**原始数据来源说明**

在模拟中，我们用到了如下数据：

**国家间贸易数据。**我们使用中国海关数据得到各省份在各行业通过一般或者加工贸易方式开展的进出口金额。我们利用OECD的跨国贸易数据库，得到2005年各行业内国家间的贸易金额。根据这一数据，我们构建了代码中用到的trade\_matrix变量（地区间贸易矩阵的国家间贸易部分）。

**国内省份间贸易数据。**本文主要采用刘卫东等（2012）计算的2007年各行业内省份间的贸易流向，并使用中国2005年到2007年间各行业的产出增长幅度进行调整，从而得到2005年各行业省份间的贸易流向。根据这一数据，我们构建了代码中用到trade\_matrix变量（地区间贸易矩阵的国内省份间贸易份额）。

**关税数据。**我们使用UNCTAD的TRAINS关税数据库计算得到1990年和2005年国家和国家之间的各行业的关税水平。关税数据库提供了6位数HS代码层面的关税水平。我们利用贸易额作为权重，将HS代码层面的关税水平加权平均得到国家间各ISIC行业的关税水平。其中，由于我们考虑了中国厂商的不同贸易方式，根据海关规定，我们设定国内加工贸易的中间投入品不需要关税，只有通过一般贸易方式进口的中间投入品需要关税。根据这一数据，我们构建代码中用到的2005年国家间各行业关税tariffrate2005.mat以及1990年关税水平tariffrate1990.mat。

**产出数据。**我们使用OECD数据得到各个外国各行业的总产出。我们使用中国的区域投入产出表可以得到中国各省份各行业的产出水平。为了区分不同贸易方式下的产出，我们通过海关出口数据得到每个省份各行业加工贸易产出，进而将每个省份各行业的产出水平扣除加工贸易产出后，得到每个省份各行业的在一般贸易方式下的产出水平。利用这一数据，我们构建了代码中用到的各国家各行业的产出水平output变量。

**劳动力分布数据。**2005年人口抽样调查提供了每个劳动力的户籍所在地、居住省份、行业以及工资。我们根据户籍所在省份划分成30个组，并计算得到每组劳动力的总数，和每组劳动者目前居住省份和所从事行业的分布以及相对应的平均工资水平。利用这一数据，我们构建了代码中用到的中国各组劳动者的初始居住省份-行业移民矩阵china\_pi\_1（非大学劳动力）和china\_pi\_2（大学毕业生），他们对应的工资总收入水平china\_pay\_1和china\_pay\_2变量，以及相应的各组人数low\_skill和high\_skill变量。

对于外国，我们根据IPUMS–International和Luxembourg Income Study数据库得到每个外国各行业的劳动力总数以及平均工资水平。根据这一数据，我们得到代码中用到的外国各行业劳动力分布w\_pi变量和总体收入水平world\_pay变量。

**投入产出表。**我们利用中国和OECD公布的投入产品表得到各地区的劳动和中间投入品占企业生产成本比例，以及各行业消费占最终消费的比重。这分别对应于代码中用到的sector\_matrix变量（投入产出矩阵）以及c\_share变量（各行业消费比重）。

在模型中，我们还用到了一些参数。反映了国际间贸易额对于贸易成本的弹性，我们取为4。表示同一产品在两种贸易方式下的生产率的相关系数，我们采用为0.7。我们采用移民工资弹性，以及行业集聚效应参数。正文中介绍了这些参数取值的来源。