are non-empty under Assumptions 2.15 and 2.16. Define

$$\mathcal{XW}_x^* \equiv \left\{ (x, w) \in \mathcal{XW} : \frac{\partial}{\partial w} \lambda(w, x) \neq 0 \right\}$$

and

$$\mathcal{XW}_x^* \mathcal{Z} \equiv \left\{ (x, w, \mathbf{z}) \in \mathcal{XWZ} : \frac{\partial}{\partial w} \lambda(w, x, \mathbf{z}) \neq 0 \right\}.$$

Theorem 2.5. *Consider model (2.27) and let Assumptions 2.1, 2.15 and 2.16 hold. Then, the component cdf's $F_i(y|\mathbf{z}, x)$, for $i = 1, 2$, are nonparametrically identified for every $(y, \mathbf{z}, x) \in \mathbb{R}^p \times \mathcal{XZ}$ and $\lambda(w, x)$ is nonparametrically identified for every $(x, w) \in \mathcal{XW}_x^*$. Moreover, the results are constructive. That is, for every $(y, w, x, \mathbf{z}) \in \mathbb{R}^p \times \mathcal{XW}_x^* \mathcal{Z}$,*

$$\begin{aligned} & F_1(y|\mathbf{z}, x) \\ = & K(y|\mathbf{z}, w, x) - \frac{-\frac{\partial}{\partial z_2} K(y_c|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_d|\mathbf{z}, w, x) + \frac{\partial}{\partial z_2} K(y_d|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_c|\mathbf{z}, w, x)}{-\frac{\partial}{\partial w} K(y_c|\mathbf{z}, w, x) \frac{\partial^2}{\partial w \partial z_2} K(y_d|\mathbf{z}, w, x) + \frac{\partial^2}{\partial w \partial z_2} K(y_c|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_d|\mathbf{z}, w, x)} \frac{\partial}{\partial w} K(y|\mathbf{z}, w, x), \end{aligned}$$

$$\begin{aligned} & F_2(y|\mathbf{z}) \\ = & K(y|\mathbf{z}, w, x) - \frac{-\frac{\partial}{\partial z_1} K(y_a|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_b|\mathbf{z}, w, x) + \frac{\partial}{\partial z_1} K(y_b|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_a|\mathbf{z}, w, x)}{\frac{\partial}{\partial w} K(y_a|\mathbf{z}, w, x) \frac{\partial^2}{\partial w \partial z_1} K(y_b|\mathbf{z}, w, x) - \frac{\partial^2}{\partial w \partial z_1} K(y_a|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_b|\mathbf{z}, w, x)} \frac{\partial}{\partial w} K(y|\mathbf{z}, w, x), \end{aligned}$$

and

$$\lambda(\mathbf{z}, w, x) = \frac{\zeta(\mathbf{z}, w, x)}{\zeta(\mathbf{z}, w, x) + 1}$$

where

$$\begin{aligned} \zeta(\mathbf{z}, w, x) &:= \frac{\left(-\frac{\partial}{\partial z_1} K(y_a|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_b|\mathbf{z}, w, x) + \frac{\partial}{\partial z_1} K(y_b|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_a|\mathbf{z}, w, x)\right)}{\left(-\frac{\partial}{\partial z_2} K(y_c|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_d|\mathbf{z}, w, x) + \frac{\partial}{\partial z_2} K(y_d|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_c|\mathbf{z}, w, x)\right)} \\ &\quad \times \frac{\left(-\frac{\partial}{\partial w} K(y_c|\mathbf{z}, w, x) \frac{\partial^2}{\partial w \partial z_2} K(y_d|\mathbf{z}, w, x) + \frac{\partial^2}{\partial w \partial z_2} K(y_c|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_d|\mathbf{z}, w, x)\right)}{\left(\frac{\partial}{\partial w} K(y_a|\mathbf{z}, w, x) \frac{\partial^2}{\partial w \partial z_1} K(y_b|\mathbf{z}, w, x) - \frac{\partial^2}{\partial w \partial z_1} K(y_a|\mathbf{z}, w, x) \frac{\partial}{\partial w} K(y_b|\mathbf{z}, w, x)\right)} \end{aligned}$$

if $(y_a, y_b) \in \mathcal{K}_1(x, w, \mathbf{z})$, $(y_c, y_d) \in \mathcal{K}_2(x, w, \mathbf{z})$, $y_a, y_b \in \mathcal{Y}_1(x, w, \mathbf{z})$ and $y_c, y_d \in \mathcal{Y}_2(x, w, \mathbf{z})$.

Remark 2.8. Once again, the identification result is constructive. Note that the sets $\mathcal{K}_1(x, w, \mathbf{z})$ and $\mathcal{K}_2(x, w, \mathbf{z})$ are identifiable from the knowledge of the joint distribution of (y, \mathbf{z}, w, x) .

Remark 2.9. Analogously to our treatment of discrete covariates in Theorem 2.3, it is possible to replace differential operators with difference operators to obtain nonparametric identification of the model (2.27) when \mathbf{Z} and W are discrete random variables. Again, such a result can be useful in nonparametric estimation even when \mathbf{Z} and W are continuously distributed: see the discussion in Remark 2.7.

Proof. Fix $(x, w, \mathbf{z}) \in \mathcal{XW}_x^* \mathcal{Z}$. As in the proof of Theorem 2.1, all the arguments remain valid conditional on $X = x$, and we drop x from the notation. Letting $F_\Delta(y^*|\mathbf{z}) := F_1(y|\mathbf{z}) - F_2(y|\mathbf{z})$, (2.27) becomes

$$(2.28) \quad K(y|\mathbf{z}, w) = F_2(y|\mathbf{z}) + \lambda(\mathbf{z}, w) F_\Delta(y^*|\mathbf{z})$$

Differentiating (2.28) with respect to w

$$(2.29) \quad \frac{\partial}{\partial w} K(y|\mathbf{z}, w) = \frac{\partial}{\partial w} \lambda(\mathbf{z}, w) \cdot F_\Delta(y^*|\mathbf{z})$$

holds for every $y \in \mathbb{R}^p$. Similarly, differentiating (2.27) with z_1 and z_2 at $y_1^* \in \mathcal{Y}_1(x, w, \mathbf{z})$ and $y_2^* \in \mathcal{Y}_2(x, w, \mathbf{z})$, respectively, by Assumption 2.15 we get

$$(2.30) \quad \frac{\partial}{\partial z_1} K(y_1^*|\mathbf{z}, w) = \frac{\partial}{\partial z_1} \lambda(\mathbf{z}, w) \cdot F_\Delta(y_1^*|\mathbf{z}) + \lambda(\mathbf{z}, w) \frac{\partial}{\partial z_1} F_1(y_1^*|\mathbf{z})$$

and

$$(2.31) \quad \frac{\partial}{\partial z_2} K(y_2^*|\mathbf{z}, w) = \frac{\partial}{\partial z_2} \lambda(\mathbf{z}, w) \cdot F_\Delta(y_2^*|\mathbf{z}) + [1 - \lambda(\mathbf{z}, w)] \frac{\partial}{\partial z_2} F_2(y_2^*|\mathbf{z}).$$

Also,

$$(2.32) \quad \frac{\partial^2}{\partial w \partial z_1} K(y_1^* | \mathbf{z}, w) = \frac{\partial^2}{\partial w \partial z_1} \lambda(\mathbf{z}, w) \cdot F_\Delta(y_1^* | \mathbf{z}) + \frac{\partial}{\partial w} \lambda(\mathbf{z}, w) \frac{\partial}{\partial z_1} F_1(y_1^* | \mathbf{z})$$

and

$$(2.33) \quad \frac{\partial^2}{\partial w \partial z_2} K(y_2^* | \mathbf{z}, w) = \frac{\partial^2}{\partial w \partial z_2} \lambda(\mathbf{z}, w) \cdot F_\Delta(y_2^* | \mathbf{z}) - \frac{\partial}{\partial w} \lambda(\mathbf{z}, w) \frac{\partial}{\partial z_2} F_2(y_2^* | \mathbf{z}).$$

By (2.32) and (2.29)

$$\begin{aligned} \frac{\partial}{\partial z_1} F_1(y_1^* | \mathbf{z}) &= \frac{1}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \left[\frac{\partial^2}{\partial w \partial z_1} K(y_1^* | \mathbf{z}, w) - \frac{\partial^2}{\partial w \partial z_1} \lambda(\mathbf{z}, w) \cdot F_\Delta(y_1^* | \mathbf{z}) \right] \\ &= \frac{1}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \left[\frac{\partial^2}{\partial w \partial z_1} K(y_1^* | \mathbf{z}, w) - \frac{\partial^2}{\partial w \partial z_1} \lambda(\mathbf{z}, w) \frac{\frac{\partial}{\partial w} K(y_1^* | \mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \right] \end{aligned}$$

Using this and (2.29) in (2.30),

$$\begin{aligned} \frac{\partial}{\partial z_1} K(y_1^* | \mathbf{z}, w) &= \frac{\frac{\partial}{\partial z_1} \lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \frac{\partial}{\partial w} K(y_1^* | \mathbf{z}, w) + \frac{\lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \left[\frac{\partial^2}{\partial w \partial z_1} K(y_1^* | \mathbf{z}, w) - \frac{\partial^2}{\partial w \partial z_1} \lambda(\mathbf{z}, w) \frac{\frac{\partial}{\partial w} K(y_1^* | \mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \right] \\ &= \frac{1}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \left[\frac{\partial}{\partial z_1} \lambda(\mathbf{z}, w) - \frac{\lambda(\mathbf{z}, w) \frac{\partial^2}{\partial w \partial z_1} \lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \right] \frac{\partial}{\partial w} K(y_1^* | \mathbf{z}, w) + \frac{\lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \frac{\partial^2}{\partial w \partial z_1} K(y_1^* | \mathbf{z}, w) \\ &:= \beta_1(\mathbf{z}, w) \frac{\partial}{\partial w} K(y_1^* | \mathbf{z}, w) + \frac{\lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \frac{\partial^2}{\partial w \partial z_1} K(y_1^* | \mathbf{z}, w) \end{aligned}$$

with

$$\beta_1(\mathbf{z}, w) := \frac{1}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \left[\frac{\partial}{\partial z_1} \lambda(\mathbf{z}, w) - \frac{\lambda(\mathbf{z}, w) \frac{\partial^2}{\partial w \partial z_1} \lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \right]$$

Evaluating this at $(y_a, y_b) \in \mathcal{K}_1 \cap (\mathcal{Y}_1(x, w, \mathbf{z}) \times \mathcal{Y}_1(x, w, \mathbf{z}))$

$$\begin{bmatrix} \frac{\partial}{\partial z_1} K(y_a | \mathbf{z}, w) \\ \frac{\partial}{\partial z_1} K(y_b | \mathbf{z}, w) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial w} K(y_a | \mathbf{z}, w) & \frac{\partial^2}{\partial w \partial z_1} K(y_a | \mathbf{z}, w) \\ \frac{\partial}{\partial w} K(y_b | \mathbf{z}, w) & \frac{\partial^2}{\partial w \partial z_1} K(y_b | \mathbf{z}, w) \end{bmatrix} \begin{bmatrix} \beta_1(\mathbf{z}, w) \\ \frac{\lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \end{bmatrix}.$$

Under Assumptions 2.15 and 2.16

$$\Delta_1 = \frac{\partial}{\partial w} K(y_a | \mathbf{z}, w) \frac{\partial^2}{\partial w \partial z_1} K(y_b | \mathbf{z}, w) - \frac{\partial^2}{\partial w \partial z_1} K(y_a | \mathbf{z}, w) \frac{\partial}{\partial w} K(y_b | \mathbf{z}, w) \neq 0$$

and

$$(2.34) \quad \frac{\lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} = \frac{N_1}{\Delta_1}$$

where

$$N_1 = -\frac{\partial}{\partial z_1} K(y_a | \mathbf{z}, w) \frac{\partial}{\partial w} K(y_b | \mathbf{z}, w) + \frac{\partial}{\partial z_1} K(y_b | \mathbf{z}, w) \frac{\partial}{\partial w} K(y_a | \mathbf{z}, w).$$

Similarly, by (2.33) and (2.29)

$$\begin{aligned}\frac{\partial}{\partial z_2} F_2(y_2^*|\mathbf{z}) &= \frac{1}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \left[-\frac{\partial^2}{\partial w \partial z_2} K(y_2^*|\mathbf{z}, w) + \frac{\partial^2}{\partial w \partial z_2} \lambda(\mathbf{z}, w) \cdot F_\Delta(y_2^*|\mathbf{z}) \right] \\ &= \frac{1}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \left[-\frac{\partial^2}{\partial w \partial z_2} K(y_2^*|\mathbf{z}, w) + \frac{\partial^2}{\partial w \partial z_2} \lambda(\mathbf{z}, w) \frac{\frac{\partial}{\partial w} K(y_2^*|\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \right]\end{aligned}$$

Using this and (2.29) in (2.30),

$$\begin{aligned}\frac{\partial}{\partial z_2} K(y_2^*|\mathbf{z}, w) &= \frac{\frac{\partial}{\partial z_2} \lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \frac{\partial}{\partial w} K(y_2^*|\mathbf{z}, w) + \frac{1 - \lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \left[-\frac{\partial^2}{\partial w \partial z_2} K(y_2^*|\mathbf{z}, w) + \frac{\partial^2}{\partial w \partial z_2} \lambda(\mathbf{z}, w) \frac{\frac{\partial}{\partial w} K(y_2^*|\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \right] \\ &= \frac{1}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \left[\frac{\partial}{\partial z_2} \lambda(\mathbf{z}, w) + \frac{(1 - \lambda(\mathbf{z}, w)) \frac{\partial^2}{\partial w \partial z_2} \lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \right] \frac{\partial}{\partial w} K(y_2^*|\mathbf{z}, w) - \frac{1 - \lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \frac{\partial^2}{\partial w \partial z_2} K(y_2^*|\mathbf{z}, w) \\ &:= \beta_2(\mathbf{z}, w) \frac{\partial}{\partial w} K(y_2^*|\mathbf{z}, w) - \frac{1 - \lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \frac{\partial^2}{\partial w \partial z_2} K(y_2^*|\mathbf{z}, w)\end{aligned}$$

with

$$\beta_2(\mathbf{z}, w) := \frac{1}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \left[\frac{\partial}{\partial z_2} \lambda(\mathbf{z}, w) + \frac{(1 - \lambda(\mathbf{z}, w)) \frac{\partial^2}{\partial w \partial z_2} \lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \right].$$

Evaluate this, this time at $(y_c, y_d) \in \mathcal{K}_2 \cap (\mathcal{Y}_2(x, w, \mathbf{z}) \times \mathcal{Y}_2(x, w, \mathbf{z}))$

$$\begin{bmatrix} \frac{\partial}{\partial z_2} K(y_c|\mathbf{z}, w) \\ \frac{\partial}{\partial z_2} K(y_d|\mathbf{z}, w) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial w} K(y_c|\mathbf{z}, w) & -\frac{\partial^2}{\partial w \partial z_2} K(y_c|\mathbf{z}, w) \\ \frac{\partial}{\partial w} K(y_d|\mathbf{z}, w) & -\frac{\partial^2}{\partial w \partial z_2} K(y_d|\mathbf{z}, w) \end{bmatrix} \begin{bmatrix} \beta_2(\mathbf{z}, w) \\ \frac{1 - \lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \end{bmatrix}.$$

Under Assumptions 2.15 and 2.16

$$\Delta_2 = -\frac{\partial}{\partial w} K(y_c|\mathbf{z}, w) \frac{\partial^2}{\partial w \partial z_2} K(y_d|\mathbf{z}, w) + \frac{\partial^2}{\partial w \partial z_2} K(y_c|\mathbf{z}, w) \frac{\partial}{\partial w} K(y_d|\mathbf{z}, w) \neq 0$$

and

$$(2.35) \quad \frac{1 - \lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} = \frac{N_2}{\Delta_2}$$

where

$$N_2 = -\frac{\partial}{\partial z_2} K(y_c|\mathbf{z}, w) \frac{\partial}{\partial w} K(y_d|\mathbf{z}, w) + \frac{\partial}{\partial z_2} K(y_d|\mathbf{z}, w) \frac{\partial}{\partial w} K(y_c|\mathbf{z}, w).$$

By (2.34) and (2.35),

$$\begin{aligned}\frac{1}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} &= \frac{\lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} + \frac{1 - \lambda(\mathbf{z}, w)}{\frac{\partial}{\partial w} \lambda(\mathbf{z}, w)} \\ &= \frac{N_1}{\Delta_1} + \frac{N_2}{\Delta_2},\end{aligned}$$

yielding

$$\frac{\partial}{\partial w} \lambda(\mathbf{z}, w) = \frac{\Delta_1 \Delta_2}{N_1 \Delta_2 + N_2 \Delta_1}.$$

By (2.34)

$$\lambda(\mathbf{z}, w) = \frac{N_1 \Delta_2}{N_1 \Delta_2 + N_2 \Delta_1}.$$

Moreover, for every $y \in \mathbb{R}^p$, by (2.29)

$$(2.36) \quad F_\Delta(y|\mathbf{z}) = \frac{N_1 \Delta_2 + N_2 \Delta_1}{\Delta_1 \Delta_2} \frac{\partial}{\partial w} K(y|\mathbf{z}, w)$$

and by (2.28)

$$\begin{aligned} F_2(y|\mathbf{z}) &= K(y|\mathbf{z}, w) - \lambda(\mathbf{z}, w) F_\Delta(y|\mathbf{z}) \\ &= K(y|\mathbf{z}, w) - \frac{N_1}{\Delta_1} \frac{\partial}{\partial w} K(y|\mathbf{z}, w), \end{aligned}$$

also yielding

$$\begin{aligned} F_1(y|\mathbf{z}) &= F_\Delta(y|\mathbf{z}) + F_2(y|\mathbf{z}) \\ &= \frac{N_1 \Delta_2 + N_2 \Delta_1}{\Delta_1 \Delta_2} \frac{\partial}{\partial w} K(y|\mathbf{z}, w) + K(y|\mathbf{z}, w) - \frac{N_1}{\Delta_1} \frac{\partial}{\partial w} K(y|\mathbf{z}, w) \\ &= K(y|\mathbf{z}, w) - \frac{N_2}{\Delta_2} \frac{\partial}{\partial w} K(y|\mathbf{z}, w). \end{aligned}$$

□

3. GENERAL MIXTURES

Section 2 dealt with the case where the distribution of the latent variable is supported on a finite number of points. We now turn to the more general setting in which the latent variable is allowed to have infinite support.

An important difference with respect to the case of finite mixtures is that now the dimension of the latent variable giving rise to the mixture plays a key role. The models considered below illustrate this point in greater detail.

Attention will be focused on four main settings: panel data models, mixed proportional hazard models, random coefficients models and random utility models.

3.1. Panel data models. A class of models in which general mixtures are used to model unobserved heterogeneity is panel data models. While there is a vast literature on parametric and semiparametric panel data, the nonparametric treatment of these models has developed more recently and is still an active area of research. We will focus on the setting considered by Evdokimov (2010).

The model takes the form

$$(3.1) \quad y_{it} = m(x_{it}, \alpha_i) + \epsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T$$

where the scalar outcome Y is modeled as a function of observable covariates X , unobserved time-invariant heterogeneity α and unobserved idiosyncratic heterogeneity ϵ . Crucially, the function m is not specified parametrically and the time-invariant heterogeneity α is not assumed to enter the outcome equation additively. The latter feature allows for heterogeneous marginal effects of the covariates on the outcome Y across individuals with different values of α but same value of X . In other words, observationally equivalent individuals are allowed to respond differently to a change in the covariates. The data in several microeconomic applications suggest that this may be the case⁷, which makes this framework particularly relevant for applied work. Note that the model can be interpreted both as a random effects model and as a fixed effects model depending on whether α_i is assumed to be independent of X_i or not, respectively.

Under the random effects assumption, nonparametric treatment of the function m requires a normalization for the distribution of α . Therefore, the unknown parameters are m and the distribution of ϵ conditional on covariates. On the other hand, in the fixed effects case, the distribution of α given covariates can be identified and estimated from the data. This is especially relevant for policy analysis.

Evdokimov (2010) provides sufficient conditions for the nonparametric identification of the parameters of interest in each of these two settings. A key assumption in each case is the monotonicity of m in α ⁸. Importantly, this imposes the implicit restriction that α be scalar-valued. While this requirement does limit the flexibility of the model, the presence of the idiosyncratic shock ϵ makes the framework suitable for modeling a variety of economic situations. For instance, if the goal is to estimate returns to education, ϵ could be thought of as the luck and measurement error component of wages, while α could be interpreted as (time-invariant) unobserved ability.

Two other features make this model appealing to applied researchers. First, only two time periods ($T = 2$) are required for identification, which allows the results above to be used with very short panels. Second, the estimation procedure proposed by Evdokimov (2010) is easy to implement, as it avoids numerical optimization and only relies on computation of empirical cdf's and quantile functions.

Finally, it may be useful to link this paper to the work by Kitamura (2003) discussed in section 2.3. In both cases, the goal is to nonparametrically identify a regression model while accounting for unobserved heterogeneity via mixtures. However, while Kitamura (2003) focuses on cross-sectional

⁷Evdokimov (2010) provides a number of references on this point.

⁸Strict monotonicity is required in the fixed effects case, while in the random effects case weak monotonicity suffices.

data and assumes the mixture is finite, Evdokimov (2010) considers panel data and allows the distribution of the latent variable to be continuous. The main restriction here consists in imposing that the unobserved heterogeneity α be scalar-valued.

3.2. Mixed proportional hazard model. Another setting in which infinite mixtures arise is mixed proportional hazard models. These models are often used in labor economics to investigate the duration of unemployment spells. The standard setup is as follows

$$(3.2) \quad \theta(t, x) = \phi(x) \psi(t)$$

$$(3.3) \quad F(t|x) = 1 - \int_0^\infty e^{-\phi(x) \int_0^t \psi(s) ds} dG(\alpha).$$

Equation (3.2) defines the hazard rate, i.e. the “intensity” with which the event of interest - leaving unemployment, in the labor literature example - occurs. The hazard rate is a function of t , the time elapsed from a given initial point (e.g. the start of the unemployment spell) and covariates X . Equation (3.3) shows that the probability that the event has not occurred yet at time t (e.g. the probability that unemployment lasts at least t) is given by a function of the hazard rate and the distribution of unobserved heterogeneity α . In the labor market example, α may be interpreted as a specification error capturing the effect of variables omitted from the covariates X .

In single spell duration models, the researcher is assumed to be able to recover the distribution $F(t|x)$ from the data. The unknown parameters are the functions ϕ and ψ , and the latent variable cdf G . Elbers and Ridder (1982) show that variation in the covariates X nonparametrically identifies all of the parameters of interest under regularity conditions. The seminal paper by Heckman and Singer (1984a) expands on these results in two main directions.

First, the authors provide an alternative set of assumptions that ensure identification. While these alternative restrictions are not globally weaker than the ones proposed by Elbers and Ridder (1982), the conditions imposed on the distribution of unobserved heterogeneity are indeed less restrictive. In particular, Heckman and Singer (1984a) allow the latent variable α not to have finite moments of any order, thus extending the analysis to fat tail distributions.

Second, the authors show that it is possible to achieve identification even when the researcher does not have access to continuous covariates data. This result comes at the cost of imposing stronger parametric assumptions on the hazard rate. However, the latent variable distribution can still be treated nonparametrically. In a different paper, Heckman and Singer (1984b) propose a nonparametric ML procedure to estimate G .

3.3. Random coefficients models.

3.3.1. *Linear models.* One of the oldest and most common ways to model unobserved heterogeneity in econometrics is to allow the coefficients of a regression to be random. The basic model is

$$(3.4) \quad y = \beta' \mathbf{x}$$

where Y is a scalar, continuously distributed outcome, \mathbf{X} is a d -dimensional vector of observed covariates and β is a random vector with density f_β . It is often assumed that $\mathbf{X} = (1, X_2, \dots, X_d)$, so that the model can be rewritten as

$$(3.5) \quad y = \beta_2 x_2 + \dots + \beta_d x_d + \epsilon$$

with $\epsilon = \beta_1$.

Equation (3.5) shows clearly that a random coefficients model can be viewed as a standard regression model in which the marginal effects of the covariates on the outcome are allowed to differ across individuals. This important feature also characterized the models in Kitamura (2003) and Evdokimov (2010) discussed in previous sections (equations (2.4) and (3.1), respectively). However, while in those models the regression function was treated nonparametrically, here we assume a linear relationship.

In this setup, the structural unknown parameter becomes the distribution of the unobserved heterogeneity β . Traditionally, the problem has been tackled in one of two ways. One approach is to impose parametric assumptions - e.g. independence among the elements of β and normality. These restrictions greatly simplify the problem, but of course misspecification becomes a serious concern. Alternatively, one may choose to rely on relatively mild assumptions on the relationship between covariates and unobserved heterogeneity, namely mean independence and conditional homoskedasticity of β given X . See, for example, the survey by Hsiao and Pesaran (2004). Under these restrictions, the average marginal effect $\mathbb{E}(\beta)$ and the variance $\text{Var}(\beta)$ are identified. However, as pointed out by Hoderlein, Klemelä, and Mammen (2010), identification is not achieved for other important features of the distribution of β . For instance, the quantiles of the marginal distributions of β , properties depending on moments of order higher than 2 (such as skewness and kurtosis), and the dependence structure among the coefficients are not uniquely identified from the data.

More recently, the literature in both statistics and econometrics has tried to find a third way between the two described above. The goal is to identify the entire distribution of β while avoiding parametric assumptions. As the discussion above makes clear, this requires stronger conditions than mean independence and conditional homoskedasticity. In particular, full independence between

the random coefficients and the covariates \mathbf{X} - or some instruments \mathbf{Z} - suffices for nonparametric identification.

A key tool for nonparametric identification of random coefficients models is the theory of Radon transforms.

The Radon transform $\mathcal{R}g$ of a function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is defined by

$$(3.6) \quad (\mathcal{R}g)(\mathbf{z}, w) = \int_{\mathbf{s}'\mathbf{z}=w} g(\mathbf{s}) d\mathbf{s}$$

The above definition applied to the random coefficients model (3.4) leads to

$$(3.7) \quad f(y|\mathbf{x}) = \int_{\beta' \mathbf{x}=y} f_{\beta}(\beta) d\beta = \mathcal{R}f_{\beta},$$

i.e., under the assumption of independence between \mathbf{X} and β , the Radon transform of f_{β} is equal to the conditional density of Y given \mathbf{X} .

The left-hand side of equation (3.7) can be recovered from the data, which suggests that one could apply the inverse Radon transform to an estimator of $f(y|\mathbf{x})$ in order to estimate the unknown f_{β} . However, the inverse of \mathcal{R} is not a continuous operator and small changes in its argument may lead to big changes in the value of the resulting estimator. In other words, this is an ill-posed inverse problem. Some regularization is thus required.

Beran, Feuerverger, and Hall (1996) work with characteristic functions and obtain a consistent and asymptotically normal estimator for f_{β} . On the other hand, Hoderlein, Klemelä, and Mammen (2010) use equation (3.7) directly along with a regularized inverse of \mathcal{R} . The estimator they propose achieves the optimal rate of convergence in a Sobolev class of functions. Further, they extend the model to allow for endogeneity of \mathbf{X} and for nonlinearities in the way \mathbf{X} affects the outcome Y .

3.3.2. Discrete choice models. The previous section analysed linear random coefficient models where the outcome variable was assumed to be continuously distributed. However, in many economics applications, one may be interested in modelling how a discrete variable - e.g. the choice whether or not to buy a good - depends on covariates, and to allow for heterogeneity in this relation across individual observations⁹. This requires extending the basic random coefficients model to the case of nonlinear regression functions.

⁹For instance, Berr and Haile (2010) discuss nonparametric identification in discrete choice models with random coefficients which are often used in the industrial organization literature.

More specifically, the model we consider is:

$$(3.8) \quad y = \mathbb{I} \{ \beta' \mathbf{x} \geq 0 \}$$

where Y is a binary outcome, $\mathbf{X} = (1, X_2, \dots, X_d)$ is a d -dimensional vector of covariates and β is again assumed to be random with unknown density f_β .

As in the linear case, the literature has traditionally employed a parametric approach to the identification and estimation of the unobserved heterogeneity distribution. For instance, Train (2003) discusses the popular choice of mixed logit specifications.

The nonparametric treatment of model (3.8) is a more recent development. An important contribution is the paper by Ichimura and Thompson (1998), who propose a nonparametric ML estimator for the distribution of β . Their procedure, however, is computationally very intensive, as it requires high-dimensional numerical optimization. Note that the same problem often arises also when the model is tackled parametrically.

A recent paper by Gautier and Kitamura (2013) proposes a novel nonparametric identification approach leading to a plug-in estimator that avoids numerical optimization and integration, and thus is easy to implement and computationally parsimonious. In the basic model, the authors assume exogeneity of the covariates \mathbf{X} with respect to unobserved heterogeneity β . This allows one to write

$$(3.9) \quad r(\mathbf{x}) \equiv \mathbb{P}(Y = 1 | \mathbf{X} = \mathbf{x}) = \mathbb{E}_\beta (\mathbb{I} \{ \beta' \mathbf{X} \geq 0 \}) .$$

Further, since all that matters is the angle between \mathbf{X} and β , without loss of generality we can normalize \mathbf{X} and β so that they both belong to the unit sphere in \mathbb{R}^d , denoted \mathbb{S}^{d-1} . Letting the density f_β of β be defined with respect to the uniform spherical measure σ on \mathbb{S}^{d-1} , we can then write

$$(3.10) \quad r(\mathbf{x}) = \int_{b \in \mathbb{S}^{d-1}} \mathbb{I} \{ b' \mathbf{x} \geq 0 \} f_\beta(b) d\sigma(b) = \int_{b \in H(\mathbf{x})} f_\beta(b) d\sigma(b) \equiv \mathcal{H}(f_\beta)(\mathbf{x}),$$

where $H(x) \equiv \{b \in \mathbb{S}^{d-1} : b' \mathbf{x} \geq 0\}$ is called a hemisphere and the mapping \mathcal{H} is called hemispherical transform.

Thus, analogously to the linear case considered in section 3.3.1, we obtain an equation where the left-hand side is estimable from the data - $r(\mathbf{x})$ is the choice probability given $\mathbf{X} = \mathbf{x}$ - and the right-hand side is the image of the unknown parameter f_β under the operator \mathcal{H} .

However, in the binary choice case, the model is not identified unless further restrictions are imposed. The reason is that the operator \mathcal{H} is in general not injective. In particular, Gautier and Kitamura

(2013) show that the even part¹⁰ of the function f_β , denoted f_β^+ , is not identified from equation (3.10). The strategy adopted in the paper consists in two steps. First, conditions are provided under which f_β^- is identified. Second, it is shown that under certain assumptions one can find a one-to-one mapping between f_β^- and f_β , so that f_β is identified as well.

As far as the first step is concerned, a sufficient condition is that the nonconstant covariates (X_2, \dots, X_d) before normalization be supported on the entire space \mathbb{R}^{d-1} . This rules out discrete regressors and regressors with bounded support, although Gautier and Kitamura (2013) also discuss a possible extension to limited-support covariates.

Turning to the second step, a sufficient condition that allows one to recover f_β from f_β^- is that the support of β is a subset of some hemisphere. More precisely, we assume that there exists a vector $\mathbf{c} \in \mathbb{S}^{d-1}$ such that $\mathbb{P}\{\beta' \mathbf{c} > 0\} = 1$. This assumption does not seem to be too restrictive in a number of economic applications. For instance, the condition is satisfied with \mathbf{c} being a vector of zeros and one 1 if the researcher is willing to assume that one element of the β vector is positive. Note that knowledge of which specific element is positive is not required. A case where the sign of a coefficient may be reasonably assumed to be known is the effect of own price on demand for a good. Moreover, the support assumption on the distribution of β imposes constraints on f_β^- , which is identified under weak conditions in the first step. Therefore, the second-step assumption can be tested based on the observables.

Even assuming identification, estimation of f_β is a nontrivial task. Similarly to the linear random coefficient model, the reason is that the operator \mathcal{H} in (3.10) has a discontinuous inverse. Thus, some regularization procedure is required. From this point of view, the main contribution of Gautier and Kitamura (2013) is to show that the operator \mathcal{H} is an analogue of convolution in \mathbb{R}^d , which implies that standard deconvolution techniques can be applied. The formula for the estimator is obtained following the constructive identification proof. Further, the estimator is shown to be consistent and pointwise asymptotically normal, and convergence rates are derived.

The extension to the case where some of the covariates \mathbf{X} are endogenous and suitable instruments are available is also discussed. On a related note, as pointed out by Hoderlein, Klemelä, and Mammen (2010), the assumption of full independence between \mathbf{X} and β is indeed strong, but it can be tested, at least in principle. One could split the support of \mathbf{X} into two subregions and estimate f_β for each subregion separately. One could then compare the resulting estimates to obtain a specification

¹⁰The even part of a function g , denoted g^+ , is defined as $g^+(x) = (g(x) + g(-x))/2$. Similarly, the odd part is defined as $g^-(x) = (g(x) - g(-x))/2$.

test for the model. This procedure could be applied both to the linear model and to the discrete choice model. In practice, however, this method may not be easy to implement due to the discontinuity of the Radon transform and of the hemispherical transform, respectively.

3.4. Random utility models. This section discusses a class of models in which the unobserved heterogeneity is infinite-dimensional. This is in contrast with all the previous cases analysed in this paper, where the latent variable giving rise to the mixture was constrained to be vector-valued (and in some setups - e.g. Evdokimov (2010) - even scalar-valued).

We consider the framework of random utility models (RUM). The fundamental problem tackled by this body of research is how to model heterogeneity in preferences across individuals. If preferences are expressed in terms of utility functions over an infinite commodity space, then it is clear that the unobserved heterogeneity has infinite dimension in this context.

Typically, the researcher has access to data on prices $\mathbf{p} \in \mathbb{R}^k$ and consumption choices $\mathbf{y} \in \mathcal{Y}$, where $\mathcal{Y} \subset \mathbb{R}^k$ is a commodity space. Income can be normalized to 1, so that each price vector \mathbf{p} is associated with a budget set $B(\mathbf{p})$. A crucial assumption that is often made and that will be maintained throughout this section is that the number of possible budget sets is finite. Extending the results from the RUM literature to the case of infinitely many budgets is not straightforward and is still an active area of research.

Each individual is associated with a utility function u , which is modeled as a random object. As mentioned above, this is often used as a modeling device to account for heterogeneity in the preferences across individuals. However, one may also employ the RUM framework to allow for stochastic elements in each individual's preferences.

For a given price vector \mathbf{p} , the consumer choice \mathbf{y} solves the standard maximisation problem

$$(3.11) \quad \mathbf{y} \in \arg \max_{\mathbf{x} \in B(\mathbf{p})} u(\mathbf{x}).$$

The researcher can estimate choice probabilities $\mathbb{P}\{\mathbf{y} \in A | \mathbf{p}\}$, for $A \subset \mathcal{Y}$, from the available data on prices and consumption choices. The question is then whether the estimated choice probabilities can be rationalised as an outcome of the RUM outlined above. In other words, the problem is to find a distribution over individual preferences (i.e. over utility functions) such that, in the aggregate, the behaviour implied by this distribution is consistent with the observed consumer choices.¹¹ This

¹¹McFadden and Richter (1991) show that a necessary and sufficient condition for the existence of such a distribution over preferences is the so-called Axiom of Revealed Stochastic Preferences (ARSP). Thus, the stochastic revealed preference problem may be reformulated as a test of the ARSP.

is the essence of the stochastic revealed preference problem, which may be viewed as a population analogue of the (deterministic) revealed preference theory pioneered by Samuelson (1938). The key difference is that, in the stochastic setting, preferences are allowed to be random, which means that the object that the researcher aims to recover from the data is no longer a single utility function, but a distribution over utility functions.

Once the distribution of the heterogeneous latent preferences has been estimated, the econometrician can use the estimates to perform counterfactual analysis. Specifically, one can compute (bounds on) the expected choice probabilities in settings for which no observations are available. In the case of consumer demand, this amounts to estimating (bounds on) choice probabilities and expected demand for budget sets that are not in the data.

Traditionally, the stochastic revealed preference problem has been tackled parametrically. This approach imposes a functional form on the utility function of each consumer; probit/logit models and their variants have been popular choices in the literature. To allow for heterogeneity across consumers, the utility function is assumed to depend on a finite dimensional parameter which is modelled as random and plays the role of the latent variable. A second parametric restriction is usually imposed on the distribution of the latent variable, which is assumed to belong to a parametric family or to be discrete (see, e.g., McFadden and Train (2000)).

More recently, econometricians have sought approaches to the stochastic revealed preference problem that do not rely on the restrictive parametric assumptions mentioned above. As is often the case, more flexibility comes at a cost and most of the papers in this body of research only achieve partial identification of the unobserved heterogeneity distribution. See, for instance, Manski and Tamer (2002), Honoré and Tamer (2006) and Manski (2007). However, in special settings, it may still be possible to obtain point-identification, as in Cosslett (1983) and Matzkin (1992).

A recent contribution by Kitamura and Stoye (forthcoming) analyses the RUM framework from a fully nonparametric standpoint. In particular, no restrictions whatsoever are imposed on the form of the utility function, except for the basic rationality and local nonsatiation conditions. As mentioned above, allowing for such a high degree of flexibility in general precludes point-identification. However, it is still possible to test the existence of a distribution over consumer types (i.e. utilities) that rationalizes the observed choice behaviour. While the task of testing rationality at the individual level has been undertaken, among others, by Epstein and Yatchew (1985), Kitamura and Stoye provide the first econometric analysis of this nonparametric testing problem in the population-level, stochastic case.

The procedure proposed by Kitamura and Stoye involves testing a potentially very high number of inequalities. Moreover, these inequalities take an “indirect” form that makes it problematic to apply the rich literature on moment inequalities. Therefore, one cannot rely on standard methods to obtain critical values, such as subsampling or moment selection. Kitamura and Stoye thus introduce a new modified bootstrap procedure to compute the critical values for the test. The tools that are developed may also be used to perform counterfactual analysis and obtain identified regions for choice probabilities and expected demand.

4. CONCLUSION

This paper considered several recent developments in the econometrics literature on mixture models. The emphasis has been placed on how these advances can contribute to the usefulness and flexibility of mixtures in a variety of economics applications. The applications in which these results may be employed include: multiple equilibria in discrete games (such as games of entry or technology adoption), analysis of panel data, random utility models and consumer behaviour analysis, measurement error, switching regression models, and proportional hazard models.

We also presented some new results on nonparametric identification of finite mixture models under exclusion restrictions. Our identification strategy requires very mild conditions on the support of the covariates. This is in contrast with the identification at infinity approach often used in the existing literature. It is shown that the basic results can be extended to deal with a mixture model with an arbitrary number of components and models with discrete covariates.

5. APPENDIX

5.1. Proof of Theorem 2.2.

Proof. Since the entire proof is conditional on $X = x$, we drop the conditioning on $X = x$ for ease of exposition. We first need some notation.

For any $\mathbf{w} \in \mathcal{W}^{J-1}$, let $\mathbf{K}(y|\mathbf{z}, \mathbf{w}) \equiv [K(y|\mathbf{z}, w_1), \dots, K(y|\mathbf{z}, w_{J-1})]'$. Let D_q denote the differential operator with respect to variable q , so that $D_w \mathbf{K}(y|\mathbf{z}, \mathbf{w}) \equiv [\frac{\partial}{\partial w} K(y|\mathbf{z}, w_1), \dots, \frac{\partial}{\partial w} K(y|\mathbf{z}, w_{J-1})]'$. Similarly, let $D_{\mathbf{z}_{-J}} \mathbf{K}(y|\mathbf{z}, \mathbf{w})$ be the $(J-1)$ -by- $(J-1)$ matrix with ij -th element $\frac{\partial}{\partial z_j} K(y|\mathbf{z}, w_i)$, i.e. each row contains the derivatives with respect to the different elements of $\mathbf{z}_{-J} \equiv [Z_1, \dots, Z_{J-1}]'$ for a given value of W and let $D_{w\mathbf{z}_{-J}} \mathbf{K}(y|\mathbf{z}, \mathbf{w})$ be the $(J-1)$ -by- $(J-1)$ matrix with ij -th element $\frac{\partial^2}{\partial w \partial z_j} K(y|\mathbf{z}, w_i)$. Analogously, let $D_{z_J} \mathbf{K}(y|\mathbf{z}, \mathbf{w})$ be the $(J-1)$ -dimensional vector with j -th element $\frac{\partial}{\partial z_J} K(y|\mathbf{z}, w_j)$ and let $D_{wz_J} \mathbf{K}(y|\mathbf{z}, \mathbf{w})$ be the $(J-1)$ -dimensional vector with j -th

element $\frac{\partial^2}{\partial w \partial z_J} K(y|\mathbf{z}, w_j)$.

Finally, let $\mathbf{F}(y|\mathbf{z}) \equiv [F_1(y|\mathbf{z}) - F_J(y|\mathbf{z}), \dots, F_{J-1}(y|\mathbf{z}) - F_J(y|\mathbf{z})]'$ and $\mathbf{F}_J(y|\mathbf{z}) \equiv F_J(y|\mathbf{z}) \cdot \mathbf{1}$, where $\mathbf{1}$ is the $(J-1)$ -dimensional column vector of ones. Using the same notation as above, we have that $D_{\mathbf{z}_{-J}} \mathbf{F}(y|\mathbf{z})$ is the $(J-1)$ -by- $(J-1)$ matrix with ij -th element $\frac{\partial}{\partial z_j} [F_i(y|\mathbf{z}) - F_J(y|\mathbf{z})]$ and $D_{z_J} \mathbf{F}_J(y|\mathbf{z})$ is the $(J-1)$ -dimensional column vector $\frac{\partial}{\partial z_J} F_J(y|\mathbf{z}) \cdot \mathbf{1}$.

Given the notation above, we can write model (2.25) as

$$(5.1) \quad \mathbf{K}(y|\mathbf{z}, \mathbf{w}) = \Lambda(\mathbf{w})' \mathbf{F}(y|\mathbf{z}) + \mathbf{F}_J(y|\mathbf{z}).$$

Differentiating (5.1) with respect to W and evaluating at \mathbf{w}^* yields

$$(5.2) \quad D_w \mathbf{K}(y|\mathbf{z}, \mathbf{w}^*) = D_w \Lambda(\mathbf{w}^*)' \mathbf{F}(y|\mathbf{z}).$$

Now we evaluate (5.1) at y^* . Differentiation with respect to \mathbf{z}_{-J} and using Assumption 2.7, we obtain

$$(5.3) \quad D_{\mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*) = \Lambda(\mathbf{w}^*)' D_{\mathbf{z}_{-J}} \mathbf{F}(y^*|\mathbf{z}).$$

Similarly, differentiation with respect to z_J yields

$$(5.4) \quad D_{z_J} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*) = [\mathbf{I} - \Lambda(\mathbf{w}^*)]' D_{z_J} \mathbf{F}_J(y^*|\mathbf{z}),$$

where \mathbf{I} is the $(J-1)$ -dimensional identity matrix. Differentiating (5.3) and (5.4) again with respect to W leads to

$$(5.5) \quad D_{w \mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*) = D_w \Lambda(\mathbf{w}^*)' D_{\mathbf{z}_{-J}} \mathbf{F}(y^*|\mathbf{z}),$$

and

$$(5.6) \quad D_{w z_J} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*) = -D_w \Lambda(\mathbf{w}^*)' D_{z_J} \mathbf{F}_J(y^*|\mathbf{z}).$$

Note that, at $y = y^*$, by Assumption 2.7, the matrix $D_{\mathbf{z}_{-J}}$ is a diagonal matrix with jj -th entry $\frac{\partial}{\partial z_j} F_j(y^*|\mathbf{z}) \neq 0$, for $j = 1, \dots, J-1$, hence it is invertible. Thus, from (5.3) it follows

$$(5.7) \quad \Lambda(\mathbf{w}^*) = [D_{\mathbf{z}_{-J}} \mathbf{F}(y^*|\mathbf{z})']^{-1} D_{\mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*)',$$

which shows that, if $D_{\mathbf{z}_{-J}} \mathbf{F}(y^*|\mathbf{z})$ is identified, then $\Lambda(\mathbf{w}^*)$ is identified as well.

Further, (5.5) and (5.6) imply

$$(5.8) \quad [D_{w \mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*)]^{-1} D_{w z_J} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*) = -[D_{\mathbf{z}_{-J}} \mathbf{F}(y^*|\mathbf{z})]^{-1} D_{z_J} \mathbf{F}_J(y^*|\mathbf{z}).$$

Note that the matrix $D_{w \mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*)$ is invertible because $D_w \Lambda(\mathbf{w}^*)$ is invertible by Assumption 2.8 and $D_{\mathbf{z}_{-J}} \mathbf{F}(y^*|\mathbf{z})$ is invertible because it is full-rank, as discussed above.

Equation (5.8), together with the fact that $D_{\mathbf{z}_{-J}} \mathbf{F}(y^*|\mathbf{z})$ is a diagonal matrix, show that, if $D_{z_J} \mathbf{F}_J(y^*|\mathbf{z})$ were known, then $D_{\mathbf{z}_{-J}} \mathbf{F}(y^*|\mathbf{z})$ would be identified.

Now, (5.4) and (5.7) imply

$$(5.9) \quad D_{z_J} \mathbf{F}_J(y^*|\mathbf{z}) = D_{z_J} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*) + D_{\mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*) [D_{\mathbf{z}_{-J}} \mathbf{F}(y^*|\mathbf{z})]^{-1} D_{z_J} \mathbf{F}_J(y^*|\mathbf{z}),$$

which, by (5.8), becomes

$$(5.10) \quad D_{z_J} \mathbf{F}_J(y^*|\mathbf{z}) = D_{z_J} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*) - D_{\mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*) [D_{w\mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*)]^{-1} D_{wz_J} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*).$$

Equation (5.10) shows that $D_{z_J} \mathbf{F}_J(y^*|\mathbf{z})$ is identified. In fact, since every element of the vector $D_{z_J} \mathbf{F}_J(y^*|\mathbf{z})$ is equal to $\frac{\partial}{\partial z_J} F_J(y^*|\mathbf{z})$, the latter partial derivative is identified by $J - 1$ distinct equations.

As noted above, identification of $D_{z_J} \mathbf{F}_J(y^*|\mathbf{z})$ implies identification of $D_{\mathbf{z}_{-J}} \mathbf{F}(y^*|\mathbf{z})$, which in turn implies identification of $\Lambda(\mathbf{w}^*)$.

We now show how to recover the component cdf's.

Combining (5.3) and (5.5), we obtain

$$(5.11) \quad D_{\mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*) [D_{w\mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*)]^{-1} = \Lambda(\mathbf{w}^*)' [D_w \Lambda(\mathbf{w}^*)']^{-1}.$$

Now, given that $\Lambda(\mathbf{w}^*)$ is identified as shown above, we can use (5.2) and (5.11) to get

$$(5.12) \quad \mathbf{F}(y|\mathbf{z}) = [\Lambda(\mathbf{w}^*)']^{-1} D_{\mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*) [D_{w\mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*)]^{-1} D_w \mathbf{K}(y|\mathbf{z}, \mathbf{w}^*),$$

which shows that $\mathbf{F}(y|\mathbf{z})$ is identified. Note that the matrix $\Lambda(\mathbf{w}^*)$ is invertible by Assumption 2.8.

Finally, using (5.1) and (5.12), we obtain

$$(5.13) \quad \mathbf{F}_J(y|\mathbf{z}) = \mathbf{K}(y|\mathbf{z}, \mathbf{w}^*) - D_{\mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*) [D_{w\mathbf{z}_{-J}} \mathbf{K}(y^*|\mathbf{z}, \mathbf{w}^*)]^{-1} D_w \mathbf{K}(y|\mathbf{z}, \mathbf{w}^*).$$

□

5.2. Proof of Theorem 2.4.

Proof. Again, given that the entire analysis is conditional on $X = x$, we omit this conditioning for brevity.

We can re-write Equation (2.8) as

$$(5.14) \quad K(y|\mathbf{z}, w^*) = F_1(y|\mathbf{z}) + [1 - \lambda(w^*)] [F_2(y|\mathbf{z}) - F_1(y|\mathbf{z})],$$

which shows that, if $[1 - \lambda(w^*)][F_2(y|\mathbf{z}) - F_1(y|\mathbf{z})]$ is identified, then $F_1(y|\mathbf{z})$ is identified as well. Differentiating Equation (2.8) and using Assumption 2.12,

$$(5.15) \quad \frac{\partial}{\partial z_2} K(y^*|\mathbf{z}, w^*) = [1 - \lambda(w^*)] \frac{\partial}{\partial z_2} F_2(y^*|\mathbf{z}).$$

Differentiating (5.15) with respect to z_1 ,

$$(5.16) \quad \frac{\partial^2}{\partial z_1 \partial z_2} K(y^*|\mathbf{z}, w^*) = [1 - \lambda(w^*)] \frac{\partial^2}{\partial z_1 \partial z_2} F_2(y^*|\mathbf{z}).$$

The cross-derivative exists and is nonzero by Assumption 2.13.

Moreover, Differentiating (5.16) with respect to w ,

$$(5.17) \quad \frac{\partial^3}{\partial z_1 \partial z_2 \partial w} K(y^*|\mathbf{z}, w^*) = -\frac{\partial}{\partial w} \lambda(w^*) \cdot \frac{\partial^2}{\partial z_1 \partial z_2} F_2(y^*|\mathbf{z}).$$

Combining (5.16) and (5.17),

$$(5.18) \quad \frac{\frac{\partial^3}{\partial z_1 \partial z_2 \partial w} K(y^*|\mathbf{z}, w^*)}{\frac{\partial^2}{\partial z_1 \partial z_2} K(y^*|\mathbf{z}, w^*)} = -\frac{\frac{\partial}{\partial w} \lambda(w^*)}{[1 - \lambda(w^*)]},$$

which shows that the right-hand side of (5.18) is identified.

Now, Differentiating (2.8) with respect to w yields

$$(5.19) \quad \frac{\partial}{\partial w} K(y|\mathbf{z}, w^*) = -\frac{\partial}{\partial w} \lambda(w^*) [F_2(y|\mathbf{z}) - F_1(y|\mathbf{z})] = -\frac{\frac{\partial}{\partial w} \lambda(w^*)}{[1 - \lambda(w^*)]} \cdot [1 - \lambda(w^*)] [F_2(y|\mathbf{z}) - F_1(y|\mathbf{z})].$$

Combining (5.18) and (5.19) show that $[1 - \lambda(w^*)][F_2(y|\mathbf{z}) - F_1(y|\mathbf{z})]$ is identified, which implies that $F_1(y|\mathbf{z})$ is identified, as claimed at the outset.

The final formula is:

$$(5.20) \quad F_1(y|\mathbf{z}) = K(y|\mathbf{z}, w^*) - \frac{\partial}{\partial w} K(y|\mathbf{z}, w^*) \cdot \frac{\frac{\partial^2}{\partial z_1 \partial z_2} K(y^*|\mathbf{z}, w^*)}{\frac{\partial^3}{\partial z_1 \partial z_2 \partial w} K(y^*|\mathbf{z}, w^*)}.$$

□

REFERENCES

- ABBRING, J. H., AND G. J. VAN DEN BERG (2003): “The non-parametric identification of treatment effects in duration models,” *Econometrica*, 71, 1491–1517.
- ANDREWS, D. W., AND M. M. A. SCHAFGANS (1998): “Semiparametric Estimation of the Intercept of a Sample Selection Model,” *Review of Economic Studies*, 65, 497–517.
- BAJARI, P., H. HONG, AND S. P. RYAN (2010): “Identification and Estimation of a Discrete Game of Complete Information,” *Econometrica*, 78, 1529–1568.
- BERAN, R., A. FEUERVERGER, AND P. HALL (1996): “On nonparametric estimation of intercept and slope distributions in random coefficient regression,” *Annals of Statistics*, 24, 2569–2692.
- BERR, S., AND P. HAILE (2010): “Identification of a heterogeneous generalized regression model with group effects,” Cowles Foundation Discussion Paper No. 1732.
- BERRY, S., M. CARNALL, AND P. SPILLER (2006): “Airline Hubbing, Costs and Demand,” in *Advances in Airline Economics, Vol. 1: Competition Policy and Anti-Trust*, ed. by D. Lee. Elsevier Press, pp. 183–214.
- BERRY, S., AND E. TAMER (2006): “Identification in Models of Oligopoly Entry,” in *Advances in Economics and Econometrics. Ninth World Congress*, ed. by R. Blundell, W. K. Newey, and T. Persson, vol. 2, pp. 46–85. Cambridge University Press.
- BOLLINGER, C. R. (2006): “Bounding mean regressions when a binary regressor is mismeasured,” *Journal of Econometrics*, 73, 387–399.
- BONHOMME, S., K. JOCHMANS, AND J. ROBIN (2012): “Nonparametric estimation of finite mixtures,” Manuscript.
- BOUND, J., C. BROWN, AND N. MATHIOWETZ (2001): “Measurement error in survey data,” in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer. Elsevier Science, pp. 3705–3843.
- BROWNING, M., AND J. M. CARRO (2007): “Heterogeneity and Microeconometrics Modelling,” in *Advances in Economics and Econometrics. Ninth World Congress*, ed. by R. Blundell, W. K. Newey, and T. Persson, vol. 3, pp. 47–74. Cambridge University Press.
- CAMERON, S. V., AND J. J. HECKMAN (1998): “Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts,” *Journal of Political Economy*, 106, 262–311.
- CHAMBERLAIN, G. (1986): “Asymptotic efficiency in semiparametric models with censoring,” *Journal of Econometrics*, 32, 189–218.
- CHEN, X., H. HONG, AND D. NEKIPELOV (2011): “Nonlinear Models of Measurement Errors,” *Journal of Economic Literature*, 49, 901–937.
- CHO, J. S., AND H. WHITE (2007): “Testing for regime switching,” *Econometrica*, 75, 1671–1720.
- CILIBERTO, F., AND E. TAMER (2009): “Market structure and multiple equilibria in airline markets,” *Econometrica*, 77, 1791–1828.
- COOPER, R. W. (2002): “Estimation and Identification of Structural Parameters in the Presence of Multiple Equilibria,” *Les Annales d'Economie et de Statistique*, 6, 1–26.
- COSSLETT, S. (1983): “Distribution-free maximum likelihood estimator of the binary choice model,” *Econometrica*, 51, 765–782.

- ELBERS, C., AND G. RIDDER (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *The Review of Economic Studies*, 49, 403–409.
- EPSTEIN, L., AND A. YATCHEW (1985): "Non-parametric hypothesis testing procedures and applications to demand analysis," *Journal of Econometrics*, 30, 149–169.
- EVDOKIMOV, K. (2010): "Identification and Estimation of a Nonparametric Panel Data Models with Unobserved Heterogeneity," Manuscript.
- FRÜHWIRTH-SCHNATTER, S. (2006): *Finite Mixture and Markov Switching Models*. Springer.
- GAUTIER, E., AND Y. KITAMURA (2013): "Nonparametric estimation in random coefficients binary choice models," *Econometrica*, 81, 581–607.
- HALL, P., AND X. ZHOU (2003): "Nonparametric Estimation of Component Distributions in a Multivariate Mixture," *The Annals of Statistics*, 31, 201–224.
- HECKMAN, J. J., AND B. SINGER (1984a): "The Identifiability of the Proportional Hazard Model," *The Review of Economic Studies*, 51, 231–241.
- (1984b): "A method of minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica*, 52, 271–320.
- HENRY, M., K. JOCHMANS, AND B. SALANIÉ (2013): "Inference on mixtures under tail restrictions," Manuscript.
- HENRY, M., Y. KITAMURA, AND B. SALANIÉ (2014): "Partial Identification of Finite Mixtures in Econometric Models," *Quantitative Economics*, forthcoming.
- HODERLEIN, S., J. KLEMELÄ, AND E. MAMMEN (2010): "Analyzing the random coefficient model nonparametrically," *Econometric Theory*, 26, 804–837.
- HOHMANN, D., AND H. HOLTZMANN (2013): "Two-component mixtures with independent coordinates as conditional mixtures: Nonparametric identification and estimation," *Electronic Journal of Statistics*, 7, 859–880.
- HONORÉ, B., AND E. TAMER (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611–630.
- HOROWITZ, J. L., AND C. MANSKI (1995): "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63, 281–302.
- HSIAO, C., AND M. H. PESARAN (2004): "Random Coefficient Panel Data Models," IZA Discussion Papers 1236.
- ICHIMURA, H., AND T. S. THOMPSON (1998): "Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution," *Journal of Econometrics*, 86, 269–295.
- KASAHARA, H., AND K. SHIMOTSU (2009): "Nonparametric identification of finite mixture models of dynamic discrete choices," *Econometrica*, 77, 135–175.
- (2011): "Sequential Estimation of Dynamic Programming Models with Unobserved Heterogeneity," Working Paper, Graduate School of Economics, Hitotsubashi University.
- (2012): "Testing the Number of Components in Finite Mixture Models," Working Paper, CIRJE-F-867.
- KEANE, M. P., AND K. I. WOLPIN (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522.
- KITAMURA, Y. (2003): "Nonparametric identifiability of finite mixtures," Manuscript.

- LAIRD, N. (1978): "Nonparametric maximum likelihood estimation of a mixing distribution," *Journal of the American Statistical Association*, 73, 805–811.
- LI, T. (2002): "Robust and consistent estimation of nonlinear errors-in-variables models," *Journal of Econometrics*, 110, 1–26.
- LINDSAY, B. G. (1988): "Mixture Models: Theory, Geometry and Applications," in *NSF-CBMS Regional Conference Series in Probability and Statistics*, vol. 5, pp. 1–163. Institute of Mathematical Statistics.
- MANSKI, C. F. (2003): *Partial Identification of Probability Distributions*. Springer-Verlag.
- (2007): "Partial Identification of Counterfactual Probabilities," *International Economic Review*, 48(4), 1393–1410.
- MANSKI, C. F., AND E. TAMER (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546.
- MATZKIN, R. (1992): "Nonparametric and Distribution free estimation of the Binary Threshold Crossing and the Binary Choice Models," *Econometrica*, 60, 239–270.
- McFADDEN, D., AND K. RICHTER (1991): "Stochastic rationality and revealed stochastic preference," in *Preferences, Uncertainty and Rationality*, ed. by J. Chipman, D. McFadden, and K. Richter. Boulder: Westview Press, pp. 161–186.
- McFADDEN, D. L., AND K. TRAIN (2000): "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 447–470.
- MOLINARI, F. (2008): "Partial identification of probability distributions with misclassified data," *Journal of Econometrics*, 144, 81–117.
- QIN, J. (1998): "Semiparametric likelihood based method for goodness of fit tests and estimation in upgraded mixture models," *Scandinavian Journal of Statistics*, 25, 681–691.
- (1999): "Empirical likelihood ratio based confidence intervals for mixture proportions," *Annals of Statistics*, 27, 1368–1384.
- SAMUELSON, P. (1938): "A Note on the Pure Theory of Consumer's Behaviour," *Economica*, 5(17), 61–71.
- TAMER, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147–165.
- TRAIN, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.
- ZOU, F., J. P. FINE, AND B. S. YANDELL (2002): "On empirical likelihood for a semiparametric mixture model," *Biometrika*, 89, 61–75.

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS, YALE UNIVERSITY, NEW HAVEN, CT-06520.

E-mail address: `yuichi.kitamura@yale.edu`

DEPARTMENT OF ECONOMICS, YALE UNIVERSITY, NEW HAVEN, CT-06520.

E-mail address: `giovanni.compiani@yale.edu`