Endogenous Risk Taking and Physical Appearance of Sex Workers*

Peter H. Egger and Andreas Lindenblatt

October 23, 2014

Abstract

Previous research found that physical appearance affects the risk-taking of sex workers through offering unprotected services. This paper utilizes a large individual-level data-set covering 16,583 pay-for-sex contracts in 2011 and 2012 by 2,517 female suppliers in Germany. Results based on instrumental variables suggest that the incentive for risk-taking is about twice as high than when assuming random assignment of risk-taking.

Keywords: Sexual workers; Risk taking; Health economics.

JEL-classification: I10; J01; J70; J71.

1 Introduction

While some countries ban the supply (e.g., Ireland, United States) or the demand (e.g., Finland, France, Ireland, Norway, Sweden, United States) of sexual services for pay, such contracts are legal elsewhere (e.g., Austria, Belgium, Brasil, Germany, Hungary, Netherlands, Switzerland, United Kingdom; see Cho, Dreher and Neumayer, 2013). Independent of legality, there is abundant evidence of existing sex work – mainly supplied by female workers to male customers – all over the world. Over the last decade supply has shifted towards the internet. Street prostitution and brothels still exist and make up a large part of the market, but with the anonymity of the internet, platforms mediating between providers and customers of sexual services offer a cheap way to advertise sexual services and have become more and more

**Egger*: ETH Zurich (KOF Swiss Economic Institute), CEPR, CESifo, GEP, and WIFO; Email: egger@kof.ethz.ch; Address: KOF Swiss Economic Institute, WEH D4, Weinbergstrasse 35, 8092 Zurich, Switzerland. Egger acknowledges financial support from GA ČR through grant no. P402/12/0982.

Lindenblatt: University of Heidelberg; Email: lindenblatt@uni-hd.de; Address: University of Heidelberg, Department of Economics, Bergheimerstrasse 58, 69115 Heidelberg, Germany.

important (Cunningham and Kendall, 2011). It is hard to get reliable data for this market (for the "online" and, especially, the "offline" segment), but estimates for the whole market speak of up to 400,000 women working full or part-time in the sector, earning per year up to 14.5 billion Euro or 0.4% of the GDP in Germany alone (Reichel and Topper, 2003).

A key issue in professional sex work is health protection for all involved parties. Recent research provides evidence that a significant fraction of sexual services are provided without protection (e.g., see Rao, Gupta, Lokshin, and Jana, 2003; Gertler, Shah, and Bertozzi, 2005; Cunningham and Kendall, 2010; de la Torre, Havenner, Adams, and Ng, 2010; Robinson and Yeh, 2011; Arunachalam and Shah, 2013), thereby encountering a risk of genital and other diseases such as HIV.[1] Most of the literature acknowledges (directly or indirectly) that risk-taking through the offering of unprotected sex is endogenous. Sex workers had been found to be willing to offer sex without condom more likely, if a price premium was paid (see Gertler, Shah, and Bertozzi, 2005; de la Torre, Havenner, Adams, and Ng, 2010) or the opportunity costs were high. E.g., de la Torre, Havenner, Adams, and Ng (2010), Robinson and Yeh (2011), and Wilson (2012) showed that sex workers offered such services more frequently after spells of bad financial shape, negative income shocks, or illness (and an associated loss of income). This evidence is consistent with the larger prevalence of unprotected sex offerings in the "offline" (e.g., street prostitution) relative to the "online" (internet) market, since suppliers in the former segment of the market tend to be less educated, to charge lower prices, and to depend more on the income generated from their services than ones in the latter segment (see Cunningham and Kendall, 2010).

The average price premia reported in the aforementioned work range from about 9% for Kenyan sex workers in Robinson and Yeh (2011) and Ecuadorian sex workers in Arunachalam and Shah (2012) over 23% in Gertler, Shah, and Bertozzi (2005) to 30% in de la Torre, Havenner, Adams, and Ng (2010) for Mexican sex workers to 194-376% for Indian sex workers in Rao, Gupta, Lokshin, and Jana (2003). It is worth mentioning that, among those studies, the one by Rao, Gupta, Lokshin, and Jana (2003) was the only one to be able to randomize about condom usage, and it found the largest price premium among all of them.

Some recent research on the size and the determinants of the price premium for unprotected sex argues that the opportunity costs of denying such services (i.e., the economic incentives of supplying them) are larger for less attractive sex workers. Gertler, Shah, and Bertozzi (2005) provide a theoretical rationale for this line of work. Their model suggests that attractiveness raises the bargaining power of sex

[1] Throughout this paper, we use the terms "risk-taking", "unprotected sex", and "sex without condom" interchangeably.

workers, whereby attractive sex workers are able to charge a higher price for sex services in general and also a higher premium for unprotected sex than less attractive ones.[2] This shows in higher price premia for unprotected sex offerings of more attractive sex workers (see Gertler, Shah, and Bertozzi, 2005) and in lower price premia for obese sex workers (see Chang and Weng, 2012). However, the associated evidence assumes that unprotected sex practice is random.[3]

This paper utilizes data on the offering and contracting of sexual services via the internet through *www.gesext.de*.[4] We downloaded all contracted services of female suppliers between January 16 of 2011 and September 9 of 2012 from this database. In this study, we focus on all contracts where the weight and offered services of the worker are known and a contract has been concluded at a known price.[5] We estimate the price premia in contracts about unprotected sex, emphasizing the importance of controlling for endogeneity. We propose an identification strategy which relies on two alternative types of instruments – one related to the frequency of prior unprotected sex offerings by other sex workers in the same region where a given contract was made and one related to personal characteristics of other sex workers offering prior unprotected sex in the same region where a given contract was made. In line with earlier work, we find that offering unprotected sexual services raises hourly wages of sex workers. Consistent with the findings of Rao, Gupta, Lokshin, and Jana (2003), we find that non-random selection (of sex workers and clients) into unprotected sex leads to a downward bias of the risk premium. Considering self-selection into risk-taking suggests that the risk premium on non-protection is more than twice as large as when assuming random risk-taking. We estimate the risk premium at about 91% of the average hourly wage. However, it is interesting to see that the magnitude of the premium is smaller than in Rao, Gupta, Lokshin, and Jana (2003), since we analyze a sample of sex workers from a developed country and in the "online" segment, where at least the purely economic opportunity costs should be relatively

[2]To some extent, the arguments in this line of research are consistent with the evidence that economic success and physical appearance are related to each other (see Hamermesh and Biddle, 1994; Fletcher, 2009; Mobius and Rosenblatt, 2006; Johnston, 2010).

[3]Gertler, Shah, and Bertozzi (2005) demonstrate in their Table 5 that the practice of unprotected sex can be explained by sex worker and client characteristics, but their price comparisons of protected and unprotected sex practices only account for endogenous selection based on time-invariant sex worker characteristics. Chang and Weng (2012, p. 482) acknowledge that their "analysis may suffer from endogeneity bias because condom use and prostitute price may be correlated due to some unobserved common factors."

[4]While this is only *one* such platform in Germany, it accounts for an annual revenue of about 4 million Euros for sexual services offered by females only.

[5]When rescinding a concluded contract, the contracted price is still due. This can be enforced since customers must fully reveal their identity to the platform owner. 15% of the contracted price is generally due as a fee for the platform services.

high. Apart from the mentioned ones, reasons for the quantitative difference between the results in this paper and the one of Rao, Gupta, Lokshin, and Jana (2003) may be that medication to treat HIV is more easily available and affordable in Germany than in India and that the time periods are rather different between the two studies (pertaining to 1993 in Rao, Gupta, Lokshin, and Jana, 2003, and to 2011-2012 in this study). Interestingly and in contrast to the results in Gertler, Shah, and Bertozzi (2005) and Chang and Weng (2012), the results in this paper suggest that appearance affects the hourly wage but not the risk premium on unprotected sex.

The remainder of the paper is organized as follows. The next section discusses the novel database on online sex contracts used in this paper. Section 3 describes the identification strategy and the econometric framework. Section 4 summarizes the estimation results, and the last section concludes.

2 Data

Before introducing the data-set employed in detail, we should emphasize two features. First, as any other study on sex workers, data on this type of work are generally selected in the sense that census-type data on sex workers are not available. Data will not be available, unless sex workers are willing to participate in a field study (as in Rao, Gupta, Lokshin, and Jana, 2003; Gertler, Shah, and Bertozzi, 2005; de la Torre, Havenner, Adams, and Ng, 2010; Robinson and Yeh, 2011; Chang and Weng, 2012; Arunachalam and Shah, 2013) or they come forward with an online posting and deliver information without knowing (as in this study and also in Moffat and Peters, 2004; Cunningham and Kendall, 2011). Second, as opposed to "offline" sex workers, "online" sex workers as used in this study tend to be better educated, financially less dependent, and part of the higher-price segment of the market (see Cunningham and Kendall, 2010).

We employ digitally collected data on the offering and contracting of sexual services by females via the internet through *www.gesext.de*. When making an offer on this platform sellers describe themselves and their services (including e.g. notes on condom use, duration of the meeting, the location, ...) and can offer sexual services either as a second price auction or at a fixed price. While in the first case the highest bidder has to pay the second highest bid, in the latter case the first bidder has to pay the fixed price. The meeting then takes place at a predefined place (hotel, flat, ...) and the payment is normally made in cash. It is unknown if the seller and buyer actually meet and engage in sexual services, but when rescinding a concluded contract, for the customer the full contracted price is due. For the seller 15% of the contracted price is generally due as a fee for the platform services. This can be enforced since customers must fully reveal their identity to the platform owner.

In general everybody can offer services on this platform, but the average price of about 148 Euro shows that this platform mainly covers a premium segment of the market. We downloaded all contracted services of female suppliers between January 16 of 2011 and September 9 of 2012 from this database and focus on all offers where a contract has been concluded at a known price.

Table 1: Descriptive statistics

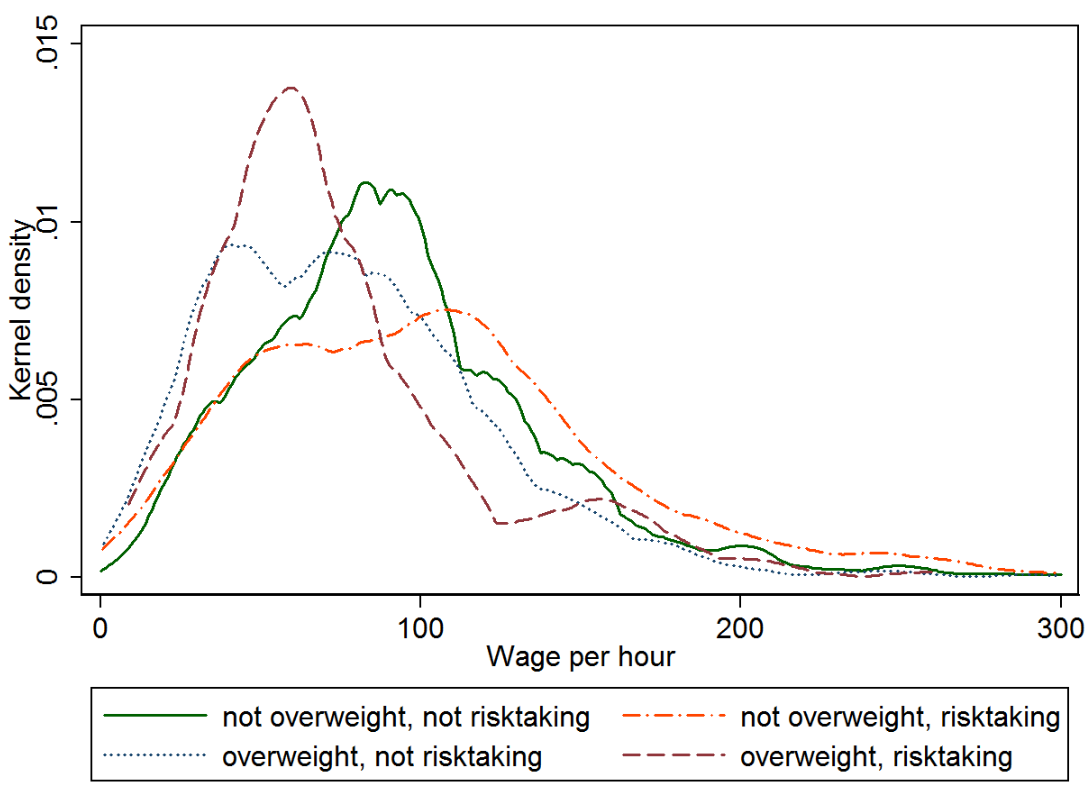
Variable	**Mean**	**Std.dev.**	**Binary**
Log price per hour	4.9970	0.7698	no
Risk-taking	0.0721	0.2587	yes
Overweight	0.2373	0.4254	yes
Height is provided	0.5815	0.4933	yes
Height in cm if height is provided	166.6254	6.0342	no
Height-squared in cm if height is provided	27,800.4400	1,994.4560	no
Age is provided	0.4098	0.4918	yes
Age in years if age is provided	29.7913	8.6382	no
Age-squared in years if age is provided	962.1291	583.7763	no
First-time supplier	0.0522	0.2224	yes
Single	0.0828	0.2756	yes

Notes: The total number of observations (transactions) is 16,583 and the underlying number of sex workers is 2,517.

Table 1 and Figure 1 provide descriptive evidence of the hourly wage of 2,517 sex workers offering sex for pay at *www.gesext.de* between January 17 of 2012 and September 9 of 2012 through altogether 16,583 individual contracts. We classify workers with a body mass index of more than 25 as overweight. Some of the workers do not provide weight and height but only pictures. For those, we classify them as being overweight or not, depending on their physical appearance (this was done by six students between March 2012 and September 2012).[6] For all of the transactions we know whether workers offer unprotected services (sex without a condom) or not. Of the 16,583 contracts, 3,935 involve overweight sexual workers and 12,648 do

[6] Of all 16,583 individual contracts covered, weight was imputed by students into the four categories "obese" (corresponding to a body-mass index, BMI, of higher than 30) an estimate BMI, "overweight" (corresponding to a BMI of higher than 25 up to 30), "normal weight" (corresponding to a BMI of higher than 18.5 and up to 25), "underweight" (corresponding to a BMI of up to 18.5) for 11,630 transactions. In order to avoid a measurement error to the largest possible extent, this categorization was done twice, by different students. Notice that subjective classification schemes of sex workers' appearances had also been used, e.g., by Moffat and Peters (2004), Gertler, Shah, and Bertozzi (2005), Cunningham and Kendall (2011).

Figure 1: Kernel densities for hourly wages of overweight/non-overweight by risk-taking/non-risk-taking female sex workers in Germany.



not, and 1,196 involve unprotected service offerings while 15,387 do not. Interestingly, about 5.5% of the contracts offered by overweight workers involve unprotected service, while 7.7% are unprotected for non-overweight workers. Table 1 provides further information on characteristics of sex workers and on the contract environment in the database. For instance, about 58% of the concluded contracts involve workers who provide information about their height and, on average, those are about 167 centimeters tall. In about 41% of the contracts sex workers mention their age and, on average, those are about 30 years old. About 5% of the contracts involve workers who indicate that they are unexperienced (they are effectively first-time suppliers on that platform), and in 8% of the contracts workers indicate that they are single. The average hourly wage across all contracts is about 148 Euros (about 5.00 in logs) with a standard deviation of about 2.16 Euros (0.77 in logs).[7]

Towards accounting for the endogeneity of risk-taking (contracting unprotected sexual services without condom), we propose two types of instruments both of which measure features of contracts that had been concluded by sex workers other than the one involved in and happening prior to transaction i but in the same two-digit district: one type of instrument is based on the prior average risk-taking of other sex workers, and one is based on the average height of other sex workers. For either type of instruments, we use the i-specific value as well as the squared value to form two identifying instrumental variables. The instruments are likely exogenous as long there is some inter-temporal habit formation, i.e., if sex workers adopt strategies of others in the same neighborhood in the past. The second type of instrument (based on others' height) is relevant and adequate when two things hold: first there is habit formation over time and height matters for risk-taking.

Figure 1 displays the distributions of log hourly wages for the four cells in the matrix overweight/non-overweight by risk-taking/non-risk-taking. The figure illustrates that the average hourly wages of non-overweight risk-taking suppliers is higher than that of non-risk-taking suppliers. Especially, the right tail of the former distribution is fatter than that of the latter distribution. These features are somewhat less clear-cut for overweight suppliers who clearly make up a smaller part of the overall distribution. However, we should bear in mind that the distributions are unconditional on explanatory variables and do not provide for direct inference about the relative importance of risk-taking on hourly wages of female sex workers.

[7]Clearly, all of the information contained in the online postings is self-reported. Hence, sex workers might strategically misreport (see Plankey, Stevens, Flegal, and Rust, 1997; and Cawley, 2014; for a general discussion of misreporting). However, the contracts are concluded based on this information provided and they become legally effective. Moreover, most sex workers are rated by earlier customers, and misreporting would show in the ratings. The ratings do not indicate strategic misreporting on average.

3 Estimation methods

One particular goal of this section is to shed light on the differences between regressions which treat $Risk - taking_i$ as exogenous – i.e., assuming that sex workers do not self-select into unprotected sex – and ones which consider $Risk - taking_i$ to be endogenous upon self-selection. For the latter, we will generally pursue an instrumental-variable (IV) two-stage least-squares (2SLS) strategy which takes the binary nature of $Risk - taking_i$. Procedures for such models had been suggested, among others, by Heckman (1978), Maddala (1983), Heckman and Vytlacil (1999), Vella and Verbeek (1999), and Wooldridge (2002). Specifically, Wooldridge (2002, pp. 621-633) discusses the assumptions and suggests four alternative parametric procedures based on those (on pages 623, 626, 629f., and 631, respectively). We will rely on such approaches, where the average treatment effect – in our case, of $Risk - taking_i$ – may vary with some covariates – here, $Overweight_i$.

For the approaches considered here, we will maintain the following set of assumptions. (*Instrument exogeneity and Conditional mean independence.*) First, we will assume that at least one instrument $Instruments_i$ exists which can be used to model the probability that $Risk - taking_i = 1$,
$P(Risk - taking_i = 1) = \Phi(Overweight_i, Controls_i, Instruments_i) = \Phi_i$, where the instrument is uncorrelated with the unobservable determinants of – i.e., the stochastic term on – the dependent variable of interest, e.g., $\ln(HourlyWage)_i$. Denoting this stochastic term by ε_i, the formal requirement is $E(Instruments_i \varepsilon_i) = 0$ and assuming that Φ_i can be estimated, the prediction $\hat{\Phi}_i$ is a natural instrument for $Risk - taking_i$ (see Wooldridge, 2002, p. 626). Conditional on $Overweight_i$, a set of control variables entering the outcome equation, $Controls_i$, which includes a constant, and the set of instruments, $Instruments_i$, $Risk - taking_i$ is then random. Formally, when using superscript 1 for $Risk - taking_i = 1$ and superscript 0 for $Risk - taking_i = 0$ (only one of which is observed for transaction i), the expectations $E(\varepsilon_i^s | Overweight_i, Controls_i, Instruments_i) = E(\varepsilon_i^s | Overweight_i, Controls_i)$ for $s \in \{0, 1\}$. (*Joint normality of the disturbances on $Risk - taking_i$ and outcome such as $\ln(HourlyWage)_i$.*) Under normality of the latent variable determining the desirability of $Risk - taking_i = 1$, the function Φ_i is the cumulative normal distribution function and $\hat{\Phi}_i$ is the prediction that $P(Risk - taking_i = 1)$. Under bivariate normality of the random component of $\ln(HourlyWage)_i$ and the random component in Φ_i, say, ν_i, the prediction of a probit model of $P(Risk - taking_i = 1)$ can be used as an instrument for $Risk - taking_i = 1$ (see Wooldridge, 2002, p. 623, Procedure 18.1). Alternatively, one can use a control function using the probability density function, $\phi(Overweight_i, Controls_i, Instruments_i) = \phi_i$, based on the latent process for $P(Risk - taking_i = 1)$ in addition to instrumenting (see Wooldridge,

2002, p. 629f, Procedure 18.3).

The same assumptions support models where the effect of $Risk - taking_i$ on outcome, e.g., $\ln(HourlyWage)_i$, varies with some or all observable elements in $(Overweight_i, Controls_i)$. In light of earlier work on the matter (see Chang and Weng, 2012), we will consider cases where the average treatment effect of $Risk - taking_i$ varies with $Overweight_i$ only. Suppose that the effect of $Risk-taking_i$ of interest on outcome is $\beta Risk-taking_i+\gamma Risk-taking_i\times Overweight_i$ (when omitting other effects for brevity). In that case, the average treatment effect of $Risk-taking_i$ is $\beta Risk - taking_i + \gamma\overline{Overweight}$, where $\overline{Overweight} = E(Overweight_i)$ can be estimated by the sample mean. Wooldridge suggests demeaning $Overweight_i$ in the interaction term and using $\widetilde{Overweight}_i \equiv Overweight_i - \overline{Overweight}$ in $\beta Risk - taking_i+\gamma Risk-taking_i\times\widetilde{Overweight}_i$, so that β measures the average treatment effect of $Risk-taking_i$, since $E(\widetilde{Overweight}_i) = 0$. Then $\hat{\Phi}(\cdot)$ and $\hat{\Phi}(\cdot)\times\widetilde{Overweight}_i$ are the suitable instruments for $Risk - taking_i$ and $Risk - taking_i \times \widetilde{Overweight}_i$, respectively (see Wooldridge, pp. 626 and 629).

Consequently, we run regressions per transaction $i = 1; ...; 16,583$ of the form

$$
\begin{aligned}
\ln(HourlyWage)_i &= \beta Risk - taking_i + \gamma Risk - taking_i \times \widetilde{Overweight}_i \\
&+ \delta Overweight_i + Controls_i\zeta + (\eta\hat{\phi}_i) + \varepsilon_i. \qquad (1) \\
\overline{Overweight} &\equiv \frac{1}{16,583}\sum_{i=1}^{16,583} Overweight_i, \\
\widetilde{Overweight}_i &\equiv Overweight_i - \overline{Overweight}. \qquad (2)
\end{aligned}
$$

Notice that the interaction term $Risk-taking_i\times\widetilde{Overweight}_i$ involves the demeaned value while the main effect involves the un-demeaned value of $Overweight_i$ as in Wooldridge (2002, pp. 626 and 629). The term $\eta\hat{\phi}_i$ is the control function based on the estimated density $\hat{\phi}_i$, which will not be included in all models. Models that exclude $\eta\hat{\phi}_i$ correspond to what Wooldridge (2002, p. 626) calls *Procedure 18.2*, whereas ones that include $\eta\hat{\phi}_i$ correspond to what Wooldridge (2002, p. 629) calls *Procedure 18.3*.

We run three pairs of versions of (1), three of them excluding $\gamma Risk - taking_i \times \widetilde{Overweight}_i$ and the other ones including it. Two models assume random assignment of sex workers into risk-taking whereby β could be estimated through ordinary least squares (OLS) on (1). Four models assume self-selection into risk-taking and aim at avoiding an associated bias by using instrumental variables in two-stage least squares (2SLS) regressions as described above. For the latter, we estimate a probability model of the form

$$P(Risk-taking_i=1)=\Phi([Instruments_i,Overweight_i,Controls_i]\theta+\nu_i), \quad (3)$$

where we use two variants of $Instruments_i$ as described in Section 2. In the main part of the paper, we employ $Instruments_i = (RegionHistory_i, RegionHistory_i^2)$, where $RegionHistory_i$ is the average value (probability) of risk-taking by other sex workers in the same two-digit zip code between January 16 of 2012 and September 8 of 2012 and at least one day prior to contract i and $RegionHistory_i^2$ is the squared value thereof. The vector θ is a conformable parameter vector on all the explanatory variables in the probit model (including fixed effects for 78 two-digit zip codes). $RegionHistory_i$ and $RegionHistory_i^2$ are valid instruments, if other sex workers than the one offering transaction i in the same district, where i is contracted, do not anticipate or are influenced by $Risk-taking_i$. Alternatively, we use $Instruments_i = (OthersHeight_i, OthersHeight_i^2)$, where $OthersHeight_i$ is the average value reported height by other sex workers in the same two-digit zip code between January 16 of 2012 and September 8 of 2012 and at least one day prior to contract i and $OthersHeight_i^2$ is the squared value thereof. Quite clearly, $OthersHeight_i$ and $OthersHeight_i^2$ cannot be influenced by $Risk-taking_i$ and, as long as there is some contagion in $Risk-taking_i$, we would expect characteristics of other sex workers – in particular, in a lagged fashion – to be suitable instruments for $Risk-taking_i$ (see Kuersteiner and Prucha, 2013; and Badinger and Egger, 2014).

4 Regression results

4.1 Main results

The results for the probit and six linear regression models are summarized in Table 2. The probit results indicate that overweight suppliers tend to offer unprotected services less likely than others. This is in some contrast with earlier work which pointed to a compensating role of risk-taking for personal appearance. However, the result is consistent with the frequencies of contracts in the 2×2 overweight/non-overweight and the risk-taking/non-risk-taking matrix as reported above. Moreover, there is some evidence of younger providers to be more likely to offer unprotected services. Finally, contagion is a strong factor in risk-taking: the more others offered unprotected services in the same region (at least one day) prior to the time contract i was concluded, the more likely will the worker involved in contract i offer unprotected services as well. The explanatory variables have a highly significant joint impact in the probit model and also the other models. Notice that all estimates reported

Table 2: Regression results

	Probit	OLS	OLS	2SLS	2SLS	2SLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Risk-taking	-	0.5063***	0.4958***	1.2277***	1.3478***	1.0673***	1.1666***
	-	0.0214	0.0215	0.1181	0.1627	0.1657	0.2202
Risk-taking×$\widetilde{\text{Overweight}}$	-	-	-0.2079***	-	0.4633	-	0.3359
	-	-	0.0537	-	0.3348	-	0.3451
Overweight	-0.1003**	-0.2716***	-0.2588***	-0.2609***	-0.2878***	-0.2579***	-0.2777***
	0.0411	0.0130	0.0134	0.0136	0.0235	0.0137	0.0247
Height is provided	15.1032***	6.4907***	6.4427***	3.2645	2.9386	3.5238	3.2677
	4.7512	2.0643	2.0634	2.1965	2.2504	2.1736	2.2257
Height in cm if height is provided	-0.1696***	-0.0842***	-0.0837***	-0.0465*	-0.0426	-0.0499*	-0.0468*
	0.0583	0.0250	0.0250	0.0265	0.0272	0.0263	0.0269
Height-squared in cm if height is provided	0.0005***	0.0003***	0.0003***	0.0002**	0.0002*	0.0002**	0.0002**
	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Age is provided	1.1090***	0.5507***	0.5497***	0.4546***	0.4439***	0.4200***	0.4149***
	0.3697	0.1216	0.1216	0.1267	0.1288	0.1284	0.1295
Age in years if age is provided	-0.0478*	-0.0203***	-0.0202***	-0.0163**	-0.0159**	-0.0147*	-0.0146*
	0.0237	0.0076	0.0076	0.0079	0.0080	0.0079	0.0080
Age-squared in years if age is provided	0.0004	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000
	0.0004	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
First-time supplier	-0.0094	0.3758***	0.3766***	0.3794***	0.3779***	0.3785***	0.3774***
	0.0707	0.0245	0.0245	0.0253	0.0257	0.0250	0.0252
Single	0.2391***	0.3215***	0.3185***	0.2956***	0.2987***	0.2894***	0.2921***
	0.0613	0.0214	0.0214	0.0226	0.0229	0.0229	0.0233
History of risk-taking in same 2-digit zipcode	5.9422***	-	-	-	-	-	-
	0.9099	-	-	-	-	-	-
History of risk-taking in same 2-digit zipcode squared	-1.4593	-	-	-	-	-	-
	1.7785	-	-	-	-	-	-
Control function	-	-	-	-	-	0.3125	0.2886
	-	-	-	-	-	0.2503	0.2588
Observations	16,583	16,583	16,583	16,583	16,583	16,583	16,583
Model degrees of freedom	88	87	88	87	88	88	89
F test on joint model significance (LR test for probit)	1,358.03	55.62	55.21	47.25	47.94	47.94	46.58
First stage							
Model degrees of freedom	-	-	-	87	88	88	89
Joint F test on $\hat{\Phi}$ and $\hat{\Phi} \times \widetilde{\text{Overweight}}$							
Equation for Risk-taking	-	-	-	598.78	315.46	296.87	163.10
Equation for Risk-taking×$\widetilde{\text{Overweight}}$	-	-	-	-	472.09	-	420.94

Notes: Figures below coefficients are standard errors. ***, ** and * are significant on the 1%, 5% and 10% level, respectively pertain to regression coefficients (parameters) rather than marginal effects. Hence, the coefficients on continuous variables measure marginal effects on the *latent* desirability of or net benefit from risk-taking, a variable which has full support in real space between minus and plus infinity.

The linear regression models all suggest that risk-taking raises the hourly wage of sex workers while being overweight reduces it. There is a premium on providing information about height or age, and there is a tendency of customers to favor medium-tall and younger, especially inexperienced (first-time) suppliers. This paper's focus is on the magnitude of the impact of $Risk{-}taking_i$ and $Risk{-}taking_i \times \widetilde{Overweight}_i$ as captured by the parameter estimates $\hat{\beta}$ and $\hat{\gamma}$ in the OLS and 2SLS models.

First of all, it stands out that $\hat{\beta}$ in the OLS models is less than half of the size of the counterpart 2SLS models. An OLS parameter of 0.4958 corresponds to a semi-elasticity of about $100[\exp(0.4958) - 1] \simeq 64\%$ for an hourly wage premium of risk-taking. According to the same model, there is a semi-elasticity of being overweight of about $100[\exp(-0.2588) - 1] \simeq -23\%$ which is *raised* (rather than *reduced*) to -37% when offering unprotected services.[8]

Second, clearly the 2SLS regressions suggest that the parameters of the OLS models are biased. The instruments work well. The F-statistics for the relevance of the single identifying instrument $\hat{\Phi}_i$ for $Risk{-}taking_i$ in the models excluding $Risk{-}taking_i \times \widetilde{Overweight}_i$ are about 599 and 297 in the first stage, and the ones for the two instruments $\hat{\Phi}_i$ and $\hat{\Phi}_i \times \widetilde{Overweight}_i$ in the other first-stage 2SLS models are between 163 and 472. Neither one of the estimated 2SLS models suggests that risk-taking has a different effect on hourly wages of overweight versus other suppliers (i.e., statistically, $\hat{\gamma}$ is not distinguishable from zero). This renders the models without the interaction term preferable for efficiency reasons. The model including the control function $\eta\hat{\phi}_i$ seems preferable from an econometric point of view, as it eliminates an endogeneity bias under less stringent assumptions than the 2SLS model without the control function. In the model with the control function, the premium on risk-taking is about $100[\exp(1.0673) - 1] \simeq 91\%$. Obviously this more than compensates being overweight for the average female sex worker in the data, which reduces the hourly wage according to the same specification by $100[\exp(-0.2579) - 1] \simeq -23\%$.

[8] The correlation of the disturbances between latent process of the probability of risk-taking and the outcome (wage) equation is negative. Hence, the downward bias of the OLS parameter on risk taking suggests that risk-taking sex workers have on average lower gains from risk-taking than non-risk-taking ones.

4.2 Sensitivity analysis

In this subsection, we assess the sensitivity of the 2SLS results along four lines. First, we assess the possible impact of using weight information that had been imputed by students. For this, we exclude all transactions, where the weight information was imputed or the self-reported body-mass index (BMI) of sex workers was in the interval $24 < BMI_i < 26$ (referred to as model 7A). Alternatively, we interact the two risk measures with a binary indicator variable which is unity if the weight information was self-reported and zero if it was imputed by students based on the provided photographs. Then, we include four risk measures – the original ones and the ones that were interacted with the aforementioned binary indicator variable (referred to as model 7B). Furthermore we run regressions, where $Instruments_i = (OthersHeight_i, OthersHeight_i^2)$ rather than $Instruments_i = (RegionHistory_i, RegionHistory_i^2)$ are used (referred to as model 7C). Finally, we present results of a 2SLS model with sex-worker-specific fixed effects (referred to as model 7D). The parameters corresponding to these regressions are reported in the aforementioned order as columns (7A)-(7D) in Table 3, and they should be compared to column (7) in Table 2.

The results from those sensitivity checks may be summarized as follows. First of all, the results for the extreme-BMI subsample in column (7A) of Table 3 suggest that the main effect and interaction effect are statistically insignificant. However, these results are based on only 3,405 rather than 16,583 observations and only half of the regions (2-digit zip codes) included in the full sample are still represented in the corresponding subsample. Secondly, the results in column (7B) suggest that both in the full sample and in the subsample of contracts with self-reported weight only there is a statistically significant and positive main effect and a statistically insignificant interaction effect. The interaction term of risk and the self-reporting indicator variable enters negatively. Hence, in the self reported sample the risk premium is smaller than in the full sample, but it is still positive. Thirdly, with the alternative instrument set in column (7C), the results are similar to the ones in Table 2. Finally, with sex-worker-specific fixed effects, none of the risk-taking effects – neither the main effect nor the interaction term – is statistically significant. The latter flows from the low degree of variation of risk-taking within sex workers over time. However, as reported in the context of the discussion of Table 2, the Hausman test statistic does not support the fixed effects estimator relative to the random effects model, since estimation of the fixed effects involves an enormous loss of degrees of freedom, while the parameter vectors between the two model types are not statistically significantly different.

Notice that one could estimate the two-stage least-squares models alternatively

Table 3: Sensitvity analysis

	Self reported without BMI 24-26	**Self reported**	**Full sample Height, Height^2 as IV**	**Full sample Seller fixed effects**
	(7A)	**(7B)**	**(7C)**	**(7D)**
	2SLS	2SLS	2SLS	2SLS
Risk-taking	0.1580	1.1925***	1.2371***	-1.6496
	0.2419	0.2184	0.2801	1.0212
Risk-taking×$\widetilde{\text{Overweight}}$	-0.6247	0.2397	0.4250	-0.0355
	0.3893	0.3434	0.3218	1.0587
Risk-taking × Self reported BMI		-0.5946**		
		0.2028		
Risk-taking×$\widetilde{\text{Overweight}}$ × Self reported BMI		-0.6115		
		0.6053		
Overweight	-0.2909***	-0.2611***	-0.2459***	0.0696
	0.0376	0.0279	0.0241	0.0601
Height is provided	-	2.6649	-0.4284	1.3415
		2.2239	2.4237	7.1665
Height in cm if height is provided	0.1860***	-0.0387	-0.0060	-0.0141
	0.0637	0.0269	0.0291	0.0862
Height-squared in cm if height is provided	-0.0005***	0.0001*	0.0001	0.0000
	0.0002	0.0001	0.0001	0.0003
Age is provided	0.7876***	0.4561***	0.0409	0.0365
	0.2467	0.1300	0.1475	0.1445
Age in years if age is provided	-0.0347***	-0.0168**	0.0020	0.0008
	0.0153	0.0080	0.0086	0.0088
Age-squared in years if age is provided	0.0002	0.0001	-0.0001	-0.0001
	0.0002	0.0001	0.0001	0.0001
First-time supplier	0.3859***	0.3820***	0.3772***	0.0890
	0.0442	0.0251	0.0256	0.0567
Single	0.3001***	0.2899***	0.2119***	0.0044
	0.0531	0.0232	0.0275	0.0338
Control function	0.0772	0.4734*	2.3352***	0.5668***
	0.3967	0.2699	0.4584	0.2162
Observations	3405	16583	16583	16583
Model degrees of freedom	54	91	89	88
F test on joint model significance (LR test for probit)	41.70	46.3	45.54	11.40
First stage				
Model degrees of freedom	54	91	89	88
F test risk	45.34	163.10	79.85	11.22
F test interaction	190.81	420.94	389.02	44.26
F test risk self reported		19.58		
F test interaction self reported		9.12		

Notes: Figures below coefficients are standard errors. ***, ** and * are significant on the 1%, 5% and 10% level, respectively by using a first-stage linear-probability model. Doing so results in a p-value for the Sargan over-identification test of 0.1358 (hence, instrument validity is not rejected in spite of a high instrument relevance as reflected in an F-statistic on the joint relevance of 46.43) and the following parameters (standard errors) of interest: 1.0573 (0.1669) for Risk-taking; 0.8702 (0.3096) for Risk-taking×$\widetilde{\text{Overweight}}$, and -0.3173 (0.0227) for Overweight.[9] However, we should be very careful with the interpretation of these results, since the standard errors and test statistics are all biased due to the limited-dependent-variable character of Risk-taking and the involved corner solutions (see Wooldridge, 2002, p. 637).

5 Conclusions

This paper analyses the price premium on unprotected sex offerings by sex workers in Germany. The data used for the study come from a large online database, where the authors downloaded the universe of transactions between January 16 of 2011 and September 9 of 2012. With an average hourly wage of about 140 Euros, the data at hand feature in the premium segment of the market. The issue of self-selection into risky behavior through unprotected sex received particular attention in the study. Using time-lagged characteristics of sex transactions by other sex workers in the geographical neighborhood, we found that disregarding endogenous risk-taking leads to down-ward biased risk-premia. The premium on unprotected services in the data is estimated at about 91% of the average hourly wage. There is no evidence in the data that over-weight sex workers receive a lower positive premium on unprotected services than other sex workers.

[9]The other model results are available from the authors upon request. Notice that further test statistics suggest that the instruments do not under-identify and are now weak. Yet, as with the other test statistics, these are biased with limited dependent endogenous right-hand-side variables.

References

Arunachalam, R. and Shah, M., 2013. Compensated for Life: Sex work and Disease Risk. *Journal of Human Resources* 48(2), 345-369.

Badinger, H. Egger, P.H., 2014. Fixed effects and random effects estimation of higher-order spatial autoregressive models with spatial autoregressive and heteroskedastic disturbances, *CESifo Working Paper* No. 4847, Munich, CESifo.

Cawley, J., 2014. The Impact of Obesity on Wages. *The Journal of Human Resources* 39(2), 451-474.

Chang, H.-H., Weng, Y., 2012. What is more important for prostitute price? Physical appearance or risky sex behavior?, *Economics Letters* 117(2), 480-483.

Cho, S.-Y., Dreher, A., Neumayer, E., 2013. Does legalized prostitution increase human trafficking?, *World Development* 41 (2013), 67-82.

Cunningham, S. and T.D. Kenndall, 2010. Risk behaviors among internet-facilitated sex workers: evidence from two new datasets. *Sexual Transmitted Infections* 86 (Supplement 3), iii100-iii105.

Cunningham, S., Kendall, T. D., 2011. Prostitution 2.0: The changing face of sex work., *Journal of Urban Economics* 69(3), 273-287.

De la Torre, A., Havenner, A., Adams, K., Ng, J., 2010. Premium sex: Factors influencing the negotiated price of unprotected sex by female sex workers in Mexico, *Journal of Applied Economics*, 13(1), 67-90.

Fletcher, J., 2009. Beauty vs. brain: early labor market outcomes of high school graduates, *Economics Letters* 105(3), 480-483.

Gertler, P. Shah, M., Bertozzi, S.M., 2005. Risky business: The market for unprotected commercial sex. *Journal of Political Economy* 113(3), 519-550.

Hamermesh, D., Biddle, J., 1994. Beauty and the labor market, *American Economic Review* 84(5), 1174-1194.

Heckman, J.J., 1978. Dummy endogenous variables in a simultaneous equations system, *Econometrica* 46(4), 931-492.

Heckman, J.J. and E. Vytlacil, 1999. Local instrumental variables and latent variable models for identifying and bounding treatment effects, *Proceedings of the National Academy of Sciences* 96, 4730-4734.

Johnston, D. W., 2010. Physical appearance and wages: Do blondes have more fun?. *Economics Letters*, 108(1), 10-12.

Kuersteiner, G., Prucha, I.R., 2013. Limit theory for panel data models with cross sectional dependence and sequential exogeneity, *Journal of Econometrics* 174(2), 107-126.

Maddala, G.S., 1983. *Limited Dependent and Qualitative Variables in Econometrics. Econometric Society Monographs No. 3*, Cambridge UK, Cambridge University Press.

Mobius, M., Rosenblatt, T., 2006. Why beauty matters?, *American Economic Review* 96(1), 222-235.

Moffat, P.G., Peters, S.A., 2004. Pricing personal services: an empirical study of earnings in the UK prostitution industry, *Scottish Journal of Political Economy* 51(5), 675-690.

Plankey, M., Stevens, J., Flegal, K. M., and Rust, P. F., 1997. Prediction Equations Do Not Eliminate Systematic Error in Self-Reported Body Mass Index. *Obesity Research* 5(4), 308-314.

Rao, V., Gupta, I., Lokshin, M., Jana, S., 2003. Sex workers and the cost of safe sex: the compensating differential for condom use among Calcutta prostitutes, *Journal of Development Economics* 71(2), 585-603.

Reichel R, Topper K., 2003. Prostitution: Der verkannte Wirtschaftsfaktor. Aufklaerung und Kritik. *Zeitschrift fuer freies Denken und humanistische Philosophie* 10(3), 2-30.

Robinson, J., Yeh, E., 2011. Transactional sex as a response to risk in western Kenya, *American Economic Journal: Applied Economics* 3(1), 35-64.

Vella, F. and M. Verbeek, 1999. Estimating and interpreting models with endogenous treatment effects, *Journal of Business and Economic Statistics* 17(4), 473-478.

Wilson, N., 2012. Economic booms and risky sexual behavior: Evidence from Zambian copper mining cities, *Journal of Health Economics* 31(6), 797-812.

Wooldridge, J.M., 2002. *Econometric Analysis of Cross Section and Panel Data*, Cambridge MA, MIT Press.