Center for Economic Institutions

Working Paper Series

No. 2018-3

"The Dynamics of Informal Employment in Urban China"

Quheng Deng and Xinxin Ma

May, 2018

Institute of Economic Research
Hitotsubashi University
2-1 Naka, Kunitachi, Tokyo, 186-8603 JAPAN
http://cei.ier.hit-u.ac.jp/English/index.html
Tel:+81-42-580-8405/Fax:+81-42-580-8333

The Dynamics of Informal Employment in Urban China[1]

Quheng Deng

Institute of Economics, Chinese Academy of Social Sciences

Xinxin Ma

Institute of Economic Research, Hitotsubashi University

Abstract

Utilizing the rotating panel data based on the Urban Household Surveys from 2002 to 2007 conducted by National Bureau of Statistics of China , this study investigates the dynamics of informal employment in urban China. It is found that the proportion of informal employment to the total employment increases continuously from 2002 to 2007. Transition rates between the informal and formal employment status indicate the probability of persistence in the informal employment is great. To consider that there may exists spurious state dependence which may overestimate the persistence in informal employment, this study utilizes the random-effects dynamic probit models to address the unobserved heterogeneity problem, and deals with the initial condition problem and the serial correlation of transitory shocks. Based on these regression results, this study disentangles the genuine from spurious state dependence. It is found that the genuine state dependence accounts for the majority of the persistence in informal employment. Genuine state dependence patterns for various subgroups are also examined.

Keywords: dynamics, informal employment, persistence, spurious state dependence, genuine state dependence

1. Introduction

[1] This paper was written while Quheng Deng was visiting Institute of Economic Research, Hitotsubashi University in 2017. This research was supported by the Joint Usage and Research Center, Hitotsubashi University. We are grateful to Professor Kazuhiro KUMO (Institute of Economic Research, Hitotsubashi University), Professor Ichiro IWAZAKI (Institute of Economic Research, Hitotsubashi University) for their helps.

Informal employment takes up a significant fraction of labor force in developing and transition economy countries. For example, Statistics show that the proportions of informal employment to total informal employment around 15% for males and 10% for females in Russia between 2002 and 2011 (Slonimczyk and Gimpelson, 2015). Bargain and Kwenda (2011) find that the proportion of informal employment (adding together informal salaried workers and self-employment) to total employment is 46% from 2002 to 2007 in Brazil. The proportion of informal employment is 21% between 2001 and 2007 in South Africa, and 59% from 2005 to 2008 in Mexico (Bargain and Kwenda, 2011). The prevalence of informal employment in China is also pointed out in Deng (2011) , Xue et al. (2014), Ma and Deng (2016), Ma(2107). Based on the 2002 urban CHIP (China Household Income Project) survey data, Deng (2011) finds that the proportion of informal employments to total employment is 22.13% for local urban workers group. Utilizing data from the China Urban Labor Survey, mainly covering large cities, Xue et al. (2004) find the proportion of informal employment is 33.44% (2005) and 29.17% (2010) for local urban workers in 2005 and 2010, respectively.

Previous research on informal employment in China focuses on the sector selection behavior of entry into informal employment and the earnings gap between formal and informal employment sectors (e.g. Deng, 2011; Xue *et al*., 2004; Ma and Deng, 2016; Ma, 2017). To the best of our knowledge, unlike the voluminous previous studies on the dynamics and transition patterns of informal employment, the persistence of informal employment in China has not been analyzed. This study aims to fill the research gap by examining the dynamics of informal employment and the genuine state dependence of informal employment in urban China.

Investigating the dynamics of informal employment sheds light on whether a group of people have been involved with informal employment from year to year or the probability of working informally is widely spread over a bunch of people. Based on our data, the probabilities of being informal workers in the period *t* among informal and formal employment worker groups in the period *t-1* are 0.8490 (informal employment) and 0.0464(formal employment), respectively, which suggests there exists a strong persistence in informal employment. However, the transition rate directly calculated from the data is not an indicator of true or genuine state dependence as the unobserved heterogeneity and serial correlation of transitory shocks may over exaggerate the true degree of state dependence in informal employment. To obtain an accurate estimation of genuine state dependence in informal employment in urban China, this study firstly utilizes various probabilities models to analyze the informal employment. Concretely, the random-effects probit model is utilized to address the unobserved heterogeneity problem, initial conditions problem, and serial correlation of transitory shocks, and yield the most reliable estimations of the effect of informal employment in the previous period (t-1) on informal employment in the current period (t). Based on these regression results, holding the characteristic constant, this study calculates predicted probabilities conditional on the employment status in the previous period (t-1), to arrive at genuine state dependence in informal employment. It is found that the genuine state dependence in informal employment is strong in urban

China and genuine state dependence accounts for the majority of the persistence in informal employment. Moreover, the genuine state dependence patterns for various subgroups are also examined.

The remainder of this paper is structured as follows. Section 2 introduces the data, reports the descriptive statistics of variables, and presents the transition patterns of informal employment in urban China. Section 3 discusses various biases related to the estimation of persistence of informal employment and introduces extensions of random-effects probit model in an attempt to deal with these biases. Section 4 presents and compares the regression results by using various regression models. Based on the estimated results by the random-effects probit model, Section 5 calculates the genuine state dependence for the whole sample as well as for various subgroups. Section 6 summarizes the conclusions of this study.

2. Data

Data comes from the annual urban household survey (UHS) from 2002 to 2007conducted by the National Bureau of Statistics of China (NBS) The UHS provides rich information about demography, work status, income as well as consumption, which has been used extensively in previous studies (See, for example, Chi and Li, 2008; Li et al., 2012; Meng, 2012; Zhang et al, 2005; Feng et al., 2017).To keep the sample from aging, about one third of the sample is replaced annually by NBS. It follows that, theoretically, an observation in UHS could be followed for 3 consecutive years. In practice, however, the stay period for some observations may be longer than 3 years and some were replaced after being surveyed for 1 year or 2 years. [2]

The UHS covers all of the 31 provinces in mainland China, out of which we obtained a dataset for 12 provinces. These 12 provinces locate in China's three regions which differ by economic development levels and are representative for the whole of urban China. Among the 12 provinces, Beijing, Liaoning, Jiangsu, and Guangdong belong to the most prosperous eastern region; Shanxi, Anhui, Henan, and Hubei represent central region; Chongqing, Sichuan, Yunnan, and Gansu locate the economically lagged behind western region.

In this study, we focus on the dynamics of informal employment and the transition between the formal and informal employment status. Accordingly, the analyzed objects are restricted to those aged 16-60 and being employed throughout the whole analyzed period.

[2] The households in the UHS are required to keep detailed diaries about their income and consumption. Failing to provide concrete diaries with high quality due to various reasons, such as low education or high opportunity costs of keeping diary, could be one explanation for some households to have been dropped from the survey more quickly.

After dropping the observations with missing values on key variables like employment status, 30124 observations can be used in the analyses.

It is pointed out that informal employment is rather heterogeneous (Günther and Launov, 2012; Nguimkeu, 2014). In this study, the informally employment includes those working as employees and employers in individually owned small size firms as well as those working in an unstable status such as family workers in the self-employed sector. The formally employment includes those working in the state-owned enterprises (SOEs), collectively-owned enterprises(COEs), joint ventures, privately owned enterprises (POEs), and foreign owned enterprises (FOEs), etc.[3] We focus on the dynamics of informal employment and analyze the transition between formal and informal employments. Therefore we drop the unemployed samples, following Akay and Khamis (2012).

Table 1 presents the descriptive statistics of variables. It is observed that men are overrepresented in formal employment than women. Men take up 59.38% and 54.49% of formal and informal workers, respectively. Compared to those informally employed, those formally employed are 1.2 years older. There is also a significant difference in education between formal and informal workers. On average, the schooling years are nearly two years shorter for informal workers than that for formal workers. More striking difference can be found in education categories between formal and informal workers. The proportion of those with tertiary education is 37.91% for formal workers, while those with tertiary education only comprise of 10.93% informal workers. Formal and informal workers also differ in parental education. For 8.47% of formal workers, at least one of their parents has received college education. However, for informal workers, the corresponding share is just 7.32%. The differences between formal and informal workers for the ethnicity and marriage status are small. The number of kids younger than six years old for informal workers is more than that for formal workers. There are also regional proportion differences between formal and informal employment groups. Unemployment rate in the local city is a little higher for informal workers than that for formal workers. Informal workers tend to be overrepresented in the eastern region while formal workers are more likely to locate in the western region.

/* Table 1 about here */

Table 2 presents unconditional probabilities of informal employment for the whole sample as well as by demographic variables such as gender, education, age, and region. As it is observed in Table 2, the average proportion of informal employment to total employment is 0.1795 in urban China from 2002 to 2007. Figure 1 illustrates the informal employment rate increases continuously from 2002 to 2007 for both male and female group.

[3] The definition of informal employment in this paper is informal sector.

/* Figure 1 about here */

As it is shown in Table 1, women are more likely to be informally employed than men. The proportion of being informal employment is 0.1672 for men and 0.1969 for women, respectively. The probability of being informally employed decreases monotonically when the education level rises. For those with only primary school education, the probability of being engaged in the informal employment sector is 0.4535. However, for those having received college education, the probability of being informal employment is only 0.0593. The probability of informal employment also differs by age. For those aged 25-35 and 45-55, the probability of working informally is 0.2016 and 0.1818, separately. The probability of informal employment exhibits regional disparity.

/* Table 2 about here */

Table 2 also lists probabilities of informal employment conditional on the employment status in the previous period, which gives us a preliminary idea of the transition pattern between formal and informal employment. For the whole sample, the probability of being informal workers in the period *t* among those informally and formally employed in the period *t-1* is 0.8490 and 0.0464, respectively, suggesting the existence of strong persistence in informal employment.

Based on the results shown in Table 2, it is found that the transition pattern between formal and informal employment status is quite similar for men and women groups as well as the two specific age groups. However, the transition patterns between formal and informal employment are greatly different between various education groups and region groups. For those with no more than primary school education, the probability of being informal workers in the period *t* between informal and formal employments in the period *t-1* is 0.9102 and 0.1627, respectively. For those with college education, the probability of being informal employment in the period *t* between informal and formal employments in the period *t-1* is 0.6456 and 0.0177, respectively. This reflects that low education worker is more likely to be trapped in informal employment than high education worker. Even if low education worker happened to be employed formally in the previous year, they are more likely to be engaged in informal employment in the next year. Comparing the transition patterns of employment status between eastern region and western region, it is shown that to compare with the worker in the western region, the worker in the eastern region is less likely to be trapped in the informal employment but are more likely to transfer to the formal employment if employed informally in the previous year, which possibly due to higher flexibility in the labor market in the eastern region.

Table 2 tends to suggest the probability of existence of persistence in informal employment is large. However, the large possibility of persistence in informal employment may be caused by unobservable heterogeneity instead of *genuine* state dependence. In the next section, we will discuss the econometric methods used to separate genuine state dependence from unobservable heterogeneity.

3. Econometric framework

As this study focuses on the dynamics of informal employment and transition between formal and informal employments, the dependent variable is a binary variable, concretely, it is a variable expressing whether or not being informal employment in a specific year. [4] For the models utilized a binary variable as dependent variable to estimate the dynamics of transition patterns, it is thought that several estimators may be available. Concretely, the dynamic linear probability model, endogenous switching model and random effects dynamic probit model would be possible alternatives.[5] In this study, the random effects dynamic probit model is adopted to estimate the dynamics of informal employment.

Formally, the latent equation for the random effects dynamic probit model of informal employment is specified as

$$y_{it}^{*} = \gamma y_{it-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + v_{it} \tag{1}$$

$(i = 1, \dots, N; t = 2, \dots, T)$, where y_{it}^{*} denotes the latent dependent variable for being informal employment and y_{it} the observed employment status.

$$y_{it} = \begin{cases} 1, & \text{if } y_{it}^{*} \geq 0 \\ 0, & \text{if } y_{it}^{*} < 0 \end{cases} \tag{2}$$

y_{it-1} is the observed binary outcome variable in the previous period (*t-1*), $\boldsymbol{x}$ is a vector of observable explanatory variables such as age, gender, and education. $\boldsymbol{\beta}$ is the vector of coefficients to be estimated. γ indicates the existence and the degree of persistence in informal employment. v_{it} is the unobservable composite error term.

[4] For the dynamics of a dichotomous variable with more than two choices, the multinomial logit model with unobserved heterogeneity is often estimated, examples of which can be found in Gong et al. (2004).
[5] Cappellari and Jenkins (2008) give a detailed description of endogenous switching model and random effects dynamic probit model in the context of modelling the dynamics of social assistance receipt. Stewart (2007) adopts the random effects dynamic probit model to examine the dynamics of low-wage employment and uses the dynamic linear probability model as one of the robustness checks. Hyslop (1999) estimates both the random effects dynamic probit and dynamic linear probability model for the dynamics of labor force participation of married women.

3.1 Spurious state dependence

If γ is significantly positive, then being informal employment in the previous period will increase one's probability of being informal employment status in the current period, indicating the presence of state dependence. However, the state dependence may be *spurious* instead of *genuine*, if (i) individual characteristics correlated with the propensity to engage in informal employment are inadequately controlled for, or (ii) the initial conditions of the employment are not taken into account.[6] To deal with these problems and identify the genuine state dependence, we control for observable individual characteristics and unobservable heterogeneity as well as account for the correlation between the initial condition and unobserved heterogeneity.

3.2 Unobserved heterogeneity

The unobserved individual-specific heterogeneity is assumed to be time-invariant and v_{it} can be decomposed into the time-invariant heterogeneity, ε_i, and a random error, u_{it}.

$$v_{it} = \varepsilon_i + u_{it} \tag{3}$$

ε_i is treated as random and $u_{it} \sim N(0, \sigma_u^2)$. [7]

In the simplest case, ε_i is assumed to be uncorrelated with the explanatory variables, x_{it}. However, this assumption is easily to be violated. For instance, unobserved ability can lower individuals' searching costs in finding a formal job. Since unobserved ability is also positively correlated with individuals' education, failing to allow for this correlation would result in an omitted variable bias. [8] To deal with this omitted variable bias, we allow ε_i to be correlated with observable explanatory variables, following Mundlak (1978) and Chamberlain (1984).

$$\varepsilon_i = \overline{\boldsymbol{x}}_i^{\prime} a + \alpha_i \tag{4}$$

where $\alpha_i \sim iid\ N(0, \sigma_\alpha^2)$ and is orthogonal to x_{it} and u_{it} for all *i* and *t*. $\overline{\boldsymbol{x}}_{\boldsymbol{i}}$ refers to the vector of means of the time-varying explanatory variables for individual *i* over time.

Substituting (4) into (1), the latent probability of being informally employed becomes:

$$y_{it}^{*} = \gamma y_{it-1} + \boldsymbol{x}_{it}^{\prime}\boldsymbol{\beta} + \overline{\boldsymbol{x}}_i^{\prime} a + \alpha_i + u_{it} \tag{5}$$

Normalizing $\sigma_u^2 = 1$ and since u_{it} is now assumed to follow the standard normal distribution, the random effects dynamic probit model is then

[6] Spurious state dependence may also arise if a single informal employment spell overlaps between two or more consecutive periods (Flaig et al., 1993; Arulampalam et al., 2000).
[7] Autocorrelation in the u_{it} would be allowed for in Section 4.4.
[8] As education tends to decrease the probability of being informally employed, which will be shown later in the paper, the omitted variable bias turns out to be downward for the coefficient of education.

$$Pr(y_{it}|\boldsymbol{x}_{it}, y_{it-1}, \alpha_i) = \Phi\left\{(\gamma y_{it-1} + \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + \overline{\boldsymbol{x}}_i^{'}\ a + \alpha_i)(2y_{it} - 1)\right\} \quad (6)$$

It is easy to see that the composite error term, v_{it}, is correlated over time due to the presence of the time-invariant individual-specific heterogeneity, even if u_{it} is assumed to be independent and identically distributed. The correlation between the v_{it} in any two different periods is

$$\lambda = corr(v_{it}, v_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_u^2} t, s = 2, \dots, T; t \neq s \quad (7)$$

It is also clear that the pooled probit model, which fails to take into account the unobserved heterogeneity, would lead to a bias in γ. If unobserved heterogeneity is positively correlated with informal employment and the lagged informal employment is also positively correlated with informal employment in the current period, then the pooled probit model tends to overestimate the genuine state dependence.

3.3 Initial conditions

The initial conditions problem arises because the employment status in the initial period, y_{i1}, may be correlated with unobserved heterogeneity, α_i, since our data cannot observe individuals' whole employment history from the very beginning of their employment. However, even if the whole history of employment can be reflected in the data, y_{i1} can still be correlated with α_i. For instance, if one has preference on flexible work schedules provided by informal employment, he would be more likely to choose informal instead of formal employment from the start of his career. Failure to account for the correlation between the initial condition and unobserved heterogeneity tends to overstate the magnitude of state dependence.

Several methods have been proposed for dealing with the initial conditions problem. Among these methods, Heckman (1981b) suggests specifying a linearized reduced form equation for the latent variable in the initial period and modelling the joint distribution of the sequence of binary outcome variables. In contrast to Heckman (1981b), Orme (1997) and Wooldridge (2005) are much less computationally intensive. Orme (1997) converts the initial conditions problem into a more tractable sample selection problem with a two-step procedure. Instead of specifying a model for y_{i1} given $\boldsymbol{x}_i$ and α_i, Wooldridge (2005) suggests specifying a model for α_i given y_{i1} and $\boldsymbol{x}_i$. In this study, we use the Heckman (1981b) method to correct for the initial conditions problem.[9] Following Heckman (1981b), the latent variable for informal employment in the initial period is specified as

[9] The methods of Orme (1997) and Wooldridge (2005) will also be used for robustness check. Akay (2012) conducts Monte Carlo experiments to compare the finite-sample performances of the Heckman (1981b) and Wooldridge (2005) methods and finds that the Wooldridge method works well for panels longer than 5-8 periods. However, since most of the observations in our sample are not observed for longer than 5 periods, we choose not to adopt the Wooldridge (2005) approach.

$$y_{i1}^{*} = \boldsymbol{z}_{i1}^{'}\delta + \varepsilon_{i1} \tag{8}$$

where $\boldsymbol{z}_{i1}$ is the vector includes $\boldsymbol{x}_{i1}$ and exogenous instruments. ε_{i1} is correlated with α_i but uncorrelated with u_{it} for $t > 1$. Using an orthogonal projection, we can write $\varepsilon_{i1} = \theta\alpha_i + u_{i1}$. The linearized reduced form for the latent variable in the initial period is thus

$$y_{i1}^{*} = \boldsymbol{z}_{i1}^{'}\delta + \theta\alpha_i + u_{i1} \tag{9}$$

Assuming u_{it} is serially independent, the joint probability of the observed binary outcome sequence for individual *i* given α_i is

$$\Phi\left\{(\boldsymbol{z}_{i1}^{'}\delta + \theta\alpha_i)(2y_{i1} - 1)\right\}\prod_{t=2}^{T}\Phi\left\{(\gamma y_{it-1} + \boldsymbol{x}_{it}^{'}\beta + \overline{\boldsymbol{x}}_{i}^{'}\,a + \alpha_i)(2y_{it} - 1)\right\} \tag{10}$$

The likelihood function is then

$$\prod_i \int_{\alpha^*}\left[\Phi\left\{\left(z_{i1}^{'}\delta + \theta\alpha_i\right)(2y_{i1} - 1)\right\}\prod_{t=2}^{T}\Phi\left\{\left(\gamma y_{it-1} + x_{it}^{'}\beta + \overline{\boldsymbol{x}}_{i}^{'}\,a + \alpha_i\right)(2y_{it}\ 1)\right\}\right]dF(\alpha^*) \tag{11}$$

where F is the distribution function of $\alpha^* = \frac{\alpha}{\sigma_\alpha}$.

To get the likelihood, α_i has to be integrated out using either the numerical integration method or the simulation method.

3.4 Autocorrelated transitory shocks

Up to now, the transitory shock, u_{it}, is assumed to be independently distributed across time. However, transitory shocks could be serially correlated, blurring the estimate of state dependence (See, for instance, Knights et al., 2002). For instance, if transitory shocks are negatively correlated across years, the state dependence estimated would be biased dowards.

In this study, we assume that the transitory shock follows a first-order autoregressive (AR(1)) process

$$u_{it} = \rho u_{it-1} + \xi_{it} \tag{12}$$

Allowing for autocorrelation in the transitory shock substantially increases the difficulty in estimation as the Heckman estimator now has to compute T-dimensional integrals of normal densities (Stewart, 2007). Following Stewart (2007), we will use the GHK simulator to evaluate the likelihood function and Halton draws as the random number generator.

4. Estimation results of probability models

Table 3 presents estimation results using different regression techniques. Firstly, to compare the coefficients of lagged informal employment, an indicator of persistence in informal employment, in different estimations. Estimation (I) are results based on the pooled probit model, it is shown that the coefficient of lagged informal employment is 2.618, and it is statistically significant. Based on the random-effects probit model (Estimation (II), the coefficient of lagged informal employment is 2.656. Since the estimation of random-effects probit model needs normalization, we have to rescale the parameters for valid comparisons between the estimates from pooled probit model and those from random-effects probit model. The scaled coefficient of lagged informal employment is 2.560, which suggests that failing to account for unobserved heterogeneity would lead to an upward estimation bias of state dependence. Indeed, the estimate of λ is significantly positive. Estimation (III) is one estimated by the random-effects probit models to address with the initial conditions problem, the coefficient of lagged informal employment is statistically significant and the rescaled coefficient is 2.217, suggesting the correlation between the initial condition and unobserved heterogeneity tends to overstate the magnitude of state dependence of informal employment. This is also reflected in the estimate of θ, which is positive and significant. The results for the random-effects probit models which corrected both the initial conditions problem and serial correlation of transitory shocks are reported in Estimation (IV). The coefficient of lagged informal employment is reported to be 2.779. After rescaling, the coefficient turns out to be 1.807. It indicates that the negative serial correlation of transitory shocks (reflected by the negative coefficient of ρ) biases the state dependence of informal employment upwards.

/* Table 3 about here */

Table 4 reports average partial effects of the estimates of the four models. We choose to comment average partial effects of the random-effects probit models which corrected both the initial conditions problem and serial correlation of transitory shocks, which is our preferred model listed in Estimation (IV). As it can be observed from Estimation (IV), once employed informally in the previous year would increase the probability of being employed informally again in the next year by 0.5148, an indicator of strong state dependence in informal employment. However, whether this strong state dependence is genuine or it is just caused by unobserved characteristics remains unknown. We will discuss it in Table 5.

/* Table 4 about here */

Based on the results shown in Table 4, it is found that education tends to reduce the probability of being informally employed. Compared to those with primary school education or below (the reference group), the probability of taking part in informal employment by 0.0465. And senior high school and college education would reduce the probability of being informally employed by 0.0987 and 0.1558, respectively. As the average probability of informal employment in the whole sample is 0.1795, the impact of education on informal employment is rather huge in magnitude. Compared to that of those unmarried, the probability of being informally employed of married couples is 0.0951 percentage point lower. Geographic variables also affect informal employment. Being in middle and western region reduces the probability of being informally employed by 0.0468 and 0.0319 percentage point, separately.

5. Genuine state dependence

As stated before, the state dependence reflected by the raw data may not be genuine as observed and unobserved characteristics would affect the persistence of informal employment, resulting in spurious state dependence. In this study, we try to disentangle the genuine state dependence from the spurious state dependence, following Arulampalam et al. (2000). We calculate predicted probabilities of being informally employed based on our preferred model, the random-effects probit models with both the initial conditions problem and serial correlation of transitory shocks corrected for. We calculate the predicted probabilities for all persons, conditional on being informally and formally employed in the previous year, respectively. It follows that the observed and unobserved characteristics are kept constant and the difference in the two probabilities can be attributed to genuine state dependence.

The predicted probabilities of informal employment are then compared with raw probabilities of informal employment from the data. As the difference in the raw probabilities of informal employment for those formally and informally employed in the previous year contains both genuine and spurious state dependence, comparing the estimate of genuine state dependence and the difference in raw probabilities gives the magnitude of spurious state dependence.

/* Table 5 about here */

Table 5 reports the calculation results for the whole sample as well as for males and females, separately. It is indicated that the raw data probability of informal employment conditional on informally and formally employed in the previous year is 0.8490 and 0.0464, respectively. And the difference between these two probabilities is 0.8026, suggesting a high degree of state dependence in informal employment. However, the spurious state dependence of informal employment contained in it needs to be purged out. The predicted probability for the whole sample is 0.5650, assuming all of them were informally employed in the previous year. The corresponding probability is 0.0502, assuming all of them were instead formally employed in the previous year. As both observed and unobserved characteristics are kept constant, the only thing that leads to the difference between these two probabilities is the change of employment status in the previous year. It follows that the difference between these two probabilities, 0.5148, measures the degree of genuine state dependence. We also calculate the proportion of genuine state dependence in total state dependence and find genuine state dependence accounts for 64.1% of total state dependence in informal employment. Table 5 also reports results for males and females. It seems that the proportion of genuine state dependence in total state dependence is a little smaller for males than for females.

/* Table 6 about here */

Regression results suggest education plays an important role in informal employment. Table 6 examines whether the patterns of persistence in informal employment differ across different education groups. Compared to other education groups, those with primary school or below are more likely to be informally employed whether they were formally or informally employed in the previous year. The state dependence of informal employment reflected from the raw data probability is highest for those with junior high school, followed by those with senior high school and primary school or below. Those finished tertiary education has the lowest state dependence of informal employment. However, genuine state dependence of informal employment depicts a rather different pattern. According to Line 6 in Table 6, genuine state dependence is highest among those with no more than primary school education and it actually declines with education. The proportion of genuine state dependence in total state dependence also declines as education increases. It could be concluded that those better educated are less likely to be trapped in informal employment.

/* Table 7 about here */

We also investigate genuine state dependence by age groups and by regional groups. Table 7 reports the results by age group. We choose those aged 26-35 as the young group and aged 46-55 as the old group. Table 7 indicates that the state dependence in informal employment between the young and old group is rather similar. However, the genuine state dependence is a little higher for the young group than for the old group. It suggests that, compared to the young group, the old group is less likely to be trapped in informal employment. Instead, the old group may choose informal employment due to their characteristics and/or preference (for flexible working schedule, for instance).

/* Table 8 about here */

Table 8 summarizes the results for the persistence in informal employment and its decomposition by regional groups, which depicts different patterns of overall state dependence and genuine state dependence. Overall state dependence in informal employment, indicated by raw data probabilities, is highest in the middle region, followed by the western and eastern region. Nevertheless, the genuine state dependence in informal employment in the middle region is the lowest while that in the eastern region is the highest. As the labor market is more fledged in eastern region than in middle and western regions, people in eastern region would be more freely to move from formal to informal employment and back forth. One possible reason for the highest genuine state dependence in eastern region would be the better payment and prestige associated with informal employment in eastern region than in other regions. Another explanation would be people in eastern region are more open to informal employment and they may change their preference about informal employment once they were informally employed.

6. Concluding remarks

Based on the annual urban household survey (UHS) data from 2002 to 2007conducted by the National Bureau of Statistics of China (NBS), this study analyzes the dynamics of informal employment in urban China. The average share of informal employment in total employment is found to be 0.1795 in urban China from 2002 to 2007. The raw data probability of being informal workers in the period *t* among those informally and formally employed in the period *t-1* is 0.8490 and 0.0464, respectively, suggesting the existence of strong persistence in informal employment. However, spurious state dependence resulting from unobserved heterogeneity, initial conditions problem, and serial correlation of transitory shocks tend to overexaggerate the persistence in informal employment. Random-effects probit models are then estimated to deal with these problems, in an attempt to purge out the biases and arrive at reliable estimates of past informal employment on informal employment in the current period.

We calculate predicted probabilities of being informally employed based on our preferred random-effects probit model for all persons, conditional on being informally and formally employed in the previous year, respectively. It follows that the observed and unobserved characteristics are kept constant and the difference in the two probabilities can be attributed to genuine state dependence. The predicted probabilities of informal employment are then compared with raw probabilities of informal employment from the data. As the difference in the raw probabilities of informal employment for those formally and informally employed in the previous year contains both genuine and spurious state dependence, comparing the estimate of genuine state dependence and the difference in raw probabilities gives the magnitude of spurious state dependence.

Genuine state dependence is found to be 0.5148 for the whole sample and it accounts for 64.1% of total state dependence in informal employment. Given the heterogeneous nature of informal employment, this study also examines genuine state dependence for various subgroups. It is found that the proportion of genuine state dependence in total state dependence is a little smaller for males than for females. And genuine state dependence is highest among those with no more than primary school education and it actually declines with education. The proportion of genuine state dependence in total state dependence also declines as education increases. It could be concluded that those better educated are less likely to be trapped in informal employment. From the raw data probabilities, the state dependence in informal employment between the young and old group is rather similar. However, the genuine state dependence is a little higher for the young group than for the old group. Different patterns of overall and genuine state dependence for regions are also observed. Overall state dependence in informal employment, indicated by raw data probabilities, is highest in the middle region, followed by the western and eastern region. Nevertheless, the genuine state dependence in informal employment in the middle region is the lowest while that in the eastern region is the highest

References

Akay, Alpaslan and Melanie Khamis, 2012, “The Persistence of Informality: Evidence from Panel Data”, in Hartmut Lehmann and Konstantinos Tatsiramos (eds.) *Informal Employment in Emerging and Transition Economies* (Research in Labor Economics, Volume 34) Emerald Group Publishing Limited, pp.229 - 255.

Akay, Alpaslan, 2012, “Finite Sample Comparison of Alternative Methods for Estimating Dynamic Panel Data Models", *Journal of Applied Econometrics,* Vol. 27, No. 7, pp. 1189-1204.

Arulampalam, Wiji, 1999, “A Note on Estimated Coefficients in Random Effects Probit Models”, *Oxford Bulletin of Economics and Statistics*, 61(4), pp.597-602.

Arulampalam, Wiji, Alison Booth and Mark Taylor, 2000, “Unemployment Persistence”, *Oxford Economic Papers*, 52(1), pp.24-50.

Bargain, Olivier and Prudence Kwenda, 2011, “Earnings Structures, Informal Employment, and Self-Employment: New Evidence from Brazil, Mexico, and South Africa”, *Review of Income and Wealth*, Vol. 57, pp. S100-S122.

Cappellari, Lorenzo and Stephen P. Jenkins, 2008, "The Dynamics of Social Assistance Receipt: Measurement and Modelling Issues, with an Application to Britain", IZA Discussion Paper No. 3765.

Chamberlain, Gary, 1984, "Panel Data", In Zvi Griliches and Michael Intrilligator (eds.) *Handbook of Econometrics*, Vol. 2, pp 1247-1318.

Chi, Wei and Bo Li, 2008, "Glass Ceiling or Sticky Floor? Examining the Gender Earnings Differential across the Earnings Distribution in Urban China, 1987—2004", *Journal of Comparative Economics*, 36(2), 243-263.

Deng, Quheng, 2011, " Informal Employment in Urban China: Heterogeneity and Selectivity", Unpublished Manuscript.

Feng, Shuaizhang, Yingyao Hu and Robert Moffitt, 2017, "Long Run Trends in Unemployment and Labor Force Participation in Urban China", *Journal of Comparative Economics*, Volume 45, Issue 2, Pages 304-324.

Flaig, Gebhard, Georg Licht and Viktor Steiner, 1993, "Testing for State Dependence Effects in a Dynamic Model of Male Unemployment Behaviour", in Henning Bunzel, Peter Jensen, Niels Westergård-Nielsen (eds.) *Panel Data and Labour Market Dynamics*, Amsterdam: North-Holland, 189-213.

Gong, Xiaodong, Arthur van Soest and Elizabeth Villagomez, 2004, "Mobility in the Urban Labor Market: A Panel Data Analysis for Mexico", *Economic Development and Cultural Change*, Vol. 53, No. 1, pp. 1-36.

Heckman, James J., 1981a, "Heterogeneity and state dependence". In *Studies in Labor Markets*, Rosen S. (ed.).University of Chicago Press: Chicago, IL; 91–139.

Heckman, James J., 1981b, "The incidental parameters problem and the problem of initial conditions in estimating a discrete time–discrete data stochastic process", In *Structural Analysis of Discrete Data with Econometric Applications*, Manski C, McFadden D (eds). MIT Press: Cambridge, MA; 114–178.

Hyslop, Dean R., 1999, "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women", *Econometrica*, Vol. 67, No. 6, pp. 1255-1294.

Knights, Stephen, Mark Harris and Joanne Loundes, 2002, "Dynamic Relationships in the Australian Labour Market: Heterogeneity and State Dependence", *Economic Record*, Volume78, Issue242, Pages 284-298.

Li, Hongbin, Lei Li, Binzhen Wu and Yanyan Xiong, 2012, "The End of Cheap Chinese Labor", *Journal of Economic Perspectives*, 26 (4), 57-74.

Maloney, William, 1999, "Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico", *World Bank Economic Review*, Volume 13, Issue 2, Pages: 275 - 302.

Meng, Xin, 2012, "Labor Market Outcomes and Reforms in China", *Journal of Economic Perspectives*, 26 (4), 75-102.

Mundlak, Yair, 1978, "On the Pooling of Time Series and Cross Section Data", *Econometrica*, Vol. 46, No. 1, 69-85.

Nguimkeu, Pierre, 2014, "A Structural Econometric Analysis of the Informal Sector Heterogeneity", *Journal of Development Economics*, 107:175-191.

Orme, Chris D., 1997, "The Initial Conditions Problem and Two-Step Estimation in Discrete Panel Data Models", Manuscript, University of Manchester.

Slonimczyk, Fabián and Vladimir Gimpelson, 2015, "Informality and Mobility: Evidence from Russian Panel Data", *Economics of Transition*, Vol. 23, Issue 2, pp. 299-341.

Stewart, Mark and Joanna Swaffield, 1999, "Low Pay Dynamics and Transition Probabilities", *Economica*, 66 (261), 23-42.

Stewart, Mark, 2007, "The Interrelated Dynamics of Unemployment and Low-Wage Employment", *Journal of Applied Econometrics*, 22: 511-531.

Wooldridge, Jeffrey, 2005, "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity", *Journal of Applied Econometrics*, Volume 20, Issue 1, Pages 39-54.

Xue, Jinjun, Wenshu Gao and Lin Guo, 2014, "Informal Employment and Its Effect on the Income Distribution in Urban China", *China Economic Review*, 31: 84-93.

Zhang, Junsen, Yaohui Zhao, Albert Park, Xiaoqing Song, 2005, "Economic Returns to Schooling in Urban China, 1988 to 2001", *Journal of Comparative Economics*, 33 (4), 730-752.

Table 1. Descriptive statistics

	Total	Formal	Informal	Significance of difference
Male	0.5850	0.5938	0.5449	***
Age	41.28	41.49	40.29	***
Education				
Primary school and below	0.0285	0.0190	0.0721	***
Junior high school	0.2446	0.2088	0.4079	***
Senior high school	0.3962	0.3930	0.4107	**
College and above	0.3307	0.3791	0.1093	***
Schooling years	12.18	12.52	10.66	***
At least one parent has received college education	0.0826	0.0847	0.0732	***
Han	0.9581	0.9574	0.9615	
Married	0.9067	0.9080	0.9011	
No. of kids aged 0-6	0.1110	0.1093	0.1187	*
Unemployment rate in the local city	0.0762	0.0754	0.0798	***
Region				
Eastern	0.4715	0.4551	0.5462	***
Middle	0.3555	0.3650	0.3121	***
Western	0.1730	0.1798	0.1416	***
No. of observations	30124	24716	5408	

Source: Calculated based on UHS from 2002 to 2007.

Table 2. Unconditional and conditional probabilities of informal employment

	(I) Unconditional	(II) Informally employed at t-1	(III) Formally employed at t-1
Male	0.1672	0.8507	0.0430
Female	0.1969	0.8469	0.0514
Education			
Primary school and below	0.4535	0.9102	0.1627
Junior high school	0.2994	0.9075	0.0770
Senior high school	0.1861	0.8514	0.0511
College and above	0.0593	0.6456	0.0177
Age group			
Aged 26-35	0.2039	0.8419	0.0509
Aged 46-55	0.1530	0.8333	0.0398
Region			
Eastern	0.2080	0.7841	0.0733
Middle	0.1576	0.9353	0.0222
Western	0.1470	0.9175	0.0244
All	0.1795	0.8490	0.0464

Source: Calculated based on UHS from 2002 to 2007.

Table 3. Estimation results

	(I)	(II)	(III)	(IV)
	Pooled probit	RE probit	Heckman estimator	Heckman estimator
Lagged informal employment (γ)	2.618***	2.656***	2.715***	2.779***
	(0.030)	(0.036)	(0.102)	(0.143)
Male	0.004	0.003	-0.038	-0.051
	(0.029)	(0.031)	(0.057)	(0.070)
Age	0.055***	0.058***	0.029	0.039
	(0.001)	(0.018)	(0.033)	(0.041)
Age squared	-0.001***	-0.001***	-0.001	-0.001*
	(0.0002)	(0.0002)	(0.0004)	(0.0005)
Junior high school	-0.380***	-0.406***	-0.341**	-0.478**
	(0.072)	(0.076)	(0.153)	(0.188)
Senior high school	-0.652***	-0.697***	-0.700***	-0.987***
	(0.071)	(0.077)	(0.173)	(0.220)
College and above	-1.209***	-1.288***	-1.299***	-1.789***
	(0.076)	(0.086)	(0.212)	(0.284)
Han	0.023	0.025	-0.127	-0.193
	(0.075)	(0.080)	(0.127)	(0.156)
Married	-0.089	-0.096	-0.630**	-0.701**
	(0.063)	(0.066)	(0.301)	(0.349)
No. of kids aged 0-6	0.067	0.070	0.317	0.316
	(0.049)	(0.052)	(0.197)	(0.234)
Unemployment rate in the local city	0.591	0.624	0.177	0.510
	(0.369)	(0.388)	(1.620)	(1.954)
Middle	-0.197***	-0.209***	-0.348***	-0.459***
	(0.032)	(0.034)	(0.074)	(0.096)
Western	-0.226***	-0.238***	-0.238***	-0.325***
	(0.043)	(0.045)	(0.081)	(0.103)
Time-averaged variables				
Married			0.644**	0.745*
			(0.325)	(0.381)
No. of kids aged 0-6			-0.332	-0.333
			(0.224)	(0.270)
Unemployment rate in the local city			2.780	3.488
			(1.860)	(2.301)
Constant	-1.695***	-1.728***	-1.344**	-1.415*
	(0.331)	(0.348)	(0.635)	(0.781)
λ		0.071**	0.334***	0.577***
		(0.028)	(0.091)	(0.072)

θ			1.013***	0.880***
			(0.196)	(0.098)
ρ				-0.214***
				(0.053)
Log likelihood	-4661.983	-4658.785	-3840.5655	-3822.4486
Sample size	21632	21632	30124	30124

Note: Standard errors in parenthesis. ***, **, and * denotes statistical significance at 1%, 5%, and 10% levels, respectively.

Source: Calculated based on UHS from 2002 to 2007.

Table 4. Average partial effects

	(I)	(II)	(III)	(IV)
	Pooled probit	RE probit	Heckman estimator	Heckman estimator
Lagged informal employment	0.7673	0.7538	0.6421	0.5148
Male	0.0006	0.0004	-0.0050	-0.0055
Age	0.0096	0.0098	0.0038	0.0043
Age squared	-0.0001	-0.0002	-0.0001	-0.0001
Junior high school	-0.0583	-0.0602	-0.0406	-0.0465
Senior high school	-0.1050	-0.1086	-0.0854	-0.0987
College and above	-0.1698	-0.1737	-0.1389	-0.1558
Han	0.0040	0.0041	-0.0177	-0.0224
Married	-0.0162	-0.0169	-0.1073	-0.0951
No. of kids aged 0-6	0.0117	0.0118	0.0415	0.0340
Unemployment rate in the local city	0.1028	0.1050	0.0232	0.0550
Central	-0.0330	-0.0340	-0.0432	-0.0468
Western	-0.0357	-0.0363	-0.0285	-0.0319

Source: Calculated based on UHS from 2002 to 2007.

Table 5. State dependence in informal employment for the whole sample and by gender

	All	Male	Female
Raw data probability			
(1) Informally employed at t-1	0.8490	0.8507	0.8469
(2) Formally employed at t-1	0.0464	0.0430	0.0514
(3) (1)-(2)	0.8026	0.8077	0.7955
Predicted probabilities holding characteristics constant			
(4) Informally employed at t-1	0.5650	0.5457	0.5919
(5) Formally employed at t-1	0.0502	0.0453	0.0577
(6) State dependence (4)-(5)	0.5148	0.5004	0.5342
(7) Genuine state dependence as a percentage of (3)	64.15	61.95	67.15

Source: Calculated based on UHS from 2002 to 2007.

Table 6. State dependence in informal employment by education groups

	(I) Primary school and below	(II) Junior high school	(III) Senior high school	(IV) College and above
Raw data probability				
(1) Informally employed at t-1	0.9102	0.9075	0.8514	0.6456
(2) Formally employed at t-1	0.1627	0.0770	0.0511	0.0177
(3) (1)-(2)	0.7475	0.8305	0.8003	0.6279
Predicted probabilities holding characteristics constant				
(4) Informally employed at t-1	0.7723	0.7056	0.5975	0.3955
(5) Formally employed at t-1	0.1445	0.1028	0.0594	0.0191
(6) State dependence (4)-(5)	0.6278	0.6029	0.5381	0.3763
(7) Genuine state dependence as a percentage of (3)	83.98	72.59	67.24	59.93

Source: Calculated based on UHS from 2002 to 2007.

Table 7. State dependence in informal employment by age groups

	Aged 26-35	Aged 46-55
Raw data probability		
(1) Informally employed at t-1	0.8419	0.8333
(2) Formally employed at t-1	0.0509	0.0398
(3) (1)-(2)	0.7910	0.7936
Predicted probabilities holding characteristics constant		
(4) Informally employed at t-1	0.6046	0.5326
(5) Formally employed at t-1	0.0616	0.0423
(6) State dependence (4)-(5)	0.5430	0.4903
(7) Genuine state dependence as a percentage of (3)	68.65	61.79

Source: Calculated based on UHS from 2002 to 2007.

Table 8. State dependence in informal employment by regional groups

	Eastern	Middle	Western
Raw data probability			
(1) Informally employed at t-1	0.7841	0.9353	0.9175
(2) Formally employed at t-1	0.0733	0.0222	0.0244
(3) (1)-(2)	0.7108	0.9131	0.8930
Predicted probabilities holding characteristics constant			
(4) Informally employed at t-1	0.6150	0.5091	0.5407
(5) Formally employed at t-1	0.0650	0.0372	0.0441
(6) State dependence (4)-(5)	0.5500	0.4719	0.4965
(7) Genuine state dependence as a percentage of (3)	77.38	51.68	55.60

Source: Calculated based on UHS from 2002 to 2007.

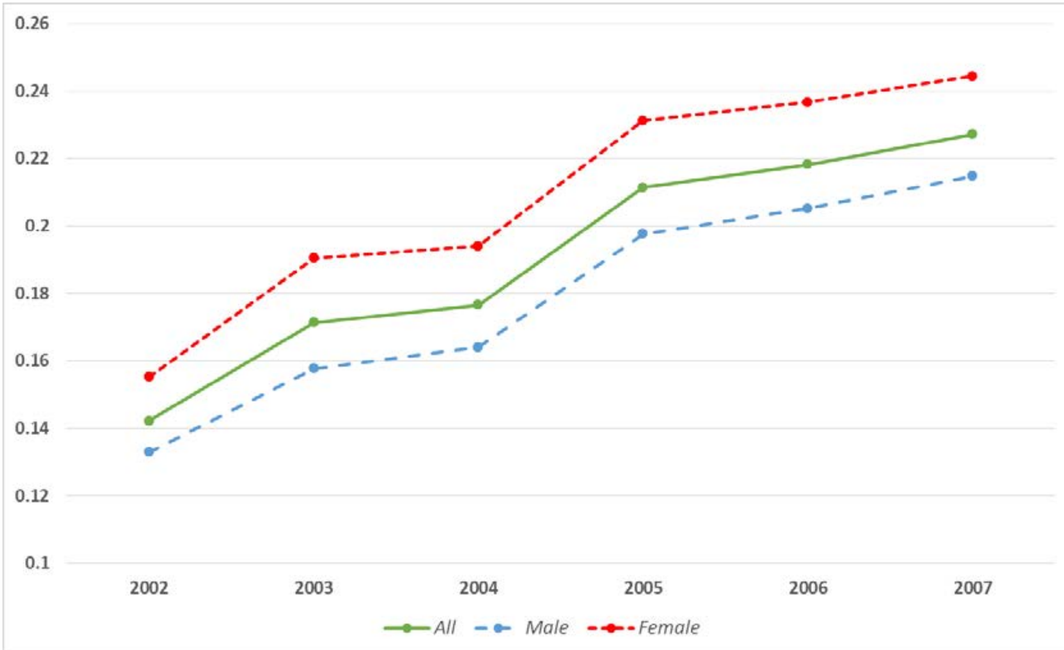

Figure 1. Proportion of informal employment

Source: Calculated based on UHS from 2002 to 2007.

Appendix Table 1. Initial conditions

	(III)	(IV)
Male	-0.001	-0.002
	(0.056)	(0.064)
Age	0.026	0.041
	(0.033)	(0.038)
Age squared	-0.001**	-0.001**
	(0.0004)	(0.0005)
Schooling years	-0.229***	-0.263***
	(0.018)	(0.021)
Han	-0.113	-0.142
	(0.128)	(0.147)
Married	0.057	0.065
	(0.134)	(0.155)
No. of kids aged 0-6	0.030	0.040
	(0.088)	(0.101)
Unemployment rate in the local city	4.098***	4.736***
	(0.721)	(0.843)
Middle	-0.441***	-0.533***
	(0.072)	(0.804)
Western	-0.349***	-0.430***
	(0.080)	(0.095)
At least one parent has received college education	-0.236**	-0.230*
	(0.119)	(0.134)
Constant	1.628	1.671**
	(0.622)	(0.710)

Note: Standard errors in parenthesis. ***, **, and * denotes statistical significance at 1%, 5%, and 10% levels, respectively.

Source: Calculated based on UHS from 2002 to 2007.