

Gutierrez, Federico H.

**Working Paper**

## A Simple Solution to the Problem of Independence of Irrelevant Alternatives in Choo and Siow Marriage Market Model

GLO Discussion Paper, No. 387

**Provided in Cooperation with:**  
Global Labor Organization (GLO)

*Suggested Citation:* Gutierrez, Federico H. (2019) : A Simple Solution to the Problem of Independence of Irrelevant Alternatives in Choo and Siow Marriage Market Model, GLO Discussion Paper, No. 387, Global Labor Organization (GLO), Essen

This Version is available at:  
<https://hdl.handle.net/10419/202040>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# A Simple Solution to the Problem of *Independence of Irrelevant Alternatives* in Choo and Siow Marriage Market Model

Federico H. Gutierrez\*  
*Vanderbilt University and GLO*

August 20, 2019

## Abstract

This paper proposes a simple solution to the independence of irrelevant alternatives (IIA) problem in Choo and Siow (2006) model, overcoming what is probably the main limitation of this approach. The solution consists of assuming match-specific rather than choice-specific random preferences. The original marriage matching function gets modified by an adjustment factor that improves its empirical properties. Using the American Community Survey, I show that the new approach yields significantly different results affecting the qualitative conclusions of the analysis. The proposed solution to the IIA problem applies to other settings in which the relative “supply” of choices is observable.

JEL: J12, J16, J10

Keywords: Independence of irrelevant alternatives, marriage market, transferable utility.

## 1 Introduction

The estimation of marriage market models is a challenging task. “Prices” are embedded in implicit transfers between husbands and wives. Market participants are highly heterogeneous, differing in multiple dimensions such as age, education, personality, and so on. Additionally, individuals of one sex may not agree on the rankings of individuals of the opposite sex.

Choo and Siow (2006) seminal paper provides a tractable model of marriage markets with transferable utilities. Their model generates a new marriage matching function with highly desirable properties. First, this matching function incorporates spillover effects. That is, the

---

\*corresponding author: federico.h.gutierrez@vanderbilt.edu - 401 Calhoun Hall, Vanderbilt University - Nashville, TN 37240

probability that a type  $j$  woman (e.g., a 25-year-old woman) marries a type  $i$  men (e.g., a 27-year-old man) not only depends on the stock of men and women of these types in the market but also on the relative supply of other types  $i' \neq i$  and  $j' \neq j$  (e.g., the number of 22-year-old women in the market). Second, the CS marriage matching function has a clear interpretation. The values of its log transformation are the average marriage gains for all the combinations of individual types in the market.

Choo and Siow (2006) marriage matching function is a valuable device in applied work. However, it has a severe limitation. It is not invariant to the set of characteristics used to classify individuals into types. A clear case of this problem is when irrelevant traits define individual types. For example, assume that people differ only in age and eye color. Men and women in the marriage market value the age of their partners (they may prefer marrying someone of similar age) but are indifferent to the color of their partner's eyes. A researcher studying the matching process using the CS approach will obtain very different conclusions if he/she classifies market participants solely on age or on the product of age and eye color. Below, I show that using the original Choo and Siow approach without solving the irrelevant trait problem leads to inconsistent and eventually contradictory results.

The problem previously mentioned is the standard one of *independence of irrelevant alternatives* (IIA) in multinomial logit models. If eye colors do not matter for marriage market participants, then the choice-specific preference shock of marrying a blue eyed woman should be identical to the choice-specific preference shock of marrying a brown eyed woman. However, the multinomial logit model assumes that these shocks are independent.

The IIA problem can potentially be solved in the usual way, by allowing choice-specific preference shocks to be correlated across alternatives. However, this approach becomes complicated if not impracticable when the choice set grows large. The solution I propose is much simpler. Instead of imposing choice-specific random shocks, I allow match-specific random components.

Choo and Siow assume that a given man is indifferent between any two women of the same type. On the contrary, I assume that women are heterogeneous in traits that are unobservable to the researcher. Then, a man is never indifferent between two women, whether they belong to the same type or not. The same is true regarding women's preferences.

The consequence of my solution to the IIA problem is that the original CS marriage matching function suffers changes. However, it does so in a simple way by incorporating a multiplicative term. All the properties of the original CS matching function are preserved, and others are added. The modified marriage matching function becomes invariant to the number of types in the market and the presence of irrelevant alternatives.

The problems with the original CS approach are prevalent in applied work for two rea-

sons. First, the characteristics that define the individual types are not well-defined from a theoretical perspective. Then, the researcher should make this decision in an arbitrary and unspecified way. Second, the estimator proposed by Choo and Siow requires a ‘good’ number of observation per type (i.e., thick cells), which imposes constraints on the number of characteristics that form the individual types.

This paper does not have the intention to undermine a valuable approach. On the contrary, the proposed changes have the intention to perfect a methodology that has the potential to reveal relevant patterns in the marriage market.

The solution to the IIA problem proposed here applies to other papers, remarkably Choo (2015) study of dynamic marriage markets and Siow (2015) paper on testing Becker’s assortative matching theory. Moreover, this approach applies to other settings beyond marriage markets, with the condition that the relative supply of choices/alternatives is observable.

The rest of the paper is structured as follows. Section 2 briefly describes the original CS model, highlights its limitations using a simple example, and shows that the proposed solution overcomes the problems. Section 3 derives the modified marriage matching function using match-specific preference shocks. It also shows how to obtain equivalent results by modifying the utility function and maintaining choice-specific shocks. Section 4 uses the American Community Survey to illustrate the differences in the original CS approach and the proposed modification. Section 5 concludes.

## 2 The original model and the proposed modification

Choo and Siow (2006) original model is the following. There are  $I$  types of men and  $J$  types of women in a transferable utility marriage market. In equilibrium, each type  $i = 1, \dots, I$  man transfers the dollar amount  $\tau_{ij}$  to his wife, who belongs to one the  $j = 1, \dots, J$  types. The combination of all types generates  $I \times J$  sub-marriage markets. The model uses a random utility framework (McFadden (1974)) to derive quasi-demands for marriage partners. More specifically, the utility of a man  $g$  belonging to type  $i$  and getting married to a type  $j$  women is the following.

$$V_{ijg} = \tilde{\alpha}_{ij} - \tau_{ij} + \epsilon_{ijg} \tag{1}$$

where  $\tilde{\alpha}_{ij}$  is the systematic gross return of any type  $i$  man who marries a type  $j$  woman;  $\tau_{ij}$  is the equilibrium transfer previously mentioned; and  $\epsilon_{ijg}$  is man  $g$  idiosyncratic component. Very importantly,  $\epsilon_{ijg}$  does not vary within  $j$  types. That is, a type  $i$  man is indifferent between any two type  $j$  women. I relax this assumption below.

The payoff of  $g$  from remaining unmarried is:

$$V_{i0g} = \tilde{\alpha}_{i0} + \epsilon_{i0g} \tag{2}$$

Man  $g$  will choose a type  $j$  woman if

$$V_{ijg} \geq V_{ij'g} \quad , \text{ for all } j' \in \{0, 1, 2, \dots, J\} \quad (3)$$

Assuming that the  $\epsilon_{ijg}$  is independent and identically distributed type I extreme-value, Choo and Siow show that:

$$\begin{aligned} \ln(\mu_{ij}^d) &= \ln(\mu_{i0}^d) + \tilde{\alpha}_{ij} - \tilde{\alpha}_{i0} - \tau_{ij} \\ &= \ln(\mu_{i0}^d) + \alpha_{ij} - \tau_{ij} \end{aligned} \quad (4)$$

where  $\mu_{ij}^d$  is the number of marriages between type  $i$  men and a type  $j$  women “demanded” by men;  $\mu_{i0}^d$  is the number of men who what to remain single. The parameter  $\alpha_{ij} = \tilde{\alpha}_{ij} - \tilde{\alpha}_{i0}$  is the systematic gross return of a type  $i$  man from marrying a type  $j$  woman relative to remaining unmarried. Following an analogous procedure from the women’s perspective gives:

$$\ln(\mu_{ij}^s) = \ln(\mu_{i0}^s) + \gamma_{ij} + \tau_{ij} \quad (5)$$

where  $\gamma_{ij}$  is the systematic gross return of a type  $j$  woman from marrying a type  $i$  man relative to remaining unmarried. In equilibrium,  $\mu_{ij}^d = \mu_{ij}^s = \mu_{ij}$ , then equations (4) and (5) yields:

$$\ln \mu_{ij} - \frac{\ln \mu_{i0} + \ln \mu_{0j}}{2} = \frac{\alpha_{ij} + \gamma_{ij}}{2} \quad (6)$$

Defining  $\ln \Pi_{ij} = (\alpha_{ij} + \gamma_{ij})/2$ , the previous equation gives the original CS marriage matching function.

$$\Pi_{ij} = \frac{\mu_{ij}}{\sqrt{\mu_{i0}\mu_{0j}}} \quad (7)$$

## Preview of the modified marriage matching function

The modified CS marriage matching function proposed in this paper is the following.

$$\Pi_{ij}^* = \frac{\mu_{ij}}{\sqrt{\mu_{i0}\mu_{0j}}} \sqrt{\frac{N_f N_m}{m_i f_j}} \quad (8)$$

where  $m_i$  and  $f_i$  are the number of type  $i$  men and the number of type  $j$  women in the marriage market. Quantities  $N_f = \sum_j^J f_j$  and  $N_m = \sum_i^I m_i$  are the total number of men and women in the market regardless of their types.<sup>1</sup>

I derive the modified marriage matching function (8) from the original CS model with the caveat that the idiosyncratic component in utility (1) varies both by  $g$  man of type  $i$  and

<sup>1</sup>The dynamic marriage market function in Choo (2015) includes a component  $\sqrt{m_i f_j}$  that is conceptually different than the one in (8). The matching function in Choo (2015) should be adjusted in a similar way as here if one wants to correct for the irrelevant trait problem.

by  $h$  woman of type  $j$ . Before showing the derivation of (8), I present a simple example that illustrates the limitations of the original matching function (7) and how the modified version (8) overcomes such problems.

**Example: Irrelevant traits.** Consider a marriage market with one thousand men and one thousand women. Everyone has the same age. The only source of observable heterogeneity is the color of the eyes. Four hundred men and four hundred women have blue eyes. The rest of the people have brown eyes. There are other sources of heterogeneity such as personality traits, physical characteristics, and cultural norms that are unobservable to the researcher but noticeable to the men and women in the market.

Assume that the color of people’s eyes is an irrelevant trait for marriage market participants and is independent of any other characteristic (i.e., the color of people’s eyes is not included in  $\tilde{\alpha}_{ij}$  and is independently distributed of  $\epsilon_{ijg}$  in equation (1)). In equilibrium, eight hundred men and eight hundred women get married. Two hundred individuals per sex remain unmarried.

Researcher A correctly believes that the color of the eyes is an irrelevant trait. Then, she considers only one type of men and one type of women ( $I = J = 1$  in the model). Figure 1a) shows the matched and unmatched individuals in matrix form (cell  $i, j$  contains  $\mu_{ij}$ , the number of matches of type  $i$  men and type  $j$  women). For example, the second-row first-column indicates that two hundred men remain unmarried. The second-row second-column indicates that eight hundred men married type 1 women (the unique type).

Figure 1: Same marriage market, different definitions of types

a) One type per sex

		WOMEN	
		Type 1 (all)	
MEN	Type 1 (all)	200	800
		200	

b) Two types per sex

		WOMEN	
		Type 1 (brown eyes)	Type 2 (blue eyes)
MEN	Type 1 (brown eyes)	120	192
	Type 2 (blue eyes)	120	128
		80	288

Researcher B believes that the color of the eyes should be used to classify individuals into different types. Figure 1b) shows the corresponding matrix. Among the six hundred men with brown eyes, 120 remain single, 288 married women with brown eyes and 192 married women with blue eyes. I chose these numbers such that the color of the eyes is irrelevant. For example, there are 60% of men and women with brown eyes. The expected number of brown-eye men getting married to brown-eye women by random chance is  $800 \times 0.6^2 = 288$ .

Since the color of the eyes is an irrelevant trait in the marriage market, then researcher A and researcher B (corresponding to a) and b) in Figure 1) should obtain the same value of the marriage market function  $\Pi_{ij}$ , which is interpreted as the exponential transformation of the average marriage gains in the CS model previously described (i.e.,  $\Pi_{ij} = \exp((\alpha_{ij} + \gamma_{ij})/2)$ ). Moreover, researcher B should find identical values of  $\Pi_{ij}$  for any  $(i, j)$  combination.

Table 1: Original CS marriage function  $\Pi_{ij}$  and proposed modified version  $\Pi_{ij}^*$

	Composition of eye colors in the market			
	40% with blue eyes		20% with blue eyes	
	$\Pi_{ij}$	$\Pi_{ij}^*$	$\Pi_{ij}$	$\Pi_{ij}^*$
	<i>CS original</i>	<i>modified</i>	<i>CS original</i>	<i>modified</i>
	(i)	(ii)	(iii)	(iv)
Researcher A - One type				
type 1-type 1	4	4	4	4
Researcher B - Two types (eye color)				
brown-brown	2.4	4	3.2	4
brown-blue	1.96	4	1.6	4
blue-brown	1.96	4	1.6	4
blue-blue	1.6	4	0.8	4

Table 1 compares the results of the original CS marriage matching function in (7) and the proposed modified version in expression (8). Columns (i) and (ii) show the outcomes of the case depicted in Figure 1, when the proportion of individuals with blue eyes is 40%. For comparison reasons, columns (iii) and (iv) show an analogous case where the size of the market is the same but the proportion of market participants with blue eyes is only 20% per gender (not shown in Figure 1). In both cases, the color of the eyes is assumed to be an irrelevant trait in the marriage market. The combination of eye colors in couples is solely the result of random chance.

The first row in column (i) corresponds to the case where the researcher correctly classifies individuals in one type per gender (Figure 1a)). The resulting quantity can be easily expressed in model parameters. The symmetry of the example implies that the transfer  $\tau_{11}$  in utility (1) is zero. The systematic gains from marriage  $\alpha_{11} = \tilde{\alpha}_{11} - \tilde{\alpha}_{10}$  are  $\ln(4) = 1.39$  for men and for women.  $\alpha_{ij}$  are ‘deep’ parameters of the model. They should not be affected by the

way researchers define individual types.

Table 1 column (i) shows that the original CS marriage matching function does not give consistent result when irrelevant traits are used to classify types (researcher B). It incorrectly indicates that the systematic gains from marriage are higher when an individual marries a brown-eyed person than a blue-eyed person. However, the modified version proposed in this paper (column ii) consistently indicates that the marriage gains are equal across all four combinations of eye colors. Moreover, the modified marriage matching function correctly shows identical marriage gains when researchers use one type or two types per gender.

The CS matching function is also sensitive to the composition of irrelevant traits in the market. The comparison of columns (i) and (iii) suggests that the marriage gains of couples with brown eyes (both the husband and the wife) increase with the share of people with brown eyes. On the contrary, columns (ii) and (iv) correctly show invariant marriage gains.

The problem with the original CS matching function in the previous example is not limited to irrelevant traits. If instead of eye colors, a researcher uses education to classify individuals into types (for example, Cornelson and Siow (2016)), then the resulting marriage gains may change as the composition of educational groups in the society changes (see illustration at the end of the paper).<sup>2</sup>

### 3 Derivation of the modified marriage matching function

Why does the original CS function yields different results in Figures 1a) and 1b)? Because, in the first case, the idiosyncratic component  $\epsilon_{ijg}$  in (1) is implicitly assumed to be perfectly correlated across eye colors, while in the second situation, this correlation is imposed to be zero. I solve this inconsistency in a very simple way.

Consider the following change in the original utility function (1).

$$V_{ijgh} = \underbrace{\tilde{\alpha}_{ij} - \tau_{ij}}_{\eta_{gh}} + \epsilon_{ijgh} \quad (9)$$

The only difference between (1) and (9) is that the random term  $\epsilon_{ijgh}$  now depends on the specific match. That is,  $V_{ijgh}$  in (9) is the utility of man  $g$  belonging to type  $i$  after getting married to woman  $h$  of type  $j$ . If  $g$  gets married to another woman  $h' \neq h$  of the same type  $j$ , then his utility will be different. Utilities (1) and (9) share the assumption that the random component and the systematic component are independent.<sup>3</sup>

Using match-specific idiosyncratic components  $\epsilon_{ijgh}$ , rather than type-specific compo-

---

<sup>2</sup>I do not imply by any means that the conclusions of Cornelson and Siow (2016) are incorrect.

<sup>3</sup>When the random component and the systematic component in the utility functions are correlated, there is a classic endogeneity problem not covered in this paper.



nents, does not only solve the IIA problem but also makes the model more realistic. There are individual characteristics that are unobservable to the research but easily perceived by marriage market participants. Then, the utility derived from marrying different women within the same type, should not be identical.

Allowing for match-specific random components expands the choice set of men (women), from  $J$  ( $I$ ) alternatives to  $N_f$  ( $N_m$ ), where  $N_f$  ( $N_m$ ) is the total number of women (men) in the market. Then, women  $h$  will be at the top of man  $g$ 's ranking if the following condition holds.

$$V_{gh} \geq V_{gh'} \quad , \text{ for all } h' \in \{0, 1, 2, \dots, N_f\} \quad (10)$$

where  $h' = 0$  is the option of remaining single. Assuming that the random term  $\epsilon_{ijgh}$  is independent and identically distributed type-I extreme-value implies that the probability that man  $g$  chooses woman  $h$  is the following.

$$P(h = \arg \max_{k=0, \dots, N_f} V_{gh}) = \frac{\exp(\eta_{gh})}{\sum_{k=0}^{N_f} \exp(\eta_{gk})} \quad (11)$$

where  $\eta_{gh}$  is the systematic component in utility (9) for man  $g$  marrying woman  $h$ . Equation (11) is the standard multinomial logit derived in McFadden (1974) as well as in the appendix of Choo and Siow (2006). Since  $\eta_{gh}$  is identical for all couples of the same type combination  $i, j$ , then the probability that man  $g$  of type  $i$  marries **any** women of type  $j$  is:

$$\begin{aligned} P(j = \arg \max_{k=0, \dots, J} V_{ijg}) &= \sum_{h \in j} \frac{\exp(\eta_{gh} - \eta_{g0})}{1 + \sum_{k=1}^{N_f} \exp(\eta_{gk} - \eta_{g0})} \\ &= \frac{f_j \exp(\alpha_{ij} - \tau_{ij})}{1 + \sum_{l=1}^J f_l \exp(\alpha_{il} - \tau_{il})} \end{aligned} \quad (12)$$

Equation (12) is the multinomial logit expression with the caveat that each choice is weighted by the number of women  $f_j$  in each type  $j = 1, \dots, J$ . Following an identical procedure as Choo and Siow used in their original paper to obtain equations (4) and (5), I derive quasi-demands from (12).

$$\ln(\mu_{ij}^d) = \ln(\mu_{i0}^d) + \alpha_{ij} - \tau_{ij} + \ln(f_j) \quad (13)$$

$$\ln(\mu_{ij}^s) = \ln(\mu_{i0}^s) + \gamma_{ij} + \tau_{ij} + \ln(m_i) \quad (14)$$

Again, in equilibrium,  $\mu_{ij}^d = \mu_{ij}^s = \mu_{ij}$ , then equations (13) and (14) yield:

$$\ln(\mu_{ij}) - \frac{\ln(\mu_{i0}) + \ln(\mu_{0j})}{2} - \frac{\ln(m_i) + \ln(f_i)}{2} = \frac{\alpha_{ij} + \gamma_{ij}}{2} \quad (15)$$

Defining  $\ln\Pi_{ij} = (\alpha_{ij} + \gamma_{ij})/2$  as before gives the following marriage matching function.

$$\Pi_{ij}^{**} = \frac{\mu_{ij}}{\sqrt{\mu_{i0}\mu_{0j}}} \frac{1}{\sqrt{m_i f_j}} \quad (16)$$

The function (16) has the problem that depends on the size of the population. Then, I redefine the utility of man  $g$  in type  $j$  from remaining unmarried as follows.

$$V_{i0g} = \tilde{\alpha}_{i0g} + \ln(N_f) + \epsilon_{ijg} \quad (17)$$

The extra term  $\ln(N_f)$  in utility (17) relative to (2) is required to obtain a stable fraction of singles as the population grows. Otherwise, the probability that a man  $g$  find a woman ‘good enough’ to get married converges to one as the number of women increases.<sup>4</sup> A similar regularization of the model would also be needed in the original CS matching function if one wanted to compare results across different definitions of types.<sup>5</sup>

Using the modified utility (17) equations (13) and (14) becomes.

$$\ln(\mu_{ij}^d) = \ln(\mu_{i0}^d) + \alpha_{ij} - \tau_{ij} + \ln(f_j) - \ln(N_f) \quad (19)$$

$$\ln(\mu_{ij}^s) = \ln(\mu_{i0}^s) + \gamma_{ij} + \tau_{ij} + \ln(m_i) - \ln(N_m) \quad (20)$$

which gives the proposed marriage matching function (8).

### 3.1 Choice-specific shocks and IIA (red bus-blue bus)

There is an alternative representation of the model that uses choice-specific idiosyncratic components and solves the IIA problem. Consider the original CS utility function (1) and add the log of participants in that type/choice.

$$V_{ijg} = \tilde{\alpha}_{ij} - \tau_{ij} + \ln(f_j) + \epsilon_{ijg} \quad (21)$$

As in the original model, the term  $\epsilon_{ijg}$  is an iid random variable distributed as a type I extreme-value. Using the modified utilities (21) and (17) in the original CS model leads to equations (19) and (20) and the modified marriage matching function (8).

The functional form  $\ln(f_j)$  in (21) is not arbitrary. It is a direct consequence of assuming

---

<sup>4</sup>Notice that the probability (12) goes to zero as the number of participants goes to infinity. With the regularization, this probability becomes:

$$P(j = \arg \max_{k=0, \dots, J} V_{ijg}) = \frac{(f_j/N_f) \exp(\alpha_{ij} - \tau_{ij})}{1 + \sum_{k=0}^J (f_k/N_f) \exp(\alpha_{ik} - \tau_{ik})} \quad (18)$$

which converges to a constant as the share of each type  $f_k/N_f$  goes to fixed value.

<sup>5</sup>The CS marriage gains tend to decrease for everyone as the number of types defined by the researcher increases. This statement is valid even if the ‘true’ number of types in the population remain fixed. This point is evident in the illustration at the end of the paper.

that the random shocks are type I extreme-value distributed. This distribution, also known as the standard Gumbel is:

$$F(\epsilon) = \exp(-\exp(-\epsilon)) \quad (22)$$

whose expected value is the Euler constant  $c \approx 0.5772$ . A well-known result is that the maximum of  $n$  independent standard Gumbel distributions is distributed as a standard Gumbel plus  $\ln(n)$ . That is, if  $X_1, X_2, \dots, X_n$  are iid standard Gumbel distributions, then  $Y = \max\{X_1, X_2, \dots, X_n\}$  has a CDF:

$$F(\epsilon) = \exp(-\exp(-\epsilon + \ln(n))) \quad (23)$$

with  $E(Y) = c + \ln(n)$ . Equation (21) says that if man  $g$  chooses to marry a woman of type  $j$ , then his expected utility  $E(V_{ijg})$  will increase with the number of women in the group because man  $g$  will select the “best” women within type  $j$  accordingly to his ranking. That is, if a man walks to a room full of type  $j$  women to choose the most beautiful one, the expected beauty of the chosen girl increases with the number of girls  $f_j$  in the room at the rate  $\ln(f_j)$ .

The solution to the IIA problem proposed in this section covers cases beyond the marriage market. However, it requires the “supply” of each choice to be known. Consider the standard example of a person choosing to commute by either train or bus. The random utilities of a worker  $g$  can be written as follows.

$$V_{g,train} = \alpha_{train} + \ln(N_{train}) + \epsilon_{g,train} \quad (24)$$

$$V_{g,bus} = \alpha_{bus} + \ln(N_{bus}) + \epsilon_{g,bus} \quad (25)$$

where  $\alpha_j$  for  $j \in \{train, bus\}$  is the systematic mean of transport component and  $\epsilon_{g,j}$  is the idiosyncratic random component distributed as type I extreme-value. The novelty in system (24)-(25) is the inclusion of the number of trains  $N_{train}$  and buses  $N_{bus}$  that arrive at man  $g$ 's stop at a given interval of time.

Assume that the government decides to paint half of the buses in blue and the other half in red. The total supply of buses is unchanged. The utility system becomes:

$$V_{g,train} = \alpha_{train} + \ln(N_{train}) + \epsilon_{g,train} \quad (26)$$

$$V_{g,blue-bus} = \alpha_{blue-bus} + \ln(N_{blue-bus}) + \epsilon_{g,blue-bus} \quad (27)$$

$$V_{g,red-bus} = \alpha_{red-bus} + \ln(N_{red-bus}) + \epsilon_{g,red-bus} \quad (28)$$

where,  $N_{blue-bus} = N_{red-bus} = N_{bus}/2$ . If people are indifferent between riding a blue or a red bus, then estimating system (26)-(28) will result in  $\alpha_{blue-bus} = \alpha_{red-bus}$ . More importantly,  $\alpha_{bus}$  in (25) will be identical (in large samples) to  $\alpha_{blue-bus}$  and  $\alpha_{red-bus}$ , correctly evidencing

the indifference of commuters regarding the color of the bus.

The previous procedure solves the problem of independence of irrelevant alternatives when the relative supply of choices is known. It treats the random preference shock **as if** it were associated with each bus rather than each type of bus. This case is specific. Further research is needed in this area to develop a general solution to the IIA problem, which is beyond the scope of this paper.

## 4 Illustration

In this section, I present two cases that illustrate the importance of adjusting the original CS approach. The first case defines individual types using a plausible irrelevant characteristic, which is the quarter of birth of marriage market participant. The second case studies a central relationship in the literature: the change in marriage gains across educational groups.

### 4.1 Data

I use the American Community Survey (ACS) to compute the original CS marriage matching function and compare it to the modified version proposed in this paper. The ACS is a multi-thematic nationally representative survey. It has a repeated cross-section structure. In each year, the sample size is approximately one percent of the U.S. population.

In this study, I pool ten rounds of the ACS covering the period 2008-2017. Since 2008, the ACS has incorporated information about changes in family structure that facilitates the analysis of marriage markets. In particular, the ACS now includes questions about new marriages (married within last years, and years since last got married), new divorces and recent spouse's death.

I define the set of marriage market participants as the sum of i) recently married individuals and ii) available individuals. Recently married individuals are those who got married within the past two years in a heterosexual union. Available individuals are all unmarried persons age 15 to 64 who have not got a divorce or become widowed in the past year. Table 2 shows summary statistics of the marriage market.<sup>6</sup>

### 4.2 Results

**An irrelevant characteristic in practice.** Table 3 shows marriage market gains when individual types are defined using the person's quarter of birth,  $i = 1(j = 1)$  if the man(woman) was born in January, February or March and  $i = 2(j = 2)$  if he(she) was born in any other

---

<sup>6</sup>Alternative definitions of the marriage market give qualitatively similar results.

Table 2: Summary statistics (marriage market observations)

ACS years: 2008-2017	
Age group	
15-24	35.88%
25-34	23.89%
35-44	14.46%
45-54	14.00%
55-64	11.76%
Birth quarter	
Jan-Feb-March	24.47%
April-May-June	24.36%
July-Aug-Sept	26.21%
Oct-Nov-Dec	24.95%
Skill level	
High school or less	48.82%
Some college	25.05%
Associate degree (2 years)	6.59%
Bachelor's degree	13.85%
Post B.A.	5.68%
Married past two years	
Yes	7.6%
No	92.40%
Tot. obs. in marriage market	10,561,324

Note: Computed using the ACS 2008-2017. Observations included in the marriage market are the sum of i) recently married individuals and ii) available individuals. Recently married individuals are those who got married within the past two years in a heterosexual union. Available individuals are all unmarried persons age 15 to 64 who did not get a divorce or become widowed in the past year.

month of the year. Marriage market participants are expected to be indifferent about their partners' month of birth (unless the person is a fierce believer in astrology and the zodiac).

Column (i) in Table 3 shows log estimates of the original CS marriage market function (7) for the four possible type combinations. This approach leads to the conclusion that the marriage gains of people born in the first quarter are lower than the marriage gains of people born in any other quarter or the year. However, this result is likely the consequence of the different composition of types in the market (as the example in section 2).

Column (ii) in Table 3 shows log estimates of the modified marriage market function (7). In this case, the marriage gains are almost identical for all combinations of types. Moreover, the marriage gains of each type combination are identical to the marriage gains of the market (where there is only one type per gender). Contrary to the original CS approach, column (ii) suggests that the quarter of birth is an irrelevant characteristic in the marriage market.

Table 3: Marriage gains for individual born in different quarters of the year

		Marriage gains	
		$\log\Pi_{ij}$	$\log\Pi_{ij}^*$
		<i>CS original</i>	<i>modified</i>
		(i)	(ii)
One type per sex			
<i>Men's type</i>	<i>Women's type</i>		
type 1	type 1	-2.50	-2.50
Two types (birth quarter)			
<i>Men's type</i>	<i>Women's type</i>		
born first quarter	born first quarter	-3.88	-2.47
born first quarter	born rest of year	-3.36	-2.51
born rest of year	born first quarter	-3.36	-2.51
born rest of year	born rest of year	-2.77	-2.49

Note: Computed using the ACS 2008-2017.

**Marriage gains across educational groups** The role of education in the marriage market has been of primary interest in economics since at least Becker (1973). Estimating marriage gains across educational groups is crucial for understanding the roots of assortative matching.

I consider two definitions of individual types to illustrate the importance of adjusting the original CS function. The first definition considers only two educational groups per sex: the unskilled (high school or less) and the skilled (at least some college). I combine these educational group with the five age categories described in Table 4, which gives a total of ten skill-age types per sex.

The second definition desegregates the skilled group in four types (some college with no degree, Associate's degree, Bachelor's degree, and Post-B.A. degree). This second definition has a total of twenty-five age-education types per sex.

Table 4 compares the results of the original CS approach (panel A) and that proposed in this paper (panel B). For the sake of clarity, the table only shows the marriage gains of individuals 25 to 34 years old. The  $2 \times 2$  matrices show the results of the first definition of types previously described. The  $5 \times 5$  matrices show the results of the second (more disaggregated) definition of types. The fact that most values in the table are negative should not be a concern. The magnitudes depend on how the market is defined. The relevant information is on the relative values.<sup>7</sup>

The comparison of results indicates that the systematic marriage gains in couples where both members are unskilled are identical across the two definitions (-3.15 for the original CS

<sup>7</sup>When the number of unmarried individuals is larger than the number of matches, marriage gains are negative.

Table 4: Marriage gains across methodological approaches and type definitions (25-34 years old)

<b>Panel A: Original CS marriage matching function - marriage gains <math>\ln(\Pi_{ij})</math></b>						
		Women				
		unskilled	skilled			
Men	unskilled	-3.15	-3.39			
	skilled	-4.00	-1.98			
		Women				
		Unskilled	Some coll.	Assoc.	B.A.	Post B.A.
Men	Unskilled	-3.15	-3.62	-4.05	-4.20	-5.09
	Some college	-4.13	-3.24	-3.66	-3.42	-4.14
	Associate's degree	-4.73	-4.05	-3.58	-3.59	-4.24
	Bachelor's degree	-5.01	-4.13	-3.97	-2.36	-2.74
	Post B.A.	-5.91	-5.16	-4.85	-2.90	-2.13

<b>Panel B: Modified marriage matching function - marriage gains <math>\ln(\Pi_{ij}^*)</math></b>						
		Women				
		Unskilled	Skilled			
Men	Unskilled	-0.69	-1.33			
	Skilled	-1.71	-0.09			
		Women				
		Unskilled	Some coll.	Assoc.	B.A.	Post B.A.
Men	Unskilled	-0.69	-1.04	-0.99	-1.64	-2.10
	Some college	-1.38	-0.36	-0.30	-0.56	-0.85
	Associate's degree	-1.40	-0.60	0.35	-0.16	-0.38
	Bachelor's degree	-2.21	-1.21	-0.56	0.55	0.60
	Post B.A.	-2.55	-1.68	-0.88	0.56	1.75

Note: Computed using the ACS 2008-2017. Types are defined as the interaction of five age groups and skill level. Only results for 25-34 year-olds are displayed.

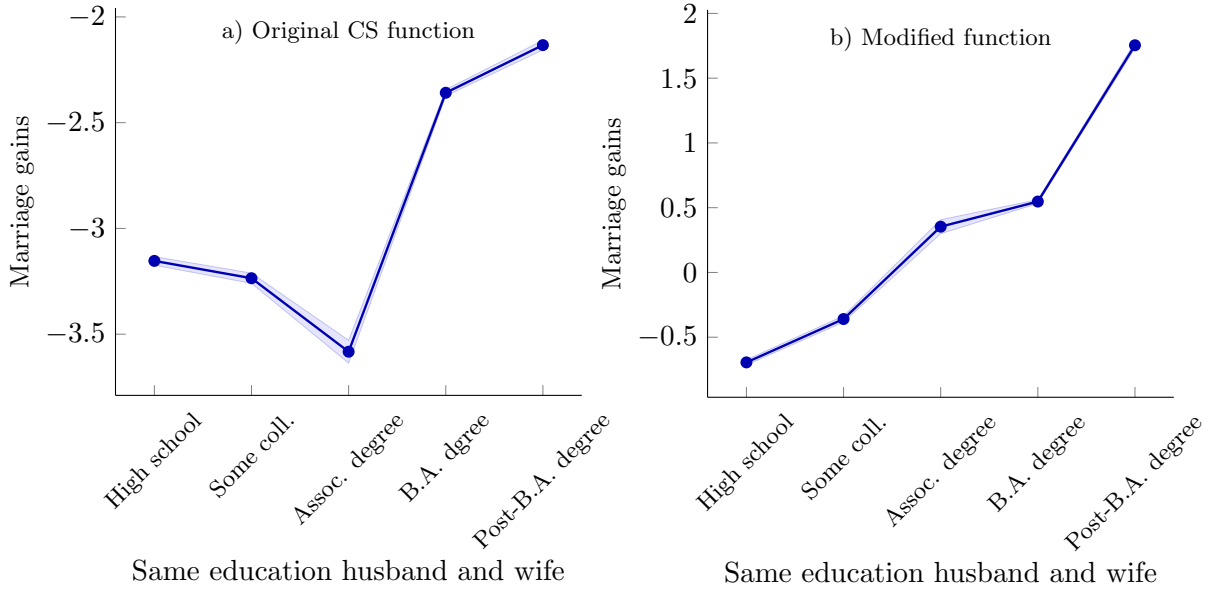
approach in panel A, and -0.69 for the modified version in panel B). This result is expected given the multinomial logit structure of the problem. However, the original CS approach shows inconsistencies when unskilled individuals marry skilled spouses.

The  $2 \times 2$  matrix in panel A indicates that the systematic marriage gains are -3.39 when an unskilled man marries a skilled woman. This value is greater than each off-diagonal value in the first row of the  $5 \times 5$  matrix in panel A (when skilled individuals are sub-classified in four types). This result is contradictory. The gains of marrying a skilled woman cannot be simultaneously higher than all marriage gains associated with each of the sub-groups that constitute the skilled type.

The problem just mentioned is the consequence that the CS matching function is not

invariant to the number of types chosen by the researcher. When types get sub-divided, the CS approach tends to show a reduction in the marriage gains for everyone. The modified approach in (8) overcomes this issue. The  $2 \times 2$  matrix in panel B indicates that the systematic marriage gains when an unskilled man marries a skilled woman (-1.33) are comparable to the mid-value of off-diagonal cells in the first row of the  $5 \times 5$  matrix.

Figure 2: Marriage gains across educational groups (25-34 years old)



Note: The values correspond to the diagonal of the matrices in Table 4. Shaded areas are 95% confidence intervals computed with 300 clustered bootstrap replications. The clusters variable is the household id.

Figure 2 plots marriage gains when couple’s members have the same level of education (i.e., the diagonal of the  $5 \times 5$  matrices in Table 4). The curve that displays the results of the CS approach is non-monotonic. For example, the marriage gains of couples with a 2-year Associate’s degree are lower than the marriage gains of couples in which both members are unskilled. Although not implausible, this pattern in the data is difficult to explain from a theoretical perspective.

Contrary to the CS approach, the curve that I derive in Figure 2b) is monotonically increasing. Some theoretical papers argue that assortative matching and homophily can both explain the tendency of individuals to marry members of the opposite sex with similar levels of education (e.g., Gihleb and Lang (2016)). Becker (1973) indicates that marriage gains should be supermodular in education to observe a positive assortative matching in this dimension. Supermodularity implies increasing marriage gains as in Figure 2b). On the other hand, educational homophily predicts that marriage gains should decline with the ‘distance’



in the educational groups of the husband and the wife. The patterns observed within rows (or columns) in the  $5 \times 5$  matrix of panel B are consistent with this prediction.

## 5 Conclusions

This paper solves the problem of using irrelevant traits when defining individual types in Choo and Siow model. As a result, the original CS marriage matching function suffers modifications. In particular, it incorporates a multiplicative adjustment term. The approach in this paper has the potential to solve the independence of irrelevant alternatives problem in general settings when the relative supply of choices is observable. More research is needed in this area.

## References

- Becker, G. S. (1973). A theory of marriage: Part i. *Journal of Political economy* 81(4), 813–846.
- Choo, E. (2015). Dynamic marriage matching: An empirical framework. *Econometrica* 83(4), 1373–1423.
- Choo, E. and A. Siow (2006). Who marries whom and why. *Journal of political Economy* 114(1), 175–201.
- Cornelson, K. and A. Siow (2016). A quantitative review of marriage markets: How inequality is remaking the american family by carbone and cahn. *Journal of Economic Literature* 54(1), 193–207.
- Gihleb, R. and K. Lang (2016). Educational homogamy and assortative mating have not increased. Technical report, National Bureau of Economic Research.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, edited by Paul Zarembka.
- Siow, A. (2015). Testing beckers theory of positive assortative matching. *Journal of Labor Economics* 33(2), 409–441.