Quantifying The Roles of Uncertainty and Preferences In Understanding Schooling Choices and Earnings Inequality

Flavio Cunha, James J. Heckman and Salvador Navarro

Escola de Pos-Graduacao em Economia

Fundacao Getulio Vargas

28 de Agosto de 2003

Goal of this research

(1) For decades, economists and others have attempted to distinguish "luck" or chance from other factors in determining earnings.

(2) crude description of the decomposition: Look at Mincer earnings equation

$$(*)\ \ln y_{it} = x_{it}\beta + \theta_i + \varepsilon_{it}$$

y_{it} =earnings of person i at time t

x_{it} =observables

θ_i =person specific fixed effect

$\varepsilon_{it} = G(L)\nu_t$

ν_t iid mean zero

(3) Components of $x_{it}, \theta_i, \varepsilon_{it}$ all contribute to inequality measurement

(4) Components of uncertainty are assigned arbitrarily

(5) Is it θ_i?ε_{it}?$\theta_i + \varepsilon_{it}$?"Innovations" in ν_t?

(6) All contribute to "heterogeneity". But how much is uncertainty?

(7) Example: Macurdy, Moffitt and Gottschalk, Abowd and Card all estimate earnings models in which versions of (*) are fit

(8) These statistical decompositions are interesting but tell us little about components unforecastable by agents

(9) Thus Gottschalk and Moffitt show rising variances in (*) over the 1980s. But this says nothing about increasing uncertainty although some (*e.g.* Ljungquist and Sargent) interpret their evidence this way

(10) Goal of this paper is to identify these components. Need a theory and a link of (*) to behavioral theory. Extract forecastable from unforecastable components

(11) Precedent: work by Flavin. Using $P|H$. Innovations in consumption linked to innovations in income

(12) This paper uses schooling.

(a) What future components of income are forecastable at the date schooling decisions are being made?

(b) How is uncertainty resolved over the life cycle?

(c) Can uncertainty and increasing uncertainty explain sluggish responses to the increase in the returns to schooling over the past 30 years?

(d) What role of uncertainty and what is the role of nonpecuniary factors?

Definitions:

(a) Heterogeneity: all sources of variability among persons

(i) observable

(ii) unobservable.

(b) Uncertainty: unforecastable components of heterogeneity.

Goal: Estimate components in agent information sets at age t. See the evolution over time.

Uncertainty: Components unforecastable at age t.

By-products: Unite and Extend Three Literatures

1. Treatment Effect and Policy Evaluation Literature (Derive Treatment Effects from Well Developed Economic Model)

2. Inequality Measurement Literature (Lift anonymity postulates)

3. Literature on Earnings Dynamics (Examine the Effect of Controlling for Selection into Schooling on the Estimated Earnings Dynamics). We present a dynamic version of the Rosen-Willis model.

1. Treatment Effect Literature

   a. Traditional focus is on means (different means emphasized in different subfields)

   b. Our focus is on *Distributions* of Treatment Effects

   c. We link the Treatment Effect Literature to the Choice-Theoretic Structural Econometric Literature using Robust Semiparametric Methods

   d. We identify and estimate both "objective" outcomes of policies and "subjective" evaluations of policies as perceived by persons who choose (self select into) treatments.

   e. We introduce analysis of uncertainty into these models

2. Inequality Measurement Literature

   a. Traditional Focus is on Overall Measures of Inequality (*e.g.,* Gini Coefficients; Variance) and subgroup decompositions (emphasis on statistical accounting).

   b. Link to Policy Analysis is Obscure. Decompositions do not correspond to well-defined policies

   c. *Anonymity Postulate:* Makes a virtue of a cross sectional necessity by comparing distributions using second order stochastic dominance

   d. Does not tell us how a policy affects anyone (*e.g.,* movement of persons from a position in the pre-policy distribution to a position in the post-policy distribution)

e. Our approach allows us to make such comparisons and informs us about which groups are affected and which are not and how policy shifts people within an aggregate distribution.

f. Allows us to examine how specific policies affect choices (*e.g.,* schooling) and how this affects distributions and movements across distributions

g. We examine both distributions of outcomes and distributions of welfare and movements across those distributions

3. Earnings Dynamics Literature.

(a) Does not correct for self selection into schooling.

(b)Silent on components of uncertainty.

(c) Can distinguish between ex ante and expost variability

Basic Framework used In Our Analysis:

Roy (1951) economy

$S = 1$ college

$S = 0$ high school

Two potential outcomes

(Y_0, Y_1) (present value of earnings)

Decision rule governing sectoral choices:

$$S = \begin{cases} 1 & \text{if } I = Y_1 - Y_0 - C \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

C is the cost tuition but also psychic costs

$Y_1 - Y_0$ is the net gain

Decompose Into Means:

$$\begin{aligned} Y_1 &= \mu_1 + U_1 \quad E(U_1) = 0 \\ Y_0 &= \mu_0 + U_0 \;\; E(U_0) = 0 \end{aligned}$$

Implicitly condition on X variables

$$C = \mu_C + U_C,$$

$$I = \mu_1 - \mu_0 - \mu_C + (U_1 - U_0 - U_C).$$

Two Kinds of Policies

(a) Those that affect (Y_0, Y_1, C) (GE Policies)

(b) Those that affect only C and hence S

(Treatment Effect Policies)

General Case

$$\text{Under policy } A : (Y_0^A, Y_1^A, C^A)$$

$$\text{Under policy } B : \left(Y_0^B, Y_1^B, C^B\right)$$

Treatment Effect Case:

Policy A and Policy B

(Y_0, Y_1, C^A) and (Y_0, Y_1, C^B)

$$\begin{aligned} Y^A &= S_A Y_1^A + (1 - S_A) Y_0^A \\ Y^B &= S_B Y_1^B + (1 - S_B) Y_0^B \end{aligned}$$

Case I

$Y_1 - Y_0$ is the same for everyone

Case II

$Y_1 - Y_0$ differs among people of the same X but

$$\Pr(S = 1|Y_1 - Y_0) = \Pr(S = 1)$$

Case III

$Y_1 - Y_0$ differs among people and people act on these differences

$$MTE : E(Y_1 - Y_0|I = 0)$$

$$ATE : E(Y_1 - Y_0)$$

$$TT : E(Y_1 - Y_0|S = 1)$$

Policy Relevant Treatment Effect
(Heckman and Vytlacil, 2001)

Why Useful to Estimate Joint Distributions of Counterfactuals?

1. The proportion of people who benefit (in terms of gross gains) from participation in the program

$$Pr(Y_1 \geq Y_0 \mid X).$$

2. Gains to participants at selected levels of the no treatment state:

$$F(Y_1 - Y_0 \mid Y_0 = y_0, X)$$

or treatment distribution

$$F(Y_1 - Y_0 \mid Y_1 = y_1, X)$$

3. The option value of social programs

4. Can estimate $E(Y_1 - Y_0 \mid X)$ (average treatment effect) and $E(Y_1 - Y_0 \mid X, D = 1)$ (treatment for those who are treated) in vector outcome cases: more than two outcomes

$E(Y_i - Y_j \mid X)$ for all i, j and $E(Y_i - Y_j \mid X, D = 1)$ for all i, j

5. $q(Y_1 - Y_0 \mid X, D = 1)$ (the qth quantile for those who receive treatment.) (Now done only through a perfect ranking assumption)

1 Estimating Distributions of Counterfactual Outcomes

Two counterfactual states (Y_0, Y_1),

$$\begin{aligned} ((Y_0, Y_1) \not\perp\!\!\!\perp X), S &= 1 \text{ if } Y_1 \text{ is observed ;} \\ S &= 0 \text{ otherwise.} \end{aligned}$$

(so Y_0 is observed)

$$Y = SY_1 + (1 - S)Y_0$$

Allow for the possibility of an instrument:

$$(Y_0, Y_1) \perp\!\!\!\perp Z \mid X$$

$$Pr(S = 1 \mid Z, X)$$

But not essential.

Goal: to recover $F(Y_0, Y_1 \mid X)$

Data Available To Analyst:

Observe Y_0 if $S = 0$ and Y_1 if $S = 1$
but we do not observe (Y_0, Y_1) for anyone.

$$F(Y_0 \mid S = 0, X),$$

$$F(Y_1 \mid S = 1, X),$$

$$\text{not } F(Y_1, Y_0 \mid S, X).$$

Three separate problems.

Selection Problem:

Know:

$$F(Y_1 \mid S = 1, X)$$

$$F(Y_0 \mid S = 0, X)$$

If solved we get:

$$F(Y_1 \mid X)$$

$$F(Y_0 \mid X)$$

not

$$F(Y_{0,} Y_1 \mid X)$$

Second Problem : **Recovering Joint Distributions**

Traditional approach:

conditional on X : Y_1 and Y_0 are deterministically related

$$Y_1 \equiv Y_0 + \Delta(X) \tag{1}$$

"Common Effect"

From means of

$$F(Y_0 \mid S = 0, X)$$
$$F(Y_1 \mid S = 1, X)$$

$$E(\Delta(X)) = E(Y_1 \mid X) - E(Y_0 \mid X)$$

Generalization

Quantiles perfectly ranked:

$$Y_1 = F_{1,X}^{-1}(F_{0,X}(Y_0))$$

$$F_{1,X} = F(y_1 \mid X)$$

$$F_{0,X} = F(y_0 \mid X)$$

$$Y_1 = F_{1,X}^{-1}(1 - F_{0,X}(Y_0))$$

This assumption widely used income equality measurement (*e.g.* Murphy, Juhn and Pierce) but it is very arbitrary and as we shall see is not based on any strong foundation.

To See Why Consider the Following

Markov kernels

$M(y_1, y_0 \mid X)$ and $\tilde{M}(y_0, y_1 \mid X)$

$$F(y_1 \mid X) = \int M(y_1, y_0 \mid X) dF_0(y_0 \mid X)$$

$$F(y_0 \mid X) = \int \tilde{M}(y_0, y_1 \mid X) dF_1(y_1 \mid X)$$

so

$$F(y_1 \mid X) = \int M(y_1, y_0 \mid X) \tilde{M}(y_0, y_1 \mid X) dF(y_1 \mid X)$$

and

$$F(y_0 \mid X) = \int \tilde{M}(y_0, y_1 \mid X) M(y_1, y_0 \mid X) dF(y_0 \mid X)$$

Approach Using Choice Theory

$$S = 1(\mu_s(Z) \geq e_s)\ , Z \perp\!\!\!\perp e_s \tag{2}$$

$$Y_1 = \mu_1(X) + U_{1,} \qquad E(U_1) = 0$$

$$Y_0 = \mu_0(X) + U_0, \qquad E(U_0) = 0$$

$$(U_1, U_0) \perp\!\!\!\perp X, Z.$$

where $S = 1(\ Y_1 > Y_0)$ (Roy model).

Identify $F(U_{1,} U_0)$ and $\mu_1(\ X)$, $\mu_0(\ X)$

$F(\ Y_{0,}\ Y_1 |\ X)$ identified

For More General Rules method Breaks Down:

Heckman (1990) and Heckman and Smith (1998):

$$(Z, X) \perp\!\!\!\perp (U_0, U_1, e_s)$$

(i) $\mu_s(Z)$ is a nontrivial function of Z conditional on X

(ii) full support assumptions on $\mu_1(X)$, $\mu_0(X)$ and $\mu_s(Z)$

Identification of $F(U_0, e_s)$, $F(U_1, e_s)$ and $\mu_1(X)$, $\mu_0(X)$ and $\mu(Z)$.

Identify $F(Y_0, S \mid X, Z)$,$F(Y_1, S \mid X, Z)$

not $F(Y_0, Y_1 | X)$ or $F(Y_0, Y_1, S \mid X, Z)$.

An alternative: Factor Models

Aakvik, Heckman and Vytlacil (1999, 2003):

$$U_0 \ = \ \alpha_0\theta + \varepsilon_{0,} \, U_1 = \alpha_1\theta + \varepsilon_1, e_s = \alpha_s\theta + \varepsilon_s, \\ \theta \perp\!\!\!\perp (\varepsilon_0, \varepsilon_1, \varepsilon_s), \ \varepsilon_0 \perp\!\!\!\perp \varepsilon_1 \perp\!\!\!\perp \varepsilon_s.$$

Can identify $\alpha_0, \alpha_1, \alpha_s, F(U_0, e_s)$ and $F(U_1, e_s)$

$$COV(U_0, e_s) = \alpha_0\alpha_s\,\sigma_\theta^2 \\ COV(U_1, e_s) = \alpha_1\alpha_s\sigma_\theta^2 \\ E(\theta) = 0, E(\theta^2) = \sigma_\theta^2$$

Third Problem: **What is in the Information Set of Agents at Date Schooling Decisions are being Made?**

Let $\widetilde{\mathcal{I}}_s$ be the information set at the date of schooling decisions. What is $\Pr(s = 1|\widetilde{\mathcal{I}}_s)$?

Is $\widetilde{\mathcal{I}}_s = (\frac{Y_1}{1+r} - Y_0, Z, X)$ Rosen-Willis.

2 Counterfactuals for the Multiple Outcome Case

State s, age a for person $\omega \in \Omega$:

$$
\begin{gathered}
Y_{s,a}(\omega) \qquad s = 1, ..., \bar{S}, a = 1, ..., \bar{A}. \\
\bar{S} \text{ states and } \bar{A} \text{ ages}
\end{gathered} \tag{3}
$$

Ceteris paribus Effect (or individual treatment effect) s' at age a'' is:

$$
\Delta((s, a), (s', a'')) = Y_{s,a}(\omega) - Y_{s',a''}(\omega). \tag{4}
$$

Average Treatment Effect:

$$
\begin{gathered}
ATE((s, a), (s', a), x) \\
= E(Y_{s,a} - Y_{s',a} \mid X = x)
\end{gathered}
$$

Across all comparisons?

Lifetime utility: $V_s(\omega), s = 1, ..., \bar{S}$.

$$\tilde{s} = \underset{s}{\text{argmax}} \left\{V_s(\omega)\right\}_{s=1}^{\bar{S}} . \tag{5}$$

$$D_s = 1 \text{ if } \bar{S} \text{ , } \sum_{s=1}^{\bar{S}} D_s = 1.$$

Marginal treatment effect:

$$\lim MTE(a, \overline{V}_{s,s'}) = E(Y_{s,a} - Y_{s',a} \mid V_s = V_{s'} = \overline{V}_{s,s'} \geq V_j, j \neq s, s') \tag{6}$$

$$s' = 1, ..., \bar{S}; s' \neq s.$$

$$MTE_s(a, \{\overline{V}_{s,s'}\}_{s'=1,s'\neq s}^{\overline{S}}) = \tag{7}$$

$$\sum_{\substack{s'=1 \\ s'\neq s}}^{\bar{S}} \lim MTE(a, \overline{V}_{s,s'}) \left(f(V_s, V_{s'} \mid V_s = V_{s'} = \overline{V}_{s,s'} \geq V_j, \;\; j \neq s, s') / \psi(a, \{\overline{V}_{s,s'}\}_{s'=1,s'\neq s}^{\overline{S}}) \right)$$

$$\psi \left(a, \{\overline{V}_{s,s'}\}_{s'=1,s'\neq s}^{\overline{S}} \right) = \sum_{\substack{s'=1 \\ s'\neq s}}^{\overline{S}} f(V_s, V_{s'} \mid V_s = V_{s'} = \overline{V}_{s,s'} \geq V_j, \; j \neq s, s') > 0$$

3 Factor Structure Models

$$V_s = \mu_s(Z) - e_s \qquad s = 1, .., \bar{S}. \tag{8}$$

$$\begin{aligned}\text{example} &: \\ V_s &= Z'\beta_s - e_s \qquad s = 1, .., \bar{S}.\end{aligned}$$

$$e_s = \alpha_s'\theta + \varepsilon_s \tag{9}$$

θ is a $L \times 1$ vector ($\theta_\ell \perp\!\!\!\perp \theta_{\ell'}, \ell \neq \ell'$)

$$\begin{gathered}\theta \perp\!\!\!\perp \varepsilon_s \quad s = 1, ..., \bar{S}; \\ \varepsilon_s \perp\!\!\!\perp \varepsilon_{s'} \quad \forall s, s' = 1, ..., \bar{S}; \quad s \neq s' \\ E(\theta) = 0; \quad E(\varepsilon_s) = 0\end{gathered} \tag{10}$$

Potential outcomes $Y^*_{s,a}$:

$$Y^*_{s,a} = \mu_{s,a}(X) + \alpha'_{s,a}\theta + \varepsilon_{s,a} \tag{11}$$

$$E(\varepsilon_{s,a}) = 0$$

$$\theta \perp\!\!\!\perp \varepsilon_{s,a}; \quad s = 1, ..., \bar{S}; \quad a = 1, ..., \bar{A}, \tag{12}$$

$$\varepsilon_{s,a} \perp\!\!\!\perp \varepsilon_{s',a''}; \quad \forall\ s \neq s', \quad \forall\ a, a''. \tag{13}$$

$$\begin{gathered} \varepsilon_{s,a} \perp\!\!\!\perp \varepsilon_{s'}, \\ \forall\ s',\ s = 1, ..., \bar{S}; \quad a = 1, ..., \bar{A}. \end{gathered} \tag{14}$$

$$X_{s,a} \perp\!\!\!\perp (\theta, \varepsilon_{s',a''}),$$
$$\forall\ s, a, s', a'',\ s = 1, ..., \bar{S};\ a = 1, ..., \bar{A}. \tag{15}$$

$Y^*_{s,a}$ vector valued

$Y^*_{s,a}$ in (11) as a latent variable

$$Y_{s,a} = 1(Y^*_{s,a} \geq 0).$$

Relationship To Matching

Matching

$$Y_{s,a} \perp\!\!\!\perp D_s \mid X = x,\ Z = z,\ \Theta = \theta \text{ for all } s$$

$$V_s \perp\!\!\!\perp D_s \mid X = x,\ Z = z,\ \Theta = \theta \text{ for all } s.$$

ATE:

$$\begin{aligned} E(Y_{s,a} - Y_{s',a} \ \mid \ X, \theta) = \\ E(Y_{s,a} \ \mid \ X, \Theta = \theta, D_s = 1) - E(Y_{s',a} \mid X, \Theta = \theta, D_{s'} = 1) \end{aligned}$$

Measurement System

$$
\begin{aligned}
M_1 &= \mu_1(x) + \beta_{11}\theta_1 + ... + \beta_{1K}\theta_K + \varepsilon_1 \\
&\vdots \\
M_L &= \mu_L(x) + \beta_{L1}\theta_1 + ... + \beta_{LK}\theta_K + \varepsilon_L
\end{aligned} \tag{16}
$$

$$
\boldsymbol{\varepsilon}_M = (\varepsilon_1, ..., \varepsilon_L), E(\boldsymbol{\varepsilon}_M) = 0 \tag{17}
$$

$\theta = (\theta_1, .., \theta_K) \perp\!\!\!\perp (\varepsilon_1, ..., \varepsilon_L), \theta \perp\!\!\!\perp \varepsilon_s,$

$\theta \perp\!\!\!\perp \varepsilon_{s,a}, \varepsilon_{s,a} \perp\!\!\!\perp \boldsymbol{\varepsilon}_M \perp\!\!\!\perp \varepsilon_s, \varepsilon_i \perp\!\!\!\perp \varepsilon_j,$

$\theta_i \perp\!\!\!\perp \theta_j, \forall i \neq j, i, j = 1, ..., K.$

Key idea: Measurements not state contingent (See Hansen, Heckman and Mullen (2003) for state contingent measurement systems).

Choice Equations

$$V(s) = \mu_s(Z, X) + \gamma_s' \theta + \varepsilon_s \tag{18}$$
$$V = \{V(s)\}_{s=1}^{\bar{S}}$$

Separable choice equations, $s = 1, ...\bar{S}$.

4 Identification of Semi-parametric Factor Models

$$G = \mu + \Lambda\theta + \boldsymbol{\varepsilon} \tag{19}$$

G: $L \times 1, \theta \perp\!\!\!\perp \boldsymbol{\varepsilon}$, μ: $L \times 1$, θ: $K \times 1$,

$\boldsymbol{\varepsilon}$: $L \times 1$ and Λ: $L \times K$. $\varepsilon_i \perp\!\!\!\perp \varepsilon_j \quad i \neq j, i, j = 1, ..., K$

Even if $\theta_i \perp\!\!\!\perp \theta_j$, the model is underidentified

$$COV(G) = \Lambda\Sigma_\theta\Lambda' + D_\varepsilon \tag{20}$$

Assume $\theta_i \perp\!\!\!\perp \theta_j \;\; \forall i \neq j$

We require that

$$\underbrace{\frac{L\,(L-1)}{2}}_{\text{Number of off-diagonal covariance elements}} \geq \underbrace{(L \times K - K)}_{\text{Number of unrestricted }\Lambda} + \underbrace{K}_{\text{Variances of }\theta}$$

A necessary condition:

$$L \geq 2K + 1$$

$$
\Lambda = \begin{pmatrix}
1 & 0 & 0 & 0 & \vdots & \ldots & \ldots & 0 \\
\lambda_{21} & 0 & 0 & 0 & \vdots & \ldots & \ldots & 0 \\
\lambda_{31} & 1 & 0 & 0 & \vdots & \ldots & \ldots & 0 \\
\lambda_{41} & \lambda_{42} & 0 & 0 & \vdots & \ldots & \ldots & 0 \\
\lambda_{51} & \lambda_{52} & 1 & 0 & \vdots & \ldots & \ldots & 0 \\
\lambda_{61} & \lambda_{62} & \lambda_{63} & 0 & \vdots & \ldots & \ldots & 0 \\
\lambda_{71} & \lambda_{72} & \lambda_{73} & 1 & \vdots & 0 & \ldots & 0 \\
\lambda_{81} & \lambda_{82} & \lambda_{83} & \lambda_{84} & \vdots & 0 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \vdots & \ldots & \ldots & \ldots \\
\lambda_{L,1} & \lambda_{L,2} & \lambda_{L,3} & \ldots & \vdots & \ldots & \ldots & \lambda_{L,K}
\end{pmatrix}. \qquad (21)
$$

$$COV(g_j, g_l) = \lambda_{j1}\lambda_{l1}\sigma^2_{\theta_1}, \;\; l = 1, 2; \;\; j = 1, ..., L; \;\; j \neq l.$$

In particular

$$\begin{aligned} COV(g_1, g_\ell) &= \lambda_{\ell 1}\sigma^2_{\theta_1} \\ COV(g_2, g_\ell) &= \lambda_{\ell 1}\lambda_{21}\sigma^2_{\theta_1}. \end{aligned}$$

Assuming $\lambda_{\ell 1} \neq 0$, we obtain

$$\frac{COV(g_2, g_\ell)}{COV(g_1, g_\ell)} = \lambda_{21}.$$

Hence, from $COV(g_1, g_2) = \lambda_{21}\sigma^2_{\theta_1}$, we obtain $\sigma^2_{\theta_1}$, and hence $\lambda_{\ell 1}$, $\ell = 1, \ldots, L$. We can proceed to the next set of two measurements and identify

$$COV(g_l, g_j) = \lambda_{l1}\lambda_{j1}\sigma^2_{\theta_1} + \lambda_{l2}\lambda_{j2}\sigma^2_{\theta_2}, \;\; l = 3, 4; \;\; j \geq 3; \;\; j \neq l.$$

Our System:

$G_s = (M, Y_s, D_s)$, G_s are of two types: (a) continuous variables and (b) discrete or censored random variables, including binary strings.

M^c,Y_s^c,M^d,Y_s^d :

Table 1: Components of G_s

	Continuous	Discrete
Variables defined for all s	M^c	M^d
Variables defined for s	Y_s^c	Y_s^d
Indicator of state	$-$	D_s

Continuous variables:

$$\begin{aligned} M^c &= \mu_m^c(X) + U_m^c \\ Y_s^c &= \mu_s^c(X) \ + U_s^c. \end{aligned}$$

Discrete variables:

$$\begin{aligned} M^{*d} &= \mu_m^d(X) + U_m^d \\ Y_s^{*d} &= \mu_s^d(X) \ + U_s^d \end{aligned}$$

U_m^d, U_s^d continuous

Data used for factor analysis :

M^c, M^{*d}, Y_s^c, Y_s^{*d} and $V(s)$, $s = 1, ..., \overline{S}$.

$$Y = \sum_{s=1}^{\overline{S}} D_s Y_s.$$

M, Y, and D_s, $s = 1, .., \overline{S}$ all contain information on θ.

Need to recover joint distribution of M^c, M^{*d}, Y_s^c, Y_s^{*d}, $V(s)$, s=1, ..., $\overline{S}$.

(A-1) $\left(U_m^c, U_m^d, U_s^c, U_s^d, \varepsilon_W\right)$ *have distribution functions that are absolutely continuous with respect to Lebesgue measure with means zero with support* $\mathbb{U}_m^c \times \mathbb{U}_m^d \times \mathbb{U}_s^c \times \mathbb{U}_s^d \times \mathbb{E}_W$ *with upper and lower limits being* $\bar{U}_m^c, \bar{U}_m^d, \bar{U}_s^c, \bar{U}_s^d, \overline{\varepsilon}_W$ *and* $\underline{U}_m^c, \underline{U}_m^d, \underline{U}_s^c, \underline{U}_s^d, \underline{\varepsilon}_W$*, respectively, which may be bounded or infinite. Thus the joint system is measurably separable (variation free). We assume finite variances. The cumulative distribution function of* ε_W *is assumed to be strictly increasing over its full support* $(\underline{\varepsilon}_W, \overline{\varepsilon}_W)$.

(A-2) $(X, Z, Q) \perp\!\!\!\perp (U, \varepsilon_W)$ *where* $U = (U_m^c, U_m^d, U_s^c, U_s^d)$, *where* Q *is a vector of state-specific regressors* $Q = (Q_1, \ldots, Q_{\overline{S}})$ *.* ε_W *has a strictly increasing distribution function.*

Theorem 1 *From data on* $F(M \mid X)$*, one can identify the joint distribution of* $\left(U_m^c, U_m^d\right)$ *(the latter component only up to scale), the function* $\mu_m^d(X)$ *is identified and* $\mu_m^c(X)$ *is identified over the support of* X *(up to scale) provided that the following assumptions, in addition to the relevant components of (A-1) and (A-2), are invoked.*

(A-3) *Order the discrete measurement components to be first. Suppose that there are* $N_{m,d}$ *discrete components, followed by* $N_{m,c}$ *continuous components. Assume Support*$\left(\mu_{1,m}^d(X), \ldots, \mu_{N_{m,d},m}^d(X)\right) \supseteq$ *Support* $\left(U_{1,m}^d, ..., U_{N_m,m}^d\right)$.

(A-4) *For each* $l = 1, ..., N_{m,d}$ $\mu_{l,m}^d(X) = X\beta_{l,m}^d$.

(A-5) *The* X *lives in a subset of* R^{N_X}*. There exists no linear proper subspace of* R^{N_X} *having probability* 1 *under* F_X*, the distribution function of* X.

Proof of Theorem 1: The case where M consists of purely continuous components is trivial. We observe M^c for each X and can recover the marginal distribution for each component. Recall that M is not state dependent.

For the purely discrete case, we encounter the usual problem that there is no direct observable counterpart for $\mu_m^d(X)$. Under (A-1)-(A-5), we can use the analysis of Manski (1988) to identify the slope coefficients $\beta_{l,m}^d$ up to scale, and the marginal distribution of $U_{l,m}^d$. From the assumption that the mean (or median) of $U_{l,m}^d$ is zero, we can identify the intercept in $\beta_{l,m}^d$. We can repeat this for all discrete components. Thus coordinate by coordinate we can identify the marginal distributions of U_m^c, $\widetilde{U}_m^d$, $\mu_m^c(X)$ and $\widetilde{\mu}_m^d(X)$, the latter up to scale ("~" means identified up to scale).

To recover the joint distribution write:

$$\begin{aligned} & \Pr\left(M_c \leq m_c, M_d = (0, ..., 0) \mid X\right) \\ = \; & F_{U_m^c, \widetilde{U}_m^d}\left(m_c - \mu_m^c(X), -\widetilde{\mu}_m^d(X)\right) \end{aligned}$$

by assumption (A-2). To identify $F_{U_m^c, \widetilde{U}_m^d}(t_1, t_2)$ for any given evaluation points in the support of $(U_m^c, \widetilde{U}_m^d)$, we know the function $\widetilde{\mu}_m^d(X)$ and using (A-3) we can find an X where $\widetilde{\mu}_m^d(X) = t_2$. Let $\widehat{x}$ denote this value, so $\widetilde{\mu}_m^d(\widehat{x}) = t_2$. In this proof, t_1, t_2 may be vectors. Thus

$$\begin{aligned} &\Pr\left(M_c \leq m_c, M_d = (0, ..., 0) \mid X = \widehat{x}\right) \\ = \ & F_{U_m^c, \widetilde{U}_m^d}\left(m_c - \mu_m^c\left(\widehat{x}\right), t_2\right) \end{aligned}$$

Let $\widehat{m}_c = t_1 - \mu_m^c\left(\widehat{x}\right)$ to obtain

$$\Pr\left(M_c \leq \widehat{m}_c, M_d = (0, ..., 0) \mid X = \widehat{x}\right) = F_{U_m^c, \widetilde{U}_m^d}\left(t_1, t_2\right)$$

We know the left hand side and thus identify $F_{U_m^c, \widetilde{U}_m^d}$ at the evaluation point t_1, t_2. Since (t_1, t_2) is any arbitrary evaluation point in the support of $U_m^c, \widetilde{U}_m^d$ we can thus identify the full joint distribution.■

Theorem 2 *For the relevant subsets of the conditions (A-1), and (A-2) (specifically, assuming absolute continuity of the distribution of ε_W with respect to Lebesgue measure and $\varepsilon_W \perp\!\!\!\perp (Z, Q)$), and the additional assumptions:*

(A-6) $c_s(Q_s) = Q_s\eta_s, s = 1, ..., \overline{S}, \varphi(Z) = Z'\boldsymbol{\beta}$

(A-7) (Q_1, Z) *is full rank (there is no proper subspace of the support* (Q_1, Z) *with probability* 1). *The* Z *contains no intercept.*

(A-8) Q_s for $s = 2, \ldots, \overline{S}$ is *full rank (there is no proper subspace of* $\left(\mathbb{R}^{Q_s}\right)$ *with probability* 1*).*

(A-9) *Support* $(c(Q_1) - \varphi(Z)) \supseteq$ *Support* (ε_W)

Then the distribution function F_{ε_W} *is known up to a scale normalization on* ε_W *and* $c_s(Q_s), s = 1, ...\bar{s}$, and $\varphi(Z)$ *are identified up to a scale normalization.*□

Proof of Theorem 2:

$$\Pr\left(D_1 = 1 \mid Z, Q_1\right) = \Pr\left(\frac{c_1\left(Q_1\right) - \varphi\left(Z\right)}{\sigma_W} > \frac{\varepsilon_W}{\sigma_W}\right)$$

Under (A-1), (A-2), (A-6), (A-7) and (A-9), it follows that $\frac{c_1(Q_1)-\varphi(Z)}{\sigma_W}$ and $F_{\widetilde{\varepsilon}_W}$ (where $\widetilde{\varepsilon}_W = \frac{\varepsilon_W}{\sigma_W}$) are identified (see Manski, 1988 or Matzkin 1992, 1993). Under rank condition (A-7), identification of $\frac{c_1(Q_1)-\varphi(Z)}{\sigma_W}$ implies identification of $\frac{c_1(Q_1)}{\sigma_W}$ and $\frac{\varphi(Z)}{\sigma_W}$ separately. Write

$$\Pr\left(D_2 = 1 \mid Z, Q_{1,} Q_2\right) = F_{\widetilde{\varepsilon}_W}\left(\frac{c_2\left(Q_2\right) - \varphi\left(Z\right)}{\sigma_W}\right) - F_{\widetilde{\varepsilon}_W}\left(\frac{c_1\left(Q_1\right) - \varphi\left(Z\right)}{\sigma_W}\right).$$

From the absolute continuity of $\widetilde{\varepsilon}_W$ and the assumption that the distribution function of $\widetilde{\varepsilon}_W$ is strictly increasing, we can write

$$\frac{c_2\left(Q_2\right)}{\sigma_W} = F^{-1}_{\widetilde{\varepsilon}_W}\left[\Pr\left(D_2 = 1 \mid Z, Q_1, Q_2\right) + F_{\widetilde{\varepsilon}_W}\left(\frac{c_1\left(Q_1\right) - \varphi\left(Z\right)}{\sigma_W}\right)\right] + \frac{\varphi\left(Z\right)}{\sigma_W}.$$

Thus we can identify $\frac{c_2(Q_2)}{\sigma_W}$ over its support and, proceeding sequentially, we can identify $\frac{c_s(Q_s)}{\sigma_W}$, $s = 3, .., \overline{S}$. Under (A-8) we can identify η_s, $s = 2, .., \overline{S}$.■

$$\Pr\left(M^c \le m^c, M^{*d} \le 0,\ Y_s^c \le y_s^c, Y_s^{*d} \le 0 | D_s = 1, X, Z, Q_s, Q_{s-1}\right) \times \Pr(D_s = 1 | Z, Q_s, Q_{s-1}) \tag{22}$$

$$= \int_{\underline{U}_c}^{m^c - \mu_m^c(X)} \int_{\underline{\widetilde{U}}_m^d}^{-\widetilde{\mu}_m^d(X)} \int_{\underline{U}_s^c}^{y_s^c - \mu_s^c(X)} \int_{\underline{\widetilde{U}}^d}^{-\widetilde{\mu}^d(X)} \int_{\frac{c_{s-1}(Q_{s-1}) - \varphi(Z)}{\sigma_W}}^{\frac{c_s(Q_s) - \varphi(Z)}{\sigma_W}} f\left(U_{m,}^c \widetilde{U}_{m,}^d U_{s,}^c \widetilde{U}_{s,}^d \widetilde{\varepsilon}_W\right) dU_m^c d\widetilde{U}_m^d dU_s^c d\widetilde{U}_s^d d\widetilde{\varepsilon}_W.$$

(A-10)

$$Support\left(-\widetilde{\mu}_m^d(X), -\widetilde{\mu}_s^d(X), \left(\frac{c_s\left(Q_s\right) - \varphi\left(Z\right)}{\sigma_W} - \frac{c_{s-1}\left(Q_{s-1}\right) - \varphi\left(Z\right)}{\sigma_W}\right)\right)$$

$$\supseteq\ Support(U_m^d, U_s^d, \widetilde{\varepsilon}_W) = (\mathbb{U}_m^d \times \mathbb{U}_s^d \times \widetilde{\mathbb{E}}_W).$$

(A-11) *There is no proper linear subspace of* (X, Z, Q_s, Q_{s-1}) *with probability one so the model is full rank.*

As a consequence of (A-6) and (A-10) we can find values of $Q_s, Q_{s-1}, \bar{Q}_s, \underline{Q}_{s-1}$ respectively so that

$$\lim_{\substack{Q_s \to \bar{Q}_s \\ Q_{s-1} \to \underline{Q}_{s-1}}} \Pr\left(D_s = 1 | Z, Q_s, Q_{s-1}\right) = 1.$$

Theorem 3 *Under assumptions (A-1), (A-2), (A-4), (A-6), (A-7),(A-8),(A-9),(A-10) and (A-11),* $\mu_m^c(X), \mu_s^c(X), \widetilde{\mu}_m^d(X), \widetilde{\mu}_s^d(X), \widetilde{\varphi}(Z), c_s(Q_s) \;\; s = 1, ..., \overline{S} - 1$ *are identified as is the joint distribution* $F(U_m^c, \tilde{U}_m^d, U_s^c, \tilde{U}_s^d, \widetilde{\varepsilon}_W)$.$\square$

Proof of Theorem 3: From (A-2), the unobservables are jointly independent of (X, Z, Q). For fixed values of (Z, Q_s, Q_{s-1}), we may vary the points of evaluation for the continuous coordinates (y_s^c) and pick alternative values of $X = \widehat{x}$ to trace out the vector $\mu^c(X)$ up to intercept terms. Thus we can identify $\mu_{s,l}^c(X)$ up to a constant for all $l = 1, ..., N_{c,s}$.(Heckman and Honoré, 1990). Under (A-2), we recover the same functions for whatever values of Z, Q_s, Q_{s-1} are prespecified as long as $c_s(Q_s) > c_{s-1}(Q_{s-1})$, so that there is interval of ε_W bounded above and below with positive probability. This identification result does not require any passage to a limit argument.

For values of (Z, Q_s, Q_{s-1}) such that

$$\lim_{\substack{Q_s \to \bar{Q}_s(Z) \\ Q_{s-1} \to \underline{Q}_{s-1}(Z)}} \Pr(D_s = 1 | Z, Q_s, Q_{s-1}) = 1.$$

where $\bar{Q}_s(Z)$ is an upper limit and $\underline{Q}_{s-1}(Z)$ is a lower limit, and we allow the limits to depend on Z, we essentially integrate out $\widetilde{\varepsilon}_W$ and obtain

$\Pr(M^c \leq m^c, \widetilde{\mu}_m^d \leq -U_m^d, U_s^c \leq y_s^c - \mu^c(X), \widetilde{U}_s^d \leq -\widetilde{\mu}_s^d(X))$

We know that this probability can be achieved by virtue of the support condition of assumption (A-10).

Then proceeding as in the proof of Theorem 1, we can identify $\widetilde{\mu}_s^d(X)$ coordinate by coordinate and we obtain the constants in $\mu_{s,l}^c(X)$, $l = 1, ..., N_{c,s}$ as well as the constants in $\widetilde{\mu}^d(X)$. From the assumption of mean or median zero of the unobservables. In this exercise, we use the full rank condition on X which is part of assumption (A-11).

With these functions in hand, under the full conditions of assumption (A-10) we can fix $y_s^c, y_m^c, \widetilde{\mu}_s^d, \widetilde{\mu}_m^d$, $\frac{c_s(Q_s)-\varphi(Z)}{\sigma_W}$, $\frac{c_{s-1}(Q_{s-1})-\varphi(Z)}{\sigma_W}$ at different values to trace out the joint distribution $F(U_m^c, \tilde{U}_m^d, U_s^c, \tilde{U}_s^d, \widetilde{\varepsilon}_W)$.■

$\overline{S}$ distributions: we factor analyze them.

$$U_s = \alpha_s' \theta + \varepsilon_s \; s = 1, ..., \overline{S}$$

$$\begin{aligned} U_m &= \alpha_m' \theta_m + \varepsilon_m \\ m &= 1, ..., N_m \end{aligned}$$

$$\varepsilon_V = \nu_s' \theta + \varepsilon_{V,s}, s = 1, ..., \overline{S}$$

Theorem 4 Under the normalizations on the factor loadings of the type in (21) for one system s under the conditions of Theorems 1-3, given the normalizations for the unobservables for the discrete components and given at least $2K+1$ measurements (Y, M, V), the unrestricted factor loadings and the variances of the factors $(\sigma^2_{\theta_i}, i = 1, ..., K)$ are identified over all systems.

Proof: The proof is implicit in the discussion surrounding equation (21). ■

Nonparametric Identification of Distributions

$$(\{U_m\}_{m=1}^{N_m}, \{U_{s,a}\}_{a=1}^{\overline{A}}, \widetilde{\varepsilon}_V) = T^s$$

First B_1 elements depend only on θ_1; next $B_2 - B_1$ elements depend on (θ_1, θ_2) and so forth.

Let T_1^s and T_2^s be the first two elements of T^s.We order the elements of T^s so that the first block depends solely on θ_1.

Theorem 5 *If*

$$T_1^s = \theta_1 + v_1$$

and

$$T_2^s = \theta_1 + v_2$$

and $\theta_1 \perp\!\!\!\perp v_1 \perp\!\!\!\perp v_2$, *the means of all three generating random variables are finite,* $E(v_1) = E(v_2) = 0$, *and the conditions of Fubini's theorem are satisfied for each random variable, and the random variables possess nonvanishing (*a.e.*) characteristic functions, then the densities of* (θ_1, v_1, v_2), $g(\theta_1), g_1(v_1), g_2(v_2)$, *respectively, are identified.*

Proof: Kotlarski (1967). See also Rao (1992). ■

$$T_1^s = \lambda_{11}^s\theta_1 + \varepsilon_1^s \text{ where } \lambda_{11}^s = 1$$
$$T_2^s = \lambda_{21}^s\theta_1 + \varepsilon_2^s \text{ where } \lambda_{21}^s \neq 0.$$

$$T_1^s = \theta_1 + \varepsilon_1^s$$
$$\frac{T_2^s}{\lambda_{21}^s} = \theta_1 + \varepsilon_2^{*,s}.$$

Proceeding to equations $B_1 + 1$ and $B_1 + 2$

$$T_{B_1+1}^s = \lambda_{B_1+1,1}^s\theta_1 + \theta_2 + \varepsilon_{B_1+1}^s$$
$$T_{B_1+2}^s = \lambda_{B_1+2,1}^s\theta_1 + \lambda_{B_1+2,2}^s\theta_2 + \varepsilon_{B_1+2}^s.$$

$$T_{B+1}^s - \lambda_{B_1+1,1}^s\theta_1 = \theta_2 + \varepsilon_{B_1+1}^s$$
$$\frac{T_{B+2}^s - \lambda_{B_1+2,1}^s\theta_1}{\lambda_{B_1+2,2}^s} = \theta_2 + \varepsilon_{B_1+2}^{*,s}$$

$\varepsilon_{B_1+2}^{*,s} = \frac{\varepsilon_{B_1+2}^s}{\lambda_{B_1+2,2}^s}$.

Example:

Two potential outcomes at $a : (Y_0, Y_1)$. Set Z big, so $D_s = 1$.

$$\begin{aligned} Y_{0a} &= \alpha_{0a}\theta + \varepsilon_{0a}, \quad a = 1, ..., A \\ Y_{1a} &= \alpha_{1a}\theta + \varepsilon_{1a}, \quad a = 1, ..., A \end{aligned}$$

$$\begin{aligned} &(\varepsilon_{0a} \perp\!\!\!\perp \varepsilon_{1a'}) \qquad \forall_{a,a'} \\ &(\varepsilon_{0a}, \varepsilon_{1a}) \perp\!\!\!\perp \theta \qquad \forall a \end{aligned}$$

Suppose that $\Pr(\ D_s = 1|Z) \doteq 1$ for $Z \in \mathcal{Z}_1$ and $\Pr(\ D_s = 0|Z) \doteq 0$, for $Z \in \mathcal{Z}_0$.

$$\begin{aligned} Y_{0a} &= \mu_{0a} + \alpha_{0a}\theta + \varepsilon_{0a}, a = 1, \ldots, A \\ Y_{1a} &= \mu_{1a} + \alpha_{1a}\theta + \varepsilon_{1a}, a = 1, \ldots, A, \end{aligned}$$

$$\begin{aligned} COV(Y_{0a}, Y_{0a'}) &= \alpha_{0a}\alpha_{0a'}\sigma_\theta^2, a \neq a', a, a' = 1, \ldots, A,\ for\ Z \in \mathcal{Z}_0, \\ COV(Y_{1a}, Y_{1a'}) &= \alpha_{1a}\alpha_{1a'}\sigma_\theta^2,\ \ a \neq a', a, a' = 1, \ldots, A,\ for\ Z \in \mathcal{Z}_1. \end{aligned}$$

$A = 3$, so if we have three panel observations, we obtain

$$\begin{aligned} COV(Y_{01}, Y_{02}) &= \alpha_{01}\alpha_{02}\sigma_\theta^2, \\ COV(Y_{01}, Y_{03}) &= \alpha_{01}\alpha_{03}\sigma_\theta^2, \text{for } Z \in \mathcal{Z}_0, \\ COV(Y_{02}, Y_{03}) &= \alpha_{02}\alpha_{03}\sigma_\theta^2, \end{aligned}$$

$$\begin{aligned} COV(Y_{11}, Y_{12}) &= \alpha_{11}\alpha_{12}\sigma_\theta^2, \\ COV(Y_{11}, Y_{13}) &= \alpha_{11}\alpha_{13}\sigma_\theta^2, \text{ for } Z \in \mathcal{Z}_1 \\ COV(Y_{12}, Y_{13}) &= a_{12}a_{13}\sigma_\theta^2. \end{aligned}$$

Assuming $\alpha_{01} = 1$ or $\sigma_\theta^2 = 1, \alpha_{03} \neq 0$

$$\frac{COV(Y_{01}, Y_{02})}{COV(Y_{01}, Y_{03})} = \frac{\alpha_{02}}{\alpha_{03}}$$

Given $\sigma_\theta^2 = 1$ and $COV(Y_{02}, Y_{03}) = \alpha_{02}\alpha_{03}$, we obtain

$$(\alpha_{03})^2 = \frac{COV(Y_{02}, Y_{03})COV(Y_{01}, Y_{03})}{COV(Y_{01}, Y_{02})}.$$

The variances of the uniquenesses:

$$\begin{aligned} Var(\varepsilon_{0a}) &= Var(Y_{0a}) - \alpha_{0a}^2 \sigma_\theta^2 \; a = 1, \ldots, A \\ Var(\varepsilon_{1a}) &= Var(Y_{1a}) - \alpha_{1a}^2 \sigma_\theta^2 \; a = 1, \ldots, A. \end{aligned}$$

$$\begin{aligned} COV(T, Y_{0a}) &= \beta \alpha_{0a} \sigma_\theta^2 \quad a = 1, ..., A \\ COV(T, Y_{1a'}) &= \beta \alpha_{1a'} \sigma_\theta^2 \quad a' = 1, ..., A \end{aligned}$$

Assume $\beta \neq 0$ and $\alpha_{0a}, \alpha_{1a'} \neq 0$:

$$\frac{COV(T, Y_{1a'})}{COV(T, Y_{0a})} = \frac{\alpha_{1a'}}{\alpha_{0a}}$$

$a = 1, ..., A, a' = 1, ..., A$ and $a \neq a'$

This fixes the sign of $COV(Y_{0a}, Y_{1a'})$ for all a and a'.

Can extend to use choice data instead of measurements.

Factor Representation of Stationary AR(1) Processes

Theorem *Let $\{y_t\}_{t=1}^{T}$ be an stationary AR(1) process with T odd integer observations where $|\rho| < 1$ and $\varepsilon_t \sim iid(0,\sigma^2)$. Then we can always find $M = \frac{T-1}{2}$ orthogonal factors θ, corresponding loadings α (which may be normalized to zero or one) and uniquenesses $u_t \sim iid(0,s_t^2)$ such that $y_t = \rho y_{t-1} + \varepsilon_t$ can be represented by $y_t = \alpha_1^t\theta_1 +\ldots+\alpha_M^t\theta_M + u_t$.*

Proof By induction. The theorem obviously holds for T = 1 observations of the AR(1) process. Assume it holds for any odd integer T number of observations. We have to show that it also holds for T+2. Indeed, let

$$y_t = \alpha_1^t\theta_1 +\ldots+\alpha_M^t\theta_M + u_t, t = 1,\ldots,T$$

be the factor representation of an stationary AR(1) process $\{y_t\}_{t=1}^{T}$. If we add two more observations, one cannot represent $\{y_t\}_{t=1}^{T+2}$ with the original M factors, because the extra number of parameters in the correlation matrix to be reproduced far exceeds the number of loadings that we get from the two extra observations. Thus, we need more factors. We show that only one more factor is necessary to represent the stationary AR(1) process $\{y_t\}_{t=1}^{T+2}$ after the addition of the two extra observations. Indeed, let

$$y_t = \alpha_1^t\theta_1 +\ldots+\alpha_M^t\theta_M + u_t, t = 1,\ldots,T-1$$

$$y_t = \alpha_1^t\theta_1 +\ldots+\alpha_M^t\theta_M + \alpha_{M+1}^t\theta_{M+1} + u_t, t = T, T+1, T+2$$

be its factor representation. Notice that

$$Corr(y_t,y_{T+1}) = \sum_{i=1}^{M}\alpha_i^t\alpha_i^{T+1}\sigma_i^2,\ t = 1,\ldots,T-1$$

$$Corr(y_t,y_{T+2}) = \sum_{i=1}^{M}\alpha_i^t\alpha_i^{T+2}\sigma_i^2,\ t = 1,\ldots,T-1$$

But this is a system of $2(T-1)$ with $2M = T-1$ unknowns. Thus, we can solve the system above to recover $(\alpha_i^{T+1},\alpha_i^{T+2})_{i=1}^{M}$. Now,

$$Corr(y_T,y_{T+1}) = \left[\sum_{i=1}^{M}\alpha_i^T\alpha_i^{T+1}\sigma_i^2\right] + \alpha_{M+1}^T\alpha_{M+1}^{T+1}\sigma_{M+1}^2 = \rho$$

$$Corr(y_T,y_{T+2}) = \left[\sum_{i=1}^{M}\alpha_i^T\alpha_i^{T+2}\sigma_i^2\right] + \alpha_{M+1}^T\alpha_{M+1}^{T+2}\sigma_{M+1}^2 = \rho^2$$

$$Corr(y_{T+1},y_{T+2}) = \left[\sum_{i=1}^{M}\alpha_i^{T+1}\alpha_i^{T+2}\sigma_i^2\right] + \alpha_{M+1}^{T+1}\alpha_{M+1}^{T+2}\sigma_{M+1}^2 = \rho$$

Notice that the terms in brackets were already determined. Thus, we have three equations to identify four unknowns. We thus need one normalization. Let $\alpha_{M+1}^T = 1$

and define $A_1 = \left[\sum_{i=1}^{M} \alpha_i^T \alpha_i^{T+1} \sigma_i^2\right]$,$A_2 = \left[\sum_{i=1}^{M} \alpha_i^T \alpha_i^{T+2} \sigma_i^2\right]$ and $A_3 = \left[\sum_{i=1}^{M} \alpha_i^{T+1} \alpha_i^{T+2} \sigma_i^2\right]$. Then, the following system must be solved for α_{M+1}^{T+1},α_{M+1}^{T+2}, and σ_{M+1}^2 :

$$\alpha_{M+1}^{T+1} \sigma_{M+1}^2 = \rho - A_1$$

$$\alpha_{M+1}^{T+2} \sigma_{M+1}^2 = \rho - A_2$$

$$\alpha_{M+1}^{T+1} \alpha_{M+1}^{T+2} \sigma_{M+1}^2 = \rho - A_3$$

and the solution is:

$$\alpha_{M+1}^{T+1} = \frac{\rho - A_3}{\rho - A_2}$$

$$\alpha_{M+1}^{T+2} = \frac{\rho - A_3}{\rho - A_1}$$

$$\sigma_{M+1}^2 = \frac{(\rho - A_2)(\rho - A_1)}{\rho - A_3}$$

Finally, we have two linear equations to determine the variances of the two uniquenesses.