China Economic Review 37 (2016) 27–39

Contents lists available at ScienceDirect

China Economic Review

The dynamic effect of rural-to-urban migration on inequality in source villages: System GMM estimates from rural China☆

Wei HA [a], Junjian YI [b,*], Ye YUAN [b], Junsen ZHANG [c]

[a] Institute of Economics of Education and Graduate School of Education, Peking University, China
[b] Department of Economics, National University of Singapore, Singapore
[c] Department of Economics, Chinese University of Hongkong, China

ARTICLE INFO

Article history:
Received 25 February 2015
Received in revised form 10 September 2015
Accepted 10 September 2015
Available online 21 September 2015

JEL classification:
C33
D31
J61
O15

Keywords:
Inequality
Migration
System GMM

ABSTRACT

Using a newly constructed panel dataset that covers the 14-year period from 1997 to 2011 for more than 100 villages in China, this study analyzes the dynamic effect of rural-to-urban migration on inequality in source villages. Given that income inequality is time persisting, we use a system GMM framework. We found that the dynamic relationship between migration and income inequality is inversely U-shaped. Specifically, contemporary migration increases income inequality, whereas lagged migration has a strong income inequality-reducing effect on the sending villages. A 50 percent increase in the lagged migration rate translates into a one-ninth to one-tenth standard deviation reduction in income inequality.

1. Introduction

Inequality is closely and reciprocally intertwined with migration. On the one hand, income inequality between source and destination areas is widely believed to be one of the key motivating factors that drive economic migration. This effect is inherent in the Lewis model of dual economy and has been made more explicit in the Harris–Todaro model of rural–urban migration (Harris & Todaro, 1970; Lewis, 1954). Internal rural–urban migration is modeled as a response to wage disparities between the urban and rural sectors (Ray, 1998).[1] By the same token, international migration can be viewed as an outcome of global inequality (Black, Natali, & Skinner, 2006). Large and increasing wage gaps across countries are cited as an irresistible force impelling greater labor mobility across national boundaries (Pritchett, 2006). In its flagship human development report, UNDP (2009) reminded the world that although international migration often captures the news headlines, internal migration is the primary form of migration globally.[2]

☆ The research was supported by NNSFC Fund (No.71373057) and NUS Start-up Grant (WBS No.:R-122-000-211-133).
* Corresponding author at: Department of Economics, National University of Singapore, Faculty of Arts & Social Sciences, AS2 Level 6, 1 Arts Link, 117570, Singapore.
E-mail addresses: wha@pku.edu.cn (W. Ha), junjian@nus.edu.sg (J. Yi), ye.yuan@u.nus.edu (Y. Yuan), jszhang@cuhk.edu.hk (J. Zhang).

[1] Needless to say, inequality among the other dimensions of life opportunities, such as education, health, civil liberty, and other dimensions, are also drivers of migration.

[2] No sharp distinction may be found between the international migration and internal migration because both involve labor mobilization across borders. Moreover, many provinces in China have a population that is equal or larger than an ordinary European country. The scale of the cross-provincial migration in China is comparable to the cross-border migration in Europe. Thus, by reviewing the literature of international migration, we gain a better understanding of the mechanism underlying the large internal migration in China.

http://dx.doi.org/10.1016/j.chieco.2015.09.002
1043-951X/

The growth of literature on the New Economics of Labor Migration has brought about focus on the inequality within sending communities as drivers of migration. The relative position of households with respect to a specific reference group and the household's absolute income serve as strong motivations for migration (Stark & Bloom, 1985; Stark & Taylor, 1991). Conversely, labor mobility generates a feedback effect on income inequality in both sending and receiving communities. Conceptually, the effect depends on the structural factors that influence the distribution of the costs and benefits of migration and the associated selectivity of migration itself (Black et al., 2006). If the costs of migration are sizeable and the poor face binding credit constraints, which is often the case in the context of developing countries, migration will be positively selected by the poor who are trapped in nonproductive activities at the source communities. This implies that the rich will benefit the most from migration, and therefore, migration will widen rather than narrow the income gaps in the sending communities. This is further illustrated in earlier research (Adams, 1993, 1998; Lipton, 1980; Stark, Taylor, & Yitzhaki, 1988).

Recent research has increasingly recognized that although pioneer migrants may come from relatively wealthy households, and as such can afford the cost of migration and have better information on outside employment, their migration is likely to induce more migration from people in the bottom of the income distribution for two reasons (Jones, 1998). First, the increase in relative deprivation among non-migrants tends to boost their desire to migrate. Second, the establishment of migrant networks in the destination areas lowers the cost and risk of migration, which in turn facilitates more waves of migration of the poor. As a result, the initial negative effect of remittances on income equality might be dampened or even reversed (De Haas, 2010; Jones, 2013; McKenzie & Rapoport, 2007; Rapoport & Docquier, 2006; Stark et al., 1988).[3] This phenomenon is the basis for migration diffusion theory as first proposed by Stark, Taylor, and Yitzhaki (1986) and expanded by Jones (1998).

However, the usefulness of earlier literature is tempered by its cross-sectional nature and small sample size. The lack of panel data at the community level seriously limits researchers' ability to quantify the temporal dimension of migration and inequality. The alleged inequality-reducing effect of migration over time remains elusive. Moreover, this effect is primarily based on anecdotes rather than evidence. Owing to the lack of a large sample of communities, previous literature has focused mainly on the examination of the effect of migration on inequality in only a couple of communities (see Jones, 2013; McKenzie & Rapoport, 2007; and Mendola, 2012 for detailed reviews). The external validity of these studies is questionable, at best. McKenzie and Rapoport (2007) contribute to the literature by constructing cross-section data of 57 rural communities from the Mexican Migration Project (MMP) survey and short panel data of 214 rural communities in Mexico from the national demographic dynamics survey (ENADID) in 1992 and 1997. This identification strategy essentially estimates the effect of the development on migration on the change in inequality. Community-level migration rates are instrumented by the historical state-level migration rates and the U.S. labor market conditions to deal with the endogeneity of migration.

McKenzie and Rapoport (2007) find that further migration reduces inequality among communities with reasonably high initial levels of migration experience. Furthermore, migration has positive but insignificant effects on inequality in communities with smaller migration networks. Employing panel data for a sample of communities observed in 1992 and 1997, they find suggestive evidence for an inverse U-shaped relationship between migration and inequality, with migration increasing inequality at first before the subsequent migration lowers it. However, since they have observed the same communities only twice in time, they essentially use contemporary variations across communities in migration to proximate the effect of changes in migration on inequality.

The Chinese literature on internal migration is primarily concerned about the inequality between urban residents and rural–urban migrants (Cai & Wang, 2012; Lee, 2012; Lu & Wang, 2013; Messinis, 2013; Xue, Gao, & Lin, 2014; Zhu, 2016). Minimal attention has been paid to the impact of rural–urban migration on inequality in the sending communities. This study fills this gap by using the panel data of more than 100 villages in rural China over a 14-year period (1997, 2000, 2004, 2006, 2009, and 2011). This study is one of the first to examine the dynamic aspects of migration and income inequality using the dynamic panel data analysis in the Chinese context. First, we are able to construct a relatively long panel of variables of many communities with a range of different migration experiences from individual- and community-level panel surveys on both income and migration, which are ideal in studying the dynamics of migration and income inequality (McKenzie & Rapoport, 2007). In our study, the large sample size (N) is critical because we exploit both cross-section and time-series variations in the panel data. The large N provides the precondition for the asymptotic property of linear regressions to hold. Furthermore, the relatively long panel allows the examination of the dynamic aspects of migration and income inequality using a linear dynamic panel analysis.[4] Second, unlike earlier studies focusing exclusively on remittances, our data include the total labor earnings of migrants in the destination areas, which allow for capturing the general equilibrium effects.[5] Third, we

[3] Rapoport and Docquier (2006) propose a dynamic theoretical framework that goes part of the way towards reconciling the conflicting results from empirical studies and complements the view of "networks" by showing that the same predictions may be obtained with exogenous (i.e., constant) migration costs. They investigate the impact of migration on income and inequality both via the direct effect of the increase in the income of migrant households through higher wages abroad, and the indirect effects of the outbound flow of individuals on the local labor market. They do so in a way that demonstrates the importance of the pre-migration distribution of wealth in determining the impact of migration on the dynamic path and long-run levels of income and wealth inequality. They show that migration and remittances always lower wealth inequality. In contrast, although income inequality is also reduced in the long run, it may either increase or decrease in the short run, depending on the initial distribution of endowments. That is to say, the short- and the long-run effects on the income distribution may be of opposite signs and display an inverse U-shaped relationship.

[4] Recall that the panel data used by McKenzie and Rapoport (2007) only observed the same villages twice in 1992 and 1997.

[5] Earlier studies treated remittance income as an exogenous transfer, and compared Gini coefficients with and without the inclusion of remittance income, whereas more recently, remittances have been treated as potential substitute for home earnings. In addition, the observed income distribution with remittances is compared to a counterfactual scenario in which no migration takes place, but includes an imputed level of home earnings (Black et al., 2006; Zhu & Luo, 2010). The earlier approach is unrealistic in assuming that remittance-earning migrants are separate entities from their households in the rural areas and excludes their income in destination areas. However, the counterfactual model only provides limited improvement because the selection into migration is difficult, if not impossible, to model. This finding has led some researchers, after reviewing the literature, to conclude that any overarching generalization about the impact on inequality is unlikely to be robust (Black et al., 2006).

study the impact of migration on gender wage inequality within the sending communities, which is calculated from the key informant interview in the community panel survey. Finally, the massive number of rural–urban migrants in China, since its reform in 1980s, provides a unique context to test the relationship between migration and inequality at the community level. The structural barriers of integration into the urban society and the economic and psychological security offered by the home villages cause temporary migrants to maintain strong linkage with the source communities through remittances and home-returning (Murphy, 2002). Therefore, the impact of migration on the sending communities is more palpable than in other contexts. Moreover, evidence proves that selectivity for temporary migrants in particular has declined based on the 1990 and 2000 Census data (Sun & Fan, 2011). This decline in selectivity for temporary migrants provides suggestive evidence that migration, among other factors, has the potential to reduce inequalities within the sending communities in the long term.

Since income inequality is time persisting, we use the system GMM to control the lagged income inequality in estimating the effect of migration on income inequality in the sending villages. At the same time, contemporary migration is instrumented in the GMM framework because of the unobserved time-varying community shocks that correlate with migration and income inequality, and the potential reverse causality from income inequality to migration. We find an inverse U-shaped pattern between migration and income inequality in the sending communities. Specifically, contemporary migration increases income inequality, while lagged migration has strong income inequality-reducing effect in the sending villages. A 50% increase in lagged migration rate translates into a one-ninth to one-tenth standard deviation reduction in inequality. Contemporary migration has positive but statistically insignificant effects on raising the income inequality within the villages. These effects are robust to different specifications and measures of inequality. More interestingly, the estimated relationship between migration and the gender wage gap also has an inverse U-shape. Migration tends to increase the gender wage gap initially, and then tends to decrease it in the sending villages.

The remainder of the paper is structured as follows: Section 2 provides the background of the rural–urban migration in China and briefly reviews the literature on migration and inequality in China. Section 3 presents the empirical strategy, which is a linear dynamic panel data analysis. Sections 4 and 5 describe the data and report the empirical results. Section 6 briefly discusses the implications and concludes the paper.

2. Background: rural–urban migration in China

2.1. Hukou system and the rural–urban migration

Prior to the 1984 reform, the Residence Registration System (*hukou*) ties citizens to a specific location within China through residency permits (Chan & Buckingham, 2008). The reform liberalized the movement of the rural poor, but did not change the *hukou* system. Without a local *hukou*, temporary migrants are not fully entitled to social benefits (e.g., government housing), public services (e.g., urban education system), and access to jobs in destination areas. This situation is referred to as "leaving the land but not the hometown" (*li tu bu li xiang*) (Cai, 2003; Fan, 2008a, 2008b; Ha, Yi, & Zhang, forthcoming). Furthermore, the end of the food rations in 1992 greatly reduced the migration cost for agricultural *hukou* holders; it also facilitated rural-to-urban migration. In addition to the liberalization of regulations, the upsurge in the movement was driven by the rapid growth in manufacturing jobs and higher pay in the coastal areas. The volume of rural-to-urban migrants more than doubled to 23 million in 1994. Based on the 2000 population census, Cai (2003) reports that 77 million rural migrants were in urban areas during that survey year. Generally, the number of temporary migrants is larger than that of permanent migrants. In the 2000 census, for example, 74.4% of inter-county migrants were temporary migrants (Fan, 2008a, 2008b). The five-year interprovincial migration flows from 1995 to 2000 almost tripled from 12 to 32 million, most of whom were temporary migrants without *hukou*. Therefore, we use rural–urban migrants and temporary migrants interchangeably from here on.

Structural barriers to integration into urban society and the economic and psychological security offered by the home villages impel temporary migrants to maintain strong linkages with the source communities through remittances and frequent home-returning (Murphy, 2002). Rural–urban migrants remit around 200 to 250 billion RMB (approximately US$25–30 billion) back to their families in the countryside, which is more than half of the central government's budget on agricultural development (Li, Mao, & Zhang, 2008). A 2004 national representative survey conducted by the National Statistical Bureau shows that seasonal migrants account for 20% of the total rural–urban migrants. Small-scale household surveys show that migration tends to have a cyclical character with more than one-third of migrants spending at least three months a year at their home villages (Song & Bai, 2002). Therefore, the impact of migration on the sending communities is more palpable in China than in other contexts.

2.2. Declining selectivity and social network in rural–urban migration in China

The characteristics of migrants are distinct from the general population. Even though migrants are a selected group, evidence shows that selectivity for the temporary migrants in particular has declined based on the 1990 and 2000 Census data (Sun & Fan, 2011). Increasingly, rural–urban migrants are drawn from a more diverse background. They are younger, less sex-selective, and less well-educated, especially compared with permanent migrants.[6] The sex ratio of temporary migrants has dropped from 2.01 (males/females) in 1990 to 1.25 in 2000, which indicates increased female participation in migration. Over 80% of temporary migrants

[6] We exclusively focus on the rural-to-urban working-age migration in our sample.

only have junior secondary education or below. This decline in selectivity for temporary migrants provides suggestive evidence that migration in the long term may indeed reduce inequalities within the sending communities.[7]

Sun and Fan (2011) argue that this decline in selectivity was facilitated by the social networks among migrants. Existing social networks are known to overcome uncertainties and reduce the risk and cost for the new generation of international migrants (See Sun & Fan, 2011 for a comprehensive review). The rural-to-urban migration bears a close resemblance to international migration in this regard due to the *hukou* barriers. Temporary migrants have to rely on the pioneers in their social networks, friends, relatives, or fellow villagers, who have set foot or even settled in the urban areas to identify prospective destinations, gather information on job opportunities, form social groups, and ease adjustments. This observation is confirmed by several empirical studies. Zhao (2000); Meng (2000); Fan (2008b), and Chen, Jin, and Yue (2010) show that social networks increase the propensity to migrate to urban areas. Zhang and Zhao (2015) analyze the role of social networks in promoting the self-employment of rural-to-urban migrants in Chinese cities, while taking into account the endogenous formation of migrant networks in destination areas. Long, Appleton, and Song (2013) show that the influence of social networks is not only limited to migration decisions but also on wages of migrants. Giulietti, Wahba, and Zenou (2014) enrich the literature by examining the role played by different types of social networks, namely weak and strong ties, on migration decision and find support for both. Furthermore, weak and strong ties complement each other in facilitating migration decision. As social networks strengthen in the urban areas, one would expect that the prospective migrants with less charming characteristics could find it easier to settle in urban areas.

2.3. Impact of rural–urban migration on sending communities

The overall impact of internal migration on the migrants themselves and their families in the rural areas are generally believed to be positive. Migration is shown to improve consumption and income levels and reduce poverty among migrant households (Du, Park, & Wang, 2005; Zhu & Luo, 2010). Most studies find positive effects on education, along with some negative effects related to rigid policies (de Brauw & Giles, 2008b). These studies tend to take a micro perspective by focusing on the effect of migration at the household level. The importance of migration on the transformation of the rural areas has been largely overlooked by the existing literature (Murphy, 2002).

The conventional wisdom is that migration contributes to higher inequality in the places of origin because of positive selection and the unequal nature of non-farm income (Zhu & Luo, 2010). However, some studies suggest the opposite. Following Barham and Boucher (1998); Zhu and Luo (2010) simulate the counterfactual income distribution in the absence of migration and remittance and find that migration reduces the Gini coefficient by 16.7%. Moreover, de Brauw and Giles (2008a) suggest that migration from the village leads to increasing per capita income and consumption, and that migrant opportunity is contributing to a more rapid economic growth among poorer households in the village. Consistent with this result, ethnographic studies show that migration promotes equality in the natal communities by obtaining resources outside the power-based distributional mechanisms of the village. Thus, migrants make the boundaries of stratification more fluid (Benjamin, Brandt, & Giles, 2005; Murphy, 2000).

3. Empirical strategy

This section specifies the system GMM model to test the effect of rural-to-urban migration on income inequality in the sending communities in China. We use panel data at the village level. Specifically, we follow Blundell and Bond (1998) and estimate the following regression equation:

$$Ineq_{i,t} = \beta_0 + \beta_1 Ineq_{i,t-1} + \beta_2 Emig_{i,t} + \beta_3 Emig_{i,t-1} + X_{i,t}\beta_4 + \eta_i + v_{it} \tag{1}$$

where i and t index village and time period ($i=1,\ldots,I$ and $t=1,\ldots,T$), respectively; *Ineq* measures inequality, such as the Gini coefficient, Theil index, or gender wage ratio; *Emig* is the share of migrants out of the total labors whose *hukou* is in the village while working outside; X is a vector of the other control variables; η is community fixed effect; v is an error term; and $\beta_j (j=0,\ldots,4)$ are the coefficients to be estimated.

We include lagged inequality in the regression equation because aggregate variables, such as GDP and inequality, are time persisting. Hence, these variables are serially correlated over time. Therefore, we estimate a linear dynamic panel data model. In addition, we include both current migration and lagged migration in Eq. (1) because current migration and lagged migration may have different effects on inequality. The current migration reflects the immediate effect, whereas the lagged migration reflects the accumulated effects of migration with a build-up of migration-specific human capital and networks.

There are two potential pitfalls in Eq. (1), which will bias the ordinary least square (OLS) estimates. First, the unobserved community heterogeneities (η) may be correlated with the other independent variables in the right hand side of this equation, which creates an omitted variable bias problem. Second, the "relative deprivation" model (Stark & Taylor, 1991; Stark & Bloom, 1985) states that the relative position of households with respect to a specific reference group serves as strong motivation for migration. Therefore, causality may also go from inequality to migration ($Emig_{i,t}$).

[7] The earliest wave through which our data can identify the migrants is 1997. From 1997 to 2011, the migrants have become younger and less sex-selective. However, the education level of migrants, especially the percentage of middle school-educated and above, has increased sharply in recent years. The uprising trend of migrant education attainment from the 2000s does not contradict Sun and Fan (2011) because the census data cover the period from 1990 to 2000.

In discussing the problem of community heterogeneities, we assume that $Emig_{i,t}$ is exogenous for the moment. Unlike the static model, the fixed effects method could not eliminate the inconsistency induced by the community heterogeneities in the dynamic model of Eq. (1), because $v_{i,t}$ will be correlated with the future value of the regressors due to the presence of the lagged dependent variable in the right hand of the regression equation. In other words:

$$E\left[\left(Ineq_{i,t-1}-\overline{Ineq_{i,t-1}}\right)\left(v_{i,t}-\overline{v_{i,t}}\right)\right]\neq 0$$

where $\overline{Ineq_{i,t-1}}$ and $\overline{v_{i,t}}$ are the within the group mean values of $Ineq_{i,t-1}$ and $v_{i,t}$, respectively.

Arellano and Bond (1991) develop a difference GMM to deal with this kind of community heterogeneities. To illustrate this method clearly, we make a first difference with respect to Eq. (1),

$$\Delta \text{Ineq}_{i,t} = \beta_1 \Delta \text{Ineq}_{i,t-1} + \beta_2 \Delta \text{Emig}_{i,t} + \beta_3 \Delta \text{Emig}_{i,t-1} + \beta_4 \Delta X_{i,t} + \Delta v_{it} \tag{2}$$

where Δ is the operator of the first difference. $\Delta \text{Ineq}_{i,t-1}$ is endogenous because

$$E\left(\Delta \text{Ineq}_{i,t-1}\Delta v_{i,t}\right) = E\left[\left(Ineq_{i,t-1}-Ineq_{i,t-2}\right)\left(v_{i,t}-v_{i,t-1}\right)\right] = -E\left(Ineq_{i,t-1}v_{i,t-1}\right)\neq 0.$$

We assume that (1) error terms ($v_{i,t}$) are serially uncorrelated,[8] i.e., $\text{cov}(v_{i,t},v_{i,t-s})=0$ if $s\neq 0$; (2) initial condition, $E(Ineq_{i,1}v_{i,t})=0$ for $t\geq 2$; and (3)$E(\eta_i v_{i,t})=0$.[9] Under these three assumptions, we can derive the moment conditions for the difference GMM method as follows:

$$E\left[Ineq_{i,t-s}\Delta v_{i,t}\right] = 0, \text{ when } t = 3, \ldots, T \text{ and } s \geq 2. \tag{3}$$

Thus, all $Ineq_{i,t-s}$ for $s\geq 2$ are valid instruments for $\Delta \text{Ineq}_{i,t-1}$ in Eq. (2).

Subsequently, we use a similar method to deal with the second problem, that is, the endogeneity of $Emig_{i,t}$ and $Emig_{i,t-1}$ in Eq. (1). Suspecting that simultaneity and feedback exist between migration and inequality in the sending communities, we assume $E(Emig_{i,t}v_{i,s})\neq 0$ for $s\leq t$ and $E(Emig_{i,t}v_{i,s})=0$ for $s>t$.[10] Under this assumption, the moment condition is:

$$E\left[Emig_{i,t-s}\Delta v_{i,t}\right] = 0, \quad \text{when } t = 3, \ldots, T \text{ and } s\geq 2\,. \tag{4}$$

Thus, all $Emig_{i,t-s}$ for $s\geq 2$ are valid instruments for both $\Delta \text{Emig}_{i,t}$ and $\Delta \text{Emig}_{i,t-1}$ in Eq. (2).[11] Therefore, the difference GMM exploits the two sets of moment conditions (3) and (4) to estimate Eq. (2).

Blundell and Bond (1998) show that when the time period is short (T is small) or the dependent variable is highly time persisting ($|\beta_1|$ is close to 1), the standard difference GMM suffers from the problem of weak instrumental variables. They label moment conditions, such as Eqs. (3) and (4), as moment conditions in differences. In addition to these moment conditions in differences, they also exploit another set of moment conditions, which are called moment conditions in levels.

Under the same assumptions discussed above, Blundell and Bond (1998) derive the following moment conditions for Eq. (1):

$$E\left[\Delta \text{Ineq}_{i,t-1}\left(\eta_i + v_{i,t}\right)\right] = 0, \quad \text{for } t = 2, 3, \ldots, T \tag{5}$$

and

$$E\left[\Delta \text{Emig}_{i,t-1}\left(\eta_i + v_{i,t}\right)\right] = 0, \quad \text{for } t = 2, 3, \ldots, T\,. \tag{6}$$

In other words, $\Delta \text{Ineq}_{i,t-1}$ and $\Delta \text{Emig}_{i,t-1}$are used to instrument $Ineq_{i,t-1}$ and $Emig_{i,t-1}$ in Eq. (1) for $t\geq 2$, respectively.

Blundell and Bond (1998) call this method the system GMM that estimates Eqs. (1)–(2) simultaneously by exploiting the moment conditions (3)–(6), and demonstrate that the system GMM estimators are very robust even in a finite sample.

---

[8] In Eq. (1), we define $u_{it}=\eta_i+v_{it}$. After the estimation, unlike $\hat{u}_{it}$ that can be computed, $\hat{v}_{it}$ cannot be computed because η_i is unobservable. To test the serial autocorrelation of v_{it}, the Arellano–Bond test is implemented to examine the serial autocorrelation of Δu_{it} as$\Delta u_{it}=\Delta v_{it}$. The null hypothesis of the Arellano–Bond test for the first-order autocorrelation is H_0: $\text{cov}[\Delta v_{i,t},\Delta v_{i,t-1}]=0$. Mathematically, this null hypothesis is always rejected because $\text{cov}[\Delta v_{i,t},\Delta v_{i,t-1}]=-\sigma^2_{v_{it}}\neq 0$. The null hypothesis of the Arellano–Bond test for the second-order autocorrelation is H_0: $\text{cov}[\Delta v_{i,t},\Delta v_{i,t-2}]=0$. Rejection of this null hypothesis fails if and only if $\text{cov}[v_{i,t},v_{i,t-1}]=0$, which means $v_{i,t}$ is not first-order serially correlated. All $Ineq_{i,t-s}$ for $s\geq 2$, are implied as valid instruments for $\Delta \text{Ineq}_{i,t-1}$ in Eq. (2), and all $\Delta \text{Ineq}_{i,t-s}$ for $s\geq 1$, are valid instruments for $Ineq_{i,t-1}$ in Eq. (1). For more details on the Arellano–Bond test, see Roodman (2009), Section 3.5.

[9] For further discussion, see Arellano and Bond (1991).

[10] This equation implies that $E(Emig_{i,t}v_{i,t})\neq 0$ and $E(Emig_{i,t-1}v_{i,t})=0$ in Eq. (1).

[11] Under the assumption of $E(Emig_{i,t}v_{i,s})\neq 0$ for $s\leq t$ and $E(Emig_{i,t}v_{i,s})=0$ for $s>t$, $Emig_{i,t-1}$ is an exogenous variable in Eq. (1) because $E(Emig_{i,t-1}v_{i,t})=0$. However, $\Delta \text{Emig}_{i,t-1}$ is not an exogenous variable in Eq. (2) because $E(\Delta \text{Emig}_{i,t-1}\Delta v_{i,t})=E[(Emig_{i,t-1}-Emig_{i,t-2})(v_{i,t}-v_{i,t-1})]$, which is $-E(Emig_{i,t-1}v_{i,t-1})$ by manipulation and it is not equal to zero.

In summary, several advantages are derived from using system GMM to estimate the effects of migration on income inequality in the sending communities.[12] First, the unobservable community heterogeneities that may affect migration and income inequality simultaneously are safely swept out. Second, the lagged inequality is controlled for in the regression equation. To our best knowledge, no other study has treated lagged income inequality properly when estimating the effect of migration on inequality. Since income inequality, like most aggregate time series, exhibits strong time persistency, controlling for lagged inequality is necessary when estimating the effect of migration on inequality. Lastly and most importantly, contemporary migration is validly instrumented, and the concern of a potential reverse causality running from income inequality to contemporary migration is cleared out in our system GMM framework. Thus, the system GMM estimate of the effect of migration on inequality suggests a causal relationship.[13]

4. Data

We use the Chinese Health and Nutrition Survey (CHNS), which is a panel dataset with nine survey waves (1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009, and 2011).[14] The CHNS is conducted by the Carolina Population Center (CPC) at the University of North Carolina, Chapel Hill, The Institute of Nutrition and Food Hygiene, and the Chinese Academy of Preventive Medicine.[15] The CHNS surveys were administered by an international team of researchers whose backgrounds include nutrition, public health, sociology, Chinese studies, demography, and economics. The CPC expended considerable effort in staff training and quality control to ensure high quality of data.

The survey was conducted at both community and household levels. A community is the basic level of China's administrative hierarchy and refers to a village in a rural area or a neighborhood in an urban area. The sampled communities were randomly drawn from nine provinces, namely, Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning, and Shandong.[16] These provinces vary substantially in terms of geography, economic development, public resources, and health indicators.[17] A multistage random cluster process was used to draw the sample surveyed in each province. A total of 190 primary sampling units (i.e., 32 urban neighborhoods, 30 suburban neighborhoods, 32 towns, and 96 rural villages) were selected from 1989 to 1993. The primary sampling units have increased to 216 (i.e., 36 urban neighborhoods, 36 suburban neighborhoods, 36 towns, and 108 villages) since 1997.[18] Currently, about 4400 households are included in the survey, with a total of 19,000 individuals.[19]

Between 20 and 35 households were randomly drawn from each community, and the CHNS survey covers all the members formally registered in a household or those with permanent residence or *hukou*. The CHNS survey is designed to examine the effects of health, nutrition, and family planning policies in China and collects detailed information on economic, demographic, and social characteristics of individuals, households, and communities.

This dataset provides a valuable and unique opportunity to examine the dynamic relationship between migration and inequality. The CHNS has compiled longitudinal files of constructed household and individual income for all the nine waves. Thus, this paper is one of the first studies that have used the CHNS longitudinal income file to conduct economic research on the relationship between migration and income inequality.

We restrict our analysis to the six waves of the CHNS survey (1997, 2000, 2004, 2006, 2009 and 2011) because it is only in these waves that specific questions on labor migration were asked at the household level. A community panel dataset is constructed by aggregating household information to the village level. A total of 733 observations are eventually obtained from six different waves, in which 122 unique communities sequentially appear at least three times in the six waves.[20] Table 1 presents the summary statistics for the main variables.

4.1. Inequality

The Gini coefficient and Theil index are calculated using household income from the income module of the household survey.[21] Household income is constructed as the sum of all of the sources of income and revenue. There are nine potential sources of income: business, farming, fishing, gardening, livestock, non-retirement wages, retirement wages, subsidies, and other income. The income at each wave is then inflated to 2011 currency values. In our paper, we further categorize household income into household gross income, household per capita income, and household net income, that is, income minus expenditure. The mean value of the Gini

[12] System GMM is one of the simultaneous equation models. This model uses all the moment conditions and is the most efficient among all linear regression methods. Hansen's overidentification test and Arellano–Bond test for first- and second-order autocorrelations are carried out after estimation of system GMM to test the validity of the instruments and the necessity of controlling for lagged dependent variables.

[13] These estimates capture the general equilibrium effects of migration on inequality, which include effects through direct remittances, multiplier effects of remittances from the spending of remittances on local non-tradable products and services, and network effects (McKenzie & Rapoport, 2007).

[14] We are grateful to an anonymous referee for the advice to add the latest two waves into our data series.

[15] We are grateful to the UNC Carolina Population Center for providing the data.

[16] Beijing, Shanghai and Chongqing were added in the 2011 wave.

[17] Jiangsu, Liaoning, and Shandong are among the richest; Henan and Hunan are among the middle; and Guangxi and Guizhou are among the poorest provinces. Geographically, Jiangsu, Liaoning, Shandong, and Guangxi are coastal regions, whereas the other provinces are inland regions (Chen and Zhou, 2007).

[18] For further information, see http://www.cpc.unc.edu/projects/china/design/survey.html.

[19] Attrition problems probably exist in the CHNS data because households and communities are replaced. However, we find that more than 90% of the communities surveyed in 1997 were also surveyed in 2011.

[20] The system GMM estimation needs at least three waves.

[21] As the majority of Chinese rural-to-urban migrants are temporary migrants, migrants are counted as household members, and their incomes are counted as part of the household income.

coefficient of the household gross income in our study period is 0.40. The value is highly consistent with the calculated Gini coefficient of the rural areas obtained by Gustafsson, Li, and Sicular (2008), which indicates our use of a representative sample. The Gini coefficient is slightly smaller when calculated using the household per capita or the household net incomes. Figs. 1 and 2 show the income inequality measured by the Gini coefficients and Theil indices based on the CHNS data.

4.2. Gender wage gap

Gender wage gap is defined as the ratio of log male wage over log female wage. We obtain the information on the gender wage from two sources: the community survey as the official record and the self-reported wage in the adult survey.[22] Using the officially reported wage rate, the mean gender wage gap in the entire study period is 1.09. However, the gap decreased over the years, starting from more than 1.30 in the early 1990s to approximately 1.15 in 2000 and 1.07 in 2011. The gender wage gap calculated from the self-reported wage rate follows a similar pattern. However, the wage information calculated from both sources is subjected to a higher missing-value rate and no-response rate than the household income data used to calculate income inequality. The number of observations in wage gap regression is about half of those in the income inequality regression.

4.3. Migration

We define a migrant as one whose *hukou* was in the village while he or she was working or seeking employment outside the village at the time of the survey. In this sense, we are analyzing the effect of *temporary* migration on the inequality in the sending village because the *hukou* of these migrants remains in the village.[23] The migration rate is defined as the ratio of migrants to the total number of laborers in the village who are older than 16 and younger than 65.[24] Table 1 shows that 17% of the total labor in the village working outside the village on the average during the study period. Fig. 3 shows that migration rates increased dramatically from 1997 to 2006 and then declined slightly from 2006 to 2011.

4.4. Additional controls

Table 1 also presents the summary statistics for the other control variables, which are time varying at the community level. They include demographic structural variables such as shares of elderly people (older than 65 years old), shares of young people (younger than 15 years old), the indicator for the Han ethnic group, and shares of people who were born in other provinces, while their *hukou* was in the village. Educational compositional variables, such as the shares of people with different educational attainments are also considered. Moreover, we include the average household gross income, measured in the 2011 constant price, in the regression as control because the Gini measure of income inequality is more sensitive to variations in the middle section of the income distribution.

5. Empirical results

This section presents our main empirical results. We first report the estimated effects of migration on income disparity across households in the sending villages. Second, we present the estimation results of the effect of migration on gender wage gap in the sending villages.

5.1. The effect of migration on income inequality in the sending villages

Table 2 presents the system GMM estimates of the effects of migration on the Gini coefficient and Theil index calculated based on household income. As discussed in the Empirical Strategy section, we use all of the lagged inequality and migration to instrument the difference in the lagged inequality as well as the differences in current migration and the lagged migration in Eq. (2). At the same time, the difference in the lagged inequality and the difference in migration are used to instrument the lagged inequality and migration in Eq. (1). We then estimate Eqs. (1) and (2) simultaneously.[25]

All the columns include the lagged Gini coefficient, contemporary migration, and lagged migration, a set of controls, and survey year fixed effects. Column (1), (2), and (3) use the Gini coefficient of household per capita income, household gross income, and household net income as dependent variables, respectively. Column (4), (5) and (6) use the Theil index as the dependent variable.

First, the estimated coefficients on the lagged Gini coefficient in columns (1) to (3) are positive and statistically significant at a high level of 1%. This result justifies the use of the system GMM model as it shows that the dependent variable of the Gini coefficient is serially correlated. Thus, omitting the lagged dependent variable from Eq. (1) will result in biased fixed effects estimates. The test result

[22] Measuring the gender wage gap in the rural areas is difficult because a significant number of females only work within the household. However, the community survey in the CHNS provides information on the daily wages of males and females in the village as reported by the village leader. Moreover, the adult survey also asks adult respondents about their monthly wages.

[23] Given that a community in the CHNS refers to a village in a rural area or a neighborhood in an urban area and that our study only focuses on rural-to-urban migration, the present paper interchangeably uses the word community level with village level.

[24] We have used the international definition of 16 to 65 years old as working age. We have also conducted robustness checks with the Chinese definition of 16 to 60 years old. The regression results are qualitatively the same.

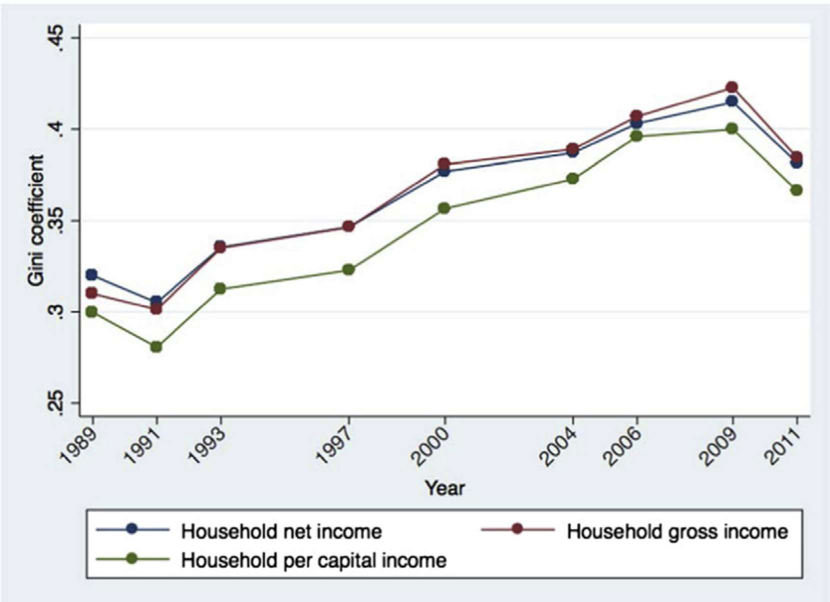
[25] Notably, the number of observations in Tables 2 is smaller than that of Table 1. Eq. (1), being a dynamic equation accounts for this. Thus, one wave of observations is not counted. Moreover, system GMM estimation requires that a village appears at least three sequential times in the waves of 1997, 2000, 2004, 2006, 2009, and 2011.

Table 1
Summary statistics.

Variables	Obs	Mean	Std. dev.
Dependent variables: income inequality			
Gini coefficient (household net income)	733	0.401	0.0970
Gini coefficient (household gross income)	733	0.405	0.106
Gini coefficient (household per capita income)	733	0.384	0.101
Theil index (household net income)	733	0.308	0.178
Theil index (household gross income)	733	0.322	0.206
Theil index (household per capital income)	733	0.286	0.185
Dependent variables: gender wage gap			
Official reported ln(male wage)/ln(female wage)	660	1.089	0.110
Self-reported ln(male wage)/ln(female wage)	430	1.050	0.0689
Interested independent variables			
Migration:	733	16.57	13.07
Proportion working outside the village in total labor (%)			
Control variables			
Share of elderly people (age > 64) (%)	733	9.643	6.128
Share of young people (age < 15) (%)	733	12.57	6.271
Share of Han people (%)	733	86.30	28.61
Share of people born in other provinces (%)	733	2.630	5.402
Share of illiterate or semi-illiterate (%)	733	24.01	15.44
Share of primary educated (%)	733	23.57	11.62
Share of middle school educated only (%)	733	34.12	11.71
Share of high school and above educated (%)	733	6.972	12.32
Household gross income (RMB/2011)	733	29,738	17,219

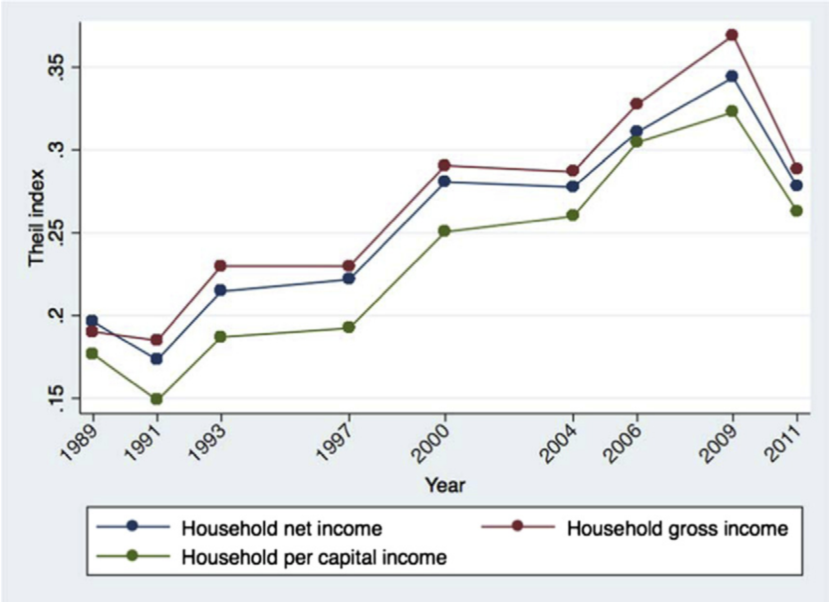
Data source: CHNS data (1997, 2000, 2004, 2006, 2009, and 2011).
Note: Sample is restricted to rural communities. The Gini coefficient and Theil index are calculated using household total income, per capita income, and net income from the income module of the household survey. Gender wage gap is defined as the ratio of log male wage rate over log female wage rate. Official wage rate is reported by the community head in the community survey. Self-reported wage rate is reported by the individual adult in the adult survey. The smaller sample size of gender wage gap measure is due to the higher non-response rate of the gender wage questions in the community and adult survey.

of the Arellano–Bond test for the first-order serial correlation further substantiates the necessity of using the system GMM model because the null hypothesis of no first-order serial correlation is statistically rejected at a high level of 1%. The result of the Arellano–Bond test for the second-order serial correlation is insignificant. Moreover, the estimated coefficient on the Gini coefficient is more than 0.20 and less than 1. Therefore, the economic pattern of inequality follows a standard conditional convergence.



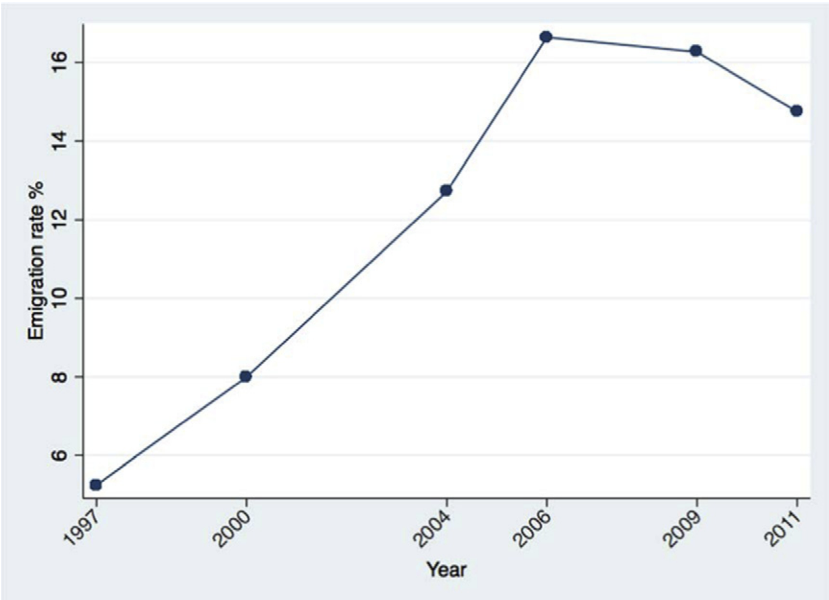
Data source: CHNS data (1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009, and 2011)

Fig. 1. Gini coefficients of household income in rural China (1989–2011).



Data source: CHNS data (1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009, and 2011)

Fig. 2. Theil indices of household income in rural China (1989–2011).



Data source: CHNS data (1997, 2000, 2004, 2006, 2009, and 2011)

Note: The migration rate is defined as the share of population aged 16 to 65 to total labor force in the village who work outside the village where their *hukou* is registered.

Fig. 3. Migration rates in rural China (1997–2011).

Second, the estimated coefficient on the current migration is positive. On the contrary, the estimated coefficient on the lagged migration is negative and statistically significant at a high level of 5%.[26] Interestingly, the positive effect of current migration and the negative effect of the lagged migration on inequality are consistent with the inverse U-shaped effect of migration on inequality

[26] We have included time dummies in all the regressions to control for time fixed effects. To check for robustness, we have also included a dummy variable indicating whether an economic zone exists near the village and a province dummy to control for any economic shocks that may simultaneously affect both migration and income inequality. The regression results are qualitatively the same.

Table 2
System GMM estimates of the effects of migration on income inequality in source villages in rural China.

Variables	(1) Gini income (per capita)	(2) Gini income (gross)	(3) Gini income (net)	(4) Theil income (per capita)	(5) Theil income (gross)	(6) Theil income (net)
Lagged dependent variable	0.215***	0.281***	0.200***	0.224**	0.261***	0.186**
	(0.072)	(0.066)	(0.065)	(0.093)	(0.081)	(0.081)
ln (% Migration)	0.013	0.002	0.010	0.029	0.004	0.026
	(0.010)	(0.009)	(0.009)	(0.021)	(0.023)	(0.021)
ln (% Migration)(t − 1)	−0.020**	−0.022**	−0.019**	−0.036**	−0.038**	−0.036**
	(0.009)	(0.009)	(0.008)	(0.017)	(0.019)	(0.017)
ln [% Old (age > 64)]	0.000	0.002**	0.001	0.000	0.003**	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
ln [% Young (age < 15)]	0.001	0.002**	0.001	0.002	0.004**	0.002
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
ln (% Han)	−0.000	−0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ln (% Born in other provinces)	0.001	0.001	0.000	−0.001	0.002	−0.001
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
ln (% Primary)	−0.000	−0.001**	−0.001*	−0.001	−0.002**	−0.002*
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
ln (% Middle school)	0.000	−0.001	−0.000	−0.000	−0.002*	−0.001
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
ln (% Higher education)	−0.002***	−0.003***	−0.003***	−0.003***	−0.006***	−0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ln (gross income)	0.013	0.063***	0.034***	0.057*	0.158***	0.088***
(RMB at year 2011)	(0.013)	(0.013)	(0.011)	(0.030)	(0.033)	(0.027)
Hansen over-identification test	46.789	51.914	66.628	47.296	49.296	58.136
p-value	0.481	0.288	0.031	0.460	0.381	0.128
Arellano–Bond test for first-order serial autocorrelation	−5.176	−5.410	−5.932	−3.874	−3.697	−3.964
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Arellano–Bond test for second-order serial autocorrelation	1.503	1.353	1.275	0.854	1.243	0.710
p-value	0.133	0.176	0.202	0.393	0.214	0.477
Observations	521	521	521	521	521	521
Number of villages	122	122	122	122	122	122

Data source: CHNS data (1997, 2000, 2004, 2006, 2009, 2011).
Note: Robust standard errors in parentheses. Sample is restricted to rural communities. The Gini coefficient and Theil index are calculated using household total income, per capita income and net income from the income module of the household survey.
*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

(McKenzie & Rapoport, 2007). The current migration reflects the immediate effect of migration, which is likely to be small. By contrast, the lagged migration reflects a highly accumulated effect of migration with a build-up of migration-specific human capital and networks. Thus, the lagged migration has a much larger effect than the current migration.

Finally, we conduct a Hansen overidentification test to examine the validity of the additional instruments. The p-value of the Hansen tests suggests that these instruments are statistically valid.[27]

The estimated positive effect of the current migration and the negative effect of the lagged migration on income inequality are strongly robust to the inclusion of different measures of household income inequality. Neither the magnitudes nor the standard errors of the estimated coefficients on both contemporary migration and lagged migration have changed substantially under different specifications.

In terms of the control variables, the share of old people tends to increase the income inequality, whereas the share of young people tends to decrease the income inequality. Compared with the omitted baseline group (the illiterate or semi-illiterate), the shares of primary school-educated, middle school-educated, and other higher level-educated people tend to decrease income inequality. Specifically, the coefficients before the share of higher level-educated people are highly significant in the inequality equation, which suggests a highly robust inequality-reducing effect of higher education.

The last control variable is the mean household gross income. Mean gross income tends to increase the income inequality at the village level, as the estimated coefficient is highly significant. However, caution should be exercised in interpreting the effect of gross income on income inequality because inequality can simultaneously affect economic growth and income (Banerjee & Duflo, 2003; Forbes, 2000). Given that this paper focuses on the relationship between migration and inequality, and that no good instrument for gross income is available, we include the mean gross income in the inequality equation only as a robustness check. The regressions without this gross income control present very similar results.

[27] The Hansen test is a test of overidentification restrictions, which allows for heteroskedasticity robust standard errors. The joint null hypothesis is that the excluded instruments are correctly excluded from the structural growth equation, and the structural equation is correctly specified. Under the null hypothesis, the test statistic is asymptotically distributed as chi-squared with the degrees of freedom equal to the number of overidentification restrictions. See Hayashi (2000) for further discussion.

Although considered as the most popular measure of inequality in the past years, the Gini coefficient possesses many undesirable properties (Deaton, 1997).[28] Thus, we replace it with the Theil index, which is another popular measure of inequality in practice, and repeat the exercises in columns (4) to (6). We find that the estimated inverse U-shaped relationship between migration and income inequality is robust across different measures of inequality. Specifically, the estimated coefficients on the lagged migration are consistently negative and statistically significant at the level of 5% across all the specifications. The estimated coefficients on the contemporary migration are consistently positive, although they are statistically insignificant.

5.2. The effect of migration on the gender wage gap in the sending villages

Table 3 reports the estimated effect of migration on the gender wage gap in the sending villages under both system GMM and fixed effect estimation. Before discussing the regression results, we note that we have less information on gender wage gap than on income inequality. The number of observations in Table 3 is about half of that in Table 2. The smaller sample size is caused by the higher no-response rate for wage questions in both the community and adult surveys.

In columns (1) and (2), we present the regression results derived using the officially reported gender wage gap as dependent variable. The gender wage gap is defined as the ratio of log male wage over log female wage. Average daily male and female wages were directly solicited from the community head during the community survey. We suspect the gender wage gap is time persistent. Thus, we control for the lagged gender wage gap and conduct system GMM estimation, the results of which are presented in column (1). The coefficient before contemporary migration is positive and insignificant, whereas the coefficient before lagged migration is negative and significant at 1%. Moreover, the reducing effect of lagged migration on gender wage gap is stronger than the reinforcing effect of contemporary migration (−0.016 versus 0.011).

The coefficient of the lagged dependent variable in column (1) is very small and insignificant, contrary to that of the income inequality regression. This suggests that the time persistency of gender wage gap is weak, and the use of system GMM to control for the lagged gender wage gap is not needed. As a robustness check, we have also conducted fixed effect estimation without including the lagged dependent variables. In column (2), FE coefficient before the lagged migration remains significant at 10%. The reverse causality from lower gender wage gap to higher migration rate continues to endanger the consistency of the fixed effect estimates even though the insignificant coefficient before the lagged gender wage gap disregarded the concerns of the time-persistent dependent variable. Thus, the system GMM is preferred because it resolves the potential reverse causality bias by instrumenting the contemporary and lagged migration.[29]

In columns (3) and (4), we replace the official-reported gender wage gap with the self-reported gender wage gap. The self-reported wage rate is obtained from the adult survey. The coefficient before the lagged migration remains significant at the 1% level in the system GMM in column (3), which confirms the reducing effect of lagged migration on gender wage disparity. The coefficient before contemporary migration is positive and insignificant but larger in magnitude than the estimate of column (1). Moreover, the coefficient of lagged dependent variable remains insignificant although larger than that of column (1), and its t-value exceeds 1. The fixed effect estimation in column (4) presents consistent results.

The comparison between columns (3) and (1) reveals the nature of the official-reported and self-reported wage data. First, the coefficients of the lagged dependent variable are compared. The self-reported gender wage gap appears more time persistent than the official measure because both the magnitude and the t-value of the lagged dependent variable are higher in column (3) than in column (1). This may be attributed to the fact that the community heads are replaced every few years; thus, different community heads may have different levels of understanding over the local labor market. Second, the reducing effect of lagged migration and the reinforcing effect of contemporary migration are stronger in magnitude and more significant on the self-reported wage gap than on the official-reported wage gap, which suggests that official-reported wage gap estimates might harbor more measurement errors, or that the community heads are more likely to report an underestimated gender wage disparity. However, the disadvantage of using self-reported gender wage gap is the smaller number of observations in column (3) than in column (1). This is mainly a result of higher no-response rate in the individual wage question of the adult survey, which may have led to more serious sample selection. Nevertheless, the significantly negative coefficients in both columns (1) and (3) confirm the reducing effect of lagged migration on gender wage gap in the source villages in rural China.

Thus, the relationship between migration and gender wage gap seems to be inverse U-shaped, which is similar to the relationship between migration and income inequality. This inverse U-shaped relationship between migration and gender wage gap resonates with the finding of Sun and Fan (2011) that earlier migrants in the 1990 census are male-biased, whereas the latter migrants in the 2000 census are more gender-neutral.

[28] For example, economies with similar incomes and Gini coefficients can still have very different income distributions. The ability of the Lorenz curves to have different shapes and yet still yield the same Gini coefficient is the reason.

[29] We have included the sex ratio of migrants as a control variable in wage gap regression. The sex ratio is defined as the percentage of male migrants over the total migrants at the community level. The coefficient before the sex ratio is negative, whereas a conventional supply–demand story predicts a positive coefficient. Conventionally, higher male migration leads to shorter supply of male labor in source villages, thus forming a higher gender wage gap. The negative effect of sex ratio of migrants on gender wage gap suggests that the lower male wage in local labor market could reversely drive a higher proportion of male migration. Thus, OLS and the fixed effect could be biased in this case. System GMM resolves this problem by instrumenting the contemporary and lagged migration rates. Nevertheless, the effect of lagged migration remains negative and significant at the 5% level even after controlling for the sex ratio of migrants.

Table 3
System GMM and FE estimates of effects of migration on gender wage gap in source villages in rural China.

	(1)	(2)	(3)	(4)
Variables	Official-reported wage gap ln(male wage)/ln(female wage)		Self-reported wage gap ln(male wage)/ln(female wage)	
Model	GMM	FE	GMM	FE
Lagged dependent variable (t − 1)	0.029		0.086	
	(0.063)		(0.093)	
ln (% Migration)	0.011	0.000	0.014	0.016**
	(0.013)	(0.010)	(0.010)	(0.008)
ln (% Migration) (t − 1)	−0.016**	−0.021*	−0.027***	−0.019**
	(0.008)	(0.012)	(0.010)	(0.009)
Male migrants/total migrants	−0.060	−0.028	−0.002	−0.031
	(0.046)	(0.037)	(0.020)	(0.025)
ln (% Primary)	−0.000	0.001	−0.000	−0.001
	(0.001)	(0.001)	(0.000)	(0.001)
ln (% Middle school)	−0.000	0.001	−0.000	−0.001
	(0.001)	(0.001)	(0.000)	(0.001)
ln (% Higher education)	−0.001	0.001	−0.001***	−0.000
	(0.001)	(0.001)	(0.000)	(0.001)
ln (gross income)	−0.043***	−0.045**	0.046***	0.034**
(RMB at year 2011)	(0.016)	(0.018)	(0.013)	(0.016)
Hansen over-identification test	50.305		49.390	
Hansen p-value	0.307		0.378	
Arellano–Bond test for first-order serial autocorrelation	−1.240		−2.922	
p-value	0.215		0.003	
Arellano–Bond test for second-order serial autocorrelation	1.086		−1.153	
p-value	0.277		0.249	
Observations	284	324	220	293
Number of villages	108	113	78	102

Data source: CHNS data (1997, 2000, 2004, 2006, 2009, and 2011).
Note: Robust standard errors in parentheses. Sample is restricted to rural communities. Gender wage gap is defined as the ratio of log male wage rate over log female wage rate. The official wage rate is reported by the community head in the community survey. The self-reported wage rate is reported by the individual adult in the adult survey. The smaller sample size of gender wage gap measure is due to the higher non-response rate of the gender wage questions in the community and adult surveys.
*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

6. Conclusion and discussion

Researchers have long contemplated an inverse U-shaped pattern between migration and inequality in the sending communities, that is, inequality rises in the beginning of the migration process and drops after migration becomes more established. However, the literature offers minimal solid empirical evidence due to both data limitations and methodology shortcomings. This study analyzes the impact of rural-to-urban migration on inequality using a newly constructed panel for more than 100 villages over a 14-year period from 1997 to 2011 in China. This is one of the first studies to examine the dynamic aspects of migration and income inequality using panel data analysis. Unlike earlier studies focusing exclusively on remittances, our analysis takes into account the total labor earnings of migrants in destination areas. Furthermore, we look at the gender dimension of the impact of migration on wage inequality within the sending communities.

We use a system GMM framework to control for the lagged income inequality in estimating the effect of migration on income inequality in the sending villages because income inequality is time persisting. Contemporary migration is instrumented in the GMM framework because of the unobserved time-varying community shocks that correlate with migration and income inequality and the potential reverse causality from income inequality to migration. We find an inverse U-shaped pattern between migration and income inequality in the sending communities. Contemporary migration increases income inequality, whereas lagged migration has strong income inequality-reducing effect in the sending villages. A 50% increase in the lagged migration rate translates into a one-ninth to one-tenth standard deviation reduction in income inequality. Contemporary migration has positive but statistically insignificant effects on raising the income inequality within the villages. These effects are robust to the different specifications and measures of inequality. Interestingly, the estimated relationship between migration and gender wage gap also has an inverse U-shape. Migration tends to increase the gender wage gap and then tends to decrease it in the sending villages.

References

Adams, R.H., Jr. (1993). The economic and demographic determinants of international migration in rural Egypt. *Journal of Development Studies*, *30*(1), 146–167.
Adams, R.H., Jr. (1998). Remittances, investment, and rural asset accumulation in Pakistan. *Economic Development and Cultural Change*, *47*(1), 155–173.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2), 277–297.

Banerjee, A.V., & Duflo, E. (2003). Inequality and growth: What can the data say? *Journal of Economic Growth*, *8*(3), 267–299.

Barham, B., & Boucher, S. (1998). Migration, remittances, and inequality: Estimating the net effects of migration on income distribution. *Journal of Development Economics*, *55*(2), 307–331.

Benjamin, D., Brandt, L., & Giles, J. (2005). The evolution of income inequality in rural China. *Economic Development and Cultural Change*, *53*(4), 769–824.

Black, R., Natali, C., & Skinner, J. (2006). Migration and inequality. Background paper, equity and development. *World Development Report*.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, *87*(1), 115–143.

de Brauw, A., & Giles, J. (2008a). Migrant labor markets and the welfare of rural households in the developing world evidence from China. *World Bank Policy Research Working Paper. 4585*.

de Brauw, A., & Giles, J. (2008b). Migrant opportunity and the educational attainment of youth in rural China. *World Bank Policy Research Working Paper 4526*.

Cai, F. (2003). Removing the barriers to labor mobility: labor market development and its attendant reforms. *World Bank Workshop on National Market Integration in China, Beijing*.

Cai, F., & Wang, M. (2012). Labor migration and income inequality. In J. Xue (Ed.), *Growth with inequality: An international comparison on income distribution. 3*.

Chan, W.K., & Buckingham, W. (2008). Is China abolishing the hukou system? *China Quarterly*, *195*, 582–606.

Chen, Y., Jin, G.Z., & Yue, Y. (2010). *Peer migration in China. No. w15671*. National Bureau of Economic Research.

Chen, Y., & Zhou, L. A. (2007). The long-term health and economic consequences of the 1959–1961 famine in China. *Journal of health economics*, *26*(4), 659–681.

De Haas, H. (2010). Migration and development: A theoretical perspective. *International Migration Review*, *44*(1), 227–264.

Deaton, A. (1997). *The analysis of household surveys: A microeconometric approach to development policy*. World Bank Publications.

Du, Y., Park, A., & Wang, S. (2005). Migration and rural poverty in China. *Journal of Comparative Economics*, *33*(4), 688–709.

Fan, C.C. (2008a). Migration, hukou, and the city. *China urbanizes: Consequences, strategies, and policies* (pp. 65–89).

Fan, C.C. (2008b). China on the move: Migration, the state, and the household. *China Quarterly*, *196*, 924–956.

Forbes, K.J. (2000). *American economic review*, 869–887.

Giulietti, C., Wahba, J., & Zenou, Y. (2014). *Strong versus weak ties in migration*.

Gustafsson, Björn A., Li Shi, and Terry Sicular, eds. Inequality and public policy in China. Cambridge University Press, 2008.

Ha, W., Yi, J., & Zhang, J. (2015). Brain drain, brain gain, and economic growth in China. *China Economic Review*. http://dx.doi.org/10.1016/j.chieco.2015.02.005.

Harris, J.R., & Todaro, M.P. (1970). Migration, unemployment and development: A two-sector analysis. *The American Economic Review*, *126-142*.

Hayashi, F. (2000). *Econometrics*. Princeton: Princeton University Press.

Jones, R.C. (1998). Remittances and inequality: A question of migration stage and geographic scale. *Economic Geography*, *74*(1), 8–25.

Jones, R.C. (2013). Migration stage and household income inequality: Evidence from the Valle Alto of Bolivia. *The Social Science Journal*, *50*(1), 66–78.

Lee, L. (2012). Decomposing wage differentials between migrant workers and urban workers in urban China's labor markets. *China Economic Review*, *23*(2), 461–470.

Lewis, W.A. (1954). Economic development with unlimited supplies of labor. *Manchester School of Economic and Social Studies*, *22*(2), 139–191.

Li, Q., Mao, X., & Zhang, T. (2008). Migrant worker's remittances: Quantity and usage [Nongmingong Huikuan De Juece Shuliang Yu Yongtu]. *China rural observerZhongguo Nongcun Guancha (pp. 2–12)3*. (pp. 2–12).

Lipton, M. (1980). Migration from rural areas of poor countries: The impact on rural productivity and income distribution. *World Development*, *8*(1), 1–24.

Long, W., Appleton, S., & Song, L. (2013). Job contact networks and wages of rural urban migrants in China. *IZA Discussion Paper No. 7577*.

Lu, Y., & Wang, F. (2013). From general discrimination to segmented inequality: Migration and inequality in urban China. *Social Science Research*, *42*(6), 1443–1456.

McKenzie, D., & Rapoport, H. (2007). Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. *Journal of Development Economics*, *84*(1), 1–24.

Mendola, M. (2012). Rural out-migration and economic development at origin: A review of the evidence. *Journal of International Development*, *24*(1), 102–122.

Meng, X. (2000). *Labor market reform in China*. New York: Cambridge University Press.

Messinis, G. (2013). Returns to education and urban-migrant wage differentials in China: IV quantile treatment effects. *China Economic Review*, *26*, 39–55.

Murphy, R. (2000). Migration and inter-household inequality: Observations from Wanzai County, Jiangxi. *China Quarterly*, *164*, 965–982.

Murphy, R. (2002). *How migrant labor is changing rural China*. Cambridge University Press.

Pritchett, L. (2006). *Let their people come: Breaking the gridlock on global labor mobility*. Washington, DC: Center for Global Development.

Rapoport, H., & Docquier, F. (2006). The economics of migrants' remittances. *Handbook of the economics of giving, altruism and reciprocity. 2*. (pp. 1135–1198).

Ray, D. (1998). *Development economics*. Princeton University Press.

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal*, *9*(2009), 86–136.

Song, H., & Bai, N. (2002). *Stay or return: Return migration of Chinese migrant workers [Huixiang or Jincheng: Zhongguo Nongcun Waichu Laodongli Huiliu Yanjiu]*. Beijing: China Fiscal Economics Publishing House (Zhongguo Caizheng Jingji Chubaishe).

Stark, O., & Bloom, D.E. (1985). The new economics of labor migration. *American Economic Review*, *75*(2), 173–178.

Stark, O., & Taylor, J.E. (1991). Migration incentives, migration types: The role of relative deprivation. *The Economic Journal*, 1163–1178.

Stark, O., Taylor, J.E., & Yitzhaki, S. (1986). Remittances and inequality. *The Economic Journal*, 722–740.

Stark, O., Taylor, J.E., & Yitzhaki, S. (1988). Migration, remittances and inequality: A sensitivity analysis using the extended Gini index. *Journal of Development Economics*, *28*(3), 309–322.

Sun, M., & Fan, C.C. (2011). China's permanent and temporary migrants: Differentials and changes, 1990–2000. *The Professional Geographer*, *63*(1), 92–112.

United Nations Development Programme (2009). *Human development report: Overcoming barriers: Human mobility and development*. United Nations Development Programme, 2009.

Xue, J., Gao, W., & Lin, G. (2014). Informal employment and its effect on the income distribution in urban China. *China Economic Review*, *31*, 84–93.

Zhang, J., & Zhao, Z. (2015). Social-family network and self-employment: evidence from temporary rural–urban migrants in China. *IZA Journal of Labor & Development*, *4*(1), 1–21.

Zhao, S. (2000). Organizational characteristics of rural labor mobility in China. *Rural labor flows in China* (pp. 231–250). Berkeley: Institute of East Asian Studies, University of California.

Zhu, N., & Luo, X. (2010). The impact of migration on rural poverty and inequality: A case study in China. *Agricultural Economics*, *41*(2), 191–204.

Zhu, R. (2016). Wage differentials between urban residents and rural migrants in urban China during 2002–2007: A distributional analysis. *China Economic Review*, *37*, 2–14.