Stochastic Sorting*

Hector Chade† and Jan Eeckhout‡

June 22, 2017

Abstract

We analyze matching markets with stochastic characteristics: agents' payoff-relevant characteristics are realized after matching takes place, and matches are formed based on ex-ante attributes that affect the distributions of ex-post ones. This generates *stochastic sorting* patterns and naturally gives rise to ex-post mismatch and correlation in realizations due to selection. We derive conditions for positive and negative sorting in the standard setting with transferable utility, as well as in the more complex case with nontransferable utility, such as risk sharing and moral hazard problems. In addition to complementarity in the match output function, the conditions for assortative matching now also depend on the properties of the stochastic order imposed on the distributions of the agents' ex-post characteristics. We provide two applications of the model: one analyzes mismatch in the labor market for executives, and the other decomposes the sources of increased inequality of married household into marital and stochastic sorting.

Keywords. Assortative Matching. Mismatch. Selection. Correlation. Stochastic Orders. Supermodularity. Executive Compensation. Household Inequality.

*We would like to thank Marco Bassetto, Nezih Guner, Sergei Guriev, Ilse Lindenlaub, Jeremy Lise, Alejandro Manelli, Costas Meghir, Fabien Postel-Vinay, and Jean-Marc Robin for insightful discussions and comments as well as seminar audiences at UCLA, Cal Tech, Chicago Booth, University of Arizona, Oxford, 2013 SED Seoul, NBER Summer Institute, Saint Louis Fed, Barcelona Summer Forum, Sciences Po, UCL, Stockholm, Basel, Melbourne, Nagoya, Hong Kong and Banco Central de Chile. Christoph Hedtrich, Alirezah Sepahsalari, Thomas Minten, and Patricia de Cea provided superb research assistance. Eeckhout gratefully acknowledges support from the ERC, Advanced Grant 339186. Chade is grateful to Universitat Pompeu Fabra for its hospitality, where part of this research was conducted.

†Arizona State University, hector.chade@asu.edu

‡University College London and Barcelona GSE-UPF, j.eeckhout@ucl.ac.uk.

1 Introduction

In many economic environments, match formation takes place with limited information. In the labor market, for example, it is often argued that labor is an experience good (Nelson (1970), Jovanovic (1979)): firms hire workers whose education level is known, but their true productivity is revealed only after a significant period of time has passed. In this paper, we extend the canonical Beckerian matching model to allow for stochastic characteristics. Agents are heterogeneous, but their characteristics upon matching are noisy predictors of their payoff-relevant ex-post attributes. This setting therefore affords a useful reinterpretation of the standard matching problem: each agent is characterized by a conditional distribution of the ex-post type, and thus the allocation problem can be thought of as one of matching distributions, or *stochastic sorting*.[1] In the labor market example, if the ex-ante characteristic is education and the ex-post one is productivity, then the distribution of productivity of those with an MBA from Chicago is different from that of those with a degree from Northern Illinois. Hence, with sorting based on education, matched partners are pairs of distributions over productivities. This stochastic sorting logic extends to other environments such as marriage, risk sharing or education, where there is ample evidence of uncertainty.[2]

An important feature of our model is that, even if there is assortative matching based on ex-ante types, there will be *equilibrium mismatch* in ex-post types. While on average they are less productive, some graduates from lower ranked schools are likely to stand out. Because they were matched based only on the average quality of their college, this implies that given the actual realization of their type, their allocation to a job is not optimal ex post. Matches are formed between pairs that ex post contradict the equilibrium properties of the deterministic Beckerian model. This is an appealing feature of the model since mismatch is a universal characteristic of the data in any application. Very often, this is dealt with by assuming there is some additive measurement error, drawn from a particular distribution.[3] In our setting, the noise not only determines the mismatch, but it is an *integral part* of the determination of equilibrium sorting patterns. This is the main contribution of our analysis.

By focusing on a static setup with one-dimensional characteristics, we can accommodate features that would be intractable in a search or a multidimensional model, such as risk sharing and moral hazard. In addition, we derive novel results about sorting patterns. For example, there can be negative sorting due to distributional properties despite the match surplus being supermodular. Also, *selection due to matching* gives rise to interesting patterns of dependence in ex-post types across matches.

[1] The Becker (1973) model, extended to incorporate search goods (Nelson (1970)'s other good category), has become one of the workhorse models in the labor literature (see for example Shimer and Smith (2000)). Yet, there is little foundational analysis of sorting with experience goods beyond some models with additive or match-specific shocks. In the seminal Jovanovic (1979) learning model, labor is both an experience good and a search good, but there is no sorting.

[2] See Athey, Katz, Krueger, Levitt, and Poterba (2007) and Conley and Önder (2014) who analyze the performance of economics PhD students and assistant professors and who find that there is a substantial amount of uncertainty. For example, the top graduate from North Carolina performs better as an assistant professor than the number two or three from MIT, even though the distribution of MIT graduates dominates that of North Carolina.

[3] While we do not per se think of the stochastic outcomes as measurement error, they can be interpreted as such.

We derive general conditions on distributions as well as on the properties of the match payoff function that induce monotone matching (positive or negative assortative matching). We do so both in the standard transferable utility (TU) world, where agents transfer utility at a constant rate, and in the nontransferable utility (NTU) case, where the rate at which agents transfer utility can vary. Intuitively, the results differ depending on the stochastic order assumed for the conditional distributions of ex-post types. For the two textbook orders – first-order stochastic dominance (FOSD) and increase in risk (IR) – we derive appropriate complementarity properties of the match payoff that yield monotone sorting.

We first derive a general characterization result on sorting patterns under TU, which involves conditions on match output and the ex-post distribution of types that lead to positive assortative matching (PAM) and negative assortative matching (NAM). In particular, when we specialize the result to conditionally independent ex-post types whose distributions are ordered by FOSD, we obtain that PAM (NAM) emerges if and only if the match output function is supermodular (submodular) in ex-post types. This subsumes Becker's deterministic model as a special case when distributions are degenerate, for then FOSD holds trivially. If instead the stochastic order is IR, then PAM (NAM) obtains if and only if match output complementarities (that is, its cross partial derivative with respect to ex-post types) is supermodular, (submodular) in ex-post types. We provide an interpretation in terms of risk attitudes when agents differ by the riskiness in their distributions.

We pay particular attention to the *stochastic sorting patterns* that emerge in our settings, since they are useful in economic applications.[4] Suppose that an econometrician only has data on ex-post types realizations of matched pairs. Under FOSD, she would find a cloud of pairs of ex-post types that have a monotone trend (positive if PAM, negative if NAM). That is, the covariance of ex-post types will be positive for PAM and negative for NAM, despite the absence of correlation within each matched pair. This is due to the selection induced by the matching pattern ex ante. The pattern is more subtle for IR, since under PAM high riskiness individuals are matched together and so are low riskiness ones. Therefore, what an econometrician will observe is that high or low realizations of ex-post types on one side of the market are associated with similarly extreme realizations on the other side, while intermediate realizations are clustered with intermediate ones. The opposite is true under NAM.

We next turn to the more challenging case of NTU where agents can transfer utility but at varying rates, so a matched pair faces a nonlinear Pareto frontier (see Legros and Newman (2007)). The presence of uncertainty in the model brings to the forefront a risk-sharing problem between risk averse partners, namely, that of dividing the match output in a way that shares risks efficiently. Following the work by Ackerberg and Botticini (2002) on landowners in early Renaissance Tuscany, it has long been recognized that an important motive for pairwise matching is risk sharing. Schulhofer-Wohl (2006), Chiappori and Reny (2005) and Legros and Newman (2007) have modeled matching under uncertainty among risk averse agents. Whereas this literature so far has been concerned with the allocation problem of agents with different *preferences*, we analyze the distinctive case of matching agents with different conditional

[4] What follows apply to both TU and NTU models.

distributions of ex-post types, i.e., they differ in their risky *endowments*. We derive a general result for the class of hyperbolic absolute risk aversion utility functions (HARA), a tractable and general class of utility functions that subsumes as special cases the most common utility functions used in applications. Moreover, it allows us to solve for the sorting patterns using similar tools as in the TU case, which dovetails well with the analysis above. We show that monotone sorting still depends on the properties of ex-post types distributions and the match output function, suitably modified to account for risk aversion. We derive several classes of primitives under which PAM or NAM obtains.[5]

Besides risk sharing, the literature on NTU models has also paid attention to how incentives interact with matching (see Ackerberg and Botticini (2002), Serfes (2006), and Legros and Newman (2007)). We analyze another NTU setting that adds moral hazard to the risk sharing problem. Keeping in mind the applied motivation that guides this project, we build a tractable and empirically implementable variation of the principal-agent model of Holmström and Milgrom (1987), which once again allows us to analyze sorting in a way similar to the TU case above. We embed it in a matching problem where principals who are heterogeneous in their technology (both ex ante and ex post) match with workers who are heterogeneous in their marginal productivity of effort (both ex ante and ex post). Principals match with agents before the realization of their ex-post types and sign a contract. Since effort is not contractible, the contract is contingent on the realization of output and hence the firm type. Optimal contracting pins down the ex-ante matching pattern as well as the ex-post payoffs. We derive explicit expressions for wages and profits, and the conditions under which there is positive or negative assortative matching. The distribution of worker and firm types jointly determine whether there are complementarities in output, as well as the extent of the ex-post mismatch. This setup is well-suited for identifying the technological features of a matching market. Unlike the deterministic matching model, there is variation in the observed match outcomes, because mismatch occurs in equilibrium. As a result, once we know the technology – the match surplus function as well as the distribution – we can quantify the extent of the complementarities between different matched types.

We illustrate our stochastic sorting model with two applications. First, we use the principal-agent set up to analyze the matching of CEOs to firms. Using Compustat data on CEO compensation and firm valuation, we estimate the underlying technology, match surplus and distribution of ex-post types. Like in Terviö (2008) and Gabaix and Landier (2008), we find that the wage variation for executives is mainly driven by the variation in the characteristics of the firms where they work. The CEOs themselves are only marginally different, but a CEO whose decisions affect the sales of a firm that is 100 times larger will as a result have an equally proportional impact on firm revenues compared to a CEO in the small firm. As a result, these marginal differences between CEOs translate into large differences in compensation. But skills and performance are also subject to uncertainty. The novel insight here is

[5]In independent work, Li, Sun, Wang, and Yu (2016) also analyze risk sharing and matching among agents who differ in their distribution of income and focus on HARA as we do. They derive a different set of results based on either perfect correlation or conditional independence with distributions by second-order stochastic dominance. Moreover, we allow for a more general production structure within a match. Hence, given the negligible overlap, our results complement theirs.

that there is a significant amount of mismatch due to this uncertainty, which is costly in terms of lost output. While CEOs are fairly similar compared to each other ex ante, their ex-post realization is very different from what was expected ex ante. Our results indicate that firms should be more concerned about the adequate selection of CEOs than about incentive provision.

Second, we use the general framework to analyze the determinants of the rise in household inequality of married couples. We perform a standard decomposition exercise that distinguishes between the sources of inequality that are due to marital sorting based on ex-ante characteristics (education) and stochastic sorting based on ex-post characteristics (earnings). We find that over 80% of the increase in household inequality is due to stochastic sorting. This is mainly driven by the fact that the variance of individual earnings – and to a lesser extent, the correlation of earnings – have increased significantly. The role of marital sorting is much less pronounced, and it is mainly driven by the fact that females have overtaken males in educational achievement. This accounts for 13%. ex-ante mismatch has barely changed and hence does not contribute to the change in household income inequality.

Our work is related to a vast literature on assortative matching. Existing matching models incorporate mismatch in different ways. First, models of random search with two-sided heterogeneity exhibit equilibrium mismatch (e.g., Shimer and Smith (2000) or Gautier and Teulings (2004)). This is due to the inability to meet new trading partners fast enough, and to the random and undirected nature of meetings between agents. Given the opportunity cost of delay, agents are willing to accept a less than perfect partner. The implication of these models is that while there is mismatch, the model predicts a sharp demarcation of the range of matches: mismatches occur only within a region bounded below by the reservation type of each agent and above by the largest type that accepts each agent. This is often hard to reconcile with observed data. Chade (2006) gets beyond this by considering a matching model with search frictions and noisy types, albeit in a strictly nontransferable utility setting.[6] Second, mismatch may be due to unobserved heterogeneity. For example, when types are multidimensional and at least one dimension is not observed to the econometrician, then the observed outcome appears like mismatch (see for example Choo and Siow (2006), Galichon and Salanié (2010) and Lindenlaub (2013)). From the agents' viewpoint, however, there is no mismatch since they observe the entire bundle of the characteristics of the partners. By contrast, in our model ex ante and in expectation there is perfect matching, but ex-post mismatch is realized also for the matched agents.

Finally, there are a number of interesting issues that we have not analyzed, but that are nonetheless very promising for future research. First, we do not allow for the possibility of rematching. Upon the realization of mismatch of their ex-post types, agents would typically mutually prefer to form new matches if allowed to do so. As a result, the existing realized matches would be unstable. When such

[6] A search model that is also related to our setting is Cheremukhin, Restrepo-Echavarria, and Tutino (2016), who analyze targeted search where agents choose distributions over possible match partners, which results in search frictions and mismatch. Their mismatch, which is the result of the strategic choice of agents, is stochastic in nature, like ours. Other models with dynamic matching and shocks but no frictions such as those in Anderson and Smith (2010) and Anderson (2015) are stochastic in nature but they do not exhibit mismatch. In the latter two papers conditions for positive sorting are not simply supermodularity because types change depending on who the matched partner is.

rematching is completely frictionless, the problem trivially reduces to a sequence of Beckerian static matching models. If instead we realistically assume that there is some cost, then only those for whom the mismatch is sufficiently large will be willing to incur the cost, and we conjecture that our insights will be robust in this case. This is akin to the search model of Eeckhout and Kircher (2011) who use mismatch to identify complementarities. Second, we assume there is complete information: uncertainty is symmetric as no agent knows anyone's ex-post type. The extension to asymmetric information seems important to pursue, but it is also likely to be nontrivial given the well-known difficulties (see for example Guerrieri, Shimer, and Wright (2010)) of matching models with incomplete information.

2 The Model Setup

Consider a matching market with two populations (workers and firms, or men and women) each of measure one, whose agents seek to match pairwise for productive purposes with agents from the other population. These populations are heterogeneous: in one of them each agent has a trait $x \in [\underline{x}, \overline{x}]$, distributed according to the cumulative distribution function (cdf) $\Gamma(x)$; similarly, in the other population an agent's characteristic is a scalar $y \in [\underline{y}, \overline{y}]$ distributed according to $\Psi(y)$.

We call x and y the agents' ex-ante traits, which are observable at the matching stage. There is, however, uncertainty about agents' ex-post attributes; they are payoff-relevant and realized after matching takes place.[7] A pair (x, y) draws ex-post types (ω, σ) from $[\underline{\omega}, \overline{\omega}] \times [\underline{\sigma}, \overline{\sigma}]$ distributed according to a cdf $H(\cdot, \cdot|x, y) : \mathbb{R}^2 \to [0, 1]$ for each pair (x, y), given by $H(\omega, \sigma|x, y)$ with density $h(\omega, \sigma|x, y)$ for all (ω, σ).[8] The marginal distributions are given by $F(\omega|x) = \lim_{\sigma\to\infty} H(\omega, \sigma|x, y)$ with density $f(\omega|x)$ for all ω, and by $G(\sigma|y) = \lim_{\omega\to\infty} H(\omega, \sigma|x, y)d\omega$ with density $g(\sigma|y)$ for all σ. That is, we impose the restriction that the marginal distributions are independent of the ex-ante trait of a partner. Sometimes we will specialize the analysis further and focus on the conditional independence case $h(\omega, \sigma|x, y) = f(\omega|x)g(\sigma|y)$. Also, for convenience we assume that f, g, and h and twice continuously differentiable.[9]

These ex-post types determine match output given by a function $q : [\underline{\omega}, \overline{\omega}] \times [\underline{\sigma}, \overline{\sigma}] \to \mathbb{R}_+$, so that $q(\omega, \sigma)$ for each pair of realizations (ω, σ). This function is four times continuously differentiable, nonnegative, and strictly increasing in each argument.[10] For instance, in the marriage application in Section 4.2, x and y are the partners' education, while ω and σ denote their incomes. Education is observable at the time of the match, and is a noisy predictor of a partner's potential income.[11]

[7]For convenience, we will use the terms trait, attribute, characteristic, and type interchangeably.

[8]The compact support can be relaxed as long as the density is bounded, something we do sometimes below.

[9] Although we focus on the continuum of signal realizations and ex-ante characteristics, it is straightforward to replace integrals by sums plus some minor technical details to adapt all the results to the discrete case as well. Also, some of the strong differentiability assumptions we impose can be relaxed at the cost of more technical detail.

[10]The results are easier to interpret when q is not a function of the ex-ante traits, but the model can accommodate this case as well and we do so in the CEO application below. For example, if x indexes the university an MBA graduates from and ω is a measure of realized productive ability, then output produced can be determined not only by ω, but also by the initial type x: having gone to Harvard provides access to a network that affects future productivity.

[11]In our model with no rematching or continuation after the realization of the types, one can interpret the type as either match specific or as permanent (just like the distinction between firm-specific and general human capital). Also, we could

Matching is based exclusively on the ex-ante traits x, y, taking into account the distribution of ex-post types ω, σ. We assume that agents can freely make contingent transfers among themselves (that is, contingent on the possible realizations of the ex-post types), although they need not be able to transfer utility at a constant rate. Indeed, we will examine instances of both transferable and nontransferable utility (TU and NTU). Our equilibrium notion is standard, namely, stability, which is equivalent to the core in our setting. That is, an equilibrium will consist of a matching of the two populations and an assignment of (expected) utilities such that no individual agent or pair can block the matching. This allocation, moreover, can be decentralized as a competitive equilibrium. Let $\mu : [\underline{x}, \overline{x}] \to [\underline{y}, \overline{y}]$ denote the (measure-preserving) equilibrium matching function. If μ is monotone in x, then there is positive assortative matching (PAM) when μ is increasing, and negative assortative matching (NAM) when $\mu(x)$ is decreasing. Under PAM, from market clearing, the allocation satisfies $\Gamma(x) = \Psi(\mu(x))$ for all x. Under NAM, it satisfies $\Gamma(x) = 1 - \Psi(\mu(x))$ for all x.[12]

Notice that agents match based on x and y, taking into account the distributions of ex-post types. An alternative *interpretation* of the model is that it matches distributions $\{F(\cdot|x)\}_{x\in[\underline{x},\overline{x}]}$ and $\{G(\cdot|y)\}_{y\in[\underline{y},\overline{y}]}$. Clearly, how we order the distributions (as functions of x and y, respectively) will affect the conditions for PAM and NAM. We will focus on the two most commonly used stochastic orders in economics, first-order stochastic dominance (FOSD) and increase in risk (IR).[13] FOSD is a natural generalization of the standard matching model à la Becker with scalar types to the case of stochastic types. Under FOSD, better types x have distributions that are on average better. In Becker, the type distribution is degenerate and increasing types x and y trivially satisfy FOSD. IR instead is a notion of higher dispersion in outcomes, and thus of increasing risk.

3 Stochastic Sorting under TU and NTU

We first analyze the case of TU and derive conditions for sorting, which are fairly intuitive with FOSD and IR, both commonly used orders in economic applications. If the applied researcher is only interested in analyzing the ex-ante sorting patterns, the TU setting is most insightful and we use it to analyze the contribution of marital sorting to household inequality in one of the applications below. If, however, one also needs properties of ex-post transfers, then TU is not ideal, as utility is linear in money and thus only *expected* transfers are pinned down. We therefore analyze the NTU setting, which delivers the functional form of ex-post transfers as part of the solution. The presence of uncertainty implies

interpret ex-ante and ex-post attributes backwards: for example, x could be a signal of an unknown attribute ω.

[12]In principle, μ can be correspondence or a random variable. Since our interest is in positive and negative assortative matching, as is most of the matching literature, we define it as a function. Also, notice that a matching is a mapping between types instead of agents. But in our set up it is easy to show that in equilibrium different agents with the same types behave in the same way. Hence there is no loss of generality in defining a matching between types.

[13]Given two cdf's F and G, we say that F dominates G in FOSD if $F(z) \le G(z)$ for all z or, equivalently, if $\int u(z)dF(z) \ge \int u(z)dG(z)$ for all increasing functions u. Similarly, F is an IR of G if $\int zdF(z) = \int zdG(z)$ and $\int^t F(z)dz \ge \int^t G(z)dz$ for all t or, equivalently, if $\int u(z)dF(z) \le \int u(z)dG(z)$ for all increasing and concave functions u.

that NTU entails an efficient risk sharing problem. We derive results for a large class of risk sharing problems. In addition, we enrich the NTU environment by adding incentives, and study a model that matches principals and agents under moral hazard. The latter provides us with a flexible framework for empirical work in the labor market, which we apply to executive compensation below.

3.1 General Results with Transferable Utility

Assume agents can transfer utility within a match at a constant rate (quasilinear preferences, linear Pareto frontier), and can set up transfers contingent on the realizations of (ω, σ). Notice that, at the matching stage, all that agents care about is the expected value of those transfers.

The role of these transfers will be implicit in the analysis, since it is well known that under TU, the optimal/equilibrium matching maximizes the total payoff of the economy, i.e., the sum of the match value of all the pairs. Moreover, the sorting pattern that ensues depends on the properties of the match value of each pair (x, y), denoted by $V(x, y)$. In our set up with uncertainty this function is given by:

$$V(x, y) = \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} q(\omega, \sigma) h(\omega, \sigma | x, y) d\sigma d\omega. \tag{1}$$

Thus, optimal matching solves $\max_{\mu} \int_{\underline{x}}^{\overline{x}} V(x, \mu(x)) d\Gamma(x)$ over all possible assignments μ of the agents of the two sides of the market. And the core is a pair $(u(x), v(y))_{x,y}$ that satisfies individual rationality and pairwise stability. The Pareto frontier of a pair (x, y) when y receives utility v is given by

$$\Phi(x, y, v) = \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} q(\omega, \sigma) h(\omega, \sigma | x, y) d\sigma d\omega - v,$$

where the additive separability in v is a reflection of TU.

Our first result provides conditions for PAM and NAM in this setting.[14]

Proposition 1 (Sorting)**.** *Optimal sorting patterns are as follows:*

1. *PAM if H is spm (sbm) in (x, y) for each (ω, σ) and q is spm (sbm) in (ω, σ);*

2. *NAM if H is spm (sbm) in (x, y) for each (ω, σ) and q is sbm (spm) in (ω, σ);*

3. *PAM if $\int_{\underline{\omega}}^{\omega} \int_{\underline{\sigma}}^{\sigma} H(s, t | x, y) d\sigma d\omega$ is spm (sbm) in (x, y), $\int_{\underline{\omega}}^{\overline{\omega}} H(s, \sigma | x, y) ds$ is independent of x for all σ, and $\int_{\underline{\sigma}}^{\overline{\sigma}} H(\omega, s | x, y) ds$ is independent of y for all ω, and if $q_{\omega\sigma}$ is spm (sbm) in (ω, σ);*

4. *NAM if $\int_{\underline{\omega}}^{\omega} \int_{\underline{\sigma}}^{\sigma} H(s, t | x, y) d\sigma d\omega$ is spm (sbm) in (x, y), $\int_{\underline{\omega}}^{\overline{\omega}} H(s, \sigma | x, y) ds$ is independent of x for all σ, and $\int_{\underline{\sigma}}^{\overline{\sigma}} H(\omega, s | x, y) ds$ is independent of y for all ω, and if $q_{\omega\sigma}$ is sbm (spm) in (ω, σ).*

The conditions on q are also necessary if it is to hold for all H.

[14] Recall that a twice continuously differentiable function $z : [a, b] \times [c, d] \to \mathbb{R}$ is supermodular (spm) in (x, y) if $z_{xy} \geq 0$ and strictly so if $z_{xy} > 0$; it is submodular (sbm) if $z_{xy} \leq 0$ and strictly sbm if $z_{xy} < 0$.

All proofs are in the Appendix. It is easy to give a sketch of the proof here.[15] Integrating (1) by parts twice and then differentiating yields

$$V_{xy} = \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} q_{\omega\sigma}(\omega,\sigma) H_{xy}(\omega,\sigma|x,y) d\sigma d\omega.$$

If H is spm in (x,y) then so is V if q is spm. And V is spm in (x,y) for all H spm in (x,y) then q must be spm. Similarly for H sbm in (x,y). Part (i) of the proposition follows.

Regarding part 3. and 4., another integration by parts yields, after differentiation

$$V_{xy} = \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} q_{\omega\omega\sigma\sigma}(\omega,\sigma) \left(\int_{\underline{\omega}}^{\omega} \int_{\underline{\sigma}}^{\sigma} H_{xy}(s,t|x,y) ds dt \right) d\sigma d\omega$$

Thus, if $\int_{\underline{\omega}}^{\omega} \int_{\underline{\sigma}}^{\sigma} H(\omega,\sigma|x,y) d\sigma d\omega$ is spm in (x,y) for each (ω,σ), then so is V if (and only if) $q_{\omega\sigma}$ is spm. Similarly if the double integral of H is sbm in (x,y) for each (ω,σ). Simple algebra reveals that, if realizations are conditionally independent, then two important special cases subsumed by Proposition 1 are the FOSD and the IR cases. We now flesh out their sorting predictions.

Assume that $H(\omega,\sigma|x,y) = F(\omega|x)G(\sigma|y)$ and consider first the case where $F(\omega|\cdot)$ is decreasing in x for each ω and $G(\sigma|\cdot)$ is decreasing in y for each σ. That is, $\{F(\cdot|x)\}_{x\in[\underline{x},\overline{x}]}$ is ordered by FOSD, and similarly for $\{G(\cdot|y)\}_{y\in[\underline{y},\overline{y}]}$. Then $H(\omega,\sigma|\cdot,\cdot)$ is supermodular in (x,y) for each (ω,σ), and Proposition 1 parts 1. and 2. hold. In particular, when F and G concentrate all of their mass at one point for each x and y, respectively, the result reduces to the standard PAM and NAM in Becker (1973).

What are the properties of the distribution of ex-post types (ω,σ) under PAM or NAM with FOSD? Let μ be monotone in x (PAM or NAM). The unconditional cdf of (ω,σ) in equilibrium is

$$H(\omega,\sigma) = \int_{\underline{x}}^{\overline{x}} F(\omega|x)G(\sigma|\mu(x)) d\Gamma(x), \tag{2}$$

and the density and marginals are denoted by $h(\omega,\sigma)$, $h_1(\omega)$, and $h_2(\sigma)$, respectively. Easy algebra then reveals that the covariance between ω and σ under μ is

$$\begin{aligned} \mathrm{Cov}[\omega,\sigma] &= \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} \omega\sigma h(\omega,\sigma) d\sigma d\omega - \int_{\underline{\omega}}^{\overline{\omega}} \omega h_1(\omega) d\omega \int_{\underline{\sigma}}^{\overline{\sigma}} \sigma h_2(\sigma) d\sigma \\ &= \int_{\underline{x}}^{\overline{x}} \mathbb{E}[\omega|x]\mathbb{E}[\sigma|\mu(x)] d\Gamma(x) - \int_{\underline{x}}^{\overline{x}} \mathbb{E}[\omega|x] d\Gamma(x) \int_{\underline{x}}^{\overline{x}} \mathbb{E}[\sigma|\mu(x)] d\Gamma(x) \\ &= \mathrm{Cov}\left[\mathbb{E}[\omega|x], \mathbb{E}[\sigma|y]\right], \end{aligned} \tag{3}$$

which is nonnegative under PAM since by FOSD the expectations inside the covariance operator are increasing in x and y, respectively. Similarly, under NAM one expectation is increasing and the other

[15] The proposition could be alternatively be proved in a much more general way building on Theorem 3 Athey (1998). We impose additional structure that affords an elementary calculus proof.

one is decreasing, and thus the covariance is negative. So PAM or NAM based on ex-ante types lead to positive or negative dependence between the ex-post types, a *stochastic sorting* property.

A stronger testable implication under PAM obtains if we strenghten FOSD to log-spm F in (ω, x) and G in (σ, y).[16] For then H is log-spm in (ω, σ) (Karlin and Rinott (1980)) and thus $H_1(\omega|\sigma)$ and $H_2(\sigma|\omega)$ are *ordered* by FOSD. That is, under PAM, the cdf of ex-post types σ's associated with a given realization of ω is increasing in FOSD sense, and thus this stochastic sorting property implies that agents with better ex-post types are matched with stochastically better agents' ex-post types.

Figures 1a and 1c provide an illustration of stochastic sorting under FOSD. If an econometrician only has information about the ex-post types of the matched pairs, then she will observe mismatches scattered at both sides of the appropriate 'diagonal.'

Assume now that instead of FOSD we have IR. That is, $\int_0^1 F(\omega|x)d\omega$ is constant in x (constant mean) and $\int_0^t F(\omega|x)d\omega$ decreases in x, and thus the family $\{F(\cdot|x)\}_{x\in[0,1]}$ is ordered by IR (if $x' > x$ then $F(\cdot|x)$ is an IR of $F(\cdot|x'))$. Similarly, impose analogous assumptions on the family $\{G(\cdot|y)\}_{y\in[0,1]}$. As a result, $\int_0^1 H(s,\sigma|x,y)ds = G(\sigma|y)\int_0^1 F(s|x)ds$ is constant in x, $\int_0^1 H(\omega,s|x,y)ds = F(\omega|x)\int_0^1 G(\sigma|y)ds$ is constant in y, and $\int_0^\omega \int_0^\sigma H(s,t|x,y)d\sigma d\omega = \int_0^\omega F(s|x)ds \int_0^\sigma G(t|y)dt$ is supermodular in (x,y). Hence, Proposition 1, parts 3. and 4. apply, and $q_{\omega\sigma}$ determines PAM or NAM.

For some intuition, consider the special case $q(\omega,\sigma) \equiv Q(\omega+\sigma)$. Then a negative fourth derivative is the higher-order risk avoidance property called temperance: thus, $q_{\omega\sigma}$ submodular in (ω,σ) is equivalent to $Q^{iv} < 0$,[17] and in this case NAM, which minimizes the spread across pairs, is optimal.

Regarding the covariance of (ω,σ) under PAM or NAM, notice that $\mathbb{E}[\omega|x]$ and $\mathbb{E}[\sigma|\mu(x)]$ are constant in x and y. Hence, $\text{cov}(\omega,\sigma) = \text{Cov}\left[\mathbb{E}[\omega|x], \mathbb{E}[\sigma|y]\right] = 0$ under both PAM and NAM, so unlike FOSD, *monotone sorting is indistinguishable from random matching* under IR based on this measure. Nevertheless, there are still some *stochastic sorting* implications of this order, at least for some restricted domains, with a stronger measure of assortative matching. Recall that a mean preserving spread (MPS) is a special case of IR, characterized by the single crossing of any two cdf's in the ordered family (IR allows for multiple crossings). The joint unconditional density $h(\omega,\sigma)$ is

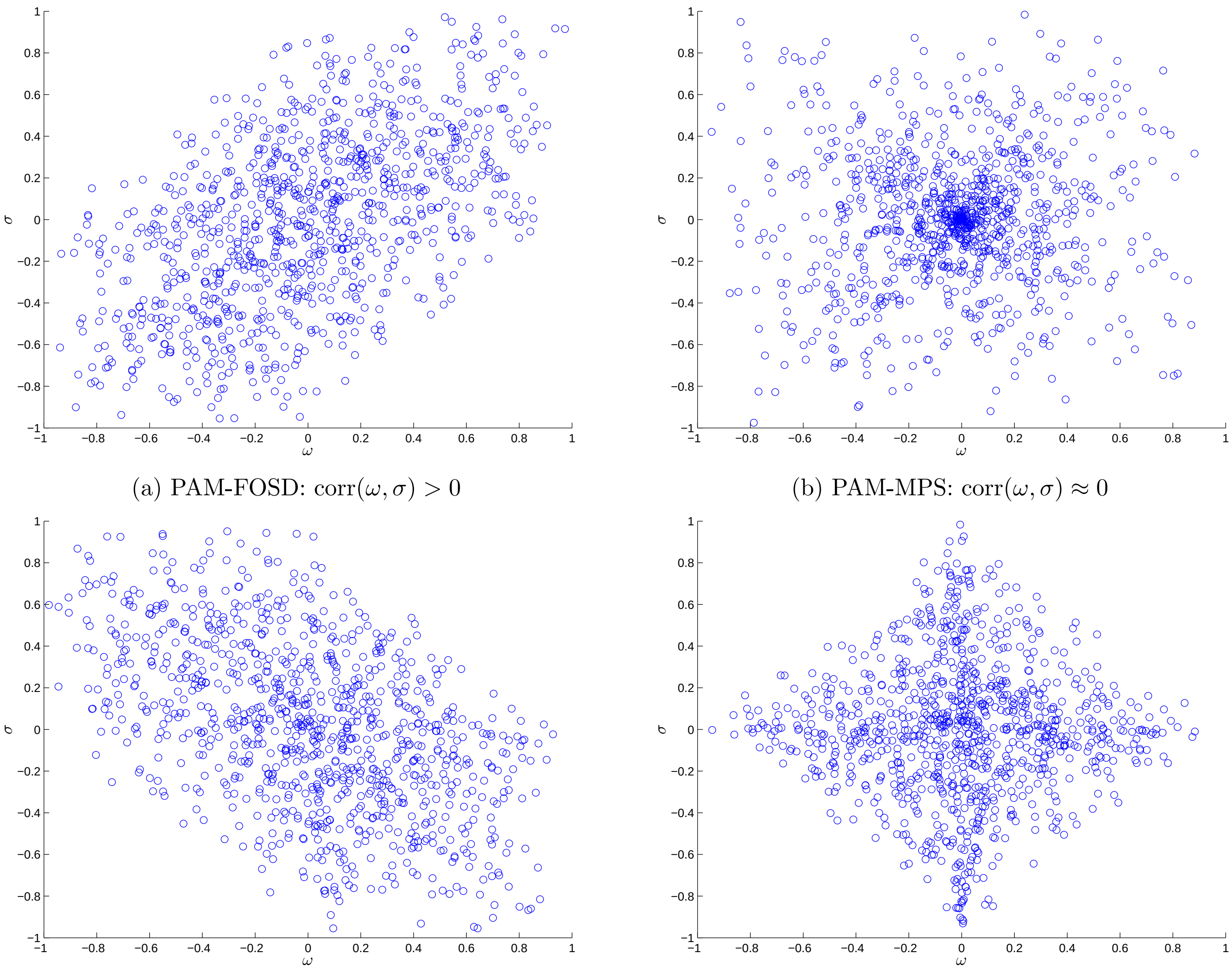
$$h(\omega,\sigma) = \int_{\underline{x}}^{\overline{x}} f(\omega|x)g(\sigma|\mu(x))d\Gamma(x), \tag{4}$$

with marginal $h_2(\sigma) = \int_0^1 g(\sigma|\mu(x))d\Gamma(x)$ and conditional $h_1(\omega|\sigma) = h(\omega,\sigma)/h_2(\sigma)$. An interesting question is whether anything can be said about $h_2(\omega|\sigma)$ as a function of σ when F is ordered by MPS as a function of x and G is ordered by MPS as a function of y.

As a first step, let us analyze the binary case. Let $x \in \{x_\ell, x_h\}$ and $y \in \{y_\ell, y_h\}$, identically distributed with $\gamma = \mathbb{P}(x = x_h) = \mathbb{P}(y = y_h) \in (0,1)$; if $x = x_i$, then $\omega \in \{\underline{\omega}_i, \overline{\omega}_i\}$, with $f(\overline{\omega}_i|x_i) = p$,

[16] A sufficient condition for F and G log-spm is that their associated densities are log-spm, or, equivalently, satisfy the monotone likelihood ratio property, which is satisfied by a large class of densities.

[17] The superscript iv denotes the fourth derivative.



(a) PAM-FOSD: $\mathrm{corr}(\omega, \sigma) > 0$

(b) PAM-MPS: $\mathrm{corr}(\omega, \sigma) \approx 0$

(c) NAM-FOSD: $\mathrm{corr}(\omega, \sigma) < 0$

(d) NAM-MPS: $\mathrm{corr}(\omega, \sigma) \approx 0$

Figure 1: Distribution of matched ω, σ realizations under PAM/NAM and with FOSD and MPS and ω, σ conditionally independent. Simulations with 1000 types. Under FOSD (both for PAM and NAM), x, y uniform on $[-.5, .5]$, and $\omega = x + \varepsilon, \sigma = y + \varepsilon$ where $\varepsilon_\omega, \varepsilon_\sigma$ are conditionally independent uniform draws on $[-.5, .5]$. Under MPS, (both for PAM and NAM), x, y uniform on $[0, 1]$, and $\omega = x \cdot \varepsilon, \sigma = y \cdot \varepsilon$ where $\varepsilon_\omega, \varepsilon_\sigma$ are conditionally independent uniform draws on $[-1, 1]$.

$i = \ell, h$; similarly, if $y = y_i$, then $\sigma \in \{\underline{\sigma}_i, \overline{\sigma}_i\}$, with $g(\overline{\sigma}_i | y_i) = q$, $i = \ell, h$. Finally, to ensure MPS, we require $\mathbb{E}[\omega | x_\ell] = \mathbb{E}[\omega | x_h]$, $\mathbb{E}[\sigma | y_\ell] = \mathbb{E}[\sigma | x_h]$, $\underline{\omega}_\ell < \underline{\omega}_h < \overline{\omega}_h < \overline{\omega}_\ell$ and $\underline{\sigma}_\ell < \underline{\sigma}_h < \overline{\sigma}_h < \overline{\sigma}_\ell$.

The goal is to compute $h_1(\omega|\sigma)$ under PAM and NAM. Notice that, for each of the four values of σ, this conditional distribution has support on $\{\underline{\omega}_\ell, \underline{\omega}_h, \overline{\omega}_h, \overline{\omega}_\ell\}$ and thus we can summarize it by a vector of four probabilities. Consider first PAM, so that $\mu(x_i) = y_i$, $i = \ell, h$. Then

$$\omega|_{\underline{\sigma}_\ell} \sim (1-p, 0, 0, p), \quad \omega|_{\underline{\sigma}_h} \sim (0, 1-p, p, 0), \quad \omega|_{\overline{\sigma}_h} \sim (0, 1-p, p, 0), \quad \omega|_{\overline{\sigma}_\ell} \sim (1-p, 0, 0, p).$$

For instance, conditional on $\underline{\sigma}_\ell$, under PAM we know that this agent's partner can only generate signal realizations $\underline{\omega}_\ell$ with probability $1-p$ and $\overline{\omega}_\ell$ with probability p. Thus, under PAM, as σ increases $h_1(\omega|\sigma)$ *exhibits first a decrease in spread when* σ *increases and then an increase.*

Consider now NAM, so that $\mu(x_i) = y_j$, $i \neq j = \ell, h$. Then

$$\omega|_{\underline{\sigma}_\ell} \sim (0, 1-p, p, 0), \quad \omega|_{\underline{\sigma}_h} \sim (1-p, 0, 0, p), \quad \omega|_{\overline{\sigma}_h} \sim (1-p, 0, 0, p), \quad \omega|_{\overline{\sigma}_\ell} \sim (0, 1-p, p, 0).$$

For instance, conditional on $\underline{\sigma}_\ell$, under NAM we know that this agent's partner can only generate signal realizations $\underline{\omega}_h$ with probability $1-p$ and $\overline{\omega}_h$ with probability p. Thus, under NAM, as σ increases $h_1(\omega|\sigma)$ *exhibits first an increase in spread when* σ *increases and then a decrease.*

Notice that although we have monotone matching based on ex-ante types (PAM or NAM), the stochastic properties of the distribution of matched ex-post types are much more subtle. Appendix A.2 generalizes the analysis to a continuum of signal realizations and shows the following intuitive property, which we conjecture holds for IR as well: under PAM, a very 'low' or very 'high' value of σ is more likely to come from a 'low' type y, who is paired with a 'low' type x; and since low types have a larger spread in their distributions, the conditional distribution of ω given a very low or high value of σ should exhibit more spread than those of 'middle' types. A similar intuition holds for NAM. Figures 1b and 1d provide an illustration based on simulations. In both cases the correlation is zero and under PAM we observe a pattern of ex-post realizations akin to the number five on a dice, while under NAM we observe a pattern of ex-post realizations that appears like a cross. The absence of some form of stochastic monotonicity in ex-post types *does not* allow the econometrician to reject assortative matching, as it could come from a setting where distributions are ordered by spread. We thus have:

Proposition 2. *Assume that* ω *and* σ *are conditional independent. Then the following results hold:*

1. *If* F *and* G *are ordered by FOSD, then* $Cov[\omega, \sigma] > 0$ *under PAM and* $Cov[\omega, \sigma] < 0$ *under NAM;*

2. *If* F *and* G *are log-spm, then under PAM* $\{H_1(\cdot|\sigma)\}_{\sigma \in [\underline{\sigma}, \overline{\sigma}]}$ *and* $\{H_2(\cdot|\omega)\}_{\omega \in [\underline{\omega}, \overline{\omega}]}$ *are ordered by FOSD in* σ *and* ω*, respectively;*

3. *If* F *and* G *are ordered by IR, then* $Cov[\omega, \sigma] = 0$*;*

4. *If* F *and* G *are ordered by MPS and ex-ante traits are binary, then under PAM the riskiness of* $\{H_1(\cdot|\sigma)\}_{\sigma \in [\underline{\sigma}, \overline{\sigma}]}$ *and* $\{H_2(\cdot|\omega)\}_{\omega \in [\underline{\omega}, \overline{\omega}]}$ *is higher for low and high ex-post types than for middle ones, and the opposite holds under NAM.*

Except for 4., whose proof is in Appendix A.2, the other results were proven above in the text. *It is worth noting that the FOSD and IR (or MPS) stochastic sorting predictions apply to both TU and NTU settings, as they depend solely on the distributions of ex-post types and monotone matching.*

We close with two comments about the model. The first one is about the possibility of *reordering* ex-ante types. Suppose for example that the order of traits is years of schooling, and that MBAs have less years of schooling than PhDs, yet MBAs get higher ranked jobs as executives and get higher wages. Suppose also that for any two graduates we can always pairwise rank the distribution F by FOSD. Despite the pairwise ordering, FOSD fails for the *entire* population of graduates: PhDs draw from a distribution of ex-post types that is dominated by that of the MBAs, even though PhDs have more years of schooling. We can nonetheless still reorder types according to the FOSD in which MBAs are ranked above PhDs. In that case, Proposition 1 still applies. The key observation here is that we can obtain this sorting outcome without changing the match output function q as long as it is a function only of the ex-post types. This is not the case in Becker's deterministic model. If the PhD is ranked higher in years of schooling but obtains lower ranked jobs and gets lower wages, the technology is at least locally submodular. We can of course also reorder MBAs and PhDs, but this will also require that we adjust the technology, which must now become supermodular.

The second comment highlights the importance of *both* technology and distributions. Unlike the standard Becker model where PAM obtains under supermodularity, q spm is not sufficient for PAM. As Proposition 1 establishes, we also need conditions on the distributions. For an easy example, consider a matching market with q given by $q(\omega, \sigma) = \omega\sigma^2$ and with conditionally independent signals where F is ordered by FOSD ($F_x < 0$), and G is normally distributed with mean $\mu(y)$ and variance $s^2(y)$, where $\mu'(y) > 0$ and $s'(y) < 0$. Then $V(x, y) = \mathbb{E}[\omega|x]\mathbb{E}[\sigma^2|y] = \mathbb{E}[\omega|x] \left(s^2(y) + \mu^2(y)\right)$ and thus

$$V_{xy}(x, y) = \mathbb{E}_x[\omega|x] \left(\frac{ds^2}{dy} + \frac{d\mu}{dy} 2\mu \right).$$

Hence, $V_{xy} < 0$ and there is NAM if $ds^2/dy < -(d\mu/dy)2\mu$, despite q being *supermodular* in (ω, σ).

3.2 Risk Sharing and Nontransferable Utility

We now turn to the more challenging setting with NTU (see Legros and Newman (2007)). Although agents can transfer utility among them, they cannot do so at a constant rate. Since matching is based on ex-ante types, agents face uncertainty at the matching stage. That uncertainty, coupled with the standard assumption in economic applications that agents are risk averse, invariably leads to a nontrivial risk sharing problem. Although there could be other decisions that a pair takes, we will assume that they are summarized in the output that is to be shared after the realization of the ex-post characteristics. As a result, we will focus on the properties of efficient risk sharing between partners (who can commit to ex-post transfers), and derive sufficient conditions for PAM or NAM. This will broaden the applications of our stochastic sorting model to, for instance, marriage markets, where risk sharing among partners is considered to be of significant importance.

Agents with ex-ante characteristics x and y match and commit to ex-post efficient sharing of the joint output q, which depends upon the ex-post characteristics of the pair (ω, σ). That is, $q(\omega, \sigma)$ is divided such that agent x consumes $c_x(\omega, \sigma)$ and y consumes $c_y(\omega, \sigma)$, with $c_x(\omega, \sigma) + c_y(\omega, \sigma) \leq q(\omega, \sigma)$. An important special case is when ω and σ are the agents' incomes, so that $q(\omega, \sigma) = \omega + \sigma$. But note that we allow for a more general interpretation of the ex-post types as inputs of a production function q.

Consider a pair (x, y), with $u_i : \mathbb{R} \to \mathbb{R}$ being the utility function of $i = x, y$, assumed strictly increasing and concave in consumption c_i. The risk sharing problem that a pair (x, y) solves is:

$$\begin{aligned}\Phi(x,y,v) \quad &= \quad \max_{c_x,c_y} \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} u_x(c_x(\omega,\sigma))h(\omega,\sigma|x,y)d\sigma d\omega \\ \text{s.t.} \quad & \quad c_x(\omega,\sigma)+c_y(\omega,\sigma) \le q(\omega,\sigma) \;\; \forall \;\; (\omega,\sigma) \\ & \quad \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} u_y(c_y(\omega,\sigma))h(\omega,\sigma|x,y)d\sigma d\omega \ge v,\end{aligned}$$

where v is the reservation utility of agent y (which is pinned down in equilibrium), and $\Phi(x, y, v)$ is the maximum expected utility from the match for an agent with type x who matches with an agent with type y whose reservation utility is v.

It is clear that both constraints will bind at the optimum, so the problem becomes

$$\begin{aligned}\Phi(x,y,v) \quad &= \quad \max_{c} \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} u_x(q(\omega,\sigma)-c(\omega,\sigma))h(\omega,\sigma|x,y)d\sigma d\omega \\ \text{s.t.} \quad & \quad \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} u_y(c(\omega,\sigma))h(\omega,\sigma|x,y)d\sigma d\omega = v.\end{aligned}$$

Let λ be the Lagrange multiplier associated with the constraint. Then maximizing pointwise we obtain:

$$u_x'(q(\omega,\sigma)-c(\omega,\sigma)) = \lambda u_y'(c(\omega,\sigma)), \;\; \forall\, (\omega,\sigma),$$

which along with the constraint determine the function c and the value of λ. The standard procedure in risk sharing problems is to solve for $c(\omega, \sigma, \lambda)$ for each $(\omega, \sigma, \lambda)$ from the first-order condition, insert the solution into the constraint, and find for each v the unique value of $\lambda(x, y, v)$ that satisfies it. Then the optimal consumption is $c^*(\omega, \sigma, x, y, v)$ for each $(\omega, \sigma, x, y, v)$. Hence,

$$\Phi(x,y,v) = \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} u_x(q(\omega,\sigma)-c^*(\omega,\sigma,x,y,v))h(\omega,\sigma|x,y)d\sigma d\omega. \tag{5}$$

Notice that $\Phi(x, y, \cdot)$ traces the Pareto frontier of the problem as we vary v.

Legros and Newman (2007) show that the properties of Φ determine when PAM or NAM is optimal, and derived sufficient (and necessary) conditions for monotone sorting. We will use a differential version of their condition (see Chade, Eeckhout, and Smith (2016) for details), which asserts that PAM is optimal if $\Phi_{xy} \ge (\Phi_y/\Phi_v)\Phi_{vx}$ for all (feasible) (x, y, v), and NAM is optimal if the inequality is reversed. Not only type-complementarity is important for sorting, but also type-utility complementarity, which determines how costly it is to transfer utility to a partner. In particular, if Φ is additively separable in v, then $\Phi_{vx} = 0$ and we obtain the usual $\Phi_{xy} \ge 0$ for PAM and $\Phi_{xy} \le 0$ for NAM.

At this level of generality, it is extremely difficult to derive results since Φ depends in a complicated fashion on its arguments.[18] One case that is immediate is when one party, say x, is risk neutral. In this case, it is well-known that x bears all the risk and gives y a constant consumption determined by $u_y(c^*) = v$, so $c^* = u_y^{-1}(v)$. Hence, $\Phi(x,y,v) = \int\int(q(\omega,\sigma)h(\omega,\sigma|x,y)d\sigma d\omega - u_y^{-1}(v) = V(x,y) - v'$, and we are essentially back in the TU world. Thus, Proposition 1 applies.

Another interesting case for which we can shed light on sorting patterns is the important HARA (hyperbolic absolute risk aversion) class of utility functions, given by $u(c) = ((1-\alpha)/\alpha)(((ac/(1-\alpha)) + b)^\alpha - 1)$, with $a > 0$ and $(ac/(1-\alpha)) + b > 0$. It subsumes all the utility functions commonly used in economic applications: linear utility (as α goes to one), log-utility (as α goes to zero), quadratic utility (when $\alpha = 2$), constant relative risk aversion utility or CRRA (when $b = 0$ and $a = 1 - \alpha$), and constant absolute risk aversion utility or CARA (when $b = 1$ and α goes to $-\infty$). We show in the Appendix that in this case Φ is given by

$$\Phi(x,y,v) = \frac{1-\alpha}{\alpha}\left(\left(V^{\frac{1}{\alpha}}(x,y) - \left(\frac{v\alpha}{1-\alpha}+1\right)^{\frac{1}{\alpha}}\right)^{\alpha} - 1\right), \tag{6}$$

where

$$V(x,y) = \int_{\underline{\omega}}^{\overline{\omega}}\int_{\underline{\sigma}}^{\overline{\sigma}}\left(\frac{aq(\omega,\sigma)}{1-\alpha} + 2b\right)^{\alpha} h(\omega,\sigma|x,y)d\sigma d\omega. \tag{7}$$

One can verify that the limit of Φ as α goes to one is $V(x,y) - v$, that is, the TU case. Hence, (6) is a generalization of TU. We can now state a general sorting result in this case:

Proposition 3. *Let agents have a common HARA utility function with $\alpha \neq 0$. Then PAM is optimal if $((1-\alpha)/\alpha)(VV_{xy} + ((1-\alpha)/\alpha))V_xV_y) \geq 0$. NAM is optimal if the inequality is reversed.*

To explain this result, assume first that $(1-\alpha)/\alpha > 0$. Then by a strictly increasing transformation of Φ, we can obtain an equivalent (for sorting purposes) representation of (6) as $\hat{\Phi}(x,y,v) = V^{\frac{1}{\alpha}}(x,y) - v^{\frac{1}{\alpha}} = \hat{V}(x,y) - \hat{v}$. In the language of Legros and Newman (2007) (see their Proposition 2), the problem is TU representable, and PAM or NAM emerges depending on whether $\hat{V}$ is spm or sbm, which reduces to the sign of $((1-\alpha)/\alpha)(VV_{xy} + ((1-\alpha)/\alpha))V_xV_y)$. Similar remarks apply when $(1-\alpha)/\alpha$ is negative. Thus, sorting requires more stringent conditions under risk aversion. Regarding V, notice that unlike the TU case where q is in the integrand, equation (7) contains what can be interpreted as a 'modified' output $\hat{q} = ((aq/(1-\alpha)) + 2b)^\alpha$, which has the same form as the utility function of each agent except for the number two in front of b.[19] Hence, $V = \int\int \hat{q}h$, and we can adapt some of the properties on q used in Proposition 1 to this case using $\hat{q}$.

Obviously, the usefulness of Proposition 3 hinges upon the ease with which one can find primitives that satisfy the conditions for monotone sorting. Although a sweeping result like Proposition 1 is not

[18] To see this, note that $c(\omega,\sigma,\lambda)$ as well as $\lambda(x,y,v)$ (and hence c^*) are implicitly defined, and so are their cross-partials, which depend on u_i, $i = x,y$, in an involved way.

[19] This can be formally interpreted as the utility function the pair would use to rank alternative q's if there were many available to choose from. It is well-known that HARA aggregates partners' preferences in such a way (see Wilson (1968) and Amershi and Stoeckenius (1983)).

available, we can derive several interesting cases where we can exactly pin down the sorting pattern for utility functions commonly used in applications. (As mentioned, Proposition 2 applies with no change.)

CRRA. Consider the case with $\alpha \in (0,1)$, $b > 0$, conditional independent signals, and FOSD. Then $V > 0$, $V_x V_y > 0$, and so is $(1-\alpha)/\alpha$. Thus, PAM is optimal if $V_{xy} > 0$, and this holds if $\hat{q}$ is spm in (ω, σ). One can verify that a sufficient condition for $\hat{q} = q^\alpha$ spm is that q be root-spm in (ω, σ), i.e. the α-root of q is supermodular: $\partial^2 q^\alpha / \partial\omega\partial\sigma > 0$. *So PAM obtains if $\alpha \in (0,1)$, $b > 0$, signals are conditionally independent and ordered by FOSD, and q is root-spm.*

Consider now the same assumptions but with $\alpha > 1$, so that $-1 < (1-\alpha)/\alpha < 0$. Then it suffices for NAM that $VV_{xy} - V_x V_y > 0$, or that V be log-spm in (x, y) (recall that it is multiplied by $(1-\alpha)/\alpha$, which is negative). Since log-supermodularity is preserved by integration (Karlin and Rinott (1980)), *NAM obtains in this case if $\hat{q}$ is log-spm in (ω, σ) (which follows if q is sufficiently log-spm in (ω, σ)) and f and g are log-spm in (ω, x) and (σ, y), respectively.*

Log Utility. Regarding the log utility case $\alpha = 0$. We show in the Appendix that in this case $\log\left(e^{V(x,y)} - e^v\right)$, where $\int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} \log(aq(\omega,\sigma) + 2b) h(\omega, \sigma | x, y) d\sigma d\omega$, and that PAM obtains if $V_{xy} + V_x V_y \geq 0$ and NAM if this expression is nonpositive. Assume conditionally independent signals ordered by FOSD, and $b > 0.5$ to ensure the integrand of V is nonnegative. In this case we have that both V_x and V_y are positive, so PAM ensues if $V_{xy} > 0$. From Proposition 1 with $\hat{q} = \log(aq + 2b)$, it follows that all we need is $\hat{q}$ spm in (ω, σ), which holds if q is log-spm in (ω, σ). *Hence, PAM is optimal if $b > 0.5$, signals are conditionally independent and ordered by FOSD, and q is log-spm in (ω, σ).*

We close with a couple of interesting special cases involving $q(\omega, \sigma) = \omega + \sigma$, which would emerge if an agent's ex-post type is income and pairs pool their incomes.

CARA. Assume agents have CARA utility $u(c) = -e^{\rho c}$ and ex-post types are conditionally independent. Then there is *no sorting* prediction in this case, as $\Phi_{xy} = (\Phi_x / \Phi_v)\Phi_{xv}$. To see this, notice that as α goes to minus infinity, the condition for monotone sorting depends on $V_{xy}V - V_x V_y$ (as the limit of $(1-\alpha)/\alpha$ is -1). Also, in the limit we obtain $V(x,y) = \int\int e^{-0.5\rho(\omega+\sigma)} f(\omega|x) g(\sigma|y) d\sigma d\omega = \mathbb{E}[e^{-0.5\rho\omega}|x]\mathbb{E}[e^{-0.5\rho\sigma}|y]$. Thus, V is log-modular and hence there is no monotone sorting prediction.

Location Family. With $q(\omega, \sigma) = \omega + \sigma$, assume instead that utility is CRRA with $\alpha \in (0,1)$, which by an affine transformation can be written as $u(c) = c^\alpha / \alpha$. Suppose that ex-post types are conditionally independent and their distributions are location families $F((\omega - x)/s)$ and $G((\sigma - y)/\nu)$, respectively, where s and ν are scale parameters that we can wlog normalize to one since they are not dependent on types. Then *the optimal sorting pattern is NAM.* To show it, notice that V in this case is

$$V(x,y) = \frac{1}{\alpha} \int \int f(\omega - x)\, g(\sigma - y)\, (\omega + \sigma)^\alpha d\sigma d\omega = \frac{1}{\alpha} \mathbb{E}\left[(\omega + \sigma)^\alpha\right].$$

Integrating by parts and differentiating V with respect to x yields

$$V_x(x,y) = \int\int f(\omega - x)\, g(\sigma - y)\, (\omega+\sigma)^{\alpha-1} d\sigma d\omega = \mathbb{E}\left[(\omega+\sigma)^{\alpha-1}\right],$$

and proceeding analogously one can show that $V_y = V_x$.

Another integration by parts and differentiation with respect to y gives

$$V_{xy}(x,y) = -(1-\alpha)\int\int f(\omega - x)\, g(\sigma - y)\, (\omega+\sigma)^{\alpha-2} d\sigma d\omega = \mathbb{E}\left[(\omega+\sigma)^{\alpha-2}\right].$$

The condition for negative sorting is $V_{xy}V + ((1-\alpha)/\alpha)V_x V_y < 0$, which reduces to

$$(\mathbb{E}[(\omega+\sigma)^{\alpha-1}])^2 - \mathbb{E}[(\omega+\sigma)^{\alpha-2}]\mathbb{E}[(\omega+\sigma)^{\alpha}] < 0.$$

Add and subtract $\mathbb{E}[(\omega+\sigma)^{\alpha-2}(\omega+\sigma)^{\alpha}]$ to the right-hand side of this expression to obtain

$$(\mathbb{E}[(\omega+\sigma)^{\alpha-1}])^2 - \mathbb{E}[(\omega+\sigma)^{\alpha-2}]\mathbb{E}[(\omega+\sigma)^{\alpha}] = -\mathrm{Var}[(\omega+\sigma)^{\alpha-1}] + \mathrm{Cov}[(\omega+\sigma)^{\alpha-2}(\omega+\sigma)^{\alpha}],$$

which is clearly negative since $\alpha \in (0,1)$ implies that $\mathrm{Cov}[(\omega+\sigma)^{\alpha-2}(\omega+\sigma)^{\alpha}] < 0$. For an intuition of this result, notice that in a location family, the riskiness is kept constant while the traits x and y change. As a result, for a type x the only difference between matching with a partners with a higher trait y is that she has a higher mean realization. With the riskiness constant, the insurance motive with CRRA comes from the income effect. Partners with higher traits most efficiently insure their high realization by matching with low traits match who have low realizations, i.e., NAM. Of course, prices are such that the share of the higher trait partners is proportionately higher.

As a final remark, the proof of Proposition 3 derives in *closed form* both $c(\omega,\sigma)$ and $q(\omega,\sigma)-c(\omega,\sigma)$ for each (ω,σ). Unlike the TU case where only the expected value of ex-post transfers is pinned down, under NTU we obtain a *precise* derivation of the ex-post transfers. This can be useful in empirical work if these transfers represent, say, wages and profits, as in one of our applications below. Needless to say, the additional information comes at the cost of a more involved analysis than in the case of TU.

3.3 Matching Principals and Agents

Another important class of problems that usually exhibit NTU is the matching of principals and agents under moral hazard. In this section we develop a model that is a variation of Holmström and Milgrom (1987) embedded in a matching setting. Besides being of independent interest, the model has enough structure to be an NTU model that is TU representable (which drastically simplifies the derivation of sorting conditions) and empirically implementable (the next section illustrates its usefulness).

There are two populations, principals and agents, with heterogeneous characteristics. If a principal with ex-ante type y matches with an agent with ex-ante type x, then the following sequence of events unfold: The principal offers an incentive contract; the agent accepts or rejects it; if rejected, the agent takes his outside option; if accepted, then the agent ex-post type ω realizes and is observed by both parties; the agent chooses a level of effort that is unobservable to the principal (moral hazard); a stochastic output realizes, which depends on the (unobservable) ex-post type of the principal; and payments are determined based on the output realization and the contract.

Principals are risk neutral, and agents have CARA utility function $-e^{-r(w-0.5e^2)}$, where r is the coefficient of risk aversion, w is wage, and $e \geq 0$ denotes effort level, which has a disutility cost that is quadratic in e. Output is stochastic and depends on the agent's effort level, his ex-post type, the principal's ex-ante type, and a shock that represents the principal's ex-post type: that is,

$$q = \omega(e + t(y) + \sigma) \tag{8}$$

where σ is normally distributed $\mathcal{N}(0, s^2(y))$ and where the variance $s^2(y)$ depends on the principal's type y as does the "average" principal's type $t(y)$. In turn, ω is distributed with cdf $F(\cdot|x)$, with $\underline{\omega} \geq 0$.

We appeal to the (dynamic) Holmström and Milgrom (1987) justification for the optimality of linear contracts, which can be solved 'as if' it was a static problem with an ad-hoc restriction to linear contracts and where effort is chosen once and for all. Hence, we restrict attention to linear contracts in output, where the slope and intercept can be contingent on the agent's ex-post type: that is, for each realization of (q, ω), the contract pays a wage $w(q, \omega) = \beta(\omega) + \alpha(\omega)q$.

An important difference with Holmström and Milgrom (1987) is that matching affects the performance of the pair through ω (whose distribution depends on x) and y. This introduces a reason for sorting based on ex-ante types. Higher ex-post agent types ω produce more output per unit of effort. The reservation wage of an agent in Holmström and Milgrom (1987) is set exogenously at some level a. Here, the reservation wage $a(x)$ depends on x with corresponding reservation utility $z(x) = -e^{-ra(x)}$. This outside option will be determined *endogenously* at the matching stage and effectively corresponds to the expected wage determined in the matching equilibirum.

The principal's contracting problem is as follows (for simplicity, we supress the dependence on x and y from the contracting variables, since they are held fixed in the analysis of the optimal contract):

$$\max_{\beta,\alpha,e} \quad \int_{\underline{\omega}}^{\overline{\omega}} \left(\mathbb{E}[q(\omega,\sigma,y)|e] - (\beta(\omega) + \alpha(\omega)\mathbb{E}[q(\omega,\sigma,y)|e])\right) dF(\omega|x) \tag{9}$$

$$\text{s.t.} \quad \int_{\underline{\omega}}^{\overline{\omega}} \left(\mathbb{E}\left[-e^{-r\left(\beta(\omega)+\alpha(\omega)q(\omega,\sigma,y)-0.5e(\omega)^2\right)}\right]\right) dF(\omega|x) \geq z(x) \tag{10}$$

$$e(\omega) \in \text{argmax}_{\hat{e}} \mathbb{E}\left[-e^{-r\left(\beta(\omega)+\alpha(\omega)q(\omega,\sigma,y)-0.5\hat{e}^2\right)}\right] \quad \forall\, \omega \tag{11}$$

where (10) is the participation constraint, (11) is the incentive constraint, and the expectation is taken with respect to the distribution of σ, which is normal with zero mean and variance $s^2(y)$.

Notice that the participation constraint only needs to hold in expectation, since at the time of contracting ω is not known. In turn, the incentive constraint needs to hold for each realization of ω, for this is known at the time the agent chooses effort. Despite the apparent complication, we show in the Appendix that the solution is as in Holmström and Milgrom (1987) since, after some algebra, it can be solved separately for each value of ω. The resulting optimal contract is given by:

$$\alpha(\omega) = \frac{1}{1+rs^2(y)}, \quad \beta(\omega) = a(x) - \frac{\omega t}{1+rs^2(y)} + \frac{\omega^2}{2\left(1+rs^2(y)\right)^2}\left(rs^2(y)-1\right), \quad e(\omega) = \frac{\omega}{1+rs^2(y)},$$

and, consequently, agent's wage, principal's profit, and output, for each (ω, σ) are given by

$$w(\omega,\sigma,x,y) = a(x) + \frac{\omega^2}{2\left(1+rs(y)^2\right)} + \frac{\omega\sigma}{1+rs(y)^2} \tag{12}$$

$$\pi(\omega,\sigma,x,y) = \omega t(y) - a(x) + \frac{\omega^2}{2\left(1+rs^2(y)\right)} + \frac{rs^2(y)\omega\sigma}{1+rs^2(y)} \tag{13}$$

$$q(\omega,\sigma,y) = \frac{\omega^2}{1+rs^2(y)} + \omega(t(y)+\sigma). \tag{14}$$

Inserting these expressions into (9) yields the maximum expected profit a principal with ex-ante type y can achieved when matched to an agent with type x who receives a level of utility z:

$$\begin{aligned}
\Pi(y,x,z) &= V(x,y) - \frac{1}{r}\log(-z) \\
&= \int_{\underline{\omega}}^{\overline{\omega}} \int_{-\infty}^{\infty} q(\omega,\sigma,y) dF(\omega|x) dG(\sigma|y) - \frac{1}{r}\log(-z) \\
&= \int_{\underline{\omega}}^{\overline{\omega}} \left(\frac{\omega^2}{1+rs(y)^2} + \omega t(y) \right) dF(\omega|x) - \frac{1}{r}\log(-z) \\
&= \int_{\underline{\omega}}^{\overline{\omega}} \overline{q}(\omega,y) dF(\omega|x) - \frac{1}{r}\log(-z),
\end{aligned}$$

where $\overline{q}(\omega,y) = \int_{-\infty}^{\infty} q(\omega,\sigma,y)dG(\sigma|y) = (\omega^2/(1+rs(y)^2)) + \omega t(y)$ is the match output after integrating out σ. Notice that the model is TU representable, and thus sorting depends on the sign of $\Pi_{xy} = V_{xy}$. And although we cannot apply Proposition 1 directly (as q depends on y), a simple modification of the analysis there delivers sorting conditions for the moral hazard case:

Proposition 4 (Moral Hazard)**.** *Optimal sorting patterns are as follows:*

1. *If F is ordered by FOSD, then PAM (NAM) obtains if $\overline{q}$ is spm (sbm) in (ω, y).*
2. *If F is ordered by IR, then PAM (NAM) obtains if $\overline{q}_{\omega y}$ is decreasing (increasing) in ω.*

Consider part 1. and the functional form of $\overline{q}$. The first term $\omega^2/(1+rs(y)^2)$ comes from the moral hazard problem, while the second term $\omega t(y)$ comes from technology, the standard Beckerian component. If t is increasing and s^2 decreasing in y, then both are spm and PAM obtains. But if s^2 is instead increasing in y, then moral hazard is a force towards NAM while technology pushes towards PAM. The CEO application below exhibits this property. Regarding part 2., it differs from the corresponding part in Proposition 1 simply because q is linear in σ and has mean zero, which implies that G disappears from V, and also because q depends on y, an effect not present in the set up of Proposition 1. Interestingly, whether PAM or NAM obtains in this case is driven by the monotonicity of s^2: PAM obtains if s^2 is increasing and NAM if it is decreasing in y.

Notice that the outside option $a(x)$ or its utility value $z(x)$ does not play any role in the determination of optimal/equilbrium sorting patterns. This is obviously due to the TU-representable nature of the problem, which owes to CARA, normality, and the linearity of the contract. But if we are interested not just on the matching but also on the ex-ante transfers that constitute a competitive equilibrium function, then we must determine $a(x)$ for each x endogenously. This is done at the matching stage when principals and agents match. The certainty equivalent $a(x)$ is equal to the ex-ante wage an agent with type x receives in equilibrium, which reflects the next best alternative of this agent. As before, the matching stage is exactly as in the Becker model with deterministic types, since it is based on x and y. Thus, the ex-ante wage $a(x)$ solves the first-order condition of the maximization $\max_x V(x, y) - a(x)$, evaluated along the equilibrium matching $\mu(x)$, that is, $a'(x) = V_x(x, \mu(x))$. Therefore

$$a(x) = a(\underline{x}) + \int_{\underline{x}}^{x} V_x(z, \mu(z))dz, , \tag{15}$$

where $a(\underline{x}) \in [0, V(\underline{x}, \mu(\underline{x}))]$ is a constant of integration. The matching μ along with utilities $z(x) = -e^{-ra(x)}$ for all x and $\Pi(y, \mu^{-1}(y), z(\mu^{-1}(y)))$ for all y fully describe the competitive equilibrium.

4 Applications

In this section, we illustrate the usefulness of our stochastic sorting framework with two economic applications. The first analyzes the role of mismatch between executives and firms. The second sheds light on the importance of assortative matching for household earnings inequality: we provide a decomposition of household inequality into marital sorting and stochastic sorting.

4.1 Mismatched Executives

First, we apply the model of the previous section to the matching of CEOs to firms: heterogeneous firms and CEOs match and their relationship and compensation is regulated by an incentive contract. Using the optimal contract and the conditions for PAM, we estimate the model using US data on firms and CEOs, and we shed light on several stylized facts.[20]

US Data on Executive Compensation and Firm Profits. We construct data on CEO compen-

[20] See also Edmans and Gabaix (2011) for an analysis of the role of risk in the market for CEO's. Recent work by Bandiera, Hansen, Prat, and Sadun (2017) performs a similar exercise to ours for data on CEOs from multiple countires.

sation by imputing the CEOs job performance in the firm's stock market value.[21] To be consistent with our model, we exclusively consider new hires and we do not envisage rematching or separation. We further assume that output, wages and profits are determined as in our principal-agent model.

The data is from the Execucomp (Compustat) database for US publicly traded firms. We construct the sample by selecting all newly hired CEOs during 2010. We then construct wages and profits based on the period 2011-2012. This gives us a sample of $n = 80$ observations.[22] The two-year interval (2011-2012) we have chosen reflects the tradeoff between choosing a longer interval, which would give us a better estimate of the ex-post outcome of the match, and a shorter interval, which ensures that fewer CEO-firm matches are separated endogenously.[23]

To construct the firm type y, we follow Gabaix and Landier (2008) and Terviö (2008) and rank firms by 2011 market capitalization (as of December 31) and define the type y as the log of market capitalization. Under the assumption of frictionless matching and positive sorting, we rank workers by the firm they are matched with: $x = y$. This is a normalization, since we use no other ex-ante information on x. It is well known that the sorting pattern of a one-to-one matching model as in Becker (1973) is normalized relative to the distribution of types x and y. Of course ex post, once we have estimated the model, we need to verify that indeed the expected value of a match $V(x, y)$ is supermodular, thus justifying our identifying assumption that there is PAM.

For wages we use total compensation – denoted by TDC1 and including salary, bonus, restricted stocks, stock options, and long-term incentive payouts – and for job performance we use firm profits, as measured by the change in the stock market valuation (from Compustat) over the course of 2011 and 2012. Over this period of two years, our wage variable w is then TDC1(2011)+TDC1(2012), and our profit variable π is derived from the change in the firm's market value (MkVal) equal to MkVal(2012)-MkVal(2010), both as measured on December 31 of the year. Observe that wages are always positive in the sample, and that profits take on both negative and positive values. Figure 2a below depicts the scatter plot of the wage data together with the model estimate of the average wage by CEO type x. Total CEO compensation over the two-year period varies between less than a million up to 45 million. While compensation on average is increasing in ex-ante type x, the variation is noteworthy.

[21] It is well known that data on payoffs to both sides of the market are hard to come by. Even in matched employer-employee data obtained from exhaustive administrative sources there is usually no good information on the productivity at the job level. This has lead to a recent literature identifying sorting from wage data alone (see Eeckhout and Kircher (2011) and Lamadon, Lise, Meghir, and Robin (2013).

[22] In the entire Execucomp sample there are 1,931 firm-CEO pairs in 2010, of which 106 were newly formed during that year. Within this sample, there are missing observations for at least one of the variables we use and there are 4 separations before the end of 2012, which gives us a sample of 80.

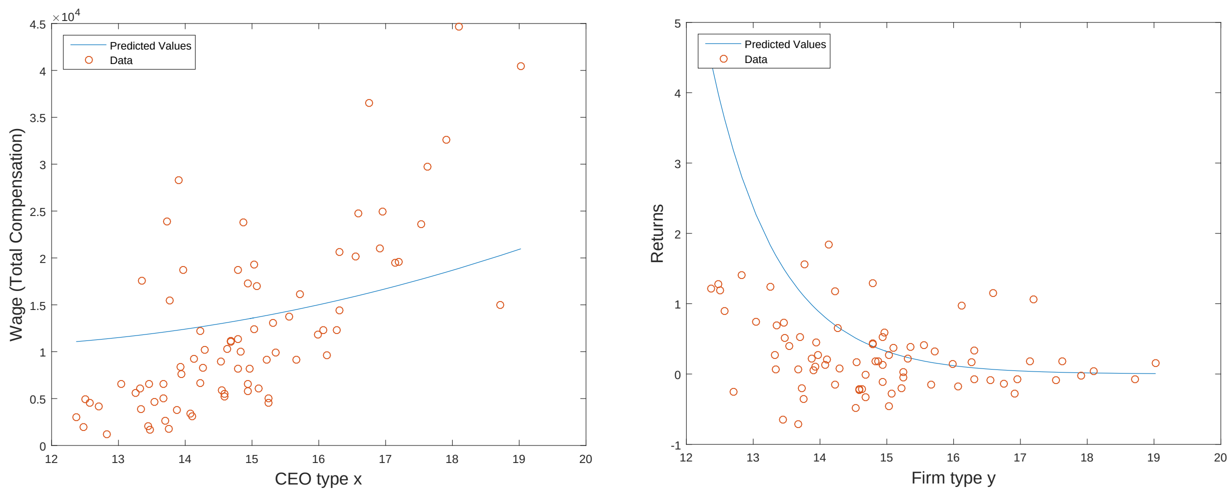
[23] Most CEO tenures last well beyond two years. In our sample, there are 4 separations within that two year window. Our model does not incorporate endogenous separations, either due to a bad realization of the CEO performance that leads to his firing, or due to a good one that leads to poaching by a more productive competitor. Since attrition would alter the continuation value and therefore the value of match formation, it could induce a bias in the estimates, especially in the variance of the estimated ex-post heterogeneity (which would be lower in the presence of endogenous separation since extreme realizations are likely to lead to separation). But it is not clear whether and if so in which direction there would be a bias in the mean of the ex-post CEO type since both too bad and too good CEOs would separate.

Figure 2b reports the same for the return data (profits over market capitalization).[24] The average stock market return over the two year period from the end of 2010 until the end of 2012 appears to be somewhat higher for the low capitalization firms in this sample, and also the variance of returns appears to be decreasing in firm capitalization.

Estimation Procedure. The model predicts a relation between observable outcomes (wages w, profits π and matched pairs $x = y$) and unobservable primitives (distributions and technology). The distribution function $F(\cdot|x)$ is assumed lognormal with parameters $k(x)$ and $m(x)$, and $G(\cdot|y)$ is normal with mean zero and variance $s^2(y)$. The model also generates an ex-post match surplus $q(\omega, \sigma, y) = (\omega^2/(1 + rs^2(y))) + \omega(t(y) + \sigma)$ as well as the split into wages and profits. Notice that once we know $k(x), m(x), t(y), s(y)$, we know both the distributions as well as the ex-post production function.

Our estimation procedure starts by positing PAM, so that $x = y$. Then we can express the primitives in terms of x: $k(x), m(x), t(x), s(x)$. Once we have obtained the estimates for these primitives, we need to verify whether the obtained technology is indeed supermodular so as to validate the PAM assumption.

The challenge is that the primitives are functions of x, and for each x we have one observation. We therefore assume polynomial forms: for $\xi \in \{k, m, t, s\}$, let $\xi(x) = \xi_0 + \xi_1 x + \xi_2 x^2$. Then we estimate these 12 parameters using maximum likelihood where we invert (12) and (13) to obtain expressions for ω and σ. The likelihood function and its derivation is reported in the Appendix.



(a) Wages (b) Firm Returns $\frac{\pi}{V_0}$

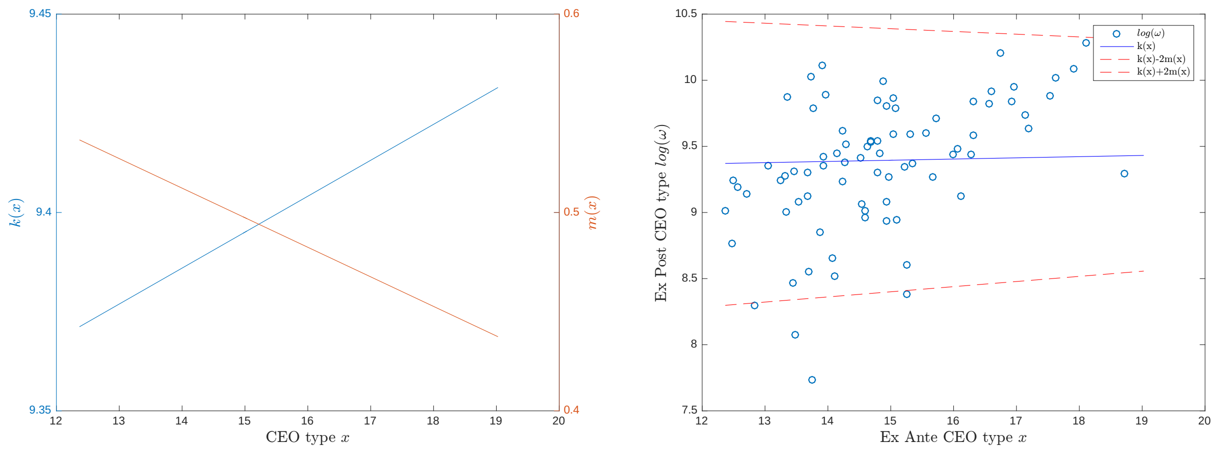
Figure 2: Data and Estimates.

Results. Our objective is to back out the distributions $F(\cdot|x)$ for all x and $G(\sigma|y)$ for all y, as well as the technology q. The model is fully specified by the 12 parameters. We present a series of plots to interpret the estimated model.[25] In Figure 2 we plot the predicted wages estimated from the model, as

[24] We choose to represent the firm profits in terms of returns rather than logs, in order to make profits of large and small firms comparable on the same scale in the graphs. We cannot take logs because some profit realizations are negative.

[25] The parameter estimates are reported in Table 2 in the Appendix.

well as the returns as a function of type x or y, together with the data. The model captures the increase in average wages and the decrease in returns as a function of the type, however the wage schedule is flatter and the return schedule is steeper than the data. Note that returns across different firm types are not necessarily the same because also the variance changes with firm type. The model does not fit the data perfectly, and below in Figures 3 and 4 we will see that this is due to the fact that the variance on ω and σ is so large that it swamps the mean of wages and returns. And of course, the variance is important for determining wages and profits given our moral hazard model with matching. We then



(a) Predicted mean $k(x)$ and predicted standard deviation $m(x)$ (of $\log \omega$).

(b) Realized ω with mean and s.d. of estimated $F(\omega|x)$ (in logs)

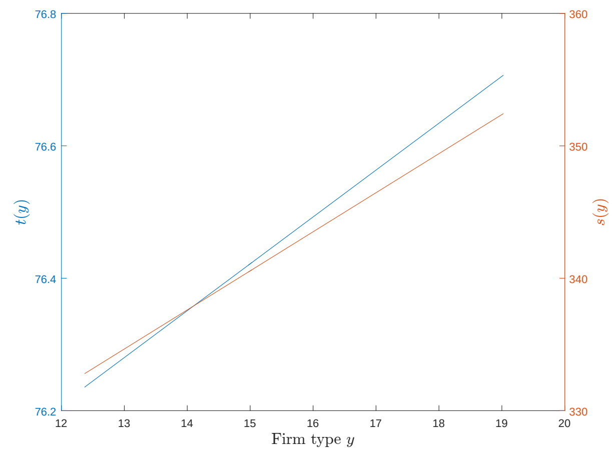
Figure 3: Estimated Parameters k, m.

plot the estimated values of k, m, t, s as a function of x or y: in Figure 3 those parameters that pertain to the distribution of ω – $k(x)$ and $m(x)$ – and in Figure 4 those that pertain to the distribution of σ – $t(y)$ and $s(y)$. In both figures, in panel (a) we plot the predicted mean as well as the predicted standard deviation, and in panel (b) we plot for each type the predicted value of the unobserved characteristic (ω and σ respectively) together with the mean and confidence bands of the distribution. A nice feature of our model is that we can not only back out the parametric distributions of unobserved types but also the actual unobserved types ω and σ given the parametric estimates.

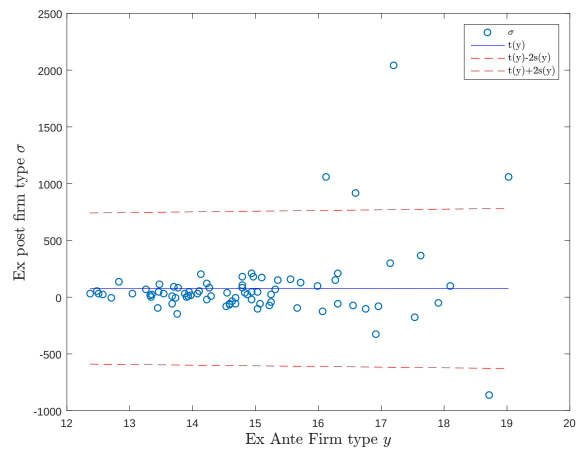
What we learn from these estimates is that k is slightly increasing, but only very slightly so, and m is basically flat. Of course, given that ω is log-normally distributed, the expected value of ω is $e^{k+0.5m^2}$ and is increasing in x (this can be seen from Figure 3b).

The following insights emerge from the results. First, the variation *across* workers of different types x is minor. This salient feature implies that the contribution of CEO marginal product is not driven by the CEO type but by the firm type. This is in line with the findings by Terviö (2008) and Gabaix and Landier (2008): wages for executives vary a lot but it is mainly driven by the variation in the characteristics of the firms they work. The CEOs themselves are only marginally different, but a CEO whose decisions affect the sales of a firm that is 100 times larger will as a result have an equally proportional economic impact compared to a CEO in the small firm. As a result, these marginal differences between CEOs translate in huge differences in compensation.

Second, there is uncertainty about CEO types – for a *given* type x – that is orders of magnitude larger than the increase in the predicted type. The "noise" around $k(x)$ in Figure 3b is substantially bigger than the increase of k in x. This is the *main insight* of this application. Not only is the wage increase of CEOs driven nearly exclusively by the firm characteristics, but there is also so much variation in the ex-post CEO type that we can barely predict who will be a good CEO ex ante. Moreover and as a result, there is significant mismatch ex post. We will return to the issue of mismatch below. The



(a) Predicted mean $t(y)$ and predicted standard deviation $s(y)$ of σ.

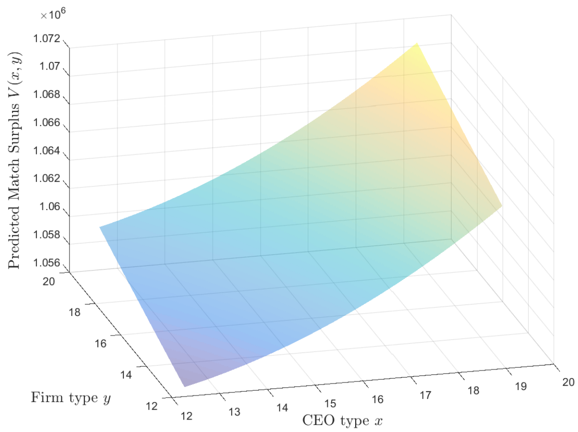


(b) Realized σ with mean $t(y)$ and s.d. of estimated $G(\sigma|y)$.

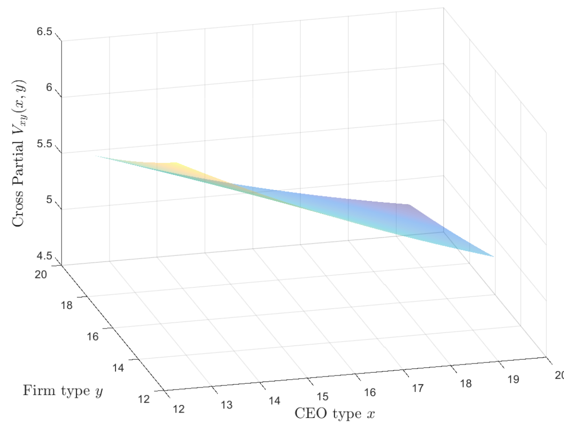
Figure 4: Estimated Parameters t, s.

estimates of the firm characteristics are as expected. Higher ranked firms generate more match surplus – $t(y)$ is increasing – but there is a lot of noise – the predicted standard deviation $s(y)$ is large, and also increasing in y – (see Figure 4a). The estimated values of σ confirm the extent of the variance as evident in the large confidence bands in Figure 4b. Observe also that there is high variance across all firm types y. The high variance in the distribution of σ has important implications for the optimal contract in our model. Variance in output implies that the CEO contribution to output is subject more to luck than to her effort. Moreover, it becomes more costly for the firm to induce effort due to CEOs risk aversion. As a result, the optimal contract will entail less powered incentives.

We now turn to the estimated match surplus. The ex-post output is equal to the sum of wages and profits. With the estimated parameters, we can now also construct the ex-ante match surplus from equation (28) as well as the cross-partial V_{xy} from equation (29), both in the Appendix. While we only have observations along the equilibrium allocation where $x = y$, the estimated model parameters allow us to reconstruct the match expected match surplus $V(x, y)$ for the entire domain of $(x, y) \in \mathbb{R}^2_+$. Figure 5 plots both the estimated $V(x, y)$ and V_{xy} in three dimensions. Figure 5a reveals that expected output is increasing both in CEO type x and in firm type y. Moreover, from the convexity of the



(a) The estimated expected match surplus $V(x, y)$



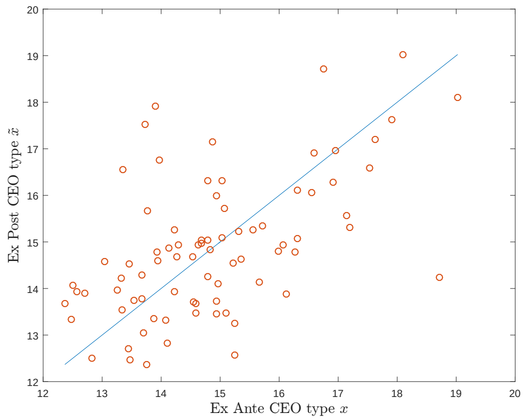
(b) Complementarity of the match surplus $V_{xy}(x, y)$

Figure 5: Estimated Match Surplus.

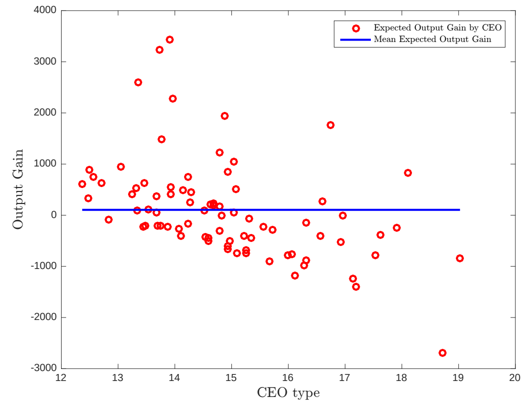
hyperplane it appears that there are complementarities and $V(x, y)$ is supermodular. To verify this, we evaluate the cross-partial derivative (29) at the estimated parameters. The plot of the cross-partial in Figure 5b confirms that over the entire domain of x and y, $V(x, y)$ is supermodular. This is crucial for our estimation strategy, since we estimated the model assuming PAM, which allowed us to start from the premise that $x = y$. This identifying assumption is justified so long as the estimated technology is indeed supermodular, as we have just shown is the case.

The Cost of Mismatch. Our main findings is that there is substantial ex-post mismatch. This was apparent already from the difference between observed and predicted wages. But we can also calculate the monetary cost of mismatch. For that purpose, we will perform a simple experiment in which we reallocate the mismatched CEOs to new firms based on their ex-post realized type. For example, if a CEO was ranked at position ninth initially, but her ex-post realization of ω is such that she is ranked fifth, then in the experiment we will allocate her to the firm with rank five. If at rank nine her initial type was x then at rank five her type will be denoted by $\tilde{x}$. Given supermodularity, this will lead to an increase in output as with her ex-post type she is better matched to the firm. Observe that we treat the realization of the CEO type ω as permanent, as if it is the same in all firms.[26] In Figure 6a we plot the relation between the ex-ante type x and the type $\tilde{x}$ assigned based on the ex-post realization of ω. We will use the ex-post realization $\tilde{x}$ to calculate the cost of mismatch. We do this by calculating the difference between the output generated under the new, virtual matching where $\tilde{x} = y$ compared to the ex-ante matching where $x = y$. Moreover, we take into account that in their virtual matches CEOs would optimally change the effort they provide since the firm's type is different. As a result, we can decompose the output gain into a component that is due to the mismatch of types, and one that is due

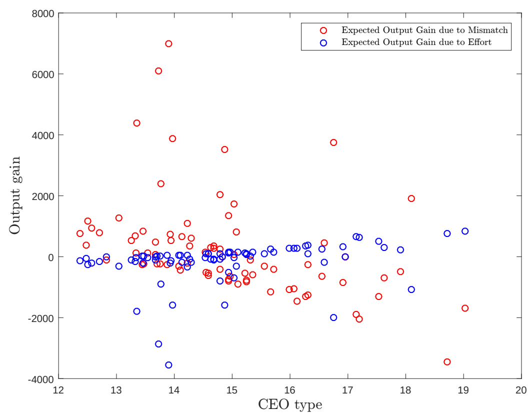
[26]We cannot know from our analysis whether the type is permanent or match specific, and it is likely that both components are present. The results of this exercise should be taken as indicative of the cost of mismatch.



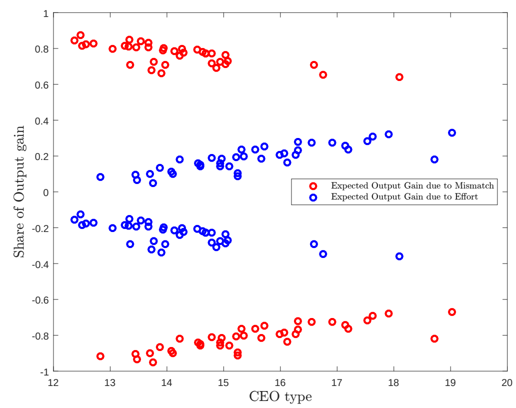
(a) ex-post rank $\tilde{x}$.



(b) Output Gain



(c) Output Gain Decomposition: mismatch and effort.



(d) Output Gain Decomposition: gain mismatch $\frac{\Delta_m}{|\Delta_e|+|\Delta_m|}$; gain effort $\frac{\Delta_e}{|\Delta_e|+|\Delta_m|}$

Figure 6: Mismatch.

to the adjustment of effort. The exact decomposition is in the Appendix.

Figure 6b plots the output gain for each type x. The gain is negative for some CEOs because they were "overmatched" and hence will produce less output than before as $\tilde{x} < x$. Over all CEOs, however, the average of the output changes is positive. This follows from the supermodular ex-post output.

Figure 6c plots the decomposition of the output gain due to pure mismatch and due to effort. These changes can be positive or negative. Interestingly, the output effect of effort tends to have the opposite sign of the output effect of mismatch. What is more relevant economically is that the output effect of effort is small compared to that of mismatch. This is evident in Figure 6d where we show the percentage contribution of each output gain. Because of the negatives, we represent this percentage change as a fraction of the absolute value of the change in each of the two components (effort and mismatch).

This counterfactual exercise teaches us that *the output loss is mainly driven by mismatch.* The contribution from the adjustment of effort is a small fraction of the total change in output (in absolute value). For the low productivity firms, the contribution to output (whether positive or negative) of mismatch is over 80%. This indicates that there is little loss due to incentives. In part this can be attributed to the amount of noise in σ, which makes the incentive component of compensation small. For higher types x the share of output change due to mismatch gets smaller. Incentives seem to matter more for the highest CEO types. Finally, in Figure 6b that there seems to be regression to the mean in the sense that low types x tend to post net gains whereas higher CEO types tend to post losses.

This application sheds new light on the role of mismatch and CEO stochastic sorting. Like Terviö (2008) and Gabaix and Landier (2008) we find that the contribution to wage inequality of CEOs is mainly driven by the heterogeneity of firms. What we add to this debate is that CEO productivity is also highly stochastic, and that the magnitude of this randomness is larger than the ex-ante variation in CEO type. As a result, ex post many firm-CEO pairs are highly mismatched. Output loss is mainly due to direct mismatch and not to suboptimal effort. Our results indicate that firms should be more concerned about the *selection* of their CEOs than about providing them with *incentives* to work hard.

4.2 Household Income Inequality: Marital vs. Stochastic Sorting?

The second application uses our framework to shed light on the determinants of household inequality and its dramatic increase in recent decades. The variance of household income has increased eightfold between 1960 and 2014 (see Table 1), while average household income has merely doubled. Many different components contribute to this increase. Our objective here is to decompose the increase in the variance of household income into the components due to *marital sorting* based on ex-ante education – the education of married partners is more similar – versus those due to *stochastic sorting* based on ex-post earnings – the distribution of individual income has become more unequal and correlated. Marital sorting is often cited as a major cause for the increased earnings inequality of married households.

We will further decompose the increase in household inequality due to stochastic sorting into: increased earnings inequality of males and females; increased correlation between male and female earnings. Similarly, we decompose marital sorting into: increased educational attainment of males and females; increased female educational attainment *relative to males*; higher assortativeness (or lower ex-ante mismatch). This decomposition is a simple accounting exercise that has a clean interpretation and foundation in our stochastic sorting framework.[27] The decomposition is inspired by the work of Greenwood, Guner, Kocharkov, and Santos (2014) and Lam (1997).[28]

For the years between 1960 and 2014, we use data from the US Decennial Census (1970, 1980, 1990,

[27] We have no data on the ex-post division of the surplus within the marriage, so we cannot estimate the match surplus technology $q(\omega, \sigma)$ using, for example, a model of risk sharing. Hence, we restrict our attention to the analysis of the decomposition of household inequality.

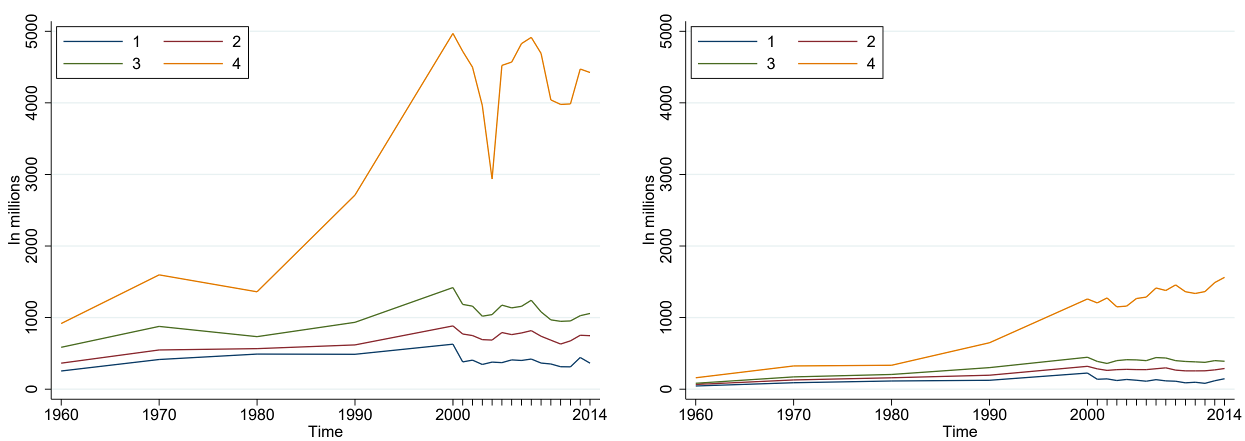
[28] Eika, Mogstad, and Zafar (2014) and Pilossoph (2016) also find that most of the inequality is due to increased earnings inequality and not due to marital sorting. We discovered their work posterior to finishing a draft that contains our results. Our focus is to link marital sorting to stochastic sorting with special attention to the matching patterns. For example, we do not assume that frictionless matching occurs along the diagonal.

2000) and the American Community Service (ACS) (2005-2014) to obtain information on married heterosexual couples between the ages of 25 and 55. For each married couple we observe each partner's education and annual earnings. In line with the notation of the model, education is denoted x for the male and y for female and distributed according to $\Gamma(x)$ and $\Psi(y)$. We group the level of education into four categories: less than high school (1=HS-), high school (2=HS), some college (3=C-) and college or more (4=C).[29] Earnings are expressed in 2014 US dollars. Earnings for males are denoted by ω and by σ for the females. The joint distribution of earnings is $H(\cdot,\cdot|x,y)$ for each couple with ex-ante types (x,y), and the marginals $F(\cdot|x)$ and $G(\cdot|y)$.

I. Stochastic Sorting

This part bundles the sources of inequality that are due to the ex-post (unconditional) distribution of income, with cdf given by $H(\omega,\sigma)$ for each (ω,σ). We distinguish between individual earnings inequality and the correlation between partners' earnings.

I.a. Increased Individual Earnings Inequality. For each level of education, in Figure 7 we report the variance of earnings for males and females. The variance has gone up for all education categories, but increasingly so for those with more years of schooling. Clearly, if each of the partners in marriage has a more volatile income process, then this will lead to higher household inequality.



(a) Male Earnings: $F(\omega|x)$ (b) Female Earnings: $G(\sigma|y)$

Figure 7: Variance of Earnings by Education over Time: $x, y \in \{1,2,3,4\}$.

I.b. Increased Correlation between Male and Female Earnings. Household earnings inequality will also be affected by the extent to which the spouses' earnings are correlated. In line with the findings by Lam (1997), we see an increase in the correlation for all education combinations of

[29]In the Appendix we report the results for the analysis with 11 education categories, one category for each year of schooling. This categorization is often used to obtain a measure of education closer to a continuum. We find that the results are robust to the choice of education category.

marriages. Over time, the correlation in incomes of husband and wife over the entire sample of married couples has increased from -0.06% in 1960 to 11.17% in 2014. The correlation in income of the spouses reflects selection due to unobserved heterogeneity. Spouses with a college degree that have met while studying finance or law end up in high paying jobs on Wall Street or in law firms. The correlation also reflects *choices*, most notably the amount of labor supply, both at the extensive and at the intensive margin. In the 1960s female labor force participation was a lot lower than in current times, introducing negative correlation in earnings. Even today, females tend to work fewer hours than males. But labor supply may also introduce positive correlation. Partners with high productivity will farm out more child care freeing up more time to work and generate higher earnings, while partners with low earnings both reduce labor supply to care for the children. In both cases correlation in earnings is positive.

II. Marital Sorting

In the literature, marital sorting on education is typically measured by regressing male education on female education. An increase in the regression coefficient is interpreted as an increase in sorting. However, this regression mixes the effect of two distinct mechanisms. First, there is the change in the *relative* distribution of education between men and women. There has been *gender equalization* in schooling. While educational attainment has gone up for both men and women, schooling for women has increased substantially relative to that of men. Second, ex-ante matching on education is not perfect and we might expect a change in *ex-ante mismatch*. While our model does not have much to say about ex-ante mismatch, we can account for its role in the decomposition of inequality.

II.a. Increased Female Educational Attainment. Educational attainment for both males and females has increased. Figure 8 shows the cumulative frequency of years of schooling for both males and females for 1960 and 2014. As is evident from the figure, both distributions have shifted massively to the right, but the educational distribution for females has shifted more relative to that of males. So much so that the distribution of education of females in 2014 stochastically dominates that of males. As a result, under PAM without any ex-ante mismatch, in 1960 we would see matches of males with females that have lower education than themselves, whereas in 2014 we expect to see matches with females that have higher education. In terms of the notation of the model, this implies that the equilibrium matching $\mu(x) = \Psi^{-1}[\Gamma(x)]$ has moved upwards from below the diagonal to above the diagonal. We must stress that this is due to the change in the distributions and has nothing to do with mismatch.

In Figure 9 we plot $m(x, y)$, the actual number of (x, y) matches observed in the data, as well as frictionless matching $m^f(x, y)$, the virtual number of (x, y) (Table 3 in the Appendix has the corresponding matrices), constructed as if there is PAM and no mismatch, which can be derived from $\Gamma(x)$ and $\Psi(y)$. Two properties can be observed: higher levels of education of males has led to a rightward move of the matching mass in both m and m^f; and the relatively higher increase in educational attainment of females has led to an upward shift of the equilibrium allocation of matches $\mu(x)$.

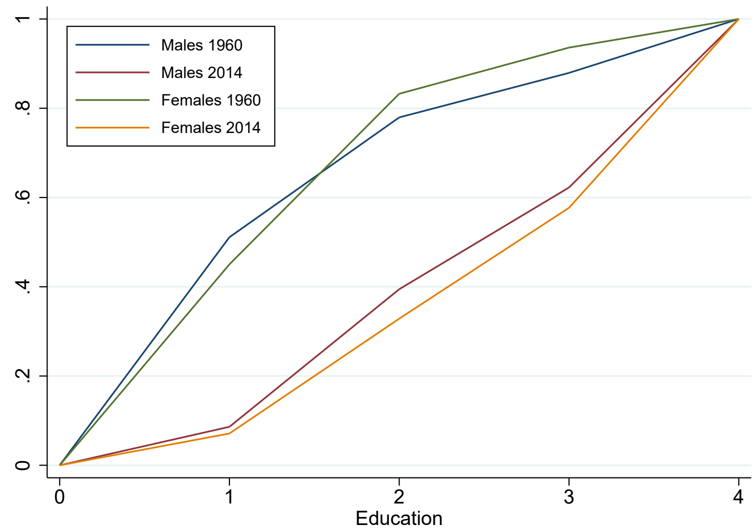
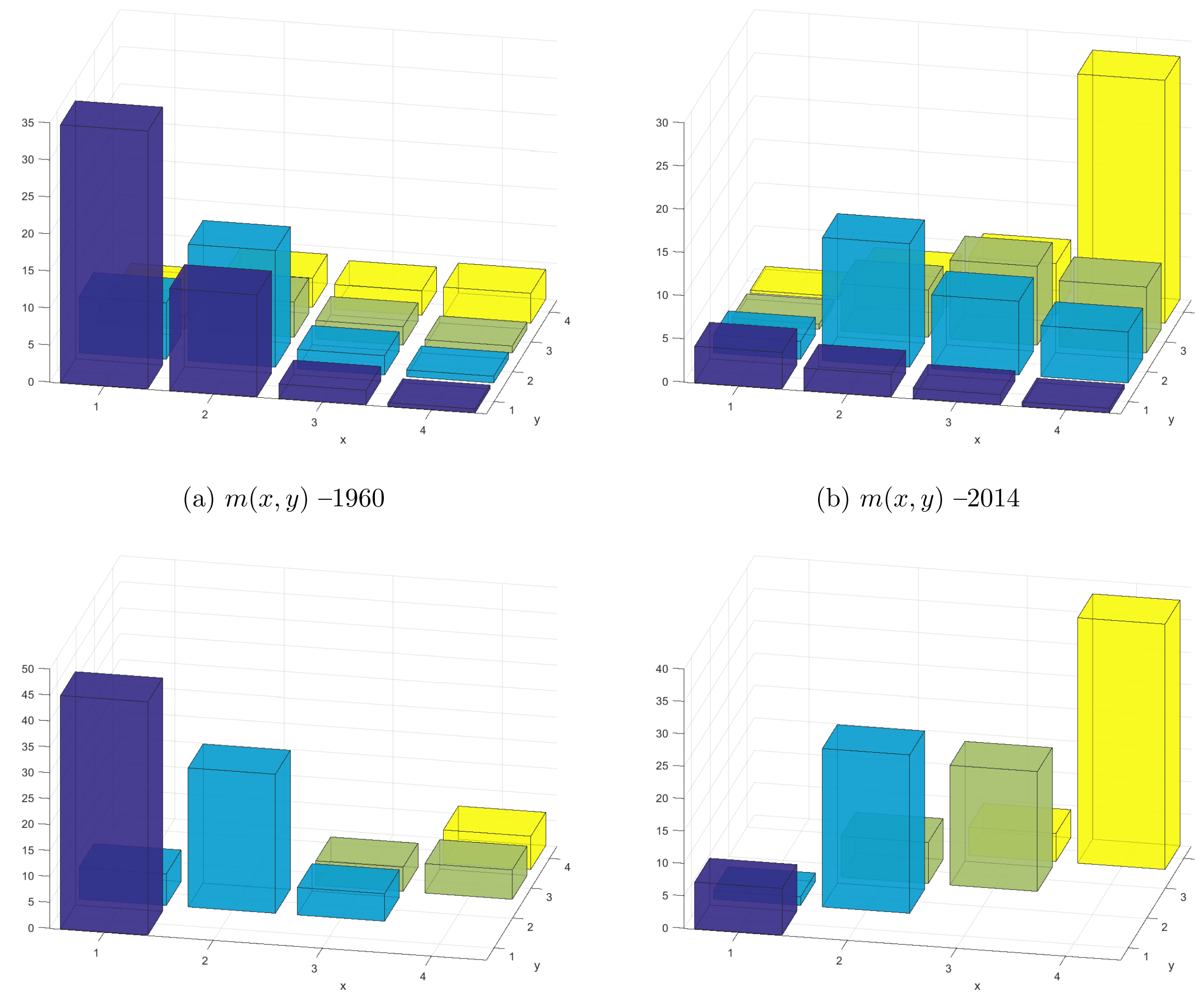


Figure 8: Distribution of Education: $\Gamma(x)$ for males and $\Psi(y)$ for females in 1960 and 2014.

II.b. Increased Marital Assortativeness (ex ante). A separate source of marital sorting is the decrease in assortativeness or ex-ante mismatch. Although our model has nothing to say about ex-ante mismatch, we can quantify the extent of its presence. To do so, we use a measure of distance between the actually observed matching matrix $m(x, y)$ and the frictionless matching matrix $m^f(x, y)$. We calculate a measure d of distance between these two matrices as the sum of the absolute value of the difference between each element, that is, $d = \sum_x \sum_y |m(x, y) - m^f(x, y)|$.

If the matching m is perfectly frictionless and equal to m^f, then the distance is zero. Using d, we can reconstruct a virtual allocation of matches in which we adjust the weight on the frictionless matching $m^f(x, y)$ with the distance measure from a different year. This is a way to control for marital sorting due to a change in assortativeness and that due to a change in $\Gamma(x)$ and $\Psi(y)$. When calculating d for our two focal years, we find that $d_{1960} = 0.3326$ and $d_{2014} = 0.3317$. The difference is negligible. Quite surprisingly, ex-ante mismatch has remained unchanged. This runs counter to the conventional wisdom that assortativeness in marriage has increased and has thus contributed to household inequality. Part of the reason why there is the belief that marital sorting has increased is because there have been substantial changes in the marriage market that are confounded with marital sorting: the overall increase in educational attainment; the relative larger increase for females than for males; the increase in the correlation of earnings; the increase in the variance of earnings. The objective of what follows is precisely to decompose the increase in household inequality into these different components.

Decomposing Household Inequality. We now decompose the different sources of inequality by calculating the variance of household income in 2014 while setting one component to its 1960 value. The objective is to gauge the share of the eightfold increase in the variance between 1960 and 2014 that



(a) $m(x,y)$ –1960

(b) $m(x,y)$ –2014

(c) $m^f(x,y)$ – 1960

(d) $m^f(x,y)$ – 2014

Figure 9: Actual Distribution of Marriages by Education $m(x,y)$ and Hypothetical Distribution without ex-ante Mismatch: $m^f(x,y)$.

can be attributed to each component. In a nutshell, we calculate that in total the earnings distribution accounts for more than 80% of the increase in household inequality. This can be attributed nearly entirely to an increase in the variance of individual earnings (about 80%), and to a minor extent to an increase in the correlation in earnings (about 10%). That these two components do not add up to what is explained by the the 1960 sample distribution of earnings is due to the fact that we use a model to do the decomposition.

The role of marital sorting is rather minor. First, while the total effect of marital sorting appears to explains 58% of the increase in household inequality, this increase cannot be separated from the increase in the variance of income (i.e. stochastic sorting). We can decompose the contribution of assortativeness and the marginal distribution. Since assortativeness has barely changed (it even decreased slightly), it

Baseline	Sample	Normal Model*		
$\text{Var}_{2014}(\omega+\sigma)$	3.85×10^9	3.33×10^9		
$\text{Var}_{1960}(\omega+\sigma)$	0.50×10^9	0.46×10^9		
$\frac{\text{Var}_{2014}(\omega+\sigma)}{\text{Var}_{1960}(\omega+\sigma)}$	7.72	7.31		
	$X_{1960}[x,y]$	$\text{Var}_{2014}(\omega+\sigma; X_{1960})$	$\frac{\text{Var}_{2014}(\omega+\sigma;X_{1960})}{\text{Var}_{1960}(\omega+\sigma)}$	X_{2014} explains
A. Stochastic Sorting (ex post)				
1. Marginals Earnings*	F_{1960}, G_{1960}	0.80×10^9	1.76	76%
2. Correlation Earnings*	ρ_{1960}	3.06×10^9	6.72	8%
Total (Normal model)*	$F_{1960}, G_{1960}, \rho_{1960}$	0.75×10^9	1.64	78%
Total	H_{1960}	0.80×10^9	1.74	76%
B. Marital Sorting (ex-ante)				
1. Marginals Education	$\Gamma_{1960}, \Psi_{1960}$	1.52×10^9	3.05	61%
Allocation: Gender Equal.	$\Gamma_{2014}, \tilde{\Psi}_{2014}, \mu^f_{1960}$	3.46×10^9	6.94	10%
All more educated	$\Gamma_{1960}, \tilde{\Psi}_{1960}, \mu^f_{2014}$	2.29×10^9	4.58	41%
2. Assortativeness Education	d_{1960}	3.98×10^9	7.99	3%
Total	M_{1960}	1.69×10^9	3.39	56%

* This decomposition assumes for each education pair $[x, y]$ normality of the joint distribution of ω, σ, which implies normality of the marginal distributions F, G, as well as normality of the distribution of $\omega + \sigma$, in order to single out a constant correlation coefficient $\rho[x, y]$.

Table 1: Decomposition of Household Income Inequality. We calculate the Variance of household income in 2014 in each case holding fixed one of the determinants at 1960 levels.

is no surprise that it does not contribute to the change in inequality. The change in Γ and Ψ contributes in two ways. First, it changes equilibrium matching because married females are more educated *relative* to males, so even with perfect assortativeness and no ex-ante mismatch, every male will be married with a (weakly) higher educated female. The effect of gender equalization explains 13% of the increase in the variance. Second, both males and females are more educated, and we know from the marginals of the earnings distributions that the variance of the highly educated earners is higher. This explains 43% of the increase in variance. But again, this is not really due to marital sorting but to more people being more educated: the composition of matches now puts more weight on high variance households (those with high education). This second effect should therefore be attributed to stochastic and not to marital sorting. We find that about 14% can be attributed to pure marital sorting, with most of the change stemming from females becoming more educated relative to males.

5 Concluding Remarks

To capture mismatch due to evolving types, we have proposed a simple model that generalizes the frictionless matching model with deterministic types (Becker (1973)) to a matching model where types are stochastic. The sorting pattern – whether there is PAM or NAM – now not only depends on the characteristics of the technology through the ex-post match value as in the deterministic matching model, but it also depends on the stochastic order imposed on the distributions of the stochastic characteristics. For instance, in the baseline case with TU we find that FOSD together with spm match output ensures positive sorting. Our results link a rich set of combinations of technological features and stochastic orders on the distributions to the optimal sorting patterns that emerge. The setup is amenable to empirical applications for two reasons: first, mismatch is an integral part of the description and the equilibrium of the model; and second, it provides a theoretical framework that rationalizes selection from matching since the distribution of ex-post outcomes across matches inherit correlation patterns even if the outcomes at the match level are uncorrelated.

We also analyze the model with NTU, focusing on important classes of problems that are TU representable, and derive conditions for PAM and NAM in two economically relevant settings: risk sharing and a principal-agent problems. First, the stochastic nature of the model leads naturally to a problem of how to efficiently share risk. Equilibrium now also depends on the properties of the partners' utility functions. We provide a fairly general result for the oft-used HARA class, which subsumes most of the standard utility functions. The sorting pattern now depends on a simple condition that suitably generalizes the TU case. Second, we embed the canonical principal-agent setting in this assignment model with stochastic types and characterize the equilibrium allocation.

Since mismatch is inherent in real world matching problems, our model is a useful extension of Becker's when it comes to taking it to data. We illustrate with two applications how the model is amenable to address significant economic questions about mismatch, incentives, and inequality.

Finally, our model provides a formal justification for sorting in the presence of measurement error, a common assumption to deal with mismatch in empirical work. We show that not all distributions of the errors induce PAM, but we provide conditions for this to be the case. For example, if the distribution of error terms satisfy FOSD and the ex-post match surplus is spm, then PAM ensues.

Appendix A Omitted Proofs

A.1 Proof of Proposition 1

1. and 2. Integrating (1) by parts, and then the second term of the resulting expression again yields:

$$
\begin{aligned}
V(x,y) &= \int_{\underline{\sigma}}^{\overline{\sigma}} \int_{\underline{\omega}}^{\overline{\omega}} q(\omega,\sigma) h(\omega,\sigma|x,y) d\omega d\sigma \\
&= \int_{\underline{\sigma}}^{\overline{\sigma}} q(\overline{\omega},\sigma) g(\sigma|y) d\sigma - \int_{\underline{\sigma}}^{\overline{\sigma}} \int_{\underline{\omega}}^{\overline{\omega}} q_\omega(\omega,\sigma) \left(\int_{\underline{\omega}}^{\omega} h(s,\sigma|x,y) ds \right) d\omega d\sigma \\
&= \int_{\underline{\sigma}}^{\overline{\sigma}} q(\overline{\omega},\sigma) g(\sigma|y) d\sigma - \int_{\underline{\omega}}^{\overline{\omega}} q_\omega(\omega,\overline{\sigma}) F(\omega|x) d\omega + \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} q_{\omega\sigma}(\omega,\sigma) H(\omega,\sigma|x,y) d\sigma d\omega,
\end{aligned}
$$

where we have used the assumption that the marginals are independent of partner's traits. Hence, V is spm in (x,y) if and only if

$$
\int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} q_{\omega\sigma}(\omega,\sigma) H(\omega,\sigma|x,y) d\sigma d\omega \tag{16}
$$

is spm in (x,y), which holds if either H is spm in (x,y) and q is spm in (ω,σ), or if both are sbm. Under these conditions, PAM is optimal. Similarly, V is sbm in (x,y) if and only if (16) is nonpositive, and this holds if H and q have opposite cross partials, in which case NAM is optimal. It is clear that the conditions on q are also necessary we want the results to hold for all H spm or sbm in (x,y).

3. and 4. The expression for $V(x,y)$ above reveals that (16) is the only term that depends on both x and y and thus we focus on this term. Let $\hat{q} \equiv q_{\omega\sigma}$ and $\hat{h} \equiv H$, and let $\hat{H} \equiv \int_{\underline{\omega}}^{\omega} \int_{\underline{\sigma}}^{\sigma} \hat{h}$. Then

$$
\int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} \hat{q}(\omega,\sigma) \hat{h}(\omega,\sigma|x,y) d\sigma d\omega, \tag{17}
$$

has the same functional form as V above, and so does the resulting expression after integrating by parts with respect to σ and then with respect to ω. Since by assumption $\int_{\underline{\omega}}^{\overline{\omega}} \hat{h} = \int_{\underline{\omega}}^{\overline{\omega}} H$ is independent of y and $\int_{\underline{\sigma}}^{\overline{\sigma}} \hat{h} = \int_{\underline{\sigma}}^{\overline{\sigma}} H$ is independent of x, it follows that the only relevant term for the cross-partial derivative of (17) with repect to x and y is the cross-partial derivative with respect to x and y of

$$
\int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} \hat{q}_{\omega\sigma}(\omega,\sigma) \hat{H}(\omega,\sigma|x,y) d\sigma d\omega = \int_{\underline{\omega}}^{\overline{\omega}} \int_{\underline{\sigma}}^{\overline{\sigma}} q_{\omega\omega\sigma\sigma}(\omega,\sigma) \left(\int_{\underline{\omega}}^{\omega} \int_{\underline{\sigma}}^{\sigma} H(s,t|x,y) ds dt \right) d\sigma d\omega.
$$

If either $\int_{\underline{\omega}}^{\omega} \int_{\underline{\sigma}}^{\sigma} H$ is spm in (x,y) and $q_{\omega\sigma}$ is spm in (ω,σ), or if both are sbm, then PAM is optimal. Similarly, NAM is optimal if $q_{\omega\sigma}$ and $\int_{\underline{\omega}}^{\omega} \int_{\underline{\sigma}}^{\sigma} H$ have opposite cross partials. It is clear that the conditions on $q_{\omega\sigma}$ are necessary if we want the results to hold for all $\int_{\underline{\omega}}^{\omega} \int_{\underline{\sigma}}^{\sigma} H$ spm or sbm in (x,y), with $\int_{\underline{\omega}}^{\overline{\omega}} H$ and $\int_{\underline{\sigma}}^{\overline{\sigma}} H$ independent of y and x, respectively. $\square$

A.2 Proof of Proposition 2

As mentioned, the proofs of parts 1.–3. are in the text.

To prove part 4., assume that $x \in \{x_\ell, x_h\}$, $\gamma(x_h) = \gamma \in (0,1)$, and $y \in \{y_\ell, y_h\}$, $\psi(y_h) = \psi \in (0,1)$. Ex-post types ω and σ, however, can take on a continuum of values.

Consider first PAM. Then the conditional cdf $H_1(\omega|\sigma)$ is given by

$$H_1(\omega|\sigma) = m(x_h|\sigma)F(\omega|x_h) + (1 - m(x_h|\sigma))F(\omega|x_\ell) = m(x_h|\sigma)\left(F(\omega|x_h) - F(\omega|x_\ell)\right) + F(\omega|x_\ell),$$

where

$$m(x_h|\sigma) = \frac{\gamma g(\sigma|\mu(x_h))}{\gamma g(\sigma|\mu(x_h)) + (1-\gamma)g(\sigma|\mu(x_\ell))} = \frac{1}{1 + \frac{(1-\gamma)}{\gamma}\frac{g(\sigma|\mu(x_h))}{g(\sigma|\mu(x_\ell))}}.$$

It follows that

$$\int_{\underline{\omega}}^{t} H_1(\omega|\sigma)d\omega = m(x_h|\sigma)\int_{\underline{\omega}}^{t}\left(F(\omega|x_h) - F(\omega|x_\ell)\right)d\omega + \int_{\underline{\omega}}^{t} F(\omega|x_\ell)d\omega. \tag{18}$$

Notice that the behavior of $\int_{\underline{\omega}}^{t} H_1$ as a function of σ is determined by the behavior of $m(x_h|\sigma)$.

Consider the following characterization of MPS (Muller and Stoyan p. 28): " A cdf G differs from F by a MPS if and only if there exist constants $a < b$ such that $G - F$ is increasing on $(-\infty, a)$ and (b, ∞) and decreasing on (a, b)."

Let us apply it to our case. Assume that $G(\sigma|\mu(x_\ell))$ differs from $G(\sigma|\mu(x_h))$ by a MPS. Then

$$\frac{d\left(G(\sigma|\mu(x_\ell)) - G(\sigma|\mu(x_h))\right)}{d\sigma} = g(\sigma|\mu(x_\ell)) - g(\sigma|\mu(x_h)) = g(\sigma|\mu(x_\ell))\left(1 - \frac{g(\sigma|\mu(x_h))}{g(\sigma|\mu(x_\ell))}\right).$$

By the definition above, this expression is nonnegative on $[\underline{\sigma}, \hat{\sigma}) \cup (\tilde{\sigma}, \overline{\sigma}]$ and nonpositive on $[\hat{\sigma}, \tilde{\sigma}]$ for some constants $\hat{\sigma} < \tilde{\sigma}$. It follows that

$$\frac{g(\sigma|\mu(x_h))}{g(\sigma|\mu(x_\ell))} = \begin{cases} \leq 1 & \text{if } \sigma \in [\underline{\sigma}, \hat{\sigma}) \cup (\tilde{\sigma}, \overline{\sigma}] \\ \geq 1 & \text{if } \sigma \in [\hat{\sigma}, \tilde{\sigma}], \end{cases}$$

and thus $m(x_h|\sigma)$ is larger if $\sigma \in [\hat{\sigma}, \tilde{\sigma}]$ than if $\sigma \in [\underline{\sigma}, \hat{\sigma}) \cup (\tilde{\sigma}, \overline{\sigma}]$. Since $\int_{\underline{\omega}}^{t}\left(F(\omega|x_h) - F(\omega|x_\ell)\right)d\omega \leq 0$ by MPS, we obtain that $\int_{\underline{\omega}}^{t} H_1(\omega|\sigma')d\omega \geq \int_{\underline{\omega}}^{t} H_1(\omega|\sigma'')d\omega$ for all $\sigma' \in [\underline{\sigma}, \hat{\sigma}) \cup (\tilde{\sigma}, \overline{\sigma}]$ and $\sigma'' \in [\underline{\sigma}, \hat{\sigma}) \cup (\tilde{\sigma}, \overline{\sigma}]$. We have thus shown that *under PAM and MPS, $H_1(\omega|\sigma)$ has a higher riskiness when σ is either low or high than for intermediate values*, as asserted in the text.

Consider now the NAM case. All we need to modify is the fact that now $G(\sigma|\mu(x_h))$ differs from $G(\sigma|\mu(x_\ell))$ by a MPS (since x_h is now matched to y_ℓ, who has a higher riskiness in G). Then

$$\frac{d\left(G(\sigma|\mu(x_h)) - G(\sigma|\mu(x_\ell))\right)}{d\sigma} = g(\sigma|\mu(x_\ell)) - g(\sigma|\mu(x_h)) = g(\sigma|\mu(x_\ell))\left(\frac{g(\sigma|\mu(x_h))}{g(\sigma|\mu(x_\ell))} - 1\right).$$

Proceeding as before, we obtain that *under NAM, $H(\omega|\sigma)$ has a lower riskiness when σ is either low or high than for intermediate values*, completing the stochastic sorting analysis of this case. □

A.3 Proof of Proposition 3

The first-order condition of the risk sharing problem is

$$\left(\frac{a(q(\omega,\sigma)-c(\omega,\sigma))}{1-\alpha}+b\right)^{\alpha-1}=\lambda\left(\frac{ac(\omega,\sigma)}{1-\alpha}+b\right)^{\alpha-1}\quad\forall(\omega,\sigma),$$

where λ is the multiplier of the constraint. Solving for $c(\omega,\sigma)$ we obtain

$$c(\omega,\sigma)=\frac{aq(\omega,\sigma)+b(1-\alpha)\left(1-\lambda^{\frac{1}{\alpha-1}}\right)}{a\left(1+\lambda^{\frac{1}{\alpha-1}}\right)}.$$

Inserting $c(\omega,\sigma)$ into the constraint

$$\frac{1-\alpha}{\alpha}\left(\int_{\underline{\omega}}^{\overline{\omega}}\int_{\underline{\sigma}}^{\overline{\sigma}}\left(\frac{ac(\omega,\sigma)}{1-\alpha}+b\right)^{\alpha}h(\omega,\sigma|x,y)d\sigma d\omega-1\right)=v,$$

reveals that

$$1+\lambda^{\frac{1}{\alpha-1}}=\left(\frac{\int_{\underline{\omega}}^{\overline{\omega}}\int_{\underline{\sigma}}^{\overline{\sigma}}\left(\frac{aq(\omega,\sigma)}{1-\alpha}+2b\right)^{\alpha}h(\omega,\sigma|x,y)d\sigma d\omega}{\frac{v\alpha}{1-\alpha}+1}\right)^{\frac{1}{\alpha}}.$$

The function Φ then becomes

$$\begin{aligned}\Phi(x,y,v)&=\frac{1-\alpha}{\alpha}\left(\int_{\underline{\omega}}^{\overline{\omega}}\int_{\underline{\sigma}}^{\overline{\sigma}}\left(\frac{a(q(\omega,\sigma)-c(\omega,\sigma))}{1-\alpha}+b\right)^{\alpha}h(\omega,\sigma|x,y)d\sigma d\omega-1\right)\\&=\frac{1-\alpha}{\alpha}\left(\frac{\lambda^{\frac{\alpha}{\alpha-1}}}{\left(1+\lambda^{\frac{1}{\alpha-1}}\right)^{\alpha}}\int_{\underline{\omega}}^{\overline{\omega}}\int_{\underline{\sigma}}^{\overline{\sigma}}\left(\frac{aq(\omega,\sigma)}{1-\alpha}+2b\right)^{\alpha}h(\omega,\sigma|x,y)d\sigma d\omega-1\right)\\&=\frac{1-\alpha}{\alpha}\left(\left(V^{\frac{1}{\alpha}}(x,y)-\left(\frac{v\alpha}{1-\alpha}+1\right)^{\frac{1}{\alpha}}\right)^{\alpha}-1\right),\end{aligned}$$

where the second line follows by substituting the expression for $c(\omega,\sigma)$, and the third by using the expression for λ and equation (7). This shows equation (6).

Differentiating Φ yields, after manipulation, that the sign of $\Phi_{xy}-(\Phi_y/\Phi_v)\Phi_{vx}$ is equal to the sign of $((1-\alpha)/\alpha)((1-\alpha)/\alpha)V_xV_y+VV_{xy})$, completing the proof of the proposition. □.

Regarding the case $\alpha = 0$ mentioned in the text, we first note that

$$\lim_{\alpha\to 0}\frac{(1-\alpha)}{\alpha}\left(\left(\frac{ac}{1-\alpha}+b\right)^{\alpha}-1\right) = \lim_{\alpha\to 0}(1-\alpha)^{1-\alpha}\left(\frac{(ac+b(1-\alpha))^{\alpha}-1}{\alpha}\right)$$
$$= \lim_{\alpha\to 0}\frac{(ac+b(1-\alpha))^{\alpha}\log(ac+b(1-\alpha))}{1}$$
$$= \log(ac+b),$$

where the first equality follows by rewriting, the second by L'Hopital's rule, and the third by taking limit. From the first-order condition we obtain $(a(q(\omega,\sigma)-c(\omega,\sigma))+b)^{-1} = \lambda(ac(\omega,\sigma)+b)^{-1}$ and thus

$$c(\omega,\sigma) = \frac{\lambda aq(\omega,\sigma)+(\lambda-1)b}{a(1+\lambda)}.$$

Inserting this expression in the constraint we obtain after some manipulation that $\lambda/(1+\lambda) = e^{v-V(x,y)}$, where $V(x,y) = \int_{\underline{\omega}}^{\overline{\omega}}\int_{\underline{\sigma}}^{\overline{\sigma}}\log(aq(\omega,\sigma)+2b)h(\omega,\sigma|x,y)d\sigma d\omega$. Now we can derive Φ as follows:

$$\Phi(x,y,v) = \int_{\underline{\omega}}^{\overline{\omega}}\int_{\underline{\sigma}}^{\overline{\sigma}}\log(a(q(\omega,\sigma)-c(\omega,\sigma)+b)h(\omega,\sigma|x,y)d\sigma d\omega$$
$$= \int_{\underline{\omega}}^{\overline{\omega}}\int_{\underline{\sigma}}^{\overline{\sigma}}\log\left(\frac{1}{1+\lambda}(aq(\omega,\sigma)+2b)\right)h(\omega,\sigma|x,y)d\sigma d\omega$$
$$= \log\left(1-\frac{\lambda}{1+\lambda}\right)+V(x,y)$$
$$= \log\left(e^{V(x,y)}-e^{v}\right).$$

This is TU representable as $\hat{\Phi}(x,y,v) = e^{V(x,y)}-e^{v}$, and thus sorting depends on the sign of the cross partial of $e^{V(x,y)}$, which reduces to the sign of $V_{xy}+V_xV_y$ as asserted in the text.

A.4 Derivation of Optimal Contract under Moral Hazard

This problem can be written as[30]

$$\max_{\beta,\alpha,e} \quad \int\left[(1-\alpha(\omega))(\omega t+\omega e(\omega))-\beta(\omega)\right]dF(\omega|x) \tag{19}$$

$$\text{s.t.} \quad \int -e^{-r\left(\beta(\omega)+\alpha(\omega)(\omega t+\omega e(\omega))-\frac{e^2}{2}-\frac{r\alpha(\omega)^2\omega^2 s^2}{2}\right)}dF(\omega|x) \geq -e^{-ra} \tag{20}$$

$$\alpha(\omega) = \frac{e(\omega)}{\omega}, \quad \forall\omega. \tag{21}$$

[30]Where we use the fact that if X is normally distributed with mean μ and variance σ^2, then $\mathbb{E}\left[e^{\lambda X}\right] = e^{\lambda\mu+\frac{1}{2}\lambda^2\sigma^2}$.

Substituting for α from the IC constraint, this problem is equivalent to:

$$\max_{\beta,e} \quad \int \left[\left(1 - \frac{e(\omega)}{\omega}\right) (\omega t + \omega e(\omega)) - \beta(\omega) \right] dF(\omega|x) \tag{22}$$

$$\text{s.t.} \quad 1 - \int e^{-r\left(\beta(\omega)+e(\omega)(t+e(\omega))-\frac{e^2}{2}-\frac{re(\omega)^2 s^2}{2}-a\right)} dF(\omega|x) \geq 0. \tag{23}$$

Denoting the multiplier on the constraint in the Lagrangian L by λ, and maximizing pointwise for each ω, we obtain:

$$\begin{aligned} \frac{\partial L}{\partial \beta} = -1 f(\omega|x) + r\lambda e^{-rA} f(\omega|x) = 0 \\ \Rightarrow \lambda = \frac{1}{r} e^{rA} \end{aligned} \tag{24}$$

where $A = \beta(\omega) + e(\omega)\left(t + e(\omega)\right) - \frac{e^2}{2} - \frac{re(\omega)^2 s^2}{2} - a$.

$$\begin{aligned} \frac{\partial L}{\partial e} &= \left(-(t+e) + (\omega - e) + r\lambda e^{-rA}\left(t + 2e - e - res^2\right)\right) f(\omega|x) = 0 \\ &= \left(-e + \omega - res^2\right) f(\omega|x) = 0 \\ &\Rightarrow e(\omega) = \frac{\omega}{1+rs^2}, \quad \forall \omega, \end{aligned}$$

where we have used $\lambda r e^{rA} = 1$ from (24).

Using this solution for e in the first order condition (24), we obtain:

$$\lambda = \frac{1}{r} e^{r\left[\beta + \frac{1}{1+rs^2}\left(\omega t + \frac{\omega^2}{1+rs^2}\right) - \frac{\omega^2}{2\left(1+rs^2\right)^2} - \frac{r\omega^2 s^2}{2\left(1+rs^2\right)^2} - a\right]}.$$

The Lagrange multiplier is by construction a constant. Now this equation has to hold for each ω, so that the only way this can be satisfied is if the exponent is equal to 0 so the RHS is a constant as well for all ω.[31] This implies that $\lambda = \frac{1}{r} > 0$. Hence,

$$\begin{aligned} \beta &= a - \frac{\omega t}{1+rs^2} - \frac{\omega^2}{\left(1+rs^2\right)^2} + \frac{\omega^2}{2\left(1+rs^2\right)^2} + \frac{r\omega^2 s^2}{2\left(1+rs^2\right)^2} \\ &= a - \frac{\omega t}{1+rs^2} + \frac{\omega^2}{2\left(1+rs^2\right)^2}\left(rs^2 - 1\right). \end{aligned}$$

[31] Notice that for each ω, this is exactly the same solution as if there was no uncertainty about ω, or equivalently, for each ω we solve the standard Holmström and Milgrom (1987) problem without the "expected" participation constraint (integrated over ω). This is a *result*, not an assumption.

The solution to the contracting problem is the quadruple $(\alpha(\omega), \beta(\omega), e(\omega), \lambda)$:

$$\begin{aligned}
\alpha(\omega) &= \frac{1}{1+rs^2} \\
\beta(\omega) &= a - \frac{\omega t}{1+rs^2} + \frac{\omega^2}{2\left(1+rs^2\right)^2}\left(rs^2-1\right) \\
e(\omega) &= \frac{\omega}{1+rs^2} \\
\lambda &= \frac{1}{r}
\end{aligned}$$

Using the expression for the optimal effort level, we obtain:

$$q(\omega, \sigma, y) = \frac{\omega^2}{1+rs^2} + \omega(t+\sigma), \tag{25}$$

which depends on y through s^2 and t. Since $w(\omega, q) = \beta(\omega) + \alpha(\omega)q$, it follows that

$$w(\omega, q) = a + \frac{\omega^2}{2\left(1+rs^2\right)} + \frac{\omega\sigma}{1+rs^2}. \tag{26}$$

Finally, using the fact that $\pi(\omega, q) = q - w(\omega, q)$ we obtain:

$$\pi(\omega, q) = \omega t - a + \frac{\omega^2}{2\left(1+rs^2\right)} + \frac{rs^2\omega\sigma}{1+rs^2}, \tag{27}$$

which completes the derivation of the optimal contract.

A.5 Proof of Proposition 4

Integration by parts yields

$$V(x, y) = \overline{q}(\overline{\omega}, y) - \int_{\underline{\omega}}^{\overline{\omega}} \overline{q}_\omega(\omega, y) F(\omega|x) d\omega.$$

Part 1. follows from the cross partial derivative $V_{xy} = \int_{\underline{\omega}}^{\overline{\omega}} \overline{q}_{\omega y}(\omega, y)(-F_x(\omega|x))d\omega$, which under FOSD is nonnegative if q is spm and nonpositive if q is sbm.

Another integration by parts yields

$$V(x, y) = \overline{q}(\overline{\omega}, y) - \overline{q}_\omega(\overline{\omega}, y) \int_{\underline{\omega}}^{\overline{\omega}} F(\omega|x) d\omega + \int_{\underline{\omega}}^{\overline{\omega}} \overline{q}_{\omega\omega}(\omega, y) \left(\int_{\underline{\omega}}^{\omega} F(s|x) ds \right) d\omega.$$

Since under IR the mean remains constant and $\int_{\underline{\omega}}^{\omega} F(s|x)ds$ decreases in x, it follows that the second term is independent of x the cross partial derivative of V is given by $V_{xy} = \int_{\underline{\omega}}^{\overline{\omega}} \overline{q}_{\omega\omega y}(\omega, y) \left(\int_{\underline{\omega}}^{\omega} F_x(s|x) \right) ds d\omega$, which is nonnegative if $q_{\omega\omega y} \leq 0$ and nonpositive if $q_{\omega\omega y} \geq 0$. This proves part 2. of the proposition. □

Appendix B Mismatched CEOs

B.1 Derivation of Expected Wages, Profits, and Output

Given that $\sigma \sim \mathcal{N}(0, s^2(x))$ and $\omega \sim \mathcal{LN}(k(x), m^2(x))$, we can derive explicit expressions for expected wages, expected profits, and expected output:[32]

$$\begin{aligned}
\mathbb{E}w(x) &= \int\int w(\omega,\sigma;x)dF(\omega|k(x),m(x))dG(\sigma|0,s(x)) \\
&= a(x) + \int \frac{\omega^2}{2(1+rs^2)}dF(\omega|k,m) \\
&= a(x) + \frac{e^{2(k+m^2)}}{2(1+rs^2)} \\
\mathbb{E}\pi(y) &= \int\int \pi(\omega,\sigma;y)dF(\omega|k(y),m(y))dG(\sigma|0,s(y)) \\
&= \int \left(\omega t - a(x) + \frac{\omega^2}{2\left(1+rs^2\right)}\right) dF(\omega|k(y),m(y)) \\
&= e^{k+\frac{m^2}{2}}t - a(x) + \frac{e^{2(k+m^2)}}{2(1+rs^2)}.
\end{aligned}$$

Since $V(x,y) = \mathbb{E}w(x) + \mathbb{E}\pi(y)$ we immediately obtain:

$$V(x,y) = \mathrm{e}^{k+\frac{m^2}{2}}t + \frac{\mathrm{e}^{2(k+m^2)}}{1+rs^2}. \tag{28}$$

Expected output is evaluated over the entire domain of pairs (x,y). Notice that it can be decomposed into two components. The first, $\mathrm{e}^{k+\frac{m^2}{2}}t$, measures the contribution to the match value from the expected types $\mathbb{E}\omega$ and $\mathbb{E}\sigma$. This term is similar to the standard one in Becker (1973) and in the applied models of Gabaix and Landier (2008) and Terviö (2008). The second component is due to moral hazard and incentive provision. As is well known in the Holmström and Milgrom (1987) model, a higher variance in output s^2 reduces match output since incentive provision is weaker for any level of effort, and the effort implemented is also lower. But this term is increasing in the expected value of ω^2, which is equal to $e^{2(k+m^2)}$ since ω is lognormally distributed.

We can also immediately derive the expected return:

$$\mathbb{E}R(y) = \frac{\mathbb{E}\pi(y)}{V_0(y)} = \frac{e^{k+\frac{m^2}{2}}t}{V_0} - \frac{a}{V_0} + \frac{e^{2(k+m^2)}}{2V_0(1+rs^2)}.$$

We can also explicitly derive $a(x)$, the expression which is given by $a(x) = a(\underline{x}) + \int_{\underline{x}}^{x} V_x(z,\mu(z))dz$,

[32] We use the fact that with $F(x) \sim \mathcal{LN}(\mu,\sigma^2)$, $\int x^2 dF(x) = e^{2(\mu+\sigma^2)}$ and $\int x dF(x) = e^{e^{\mu+\frac{\sigma^2}{2}}}$.

where $\mu(z) = z$. From $V(x,y)$ we can write:

$$V_x(z,z) = e^{k+\frac{m^2}{2}}(k' + mm')t + \frac{e^{2(k+m^2)}(k' + mm')}{1 + rs^2},$$

and hence

$$a(x) = a(\underline{x}) + \int_{\underline{x}}^{x} \left(e^{k(z)+\frac{m(z)^2}{2}}(k'(z) + m(z)m'(z))t(z) + \frac{e^{2(k(z)+m^2(z))}(k'(z) + 2u(z)m(z)')}{1 + rs(z)^2} \right) dz.$$

To verify that the identifying assumption – that there is PAM – is satisfied, we need to check whether the match surplus function is spm, that is, $V_{xy} \geq 0$, where

$$V_{xy}(x,y) = \mathrm{e}^{\frac{m^2}{2}+k} \left(m\,m' + k'\right) t' - \frac{4\,r\,\mathrm{e}^{2\,m^2+2\,k} \left(2\,m\,m' + k'\right) s\,s'}{\left(1 + r\,s^2\right)^2} \tag{29}$$

Notice that even if k, m, t, s are increasing in x or y, equation (29) may be negative due to the second term. The first term, when positive, is the standard Becker force towards PAM. The second term, due to moral hazard, is submodular because s is increasing in y, and thus it is a force towards NAM. The monotonicity of the variance makes the verification of the PAM assumption nontrivial.

B.2 Likelihood Function

With F log-normal and G normally distributed, the log-likelihood function to be maximized can be written as:

$$\ln \mathcal{L}(\theta|w,\pi,x) \;\; = \;\; -\sum_{i=1}^{n} \frac{1}{\ln \omega_i} \frac{\left[\ln \omega_i - k(x)\right]^2}{2u(x)^2} - \sum_{i=1}^{n} \frac{\left[\sigma_i - t(x)\right]^2}{2s(x)^2} + \ln|J| - n\left[\ln m(x) + \ln s(x) + \ln(2\pi)\right],$$

where $k(x)$ and $u(x)$ are the type-dependent mean and standard deviation of ω, and $t(x)$ and $s(x)$ are the type-dependent mean and standard deviation of σ, and the latter are evaluated along the equilibrium allocation $y = x$. We obtain expression for ω and σ from solving (12)–(13):

$$\begin{aligned} w &= a + \frac{\omega^2}{2\left(1 + rs^2\right)} + \frac{\omega}{1 + rs^2}\sigma \\ \pi &= \omega t - a + \frac{\omega^2}{2\left(1 + rs^2\right)} + \frac{rs^2}{1 + rs^2}\omega\sigma \end{aligned}$$

substituting one in the other:

$$\pi \;\; = \;\; \omega t - a + \frac{\omega^2}{2\left(1 + rs^2\right)} + rs^2 \left(w - a - \frac{\omega^2}{2\left(1 + rs^2\right)} \right)$$

gives a quadratic equation in ω.

$$
\begin{aligned}
-\pi + \omega t - (1+rs^2)a + rs^2 w + (1-rs^2)\frac{\omega^2}{2\,(1+rs^2)} &= 0 \\
\frac{1-rs^2}{2\,(1+rs^2)}\omega^2 + \omega t - (1+rs^2)a + rs^2 w - \pi &= 0 \\
(1-rs^2)\omega^2 + 2\left(1+rs^2\right)t\omega - 2(1+rs^2)^2 a + (rs^2 w - \pi)2\left(1+rs^2\right) &= 0 \\
(-1+rs^2)\omega^2 - 2t\left(1+rs^2\right)\omega + 2(1+rs^2)\left(\pi + a - rs^2(w-a)\right) &= 0
\end{aligned}
$$

The discriminant (divided by 4) of this quadratic form is:

$$
\begin{aligned}
D &= t^2(1+rs^2)^2 - 2(1+rs^2)(-1+rs^2)\left(\pi + a - rs^2(w-a)\right) \\
&= (1+rs^2)\left[t^2(1+rs^2) - (-1+rs^2)2\left(\pi + a - rs^2(w-a)\right)\right]
\end{aligned}
$$

and then the solutions for ω are:

$$
\omega = \frac{t\left(1+rs^2\right) \pm \sqrt{D}}{-1+rs^2}
$$

and using (12) to solve for σ we obtain:

$$
\begin{aligned}
\sigma &= (w-a)\frac{1+rs^2}{\omega} - \frac{\omega}{2} \\
&= (w-a)\frac{1+rs^2}{\frac{t(1+rs^2)\pm\sqrt{D}}{-1+rs^2}} - \frac{\frac{t\left(1+rs^2\right)\pm\sqrt{D}}{-1+rs^2}}{2} \\
&= (w-a)\frac{(1+rs^2)(-1+rs^2)}{t\,(1+rs^2)\pm\sqrt{D}} - \frac{t\left(1+rs^2\right)\pm\sqrt{D}}{2(-1+rs^2)}
\end{aligned}
$$

So we get:

$$
\begin{aligned}
\omega &= \frac{t(1+rs^2)+\sqrt{D}}{-1+rs^2} \\
\sigma &= (w-a)\frac{(1+rs^2)(-1+rs^2)}{t\,(1+rs^2)\pm\sqrt{D}} - \frac{t\left(1+rs^2\right)\pm\sqrt{D}}{2(-1+rs^2)}
\end{aligned}
$$

(with $D = \left(rs^2+1\right)\left(2a + 2\pi + t^2 - 2\pi rs^2 - 2rs^2 w - 2ar^2s^4 + rs^2t^2 + 2r^2s^4 w\right)$) and where $|J|$ is the Jacobian of the transformation:

$$
|J| = \left| \begin{array}{cc} \frac{\partial\omega}{\partial w} & \frac{\partial\omega}{\partial\pi} \\ \frac{\partial\sigma}{\partial w} & \frac{\partial\sigma}{\partial\pi} \end{array} \right|.
$$

Mean ω			st. dev. ω		
k_0	k_1	k_2	m_0	m_1	m_2
9.26	0.01	0.00	0.72	-0.02	0.00
Mean σ			st. dev. σ		
t_0	t_1	t_2	s_0	s_1	s_2
75.36	0.07	0.00	296.31	2.95	0.00

Table 2: Estimated Parameters.

B.3 Estimated Parameters

B.4 Cost of Mismatch: Decomposition

Denote the realization of ω, σ by $\hat{\omega}, \hat{\sigma}$. Then we can construct the order of the ex-post types $\tilde{x}$ and rematch $\tilde{x} = y$. This implies not only a new matched partner but also a new level of effort by the CEO $\tilde{e} = \hat{\omega}/(1+rs(\tilde{x})^2)$. ex-post output is equal to $q = \omega\left(e + t(y) + \sigma\right)$, and in the case of the original allocation we write it as $q(x,e)$. The new output after rematching and adjusting effort is written as $q(\tilde{x},\tilde{e})$. We can also write output *as if* there was a new allocation $\tilde{x} = y$ but where effort was as under the original allocation, $e(x) = \hat{\omega}/(1 + rs(x)^2)$ denoted by $q(\tilde{x}, e)$. These three expressions are equal to:

$$\begin{aligned}
q(x,e) &= \hat{\omega}\left(\frac{\hat{\omega}}{1+rs(x)^2} + t(x) + \hat{\sigma}\right) \\
q(\tilde{x},e) &= \hat{\omega}\left(\frac{\hat{\omega}}{1+rs(x)^2} + t(\tilde{x}) + \hat{\sigma}\right) \\
q(\tilde{x},\tilde{e}) &= \hat{\omega}\left(\frac{\hat{\omega}}{1+rs(\tilde{x})^2} + t(\tilde{x}) + \hat{\sigma}\right).
\end{aligned}$$

We are interested in the output change after ω is realized, and before σ. We therefore take the expectation of match output with respect to σ: $\mathbb{E}_\sigma q(x,e) = \hat{\omega}\left(\frac{b(x)\hat{\omega}}{1+rs(x)^2} + t(x)\right)$, where σ simply cancels because it is additive and has zero mean.

Our objective is to decompose the total output change from rematching, and which part is due to the change in the allocation $\tilde{x}$ and which part is due to the change in effort $\tilde{e}$. Denote by $\Delta(\tilde{x},\tilde{e}|x,e) = \mathbb{E}_\sigma q(\tilde{x},\tilde{e}) - \mathbb{E}_\sigma q(x,e)$ the change in output from both a change in the allocation and the effort; by $\Delta(\tilde{x},\tilde{e}|\tilde{x},e) = \mathbb{E}_\sigma q(\tilde{x},\tilde{e}) - \mathbb{E}_\sigma q(\tilde{x},e)$ the change in output from a change in effort while keeping the allocation unchanged at the ex-post level; and by $\Delta(\tilde{x},e|x,e) = \mathbb{E}_\sigma q(\tilde{x},e) - \mathbb{E}_\sigma q(x,e)$ the change in output from a change in the allocation while keeping the effort unchanged at the original level. Then:

$$\begin{aligned}
\Delta(\tilde{x},\tilde{e}|x,e) &= \Delta(\tilde{x},\tilde{e}|\tilde{x},e) + \Delta(\tilde{x},e|x,e) \\
&= \underbrace{\hat{\omega}\frac{\hat{\omega}}{1+rs(\tilde{x})^2}}_{\text{output change due to effort}} + \underbrace{\hat{\omega}\frac{b(x)\hat{\omega}}{1+rs(x)^2} + \hat{\omega}t(\tilde{x}) - \hat{\omega}t(x)}_{\text{output change due to mismatch}}.
\end{aligned}$$

And the total output gain from this exercise is:

$$\sum_x \Delta(\tilde{x}, \tilde{e}|x, e) = \sum_x \Delta(\tilde{x}, \tilde{e}|\tilde{x}, e) + \sum_x \Delta(\tilde{x}, e|x, e),$$

where $\tilde{x}$ and $\tilde{e}$ depend on x. Observe that the output gain of this rematching is positive provided expected output $\mathbb{E}_\sigma q(\omega, y)$ is supermodular. We can readily verify that this is the case for the estimated technology and for the domain of observed realizations of ω. Note that supermodularity $V(x, y)$ does not necessarily imply supermodularity of $\mathbb{E}_\sigma q(\omega, y)$, though they are closely related.

Appendix C Household Income Inequality

C.1 Additional Table – Table 3

	1	2	3	4
4	0.5140	0.9007	0.9537	4.0381
3	1.9580	2.6024	2.4850	3.3010
2	13.7884	15.7335	4.7507	3.9461
1	34.8495	7.6044	1.7847	0.7898

(a) $m(x, y)$ – 1960

	1	2	3	4
4	0.5180	5.9007	7.6423	28.1259
3	1.1865	8.5312	9.1192	5.9766
2	2.6114	14.2750	5.4522	3.4372
1	4.2252	2.0709	0.5920	0.3356

(b) $m(x, y)$ – 2014

	1	2	3	4
4	0	0	5.6685	6.4064
3	0	5.2962	4.6779	0
2	0	26.841	0	0
1	45.028	6.0814	0	0

(c) $m^f(x, y)$ – 1960

	1	2	3	4
4	0	0	0	37.875
3	0	0	18.494	4.3116
2	0	24.458	6.3194	0
1	7.2238	1.3173	0	0

(d) $m^f(x, y)$ – 2014

Table 3: Distribution of Marriages by Education (in percentage): actual and frictionless.

C.2 Derivation of the Formulas for Table 1

We are assuming that for each pair (x, y), wages are jointly normally distributed, i.e.

$$(\omega, \sigma; x, y) \sim \mathcal{N}(\mu_\omega[x, y], \mu_\sigma[x, y], s_\omega[x, y], s_\sigma[x, y], \rho[x, y]).$$

Then the marginals $F(\omega|x, y) = \mathcal{N}(\mu_\omega[x, y], sd_\omega[x, y])$ and $G(\sigma|x, y) = \mathcal{N}(\mu_\sigma[x, y], sd_\sigma[x, y])$ are normal. Using the general formula:

$$\begin{aligned} \mathrm{Var}(\omega + \sigma|x, y) &= \mathrm{Var}(\omega|x, y) + \mathrm{Var}(\sigma|x, y) + 2\mathrm{Cov}(\omega, \sigma|x, y) \\ &= s_\omega^2[x, y] + s_\sigma^2[x, y] + 2\rho[x, y] s_\omega[x, y] s_\sigma[x, y] \end{aligned} \tag{30}$$

Normality of the joint distribution of (ω, σ) implies that $\omega + \sigma$ is normally distributed. To calculate (30) for each pair (x, y) we use the sample means and standard deviations as well as the sample covariance:

$$\begin{aligned}
\hat{\mu}_\omega[x,y] &= \mathbb{E}[\omega|x,y] \\
\hat{s}^2_\omega[x,y] &= \text{Var}[\omega|x,y] \\
\hat{\mu}_\sigma[x,y] &= \mathbb{E}[\sigma|x,y] \\
\hat{s}^2_\sigma[x,y] &= \text{Var}[\sigma|x,y] \\
\hat{\rho}[x,y] &= \frac{\text{Cov}(\omega,\sigma|x,y)}{\sqrt{\text{Var}\ [[\omega|x,y]]\ \text{Var}\ [[\sigma|x,y]]}}.
\end{aligned}$$

To calculate the $\text{Var}[\omega+\sigma]$ for the *entire sample*, we use again the general formula for the variance of the sum of two random variables:

$$\text{Var}[\omega+\sigma] = \text{Var}[\omega] + \text{Var}[\sigma] + 2\text{Cov}[\omega,\sigma]. \tag{31}$$

where

$$\begin{aligned}
\text{Cov}[\omega,\sigma] =& \mathbb{E}\left[\mathbb{E}[\omega\sigma|x,y]\right] - \mathbb{E}\left[\mathbb{E}[\omega|x,y]\right]\mathbb{E}\left[\mathbb{E}[\sigma|x,y]\right] \\
=& \mathbb{E}\left[\text{Cov}[\omega,\sigma|x,y] + \mu_\omega[x,y]\mu_\sigma[x,y]\right] - \mathbb{E}\left[\mu_\omega[x,y]\right]\mathbb{E}\left[\mu_\sigma[x,y]\right] \\
=& \sum_x\sum_y \rho[x,y]s_\omega[x,y]s_\sigma[x,y]m(x,y) + \text{Cov}(\mu_\omega[x,y],\mu_\sigma[x,y]) \\
\text{Var}[\omega] =& \sum_x\sum_y s^2_\omega[x,y]m(x,y) \\
\text{Var}[\sigma] =& \sum_x\sum_y s^2_\sigma[x,y]m(x,y).
\end{aligned}$$

Then, substituting these in the expression for the variance (31), we obtain:

$$\begin{aligned}
\text{Var}[\omega+\sigma] =& \sum_x\sum_y s^2_\omega[x,y]m(x,y) + \sum_x\sum_y s^2_\sigma[x,y]m(x,y) + 2\sum_x\sum_y \rho[x,y]s_\omega[x,y]s_\sigma[x,y]m(x,y) \\
&+ 2\sum_x\sum_y \mu_\omega[x,y]\mu_\sigma[x,y]m(x,y) - 2\sum_x\sum_y \mu_\omega[x,y]m(x,y)\sum_x\sum_y \mu_\sigma[x,y]m(x,y).
\end{aligned} \tag{32}$$

To calculate the variance using this formula for a given year (1960 or 2014), we use the sample means, sample variances, sample correlations and the matchings for that year: $\hat{\mu}_\omega[x,y], \hat{s}^2_\omega[x,y], \hat{\mu}_\sigma[x,y], \hat{s}^2_\sigma[x,y], \hat{\rho}[x,y], m(x,y)$.

C.2.1 Stochastic Sorting, Marginals: $\text{Var}_{2014}(\omega+\sigma; F_{1960}[x,y], G_{1960}[x,y])$

Apply equation (32) with all variables from 2014, except those that relate to F and G:[33]

$$\begin{aligned}
\text{Var}[\omega+\sigma] =& \sum_x\sum_y s^2_{\omega,1960}[x,y]m(x,y) + \sum_x\sum_y s^2_{\sigma,1960}[x,y]m(x,y) + 2\sum_x\sum_y \rho[x,y]s_{\omega,1960}[x,y]s_{\sigma,1960}[x,y]m(x,y) \\
&+ 2\sum_x\sum_y \mu_{\omega,1960}[x,y]\mu_{\sigma,1960}[x,y]m(x,y) - 2\sum_x\sum_y \mu_{\omega,1960}[x,y]m(x,y)\sum_x\sum_y \mu_{\sigma,1960}[x,y]m(x,y).
\end{aligned}$$

[33] If there is no year subscript in the variable, it is assumed to be for 2014. When it is for 1960 it is made explicit.

C.2.2 Stochastic Sorting, Correlation: $\mathbf{Var}_{2014}(\omega + \sigma; \rho_{1960}[x, y])$

Apply equation (32) with all sample means from 2014, except those that relate to ρ:

$$\begin{aligned}\mathrm{Var}[\omega+\sigma] = &\sum_x\sum_y s_\omega^2[x,y]m(x,y) + \sum_x\sum_y s_\sigma^2[x,y]m(x,y) + 2\sum_x\sum_y \rho_{1960}[x,y]s_\omega[x,y]s_\sigma[x,y]m(x,y)\\ &+ 2\sum_x\sum_y \mu_\omega[x,y]\mu_\sigma[x,y]m(x,y) - 2\sum_x\sum_y \mu_\omega[x,y]m(x,y)\sum_x\sum_y \mu_\sigma[x,y]m(x,y).\end{aligned}$$

C.2.3 Total Earnings: $\mathbf{Var}_{2014}(\omega + \sigma; H_{1960}[x, y])$

Here we use the sample distribution of joint earnings H in 1960, without the normality assumption:

$$\begin{aligned}\mathrm{Var}_{2014}[\omega+\sigma; H_{1960}[x,y]] = &\left(\sum_\omega\sum_\sigma(\omega+\sigma)^2\sum_x\sum_y h_{1960}(\omega,\sigma|x,y)m_{2014}(x,y)\right)\\ &-\left(\sum_\omega\sum_\sigma(\omega+\sigma)\sum_x\sum_y h_{1960}(\omega,\sigma|x,y)m_{2014}(x,y)\right)^2\\ =&\sum_x\sum_y\left(\sum_\omega\sum_\sigma(\omega+\sigma)^2 h_{1960}(\omega,\sigma|x,y)m_{2014}(x,y)\right)\\ &-\left(\sum_x\sum_y\left(\sum_\omega\sum_\sigma(\omega+\sigma)h_{1960}(\omega,\sigma|x,y)\right)m_{2014}(x,y)\right)^2.\end{aligned}$$

C.2.4 Marital Sorting, Marginals: $\mathbf{Var}_{2014}(\omega + \sigma; \Gamma_{1960}[x, y], \Psi_{1960}[x, y])$

We now keep all the features of the income distribution H constant, and we vary the distribution of matches M. To decompose the change in matching due to a change in the marginals $\Gamma(x)$ and $\Psi(y)$ and the mismatch, we use the total variation norm as the measure of distance $d = \frac{1}{2}\sum_x\sum_y |m(x,y) - m^f(x,y)|$. Adjusting a distribution for d, while keeping the marginals the same has a continuum of such solutions. Because the measure d can be approximated by $d \approx 1 - \delta$ where $\delta = \sum_x\sum_y m(x,y)\mathbb{I}_{m^f>0}$, i.e., δ is the sum of values of $m(x,y)$ that are on the frictionless allocation.[34] In the data, these values are equal to $d_{1960} = 0.3326, 1 - \delta_{1960} = 0.3173, d_{2014} = 0.3317, 1 - \delta_{2014} = 0.3075$.

To that effect, we calculate the variance in for 2014 assuming that the marginals $\Gamma(x), \Psi(y)$ are from 1960.

[34] We can prove that d is exactly equal to $1 - \delta$ under the assumption that $m > m^f$ if $m^f = 0$ and $m < m^f$ if $m^f > 0$. To see this, write

$$\begin{aligned} d &= \frac{1}{2}\left(\sum_x\sum_y[m(x,y) - m^f(x,y)]\mathbb{I}_{m^f=0} - [m(x,y) - m^f(x,y)]\mathbb{I}_{m^f>0}\right)\\ &= \frac{1}{2}\left(\sum_x\sum_y[1 - m(x,y)]\mathbb{I}_{m^f>0} - [m(x,y) - 1]\mathbb{I}_{m^f>0}\right)\\ &= \frac{1}{2}\left(\sum_x\sum_y[2 - 2m(x,y)]\mathbb{I}_{m^f>0}\right) = 1 - \sum_x\sum_y m(x,y)\mathbb{I}_{m^f>0} = 1 - \delta,\end{aligned}$$

since $\sum_x\sum_y m(x,y) = 1$ and $\sum_x\sum_y m^f(x,y)\mathbb{I}_{m^f>0} = 1$.

We use the distributions from 1960 adjusting them so that the mismatch is as in 2014:

$$m_{marginals} = \begin{cases} m_{1960}(x,y)\frac{1-d_{2014}}{1-d_{1960}} & \text{if } m^f_{1960}(x,y) > 0 \\ m_{1960}(x,y)\frac{d_{2014}}{d_{1960}} & \text{if } m^f_{1960}(x,y) = 0 \end{cases}$$

$$\begin{aligned}\text{Var}_{2014}[\omega+\sigma; M_{1960}[x,y]] = &\left(\sum_\omega\sum_\sigma(\omega+\sigma)^2\sum_x\sum_y h_{2014}(\omega,\sigma|x,y)m_{marginals}(x,y)\right)\\ &-\left(\sum_\omega\sum_\sigma(\omega+\sigma)\sum_x\sum_y h_{2014}(\omega,\sigma|x,y)m_{marginals}(x,y)\right)^2.\end{aligned}$$

C.2.5 Marital Sorting, Allocation: Gender equalization: Var$_{2014}(\omega+\sigma; \Gamma_{1960}, \tilde{\Psi}_{1960}, \mu_{2014}[x,y][x,y])$

We change the mass of education to that of 1960, but we keep the equilibrium matching $\mu_{2014}(x)$. To do so, we need to adjust at least one of the distributions, in this case we change $\Psi(y)$ to $\tilde{\Psi}(y)$. Hence, we have $\Gamma_{1960}, \mu_{2014}$ and $\tilde{\Psi}_{1960}$ such that: $\tilde{\Psi}_{1960}(\mu_{2014}(x)) = \Gamma_{1960}$. With discrete type distributions, this means that the new matching allocation μ must yield this different distribution of matches. Because of simplicity, we focus on the variance under the frictionless matching m^f to adjust the matching. Our procedure is as follows. We construct m^f_{mass} to account for the distributions $\Gamma_{1960}(x)$ and $\Psi_{1960}(y)$ as follows. We first pin down the off-diagonal elements using the proportional rule $m^f_{mass}(x,y) = m^f_{1960}(x,y)\gamma_{2014}(x)/\gamma_{1960}(x)$. Then, starting at the top we recursively pin down the frictionless allocation along the diagonal. This procedure results in the following $m^f_{mass}(x,y)$:

	$m^f(x,y)$	x: 1	2	3	4	$\Psi(y)$
y	4				20.51	20.51
	3			10.59	18.11	28.70
	2	1.66	31.42	11.92		45.00
	1	5.80				5.80
	$\Gamma(x)$	7.46	31.42	22.51	38.62	100.00

Then we calculate the Variance as:

$$\begin{aligned}\text{Var}_{2014}[\omega+\sigma; \Gamma_{1960}, \tilde{\Psi}_{1960}, \mu_{2014}[x,y]] = &\left(\sum_\omega\sum_\sigma(\omega+\sigma)^2\sum_x\sum_y h_{2014}(\omega,\sigma|x,y)m^f_{mass}(x,y)\right)\\ &-\left(\sum_\omega\sum_\sigma(\omega+\sigma)\sum_x\sum_y h_{2014}(\omega,\sigma|x,y)m^f_{mass}(x,y)\right)^2.\end{aligned}$$

C.2.6 Marital Sorting, All More Educated: Var$_{2014}(\omega+\sigma; \Gamma_{2014}, \tilde{\Psi}_{2014}, \mu_{1960}[x,y][x,y])$

Now we change the equilibrium matching $\mu_{1960}(x)$ and keep the mass of education at the level of 2014. Again, we need to adjust one of the distributions to allow for the μ to change, so we change $\Psi_{2014}(y)$ again to $\tilde{\Psi}_{2014}(y)$ such that $\tilde{\Psi}_{2014}(\mu_{1960}(x)) = \Gamma_{2014}$. For simplicity, we will again adjust the frictionless matching allocation $m^f(x,y)$ and not $m(x,y)$. As before, we calculate $m^f_\mu(x,y)$ starting with the off diagonal using the proportional formula $m^f_\mu(x,y) = m^f_{2014}(x,y)\gamma_{1960}(x)/\gamma_{2014}(x)$. Then, from the top we recursively pin down the frictionless allocation. This procedure results in the following $m^f_\mu(x,y)$:

	$m^f(x,y)$	1	2	3	4	$\Psi(y)$
y	4			2.17	12.07	14.24
	3		5.81	7.80		13.61
	2	9.39	21.03			30.42
	1	41.72				41.72
	$\Gamma(x)$	51.11	26.84	9.97	12.07	100.00

(Columns indexed by x.)

$$\mathrm{Var}_{2014}[\omega+\sigma;\Gamma_{2014},\tilde{\Psi}_{2014},\mu_{1960}[x,y]] = \left(\sum_{\omega}\sum_{\sigma}(\omega+\sigma)^2\sum_{x}\sum_{y}h_{2014}(\omega,\sigma|x,y)m^f_{\mu}(x,y)\right) - \left(\sum_{\omega}\sum_{\sigma}(\omega+\sigma)\sum_{x}\sum_{y}h_{2014}(\omega,\sigma|x,y)m^f_{\mu}(x,y)\right)^2.$$

C.2.7 Marital Sorting, Assortativeness: Var$_{2014}(\omega+\sigma;\alpha_{1960}[x,y][x,y])$

We calculate the variance assuming the mismatch is as in 1960 but the marginals are from 2014. With the distance measure d we construct a distribution of matches in 2014 that inherits the amount of mismatch from 1960. To do so, we endow the elements of M that are non-zero in M^f with a the value $m(x,y)(1-d_{1960})/(1-d_{2014})$ that adjusts the distribution with the distance measure from 1960, and likewise for the masses that are zero in M^f:

$$m_{mismatch} = \begin{cases} m_{2014}(x,y)\frac{1-d_{1960}}{1-d_{2014}} & \text{if } m^f_{2014}(x,y) > 0 \\ m_{2014}(x,y)\frac{d_{1960}}{d_{2014}} & \text{if } m^f_{2014}(x,y) = 0, \end{cases}$$

$$\mathrm{Var}_{2014}[\omega+\sigma;M_{1960}[x,y]] = \left(\sum_{\omega}\sum_{\sigma}(\omega+\sigma)^2\sum_{x}\sum_{y}h_{2014}(\omega,\sigma|x,y)m_{mismatch}(x,y)\right) - \left(\sum_{\omega}\sum_{\sigma}(\omega+\sigma)\sum_{x}\sum_{y}h_{2014}(\omega,\sigma|x,y)m_{mismatch}(x,y)\right)^2.$$

C.2.8 Total Marital Sorting: Var$_{2014}(\omega+\sigma;M_{1960}[x,y])$

Here we use the sample distribution of matches M for each subgroup (x,y) in 1960.

$$\mathrm{Var}_{2014}[\omega+\sigma;M_{1960}[x,y]] = \left(\sum_{\omega}\sum_{\sigma}(\omega+\sigma)^2\sum_{x}\sum_{y}h_{2014}(\omega,\sigma|x,y)m_{1960}(x,y)\right) - \left(\sum_{\omega}\sum_{\sigma}(\omega+\sigma)\sum_{x}\sum_{y}h_{2014}(\omega,\sigma|x,y)m_{1960}(x,y)\right)^2.$$

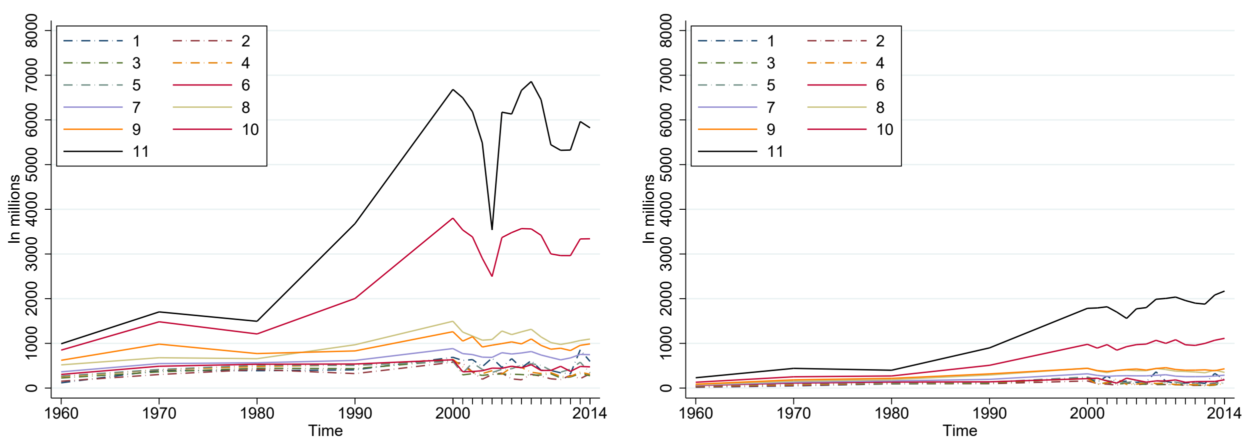
C.3 Robustness: Eleven Education Categories

In order to investigate whether our findings are robust to the choice of education categories, we repeat the analysis with 11 education categories. The categories are based on the CPS variable "EDUC" and are aimed at capturing the actual years of schooling. This categorization is often used to obtain a continuous measure of education. The categories are reported in Table 4.

1	N/A or no schooling
2	Nursery school to grade 4
3	Grade 5, 6, 7, or 8
4	Grade 9
5	Grade 10
6	Grade 11
7	Grade 12
8	1 year of college
9	2 years of college
10	3 or 4 years of college
11	5+ years of college

Table 4: 11 Education Categories

We now report the same sets of figures as in the main text. Because of the normalization, the distribution of educational attainment have a somewhat different shape, but qualitatively they capture the same relation between male and female attainment and how it has changed between 1960 and 2014.



(a) Male Earnings: $F(\omega|x)$ (b) Female Earnings: $G(\sigma|y)$

Figure 10: Variance of Earnings by Education over Time: $x, y \in \{1, ..., 11\}$.

The distance measures between 1960 and 2014 are similar also for 11 education categories: $d_{1960} = 0.59$ and $d_{2014} = 0.51$. Of course, given a finer grid of education categories, the distance measure is larger than under only 4 education categories. But the main point is that the distance measure has not changed much between 1960 and 2014. In Figure 12 we see the evolution of the matching patterns and how they compare to the frictionless outcome. The main change is due to the change in the marginal distributions $\Gamma(x)$ and $\Psi(y)$, which implies that there is more mass on the low education marriages in 1960 and more on the high education marriages in 2014. Moreover, the frictionless matching allocation has moved from above the diagonal in 1960 to on or below the diagonal in 2014, reflecting the fact that female educational attainment has increased more than that of males.

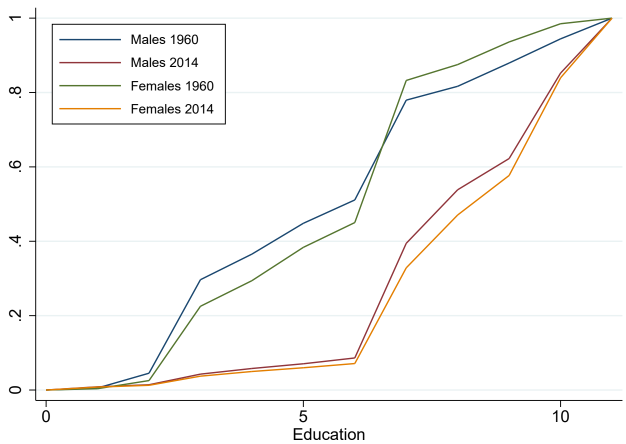


Figure 11: Distribution of Education: $\Gamma(x)$ for males and $\Psi(y)$ for females in 1960 and 2014.

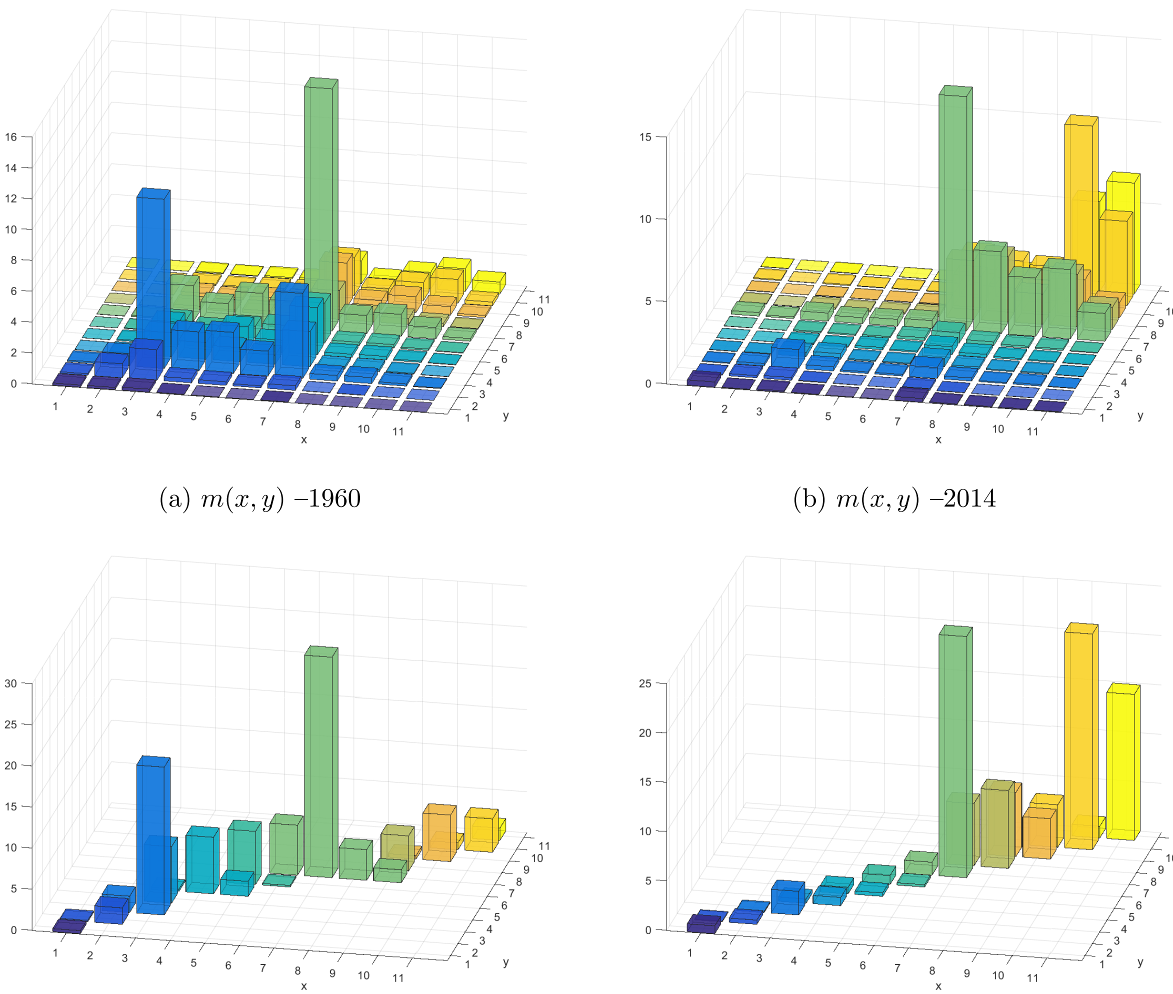
But again, even though the distributions and the matching patterns have changed, we see little evidence of a change in the marital mismatch.

Finally, we perform the same decomposition exercise on the setup with eleven education categories, reported in Table 5. It is striking how close the values are that we obtain to those in Table 1 for four education categories. In particular, the last column that summarizes those changes only differs by at most a few percentage points.

Baseline	Sample	Normal Model*		
$\text{Var}_{2014}(\omega+\sigma)$	3.85×10^9	3.27×10^9		
$\text{Var}_{1960}(\omega+\sigma)$	0.50×10^9	0.45×10^9		
$\frac{\text{Var}_{2014}(\omega+\sigma)}{\text{Var}_{1960}(\omega+\sigma)}$	7.72	7.31		
	$X_{1960}[x,y]$	$\text{Var}_{2014}(\omega+\sigma; X_{1960})$	$\frac{\text{Var}_{2014}(\omega+\sigma;X_{1960})}{\text{Var}_{1960}(\omega+\sigma)}$	X_{2014} explains
A. Stochastic Sorting (ex post)				
1. Marginals Earnings*	F_{1960}, G_{1960}	0.78×10^9	1.75	76%
2. Correlation Earnings*	ρ_{1960}	3.02×10^9	6.75	9%
Total (Normal model)*	$F_{1960}, G_{1960}, \rho_{1960}$	0.73×10^9	1.64	78%
Total	H_{1960}	0.79×10^9	1.77	77%
B. Marital Sorting (ex ante)				
1. Marginals Education	$\Gamma_{1960}, \Psi_{1960}$	1.70×10^9	3.40	56%
Allocation: Gender Equal.	$\Gamma_{2014}, \tilde{\Psi}_{2014}, \mu^f_{1960}$	3.39×10^9	6.80	12%
All more educated	$\Gamma_{1960}, \tilde{\Psi}_{1960}, \mu^f_{2014}$	2.42×10^9	4.85	37%
2. Assortativeness Education	d_{1960}	4.07×10^9	8.16	6%
Total	M_{1960}	1.69×10^9	3.39	56%

* This decomposition assumes for each education pair $[x, y]$ normality of the joint distribution of ω, σ, which implies normality of the marginal distributions F, G, as well as normality of the distribution of $\omega + \sigma$, in order to single out a constant correlation coefficient $\rho[x, y]$.

Table 5: Decomposition of Household Income Inequality. We calculate the Variance of household income in 2014 in each case holding fixed one of the determinants at 1960 levels.



(a) $m(x, y)$ –1960

(b) $m(x, y)$ –2014

(c) $m^f(x, y)$ – 1960

(d) $m^f(x, y)$ – 2014

Figure 12: Actual Distribution of Marriages by Education $m(x, y)$ and Hypothetical Distribution without ex-ante Mismatch: $m^f(x, y)$.

The main conclusion therefore continues to hold. Marital sorting only explains a minor share of the increase in the increase in household income inequality. Most is due to stochastic sorting, i.e., the increase in the variance of income for a given education category, and in particular those in the category with 3 or 4 years of college and 5+ years. In addition, the change in the distribution of education with a big first order stochastic dominance shift for both males and females implies that more people in the sample are subject to high inequality incomes. The fact that females have become relative more educated explains some of the increase in inequality.

References

Ackerberg, D., and M. Botticini (2002): "Endogenous matching and the empirical determinants of contract form," *Journal Political Economy*, 110(3), 564–591.

Amershi, A. H., and J. H. Stoeckenius (1983): "The theory of syndicates and linear sharing rules," *Econometrica: Journal of the Econometric Society*, pp. 1407–1416.

Anderson, A. (2015): "A dynamic generalization of Becker's assortative matching result," *Journal of Economic Theory*, 159, 290–310.

Anderson, A., and L. Smith (2010): "Dynamic matching and evolving reputations," *The Review of Economic Studies*, 77(1), 3–29.

Athey, S. (1998): "Characterizing Properties of Stochastic Objective Functions," MIT Working Paper.

Athey, S., L. F. Katz, A. B. Krueger, S. Levitt, and J. Poterba (2007): "What Does Performance in Graduate School Predict? Graduate Economics Education and Student Outcomes," *The American Economic Review*, 97(2), 512–518.

Bandiera, O., S. Hansen, A. Prat, and R. Sadun (2017): "CEO Behavior and Firm Performance," Discussion paper, National Bureau of Economic Research.

Becker, G. (1973): "A Theory of Marriage: Part I," *Journal Political Economy*, 81(4), 831–846.

Chade, H. (2006): "Matching with Noise and the Acceptance Curse," *Journal of Economic Theory*, 129(1), 81–113.

Chade, H., J. Eeckhout, and L. Smith (2016): "Sorting Through Search and Matching Models in Economics," *Journal of Economic Literature*.

Cheremukhin, A., P. Restrepo-Echavarria, and A. Tutino (2016): "Targeted Search in Matching Markets," mimeo.

Chiappori, P.-A., and P. Reny (2005): "Matching to Share Risk," University of Chicago Working Paper.

Choo, E., and A. Siow (2006): "Who Marries Whom and Why," *Journal of Political Economy*, 114(1), 175–201.

Conley, J. P., and A. S. Önder (2014): "The Research Productivity of New PhDs in Economics: The Surprisingly High Non-Success of the Successful," *The Journal of Economic Perspectives*, 28(3), 205–215.

Edmans, A., and X. Gabaix (2011): "The effect of risk on the CEO market," *Review of Financial Studies*, 24(8), 2822–2863.

Eeckhout, J., and P. Kircher (2011): "Identifying Sorting – In Theory," *Review of Economic Studies*, 78, 872–906.

Eika, L., M. Mogstad, and B. Zafar (2014): "Educational assortative mating and household income inequality," Discussion paper, National Bureau of Economic Research.

Gabaix, X., and A. Landier (2008): "Why has CEO Pay Increased So Much?," *The Quarterly Journal of Economics*, 123(1), 49–100.

Galichon, A., and B. Salanié (2010): "Matching with Trade-offs: Revealed Preferences over Competing Characteristics," CEPR Discussion Papers 7858, C.E.P.R. Discussion Papers.

Gautier, P. A., and C. N. Teulings (2004): "The right man for the job," *The Review of Economic Studies*, 71(2), 553–580.

Greenwood, J., N. Guner, G. Kocharkov, and C. Santos (2014): "Marry your like: Assortative mating and income inequality," *The American Economic Review*, 104(5), 348–353.

Guerrieri, V., R. Shimer, and R. Wright (2010): "Adverse selection in competitive search equilibrium," *Econometrica*, 78(6), 1823–1862.

Holmström, B., and P. Milgrom (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55(2), 303–28.

Jovanovic, B. (1979): "Job matching and the theory of turnover," *The Journal of Political Economy*, pp. 972–990.

Karlin, S., and Y. Rinott (1980): "Classes of Orderings of Measures and Related Correlation Inequalities. I. Multivariate Totally Positive Distributions," *Journal of Multivariate Analysis*, 10, 467–498.

Lam, D. (1997): "Demographic Variables and Income Inequality," in *Handbook of Population and Family Economics*. Elsevier Science.

Lamadon, T., J. Lise, C. Meghir, and J.-M. Robin (2013): "Matching, sorting and wages," Discussion paper, Mimeo.

Legros, P., and A. Newman (2007): "Beauty is a Beast, Frog is a Prince: Assortative Matching with Nontransferabilities," *Econometrica*, 75, 1073–1102.

Li, S., H. Sun, T. Wang, and J. Yu (2016): "Assortative matching and risk sharing," *Journal of Economic Theory*, 163, 248–275.

LINDENLAUB, I. (2013): "Sorting of Multidimensional Types," UPF mimeo.

MEGHIR, C., AND L. PISTAFERRI (2004): "Income variance dynamics and heterogeneity," *Econometrica*, 72(1), 1–32.

NELSON, P. (1970): "Information and consumer behavior," *Journal of political economy*, 78(2), 311–329.

PILOSSOPH, L. (2016): "Assortative Matching and Income Inequality: A Structural Approach," Discussion paper, mimeo.

SCHULHOFER-WOHL, S. (2006): "Negative assortative matching of risk-averse agents with transferable expected utility," *Economics Letters*, 92, 383–388.

SERFES, K. (2006): "Risk sharing vs. incentives: contract design under two-sided heterogeneity," *Economics Letters*, 83, 343–349.

SHIMER, R., AND L. SMITH (2000): "Assortative Matching and Search," *Econometrica*, 68(2), 343–369.

TERVIÖ, M. (2008): "The difference that CEOs make: An assignment model approach," *The American Economic Review*, pp. 642–668.

WILSON, R. (1968): "The theory of syndicates," *Econometrica: journal of the Econometric Society*, pp. 119–132.