

Small Sample Bias Using Maximum Likelihood versus
Moments: The Case of a Simple Search Model of the Labor
Market*

Alice Schoonbroodt[†]

University of Minnesota, MN

March 12, 2004

Abstract

I investigate the problem of small sample biases, when using Maximum Likelihood (ML) versus Moments (MOM) to estimate the parameters of a simple search model from accepted wage and duration (to first job) data only. Using a Monte Carlo (MC) procedure I show that ML displays a much larger bias than MOM in small samples. The fact that ML estimation connects all moments makes it efficient for large samples but subject to "bias contamination effects" for small samples. MOM estimation on the other hand picks out only a few moments and can thereby avoid the problematic ones.

*I thank Zvi Eckstein, Tom Holmes, Sam Kortum and Andrea Moro, as well as participants at the micro workshop for helpful comments.

[†]Contact information: Alice Schoonbroodt, 271 19th Avenue S Heller Hall 1035, Minneapolis MN, 55455, Tel.: 1-612-204-5544, Fax: 1-612-204-5515, e-mail: alicesch@econ.umn.edu

1 INTRODUCTION

Duration of unemployment depends on the product of two probabilities: the probability to get a job offer on the one hand, and the probability that this offer is accepted on the other. Any analysis of unemployment depends heavily on the estimation of those two probabilities. One way to analyze them is through the search model. When looking at the escape rate from unemployment which determines the reservation wage and duration of unemployment, it is important to be able to identify effects that come from the arrival rate of job offers as opposed to those that come from the offer distribution of wages.¹ Therefore it is important to choose the most reliable estimation method available. One problem in empirical work is sample size. The search model is a good apparatus to analyze panel data and therefore lends itself to the comparison of different estimation methods.

In this paper, I investigate the problem of small sample biases, when using Maximum Likelihood (ML) versus Moments (MOM) to estimate the parameters of a simple search model from accepted wage and duration (to first job) data only. Using a Monte Carlo (MC) procedure I show that even though the ML estimator is consistent as the number of observations becomes very large, there is a very large bias when relatively small samples are used. Among consistent estimators, it is commonly believed that ML estimation is a better method than moments (MOM) estimation, because ML takes all the margins into account at the same time, which translates into higher efficiency - for large samples. I show that for small samples quite the opposite is true. In fact, MOM estimation is much less biased than ML precisely because the parameters influencing the two probabilities mentioned above can be estimated separately, which avoids a contamination effect present in the small sample ML estimation.

The question of small sample properties of the ML estimator has come up in previous empirical

¹For example, when considering male-female wage and duration of unemployment differentials, being able to identify the arrival rate apart from the offer distribution parameters, could give a clue as to whether there is discrimination against women or whether they simply have lower search efforts. Biased estimates of the parameters might lead to serious flaws in the policy implications.

work. It is particularly important in applied microeconomics and especially around the search model.² It turns out that the theory about it is very limited.³ Flinn and Heckman (1982) note the ML estimation bias but do not compare ML and MOM estimation in small samples. I use a Monte Carlo (MC) procedure to evaluate ML versus MOM estimation when structurally estimating a simple search model using wage and duration-to-(first)-job data only.

I find that ML displays a tremendous upward bias for the Poisson arrival rate of job offers, p , in samples smaller than 50 observations. For a bias on p smaller than 10% and 5%, the sample size should have at least 220 versus 450 observations respectively. Moreover, there is a large correlation between the parameter estimates that are obtained simultaneously from the three ML first order conditions. MOM estimation on the other hand displays a much less important bias for small samples (about half the bias from ML estimation). I show that the most important reason for this difference is a contamination effect present in the ML estimation and avoidable in the MOM estimation. With MOM estimation the parameters of the wage offer distribution can be estimated separately from the arrival rate, p . This implies that, the bias problem that occurs when estimating p , does not contaminate the estimation of the other parameters, which in turn keeps the bias on p relatively low.

The outline of the paper is as follows. Section 2 presents the a simple model based on Mortensen (1986), derives closed form solutions given a set of parameters and distributional assumptions. Section 3 and 4 describe the two estimation methods for this particular model, namely maximum likelihood estimation and moments estimation. In Section 5 I explain the Monte Carlo procedure in more detail. In Section 6 I present the small sample results from ML and MOM estimation. Section 7 shows how

²One example is Eckstein and Wolpin (1995), where wage and duration differentials between different education levels are considered by race. They estimate a discrete time search model of the labor market, where they subdivide their data set into black-white and within these categories into five education groups. This has led to quite small samples for some groups. It turns out that the parameter p , which is the probability of getting a wage offer within a given period was estimated close to one for most groups. This should make us suspicious of small sample biases.

³See Davidson and MacKinnon (1993), p. 247.

the results point to a contamination effect in the ML estimation that is not present in the MOM estimation. Section 8 concludes.

2 THE MODEL⁴

2.1 Setup

Consider the following partial equilibrium model. A worker gets wage offers and decides whether to accept the job and earn the offered wage forever or to turn down the current offer in expectation for a better one. The individual is assumed to maximize expected present value of earnings over an infinite horizon with linear preferences.

Let $F(w), w \in (0, \infty)$, be the c.d.f. of the offered wage distribution, and let $f(w)$ be its density. This distribution is taken as exogenously given. If the individual accepts the wage offer then he/she receives this wage (w) forever. If the individual rejects the wage offer or receives no offer, then he/she goes on searching and gets b , the instantaneous utility in the unemployment state.⁵ Furthermore, let p be the arrival rate of wage offers and r the rate of time preference.

There are only two states, unemployed (search) and employed. At every wage offer, the individual is thus facing a discrete choice problem. Let V^u be the stationary expected present value of being unemployed and search for a job. Let $V^e(w)$ be the expected present value of being employed (forever) at wage w . Thus, the optimal choice is to continue searching as long as $V^u > V^e(w)$.

⁴The theoretical results are mostly from Mortensen (1986). The structural estimation part is from Flinn and Heckman (1981) for ML estimation and Lecture Notes by Zvi Eckstein at the University of Minnesota, MN for MOM estimation.

⁵That is, the value of not working = leisure + unemployment compensation + other income. Note that b is assumed to be independent of time spent searching.

2.2 Solution

The solution is an *optimal stopping rule*. That is a reservation wage w^* such that if $w < w^*$ the individual rejects the wage offer, and accepts it otherwise. This rule maximizes the present value of lifetime utility and implies that w^* is defined by $V^u = V^e(w^*) = (w^*/r)$. Using the Bellman equation as in Mortensen (1986) gives the following implicit equation for w^* as a function of the parameters $(b, p, r, f(\cdot))$.

$$w^* = b + \frac{p}{r} \int_{w^*}^{\infty} (w - w^*) f(w) dw \quad (1)$$

Assuming that $F(\cdot)$ is the log-normal distribution with mean μ and variance σ^2 , the reservation wage should satisfy the following analytical closed form equation⁶

$$w^* = b - \frac{pw^*}{r} \left[1 - \Phi \left(\frac{\ln w^* - \mu}{\sigma} \right) \right] + \frac{p}{r} \exp\{\mu + 0.5\sigma^2\} \left[1 - \Phi \left(\frac{\ln w^* - (\mu + \sigma^2)}{\sigma} \right) \right] \quad (2)$$

where $\Phi(\cdot)$ is the standard normal distribution with mean 0 and variance 1.

3 MAXIMUM LIKELIHOOD

3.1 The likelihood function

The likelihood of observing a particular data set of size I , $(d_1, d_2, \dots, d_I; w_1, w_2, \dots, w_I)$, given the parameters (b, r, p, μ, σ) , is:

⁶See appendix for details (equation (20)).

$$L(\{d_i, w_i\}_{i=1}^I | \cdot) = \prod_{i=1}^I [pf(w_i) \exp\{-p(1 - F(w^*(\cdot)))d_i\}] \quad (3)$$

where $(\cdot) = (b, r, p, \mu, \sigma)$.

For the log-normal, continuous case the log-likelihood becomes⁷:

$$\begin{aligned} \ln L(\{d_i, w_i\}_{i=1}^I | \cdot) = \\ \sum_{i=1}^I \left[\ln p - \ln(w_i \sigma \sqrt{2\pi}) - 0.5 \left(\frac{\ln w_i - \mu}{\sigma} \right)^2 - p \left(1 - \Phi \left(\frac{\ln w^*(\cdot) - \mu}{\sigma} \right) \right) d_i \right] \end{aligned} \quad (4)$$

3.2 Maximum Likelihood estimation

First, I use the count estimator proposed in Flinn and Heckman (1982) to determine the reservation wage $\hat{w}^* = \min_{i=1, \dots, I} w_i$. They show that this estimator is strongly consistent. Then, given the data, $\ln L$ is maximized choosing $\hat{p}, \hat{\mu}$ and $\hat{\sigma}$. The ML estimators $(\hat{\mu}, \hat{\sigma}, \hat{p})$ are given by the solution to the following three equations:

$$\frac{d \ln L}{d \hat{\mu}} = \sum_{i=1}^I \frac{\ln w_i}{\hat{\sigma}^2} - \frac{I \hat{\mu}}{\hat{\sigma}^2} - \frac{\hat{p}}{\hat{\sigma}} * \Phi \left(\frac{\ln \hat{w}^* - \hat{\mu}}{\hat{\sigma}} \right) * \sum_{i=1}^I d_i = 0 \quad (5)$$

$$\frac{d \ln L}{d \hat{\sigma}} = \sum_{i=1}^I \frac{(\ln w_i - \hat{\mu})^2}{\hat{\sigma}^3} - \frac{I}{\hat{\sigma}} - \frac{\hat{p}}{\hat{\sigma}} * \Phi \left(\frac{\ln \hat{w}^* - \hat{\mu}}{\hat{\sigma}} \right) * \frac{\ln \hat{w}^* - \hat{\mu}}{\hat{\sigma}} * \sum_{i=1}^I d_i = 0 \quad (6)$$

$$\frac{d \ln L}{d \hat{p}} = \frac{I}{\hat{p}} - \left[1 - \Phi \left(\frac{\ln \hat{w}^* - \hat{\mu}}{\hat{\sigma}} \right) \right] * \sum_{i=1}^I d_i = 0 \quad (7)$$

checking that second derivatives are negative. Here d_i are completed unemployment spells. Now,

⁷See appendix for details (equations (14) and (19)).

taking r as given, the equation of the reservation wage gives the estimator \hat{b} as in:

$$\hat{b} = \hat{w}^* + \frac{\hat{p}\hat{w}^*}{r} \left[1 - \Phi \left(\frac{\ln \hat{w}^* - \hat{\mu}}{\hat{\sigma}} \right) \right] - \frac{\hat{p}}{r} \exp\{\hat{\mu} + 0.5\hat{\sigma}^2\} \left[1 - \Phi \left(\frac{\ln \hat{w}^* - (\hat{\mu} + \hat{\sigma}^2)}{\hat{\sigma}} \right) \right] \quad (8)$$

Note that this means that with real data sets, the model cannot tell r apart from b , given duration and wage data only. The interpretation of b in this model is the value of leisure together with potential unemployment benefits, and r is the time-preference parameter. While the value of leisure is hard to identify from outside, the time preference parameter has usually been backed out from interest rate data. This is why I take r as given and estimate b from (8).

4 MOMENTS ESTIMATION

Again the reservation wage, is obtained from Flinn and Heckman (1982)'s count estimator, $\tilde{w}^* = \min_{i=1, \dots, I} w_i$. The wage data provides two other moments, namely⁸

$$E(w|w > \tilde{w}^*) = \frac{\left(1 - \Phi \left(\frac{\ln \tilde{w}^* - \tilde{\mu} - \tilde{\sigma}^2}{\tilde{\sigma}} \right) \right) * \exp \left\{ \tilde{\mu} + 0.5\tilde{\sigma}^2 \right\}}{\left(1 - \Phi \left(\frac{\ln \tilde{w}^* - \tilde{\mu}}{\tilde{\sigma}} \right) \right)} \quad (9)$$

and

$$Var(w|w > \tilde{w}^*) = \frac{\left(1 - \Phi \left(\frac{\ln \tilde{w}^* - \tilde{\mu} - 2\tilde{\sigma}^2}{\tilde{\sigma}} \right) \right) * \exp \left\{ 2 \left(\tilde{\mu} + \tilde{\sigma}^2 \right) \right\}}{\left(1 - \Phi \left(\frac{\ln \tilde{w}^* - \tilde{\mu}}{\tilde{\sigma}} \right) \right)} - (E(w|w > \tilde{w}^*))^2 \quad (10)$$

These two equations, given \tilde{w}^* , identify $\tilde{\mu}$ and $\tilde{\sigma}$. Note that (9) and (10) are independent on \tilde{p} and don't use duration data. This will be important in the analysis of the results below.

⁸See appendix for details (equations (27) and (28)).

If w^*, μ and σ are known, and using the assumed distribution of wage offers, then p solves⁹,

$$E(d) = \frac{1}{p^* \left(1 - \Phi\left(\frac{\ln w^* - \mu}{\sigma}\right)\right)} \quad (11)$$

where $E(d)$ is a moment we can get from the duration data. So once we found $\tilde{\mu}$ and $\tilde{\sigma}$ from (9) and (10), \tilde{p} is given by

$$\tilde{p} = \frac{1}{\left(\sum_{i=1}^I \frac{d_i}{T}\right)^* \left(1 - \Phi\left(\frac{\ln \tilde{w}^* - \tilde{\mu}}{\tilde{\sigma}}\right)\right)} \quad (12)$$

Comparing equation (12) to equation (7), it is clear that for given estimates of w^*, μ and σ , the MOM estimator, \tilde{p} , coincides with its ML estimator, \hat{p} . Again d_i are completed spells.

5 MONTE CARLO PROCEDURE

First, I chose consistent parameter values because given a choice of (w^*, p, μ, σ, r) , b is uniquely determined through (1). For the sake of realism I chose the reservation wage w^* close to the minimum wage. Furthermore I chose the parameter values, μ and σ , so as to match the data for White Male High School Graduates in Eckstein and Wolpin (1995), that is a mean observed wage of \$9 and a coefficient of correlation of 0.47. Then I chose the rate of time preference to match a discount rate of 0.9. Finally for different values of p I calculated the corresponding b to keep the other parameters fixed.¹⁰ The appendix summarizes some sensitivity analysis. The chosen values are:

⁹See appendix for details (equation (26)).

¹⁰See equation (20) in appendix.

$w^* = \$5$	p	$b(\$)$
$\mu = 1.7$	0.3	-1.1295764
$\sigma = 0.6$	0.5	-5.2159606
$r = 0.111\dots$	0.7	-9.3023448

This gives $E(w) = \$6.55$ and $Std(w) = 4.31$.

Using Gauss 3.0, I then generated data on wages $(w_i)_{i=1}^I$ and duration $(d_i)_{i=1}^I$ where I denotes the sample size, according to the distributions assumed above and using the parameter vector (b, r, p, μ, σ) . I consider only complete spells to avoid any right hand-side truncation of the duration data. I then use ML and MOM estimation as described above. This data generation and estimation procedure is performed 500 times for every sample size considered. For every sample size, I , this provides us with an estimate of the mean ML-bias, $m_bias\hat{\theta}_I = \sum_{n=1}^{500} \frac{\hat{\theta}_{In} - \theta}{500}$ and the mean MOM-bias, $m_bias\tilde{\theta}_I = \sum_{n=1}^{500} \frac{\tilde{\theta}_{In} - \theta}{500}$, where $\theta \in \{w^*, b, p, \mu, \sigma\}$. I express the mean bias as a percentage of the true parameter. As the sample size increases, one can observe how fast $(m_bias\hat{\theta}_I/\theta)$ and $(m_bias\tilde{\theta}_I/\theta)$ go to zero.

In the next section, I present the results from the ML estimation in terms of biases and give minimum sample sizes for a bias on \hat{p} smaller than 10%, 5% and 1%, given the parameters chosen. I then plot histograms of the 500 estimates for every sample size and present some sensitivity analysis. Finally I address the two main reasons for the large bias on \hat{p} in small samples. In particular, I show how the count estimation of w^* and the joint ML estimation of $\hat{\mu}, \hat{\sigma}$ and \hat{p} affect the biases.

Then I present the results from MOM estimation in comparison to ML estimation. It turns out that for a given mean bias, one needs only about half as many observations when using MOM as opposed to ML.

6 RESULTS

6.1 Results from ML estimation

6.1.1 Mean bias as I increases

Figures (1) to (3) show the mean bias as a percentage of the true parameter against increasing sample sizes, for different true values of p ($p = 0.3, p = 0.5, p = 0.7$) and Table 1 shows the numbers for small samples. Here are the most important observations:

- there is an important small sample bias in particular as far as the estimation of p is concerned; that is for small samples the estimator's expectation is different from the true parameter value (this is shown since for every sample size a large number of simulations and estimates are obtained) while the larger the sample gets, the closer the mean bias comes to zero and the smaller its variance.
- for all values of p , the mean \hat{p} is greater than p , the mean $\hat{\mu}$ is smaller than μ and the mean $\hat{\sigma}$ is greater than σ ;
- for all values of p , for samples smaller than 200 observations, the mean bias as a percentage of the true parameter is very large for \hat{p} , while $\hat{\mu}$ and $\hat{\sigma}$ are relatively well estimated by ML. Table 1 shows the percentage values of the bias for $I = 50, I = 100$ and $I = 200$.
- Figures (1), (2) and (3) plot the bias for I increasing up to 12,800 observations. Clearly the bias decreases almost monotonically towards zero.
- for a bias of less than 10% on \hat{p} we need $I > 220$, less than 5% we need $I > 450$ and less than 1% we need $I > 3,200$, for all three true values of p .

Since the observations are very similar for the three values of p , I will focus on $p = 0.5$ for the remainder of the paper.

6.1.2 Histograms as I increases

The above suggests that one way of constructing a better estimator could be to subtract a constant from the ML estimate. However looking at histograms of \hat{p} for different sample sizes (see Figure (4)), the main observation is that for small samples, the estimator either does pretty well or shoots off way above the true value. The estimates close to the (arbitrary) upper bound of 1000 weigh a lot in the mean bias. In other words, the estimator does not hit its average value very often. Therefore we cannot expect to get a more useful estimator by subtracting some constant from any given ML estimate.

6.1.3 Effect of the count estimate for w^*

Since $\frac{dw^*}{dp}$ is positive and $\hat{w}^* = \min w_i$ is an upward biased estimator for w^* ¹¹, the obvious next step is to see how much of the bias on \hat{p} is due to the count estimation of w^* . Therefore I estimated μ, σ and p taking the true value of the reservation wage as given, i.e. $\hat{w}^* = w^*$, using the same simulated data as for the experiment above. This allows us to see the gain from a hypothetical unbiased estimation of the reservation wage.

It turns out that the bias is reduced to half for small samples (see Figure (5) where ML stands for maximum likelihood and MLRWG stands for ML taking the reservation wage as given). However, the bias on the \hat{w}^* is small and goes away quite quickly as sample size increases. Therefore we cannot expect to gain much from this adjustment for samples greater than 150 observations. The size of

¹¹See Flinn and Heckman (1982), p.130.

the small sample problem coming from the count estimate of w^* could be diminished by introducing measurement error.¹²

6.1.4 Correlation between \hat{p} , $\hat{\mu}$ and $\hat{\sigma}$

The ML estimates of \hat{p} , $\hat{\mu}$ and $\hat{\sigma}$ are highly correlated. The correlation coefficients up to $I = 400$ are in Table 2. Also, whenever the estimates for the mean and the standard deviation are close to their true value, i.e. $\hat{\mu} \simeq \mu$ and $\hat{\sigma} \simeq \sigma$, then so is the estimate for the arrival rate, i.e. $\hat{p} \simeq p$. This suggests that if μ and σ could be estimated more precisely and independently of the estimate for arrival rate of job offers, \hat{p} , then \hat{p} would display a much smaller bias when estimated by ML. This is exactly what MOM estimation does: equations (9) and (10) are independent on \tilde{p} .

6.2 Results from MOM estimation in comparison with ML estimation

Again I plotted the mean bias as a percentage of the true parameter against increasing sample size for different estimation methods. The results are displayed in Figure (6), where ML stands for maximum likelihood, MLRWG stands for ML taking the reservation wage as given, MOM stands for moments estimation and MOMRWG stands for MOM taking the reservation wage as given. Table 3 gives the small sample bias for each experiment.

Clearly MOM estimation does much better than ML estimation for small samples. Still the bias on \tilde{w}^* remains. If we take $\tilde{w}^* = w^*$ (MOMRWG), then there is a bias of about 13% left even for a sample size as small as $I = 100$.

Finally, the variance of the estimators is much higher for ML than for MOM estimations for very small samples. However, once a sample size of 400 or more observations is reached, the results of

¹²See Wolpin (1987).

the MC method are consistent with the theory in that ML estimation is more efficient than MOM estimation.

7 CONTAMINATION

Despite the higher efficiency of ML estimation over MOM estimation for large samples, the above results clearly show that for small samples MOM estimation is by far less biased than ML. The main reason for this result is the fact that in the ML formulation $\hat{p}, \hat{\mu}$ and $\hat{\sigma}$ are estimated simultaneously. The argument is the following. Given estimates for the reservation wage, w^* and the wage offer distribution parameters, μ and σ , equations (7) (ML) and (12) (MOM) give the same estimate for p . Thus for the bias on p to be so large under ML as opposed to MOM, it must be that the estimation of μ and σ is less good under ML than MOM. Now the main difference between the two estimation methods in the present framework is that, to estimate μ and σ , MOM uses wage data only, while the simultaneous estimation of μ, σ and p under ML estimation requires the use of both, wage and duration data at the same time. Therefore the bias lies in the duration data. So with ML, the estimation of the wage offer distribution parameters are also contaminated which in turn aggravates the bias on p . In fact, given that the correlation between these estimates is so high, any bias in the estimation of p , contaminates the estimation of μ and σ . This in turn feeds back into the bias on p . Unlike in the ML estimation of μ and σ , their MOM estimation is independent on p . Since their estimation is not contaminated by the estimation of p , it follows that the bias on p is also smaller.

8 CONCLUDING REMARKS

In this paper, I investigated the problem of small sample biases, when using Maximum Likelihood (ML) versus Moments (MOM) to estimate the parameters of a simple infinite horizon partial equilibrium search model in continuous time from accepted wage and duration (to first job) data only. Using a Monte Carlo (MC) procedure I show that there is a serious small sample bias when using Maximum Likelihood. I documented different dimensions of this bias: how fast it decreases as sample size increases, what patterns the variance of the estimates displays and how the bias changes for different true values chosen in the Monte Carlo simulation. Two interesting features of the bias on the offer arrival rate have been analyzed in more detail. The first one, namely the biased count estimation of the reservation wage and the positive relationship of the latter with the offer arrival rate in the model's equations, is easily taken care of ($\hat{w}^* = w^*$). However, the second feature, namely the high correlation between the ML estimates of the offer arrival rate and the two offer distribution parameters, which are jointly estimated under the ML procedure, is more subtle. There is a "contamination effect": if we get one parameter slightly wrong, the other shoots off completely and vice versa. Moments estimation separates the estimation of the wage offer distribution parameters and the offer arrival rate so that there is no contamination. I show that the limiting results of consistency of the ML estimators and their higher efficiency compared to MOM estimators hold true when large samples are considered. In small samples however MOM estimation is both, less biased and more efficient because the contamination effect can be avoided. The strength of ML over MOM in large samples actually constitutes its weakness in small samples.

9 REFERENCES

- Burdett, K., (1981), “A useful restriction on the offer distribution in job search models”, in G. Eliasson, B. Holmlund and F. P. Stafford, eds., *Studies in Labor Market Behavior: Sweden and the United States*, Stockholm, Sweden: I.U.I. Conference Report.
- Davidson R. and J.G. MacKinnon, (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York, NY.
- Eckstein, Z. and K.I. Wolpin, (1995), “Duration to first job and return to schooling: estimates from a search-matching model”, *The Review of Economic Studies*, 62, 263–286.
- Flinn, C.J. and J.J. Heckman, (1982), “New methods for analyzing structural models of labor force dynamics”, *Journal of Econometrics*, 18, 115–168.
- Mortensen, D.T., (1986), “Job Search and Labor Market Analysis”, in *Handbook of Labor Economics*, Volume 2, 849-919.
- Wolpin, K.I., (1987), “Estimating a structural job search model: the transition from school to work”, *Econometrica*, 55, 801–818.

10 APPENDIX

10.1 Functional forms and comparative statics

10.1.1 The reservation wage and the distribution of accepted wages

In equation (1), let the offered wage w be log-normally distributed. That is $w = e^y$; $y \sim N(\mu, \sigma^2)$ or $\ln w = y \sim N(\mu, \sigma^2)$. Let $\psi(y)$ denote the distribution of $y = \ln w$, i.e.

$$\psi(y) = (\sigma\sqrt{2\pi})^{-1} \exp\left\{-0.5\left(\frac{y-\mu}{\sigma}\right)^2\right\} \quad (13)$$

and let $\Psi(y)$ denote its c.d.f., i.e. $\Psi'(y) = \psi(y)$. Then $f(w)$, the density of the wage, w , is

$$\begin{aligned} f(w) &= \psi(y)|\partial y/\partial w| = w^{-1}\psi(y) \\ &= (w\sigma\sqrt{2\pi})^{-1} \exp\left\{-0.5\left(\frac{\ln w - \mu}{\sigma}\right)^2\right\} \end{aligned} \quad (14)$$

Moreover $F(w) = \Psi(\ln w)$ for all $w \in (0, \infty)$.

The mean offered wage is given by $E(w) = E(e^y) = \int_{-\infty}^{\infty} e^y \psi(y) dy$. Substituting $\psi(y)$ and rearranging terms provides the result:

$$E(w) = \exp\{\mu + 0.5\sigma^2\} \int_{-\infty}^{\infty} (\sigma\sqrt{2\pi})^{-1} \exp\left\{-0.5\left(\frac{y-\mu-\sigma^2}{\sigma}\right)^2\right\} dy \quad (15)$$

Since the last term is the integral of the normal density that is equal to one we get the result that

$$E(w) = \exp\{\mu + 0.5\sigma^2\} \quad (16)$$

An its variance is given by

$$Var(w) = \exp \{2\mu + \sigma^2\} * (\exp \{\sigma^2\} - 1) \quad (17)$$

In addition we can write the following:

$$\begin{aligned} \int_{w^*}^{\infty} wf(w)dw &= \Pr(w \geq w^*)E(w|w \geq w^*) = \int_{\ln w^*}^{\infty} e^y \psi(y) dy \\ &= \exp \{ \mu + 0.5\sigma^2 \} \left[1 - \Phi \left(\frac{\ln w^* - (\mu + \sigma^2)}{\sigma} \right) \right] \end{aligned} \quad (18)$$

where Φ is the c.d.f. of standard normal.

We also have:

$$\int_{w^*}^{\infty} f(w)dw = \int_{\ln w^*}^{\infty} g(y)dy = \left[1 - \Phi \left(\frac{\ln w^* - \mu}{\sigma} \right) \right] = 1 - F(w^*) \quad (19)$$

With the above two equations the integral $\int_{w^*}^{\infty} (w - w^*)f(w)dw$ can be solved analytically such that the reservation wage should satisfy the following analytical closed form equation

$$w^* = b - \frac{pw^*}{r} \left[1 - \Phi \left(\frac{\ln w^* - \mu}{\sigma} \right) \right] + \frac{p}{r} \exp \{ \mu + 0.5\sigma^2 \} \left[1 - \Phi \left(\frac{\ln w^* - (\mu + \sigma^2)}{\sigma} \right) \right] \quad (20)$$

This is equation (2) in the text.

10.1.2 The hazard rate and the distribution of duration

The model above is one basic tool to analyze the determinants of the duration of unemployment. To this end, let us first derive the distribution of unemployment duration given the parameters $(p, b, f(\cdot), r)$.

The probability of receiving offers during an interval of d_i that are all rejected, i.e. less than w^* , is

$$\Pr(d \geq d_i) = \sum_{j=1}^{\infty} (pd_i)^j \left(\frac{e^{-pd_i}}{j!} \right) [F(w^*)]^j \quad (21)$$

multiplying by $\exp\{-pF(w^*) + pF(w^*)\}$ implies that the survivor function is:

$$\Pr(d \geq d_i) = \exp\{-p(1 - F(w^*)) d_i\}. \quad (22)$$

Let h denote the escape rate from unemployment or simply the hazard rate. Then $h = p(1 - F(w^*))$ and thus duration d_i is distributed exponentially with parameter h . Thus its C.D.F. is:

$$G(d_i) = \Pr(d < d_i) = 1 - \exp\{-hd_i\} \quad (23)$$

and density

$$g(d_i) = h \exp\{-hd_i\}. \quad (24)$$

Therefore expected unemployment duration is

$$E(x) = \int_0^{\infty} xg(x)dx = \int_0^{\infty} xh \exp\{-hx\}dx. \quad (25)$$

Using integration by parts, we get

$$E(d) = \frac{1}{h} = \frac{1}{p(1 - F(w^*))} \quad (26)$$

which is exactly (11) in the text.

10.1.3 Comparative Statics

The arrival rate of job offers p and the parameters of the wage offer distribution $f(\cdot)$, μ and σ , affect unemployment duration in two ways: a direct way, that is keeping the reservation wage constant, and in an indirect way, that is through their affect on the reservation wage. On the other hand b and r , affect expected duration through the reservation wage only.

It is easy to show that $\frac{dw^*}{dp} > 0$, $\frac{dw^*}{d\mu} > 0$, $\frac{dw^*}{d\sigma} > 0$, $\frac{dw^*}{db} \in (0, 1)$ and $\frac{dw^*}{dr} < 0$. Since $\frac{\partial E(d)}{\partial w^*} > 0$, an increase in b , the value of unemployment increases expected duration of unemployment, whereas an increase in r , the rate of time preference, decreases expected unemployment duration. For the parameters of the wage offer distribution, μ and σ , Mortensen (1986) shows that $\frac{\partial E(d)}{\partial \mu} < 0$, whereas $\frac{\partial E(d)}{\partial \sigma^2}$ has an ambiguous sign in general. Burdett(1981) shows that a sufficient condition for the intuitive result that an increase in job availability should lower unemployment duration i.e. $\frac{\partial E(d)}{\partial p} < 0$, is for $f(\cdot)$ to be "log-concave".

10.2 Moments estimation

Equation (9) comes from:

$$\begin{aligned} E(w|w > \tilde{w}^*) &= \frac{\int_{\tilde{w}^*}^{\infty} wf(w)dw}{\Pr(w \geq \tilde{w}^*)} = \frac{\int_{\ln \tilde{w}^*}^{\infty} e^y \psi(y) dy}{\Pr(y \geq \ln \tilde{w}^*)} \\ &= \frac{\left(1 - \Phi\left(\frac{\ln \tilde{w}^* - \tilde{\mu} - \tilde{\sigma}^2}{\tilde{\sigma}}\right)\right) * \exp\left\{\tilde{\mu} + 0.5\tilde{\sigma}^2\right\}}{\left(1 - \Phi\left(\frac{\ln \tilde{w}^* - \tilde{\mu}}{\tilde{\sigma}}\right)\right)} \end{aligned} \quad (27)$$

and equation (10) can be derived as in:

$$\begin{aligned}
Var(w|w > \tilde{w}^*) &= E \left[(w - E(w|w > \tilde{w}^*))^2 | w > \tilde{w}^* \right] \\
&= E \left[w^2 - 2wE(w|w > \tilde{w}^*) + [E(w|w > \tilde{w}^*)]^2 | w > \tilde{w}^* \right] \\
&= E(w^2|w > \tilde{w}^*) - (E(w|w > \tilde{w}^*))^2 \\
&= \frac{\int_{\tilde{w}^*}^{\infty} w^2 f(w) dw}{Pr(w \geq \tilde{w}^*)} - (E(w|w > \tilde{w}^*))^2 \\
&= \frac{\int_{\ln \tilde{w}^*}^{\infty} e^{2y} \psi(y) dy}{Pr(y \geq \ln \tilde{w}^*)} - (E(w|w > \tilde{w}^*))^2 \\
&= \frac{\int_{2 \ln \tilde{w}^*}^{\infty} (1/2) e^z \psi(z/2) dz}{Pr(z \geq 2 \ln \tilde{w}^*)} - (E(w|w > \tilde{w}^*))^2 \\
&= \frac{\int_{2 \ln \tilde{w}^*}^{\infty} e^z \xi(z) dz}{Pr(z \geq 2 \ln \tilde{w}^*)} - (E(w|w > \tilde{w}^*))^2 \\
&= \frac{\left(1 - \Phi \left(\frac{\ln \tilde{w}^* - \tilde{\mu} - 2\tilde{\sigma}^2}{\tilde{\sigma}} \right) \right) * \exp \left\{ 2 \left(\tilde{\mu} + \tilde{\sigma}^2 \right) \right\}}{\left(1 - \Phi \left(\frac{\ln \tilde{w}^* - \tilde{\mu}}{\tilde{\sigma}} \right) \right)} - [E(w|w > \tilde{w}^*)]^2
\end{aligned} \tag{28}$$

10.3 Sensitivity to the choice of true parameters

The results presented in Section 6 are not very sensitive to changes in the choice of the original parameter values. In particular, different values for the rate of time preference, r , the leisure/unemployment benefit parameter, b , the mean log wage, μ and , as we have seen, the offer arrival rate, p , do not significantly change the magnitudes of the results, and in no way change the qualitative findings.

However, choosing a very low standard deviation of the log offered wage, σ , does reduce the bias on all three estimates from ML significantly. This suggests an interesting trade-off. Consider Eckstein and Wolpin (1995) for example. Had they controlled for more observables, say marital status, location,..., they might have ended up with lower coefficients of variation, i.e. lower wage variance, within groups.

But also with even fewer observations per group. On the one hand there would have been hope for more reliable estimation due to the lower coefficients of correlation. But at the risk of less reliable estimation due to the even smaller samples.

It is true that, keeping the number of observations constant, a lower log offered wage standard deviation reduces the biases in the MC simulations with ML estimation.

Table 1: Small sample ML bias on \hat{p} , $\hat{\mu}$ and $\hat{\sigma}$ as a percentage of true parameter

for different values of true p ,

$I = 50$	$p = 0.3$	$p = 0.5$	$p = 0.7$
$m_bias\hat{p}$	2683%	2244%	2681%
$m_bias\hat{\mu}$	-19%	-16%	-17%
$m_bias\hat{\sigma}$	11%	9%	10%

$I = 100$	$p = 0.3$	$p = 0.5$	$p = 0.7$
$m_bias\hat{p}$	52%	43%	350%
$m_bias\hat{\mu}$	-7.4%	-7%	-7%
$m_bias\hat{\sigma}$	5%	4.8%	4.3%

$I = 200$	$p = 0.3$	$p = 0.5$	$p = 0.7$
$m_bias\hat{p}$	10.5%	15%	11.5%
$m_bias\hat{\mu}$	-2.5%	-2.7%	-3%
$m_bias\hat{\sigma}$	1.7%	2.7%	1.7%

Table 2: Coefficient of correlation between ML estimates of p , μ and σ

for different sample sizes

I	coef. corr. $(\hat{p}, \hat{\mu})$	coef. corr. $(\hat{p}, \hat{\sigma})$
50	-0.7	0.5
100	-0.8	0.67
200	-0.94	0.8
400	-0.95	0.8

Table 3: Small sample ML, MLRWG, MOM, MOMRWG bias on \hat{p} , $\hat{\mu}$ and $\hat{\sigma}$ as a percentage of true

parameter					
$I = 50$	ML	MLRWG	$I = 50$	MOM	MOMRWG
$m_bias\hat{p}$	2244%	500%	$m_bias\tilde{p}$	49.7%	25%
$m_bias\hat{\mu}$	-16%	-7.5%	$m_bias\tilde{\mu}$	-5.5%	-1.1%
$m_bias\hat{\sigma}$	9%	2.3%	$m_bias\tilde{\sigma}$	-0.04%	-3.7%
$I = 100$	ML	MLRWG	$I = 100$	MOM	MOMRWG
$m_bias\hat{p}$	43%	20%	$m_bias\tilde{p}$	20%	13.2%
$m_bias\hat{\mu}$	-7%	-2.9%	$m_bias\tilde{\mu}$	-2.2%	-1.4%
$m_bias\hat{\sigma}$	4.8%	1%	$m_bias\tilde{\sigma}$	-1%	-1%
$I = 200$	ML	MLRWG	$I = 200$	MOM	MOMRWG
$m_bias\hat{p}$	15%	7.2%	$m_bias\tilde{p}$	9.3%	4.9%
$m_bias\hat{\mu}$	-2.7%	-1.2%	$m_bias\tilde{\mu}$	-0.2%	-0.2%
$m_bias\hat{\sigma}$	2.7%	-0.3%	$m_bias\tilde{\sigma}$	-0.1%	-0.7%

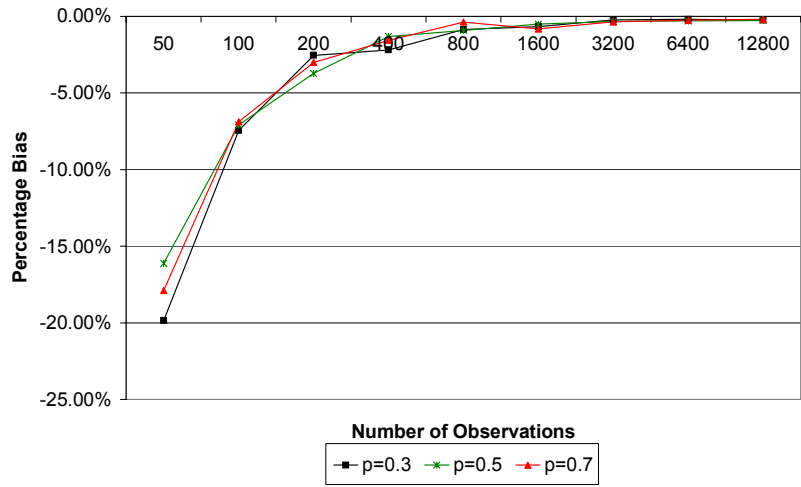


Figure 1: Mean Bias on $\hat{\mu}$ for Increasing Sample Size I

(ML estimation, 500 iterations for every I)

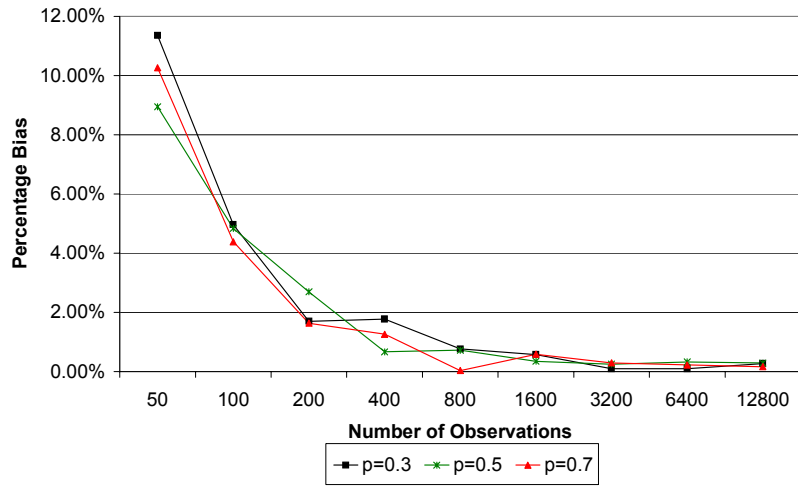


Figure 2: Mean Bias on $\hat{\sigma}$ for Increasing Sample Size I

(ML estimation, 500 iterations for every I)

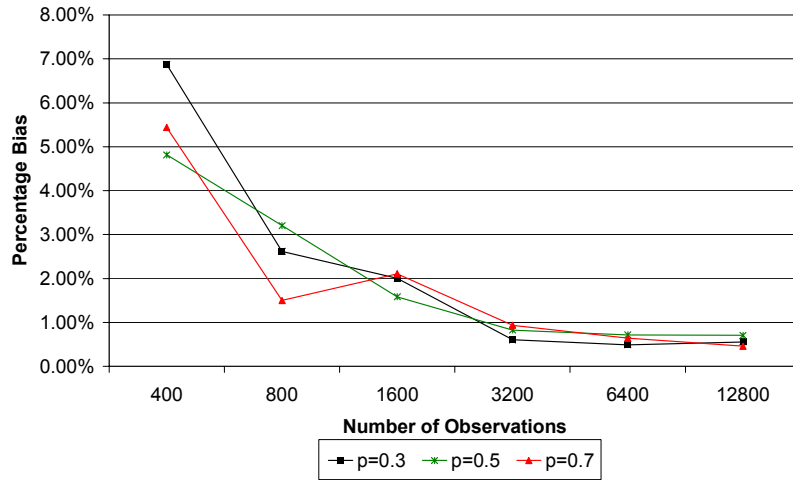


Figure 3: Mean Bias on \hat{p} for Increasing Sample Size I

(ML estimation, 500 iterations for every I) [$I \geq 400$]

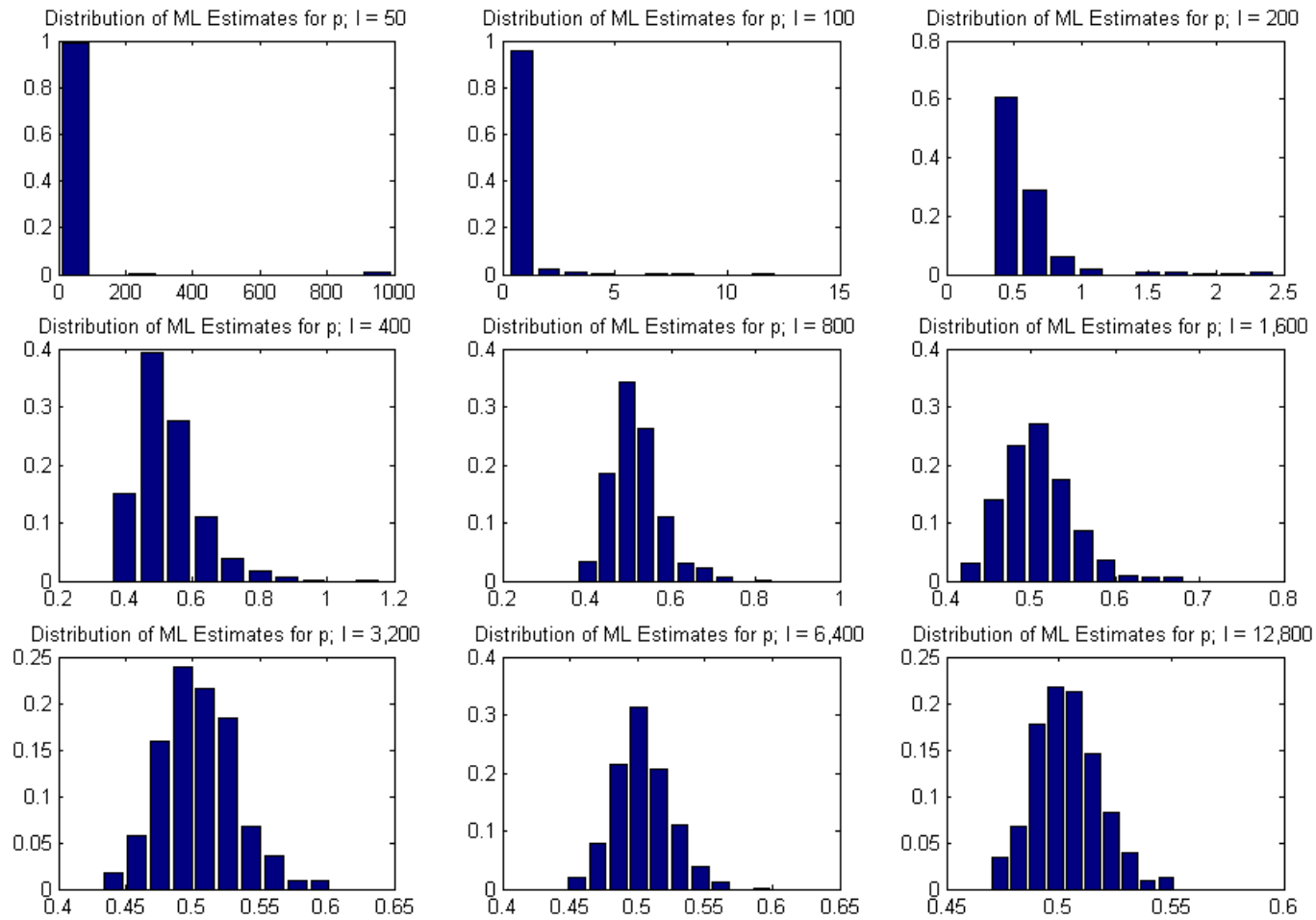


Figure 4: Histograms of 500 \hat{p} Estimates for Increasing Sample Size, I (true $p = 0.5$)

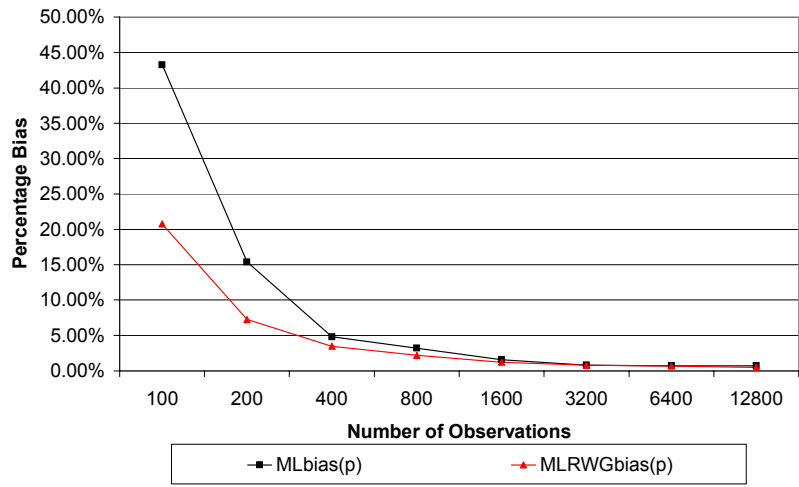


Figure 5: Mean Bias on \hat{p} for Increasing Sample Size I

$$(\hat{w}^* = \min w_i \text{ vs. } \hat{w}^* = w^*)$$

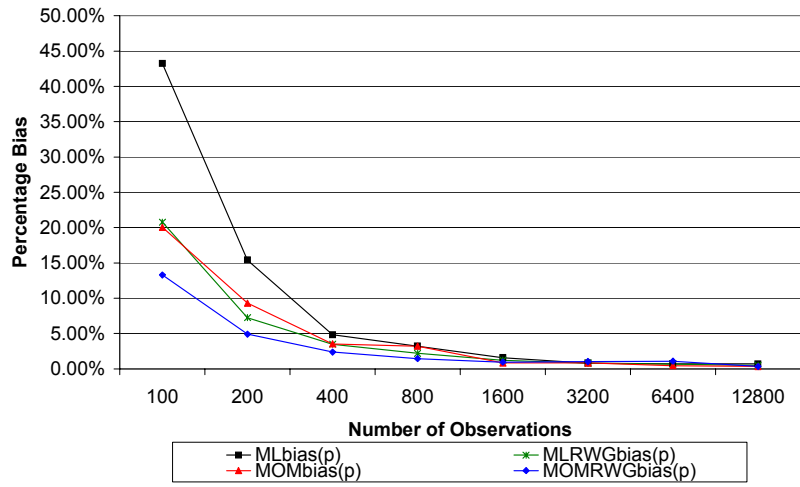


Figure 6: Mean Bias on \hat{p}, \tilde{p} for Increasing Sample Size I

(ML vs. MOM) [$I \geq 100$]