Climate Anomalies and Migration between Chinese Provinces: 1987–2015

M. R. Barassi, M. G. Ercolani,*† M. J. Herrerias,* and Z. Jin***

ABSTRACT

Internal migration between Chinese provinces has increased substantially since the mid–1980s. Though it is generally agreed that this has been driven by economic factors, climatic factors might also have had a part to play. The challenge is to evaluate the impact of climatic factors on migration in the simultaneous presence of changing socio-economic influences. We resolve this challenge by carrying out a statistical multivariate regression analysis on bilateral migration rates between Chinese provinces. The analysis simultaneously includes climate change in the form of climate anomalies (temperature, precipitation, sunshine) and various socio-economic factors including energy consumption. To this end we have constructed a unique three-dimensional panel dataset (time, sending province, receiving province) with bilateral migration rates between 30 provinces for the period 1987–2015. Due to the distributional properties of the data and underlying theory we use a Poisson Pseudo Maximum Likelihood (PPML) estimator but include OLS estimates for comparison. The results suggest that increases in temperature and precipitation are significant migration push factors while increased sunshine discourages push migration. Provincial differentials in per capita energy consumption and Gross Regional Product (GRP) are also significant drivers of migration.

Keywords: Migration; climate change; energy consumption; Chinese provinces;

https://doi.org/10.5547/01956574.39.SI1.merc

1. INTRODUCTION

Human migration is currently one of the major challenges facing both developed and developing countries. However, the forces driving this migration have changed with the increase in climate change. This is evidenced by the inclusion of these issues in the United Nations Framework Convention on Climate Change (UNFCCC, 1992) negotiations and in the comprehensive overview of the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change[1] (Warner, 2012; Marchiori et al., 2012). Other issues arising from climate change include poverty, poor health, and food insecurity, which increase the pressure to migrate (Joseph and Wodon, 2013). In addition, in a recent paper, Webster and Clarke (2017) suggested that a one degree centigrade increase

1. This report highlights the likely increase in global temperatures over the next 60 years and that increased climate variability is creating more frequent droughts, storms and other extreme weather events (IPCC, 2015).

* University of Birmingham, Department of Economics

** University of Nottingham, School of Economics

† Corresponding author. Department of Economics, Birmingham Business School, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom. E-mail: m.g.ercolani@bham.ac.uk

 Copyright © 2018 by the IAEE. All rights reserved.

in global average temperature will incur a 1% cost to US GDP.[2] The same authors estimated a 0.4% annual cost to world GDP between 1980 and 2010, which is equivalent to $1 trillion of costs, growing at 0.1% to 0.2% per decade. Though this topic has received increased attention in the media and policy-forums, the related empirical literature is still at an early stage (Marchiori et al, 2012). The goal of this paper is to better understand the underlying mechanisms through which climate change induces migration between Chinese provinces. This should help the Chinese authorities, and other governments, to design better policy responses to these migration pressures.

To the best of our knowledge, few papers have assessed the joint effects on migration arising from climate change, energy consumption and other socio-economic factors. Nonetheless, there is a growing awareness of the effects of long-run anomalies in temperatures, rainfall and sunshine on migration. The mechanisms seem evident; climate anomalies are associated with floods, storms and other events which can provoke energy outages thus reducing energy production and consumption. This, in turn, can impact on economic activity and consequently on social welfare.[3] Climate anomalies will affect energy efficiency, particularly in poor areas, because they affect energy costs for government agencies and private firms. Increased costs can lead to job losses and, in more severe cases, in the proliferation of diseases thus increasing health expenditure and diverting resources. Some of these effects are also evident at a micro-economic level, where areas affected by climate anomalies experience higher insurance premiums and lower profitability for firms (Webster and Clarke, 2017). In the financial sector insurance problems linked to a lack of access to finance will stop investment projects from being undertaken. These negative 'destructive effects' of climate anomalies might induce people to migrate to other provinces in search of better opportunities.

It is also possible, however, to find 'agglomeration effects' in areas which are less affected by climate anomalies and have a higher degree of development. Migration to these areas increases levels of energy consumption and the subsequent economic development leads to increased industrialization and urbanization. These areas have higher capacity and more resources, and might therefore be in a better position to confront the negative effects of climate anomalies. In the particular case of China, the relaxation of the restrictive Hukou system facilitated the increase in inter-provincial migration. This increase in migration is now further facilitated by the current Five-Year Plan which uses urbanization as a way of increasing household consumption and promoting internal demand thus achieving a more balanced economic growth.

The main contribution of our paper to the literature is to investigate the effect of climate change, energy consumption and other socio-economic factors on migration flows between Chinese provinces. Traditional research on migration has focused on climatic and socio-economic factors to explain international migration flows (Backhaus et al, 2015). However, to the best of our knowledge, this is the first attempt in the empirical literature to provide new insights on this issue for migration within China and the very first to include energy consumption in the analysis. We also believe this is the first empirical paper to include three climatic factors in the analysis in the form of temperature, precipitation and sunshine anomalies.

Understanding the impact of climate change on migration is important for policy-makers in the design of migration, energy and environmental policies. For example, high risk regions, that

2. The US Department of Energy (2014) highlighted that severe droughts affected more than one third of the US in 2012, which in turn affected water supply and the generation of electricity. The proliferation of shale gas and the water requirements to generate electricity are also driving a significant debate on the energy-water nexus and climate change.

3. This is particularly relevant in China as the energy mix is still highly dominated by coal in the North and hydropower in the South (Herrerias and Girardin, 2013). Moving towards the low carbon economy and a diversification of the sources of energy, in particular the promotion of renewable energy, should help to alleviate these effects.

All rights reserved. Copyright © 2018 by the IAEE.

suffer frequent floods or droughts, push Chinese citizens to migrate to less vulnerable regions. Furthermore, the high degree of urbanization in some regions presents a challenge to the effectiveness of environmental policies. Linking policies on migration, urbanization and the environmental can benefit the sustainable development of a green, low carbon economy (Barnett and Webber, 2010). This research can help to better understand past migration patterns and can even inform policy-makers on how best to respond to climate change that is induced by migration (Klaiber, 2014).

To estimate the impact of climate change on migration flows, we first consider the inter-provincial migration flows evident in China's census data (1987, 1990, 1995, 2000, 2005, 2010 and 2015). The advantage of these data is the highly detailed province-level information for the seven available censuses. We complement this information with data on climate change (temperature, rainfall, sunshine hours). We also control for socio-economic factors to isolate the climatic effects.

Our model estimates, based on Poisson Pseudo Maximum Likelihood (PPML) and confirmed using OLS, suggest that there are significant climatic and other effects on migration. Our results indicate that temperature increases and, to a lesser extent, increased rainfall are significant migration push factors. Interestingly, our results also indicate that increased sunshine in a province discourages push migration if other factors are simultaneously controlled for. When it comes to energy consumption, our results indicate that provincial differentials in energy consumption are significant drivers of migration given that they might be a proxy for labour market activity.[4] Similarly, we find that provincial differentials in Gross Regional Product (GRP) per capita are also significant drivers of migration.

The reminder of the paper is structured as follows. In Section 2 we provide a background to the Chinese experience of climatic change. In Section 3 we describe the data. In Section 4 we present the estimation methodology and discuss the regression results. A final section concludes and draws policy implications.

2. LITERATURE

Climate change is now considered as one of the main drivers of increased urbanization (Joseph and Wodon, 2013). Urbanization typically raises per capita income as a result of specialization and trade cost reductions. As a consequence, individuals have more resources and governments can provide more public goods (Khah, 2014). In 2009 the global urban population exceeded 50% for the first time and it is forecast is that by 2050 it will reach 70% (Klaiber, 2014). Furthermore, productivity in urban areas is typically less vulnerable and less affected by climate 'anomalies' and natural hazards. Productivity and the population in rural areas are vulnerable to these anomalies because they are largely reliant on agriculture (Burrows and Kinney 2016, Reuvery 2017). This leads to rural-to-urban migration which, in turn, exerts a downward pressure on real urban wages which in turn provokes international migration (Klaiber, 2014).

Estimates vary widely on the global size of the 'environmental migration' caused by climate change and other environmental factors. For example, projections range from 150 million to 300 million environmental migrants by 2050 (Stern 2007, Christian Aid 2007, Government Office for Science 2011 p.28). However, there is still a lack of a full understanding about the mechanisms through which climate variability affects migration. Therefore, understanding the linkages between

4. See Herrerias, Joyeux and Girardin (2013) for the empirical relationship between energy consumption and economic activity across Chinese provinces.

Copyright © 2018 by the IAEE. All rights reserved.

climate change, its variability, and migration flows should help policy-makers and policy-advisers in designing suitable policy packages.

Burrows and Kinney (2016) provide a comprehensive evaluation of the new challenges arising from climate change and the water-energy nexus. They argue that there is a need to consider additional factors and their impact on the economy. First, the urgent need to integrate water and energy supply in the same policy framework which is in line with the debate in Europe on energy needs and climate change. Second, the negative effects arising from climate change, i.e. how increasing temperatures and precipitation affect the water-energy nexus. Third, population growth, its structure and migration flows. Fourth, the introduction of more efficient energy technology and the adoption of green technology, which might alleviate some of these problems. Perhaps, among the aforementioned factors, those related to climate change (temperature, precipitation, drought, extreme events) are the most challenging to model.

Traditionally, research into migration has focused on factors such as political instability, conflict, lack of employment, family networks, personal motivations or transportation (Burrows and Kinney 2016, Black et al. 2007, Massey 2015). Less attention has been placed on climate variability as a driver of migration. Those studies that have looked at the link between climate change and migration have used climate variability as a proxy variable to climate change (Burrows and Kinney, 2016). However, most of the empirical work has focused on droughts, soil degradation and floods (Benson et al. 2006, Gray 2011, Mallick and Vogt 2012) and less attention has been paid to temperature extremes (Burrows and Kinney, 2016).

Despite the growing interest, the empirical evidence is mixed (Cai et al. 2014). Some of the empirical evidence finds a positive relationship between migration and climate change (Reuveny and Moore 2009, Feng and Oppenheimer 2012). Others argue that climate change is inconsequential compared with other drivers of migration (Morteux and Barnett 2009, Naudé 2010). In addition to this, yet others have analysed permanent and transitory effects of natural disasters (Costello et al., 2009) and climate change on migration flows and whether these effects are direct or indirect (Beine and Parsons, 2015). The empirical evidence suggests that, in developed economies at least, the negative effects from climate change tend only to be transitory (Noy, 2009). In their recent paper Burrows and Kinney (2016) analysed the drivers of migration providing a comprehensive review of the literature. This literature highlighted that climate variability and the negative consequences of climate change such as floods, storms, droughts, and rising sea-levels could lead to international migration and that the regional context socio-economic context is of enormous relevance to this (Dun et al., 2008; Morton et al., 2008). It might therefore be that much of the ambiguity over the various empirical evidence is due to different methods used, countries analysed and time-periods considered (Cai et al., 2014).

Our paper seeks to fill these gaps and deal with some of these ambiguities by contributing new findings in the linkage between climate variability, socio-economic factors and migrations flows. Our modelling strategy is based on the premise that climatic factors can affect human migration directly by displacing people through natural disasters and indirectly by affecting their living standards (Beine and Parsons, 2015). Migration is therefore one channel through which individuals can respond to climate change. Migration can occur suddenly as a result of natural disasters or gradually as individuals progressively update their expectations about the economic opportunities and climate change (Klaiber, 2014). In related research, Saldaña-Zorrilla and Sanberg (2009) analyse recurrent natural disasters as a main determinant of outward migration from Mexico. They argue that income inequality plays an important role in explaining emigration. This is mainly driven by

All rights reserved. Copyright © 2018 by the IAEE.

natural disasters which reduce future income expectations, especially in rural areas where there is less adaptive capacity due to the high dependency on agriculture.

Analysing climate change as a determinant of migration does present some specific challenges (Joseph and Wodon, 2013). First, climate change can take different forms such as changes in temperature or rises in sea levels, but it can also lead to an increase of the frequency of extreme events such as droughts, floods, earthquakes and typhoons. Second, despite the significant improvements in measurement and early-warning systems, there can be a high degree of uncertainty on the exact geographical location of past and future natural disasters, and their impact on the economy.

The motivation for looking at the effect of climate patterns on migration seems therefore clear, however, some ambiguity exists in the literature on the definition of long-run weather or climatic anomalies. These definitions imply that something out of the ordinary is occurring, usually defined as deviations from the mean or sometimes defined as unusual intra-year events. In the natural sciences climatic anomalies sometimes simply refer to annual changes (e.g. Parry et al. 2007, Figure TS.6) or more often as annual deviations from some arbitrary multi-decade mean (e.g. Jones et al. 2009). Sometimes, particularly in large institutional reports, extreme short-run climatic events are referred to as climate anomalies (e.g. World Bank 2010, Map 1.2, deaths in the 2003 European heatwave).

In our case we adopt the definition of "weather anomalies" as specified by Marchiori et al. (2012) and defined in equation (5) in subsection 3.3. These anomalies are the 'normalised' annual deviations of any climate measurement from its whole-sample-period-mean. This allows us to normalise three different climatic measures: temperature, rainfall and sunshine; so that broadly comparable parameters can be estimated. Therefore, Marchiori et al.'s "weather anomalies" are not so much measures of weather anomalies but more akin to the measures of climate change used in the natural science literature. For this reason we prefer to use the term climate anomalies in our discussion. These anomalies could be one-offs, long-run trends, or even cycles but they are all annual deviations from the mean.

In summary, in this paper we examine the nexus between climate change, energy consumption and migration in a single integrated statistical framework. This analysis seems timely in the current climate-related policy debate and it also opens the discussion to further research questions.

3. DATA

Our empirical modelling strategy follows that found in Marchiori et al. (2012), Cattaneo and Peri (2015), and Beine and Parsons (2015). As in these papers, we carry out an empirical analysis of bilateral migration between geographical zones and correlate these movements with climatic factors.

As discussed above, Marchiori et al. (2012) look at the effect of climate change in the form of climate anomalies on migration from sub-Saharan countries. Cattaneo and Peri (2015) consider the effect of increasing temperatures on migration between 116 countries over a 41 year period. Beine and Parsons (2015) look at the effect of climatic trends, also measured by anomalies, on the migration between 166 countries over a 40 year period.

Our methodology follows most closely that by Beine and Parsons (2015) but our analysis is for 30 Chinese provinces over a 29 year period. We also make very minor adjustment to the notation in the model for the sake of clarity and we use a slightly different set of explanatory variables due to data availability.

Copyright © 2018 by the IAEE. All rights reserved.

3.1 Data structure

We use a near-balanced panel dataset for 30 Chinese provinces in the period 1987 to 2015. We examine the effect of natural hazards and climate change on migration rates while controlling for socio-economic determinants.

The data panel is not fully-balanced because in 1988 Hainan split off from Guangdong and in 1997 Chongqing gained a separate city status from Sichuan. Therefore, for Hainan and Guangdong the data start in census year 1990 and for Chongqing and Sichuan the data start in census year 2000. We were unable to use the alternative strategy of re-combining these four provinces into two because for some variables (e.g. climatic ones) we could not re-calculate the re-combined province level values. Furthermore, we have entirely excluded the 31st province, Tibet, because no data on energy consumption are available for it.

The data on migration rates originate from the 1990, 2000, 2010 censuses and the 1987, 1995, 2005, 2015 'intra-censuses' based on 1% population samples; henceforth all referred to as 'censuses'. Data on demographic characteristics in census and non-census years come from the China Compendium of Statistics for 1949–2008 and the China Statistical Yearbook for 2002–2016. Any overlap is used to check for data consistency. Data for energy consumption in census and non-census years come from the China Energy Statistical Yearbook. The data on annual provincial weather (climate) averages are from the China Statistical Yearbook for 1986–2016.

For each model specification we will present two sets of results, one limited to just the seven years for which we have census data and the other at the yearly frequency. In the latter case, we exploit the fact that all the explanatory variables are available for every year and we use geometric interpolation to fill in missing values for the dependent variable (migration) which are only collected in census years.

In addition, in some rare cases we have carried out linear or geometric interpolation to recover some isolated missing observations for the explanatory variables. Below we provide a detailed description on how and where this was done, and the affected proportion of observations. We have not recovered any missing values by extrapolation.

3.2 The dependent variable

Table 1 provides summary statistics for the dependent variable, *i.e.* the bilateral migration rate between provinces, in its seven census-years version and in its 29 years version. Though the number of people migrating between provinces each year can be large, when calculated as population proportions they are small averaging just 0.08%. The three-dimensional panel is not perfectly balanced for the reasons already explained above. The bottom half of Table 1 also summarizes the average bilateral migration rates between provinces for each individual census year. From this we can see that average bilateral migration has been on the increase since 1987, though has diminished slightly in 2015. As discussed above, though the 1990, 2000 and 2010 censuses are based on attempting to survey the entire population, the surveys for the other years is on 1% of the population so most of these migration shares are based on sample proportions.

Figure 1 illustrates a histogram of the dependent variable for the census years. We can see that its distribution is far from normal and is censored at zero. The zero cases (zero arrival rates) form a large, though not huge, number of cases. As discussed in section 3.4, for statistical and theoretical reasons we therefore model this variable using a Poisson Pseudo Maximum Likelihood (PPML) estimator. For comparison, in the appendix, we report OLS estimates on the natural logarithm of the dependent variable.

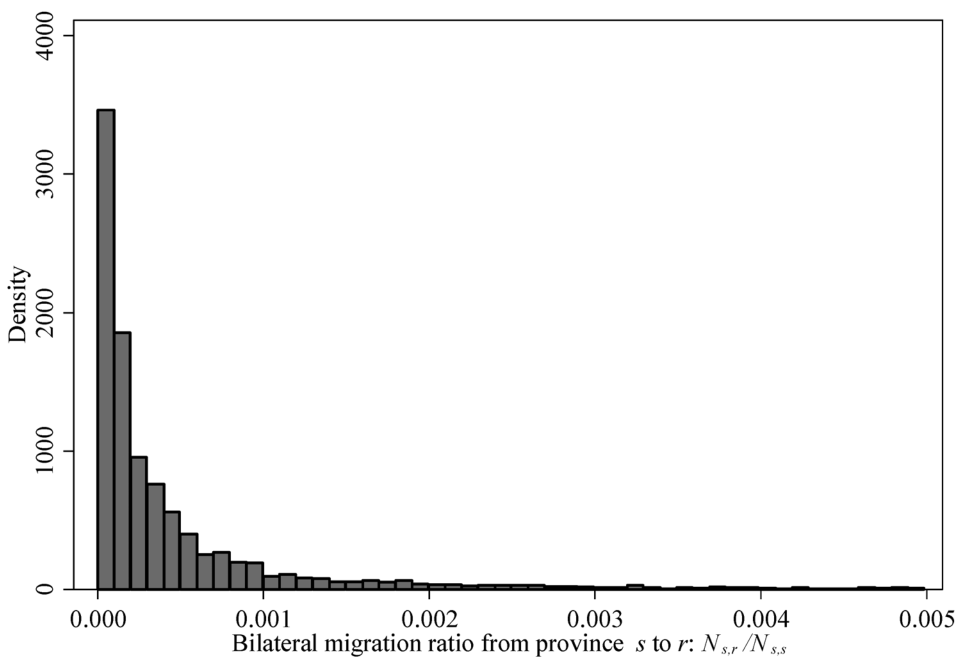
All rights reserved. Copyright © 2018 by the IAEE.

Table 1: Summary statistics for dependent variable: Bilateral Migration Rate

	Observations*	Mean	Standard deviation	Min	Max
Seven census years	5,860	0.00080	0.00252	0	0.05376
All 29 years	24,308	0.00081	0.00252	0	0.05376
Census year:					
1987	756	0.00028	0.00052		
1990	812	0.00039	0.00075		
1995	812	0.00033	0.00074		
2000	870	0.00088	0.00296		
2005	870	0.00092	0.00286		
2010	870	0.00140	0.00393		
2015	870	0.00130	0.00280		

*Unbalanced panel due to Hainan being part of Guangdong before 1988 and Chongqing being part of Sichuan before 1997. Excludes Tibet due to missing data on energy consumption.
30 provinces: Anhui, Beijing, Chongqing, Fujian, Gansu, Guangdong, Guangxi, Guizhou, Hainan, Hebei, Heilongjiang, Henan, Hubei, Hunan, Inner Mongolia, Jiangsu, Jiangxi, Jilin, Liaoning, Ningxia, Qinghai, Shaanxi, Shandong, Shanghai, Shanxi, Sichuan, Tianjin, Xinjiang, Yunnan, Zhejiang.

Figure 1: Histogram of dependent variable



Note: values >0.005 are excluded from this graph.

The dependent variable for the bilateral migration rate is calculated using:

$$n_{s,r,t} = \frac{N_{s,r,t}}{N_{s,t}}, \quad s \neq r \tag{1}$$

where $N_{s,r,t}$ is the number of migrants from the sending province s to the receiving province r in year t and $N_{s,t}$ is mid-year t population N in the sending province s. Note that we do not consider intra-province migration and hence we exclude the cases where $s = r$.

Copyright © 2018 by the IAEE. All rights reserved.

These observations are available across 30 provinces, so for each year there are at most $30^2-30=870$ bilateral migration rates. As described above, some of the censuses are missing for four provinces so we end up with just 5,860 Census observations on the dependent variable to estimate our most basic model.[5]

However, in order to exploit the fact that we also have non-census year observations for all the explanatory variables, we generate another dependent variable where the missing non-census values have been generated by geometric interpolation. This serves to estimate a supplementary set of models for comparison. For every missing observation on $n_{s,r,t}$ we generate the absent values using:

$$n_{s,r,t} = n_{s,r,t-1}R_{s,r,t} \tag{2}$$

where $R_{s,r,t}$ is the geometric growth ratio which can also be defined as $R=1+r$ where r is the growth rate. This yearly geometric ratio is assumed constant between any two non-missing, non-consecutive yearly observations within the same province. In order to calculate these constant geometric ratios for the gap between any two non-missing, non-consecutive years t_2 and t_1, we use the formula:

$$R_{s,r,t_2-t_1} = \left(\frac{n_{s,r,t_2}}{n_{s,r,t_1}}\right)^{1/(t_2-t_1)} \tag{3}$$

For example, if we wanted to calculate the yearly geometric ratios in the years between 1995 and 1990 of any pair of provinces s and r, we would use:

$$R_{s,r,1995-1990} = \left(\frac{n_{s,r,1995}}{n_{s,r,1990}}\right)^{1/5} \tag{4}$$

and then apply equation (2). Using this approach, we generate a dataset across 29 years that, due to missing cases, gives a total of 24,308 observations. In 0.8% of cases this geometric interpolation did not work because the recorded census migration ratio is zero. In this 0.8% of cases we used linear interpolation instead. For example in the 1995 Census the recorded migration rate from Shanghai to Inner Mongolia was 0% and in the 2000 Census it was 0.002%, so the values in-between were resolved by linear interpolation.

Though geometric interpolation is our favoured technique for recovering missing observations on the dependent variable, as a robustness check we also generate an alternative dependent variable based on linear interpolation. Regression results for this alternative dependent variable are presented in tables A3 and A4 of the Appendix.

3.3 The explanatory variables

Table 2 presents the summary statistics for each of the explanatory variables in the three dimensional panel dataset. We report summary statistics on these explanatory variables for the full 1987–2015 sample because they cover the entire sample period and not just the census years. As explained above, the panel is not perfectly balanced because of Hainan being part of Guangdong before 1988 and Chongqing being part of Sichuan before 1997.

5. The maximum number of census observations would have been 6,090 = 870 migration rates × 7 censuses, had there been no missing cases.

All rights reserved. Copyright © 2018 by the IAEE.

Table 2: Summary statistics for three-dimensional panel explanatory variables (Receiving to sending province log ratios, see formula 6)

24,308 observations on:	Mean	Min	Max
ln(*pc*Energy$_r$ /*pc*Energy$_s$)	0	–2.609	2.609
ln(*pc*GRP$_r^{2010pp}$ /*pc*GRP$_s^{2010pp}$)	0	–2.162	2.162
ln(%Secondary$_r$ / %Secondary$_s$)	0	–1.255	1.255
ln(*pc*Female$_r$ /*pc*Female$_s$)	0	–0.098	0.098
ln(*pc*Aged 0–14$_r$ /*pc*Aged 0–14$_s$)	0	–1.229	1.229
ln(*pc*Aged 65+$_r$ /*pc*Aged 65+$_s$)	0	–1.252	1.252

Notes: r and *s* subscripts indicate receiving and sending province respectively.
pc prefixes indicate *per capita* values.
Energy is energy consumption in tonnes of coal equivalent.
GRP is Gross Regional Product.

Table 3: Summary statistics for two-dimensional panel explanatory variables

838 observations on:	Mean	Min	Max
Climate anomalies, see formula (5).			
Anomaly[Temper.]$_s$	0.101	–2.393	3.102
Anomaly[Precip.]$_s$	–0.002	–2.430	3.473
Anomaly[Sunshi.]$_s$	–0.030	–4.154	3.517
Per capita (*pc*) *and percentage* (%) *values used to construct log-ratios or logs:*			
*pc*Energy$_p$	2.059	0.185	8.093
*pc*GRP$_p^{2010pp}$	19200	1921	106838
%Secondary$_p$	0.447	0.190	0.669
*pc*Female$_p$	0.489	0.454	0.513
*pc*Aged 0–14$_p$	0.220	0.076	0.364
*pc*Aged 65+$_p$	0.077	0.030	0.164
Variables used to construct some variables above:			
Temperature (°C annual average)	14.2	3.4	25.4
Precipitation (mm annual total)	886	75	2,679
Sunshine (hours annual total)	2,015	598	3,093
GRP at 2010 provincial prices (in Yuan)	7,902	598	69,178
Population (10,000)	4,104	426	10,849

Notes: p subscripts indicate province, used for sending or receiving province.
s subscripts indicate sending province.
pc prefixes indicate *per capita* values.
Energy is energy consumption in tonnes of coal equivalent.
GRP is Gross Regional Product.

Table 3 presents the summary statistics for the two-dimensional panel variables (hence only 688 observations). These are used either to construct the three-dimensional panel variables in Table 2 or are mapped directly, after taking logs, unto three dimensions for the regressions. Table 3 also includes summary statistics on some of the primitive variables used to construct the other variables. Note that many of the per capita values are based on data in years where 1% or less or the population were surveyed and they are therefore estimates based on sample proportions (e.g. the female population share).

We now provide a description of the procedures and formulae used to construct these explanatory variables.

Copyright © 2018 by the IAEE. All rights reserved.

Just as in Beine and Parsons (2015) we include “climatic factors” in the form of climate anomalies for rainfall and temperature for each sending province. We also have sunshine data at our disposal so we include sunshine anomalies too. For all three variables, these are defined using the following formula:

$$\text{Anomaly}_{s,t} = (C_{s,t} - \overline{C}_s) / s.d.(C_s) \tag{5}$$

where $C_{s,t}$ is any one of three annual climate variable (rainfall, temperature or sunshine) for the sending province *s* in year *t*. The mean value for the entire sample period is $\overline{C}_p$ and $s.d.(C_s)$ is its standard deviation. The formula above is the one typically used in the literature to define climate anomalies and it creates a standardized measure of the direction of change in ‘long-run weather’ and therefore climate patterns.

It is standard practice in the literature to include climate anomalies from the sending country and not to consider climate anomalies in the receiving country. We have not seen this issue discussed elsewhere but the obvious justification is that potential migrants are more likely to be basing their migration decisions on the climate they experience in the province where they might migrate from rather than an analysis of climate reports for the provinces they might migrate to. So our justification is one based on informational asymmetry.

Beine and Parsons (2015, eq.4) include a measure of the log-ratio (see equation 6 below) of the wage rate between the receiving and the sending countries but all their other variables are included as log proportions for both the sending and receiving countries. We on the other hand, use two separate model specifications. In one we include all variables as log-ratios between the sending and receiving province as shown in equation (6). In the other model we include the log-proportions for both the sending and receiving province.

The formula for each log-ratio is simply:

$$\ln(X_r / X_s) \tag{6}$$

where X_r is the proportion (in per capita or percentage values) of the variable in the receiving province and X_s is the equivalent variable for the sending province.

Equation (6) can clearly be rewritten as $\ln(X_r) - \ln(X_s)$ and this gives rise to our second model specification. As we will clearly see by comparing equations (7) and (8), these log-differences are just a restricted versions of the model specification with log rates for both sending and receiving province $\gamma_{11}\ln(X_r) - \gamma_{12}\ln(X_s) = \gamma_1 \ln(X_r / X_s)$ thus imposing the restriction $\gamma_{11} = -\gamma_{12}$.

In a small proportion of cases we have used, as appropriate, linear or geometric interpolation to fill in for missing cases. For some variables, such as female proportions, which are stationary or at most have a linear deterministic trend, we have used simple linear interpolation. For other variables with growth rates, such as energy consumption per capita, we have used geometric interpolation as already described for the dependent variable.

Most of the variables were available for most years from 1987 to 2015 because mostly based on yearly provincial surveys. These provincial surveys only rely on sample proportions and these are used as proxies for the population proportions. In 4.1% of cases data on the female ratio, dependency ratio for children (aged 0 to 14) and dependency ratio for older people (proportion aged 65 plus) are missing because 1988 values are unavailable. We recovered these missing cases by linear interpolation.

Some difficulties emerged when dealing with measures of population size for non-census years. The only variable for which we really needed population size was to calculate energy consumption per capita because only data on total energy consumption are available. However,

All rights reserved. Copyright © 2018 by the IAEE.

the available estimates for provincial population size seemed inconsistent between different official sources.[6] In the end the most reliable way of obtaining consistent provincial population values seemed to be by starting with the most recent data from the China Statistical Yearbook 2016, then adding missing 2003 values using the China Statistical Yearbook 2015. Then filling as many missing population values as possible by taking the ratio of official GRP to official GRP per capita.[7] Finally, any remaining missing population values were recovered using the China Compendium of Statistics 1949–2008.

The resulting population values were used to calculate energy consumption per capita. The resulting energy consumption per capita values had 2.5% of cases missing and thus we used geometric interpolation to recover 1.3% of them.

Real gross regional product (GRP) per capita was generated by dividing nominal GRP by provincial population and then deflating it to 2010 prices using a province-level deflator.[8] Thought official (nominal) GRP per capita was available, we chose to be consistent and to divide our real GRP measure by the same population measure we used to create energy consumption per capita.[9]

The provincial percentage of secondary industry to GRP (%Secondary) was readily available for 1987–2015. We included the secondary sector as a proxy for pollution and industry effects that might indirectly discourage migration even though GRP and energy consumption might directly encourage migration.[10] The secondary sector adequately captures the existence of the manufacturing and assembly industries but does not capture any adverse effects caused by ‘primary’ mining industries. However, primary industries also include a large share of agriculture and fishing in the Chinese economy so we did not use the primary industry share to proxy industry and pollution effects.

In addition to the variables specified above, we also included fixed-effects for sending and receiving provinces in the form of 29 dummies for the 30 sending provinces and 29 dummies for the 30 receiving provinces. Sometimes papers in this literature use a large number of dummies and interaction dummies (dyads) in their regressions, particularly if they do not have many socio-economic variables at their disposal. For example, Beine and Parsons (2015) include fixed effects for the sending countries and dyads for the destination-year interactions. Marchiori et al. (2012) include sending region dummies, time dummies, sending region and time dyads, and an extra set of dummies for Ghana and Nigeria in 1983–1985. We did try adding region-year dyads in our specifications but found that the significance of our explanatory variables collapsed due to the large number of dyads resulting for our particular dataset. We therefore do not include any fixed effects in the time-dimension and just allow the included regressors to explain the temporal variation in migration.

4. RESULTS

In this section we describe the regression model specifications, the estimation methodology and discuss the regression results.

6. Provincial population values for non-census years seemed to be obvious miss-estimates because of sudden, discrete shifts. These intra-census measurement issues are not uncommon for older datasets. We think that they come about because population values for non-census years are generated using population projections which turn out to be imprecise when new census data become available.

7. It is possible that the statistical authorities used some degree of smoothing and linking to generate the population values in non-census years when they generated the official GRP per capita measure.

8. We are grateful to an anonymous reviewer for highlighting the availability of these price indices.

9. Our (nominal) GRP per capita measure was virtually the same as the official one, and identical to it in those years where our population measure had originally come from the ratio of official GRP to official GRP per capita (see above).

10. We are grateful to an anonymous reviewer for suggesting this variable.

Copyright © 2018 by the IAEE. All rights reserved.

4.1 Estimation procedure

We estimate models of bilateral migration based on the following two general specifications. The first of these is based on climate anomalies as previously defined in equation (5) and on the logs-ratios of the socio-economic variables previously defined in equation (6):

$$\ln\left(\frac{N_{s,r,t}}{N_{s,t}}\right) = \sum_{i=1}^{3}\beta_i \text{Anomaly}[C_i]_{s,r,t} + \sum_{j=1}^{6}\delta_j \ln([X_j]_{r,t}/[X_j]_{s,t}) + \alpha_s + \alpha_r + \varepsilon_{s,r,t} \qquad s \neq r \tag{7}$$

where C indexes the three climate anomalies and X indexes the six socio-economic variables and where $r \neq s$ because we are not considering intra-province migration. The second model specification is an unrestricted version of the first where the receiving and sending provide log-demographics are allowed unrestricted coefficients $\delta_j \neq -\gamma_j$:

$$\ln\left(\frac{N_{s,r,t}}{N_{s,t}}\right) = \sum_{i=1}^{3}\beta_i \text{Anomaly}[C_i]_{s,r,t} + \sum_{j=1}^{6}\delta_j \ln([X_j]_{r,t}) + \sum_{j=1}^{6}\gamma_j \ln([X_j]_{s,t}) + \alpha_s + \alpha_r + \varepsilon_{s,r,t} \qquad s \neq r \tag{8}$$

In line with most of the recent research in this field we have chosen to estimate the models using Poisson Pseudo Maximum Likelihood (PPML). This is due to the distribution of the dependent variable and those instances where its value is zero, as seen in Figure 1.

PPML helps overcome three main shortcomings present in other estimation methods. Firstly, OLS on the log-specification illustrated in equations (7) and (8) would mean that many observations would be omitted whenever a zero appears in the dependent variable (4.53% of cases). Though zero migration is unlikely in practice, it can occur as a result of sampling error if two provinces are very far apart and if a region has extremely low migration in practice. The second is that estimation by OLS or by Tobit will lead to inconsistent errors because of zeros among the regressors. This effect is described in Santos-Silva and Tenreyro (2006) in their often-cited paper "The Log of Gravity". Thirdly, the standard Poisson (not the PPML version) does not drop problematic regressors that may lead to non-existence of the maximum likelihood function. This is discussed in another paper by Santos-Silva and Tenreyro (2010) titled "On the Existence of the Maximum Likelihood Estimates in Poisson Regression".

A further incidental feature of the PPML estimator is that it automatically takes the logarithm of the dependent variable. So when the regression command is issued the dependent variable is simply specified as $N_{s,r,t}/N_{s,t}$. To carry out the PPML estimates we use the user-written code -ppml- in Stata by Santos-Silva, J.M.C. and Tenreyro (2010). We do not select any special options when issuing the -ppml- Stata command.

4.2 Estimation results

Table 4 reports the PPML regression results for the parsimonious model specification in equation (7) and Table 5 reports the results for the slightly less parsimonious model specified in equation (8).

In each table the first three regressions are for the sample covering just the seven census years and the last three regressions are for the sample covering all 29 years. If the regression results for the two samples are similar we would expect similar magnitudes for the parameters but much larger *t*-statistics due to the larger regression degrees of freedom for the data covering 29 years.

All rights reserved. Copyright © 2018 by the IAEE.

Table 4: PPML regressions of migration ratio $N_{s,r} / N_{s,s}$ on climate anomalies and socio-economic characteristics based on the parsimonious equation (7) model specification

	Census years only			Census & non-census years		
	(1)	(2)	(3)	(4)	(5)	(6)
Anomaly[Temper.]$_s$	0.167**	0.165**	0.163**	0.134**	0.136**	0.140**
	(4.6)	(4.5)	(4.5)	(8.5)	(8.6)	(9.0)
Anomaly[Precip.]$_s$	0.065*	0.068**	0.065*	–0.013	–0.011	–0.011
	(1.9)	(2.0)	(1.9)	(–0.8)	(–0.7)	(–0.7)
Anomaly[Sunshi.]$_s$	–0.192**	–0.187**	–0.201**	–0.077**	–0.075**	–0.077**
	(–5.2)	(–5.0)	(–5.4)	(–4.9)	(–4.8)	(–4.9)
ln(*pc*Energy$_r$ /	0.391**	0.596**		0.606**	0.792**	
*pc*Energy$_s$)	(2.6)	(4.6)		(8.2)	(11.6)	
ln(*pc*GRP$_r^{2010pp}$ /	0.654**		0.919**	0.646**		1.065**
*pc*GRP$_s^{2010pp}$)	(3.3)		(5.5)	(6.0)		(10.6)
ln(%Secondary$_r$ /	–0.700**	–0.552**	–0.486**	–1.135**	–0.977**	–0.763**
%Secondary$_s$)	(–4.0)	(–3.1)	(–3.2)	(–9.8)	(–8.6)	(–7.6)
ln(*pc*Female$_r$ /	4.332	6.253**	3.723	1.345	2.727**	0.590
*pc*Female$_s$)	(1.3)	(2.0)	(1.2)	(1.0)	(2.1)	(0.5)
ln(*pc*Aged 0–14$_r$ /	–0.892**	–1.160**	–0.811**	–0.619**	–0.789**	–0.583**
*pc*Aged 0–14$_s$)	(–3.4)	(–4.5)	(–3.1)	(–3.6)	(–4.8)	(–3.4)
ln(*pc*Aged 65+$_r$ /	–2.255**	–2.296**	–2.322**	–1.475**	–1.434**	–1.590**
*pc*Aged 65+$_s$)	(–7.4)	(–7.5)	(–7.6)	(–11.7)	(–11.3)	(–12.8)
R^2	0.4159	0.4104	0.4137	0.4448	0.4389	0.4405
Observations	5860	5860	5860	24308	24308	24308

Notes: t statistics in parentheses, * $p < 10\%$, ** $p < 5\%$.
Regressions include fixed effects for sending (α_s) and receiving (α_r) provinces.
2010*pp* superscripts indicate Gross Regional Product (GRP) is indexed at 2010 provincial prices.
r and *s* subscripts indicate receiving and sending provinces.
pc prefixes indicate *per capita* values.

Both the Energy consumption variables and the GRP variables are intended to capture economic or industrial activity in a province as a pull factor for migrants. One possible problem is that these two variables might be highly correlated with one another and indeed the correlation coefficient between ln(*pc*Energy$_r$ /*pc*Energy$_s$) and ln(*pc*GRP$_r^{2010pp}$ /*pc*GRP$_s^{2010pp}$) is 0.6. We therefore have three estimates for each model and dataset, one with both these variables and two with just one of each.

In the parsimonious regressions based on equation (7) and reported in Table 4. The general result with respect to climate anomalies is that an increase in temperatures in a sending province is associated with an increase in migration away from that province. This confirms that long-term relative warming between provinces is a push factor in driving migration. Precipitation anomalies also seem to be an important push factor in driving migration at around the 10% significance level but, curiously, only in the regressions based on census-years data. Interestingly, sunshine anomalies are in all cases significant in explaining migration but with a negative coefficient. This suggests that increased sunshine in a province reduces the push factor driving people away from the province. We do not know if this is as a practical consequence of increased sunshine making a province more productive (e.g. in agriculture) or as a beneficial psychological effect on its residents. In extreme cases, it might even be that there is an inverse relationship between sunshine and smog pollution but the latter is more likely to have already been captured by the variables for energy consumption, GRP, and secondary sector share.

In Table 4 the socio-economic variables are included as log-ratios as specified in equation (7). In these it appears that real GRP per capita ratios and energy consumption per capita ratios all

Copyright © 2018 by the IAEE. All rights reserved.

Table 5: PPML regressions of migration ratio $N_{s,r} / N_{s,s}$ on climate anomalies and socio-economic characteristics based on the less parsimonious equation (8) model specification

	Census years only			Census & non-census years		
	(1)	(2)	(3)	(4)	(5)	(6)
Anomaly[Temper.]$_s$	–0.021	0.003	–0.028	0.029*	0.037**	0.035**
	(–0.5)	(0.1)	(–0.7)	(1.8)	(2.3)	(2.1)
Anomaly[Precip.]$_s$	–0.029	–0.020	–0.033	–0.029*	–0.023	–0.026*
	(–0.9)	(–0.6)	(–1.0)	(–1.9)	(–1.5)	(–1.7)
Anomaly[Sunshi.]$_s$	–0.037	–0.063*	–0.058	–0.017	–0.020	–0.019
	(–1.0)	(–1.7)	(–1.6)	(–1.1)	(–1.3)	(–1.2)
ln(*pc*Energy$_r$)	0.135	0.838**		0.410**	0.916**	
	(0.7)	(5.8)		(4.4)	(12.3)	
ln(*pc*Energy$_s$)	–0.600**	–0.277		–0.616**	–0.349**	
	(–2.7)	(–1.5)		(–5.6)	(–3.9)	
ln(*pc*GRP$_r^{2010pp}$)	0.835**		0.955**	0.598**		0.885**
	(3.7)		(5.2)	(5.0)		(8.4)
ln(*pc*GRP$_s^{2010pp}$)	0.358		–0.015	0.116		–0.279**
	(1.4)		(–0.1)	(0.9)		(–2.5)
ln(%Secondary$_r$)	–0.325	–0.581**	–0.281	–0.685**	–0.942**	–0.428**
	(–1.6)	(–2.8)	(–1.5)	(–5.3)	(–7.3)	(–3.8)
ln(%Secondary$_s$)	0.291	0.237	–0.039	0.447**	0.273*	0.076
	(1.1)	(0.9)	(–0.2)	(2.8)	(1.8)	(0.6)
ln(*pc*Female$_r$)	9.453**	4.824	8.738**	3.707**	1.979	3.160**
	(2.7)	(1.4)	(2.5)	(2.6)	(1.5)	(2.3)
ln(*pc*Female$_s$)	–4.477	–8.641**	–4.202	–2.931**	–4.708**	–2.499*
	(–1.4)	(–2.4)	(–1.2)	(–2.1)	(–3.4)	(–1.8)
ln(*pc*Aged 0–14$_r$)	–1.098**	–1.472**	–0.959**	–0.462**	–0.628**	–0.391**
	(–3.9)	(–5.4)	(–3.5)	(–2.6)	(–3.7)	(–2.1)
ln(*pc*Aged 0–14$_s$)	0.627*	0.242	0.561*	0.112	–0.077	0.093
	(1.8)	(0.8)	(1.7)	(0.6)	(–0.4)	(0.5)
ln(*pc*Aged 65+$_r$)	–2.253**	–1.343**	–2.271**	–0.775**	–0.418**	–0.824**
	(–6.2)	(–4.1)	(–6.4)	(–5.1)	(–2.8)	(–5.4)
ln(*pc*Aged 65+$_s$)	0.013	0.665*	0.056	0.189	0.479**	0.275*
	(0.0)	(1.9)	(0.2)	(1.3)	(3.5)	(1.9)
R^2	0.4528	0.4381	0.4506	0.4650	0.4618	0.4643
Observations	5860	5860	5860	24308	24308	24308

Notes: t statistics in parentheses, * $p < 10\%$, ** $p < 5\%$.
Regressions include fixed effects for sending (α_s) and receiving (α_r) provinces.
2010*pp* superscripts indicate Gross Regional Product (GRP) is indexed at 2010 provincial prices.
r and *s* subscripts indicate receiving and sending provinces.
pc prefixes indicate *per capita* values.

have positive effects on migration ratios. This means that high activity in receiving provinces acts as migration pull factor and low activity in sending provinces acts as migration push factor. It seems that our concerns with the high collinearity between GRP and energy consumption were unfounded and, if anything, regressions (2), (3), (5) and (6) might be suffering from omitted variable bias. The estimated coefficients on secondary industry sector share ratios are all significant and negative if these really do serve as good proxies for pollution and other adverse industry effects.

The estimated coefficients on the demographic variables in Table 4 also seem reasonable. Higher female proportions in receiving provinces attract migration and higher female proportions in sending provinces reduce migration, *ceteris paribus*, though these results are statistically significant in only about half of the regressions. China as a whole has a slightly higher male than female popu-

All rights reserved. Copyright © 2018 by the IAEE.

lation proportion and this is also true on average across all provinces (with the exception of Jiangsu which is almost exactly half-and-half and Tibet which we do not have in our dataset). If migration is male dominated then this mechanism will work toward redressing the provincial gender imbalance. The evidence (Lu and Xia 2016) is mixed but it suggests that internal Chinese migration has been male-dominated over our analysis period. Before 1990, when migration rates were low, migration was female dominated at 54.3%–51.5% due to social reasons such as marriage or joining relatives in other provinces (Duan et al. 2008). However, this changed during the 1990s when migration became male dominated at 55.6% as a result of the increased demand for male workers in fast growing non-agricultural industries. The ratio fell to 51.7% in 2000 and then to 50.2% in 2005 as female migrants increased to meet labour demand from service industries and to reunite with husbands who had previously migrated. However, by 2010, the male migrant ratio rose again rose to 53.2% due to increased demand from the construction industry resulting from the Chinese government's stimulus package (Duan et al. 2013).[11]

The results in Table 4 also suggest that an age imbalance has the effect of reducing migration. A higher relative proportion of children aged 0–14 seems to reduce the tendency to migrate as does a higher relative proportion of people aged 65 or above. Why this should be the case seems obvious, a higher proportion of dependent people causes a reduction in migration, *ceteris paribus*. Of course, migration might arise as a need to support these dependents but the economic incentives affect both those with and without dependents. With respect to the problem of endogeneity, in this case at least, we can say that endogeneity is unlikely to be a cause of the significant results. This is because if these dependency ratios were purely a consequence of migration then we would expect positive parameter estimates on ln(*pc*Aged $0–14_r$ /*pc*Aged $0–14_s$) and ln(*pc*Aged $65+_r$ /*pc*Aged $65+_s$).

The less parsimonious regressions in Table 5 confirm some of the results in Table 4. They also provide a little more insight but are much weaker with respect to the climate anomaly variables. The only coefficients climate variables that achieve 5% statistical significance are for temperature anomalies in the regressions that use both census and non-census year data (regressions 4, 5 and 6). In Table 5 the socio-economic variables are split into separate variables for the sending and receiving province. Splitting the numerator and the denominator tends to confirm the ratios in Table 4 but we suspect that any statistical test for setting all the pairs of parameters to be equal in magnitude though opposite in sign would probably fail given the large number of restrictions (six) this implies. For example, for the energy consumption variables the estimated coefficient for the receiving province is always positive and that for the sending province is always negative. This confirms all the positive estimated coefficients on the energy rations used in Table 4. However, in two cases the coefficients on energy consumption do not achieve 10% statistical significance. Similar patterns are observed for all the pairs of estimated coefficients reported in Table 5 which leads us to believe that these models are over-parameterised when compared to those reported in Table 4. A final note of caution is warranted here with respect to both the energy and GRP variables given they might possess a degree of endogeneity with respect to inward migration also being a driver for the level of these two.

In the Appendix we also report two sets of regression results that act as robustness checks for the results reported above. The first robustness checks are based on using an OLS estimator rather than PPML and the second checks are based on an alternative dependent variable where miss-

11. We need to qualify this discussion by pointing out that the population is slightly male dominated (e.g. 51.4% in 2010 in our data), so if migration is equally male dominated, say at 51.4%, then the provincial gender imbalance will remain the same.

Copyright © 2018 by the IAEE. All rights reserved.

ing non-census years are interpolated by a simple linear method rather than our preferred geometric interpolation.

In appendix tables A1 and A2, OLS regression results are reported for all the same PPML regressions as in tables 4 and 5. To be consistent with the PPML model, the natural logarithm of the dependent variable has been used for OLS. In those cases where migration rate was zero (1.2% for the census-only data and 0.9% for the 29 years data), the missing logarithm of migration has been set to –16 given the lowest value of the log-variable is –15.7. We find that the OLS estimates are very similar to the PPML estimates in term of estimated coefficients and overall fit. The only exception is that when using the census-years data, in the OLS estimates the coefficients on precipitation anomalies achieve higher significance at the 5% level rather than the 10% level for the PPML estimates. We do not know why this should be the case but, if relevant, it suggests that that the effect of precipitation anomalies is more significant that found in the PPML estimates.

Finally, appendix tables A3 and A4 also report PPML estimates based on the alternative dependent variable with missing values interpolated by a simple linear method. Using this alternative interpolation method seems to produce results which are remarkably similar to the main ones in Tables 4 and 5. The only noticeable difference is that the coefficients of determination are very slightly larger for the PPML estimates in the appendix but we have no formal test for to choose between these two specifications.

5. CONCLUSION

The core methodological contribution of this paper has been to investigate the simultaneous effects of climatic and socio-economic factors on migration between Chinese provinces in the period 1987–2015. The choice of variables was largely guided by past research and by data availability but, in contrast to other studies on national bilateral migration, we have also included measures of energy consumption and sunshine. These results matter because they point out clearer drivers for migration and this has implications for economic and energy planning across Chinese provinces.

Our analysis was based on the Poisson Pseudo Maximum Likelihood (PPML) estimator commonly used in this literature. Most of our results (tables 4 and 5) are in line with other research in this field. We begin by discussing the results on the climate variables and follow with a discussion of the results on the energy and socio-economic variables.

Increased temperatures and precipitation (long-run anomalies) in a province are associated with increased migration away from the province. Thanks to our particular data sources we were also able to include long-run sunshine anomalies and found the new result that increased sunshine is associated with lower migration away from a province. We speculate that these results arise because, given the general rise in temperatures, any additional rises are detrimental. However, once temperature is controlled for, increases in sunshine reduce emigration and therefore greater sunshine appears to be beneficial. We remain unsure as to why the results on precipitation are (borderline) significant. In the literature precipitation anomalies are typically found to be insignificant in driving migration. We do not know if our precipitation result is important or just an isolated model anomaly.

The effects of energy and socio-economic characteristics also seem reasonable. Relatively high energy consumption per capita seems to be a pull factor for migrants into a province in the same way as relatively high Gross Regional Product (GRP) per capita seems to be. Including both Energy and GRP in our estimated models does not seem to cause collinearity problems. However, we are aware that there might be some degree of endogeneity with respect to the energy and GRP

All rights reserved. Copyright © 2018 by the IAEE.

measures, and feel that future research could focus on this as a robustness check. We suspect that limited availability of instrumental variables will prevent an instrumental variable approach but perhaps a lagged instrumenting variable approach such as the Generalised Method of Moments is viable.

The demographic variables too seem important in driving migration. A higher female population proportion in the sending province seems to discourage migration. We speculate that this is due to the slight female shortage in the Chinese population and the fact that migration during the period 1987–2015 is very slightly male dominated. Finally, higher dependency ratios from those aged 0–14 and those aged 65 plus, seem to discourage migration. Though it is true that much migration is caused by the need to support dependents in the form of remittances, some of this is already controlled for by the inclusion of energy and GNP variables in our models. Our results therefore reflect the fact that *ceteris paribus* dependents are better supported by remaining in the home province. We are not troubled by the potential for endogeneity in the case of the youth dependency ratio and older dependency ratio because the migration of middle-aged people from a province would cause the proportion of dependents to increase, and in our results the relevant coefficients are negative. So if endogeneity is an issue, the true negative dependency ratio effects are even greater than in our estimates.

APPENDIX: ALTERNATIVE MODEL ESTIMATES

Table A1: OLS regressions of migration ratio ln($N_{s,r}$ /$N_{s,s}$) on climate anomalies and socio-economic characteristics based on the parsimonious equation (7) model specification

	Census years only			Census & non-census years		
	(1)	(2)	(3)	(4)	(5)	(6)
Anomaly[Temper.]$_s$	0.333**	0.333**	0.328**	0.175**	0.174**	0.175**
	(14.4)	(14.4)	(14.2)	(19.3)	(19.2)	(19.3)
Anomaly[Precip.]$_s$	0.130**	0.133**	0.128**	0.013	0.014	0.013
	(6.3)	(6.4)	(6.2)	(1.4)	(1.5)	(1.4)
Anomaly[Sunshi.]$_s$	–0.271**	–0.264**	–0.279**	–0.088**	–0.086**	–0.087**
	(–12.6)	(–12.4)	(–13.0)	(–9.9)	(–9.7)	(–9.8)
ln(*pc*Energy$_r$ /	0.263**	0.327**		0.296**	0.322**	
*pc*Energy$_s$)	(2.9)	(3.9)		(6.6)	(7.5)	
ln(*pc*GRP$_r^{2010pp}$ /	0.271**		0.408**	0.131*		0.265**
*pc*GRP$_s^{2010pp}$)	(2.0)		(3.2)	(1.9)		(4.1)
ln(%Secondary$_r$ /	–0.265**	–0.194	–0.102	–0.398**	–0.358**	–0.185**
%Secondary$_s$)	(–2.1)	(–1.6)	(–0.9)	(–5.5)	(–5.2)	(–2.9)
ln(*pc*Female$_r$ /	–2.787*	–1.983	–3.397**	1.129*	1.413**	0.646
*pc*Female$_s$)	(–1.8)	(–1.3)	(–2.2)	(1.7)	(2.2)	(1.0)
ln(*pc*Aged 0–14$_r$ /	–0.103	–0.230	–0.011	–0.303**	–0.351**	–0.246**
*pc*Aged 0–14$_s$)	(–0.6)	(–1.3)	(–0.1)	(–3.6)	(–4.3)	(–2.9)
ln(*pc*Aged 65+$_r$ /	–0.503**	–0.531**	–0.532**	–0.237**	–0.232**	–0.268**
*pc*Aged 65+$_s$)	(–2.8)	(–2.9)	(–2.9)	(–3.2)	(–3.2)	(–3.7)
R^2	0.3893	0.3889	0.3885	0.3960	0.3959	0.3949
Observations	5860	5860	5860	24308	24308	24308

Notes: t statistics in parentheses, * p < 10%, ** p < 5%.
Regressions include fixed effects for sending (α_s) and receiving (α_r) provinces.
2010*pp* superscripts indicate Gross Regional Product (GRP) is indexed at 2010 provincial prices.
r and *s* subscripts indicate receiving and sending provinces.
pc prefixes indicate *per capita* values.

Copyright © 2018 by the IAEE. All rights reserved.

Table A2: OLS regressions of migration ratio ln($N_{s,r}$ /$N_{s,s}$) on climate anomalies and socio-economic characteristics based on the less parsimonious equation (8) model specification

	Census years only			Census & non-census years		
	(1)	(2)	(3)	(4)	(5)	(6)
Anomaly[Temper.]$_s$	0.055**	0.060**	0.057**	0.011	0.015*	0.007
	(2.4)	(2.6)	(2.5)	(1.3)	(1.7)	(0.8)
Anomaly[Precip.]$_s$	0.020	0.023	0.025	–0.012	–0.010	–0.012
	(1.0)	(1.2)	(1.3)	(–1.4)	(–1.2)	(–1.4)
Anomaly[Sunshi.]$_s$	–0.014	–0.018	–0.015	–0.014*	–0.015*	–0.013
	(–0.7)	(–0.9)	(–0.7)	(–1.8)	(–1.9)	(–1.5)
ln(*pc*Energy$_r$)	0.526**	0.636**		0.503**	0.649**	
	(4.9)	(6.8)		(9.4)	(14.1)	
ln(*pc*Energy$_s$)	–0.101	–0.023		–0.097*	0.033	
	(–0.9)	(–0.2)		(–1.8)	(0.7)	
ln(*pc*GRP$_r^{2010pp}$)	0.192		0.471**	0.193**		0.439**
	(1.4)		(3.8)	(2.9)		(7.2)
ln(*pc*GRP$_s^{2010pp}$)	0.053		0.016	0.112*		0.092
	(0.4)		(0.1)	(1.7)		(1.5)
ln(%Secondary$_r$)	–0.158	–0.169	0.183	–0.456**	–0.512**	–0.060
	(–1.0)	(–1.2)	(1.4)	(–5.4)	(–6.3)	(–0.8)
ln(%Secondary$_s$)	0.345**	0.296**	0.285**	0.274**	0.193**	0.237**
	(2.3)	(2.0)	(2.1)	(3.2)	(2.3)	(3.2)
ln(*pc*Female$_r$)	–13.379**	–13.903**	–13.657**	–6.233**	–6.770**	–6.724**
	(–7.6)	(–8.2)	(–7.8)	(–7.8)	(–8.7)	(–8.4)
ln(*pc*Female$_s$)	–9.621**	–10.535**	–8.621**	–7.490**	–8.233**	–6.976**
	(–5.4)	(–6.2)	(–4.9)	(–9.3)	(–10.5)	(–8.7)
ln(*pc*Aged 0–14$_r$)	–0.847**	–0.976**	–0.756**	–0.541**	–0.673**	–0.527**
	(–4.1)	(–5.1)	(–3.7)	(–5.7)	(–7.5)	(–5.6)
ln(*pc*Aged 0–14$_s$)	–0.086	–0.157	–0.208	0.077	–0.025	–0.026
	(–0.4)	(–0.8)	(–1.0)	(0.8)	(–0.3)	(–0.3)
ln(*pc*Aged 65+$_r$)	–0.436**	–0.296	–0.518**	–0.057	0.084	–0.135
	(–2.1)	(–1.5)	(–2.5)	(–0.7)	(1.0)	(–1.6)
ln(*pc*Aged 65+$_s$)	0.114	0.277	0.128	0.130	0.276**	0.106
	(0.5)	(1.4)	(0.6)	(1.5)	(3.4)	(1.2)
R^2	0.4998	0.4994	0.4977	0.5069	0.5063	0.5051
Observations	5860	5860	5860	24308	24308	24308

Notes: t statistics in parentheses, * $p < 10\%$, ** $p < 5\%$.
Regressions include fixed effects for sending (α_s) and receiving (α_r) provinces.
2010*pp* superscripts indicate Gross Regional Product (GRP) is indexed at 2010 provincial prices.
r and *s* subscripts indicate receiving and sending provinces.
pc prefixes indicate *per capita* values.

All rights reserved. Copyright © 2018 by the IAEE.

Table A3: PPML regressions of migration ratio $N_{s,r} / N_{s,s}$ (linearly interpolated for non-census years) on climate anomalies and socio-economic characteristics based on the parsimonious equation (7) model specification

	Census & non-census years		
	(1)	(2)	(3)
Anomaly[Temper.]$_s$	0.138**	0.140**	0.144**
	(8.8)	(8.9)	(9.3)
Anomaly[Precip.]$_s$	–0.010	–0.008	–0.008
	(–0.6)	(–0.5)	(–0.5)
Anomaly[Sunshi.]$_s$	–0.080**	–0.078**	–0.080**
	(–5.1)	(–5.0)	(–5.1)
ln(pcEnergy$_r$ /pcEnergy$_s$)	0.593**	0.780**	
	(8.1)	(11.6)	
ln(pcGRP$_r^{2010pp}$ /pcGRP$_s^{2010pp}$)	0.648**		1.057**
	(6.2)		(10.7)
ln(%Secondary$_r$ /%Secondary$_s$)	–1.108**	–0.949**	–0.744**
	(–9.7)	(–8.5)	(–7.5)
ln(pcFemale$_r$ /pcFemale$_s$)	1.601	2.988**	0.863
	(1.2)	(2.4)	(0.7)
ln(pcAged 0–14$_r$ /pcAged 0–14$_s$)	–0.573**	–0.743**	–0.538**
	(–3.4)	(–4.7)	(–3.2)
ln(pcAged 65+$_r$ /pcAged 65+$_s$)	–1.402**	–1.362**	–1.516**
	(–11.3)	(–10.9)	(–12.3)
Constant	–6.840**	–6.842**	–6.844**
	(–103.6)	(–103.4)	(–103.5)
R^2	0.4496	0.4438	0.4453
Observations	24308	24308	24308

Notes: t statistics in parentheses, * $p < 10\%$, ** $p < 5\%$.
Regressions include fixed effects for sending (α_s) and receiving (α_r) provinces.
2010*pp* superscripts indicate Gross Regional Product (GRP) is indexed at 2010 provincial prices.
r and s subscripts indicate receiving and sending provinces.
pc prefixes indicate *per capita* values.

Copyright © 2018 by the IAEE. All rights reserved.

Table A4: PPML regressions of migration ratio $N_{s,r}/N_{s,s}$ (linearly interpolated for non-census years) on climate anomalies and socio-economic characteristics based on the less parsimonious equation (8) model specification

	Census & non-census years		
	(1)	(2)	(3)
Anomaly[Temper.]$_s$	0.033**	0.042**	0.039**
	(2.0)	(2.5)	(2.4)
Anomaly[Precip.]$_s$	–0.025*	–0.020	–0.023
	(–1.7)	(–1.3)	(–1.5)
Anomaly[Sunshi.]$_s$	–0.020	–0.023	–0.022
	(–1.3)	(–1.5)	(–1.5)
ln(*pc*Energy$_r$)	0.360**	0.890**	
	(3.9)	(12.1)	
ln(*pc*Energy$_s$)	–0.619**	–0.344**	
	(–5.8)	(–3.9)	
ln(*pc*GRP$_r^{2010pp}$)	0.631**		0.884**
	(5.4)		(8.6)
ln(*pc*GRP$_s^{2010pp}$)	0.116		–0.275**
	(0.9)		(–2.5)
ln(%Secondary$_r$)	–0.670**	–0.935**	–0.452**
	(–5.3)	(–7.4)	(–4.1)
ln(%Secondary$_s$)	0.413**	0.231	0.033
	(2.7)	(1.5)	(0.2)
ln(*pc*Female$_r$)	4.094**	2.286*	3.533**
	(2.9)	(1.7)	(2.6)
ln(*pc*Female$_s$)	–2.825**	–4.704**	–2.437*
	(–2.1)	(–3.4)	(–1.8)
ln(*pc*Aged 0–14$_r$)	–0.392**	–0.568**	–0.313*
	(–2.3)	(–3.4)	(–1.8)
ln(*pc*Aged 0–14$_s$)	0.113	–0.080	0.102
	(0.6)	(–0.4)	(0.5)
ln(*pc*Aged 65+$_r$)	–0.718**	–0.341**	–0.754**
	(–4.8)	(–2.3)	(–5.0)
ln(*pc*Aged 65+$_s$)	0.174	0.478**	0.262*
	(1.2)	(3.6)	(1.8)
Constant	–14.583**	–7.759**	–12.705**
	(–8.8)	(–5.7)	(–10.1)
R^2	0.4702	0.4660	0.4692
Observations	24308	24308	24308

Notes: t statistics in parentheses, * $p < 10\%$, ** $p < 5\%$.
Regressions include fixed effects for sending (α_s) and receiving (α_r) provinces.
2010*pp* superscripts indicate Gross Regional Product (GRP) is indexed at 2010 provincial prices.
r and *s* subscripts indicate receiving and sending provinces.
pc prefixes indicate *per capita* values.

All rights reserved. Copyright © 2018 by the IAEE.

ACKNOWLEDGMENTS

This research has received Ethics Research Approval from the University of Birmingham under reference number ERN_17–1539. We are grateful to an anonymous reviewer for various suggestions which are credited in the text. Any errors remain our own.

REFERENCES

Backhaus et al. (2015). "Do climate variations explain bilateral migration? A gravity model analysis" *IZA Journal of Migration* 4: 3. https://doi.org/10.1186/s40176-014-0026-3.

Barnett, J.R. and M. Webber (2010). "Accommodating migration to promote adaptation to climate change" *World Bank Policy Research Working Paper Series*

Beine and Parsons (2015). "Climatic Factors as Determinants of International Migration" *The Scandinavian Journal of Economics*.

Benson, L., K. Peterson, and J. Stein (2006). "Anasazi (pre-Columbian Native-American) migrations during the middle–12th and late–13th centuries - Were they drought induced?" *Climate Change* 83: 187–213. https://doi.org/10.1007/s10584-006-9065-y.

Black, R., D. Kniveton, and K. Schmidt-Verkerk (2007). "Migration and climate change: Towards an integrated assessment of sensitivity." *Environment and Planning A*, 43: 431–450. https://doi.org/10.1068/a43154.

Burrows, K. and P. L. Kinney (2016). "Exploring the Climate Change, Migration and Conflict." *International Journal of Environmental Resources and Public Health* 13: 443. https://doi.org/10.3390/ijerph13040443.

Cai, W., S. Borlace, M. Lengaigne, P. van Rensch, M. Collins, G. Vecchi, A. Timmermann, A. Santoso, M. J. McPhaden, L. W., M. H. England, G. Wang, E. Guilyardi, and F-F. Jin (2014). "Increasing frequency of extreme El Niño events due to greenhouse warming." *Nature Climate Change* 4: 111–116. https://doi.org/10.1038/nclimate2100.

Cattaneo and Peri (2015). "The migration response to increasing temperatures." NBER working paper. https://doi.org/10.3386/w21622.

Costello, A., M. Abbas, A. Allen, S. Ball, S. Bell, R. Bellamy, S. Friel, N. Groce, A. Johnson, and M. Kett (2009). "Managing the health effects of climate change." *The Lancet* 373: 1693–1733. https://doi.org/10.1016/S0140-6736(09)60935-1.

China Compendium of Statistics 1949–2008 (2010). National Bureau of Statistics of China. ISBN: 9787503758942.

China Energy Statistical Yearbook 2013 (2014). ISBN: 9787503770227, Publisher: China Statistics Press

China Statistical Yearbook (2015). Publisher: China Statistics Press. ISBN: 978–7503776380

Christian Aid (2007). Human Tide: The Real Migration Crisis, https://www.christianaid.org.uk/Images/human-tide.pdf , Accessed 15/8/2015.

Duan, C., G. Yang, F. Zhang, and X. Lu (2008). "Nine Trends of China's Population Flow since the Reform and Opening Up." *Population Research* 32(6): 30–43.

Duan, C., Y. Yuan, and J. Guo (2013). "The Latest Features of China's Floating Population." *Northwest Population* 34(6): 1–12.

Dun, O. and F. Gemenne (2008). "Defining 'environmental migration'." *Forced Migration Review* 31: 10–11.

Feng, S. and M. Oppenheimer (2012). "Applying statistical models to the climate–migration relationship." *Proceedings of the National Academy of Sciences* 109 (43). https://doi.org/10.1073/pnas.1212226109.

Government Office for Science (2011). "Foresight: Migration and Global Environmental Change." Final Project Report, London. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/287717/11–1116-migration-and-global-environmental-change.pdf, Accessed 6/1/2016.

Gray, C. (2011). "Soil quality and human migration in Kenya and Uganda." *Global Environmental Change* 21(2): 421–430. https://doi.org/10.1016/j.gloenvcha.2011.02.004.

Herrerias, M. J. and E. Girardin (2013). "Seasonal Patterns of Energy in China." *The Energy Journal* 34: 169–198. https://doi.org/10.5547/01956574.34.2.6.

Herrerias, M. J., R. Joyeux, and E. Girardin (2013). "Short- and long-run causality between energy consumption and economic growth: Evidence across regions in China." *Applied Energy* 112: 1483–1492. https://doi.org/10.1016/j.apenergy.2013.04.054.

Intergovernmental Panel on Climate Change (2007). Fourth Assessment Report, https://www.ipcc.ch/report/ar4/

Intergovernmental Panel on Climate Change (2015). Fifth Assessment Report (synthesis), http://www.ipcc.ch/report/ar5/syr/

Joseph and Wodon (2013). "Is Internal Migration in Yemen Driven by Climate or Socio-economic Factors?" *Review of International Economics* 21(2): 295–310. https://doi.org/10.1111/roie.12037.

Copyright © 2018 by the IAEE. All rights reserved.

Jones, P. D., D. E. Parker, T. J. Osborn, and K. R. Briffa (2009). "Global and Hemispheric Temperature Anomalies-Land and Marine Instrumental Records." In *Trends: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, OakRidge National Laboratory, U.S. Department of Energy, Oak Ridge, TN. https://doi.org/10.3334/CDIAC/cli.002

Kahn, M.E. (2014). "Climate change adaptation: lessons from urban economics" *NBER working paper series*, Working Paper 20716, http://www.nber.org/papers/w20716.

Klaiber, H.A. (2014). "Migration and Household Adaptation to Climate: A Review of Empirical Research." *Energy Economics* 46: 539–547. https://doi.org/10.1016/j.eneco.2014.04.001.

Lu, M., and Y. Xia (2016). "Migration in the People's Republic of China." ADBI Working Paper 593, Tokyo: Asian Development Bank Institute. www.adb.org/publications/migration-people-republic-china/. https://doi.org/10.2139/ssrn.2838116.

Mallick, B. and J. Vogt (2012). "Cyclone, coastal society and migration: Empirical evidence from Bangladesh." *International Development Review* 34: 217–240. https://doi.org/10.3828/idpr.2012.16.

Marchiori, L., J-F. Maystadt, and I. Schumacher (2012). "The impact of weather anomalies on migration in sub-Saharan Africa." *Journal of Environmental Economics and Management* 63: 355–374. https://doi.org/10.1016/j.jeem.2012.02.001.

Massey, D. (2015). "Theory of migration." In *International Encyclopaedia of the Social and Behavioral Sciences*, 2nd edition by Wright, J.D., Ed.; University of Central Florida: Orlando, FL, USA, 467–471. https://doi.org/10.1016/B978-0-08-097086-8.31119-9.

Morton, A., P. Boncour, and F. Lackzo (2008). "Human security policy challenges." *Forced Migration Review* 31: 5–8.

Mortreux, C. and J. Barnett (2009). "Climate change, migration and adaptation in Funafuti, Tuvalu." *Global Environmental Change* 19(1): 105–112. https://doi.org/10.1016/j.gloenvcha.2008.09.006.

Naudé, W. (2010). "The determinants of migration from Sub-Saharan African countries." *Journal of African Economies* 19(3): 330–356. https://doi.org/10.1093/jae/ejq004.

Noy, I. (2009). "The macroeconomic consequences of disasters." *Journal of Development Economics* 88(2): 221–231. https://doi.org/10.1016/j.jdeveco.2008.02.005.

Parry, M., O. Canziani, J. Palutikof, P. J. van der Linden, and C. Hanson (2007). Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK, 433–467.

Reuveny, R. and W. H. Moore (2009). "Does environmental degradation influence migration? Emigration to developed countries in the late 1980s and 1990s." *Social Science Quarterly* 90(3): 461–479. https://doi.org/10.1111/j.1540-6237.2009.00569.x

Reuveny, R. (2007). "Climate change-induced migration and violent conflict." *Political Geography* 26: 656–673. https://doi.org/10.1016/j.polgeo.2007.05.001

Saldaña-Zorrilla, S. O. and K. Sandberg (2009). "Spatial econometric model of natural disaster impacts on human migration in vulnerable regions of Mexico." *Disasters* 33: 591–607. https://doi.org/10.1111/j.1467-7717.2008.01089.x.

Santos-Silva and Tenreyro (2006). "The Log of Gravity." *Review of Economics and Statistics.*

Santos-Silva, J. M. C. and Silvana Tenreyro (2010). "On the Existence of the Maximum Likelihood Estimates in Poisson Regression." *Economics Letters* 107(2): 310–312. https://doi.org/10.1016/j.econlet.2010.02.020.

Stern, N. (2007). *The Economics of Climate Change: The Stern Review*, Cambridge University Press: Cambridge, UK. https://doi.org/10.1017/CBO9780511817434.

United Nations Framework Convention on Climate Change (1992). United Nations, https://unfccc.int/resource/docs/convkp/conveng.pdf.

US Department of Energy (2014). "The Water-Energy Nexus: Challenges and Opportunities" https://www.energy.gov/under-secretary-science-and-energy/downloads/water-energy-nexus-challenges-and-opportunities.

Warner, K. (2012). "Human migration and displacement in the context of adaptation to climate change: the Cancun Adaptation Framework and potential for future action." *Environment and Planning C: Government and Policy* 30: 1061–1077. https://doi.org/10.1068/c1209j.

Wester, J.A. and R. H. Clarke (2017). "Insurance companies should collect a carbon levy." *Nature* 549(7671).

World Bank (2010). World Development Report 2010: Development and Climate Change, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/4387.

All rights reserved. Copyright © 2018 by the IAEE.