University of Neuchatel

Institute of Economic Research

IRENE Working paper 17-04

Correcting Measurement Errors in Transition Models Based on Retrospective Panel Data

Shaimaa Yassin

Francois Langot

UNIVERSITÉ DE NEUCHÂTEL

Institut de recherches économiques

Correcting Measurement Errors in Transition Models Based on Retrospective Panel Data

Shaimaa Yassin[*] Francois Langot[†]

March 3, 2017

Abstract

We propose in this paper a dynamic n-state transition model to correct for measurement error, that could arise for example from recall and/or design bias, in retrospective panels. Our model allows the correction of measurement errors, when very little auxiliary information is available, over a long period of time taking into consideration the conjuncture fluctuations. The technique suggested shows that it is sufficient to have population moments (for at least one point in time) to correct over- or under-reporting biases. Using a Simulated Method of Moments, one can estimate a transition- and time-specific correction matrix for the labor market flows in a biased retrospective panel. Using retrospective and contemporaneous data from Egypt, we estimate the model and show the significance and robustness of our correction. We show through a reform evaluation that neglecting measurement error in the data would have produced significantly different and misleading results.

JEL classification: C83, C81, J01, J62, J64

Keywords: Panel Data, Retrospective Recall, Measurement Error, Labor Markets, Transition Models.

[*]University of Neuchâtel (Institute of Economic Research, IRENE), University of Paris 1 Panthéon-Sorbonne (CES-Paris School of Economics), University of Le Mans (GAINS-TEPP). E-mail: shaimaa.yassin@unine.ch or yassin.shaimaa@gmail.com

[†]University of Le Mans (GAINS-TEPP & IRA), Paris School of Economics, Cepremap and IZA. E-mail:flangot@univ-lemans.fr.

1 Introduction

Economists have become increasingly aware of the presence of measurement error, especially where empirical analysis rely on surveys susceptible to the data reported being erroneous. While a huge base of the literature has been dedicated to pointing out the problem, few works have actually provided pragmatic solutions. We focus in this paper on labor market transition models. These types of models rely mainly on reported labor market statuses. Whether these labor market states and transitions can be accurately represented by current (contemporaneous) or recalled (retrospective) data is essentially an issue of measurement error. Conclusions and policy implications built upon data without taking into consideration and actually correcting for such measurement errors might therefore be problematic since they might lack accuracy and consequently credibility.

In certain cases of analysis, however, one might be obliged to use certain types of data, such as retrospective data, that might suffer from substantial biases due to recall and design measurement errors. This is true for countries for instance that lack annual panel labor market surveys due to budget constraints and even sometimes due technical problems (where unique panel id's are lost, mixed up and confused....etc). Unfortunately, the concern to obtain accurate results put retrospective panels without correction in a very limited application framework.

Previous literature on measurement error in transition models, have mainly used two approaches. The first approach, in the tradition of the seminal papers of Poterba and Summers (1986, 1995), uses either validation or reinterview data (assuming that such data is error free) to estimate the measurement error. While Poterba and Summers (1986) use the reinterview data from the Current Population Survey to study the impact of measurement error on the estimated number of labor market transitions, Magnac and Visser (1999) use contemporaneous and retrospective data for the same time period to study labor mobility of French workers with the Labor Force Survey, where the prospective data was being treated as error-free. The second approach, for example by Rendtel et al. (1998), is applied when no auxiliary (error-free) information is available. By relying on the assumption of the Independent Classification Errors[1], these methods use latent Markov models with measurement error. Magnac and Visser (1999) and Bassi et al. (2000) extended such methods to

[1] This assumption means that the errors made at two subsequent time periods are conditionally independent given the true states

the case where correlation between errors are possible.

Nevertheless, all the above methods are designed for short term analysis of the labor market. In other words, the impact of the business cycle on labor market transitions has been ignored. They use surveys where annual waves are available, and which include intra-annual information. In this perspective, the measurement error is approximated as a small noise, with an update each year at the time of the interview. Moreover, Magnac and Visser (1999) provide a method to correct for recall errors by assuming that transition rates are constant over time i.e. adopting a homogenous Markov chain.

In this paper, we propose a model to correct for measurement error in retrospective panels (i.e. recalled data) as per the first approach, where we have some sort of contemporaneous auxiliary information to trust and to consider as the true information. However, we design a dynamic non-stationary model that can be adopted for error correction over a long period of time and hence takes into consideration the conjuncture fluctuations over time. Building our model, we were concerned with the extent of availability of validation (i.e. error-free) auxiliary information. We therefore propose a general and flexible model that requires the least possible validating information as an input. We were also concerned by the increase over time of the magnitude of recall error as has been discussed in De Nicola and Giné (2014).

We therefore propose an n-state transition model to correct for measurement error, that could arise for example from recall and/or design bias, in retrospective panels. The technique suggested relies on the markovian structure of the labor market transitions. It shows that it is sufficient to have population moments - stocks and/or transitions (for at least one point in time) to correct over- or under-reporting biases. Using a Simulated Method of Moments, one can identify and estimate a transition- and year-specific correction matrix for the labor market flows in a biased retrospective dataset.

By using available retrospective panels and contemporaneous cross-section data from Egypt on worker flows and labor market stocks, we estimate the structural parameters of the model. Bootstrap sampling techniques are also used to demonstrate the robustness and statistical significance of our estimation results. As an application, we make use of a reform evaluation exercise to show the importance and significance of the proposed dynamic correction methodology. Through this application, we show that our methodology is crucial since we analyze a non-stationary path that is potentially disturbed by a structural break, in the case of Egypt this was an employment protection reform that took place during the corrected retrospective observation period. If one ignores the recall bias or corrects the error in a stationary manner, the conclusions obtained would be significantly different, misleading and contradicting reality.

The rest of the paper is divided as follows. In the second section, we build our correction model and demonstrate the identification strategy adopted for estimation. The third section briefly presents the data used in our analysis and the nature of the bias observed in the retrospective panels in question. This section also presents the estimation methodology using the Egyptian data and the estimation results of the model's parameters. Section 4 exposes an application of the proposed correction methodology by evaluating a structural reform, that took place during the observed retrospective period, using the corrected data. Then in section 5, we finally conclude.

2 Methodology

2.1 Transition models with measurement errors

We suppose that the true labor market histories are generated by a discrete-time Markov chain. The vector of the true labor market state occupied at year t is

$$\mathbf{X}(t) = \begin{bmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \\ \vdots \\ X_n(t) \end{bmatrix}$$

where $X_i(t)$ $i = 1, 2, .., n$ represent the true proportion of state i in the labor force at the point of time t. These are therefore the unbiased true moments of the population stocks obtained from the data. The vector

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ \vdots \\ x_n(t) \end{bmatrix}$$

denotes the observed empirical labor market state proportions at time t, with $x_i(t)$ for the states $i = 1, 2, .., n$ at the date t. These are the observed moments that decay as one goes back in time through the retrospective panel, i.e. get biased due to some sort of measurement error. With $\lambda_{ij}(t-1,t)$ being the transition rates from state i occupied in $t-1$ to the state j occupied in t, the matrix

$$\mathbf{\Lambda}(t-1,t) = \begin{bmatrix} \lambda_{11}(t-1,t) & \lambda_{12}(t-1,t) & \lambda_{13}(t-1,t) & \dots & \lambda_{1n}(t-1,t) \\ \lambda_{21}(t-1,t) & \lambda_{22}(t-1,t) & \lambda_{23}(t-1,t) & \dots & \lambda_{2n}(t-1,t) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{n1}(t-1,t) & \lambda_{n2}(t-1,t) & \lambda_{n3}(t-1,t) & \dots & \lambda_{nn}(t-1,t) \end{bmatrix} \tag{1}$$

gives the observed transition probabilities between the point in time $t-1$ and the point in time t. These are obtained by aggregating the expanded number of individuals making the transition lk from the time $t-1$ to time t in the retrospective panels and dividing by the stock of l in the year $t-1$. There exists a restriction on these transition rates: the sum of the elements of each row must be equal to one, i.e. $\sum_{j=1}^{n} \lambda_{ij}(t-1,t) = 1$, $\forall t, i$. Using (1), the dynamics of the labor market stocks is given by

$$\mathbf{x}(t) = \mathbf{\Lambda}'(t-1,t)\mathbf{x}(t-1) \tag{2}$$

where $\Lambda'(t-1,t)$ is the transposed matrix of $\Lambda(t-1,t)$. If the observed transition probabilities are biased due to measurement errors, then one would need to correct for such a bias. In our model, these errors are time and transition specific, such that the error matrix is given as follows;

$$\mathcal{E}(t-1,t) = \begin{bmatrix} \epsilon_{11}(t-1,t) & \epsilon_{12}(t-1,t) & \epsilon_{13}(t-1,t) & \dots & \epsilon_{1n}(t-1,t) \\ \epsilon_{21}(t-1,t) & \epsilon_{22}(t-1,t) & \epsilon_{23}(t-1,t) & \dots & \epsilon_{2n}(t-1,t) \\ \vdots & \vdots & \vdots & \ddots \ \vdots & \\ \epsilon_{n1}(t-1,t) & \epsilon_{n2}(t-1,t) & \epsilon_{n3}(t-1,t) & \dots & \epsilon_{nn}(t-1,t) \end{bmatrix} \tag{3}$$

with $\sum_{j=1}^{n} \epsilon_{ij}(t-1,t) = 0$ in order to ensure that $\mathbf{\Lambda(t-1,t)} + \mathcal{E}(t-1,t)$ is a markovian matrix. The matrix $\mathcal{E}(t-1,t)$ contains $n(n-1)$ unknowns. Hence, if the retrospective panel covers a period going from t to t_0, this leads to $(t-t_0)n(n-1)$ unknowns. Thus, in order to identify these errors, we define the following restrictive assumptions:

A1. The risk of error in the survey is a function of the time duration of the retrospective questionnaire $\varphi_{ij}(t)$, $\forall i,j,t$, with $\sum_{j=1}^{n} \varphi_{ij}(t) = 0$.

A2. This function is a common factor for all types of transitions whose origin is state i, regardless the destination state $j = 1,...,n$, i.e. $\varphi_{ij}(t) = \alpha_{ij}\varphi_i(t)$, $\forall i,j,t$. This implies $\sum_{j=1}^{n} \alpha_{ij}\varphi_i(t) = \varphi_i(t)\sum_{j=1}^{n} \alpha_{ij} = 0$.

A3. The function is given by $\varphi_i(t) = \varphi(t|\theta_i)$ where θ_i represent its k unknowns parameters.

With assumptions **A1-A3**, we have $n(k+n-1)$ unknown parameters. Hence, these restrictions allow to reduce the number of unknowns if $k+n-1 < (t-t0)(n-1)$. The terms $\varphi(t|\theta_i)$ vary in time. They are expected to increase as one goes back in history, showing the loss of accuracy and memory as older events are being reported. The parametric form of the decay depends however on the data in question and the number of available identifying unbiased points. The true matrix of transition probabilities between years $t-1$ and t can therefore be written as follows:

$$\begin{aligned} \mathbf{\Pi}(t-1,t) &= \mathbf{\Lambda(t-1,t)} + \mathcal{E}(t-1,t) \qquad (4) \\ &= \begin{bmatrix} \lambda_{11}(t-1,t)+\alpha_{11}\varphi_1(t) & \lambda_{12}(t-1,t)+\alpha_{12}\varphi_1(t) & \dots & \lambda_{1n}(t-1,t)-\varphi_1(t)\sum_{i=1}^{n-1}\alpha_{1i} \\ \lambda_{21}(t-1,t)+\alpha_{21}\varphi_2(t) & \lambda_{22}(t-1,t)+\alpha_{22}\varphi_2(t) & \dots & \lambda_{2n}(t-1,t)-\varphi_2(t)\sum_{i=1}^{n-1}\alpha_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{n1}(t-1,t)+\alpha_{n1}\varphi_n(t) & \lambda_{n2}(t-1,t)+\alpha_{n2}\varphi_n(t) & \dots & \lambda_{nn}(t-1,t)-\varphi_n(t)\sum_{i=1}^{n-1}\alpha_{ni} \end{bmatrix} \end{aligned}$$

2.2 Identifying strategy and estimation

By correcting the observed transition matrix $\Lambda(t-1,t)$ and obtaining a corrected one $\Pi(t-1,t)$, we deduce that the low of motion of the labor market stocks is given by

$$X(t) = \Pi'(t-1,t)X(t-1) \tag{5}$$

where $\Pi'(t-1,t)$ is the transposed matrix of $\Pi(t-1,t)$. We assume that the error terms $\varphi(t|\theta_i)$, for $\forall i$, depends only on θ_i at each time t. This parametric form has $dim(\theta_i) = k$ unknowns for each i. Hence, using (5), we deduce

$$X(t_M) = \left(\prod_{t=t_{M-1}+1}^{t_M} \Pi'(t-1,t) \right) X(t_{M-1}) = \widetilde{\Pi}_M(\Theta)X(t_{M-1})$$

where $\Theta = \{\theta_i, \{\alpha_{i,j}\}_{j=1}^{n-1}\}_{i=1}^{n}$ with $dim(\Theta) = kn + n(n-1)$. Our identifying assumptions, which are also shared by Magnac and Visser (1999), are the following

I1. The labor market stocks at the dates (t_μ for $\mu \in \{t_1, ...t_{M-1}, t_M\}$) of the interviews of the survey are perfectly observed. Hence $X(t_M)$,$X(t_m)$,..., are known for the available contemporaneous interviews.

I2. Given that $\varphi(T|\theta_i) = 0$ for only the contemporaneous points in time $T = t_M, t_m, ...$ This implies that the transitions at contemporaneous dates of the interviews of the survey are also perfectly observed: $\Pi(T-1,T) = \Lambda(T-1,T)$, for $T = t_M, t_m, ...$

Given **I1** and **I2**, we are able to estimate the vector of parameters Θ using a Simulated Method of Moments (SMM) developed by Duffie and Singleton (1993) and Gourieroux et al. (1993). We then solve the following problem

$$J = \left\{ \begin{array}{ll} \min_\Theta & [\psi_T - \psi(\Theta)]W[\psi_T - \psi(\Theta)]' \\ \text{s.c.} & 0 \leq \Gamma(\Theta) \leq 1 \end{array} \right. = \left\{ \begin{array}{ll} \min_\Theta & g(X_T,\Theta)Wg(X_T,\Theta)' \\ \text{s.c.} & 0 \leq \Gamma(\Theta) \leq 1 \end{array} \right. \tag{6}$$

where W is a weight matrix and $\Gamma(\Theta)$ is the set of non-linear constraints insuring that the estimated parameters allow to define a probability. Usually, W is deduced from the covariance matrix of ψ_T.

But, as will be shown in the presentation of the data, the non-stationarity of ψ_T lead us choose $W = Id$. We then compute the standard errors of the estimated parameters using a bootstrap method.[2] The function $g(X_T, \Theta)$ is given by

$$[\psi_T - \psi(\Theta)] \equiv g(X_T, \Theta) \quad = \quad \left\{ \begin{bmatrix} X(t_M) \\ X(t_{M-1}) \\ \vdots \\ X(t_1) \\ \Lambda(t_{M-1}-1, t_{M-1}) \\ \vdots \\ \Lambda(t_1 - 1, t_1) \end{bmatrix} - \begin{bmatrix} \widetilde{\Pi}_M(\Theta)X(t_{M-1}) \\ \widetilde{\Pi}_{M-1}(\Theta)X(t_{M-2}) \\ \vdots \\ \widetilde{\Pi}_1(\Theta)X(t_0) \\ \Pi(t_M, t_{M-1}) \\ \vdots \\ \Pi(t_2, t_1) \end{bmatrix} \right\}$$

The first M lines of $g(x_T, \Theta)$ are systems with only $(n-1)$ independent targets. The last M lines of $g(x_T, \Theta)$ are systems leading to $n(n-1)$ independent restrictions. Hence we have $M(n-1) + (M-1)n(n-1)$ restrictions for $kn + n(n-1)$ parameters. The parameters are identified iff $M(n-1) + (M-1)n(n-1) \geq kn + n(n-1)$.[3] The set of constraints is such that all the "corrected" transitions probabilities must be between 0 and 1. Hence,

$$\begin{cases} 0 \ \leq \ \lambda_{ii}(t-1,t) - \alpha_{ii}\phi(t|\theta_i) \ \leq \ 1 & \forall i, t \\ 0 \ \leq \ \lambda_{ii}(t-1,t) + \alpha_{ij}\phi(t|\theta_i) \ \leq \ 1 & \forall i, t \text{ and } j \in [1, n-1] \end{cases}$$

defines the set of non-linear restrictions $0 \leq \Gamma(\Theta) \leq 1$.

Numerical method. Given the strong non-linearities of the problem, we cannot provide a general proof for the existence and the uniqueness of the solution. As usual, we assume that this solution exists and is unique. From a numerical point of view, to deal with this problem and for the local versus global minima, we simply run several minimization algorithm, each of them being conditional to a particular initial guess for the parameters. We then compare the value of the objective functions

[2]See section 3.3 for a description of the bootstrap method.

[3]If *(i)* we have only two waves of the survey ($M = 2$), *(ii)* the number of states in the labor market is $N = 2$ and *(iii)* the number of structural parameters is $k = 2$ for each state, then the system could be just-identified. Nevertheless, the set of constraints $0 \leq \Gamma(\Theta) \leq 1$ must also be satisfied. This leads the restriction $g(\widehat{\Theta}) = 0$, usually used to find Θ in a just-identified system, to be not sufficient. Hence, even in this simple case, it is necessary to solve a minimization problem under constraints. See the applications for more examples.

(J), having rejected all the minima where the first order conditions (FOC) of the problem are not satisfied. The "solution" for Θ corresponds to the smallest value of J satisfying the FOC.

3 Application

3.1 Data and Sampling

As an application to the above proposed methodology, we rely in this section on worker flows and stocks data from the retrospective and contemporaneous information available in the Egypt Labor Market Panel Surveys fielded in 1998, 2006 and 2012. These rounds are the first, second and third waves of a periodic longitudinal survey that tracks the labor market and demographic characteristics of households and individuals interviewed in 1998 in Egypt. The households selected in the longitudinal data are national-representative and randomly selected [4].

With a gap of six to eight years between each wave, creating annual retrospective panels[5] is the only source one could assess the short-term labor market transitions in the Egyptian Economy.[6] Nevertheless, with a retrospective questionnaire covering a period of six to eight years, the retrospective information obtained from these surveys can suffer from a significant recall and design biases.[7]

We therefore have two retrospective panels 1997-2005 (from the round 2006) and 2005-2011 (from the round 2012). These provide us with information on both stocks and transitions. We also have

[4]The attrition cross-sectional and panel weights attributed in these data sets by Assaad and Krafft (2013) allow to expand sample figures to a macro population level.

[5]The survey provides detailed current employment and nonemployment information as well as labor market histories. It is therefore possible to assess employment and nonemployment transitions and spells' durations. Information on detailed individual characteristics as well as job (or firm) characteristics are also available. See Yassin (2014) and Yassin (2015)

[6]Assaad and Krafft (2016) and Wahba and Assaad (2016) are examples of recent studies that were ambitious to look at the labor market transitions in Egypt but have been all concerned with the measurement errors the data are suffering from.

[7]Recall bias is defined as respondents miss-reporting their retrospective trajectory because they tend to forget some events or spells, especially the short ones. The design bias arises from the fact that different types of questions are being asked for current versus recall/retrospective statuses. There is therefore a question of salience/cognitive recognition by the respondents where by asking the questions differently, respondents, or even sometimes the enumerators, can interpret or record them differently. Due to the design of the questionnaires of the ELMPS, statuses in the retrospective sections are sometimes interpreted more as job statuses rather than labor market states (Yassin, 2015).

contemporaneous data on stocks in 1998, 2006 and 2012.[8, 9]

The sample used in this paper includes employed, unemployed and inactive male individuals between 15 and 49 years of age. The sample includes those who ever worked, the young inexperienced new labor market entrants and the individuals who are permanently out of the labor force. Females in this context are being excluded since their behavior of entry and exit into/from the labor market is likely to be driven by personal motives such as marriage and child birth. Assumptions made in the recall correction model can therefore be distorted and might not be fully applicable if female workers are included in the analysis. Moreover, going back in time, our sample should have included people who were alive back then but passed away by the year of the survey i.e. 2006 and 2012 and hence did not respond to the ELMPS questionnaire. Due to this backward attrition, we were obliged to limit the age of our analysis group to what we refer to as the prime age group (i.e. between 15 and 49 years old).

As a robustness check, we apply our methodology using both a two-state (Employment [E] - Unemployment [U]) and a three-state (Employment [E] - Unemployment [U] - Inactivity [I]) transition models. In the former, The transition from employment to unemployment is referred to as job separation and the transition from unemployment to employment is referred to as job finding. In the latter, all inter- and intra- state transitions are illustrated and are used to calculate the job finding and separation rates of the three-state economy following Shimer (2012).

3.2 Descriptive Statistics: how big is the bias in the raw data?

We use in this section the link between worker flows and stocks data to expose and show the extent and the form of the recall and design bias in the raw Egyptian data. An easy way to reconstruct the labor market stocks given the observed and corrected transitions, is to assume that the theoretical

[8]As the surveys are fielded at the beginning of the survey year, the last year's transitions are not captured fully and are therefore omitted from the observation period. Stocks therefore capture the status at the end of the preceding year to the survey.

[9]Unfortunately, the ELMPS 1998 round did not contain what we require as "full" (compared to ELMPS06 and ELMPS12) retrospective accounts about the interviewed individuals. The type and different characteristics of an individual's first state in the labor market have not been collected. We choose to only use the cross-section stocks from this round in our analysis, for identification and comparability reasons in the correction model, given that it does not contain the minimal information required to extract a retrospective panel, comparable to the ones extracted from the 2006 and 2012 rounds. This is thought to be a good application to test the correction methodology, where one has the minimum possible number of identifying points (in other words very little auxiliary information).

steady state of unemployment for the year t is a good proxy for the empirical unemployment rate at this year. Using the job finding (f) and separation (s) rates obtained from the retrospective panels (from the ELMPS 2006 and 2012), we plot in figure 1 the theoretical steady state versus the empirical unemployment rate.

[Insert Figure 1 about here]

Clearly, the theoretical unemployment rate is correctly estimated and hence a good proxy for the contemporaneous unemployment rate in the economy (i.e. 2005 and 2011). This assumption does not however hold as one goes back in time; the gap between the empirical and theoretical unemployment rate widens.

A potential argument behind the reason of the backward increasing gap between the theoretical and empirical unemployment rates could be the declining growth rate of the working age population in Egypt. The Steady State theoretical unemployment rate assumes a population that increases at a constant growth rate. We therefore re-plot in figure 2 the steady state theoretical unemployment rate with a declining population growth rate n. Even after correcting for the population dynamics, an increasing backward bias in the theoretical unemployment curve persists confirming that the form of this gap very possibly originates from the presence of recall and design measurement errors as discussed above.

[Insert Figure 2 about here]

On the one hand, it is intuitive and very likely that when reporting their labor market histories, individuals would not recall their unemployment spells especially the short ones. On the other hand,the design of the ELMPS survey tends to under-record the unemployment and inactivity spells through the retrospective accounts. Consequently our estimations for the job separation rates over previous years are likely to be underestimated. On the other hand, people are more likely subject to over-recall and over-record their job finding transitions. This becomes clearly obvious as we overlap in figure 3 the job finding and separation rates from both panels, ELMPS06 and ELMPS12. Estimations for the job separation rates are increasingly being underestimated as we move backwards from the year of the survey, whilst job finding rates tend to be over-estimated. Even by adding the separation and job finding rates in 1998 obtained from the ELMPS 1998 retrospective panel which contains incomplete information, we still observe the same form of the bias. Assaad et al. (2016) carry out various decompositions of these overlaps of the retrospective panels of the different rounds confirming our results.

[Insert Figure 3 about here]

To sum up, three waves of the ELMPS survey are available. Each providing the true contemporaneous unbiased stocks. The ELMPS 2006 and 2012 retrospective panels provide the labor market transitions' rates over time. These rates, are the transitions moments, which decay as one goes back in time due to the recall and design bias. There exists however two unbiased moments of these for the most recent year of each panel i.e. 2004/2005 from the ELMPS 2006 and 2010/2011 from the ELMPS 2012 (as per figures 1 and 3).

3.3 Estimation and Results

Following the set-up and identification strategy of the model described in section 2, we build a two-state (see appendix A for the three-state model) transition model, to be able to estimate the measurement errors matrix $\mathcal{E}(t-1,t)$. The vectors of the true labor market state occupied and of the observed empirical labor market state proportions at year t in our two-state model becomes

$$X(t) = \begin{bmatrix} E(t) \\ U(t) \end{bmatrix} \qquad x(t) = \begin{bmatrix} e(t) \\ u(t) \end{bmatrix} \tag{7}$$

where $E(t)$ ($e(t)$) and $U(t)$ ($u(t)$) represent the true (observed) proportion of employed and unemployed respectively in the labor force in year t. The true proportions $X(t)$ represent therefore the unbiased true moments of the population stocks obtained from the data, which we have for three contemporaneous points in time (1998, 2006 and 2012). The observed moments $x(t)$ decay, i.e. get biased due to the recall and design measurement errors, as one goes back in time from the year of the survey. The dynamics of observed state are given by (2), whereas the dynamics of the true stocks are given by (5), where the matrix $\Pi(t-1,t)$ that includes the correction, is written as follows[10];

$$\Pi(t-1,t) = \begin{bmatrix} \lambda_{EE}(t-1,t)-\varphi(t|\theta_E) & 1-[\lambda_{EE}(t-1,t)-\varphi(t|\theta_E)] \\ 1-[\lambda_{UU}(t-1,t)-\varphi(t|\theta_U)] & \lambda_{UU}(t-1,t)-\varphi(t|\theta_U) \end{bmatrix} \tag{8}$$

Given the shape of the recall bias observed and discussed in the previous section in figures 1 and 3, the following parametric form is imposed on the error terms $\varphi(t|\theta_z)$, for $z = E, U$:

$$\varphi(t|\theta_z) = \nu_z(1-\exp(-\theta_z(T-t))) \tag{9}$$

implying $\varphi(T|\theta_z) = 0$. As suggested by the descriptive statistics in the previous section, the worker flows are correctly estimated for the most recent year T, we therefore assume that $\Pi(T-1,T) = \Lambda(T-1,T)$ for a given retrospective panel data set. For the 2012 round of the ELMPS, for instance, the assumption $\Pi(2010, 2011) = \Lambda(2010, 2011)$ is made and $\Pi(2004, 2005) = \Lambda(2004, 2005)$ is assumed for the ELMPS 2006 round, reflecting that the most recent year of the retrospective panel extracted from a survey is the most accurate one.

The first set of restrictions allowing to estimate the parameters $\Theta = \{\theta_E, \theta_U, \nu_E, \nu_U\}$ are

$$[\psi_T - \psi(\Theta)] \equiv g(x_T, \Theta) = \left\{ \begin{bmatrix} X(2011)_{ELMPS12} \\ X(2005)_{ELMPS06} \\ \lambda_{EE}(2004,2005)|_{2006} \\ \lambda_{UU}(2004,2005)|_{2006} \end{bmatrix} - \begin{bmatrix} \widetilde{\Pi}_1(\Theta) \\ \widetilde{\Pi}_2(\Theta) \\ \widetilde{\Pi}_3(\Theta) \\ \widetilde{\Pi}_4(\Theta) \end{bmatrix} \right\} \tag{10}$$

$$\text{where} \quad \begin{cases} \widetilde{\Pi}_1(\Theta) = \left(\prod_{t=2006}^{2011} \Pi'(t-1,t)\right) X(2005)_{ELMPS06} \\ \widetilde{\Pi}_2(\Theta) = \left(\prod_{t=1998}^{2011} \Pi'(t-1,t)\right) X(1997)_{ELMPS98} \\ \widetilde{\Pi}_3(\Theta) = \lambda_{EE}(2004,2005)|_{2012} - \nu_E(1-\exp(-\theta_E(2011-2005))) \\ \widetilde{\Pi}_4(\Theta) = \lambda_{UU}(2004,2005)|_{2012} - \nu_U(1-\exp(-\theta_U(2011-2005))) \end{cases}$$

This set of restrictions lead to 4 identifying equations. The fist two lines of $g(x_T, \Theta)$ are two systems

[10] In the two-state estimation model, estimating the weights α_{ij} would just change the levels of the estimated structural parameters ν_z for $z = E, U$, since given the set-up of the model $\alpha_{i1} = -\alpha_{i2}$, for $i = 1, 2$. We therefore choose to normalize the weights in the two-state model by imposing $\alpha_{i1} = -\alpha_{i2} = 1$, for $i = 1, 2$.

2×2, with one independent equation for each one:

$$\begin{aligned} E(2011) &= \widetilde{\pi}_{1,EE}E(2005) + (1-\widetilde{\pi}_{1,UU})(1-E(2005)) \qquad \text{where} \\ E(2005) &= \widetilde{\pi}_{2,EE}E(1997) + (1-\widetilde{\pi}_{2,UU})(1-E(1997)) \qquad \forall i = 1,2 \end{aligned} \qquad \Pi_i(\Theta) = \begin{bmatrix} \widetilde{\pi}_{i,EE} & 1-\widetilde{\pi}_{i,EE} \\ 1-\widetilde{\pi}_{i,UU} & \widetilde{\pi}_{i,UU} \end{bmatrix}$$

In order to estimate $\Theta = \{\theta_E, \theta_U, \nu_E, \nu_U\}$ allowing us to reproduce the true transition probabilities $\Pi(t-1,t)$ between the years 1999 and 2005 using the retrospective panel extracted from the ELMPS 2006 survey, we solve (6), where the set of constraints is defined by

$$0 \le \Gamma(\Theta) \le 1 \quad \Leftrightarrow \quad \begin{cases} 0 \;\le\; \lambda_{EE}(t-1,t) - \varphi(t|\theta_E) \;\le\; 1 & \forall t \\ 0 \;\le\; \lambda_{UU}(t-1,t) - \varphi(t|\theta_U) \;\le\; 1 & \forall t \end{cases}$$

We choose a weight matrix $W = Id$, and we compute the standard errors by re-estimating our correction model using 10000 bootstrap samples of the two-state employed and unemployed individuals. More precisely, following Efron (1979) and Efron and Tibshirani (1994), we rely on the assumption that our sample is representative of the population, and therefore, the empirical distribution function $\hat{F}$ is a nonparametric estimate of the population distribution F.[11] From the sample dataset, our estimated parameters, $\hat{\Theta} = \{\hat{\theta}_E, \hat{\theta}_U, \hat{\nu}_E, \hat{\nu}_U\}$, are therefore calculated as the empirical estimates of the true parameters. To measure the precision of the estimates, we compute the bootstrapped standard errors. We first draw 10000 random samples of individuals in the survey, who recall their employment status information, with replacement repeatedly from the sample dataset. For each boostrap sample, $b = 1, ..., 10000$, we are able to calculate the true and the observed labor market states, $X_b(t)$ and $x_b(t)$, as well as $\Lambda_b(t-1,t)$. This then allows the estimation of the structural parameters of our correction model corresponding to each bootstrap sample $\hat{\Theta}_b$ for $b = 1, ...10000$. This forms the sampling distribution of $\hat{\Theta}$, and allows us to calculate the sample standard deviation of the sampling distribution, which provides us with the standard errors of the estimated parameters. We also use the bootstrapped standard errors to construct approximate confidence levels for the recall error estimates.

Given the importance of taking into consideration the entry and exit of the labor force, and to

[11] The empirical distribution comes from the sampled male workers in each wave (1998, 2006 and 2012), employed and unemployed in the 2-state model, and employed, unemployed and inactive in the 3-state model. We restrict the sample to workers whose age is between 15 and 49 years old in at least one year during the recalled period.

test the robustness of our method, we extend our analysis to a three-state model of the labor market (employment, unemployment and inactivity) and check if the results on unemployment rates, reconstructed from a series of corrected labor market flows, are consistent. We show below that estimates of corrections then yield similar results, suggesting that our statistical correction method produces robust series.[12]

[Insert Table 1 about here.]

[Insert Table 2 about here.]

Tables 1 and 2 report our results and the bootstrapped standard errors. Figures 4 and 5 show the corrected trends of the separation and job finding rates. The separation rates are significantly under-estimated and the bias is larger when the individuals appeal to distant memory.[13] For the job finding rate, the transition rates are significantly over-estimated in the survey. The setting of our correction model induces large and significant corrections for both job finding and separation rates. This suggests that the use of the raw data can lead to biased estimates, whereas the corrected retrospective data allows us to calculate these transition rates in a convincing way in the sense that they are now (quasi-)stationary. This contrasts with the raw data where the memory bias generates a decreasing trend in these transitions rates. Finally, as we compare our corrected separation and job finding rates in 1999 in figure 4, to the empirical rates we obtain from ELMPS98 in 1998 in figure 1, we find that our methodology allows us to obtain a good proxy to the true level of these rates as we go backwards in time.

[Insert Figure 4 about here]

As we replot the steady-state unemployment rates using the corrected separation and job finding rates[14] for each of the two models, we obtain much more reasonable time series (see panel 4c). The

[12] We describe in the appendix A the estimation steps adopted to obtain the correction in the case of a three-state transition model.

[13] See appendix A for the corrected detailed transition rates of the 3-state model.

[14] Finding and Separation rates obtained in the three-state model are not of the same level as the rates in the two-state. This is pretty intuitive and normal since in the first model, an individual can only occupy one of two states (E or U), the transitions involved are therefore only EU and UE. In the three-state E,U,I model, the finding and separation rates take into consideration any other type of transition or state, an individual could have gone through before entering employment or exiting to unemployment. The probabilities calculated are therefore conditional on the existence of a third state in the labor market, namely inactivity and all related potential transitions.

corrected "theoretical" unemployment rate ($u = \frac{s}{s+f}$) is significantly different from the theoretical unemployment rate calculated from the raw data, but similar on average to the levels obtained for the corrected empirical unemployment rates (the element $U(t)$ of the the vector $X(t)$). We therefore manage to obtain corrected theoretical unemployment rates that share approximatively the same average of the corrected aggregate empirical unemployment rates. Nevertheless, as has been well established in the flow approach literature, the theoretical unemployment rate is more cyclical than the prevailing empirical unemployment rate, because it contains more information that one can exploit for different applications.

[Insert Figure 5 about here]

4 How important is correcting for measurement?

As previously mentioned, our technique not only allows the correction of measurement errors with little auxiliary information but also takes into consideration while doing so the conjuncture fluctuations over time. This becomes particularly important when the observed retrospective period of time has experienced some sort of structural break in the labor market flows. In this section, we exploit our uncorrected and corrected worker flows to examine via a reform evaluation application the significance of our correction. More flexible employment protection legislations have been introduced in the Egyptian labor market in 2004, via a New Labor Law.[15] The law's main objective was facilitating the hiring and firing processes in the Egyptian labor market. We therefore choose to conduct as an application an evaluation of the 2004 Labor Law that has taken place in Egypt during the observation period of our retrospective panels. By testing for a structural break in the worker flows time series, we show that correcting for the recall bias enables us, to investigate the "true" evolution of worker flows trends over the period 1997-2011 in both models; the two-state (E-U) and the three-state (E-U-I). To illustrate the interest of our approach, we propose, in a first "naive" estimation, the impact of the reform using non-corrected data. In a second step, using the corrected data, we provide a more robust analysis showing the true link between changes in the job finding and separation rates to the New Labor Law implemented in Egypt in 2004.

[15]See Langot and Yassin (2015) for a detailed description of this law.

4.1 A Naive Econometric Model

In a naive econometric scenario, the above recall error would be neglected: the data used in this "naive" approach are the non-corrected data. The job finding and separation rates are purged from their business cycle component. In order to account for the increase of the recall bias, we also introduce a linear and a quadratic trend: this is the "naive" method which allows this simple econometric model to have stationary residuals, given the shape of the non-corrected time series of separation and job finding rates. We hence use the following econometric model:

$$x_t = \beta_1 t + \beta_2 t^2 + b + \mathbb{I}_a \gamma + \epsilon_t \quad \text{for } x = f, s$$

where f_t and s_t are respectively the observed job finding and job separation rates. β_1 and β_2 are two constants representing the linear and quadratic trends (in our case the increasing slope) of the time series. b is a constant term that encompasses the "true" constant and the structural rate of worker flows (hiring or separation). We also introduce a dummy $\mathbb{I}_a = 1$ after the reform and 0 before. By running such a regression, one obtains the results reported in table 3.

[Insert Table 3 about here.]

By neglecting the recall error, or more precisely, by using a reduced form analysis which does not use the restrictions provided by the data and the stock-flow models, the law seems to have reduced the unemployment rate. Indeed, there has been a significant decrease in the separation rates and a non-significant effect on the job finding rates. In such case, rather than introducing more flexibility, via larger separation and hiring rates, the law seems to yield effects in the adverse direction. Yet, the policy makers would be relieved seeing the unemployment rates reduced (see panel(b) of Figure 9 of this unsatisfying correction of the unemployment rate). Unfortunately, the above naive scenario does not reflect not even part of the reality. By neglecting the structural interaction between the job finding and separation rates, and detrending each time series apart, one obtains misleading results: some data restrictions are not used in order to constraint the estimation. We show below the impact of the reform after correcting for the measurement error given the underlying interaction between the structural stock-flow approach of the labor market model and the data.

[Insert Figure 6 about here]

4.2 The evaluation of the reform with corrected data

The impact of the reform on flow data. We use in this section an econometric methodology[16] that extracts the cyclical component from the trends of the "corrected" labor market flows making it possible to detect a structural break observed in our time series showing the impact of the new labor law implemented in 2004. We first limit our analysis to individuals being either employed or unemployed. At first glance, figure 7 shows that the new labor law has lead to positive effects on both separation and job finding rates. Our regression results in table 4, however, reveal that only the increase in separation rates was significant at the 1% level.

[Insert Figure 7 about here]

With a very significant rise in the separations and a no significant change in the job findings, it becomes intuitive that the normal net effect of the reform is higher levels of unemployment. Our results are robust and coherent when we redo this econometric exercise using a three-state model. The reform leads to a significant increase in the separation rates and barely any impact on the job finding rates (See Figures 8 and 12).

We report in the table 4 the ols regression estimations, of both the two-state and three-state transition models, for the equations 11 and 12 (where Δy_t is used as an approximation for the difference between the observed and the potential output).

$$x_t = \alpha\Delta y_t + b + \epsilon_t \quad \text{for } x = f, s \tag{11}$$

$$x_t = \alpha\Delta y_t + b + \mathbb{I}_a\gamma + \epsilon_t \quad \text{for } x = f, s \tag{12}$$

[Insert Table 4 about here.]

[Insert Figure 8 about here]

[16] see appendix B for a full description of the estimated model

The impact of the reform on unemployment. Having shown the effects of the reform on the labor market flows (the components of unemployment), we were able to deduce that the dynamics of the separation rates has a much more dominant impact, especially after the new 2003 labor law, on the variability of the unemployment rate than the job finding rates. Given that the labor market flows are jump variables, one can use our estimation results to decompose the unemployment dynamics between each of its components.

[Insert Figure 9 about here]

To be able to verify this observation and using the estimates of equations 11 and 12, we construct counterfactual experiments. After extracting the cyclical component of the worker flows driven by the output gap, and then focussing only on the structural changes on the labor market, we can construct two time series: the first where it is assumed that the reform has no impact on the structural worker flows ($\widehat{\gamma} = 0$) and the other where the estimates of the 2003 reform are take into account ($\widehat{\gamma} \neq 0$). We therefore plot the evolution of unemployment rate after the reform assuming the separation rates have followed the same dynamics before the law. In other words, these time series assume that the separation rates remain unaffected by the reform. This scenario captures the impact of the reform on the variability of the unemployment rate if and only if the law had an impact on the Egyptian labor market's job finding rates. We reproduce the same exercise with the "naive" econometric model presented in the section 4.1.[17] The two panels of figure 9 show using the corrected data, that take into account the restrictions of the markovian processes of the workers flows, does not lead to the same predictions of a reduced form estimation. Our correction clearly significantly matters, even for a simple policy evaluation exercise.

Figures 9 and 10 show that, whether we take into consideration the existence of a third state of inactivity in the market or not, the relative contribution of the separation rates to the Egyptian unemployment dynamics is substantial and significant. The structural increase in the unemployment rates after the reform is mainly due to the increase in separation rates. The positive responses (decrease in unemployment) due to the insignificant increase in the job finding rates were definitely outweighed by the significant impact of the augmented separations (figure 9). Adding inactivity

[17] Given the non-stationarity of the uncorrected job flows data, the average of the job finding and separation rates are $x = \frac{1}{T}\sum_t(\beta_1 t + \beta_2 t^2) + b$.

as a third state in the economy, the positive impact of the job findings on the unemployment is still not significant and the largest share of unemployment variations are mainly attributed to the separations increase in this case. The Egyptian unemployment rate was therefore more responsive and had a larger elasticity vis-a-vis the variation in the level of the separation rate. Generally, we show using our corrected flows that, unlike the naive scenario, the law achieved only part of its double-sided mission, where the firing process was to some extent facilitated. Yet it has not been offset by a sufficiently increased and facilitated hiring. As a matter of fact, the law did not affect by any means the hiring process in the Egyptian labor market. In simple words, more jobs were being destructed, than before the law, while the same relative number of jobs were being created. A normal consequence would be a rise in the unemployment even if the economy has been experiencing rising rates of economic growth, which is what is observed in the case of Egypt Langot and Yassin (2015).

[Insert Figure 10 about here]

5 Concluding Remarks

This paper joins the previous literature on measurement error in pointing out the importance of the problem and the significance of correcting for it to ensure the accuracy of results and policy implications of empirical analyses. The primary objective of this paper however is to propose a handy and reliable solution to fix such problems not only (i) to avoid misleading conclusions based on erroneous data, but also (ii) to find outlets for research that has got to be based on survey data susceptible to presence of measurement errors and different types of biases.

We choose to focus on labor market transition models that rely mainly on reported labor market states at different points in time. Due to budget and technical constraints, many researchers are obliged to sacrifice the accuracy of their results if the only data available to use are retrospective panels i.e. recalled information. We therefore propose an n-state non-stationary transition model that can be used to correct for measurement error, that could arise for example from recall/and or design bias, (or any other type of bias), in retrospective panels. our model is particularly buit up for cases when little auxiliary (error-free) information is available. Being non-stationary, unlike models previously proposed by the literature, our corrected flows guarantee to retain the information on the long-term conjuncture fluctuations. This is particularly important when the correction involves a long recalled period of time. By relying on the markovian structure of the labor market transitions, we develop a model that shows that it is sufficient to have population moments - stocks and/or transitions (for at least one point in time) to correct over- or under-reporting biases. Using our identifying strategy, we show that one is able to estimate using a Simulated Method of Moments a transition- and time- specific correction matrix for biased labor market flows obtained from retrospective panels.

We estimate the structural parameters of our model using biased retrospective data from Egypt, and auxiliary unbiased contemporaneous cross-sectoinal data. We show using a reform evaluation exercise that our correction methodology provides significantly different results than when the data is uncorrected, which confirms the motivation of our paper. In other words, if the data is not corrected, one would make misleading conclusions based on significantly biased results.

References

Assaad, Ragui and Caroline Krafft, "The Egypt labor market panel survey: introducing the 2012 round," *IZA Journal of Labor & Development*, 2013, *2* (1), 8.

_ and _ , "Labor market dynamics and youth unemployment in the Middle East and North Africa: evidence from Egypt, Jordan and Tunisia," in "Economic Research Forum Working Paper Series" number 993 2016.

_ , _ , and Shaimaa Yassin, "Comparing Retrospective and Panel Data Collection Methods to Assess Labor Market Dynamics," in "Economic Research Forum Working Paper Series" number 994 2016.

Bassi, Francesca, Jacques A Hagenaars, Marcel A Croon, and Jeroen K Vermunt, "Estimating true changes when categorical panel data are affected by uncorrelated and correlated classification errors an application to unemployment data," *Sociological Methods & Research*, 2000, *29* (2), 230–268.

Duffie, Darrel and Kenneth J. Singleton, "Simulated moments estimation og markov models of asset prices," *Econometrica*, 1993, *61* (4), 929–952.

Efron, Bradley, "Computers and the theory of statistics: thinking the unthinkable," *SIAM review*, 1979, *21* (4), 460–480.

_ **and Robert J Tibshirani**, *An introduction to the bootstrap*, CRC press, 1994.

Gourieroux, Christian, Alain Monfort, and Eric Renault, "Indirect inference," *Journal of applied econometrics*, 1993, *8* (S1), S85–S118.

Langot, François and Shaimaa Yassin, "Reforming employment protection in Egypt: An evaluation based on transition models with measurement errors," in "Economic Research Forum Working Paper Series" number 918 2015.

Magnac, Thierry and Michael Visser, "Transition models with measurement errors," *Review of Economics and Statistics*, 1999, *81* (3), 466–474.

Nicola, Francesca De and Xavier Giné, "How accurate are recall data? Evidence from coastal India," *Journal of Development Economics*, 2014, *106*, 52–65.

Poterba, James M and Lawrence H Summers, "Reporting errors and labor market dynamics," *Econometrica: Journal of the Econometric Society*, 1986, pp. 1319–1338.

_ **and** _ , "Unemployment benefits and labor market transitions: A multinomial logit model with errors in classification," *The Review of Economics and Statistics*, 1995, pp. 207–216.

Rendtel, Ulrich, Rolf Langeheine, and Roland Berntsen, "The estimation of poverty dynamics using different measurements of household income," *Review of Income and Wealth*, 1998, *44* (1), 81–98.

Shimer, Robert, "Reassessing the ins and outs of unemployment," *Review of Economic Dynamics*, 2012, *15* (2), 127–148.

Wahba, Jackline and Ragui Assaad, "Flexible labor regulations and informality in egypt," *Review of Development Economics*, 2016.

Yassin, Shaimaa, "Job Accession, Separation and Mobility in the Egyptian Labor Market over the Past Decade," Technical Report, Economic Research Forum, Cairo 2014.

_ , "Labor Market Search Frictions in Developing Countries: Evidence From the MENA Region." PhD dissertation, University of Paris 1 Panthèon-Sorbonne 2015.

A Estimation of a three-state transition model using data from Egypt

To test for the robustness of our results, we also build a three-state model to correct for the labor market transitions between employment (E), unemployment (U) and inactivity (I). The vector of the true labor market state occupied at year t becomes

$$Y(t) = \begin{bmatrix} E(t) \\ U(t) \\ I(t) \end{bmatrix} \tag{13}$$

where $E(t)$, $U(t)$ and $I(t)$ represent the true proportion of employed, unemployed and inactive individuals respectively in year t (i.e. the unbiased moments of the population stocks). The vector

$$y(t) = \begin{bmatrix} e(t) \\ u(t) \\ i(t) \end{bmatrix} \tag{14}$$

denotes the observed empirical labor market state proportions at time t, with $e(t)$, $u(t)$ and $i(t)$ being the observed proportion of employed, unemployed and inactive in year t. With $\lambda_{lk}(t-1,t)$ being the transition rates from state l occupied in $t-1$ to the state k occupied in t, the matrix

$$N(t-1,t) = \begin{bmatrix} \lambda_{EE}(t-1,t) & \lambda_{EU}(t-1,t) & \lambda_{EI}(t-1,t) \\ \lambda_{UE}(t-1,t) & \lambda_{UU}(t-1,t) & \lambda_{UI}(t-1,t) \\ \lambda_{IE}(t-1,t) & \lambda_{IU}(t-1,t) & \lambda_{II}(t-1,t) \end{bmatrix} \tag{15}$$

gives the observed transition probabilities between the year $t-1$ and the year t. Once again, there exists a restriction on these transition rates: the sum of the elements of each row must be equal to one. Thus, one obtains:

$$\begin{aligned}
\lambda_{EI}(t-1,t) &= 1-\lambda_{EU}(t-1,t)-\lambda_{EE}(t-1,t) &\quad (16)\\
\lambda_{UI}(t-1,t) &= 1-\lambda_{UE}(t-1,t)-\lambda_{UU}(t-1,t) &\quad (17)\\
\lambda_{IU}(t-1,t) &= 1-\lambda_{IE}(t-1,t)-\lambda_{II}(t-1,t) &\quad (18)\\
& &\quad (19)
\end{aligned}$$

This transition matrix in equation 15 leads to

$$y(t) = N'(t-1,t)y(t-1) \tag{20}$$

As previously mentioned, the observed transition probabilities are biased due to recall or design issues. An error term $\varphi(t|\theta_z)$, for $z = E, U, I$, is therefore defined and associated to the z-type agents. These error terms vary in time and increase as one goes back in history, showing the loss of accuracy and memory as older events are being reported. The true matrix of transition probabilities between years $t-1$ and t can therefore be written as follows;

$$\Omega(t-1,t) = \begin{bmatrix} \lambda_{EE}+\alpha_{EE}\varphi(t|\theta_E) & \lambda_{EU}+\alpha_{EU}\varphi(t|\theta_E) & \lambda_{EI}-(\alpha_{EE}+\alpha_{EU})\varphi(t|\theta_E) \\ \lambda_{UE}+\alpha_{UE}\varphi(t|\theta_U) & \lambda_{UU}+\alpha_{UU}\varphi(t|\theta_U) & \lambda_{UI}-(\alpha_{UU}+\alpha_{UE})\varphi(t|\theta_U) \\ \lambda_{IE}+\alpha_{IE}\varphi(t|\theta_I) & \lambda_{IU}-(\alpha_{II}+\alpha_{IE})\varphi(t|\theta_I) & \lambda_{II}+\alpha_{II}\varphi(t|\theta_I) \end{bmatrix} \tag{21}$$

The above correction therefore allows to obtain:

$$Y(t) = \Omega'(t-1,t)Y(t-1) \tag{22}$$

where $\Omega'(t-1,t)$ is the transposed matrix of $\Omega(t-1,t)$. As in the two-state model, the parametric form on the fundamental error terms is $\varphi(t|\theta_z) = \nu_z(1-\exp(-\theta_z(T-t)))$, for $z = E, U, I$, implying $\varphi(T|\theta_z) = 0$. We show in figure 11 the corrected labor market flows in a three-state transition model i.e. employment, unemployment and inactivity using the described above specification and the estimated structural parameters in table 2

[Insert Figure 11 about here]

B Econometric Methodology to evaluate the 2004 labor market reform

A two-state labor market. In our time series, there are two components. The first one accounts for the business cycle, whereas the second accounts for long run component. Only this last part matters for our analysis. It is therefore necessary to purge the time series from their cyclical components. We extract the high frequency component of each series using the first difference of the observed output (in log): our final data are then the trend of the original time obtained after a projection on aggregate business cycle measures: we obtain a measure of the long run components of the worker flows.[18] Any policy, that changes the natural rate of the worker flows ($x^{\star}$), introduces an instability on the relation

$$\widehat{x}_t = \widehat{b} + \mathbb{I}_a\widehat{\gamma} + \widehat{\epsilon}_t \quad \text{for } x = f, s.$$

This allows us to test the impact of the 2004 reform in the Egyptian labor market. Without any observed policy change ($\widehat{\gamma} = 0$), the variations in $\widehat{x}_t$ are driven by unobservable changes in the matching and the separation processes. Remark that the time series $\widehat{x}_t^s$, built under the assumption of a stable relationship over time, can be interpreted as the counterfactual of an economy without any policy changes (this time series is build with $\widehat{\gamma} = 0$). If the policy change the natural rate of the worker flows, then the "true" series of the natural worker flows are given by $\widehat{x}_t$. The gap between $\widehat{x}_t$ and $\widehat{x}_t^s$ measures the impact of the reform.

Given that the unemployment rate is well approximated by its stationary value at the flow equilibrium, we can use our estimations of the natural flows to construct the implied natural unemployment. More formally, we have $u = \frac{s}{s+f}$. Thus, if we only focus on the component of the worker

[18]We tested the robustness of our statistical approach by using the cyclical component of the output (in log) extracted by the Hodrick-Prescott (HP) filter instead of the first difference of the output. We obtained similar results to the ones reported in the paper.

flows purged from the cyclical component linked to the GDP, we have $\widehat{u}_t = \frac{\widehat{s}_t}{\widehat{s}_t+\widehat{f}_t}$ and $\widehat{u}^s_t = \frac{\widehat{s}^s_t}{\widehat{s}^s_t+\widehat{f}^s_t}$.

Finally, in order to measure the relative contribution of the worker flows in the unemployment dynamics, one can compute $\widehat{u}^f_t = \frac{\widehat{s}_t}{\widehat{s}_t+\widehat{f}^s_t}$: this time series gives the unemployment dynamics if only the job finding rate is affected by the reform, or in other words, the contribution of the change in the job finding rate to the natural unemployment variation.

Extension: Entry and exit from the labor force. In a developing rigid labor market such as Egypt, flows to and from inactivity play an important role as a determinant of final labor market outcomes. Examining the gross flows of workers, between the three labor market states, employment (E), unemployment (U) and inactivity, becomes essential to portray as fully as possible the real story and the particular nature of the market.

In such case we adopt the same econometric methodology described above to measure the impact of the 2003 new labor law on the three-state labor market transitions. However, as mentioned previously, we now have a 3×3 matrix of the corrected transition probabilities, $\Omega(t-1,t)$. With $\Lambda_{ji}(t-1,t)$ being the corrected transition rates from state j occupied in $t-1$ to the state i occupied in t, we re-write $\Omega(t-1,t)$ as follows;

$$\Omega(t-1,t) = \begin{bmatrix} \Lambda_{EE} & \Lambda_{EU} & \Lambda_{EI} \\ \Lambda_{UE} & \Lambda_{UU} & \Lambda_{UI} \\ \Lambda_{IE} & \Lambda_{IU} & \Lambda_{II} \end{bmatrix}$$

This therefore extract the cyclical component of the workers flows using the first difference of the observed output (in log)[19], and we analyze the behavior of the "natural" rate of worker flows using the model

$$\widehat{x}_t = \widehat{b} + \mathbb{I}_a \widehat{\gamma} + \widehat{\epsilon}_t \quad \text{for } x = \Lambda_{EE}, \Lambda_{EU}, \Lambda_{EI}, \Lambda_{UE}, \Lambda_{UU}, \Lambda_{UI}, \Lambda_{IE}, \Lambda_{IU}, \Lambda_{II}.$$

This allows us to test if the policy changes the natural rate of the worker flows or not.

We then use our estimations of the natural flows to construct the implied natural unemployment.

[19]As previously, a robustness check is provided by the use of the HP filter instead of the first difference of the output.

Following Shimer (2012), in a three-state E-U-I model, the number of employed, unemployed and inactive individuals are determined by the following equations;

$$E = k(\Lambda_{UI}\Lambda_{IE} + \Lambda_{IU}\Lambda_{UE} + \Lambda_{IE}\Lambda_{UE})$$

$$U = k(\Lambda_{EI}\Lambda_{IU} + \Lambda_{IE}\Lambda_{EU} + \Lambda_{IU}\Lambda_{EU})$$

$$I = k(\Lambda_{EU}\Lambda_{UI} + \Lambda_{UE}\Lambda_{EI} + \Lambda_{UI}\Lambda_{EI})$$

where k is a constant set so that E, U and I sum to the relevant population. The steady-state unemployment rate ($u = \frac{s}{s+f}$) in a three-state labor market can therefore be written as

$$u = \frac{\Lambda_{EI}\Lambda_{IU} + \Lambda_{IE}\Lambda_{EU} + \Lambda_{IU}\Lambda_{EU}}{(\Lambda_{EI}\Lambda_{IU} + \Lambda_{IE}\Lambda_{EU} + \Lambda_{IU}\Lambda_{EU}) + (\Lambda_{UI}\Lambda_{IE} + \Lambda_{IU}\Lambda_{UE} + \Lambda_{IE}\Lambda_{UE})}$$

The relative contribution of the worker flows in the unemployment dynamics is then calculated. One can compute $\widehat{u}_t^f = \frac{\widehat{s}_t}{\widehat{s}_t + \widehat{f}_t^s}$, a time series that gives the unemployment dynamics if only job finding rate is affected by the reform given no change in the separation rates. In the three-state model (where individuals can also be inactive), the separation and job finding rates take into account all intermediate states/transitions, an individual could have gone through before exiting into unemployment or entering into employment. The hypothetical separation and job finding rates are therefore calculated as follows;

$$\widehat{s}_t = \overline{\Lambda}_{EI}\overline{\Lambda}_{IU} + \overline{\Lambda}_{IE}\overline{\Lambda}_{EU} + \overline{\Lambda}_{IU}\overline{\Lambda}_{EU}$$

$$\widehat{f}_t^s = \Lambda_{UI}\overline{\Lambda}_{IE} + \overline{\Lambda}_{IU}\Lambda_{UE} + \overline{\Lambda}_{IE}\Lambda_{UE}.$$

In other words, we show the unemployment dynamics if the three-state model separation rate followed the same dynamics as before the 2003 reform. In table 5, the three-state E-U-I ols regression estimations for the equations 23 and 24 are illustrated. Figure 12 shows the graphical results of these regressions.

$$x_t = \alpha\Delta y_t + b + \epsilon_t \quad \text{for } x = s, f, \Lambda_{EE}, \Lambda_{EU}, \Lambda_{EI}, \Lambda_{UE}, \Lambda_{UU}, \Lambda_{UI}, \Lambda_{IE}, \Lambda_{IU}, \Lambda_{II} \tag{23}$$

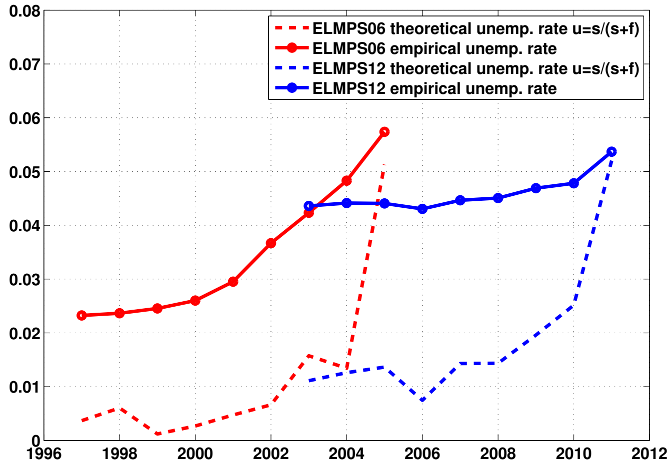
$$x_t = \alpha\Delta y_t + b + \mathbb{I}_a\gamma + \epsilon_t \quad \text{for } x = s, f, \Lambda_{EE}, \Lambda_{EU}, \Lambda_{EI}, \Lambda_{UE}, \Lambda_{UU}, \Lambda_{UI}, \Lambda_{IE}, \Lambda_{IU}, \Lambda_{II} \tag{24}$$

[Insert Table 5 about here.]

[Insert Figure 12 about here]

Figures

Figure 1: Empirical Versus Theoretical Unemployment Rate, Male Workers between 15 and 49 years of age



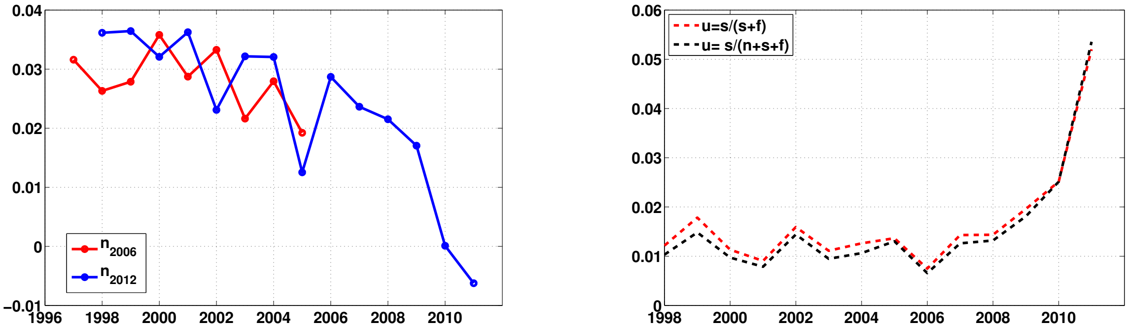
Source: Authors' own calculations using ELMPS retrospective panels extracted from rounds 2006 and 2012.
In steady-state equilibrium, flows into unemployment ("separations") equal flows from unemployment ("finds"). Using the flow balance equation, we therefore have

$$\underbrace{fU}_{\text{Probability to find a job} \times \text{no. of unemployed}} = \underbrace{sE}_{\text{Probability to quit/lose a job} \times \text{no. of employed}}$$

We can therefore show that in equilibrium, unemployment rate, with $L = U + E$, is

$$\underbrace{\frac{U}{L}}_{\text{Empirical Unemployment Rate}} = \underbrace{\frac{s}{s+f}}_{\text{Theortical Unemployment Rate}}$$

Figure 2: Steady-State unemployment rates, with a constant versus decreasing population growth rate, male workers between 15 and 49 years of age.

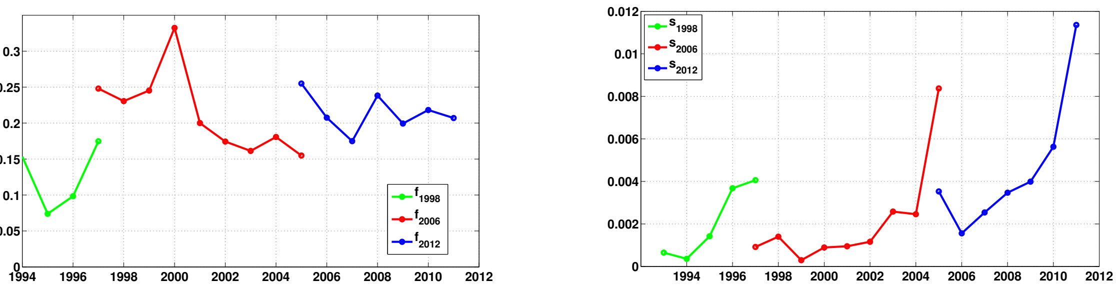


(a) Working Age Population Growth

(b) Steady-State Theoretical Unemployment Rate (ELMPS12)

Source: Authors' own calculations using ELMPS (rounds 20016 and 2012).

Figure 3: Evolution of job finding and separation rates for workers between 15 and 49 years of age over the period 1999-2011 in Egypt, using ELMPS 2006 and ELMPS2012.

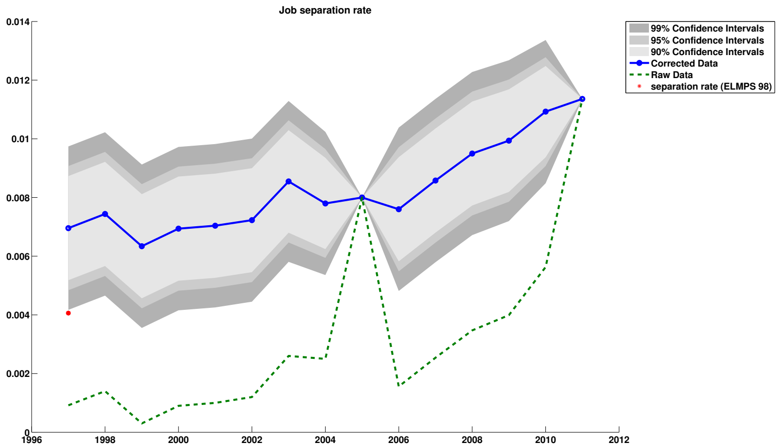


(a) Job Finding Rates

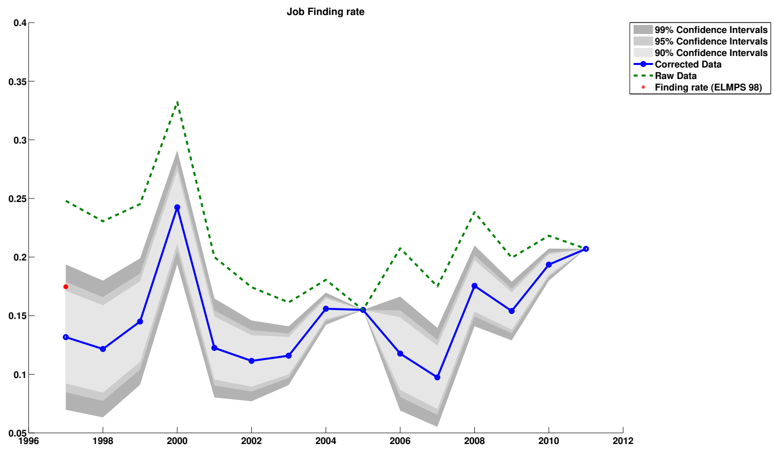
(b) Job Separation Rates

Source: Authors' own calculations using ELMPS (rounds 1998, 2006 and 2012).

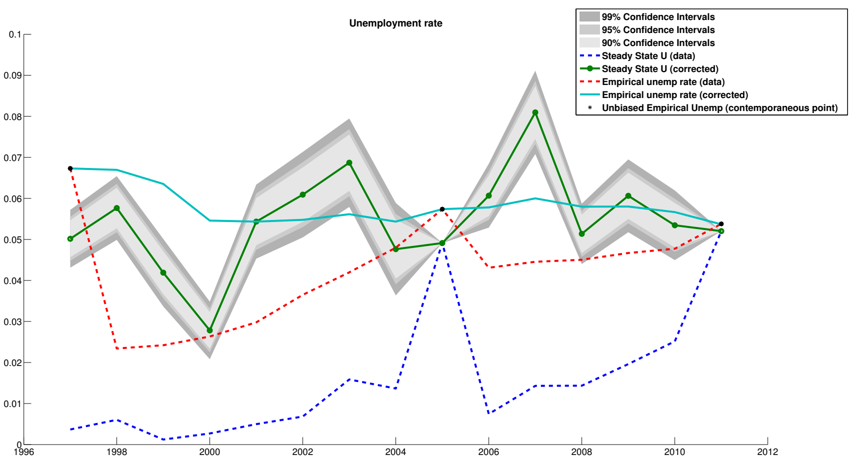
Figure 4: Corrected 2-state Job finding (f), separation (s) and unemployment rates in Egypt for male workers between 15 and 49 years of age. Bootstrapped confidence intervals are reported.



(a) Employment to unemployment separation



(b) Unemployment to employment job finding



(c) Unemployment rate

Figure 5: Corrected 3-state job finding (f) and separation (s) rates in Egypt for male workers between 15 and 49 years of age. Bootstrapped confidence intervals are reported.

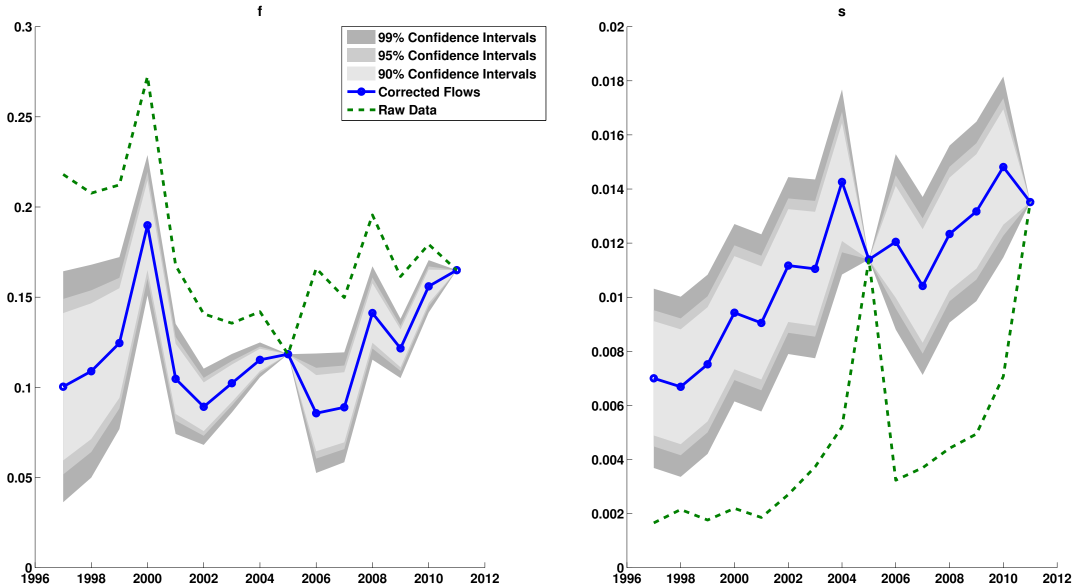
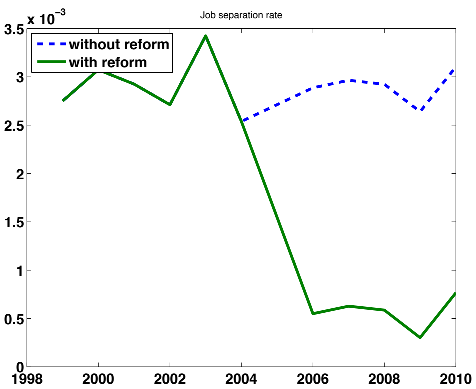
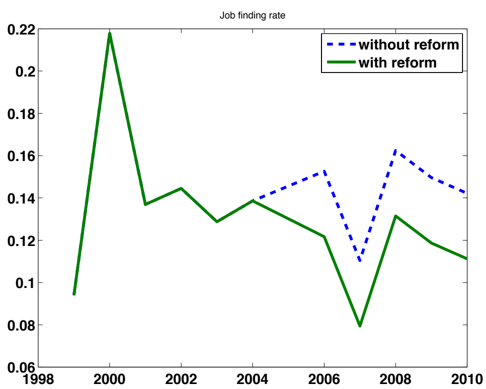


Figure 6: Job Finding and Separation Rates with and without the new labor market reform in 2004, a naive econometric model

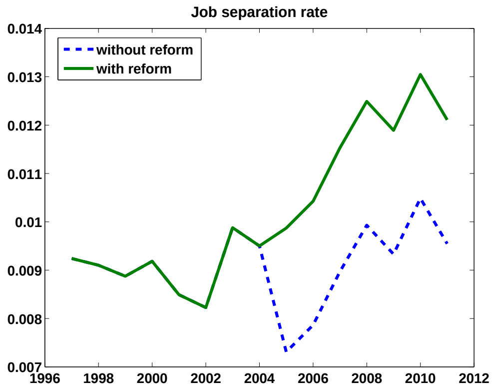


(a) Separation rate

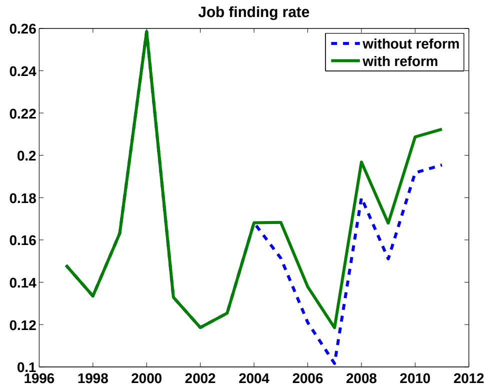


(b) Job finding rate

Figure 7: Job Finding and Separation Rates with and without the new labor market reform in 2004, a two-state E/U model

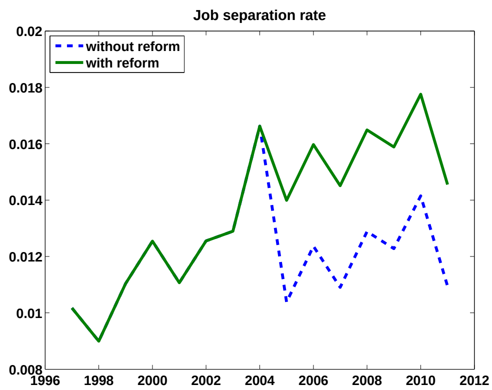


(a) E-to-U separation rate

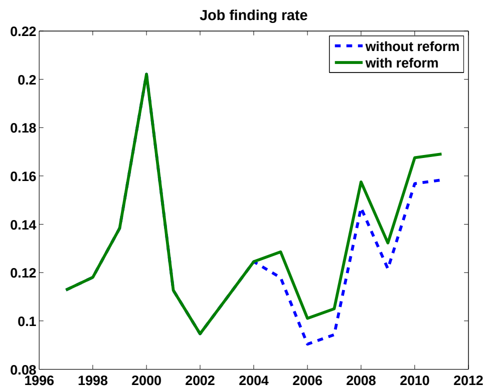


(b) U-to-E job finding rate

Figure 8: Job Finding and Separation Rates with and without the new labor market reform in 2004, a three-state E/U/I model

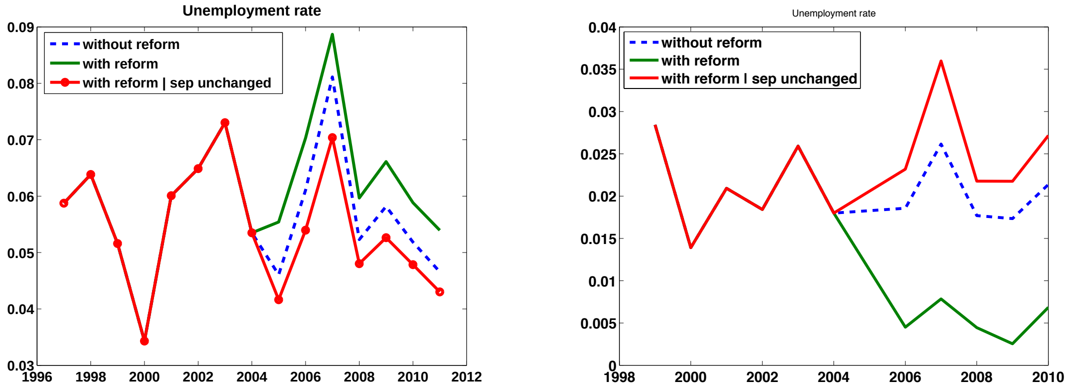


(a) Separation rate



(b) Job finding rate

Figure 9: Counterfactual evolution of unemployment rate if separation rates followed the same dynamics before the labor market reform in 2004, a two-state E-U model



(a) Unemployment rate

(b) Naive unemployment rate

Figure 10: Counterfactual evolution of unemployment rate if separation rates followed the same dynamics before the labor market reform in 2004, a three-state E-U-I model

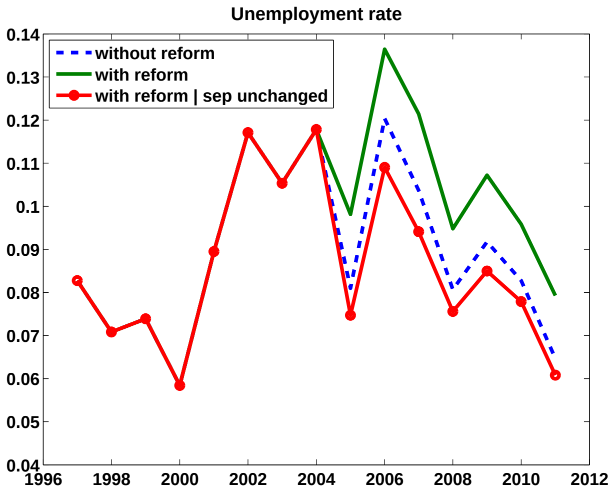


Figure 11: Corrected 3-state transitions in Egypt for male workers between 15 and 49 years of age

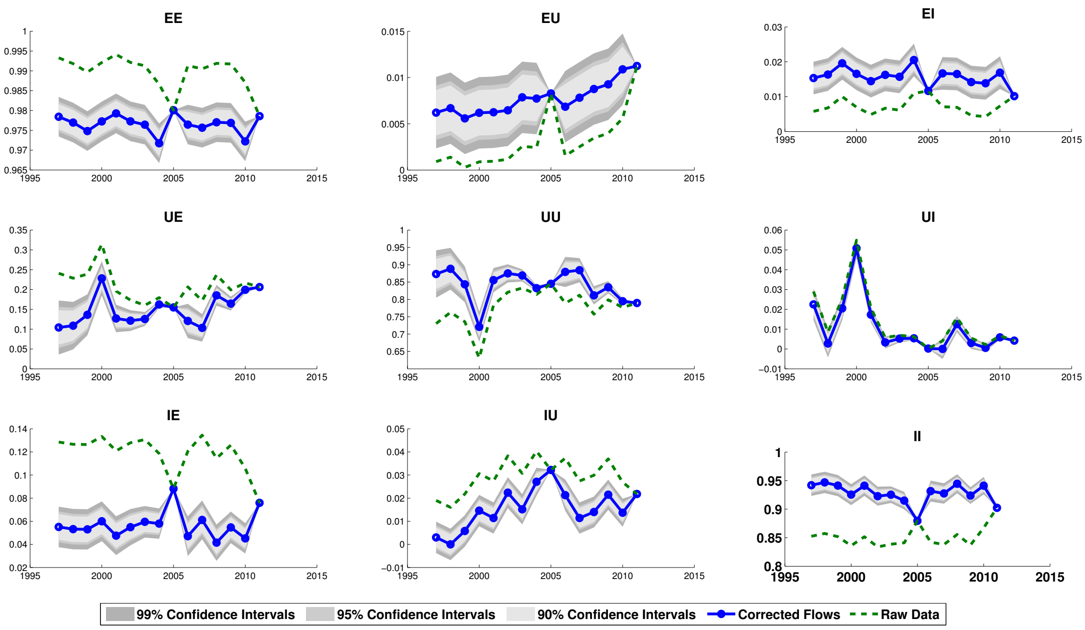
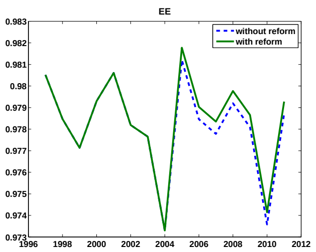
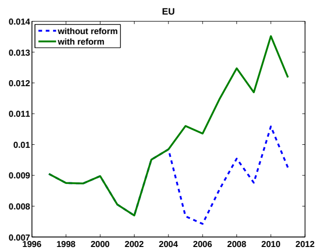


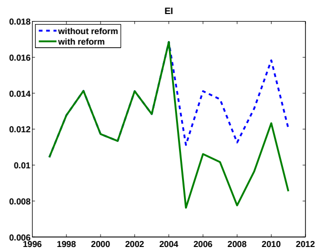
Figure 12: Trends of labor market transition rates with and without the new labor market reform in 2004, a three-state E-U-I model



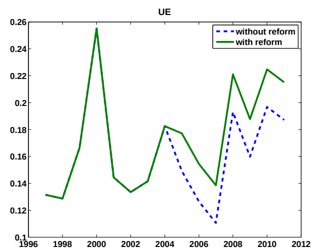
(a) Λ_{EE}



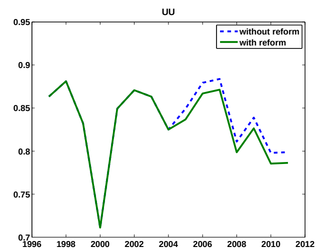
(b) Λ_{EU}



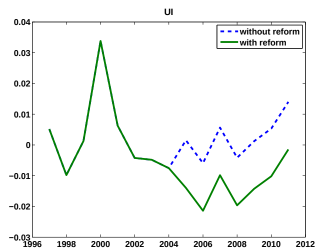
(c) Λ_{EI}



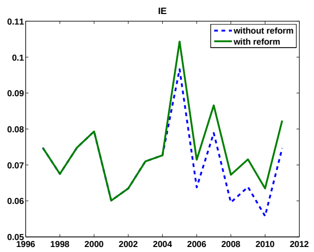
(d) Λ_{UE}



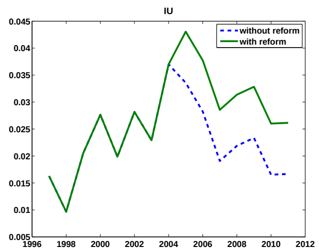
(e) Λ_{UU}



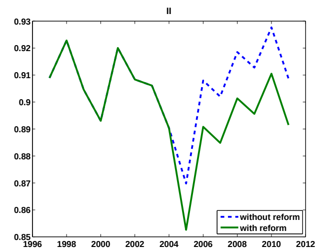
(f) Λ_{UI}



(g) Λ_{IE}



(h) Λ_{IU}



(i) Λ_{II}

Tables

Table 1: Estimation Results of the model's parameters

Two-state Model		Three-state Model			
		ν_E	0.2675	α_{EE}	0.0556
			(0.0306)		(0.0132)
ν_E	0.006	ν_U	-4.5425	α_{EU}	0.0198
	(0.0011)		(0.0028)		(0.0069)
θ_E	2.0961	ν_I	-0.1037	α_{UU}	0.9998
	(0.0140)		(0.0086)		(0.0382)
ν_U	-0.1549	θ_E	10.8866	α_{UE}	0.9543
	(0.0308)		(0.0002)		(0.0301)
θ_U	0.1735	θ_U	0.004	α_{II}	0.8623
	(0.0022)		(0.0012)		(0.0313)
		θ_I	1.7451	α_{IE}	0.7081
			(0.4014)		(0.0400)

Standard errors in parenthesis.

Table 2: Estimation Results for the fundemental error function $\varphi(t|\theta_i)$

	Two-state Model								
	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011			
$\varphi(t\|\theta_E)$	0.00604	0.00604	0.00603	0.00595	0.00530	0.00000			
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0009)	(0.0000)			
$\varphi(t\|\theta_U)$	-0.08985	-0.07751	-0.06286	-0.04542	-0.02467	0.00000			
	(0.0189)	(0.0164)	(0.0134)	(0.0097)	(0.0053)	(0.0000)			
	1996/1997	1997/1998	1998/1999	1999/2000	2000/2001	2001/2002	2002/2003	2003/2004	2004/2005
$\varphi(t\|\theta_E)$	0.00604	0.00604	0.00604	0.00604	0.00604	0.00603	0.00595	0.00530	0.00000
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0009)	(0.0000)
$\varphi(t\|\theta_U)$	-0.11625	-0.10892	-0.10021	-0.08985	-0.07752	-0.06286	-0.04542	-0.02467	0.00000
	(0.0240)	(0.0226)	(0.0209)	(0.0189)	(0.0164)	(0.0134)	(0.0097)	(0.0053)	(0.0000)
	Three-state Model								
	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011			
$\varphi(t\|\theta_E)$	0.27945	0.27945	0.27945	0.27945	0.27944	0.00000			
	(0.0306)	(0.0306)	(0.0306)	(0.0306)	(0.0305)	(0.0000)			
$\varphi(t\|\theta_U)$	-0.09001	-0.07216	-0.05423	-0.03623	-0.01815	0.00000			
	(0.0232)	(0.0187)	(0.0141)	(0.0094)	(0.0047)	(0.0000)			
$\varphi(t\|\theta_I)$	-0.12229	-0.12218	-0.12161	-0.11839	-0.10040	0.00000			
	(0.0084)	(0.0081)	(0.0075)	(0.0062)	(0.0057)	(0.0000)			
	1996/1997	1997/1998	1998/1999	1999/2000	2000/2001	2001/2002	2002/2003	2003/2004	2004/2005
$\varphi(t\|\theta_E)$	0.27945	0.27945	0.27945	0.27945	0.27945	0.27945	0.27945	0.27944	0.00000
	(0.0306)	(0.0306)	(0.0306)	(0.0306)	(0.0306)	(0.0306)	(0.0306)	(0.0305)	(0.0000)
$\varphi(t\|\theta_U)$	-0.14311	-0.12549	-0.10779	-0.09001	-0.07216	-0.05423	-0.03623	-0.01815	0.00000
	(0.0366)	(0.0322)	(0.0277)	(0.0232)	(0.0187)	(0.0141)	(0.0094)	(0.0047)	(0.0000)
$\varphi(t\|\theta_I)$	-0.12231	-0.12231	-0.12230	-0.12229	-0.12218	-0.12161	-0.11839	-0.10040	0.00000
	(0.0086)	(0.0086)	(0.0085)	(0.0084)	(0.0081)	(0.0075)	(0.0062)	(0.0057)	(0.0000)

Standard errors in parenthesis.

Table 3: OLS regression results, a naive econometric model

	f	f	s	s
β_1	-0.0360**	-0.0348	-0.000028	0.000214
β_2	0.0025*	0.0035	0.000042**	0.00044***
b	0.2534***	0.2253*	0.0017**	-0.0004
γ		-0.0310		-0.002337***

Table 4: OLS regression results

Two-state	f	f	s	s
α	-0.1240	-0.2970	-0.0153	-0.0416***
b	0.1557***	0.1560***	0.0090***	0.0091***
γ		0.0169		0.0026***
Three-state	f	f	s	s
α	-0.1167	-0.2264	-0.0208	-0.0577***
b	0.1264***	0.1266***	0.0119***	0.0120***
γ		0.0107		0.0036***

Table 5: OLS regression results, a three-state E-U-I model

	Λ_{EE}	Λ_{EE}	Λ_{EU}	Λ_{EU}	Λ_{EI}	Λ_{EI}
α	-0.0318	-0.0376	-0.0216	-0.0516***	0.0534	0.0893
b	0.9781***	0.9781***	0.0090***	0.0088***	0.0131***	0.0130**
γ		0.0006		0.0029***		-0.0035***
	Λ_{UE}	Λ_{UE}	Λ_{UU}	Λ_{UU}	Λ_{UI}	Λ_{UI}
α	-0.2110	-0.4972	0.0547	0.1823	0.1563	0.3148
b	0.1600***	0.1605***	0.8373***	0.8370***	0.0028	0.0025
γ		0.0280		-0.0125		-0.0155**
	Λ_{IE}	Λ_{IE}	Λ_{IU}	Λ_{IU}	Λ_{II}	Λ_{II}
α	-0.2798	-0.3585	-0.1443	-0.2414	0.4241	0.6000**
b	0.0703***	0.0705***	0.0226***	0.0228***	0.9071***	0.9068***
γ		0.0077		0.0095**		-0.0172*