Review of Economic Studies (2007) **74**, 477–506 0034-6527/07/00170477$02.00

Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium

E. D. GOULD
Department of Economics,
Hebrew University, CEPR, and IZA

First version received January 2004; *final version accepted May* 2006 (*Eds.*)

Workers earn higher wages in cities vs. rural areas. This gap could arise because cities make workers more productive, or it could be the result of a non-random selection of workers into cities based on their ability and their endogenous history of career choices. To untangle these issues, this paper estimates a dynamic programming model, which embeds the choice of residing in a city or rural area within a model of career choices over time. After controlling for all the sources of selection and endogeneity, the estimates indicate that a given worker does earn more in the city for white-collar work, but not for blue-collar work. In addition, city work experience is found to be worth more than rural work experience in the rural area for white-collar work, but not for blue-collar work. These results support the interpretation that cities make white-collar workers more productive and suggest that workers may consider moving to the city not only in terms of locational choice, but also as a form of human capital investment.

1. INTRODUCTION

Average wages are higher in cities than in rural areas—a phenomenon typically known as the "urban wage premium". As Glaeser and Mare (2001) argue, understanding why workers earn more in cities goes to the heart of the central question in urban economics: why do cities exist?[1] If firms pay higher wages in cities for the same quality of workers, it is hard to imagine why firms would locate themselves in cities and why all workers do not move to cities. The latter question is easier to explain if certain people have a distaste for certain characteristics of big cities (the congestion, pollution, higher prices, and other non-market amenities) and therefore, require a wage premium as compensation (Roback, 1982). The question about why firms agree to pay higher wages in large cities is harder to explain. The most likely reason is that the value of the marginal product of labour is higher in cities, which may be due to several factors. First, product prices may be higher in cities due to increased local demand or production costs may be lower, such as transportation costs from nearby suppliers. However, many firms do not rely on local suppliers and local demand, and still they choose to be in big cities and pay higher wages. Therefore, it seems more likely that wages are higher in cities due to higher levels of productivity, which may occur for several reasons: (a) workers in cities have higher ability, (b) living in a city helps foster the accumulation of human capital through social interactions (Glaeser, 1999), (c) living in a city helps match workers to the most suitable job, or (d) firms in cities learn from one another (Rauch, 1993; Lucas, 2001).

1. The motivation for this paper presented in the next few paragraphs borrows from the motivation discussed in Glaeser and Mare (2001).

All of these explanations are consistent with observed wages being higher in cities, but many of them differ in their empirical predictions which, in principle, can be tested. For example, higher ability workers may tend to live in cities because of their taste for certain amenities of the city, such as cultural activities, which may be more profitable and therefore more prevalent in higher density areas. If true, estimates of the urban wage premium should be zero after controlling for unobserved ability. If firms pay more in cities due to higher local demand or lower production costs, then estimates of the urban wage premium should be positive even after controlling for unobserved ability, which implies that workers moving to the city should experience an immediate wage increase, and workers leaving the city should experience an immediate wage reduction (*i.e.* an "intercept" effect). Similar predictions hold if firms are more productive in cities because they learn from other firms in close proximity, if workers find better matches in cities, or if workers require a compensating wage premium to live in a city. However, these theories do not predict that workers are able to acquire extra human capital in the city, which is transferable back to the rural area. If a denser population makes workers more productive somehow, perhaps through knowledge spillovers, then workers in cities should be able to at least partially transfer these productivity gains back to the rural sector.

The key distinguishing predictions of these models are whether a city wage premium exists after controlling for differences in the types of workers across locations, and if so, whether cities generate a positive effect on wages, which disappears once a worker moves away, or whether workers are able to transfer their human capital gains from the city to the rural area. By examining each of these issues, this paper contributes not only to our understanding of why cities exist, but also to the understanding of a potentially important source of human capital which has not been acknowledged in the literature. If human capital acquired in the city is transferable back to the rural area, then "living in the city" should be considered not just a decision about location of residence, but also as an investment in human capital. This type of investment may be costly, particularly for certain individuals who have a distaste for living in an urban area, but moving to a city temporarily could be a worthwhile human capital investment in the same manner that it pays to invest in other costly forms of human capital such as education, internships, training, etc. Therefore, determining whether working in a city fosters human capital accumulation has implications regarding how individuals balance their careers, human capital decisions, and location of residence over time.

The empirical strategy is to develop and estimate a structural model, which explicitly accounts for the non-random allocation of workers who grow up in cities and the self-selection of workers who choose to move to and from cities over time. In addition, the model embeds the decision to live in a city or in a rural area within a dynamic model of schooling and career choices. This dynamic interaction explicitly allows for individuals to consider any benefits that working in the city might bring to them in the future in terms of their wages and productivity in both the city and the rural area. The model endogenizes the location of residence decision with decisions in each period about labour force participation, going to school, and choosing between the white-collar and blue-collar occupations. All of these decisions are likely to be related, since they are all very correlated with one another in both the levels and transitions over time. For example, as shown in the next section, living in the city is highly correlated with education and with working in the white-collar occupation, and moving to the city is highly correlated with becoming a white-collar worker. Therefore, to account for the correlation of all these decisions, which may generate a spurious relationship between living in a city and higher wages, the model controls for unobserved heterogeneity across people and explicitly models the individual's decision in every period over occupational choice, schooling, and labour force participation jointly with the decision of whether to live in a city or rural area.

After controlling for all of the potential sources of selection bias, the results show that the observed 10·8% wage gap at the age of 30 between blue-collar workers in the city and rural area is reduced to 1·2%. For the white-collar sector, the observed 17·5% wage gap is reduced to 11·5%. Therefore, the endogenous selection of individuals into occupations and locations over time can almost completely explain why blue-collar workers earn more in the city vs. the rural area, but explains only 34% of the observed wage gap between white-collar workers in the city and rural area. In addition, the estimates show that work experience acquired in the city receives an extra reward in the white-collar sector in the rural area, but this is not the case for blue-collar work. Five years of city work experience increase the wage of a white-collar worker in the rural area by 8·5% compared to a similar worker without any work experience in the city.

Thus, the estimates indicate that while blue-collar workers earn more in cities because of endogenous selection, a true city wage premium does exist for the white-collar sector and increases with experience. In addition, the results suggest that human capital gains associated with working in the city are transferable to the rural area, but only for the white-collar occupation. Although there are many possible explanations for why white-collar work is paid more in cities after controlling for all the sources of selection and endogeneity, the finding that human capital gains associated with working in the city are rewarded in the rural area provides support for the interpretation that cities make white-collar workers more productive. These results also suggest that workers may consider moving to the city not only in terms of locational choice, but also as a form of human capital investment.

There are several empirical papers on the urban wage premium in the literature, but there is no paper that empirically examines the joint determination of city–rural status and career decisions for men. However, the empirical literature has examined this issue in a non-structural setting (see Glaeser and Mare, 2001; Wheaton and Lewis, 2002). These studies typically try to measure the urban wage premium by comparing the wages of observationally equivalent workers in urban and rural areas, although it is problematic to assume that workers in urban and rural areas do not differ in unobservable characteristics. Glaeser and Mare (2001) control for individual heterogeneity by using panel data and including individual fixed effects in the analysis. Their analysis yields three main results: (1) workers who move to the city experience a wage increase, (2) workers who leave the city do not suffer a wage decline, and (3) the "city effect" appears to increase over time (*i.e.* there is a slope and an intercept effect). They conclude that working in a city facilitates the accumulation of human capital.

However, the fixed-effect approach used by Glaeser and Mare (2001) relies on the assumption that moving across locations is random conditional on the fixed effect. This assumption is unlikely to be true if changes in wages as people move across locations are correlated with changes in the quality of the opportunities in both places.[2] For example, it is reasonable to suspect that workers who decide to move to the city were presented with better opportunities in the city in comparison to those that stayed in the rural area. In this case, the estimated city effect would be biased upwards, since the coefficient would be picking up the self-selection of individuals capitalizing on positive opportunities, rather than the effect of a typical worker moving to the city. Similarly, the lack of a wage decline for movers to the rural area could be the result of self-selection, since only workers who received relatively good offers in the rural area or relatively bad offers in the city are likely to move there. In this case, the estimated city effect would be biased down, since the coefficient will be picking up the relatively good opportunities in the

2. The issue of self-selection in the context of immigrants who move between countries is examined by Borjas (1987). In addition, several papers have attempted to estimate the returns to moving across regions after accounting for the non-random selection of individuals who choose to move. See Ham, Li and Reagan (2001) for a recent look at this issue. However, this literature does not look at the decision to live in a city or rural area and does not attempt to model and estimate the dynamic decisions over the joint migration–career choice over time.

rural area for those that chose to move there, rather than the effect of a typical person moving there. The estimation strategy in this paper overcomes these problems by modelling the decision to move locations in conjunction with the choice over various opportunities in each occupation in each location. That is, the model explicitly allows for moving across locations to be correlated with people taking advantage of relatively good wage offers across locations. As a result, this paper provides the first estimates of the urban wage premium which control for unobserved heterogeneity and the endogeneity of location and career decisions over time.

2. THE DATA AND REDUCED-FORM ESTIMATES

The data are taken from the National Longitudinal Survey of the Labor Market Experience (National Longitudinal Survey of Youth (NLSY) 1979) from 1979 until 1998. The sample consists of a random sample (the "core" sample) of white men between the ages of 14 and 21 at the start of the sample period in 1979. Each individual is followed from the age of 16 until 35, and I record the individual's education, work, and occupational status for every calendar year available. In addition, the residential location of the respondent is categorized as living in a standard metropolitan statistical area (SMSA) or not (a SMSA includes the metropolitan area of a city). The main analysis consists of a sample of 1887 men who report their joint career and city status for at least three years. Individuals appear in the sample for an average of 14·9 years.[3] In addition, the analysis uses data on weekly wages for individuals who work, as well as measures for each respondent's mental ability (the armed forces qualification test (AFQT) score adjusted for age at which the test was taken) and parental background (whether both parents were living with the respondent at age 14 and the education level of each parent).[4]

The estimation method used in the analysis does not require each individual to have complete retrospective data for any given age to be included in our sample, as in many previous dynamic models such as Keane and Wolpin (1997). Therefore, the estimation strategy will be able to use information on each individual at each point in time, regardless of the person's wage, residential location, or career status being unknown at any given time period.

In each period where sufficient information is available, each individual is characterized as living in a "city" or "rural" area, as well as one of four career sectors: schooling, white-collar employment, blue-collar employment, and the "home" sector. These career classifications are similar to those used by Keane and Wolpin (1997). To be considered working in either the white-collar or blue-collar sector, the individual had to work at least 30 weeks out of the calendar year for at least 20 hours per working week. To be classified as being in school, the individual had to complete a year of schooling during the calendar year.[5] Respondents with non-missing information about weeks worked and schooling status who were not classified into one of the other three sectors were placed into the "home" sector.

3. Ninety per cent of the men in the sample report for at least 10 years, and 56% for at least 15 years.

4. The wages for self-employed workers were treated as missing. The sample was restricted to those with non-missing AFQT scores and family background characteristics (whether both parents were living with the respondent at age 14 and the education level of each parent).

5. Specifically, the individual had to satisfy three requirements to be in the schooling sector in any given calendar year. First, the respondent must report that "Highest Grade Completed as of May 1" has increased by one since the previous calendar year (using the *created "Highest Grade Completed by May 1 of Survey Year" variable). Second, the respondent must satisfy at least one of the following two items: (a) report to be enrolled in school as of May 1 of the calendar year (using the *created "Enrollment Status as of May 1 of Survey Year" variable) or (b) report that the main activity during the interview week in the calendar year as "going to school" (using the *created "Employment Status" variable). Third, the respondent must satisfy at least one of the following two: (a) worked less than 24 weeks during the calendar year, or (b) worked more than 24 weeks but less than an average of 20 hours per working week. In addition, for the year 1979, individuals who were currently in the grade appropriate for someone who went continually to school were classified retrospectively as being in school from the age of 16 to the current age.

TABLE 1

Sector proportions over time in the NLSY sample of men

	City				Rural area				
Age	Home	School	Blue collar	White collar	Home	School	Blue collar	White collar	Sample size
16	0·03	0·60	0·02	0·00	0·03	0·30	0·01	0·00	701
17	0·05	0·58	0·03	0·00	0·03	0·28	0·02	0·00	918
18	0·08	0·41	0·15	0·03	0·05	0·19	0·09	0·01	1172
19	0·09	0·28	0·25	0·07	0·06	0·12	0·13	0·01	1394
20	0·08	0·24	0·28	0·09	0·06	0·09	0·14	0·02	1580
21	0·08	0·21	0·29	0·11	0·05	0·07	0·16	0·03	1764
22	0·07	0·12	0·33	0·17	0·05	0·05	0·16	0·05	1780
23	0·07	0·09	0·32	0·22	0·05	0·03	0·17	0·06	1752
24	0·06	0·06	0·32	0·27	0·04	0·02	0·17	0·07	1735
25	0·06	0·05	0·32	0·29	0·03	0·01	0·17	0·08	1713
26	0·05	0·03	0·32	0·31	0·03	0·01	0·17	0·08	1694
27	0·05	0·02	0·33	0·32	0·03	0·01	0·16	0·09	1684
28	0·05	0·02	0·31	0·34	0·03	0·01	0·16	0·08	1678
29	0·05	0·02	0·30	0·34	0·02	0·01	0·17	0·09	1470
30	0·05	0·02	0·29	0·34	0·03	0·00	0·18	0·10	1259
31	0·04	0·02	0·29	0·33	0·04	0·01	0·16	0·10	1011
32	0·05	0·02	0·28	0·36	0·03	0·00	0·15	0·11	783
33	0·06	0·02	0·28	0·34	0·04	0·01	0·15	0·10	557
34	0·05	0·04	0·28	0·33	0·04	0·00	0·12	0·14	370
35	0·03	0·03	0·25	0·37	0·04	0·01	0·15	0·12	183

Notes: NLSY, National Longitudinal Survey of Youth.

Table 1 displays the sector proportions over time in the sample and reveals the expected pattern of career and location choices as the cohort ages. At first, most of the sample attends school and individuals gradually move into the blue-collar and white-collar occupations over time. Interestingly, the data displays a cross-over pattern as workers in the city tend to work less in the blue-collar sector and more in the white-collar sector over time—a pattern not picked up by Keane and Wolpin's (1997) study of career choices since their study stopped at the age of 26. Although not displayed, there are also very stark differences in career patterns across levels of AFQT: men with lower AFQT scores typically work less, study less, and tend to work in the blue-collar sector more than the white-collar sector.

A descriptive analysis of the sample reveals a strong relationship between living in a city and career outcomes. The first column of Table 2 presents the results of a standard log wage regression using the sample of workers between the ages of 30 and 32. Wages are shown to be significantly associated with occupational choice and city status. Blue-collar workers earn 14% less then white-collar workers, while city workers earn a 15·4% premium over workers in rural areas. The last two columns of Table 2 present wage regressions for white-collar and blue-collar workers separately. Interestingly, the 18·3% city premium is larger for blue-collar work than the 11·5% premium for white-collar work. These findings, however, are just correlations and do not establish any causal relationships. In particular, the "city premium", which is a common result, may not be due to the effect of living in a city on wages, but rather may result from a non-random selection of men who choose to remain in cities or choose to move to cities from rural areas.

This scenario is explored in Table 3, which performs a basic wage regression after including fixed effects for each person. The estimated city premium for the whole sample is now insignificant and close to zero in magnitude. This result indicates that the observed city premium is explained by unobserved characteristics of workers who live in cities. However, when the sample

TABLE 2

Log wage regressions for NLSY sample of male workers, ages 30–32

	All workers	Blue-collar workers	White-collar workers
Intercept	5·022	5·424	4·554
	(0·399)	(0·595)	(0·528)
Education	0·064	0·036	0·087
	(0·017)	(0·026)	(0·022)
Experience	0·017	−0·000	0·034
	(0·016)	(0·023)	(0·021)
AFQT	0·022	0·025	0·021
	(0·020)	(0·029)	(0·028)
Blue collar	−0·143		
	(0·028)		
City (SMSA)	0·154	0·183	0·115
	(0·028)	(0·039)	(0·040)
# Of observations	2422	1213	1209
R-square	0·10	0·03	0·07

Notes: S.E. in parentheses. Wages are defined as weekly wages. Experience is equal to age minus education minus six. SMSA is a dummy variable for living in a SMSA. The blue-collar coefficient is relative to the white-collar sector, which is the omitted sector. SMSA, standard metropolitan statistical area; AFQT, armed forces qualification test; NLSY, National Longitudinal Survey of Youth.

TABLE 3

Fixed-effects log wage regressions

	All workers		Grew up in a city		Grew up in rural area	
Intercept	3·800	3·882	4·028	4·074	4·231	4·269
	(0·119)	(0·120)	(0·179)	(0·182)	(0·298)	(0·290)
Education	0·136	0·135	0·115	0·115	0·096	0·095
	(0·009)	(0·00)	(0·014)	(0·014)	(0·023)	(0·023)
Experience	0·047	0·036	0·059	0·052	0·041	0·038
	(0·001)	(0·002)	(0·002)	(0·006)	(0·003)	(0·003)
Blue collar	−0·030	−0·030	−0·013	−0·013	−0·015	−0·018
	(0·013)	(0·013)	(0·020)	(0·020)	(0·029)	(0·029)
City (SMSA)	0·008	−0·097	−0·009	−0·061	0·154	0·060
	(0·015)	(0·022)	(0·027)	(0·047)	(0·032)	(0·051)
City∗Experience		0·016		0·008		0·015
		(0·002)		(0·006)		(0·006)
# Of observations	16,462		7074		3157	
# Of individuals	1836		856		366	

Notes: S.E. in parentheses. Wages are defined as weekly wages. Experience is equal to age minus education minus six. SMSA is a dummy variable for living in a SMSA. The blue-collar coefficient is relative to the white-collar sector, which is the omitted sector. The sample in columns 3 and 4 is composed of workers who lived in a SMSA at age 18 or before. The sample in the last two columns is composed of workers who did not live in a SMSA at age 18 or before. The combined number of observations in the last two samples is smaller than the sample in the first two columns because of missing information about where people lived at age 18 or before. SMSA, standard metropolitan statistical area.

is broken down into men who grew up in the city and those that did not, the results show a significant city premium of 15·4% for those that did not grow up in the city and a zero city premium for those that grew up in a city. Thus, it appears that a significant city wage premium still exists for those who move to the city later on in their careers, but workers who move out of the city do not experience a cut in their wages. Table 3 also includes specifications which interact

TABLE 4

Probability of moving across locations in the next period $(t+1)$

	Living in rural area in period t	Living in city in period t
Education	0·019	0·002
	(0·002)	(0·001)
AFQT	0·028	0·001
	(0·004)	(0·002)
Mother, college graduate	0·106	0·011
	(0·014)	(0·005)
Father, college graduate	0·109	0·011
	(0·010)	(0·004)
Home sector	0·040	0·019
	(0·011)	(0·006)
Schooling sector	0·042	0·019
	(0·010)	(0·004)
Blue-collar sector	−0·064	−0·010
	(0·008)	(0·003)
White-collar sector	0·032	−0·009
	(0·009)	(0·003)
Log wage	0·043	−0·001
	(0·007)	(0·003)

Notes: S.E. in parentheses. Each entry in the table comes from a separate ordinary least squares regression where a dummy variable for moving in the next period is regressed on the independent variable in the first column. Wages are defined as weekly wages. AFQT, armed forces qualification test.

city status with experience. The interaction term is significant mainly for the sample of workers that grew up outside the city. Thus, the fixed-effect results indicate that the city wage premium exists mainly for workers moving to the city after childhood, and it appears to be a "slope" effect which increases with experience in the city, rather than a simple "intercept" effect.

These fixed-effects results are similar to those found in Glaeser and Mare (2001), who also found that the effect is mainly driven by those that move to the city and that the effect is increasing with experience in the city. They interpret their results as evidence for the idea that workers accumulate human capital faster in cities. However, these results may be due to a steeper wage profile for the type of people who tend to move from the rural area to the city. If this is the case, then the coefficient on the interaction between city status and experience may not be picking up higher human capital accumulation in the city, but rather higher returns to experience for workers who would get the same wage increases even if they did not move to the city.

In addition, Glaeser and Mare (2001) point out that the lack of a wage reduction experienced by workers leaving the city may be "explained by the selection bias deriving from workers' endogenous choice of location. If workers only leave if they are expecting solid wages outside of the city, this would explain the absence of a wage decline" (p. 337). In this scenario, the fixed-effects estimate of the city wage premium would be biased towards zero. However, there could also be endogenous moving from the rural area to the city. It is likely that the workers who choose to move to the city have been offered better opportunities in the city than the ones that choose not to leave the rural area. In this case, the fixed-effects specification will yield a city premium which is upward biased. In general, the fixed-effects specification yields biased estimates of the city wage premium if endogenous moving is important, since the city premium is identified off of the changes in wages for those workers who choose to move between locations.

The next two tables present strong evidence for the existence of endogenous moving. Table 4 analyses the differences between the characteristics of individuals who choose to move vs. those that do not move. Each coefficient in Table 4 represents a separate ordinary least squares regression where the dependent variable is a dummy variable for moving, and the independent variable is given by the variable associated with each row. The first column of results shows that individuals who move from the rural area to the city are generally more educated and of higher ability than those who choose to stay in the rural area. These findings are represented by the positive coefficients on education, AFQT, and parental education. In addition, moving to the city is positively correlated with wages and with being in the home sector, school sector, and white-collar sector, while being in the blue-collar sector is negatively correlated with moving. The last column in Table 4 shows that moving from the city to the rural area is also correlated, although to a weaker degree, with education, AFQT, parental education, and with being in the home and schooling sectors.

Table 4 shows that moving between locations is clearly not random, since moving is highly correlated with a host of background variables and sector choices. Identification of the city wage premium in a fixed-effect regression does not require moving to be random, it only requires it to be random conditional on the fixed characteristics of the individual. However, the strong correlations between moving and sectoral choices, which are not fixed, indicate that moving is part of the dynamic, endogenous process for some people as they build their career. Table 5 examines this further by comparing the sectoral transition matrix for movers vs. stayers. The main result in Table 5 is that individuals who move from the rural area to the city are much more likely to change sectors than individuals who stay in the rural area. (This is seen by comparing the diagonals for the two matrices in the top half of the table.) This finding is also present to a much lesser degree when comparing individuals in the city who move vs. those that stay in the city. Overall, there is a strong pattern that moving to the city is associated with changing career sectors, especially entering the white-collar sector. Therefore, the simultaneous choice of location and sector is likely to induce a bias in a fixed-effect regression, which assumes that city status is uncorrelated with other factors changing over time. The structural model developed in the next section will overcome these problems by explicitly modelling the endogenous choice of location and career sector, while controlling for differences in unobserved ability.

TABLE 5

Career sector transition matrix of movers and stayers

	Stayers in next period				Movers in next period			
	Home	School	Blue collar	White collar	Home	School	Blue collar	White collar
Rural								
Home	0·54	0·08	0·31	0·06	0·34	0·13	0·30	0·23
School	0·10	0·66	0·15	0·09	0·08	0·54	0·10	0·27
Blue collar	0·07	0·01	0·83	0·09	0·05	0·02	0·73	0·19
White collar	0·03	0·03	0·21	0·73	0·03	0·05	0·10	0·82
City								
Home	0·44	0·11	0·31	0·13	0·41	0·17	0·34	0·08
School	0·09	0·63	0·13	0·16	0·17	0·46	0·16	0·21
Blue collar	0·06	0·02	0·79	0·13	0·10	0·02	0·74	0·13
White collar	0·03	0·04	0·15	0·79	0·06	0·05	0·13	0·76

Notes: For the first four columns, the number in each row and column represents the fraction of men in the NLSY sample who stayed in the same location and chose the row in period t and the column in period $t+1$. The last four columns are similar but restricted to the sample of men who move across locations from period t to period $t+1$. NLSY, National Longitudinal Survey of Youth.

3. THE STRUCTURAL MODEL

This section presents the basic structure of the model and the parameterizations of each equation. The solution to the model and the estimation method are also discussed. The model corresponds to the decision problem of a single individual choosing his career option and location of residence (city vs. rural) in each time period t ($t = 1, \ldots, T$) in order to maximize his expected present discounted value of utility. Each period is associated with a certain age from 16 to 35. Each individual enters the first period with initial background variables consisting of living in a city or rural area, 10 years of schooling, an afqt score (corresponding to the quartile of the individual's age-adjusted AFQT score, so that afqt $\in 1, 2, 3, 4$), and family (family background characteristics—based on whether both parents lived with the respondent at age 14 and the schooling levels of each parent).[6] There is also unobserved heterogeneity in men, characterized by three different types of men (type $\in 1, 2, 3$).

3.1. *City and career choice set*

In each period, individuals choose whether to live in the "city" or "rural" area and to participate in one of four broadly defined career sectors: "home" (non-employment), school, blue-collar employment, and white-collar employment. The interaction of city status and the four career sectors yields the following choice set composed of eight options:

1. City, Home
2. City, School
3. City, Blue-collar
4. City, White-collar
5. Rural, Home
6. Rural, School
7. Rural, Blue-collar
8. Rural, White-collar

If a person chooses to live in the city in period t, then $\text{city}_t = 1$, and $\text{city}_t = 0$ if he chooses a rural area. The career choice of a person in period t is represented by k_t, where $k_t = 0$ if he chooses "home", $k_t = 1$ if he attends school, $k_t = 2$ if he works in blue-collar, and $k_t = 3$ if he works in white-collar. The number of years accumulated in each career choice k at the end of year t is represented by x_{kt} (*i.e.* $\sum_{\tau=1}^{t}\sum_{k=0}^{3} x_{k\tau} = t$). Initial conditions for the experience levels in each sector are normalized to zero: $x_{10} = x_{20} = x_{30} = 0$. Accumulated work experience in the city as of period t is represented by cityexp_t, where $\text{cityexp}_t = \text{cityexp}_{t-1} + 1$ if $\text{city}_{t-1} = 1$ and the person was employed ($k_t = 2$ or 3). All of the potential joint city–career choices have observable counterparts in the NLSY data.

3.2. *Parameterizations*

3.2.1. Career utility. The current period utility associated with each of the four career sectors is dependent on whether the person resides in the city or rural area (city_t), as well as

6. This paper follows Taber (2001) by assuming that AFQT (adjusted by the age at which the test was taken) is an exogenous variable. This does not mean that AFQT is equivalent to IQ or that it is inherited genetically. AFQT is simply regarded as a measure of ability at the time the test was taken. In addition, I do not find that the AFQT score is endogenous to education level at the time of the test—which suggests that AFQT is comparable across all ages in the sample. I tested this by regressing the age-adjusted AFQT score on the eventual highest grade completed and found that the relationship between AFQT and eventual schooling was stable across all age levels. Also, in order to use AFQT (which is a continuous variable) in the state space, the score was discretized into quartiles, and the score associated with each quartile is the value of a standard normally distributed variable in the middle of the quartile.

the accumulated levels of experience in each career sector as of year t (x_{0t}, x_{1t}, x_{2t}, and x_{3t}), accumulated work experience in the city (cityexp_t), the individual's type (type $= 1, 2$, or 3), and the individual's afqt quartile score. To ease the notation, the vector of experience levels in all four career sectors as of period t is represented by X_t. In general, the one-period utility of choosing career sector k with a given city_t status ($\text{city}_t = 0, 1$) is represented by $u^k_{\text{city}_t}$,which is specified for each of the four career sectors.

Home Sector Utility ($k_t = 0$).

$$u^0_{\text{city}_t} = b^0_t(\text{type, afqt}, t) + \varepsilon^0_{\text{city}_t}, \tag{1}$$

where b^0_t is the one-period deterministic value of leisure, which does not depend on city status, but does depend linearly on a quartic of the individual's age (represented by t), type, and afqt score. The $\varepsilon^0_{\text{city}_t}$ term is a stochastic shock to the value of leisure in period t, which varies independently across city status ($\text{city}_t = 0$ or 1) and is uncorrelated over time. The structure of all the shocks in the model will be discussed later.

Schooling Sector Utility ($k_t = 1$).

$$u^1_{\text{city}_t} = b^1_t(t, x_{1t}) + \text{entry}^1(x_{1t}, k_{t-1}) + \varepsilon^1_{\text{city}_t}, \tag{2}$$

where b^1_t is the one-period deterministic component of the net utility of being in school, taking into consideration both the direct monetary costs of schooling and the potential consumption value of schooling. The b^1_t component varies according to the individual's age, represented by t, and by the level of schooling, parameterized as a step function with steps for high school ($x_{1t} \leq 2$), college ($x_{1t} \leq 6$), and graduate school ($x_{1t} > 6$). The entry^1 function allows for the one-time costs of returning to school from a different career sector (*i.e.* $k_{t-1} \neq 1$) to vary with the amount of schooling. That is, the costs of returning to high school are γ^1_1,while returning to college or graduate school costs γ^1_2 and γ^1_3, respectively. The $\varepsilon^1_{\text{city}_t}$ term is a stochastic shock to the utility of schooling in period t, which varies independently across the city and rural areas ($\text{city}_t = 0$ or 1) and is uncorrelated over time.

Blue-Collar Sector Utility ($k_t = 2$).

$$u^2_{\text{city}_t} = \text{wage}^2_{\text{city}_t}(X_t, \text{type}, \text{afqt}, k_{t-1}, \text{cityexp}_t, \varepsilon^2_{\text{city}_t}) + \text{entry}^2(X_t, k_{t-1}), \tag{3}$$

where $\text{wage}^2_{\text{city}_t}$ is the blue-collar wage offer function, which is specified in a Mincer-like fashion and depends on the accumulated experience in each career choice X_t (completed schooling, blue-collar experience, and white-collar experience), the individual's afqt score, and the individual's type. The entire blue-collar wage function is parameterized separately for each city status ($\text{city}_t = 0$ or 1), which allows for all of the parameters to differ between the city and rural locations. In addition, the wage function includes a return to work experience acquired while working in the city (represented by cityexp_t). The inclusion of this variable allows for the possibility that work experience acquired in the city makes someone more productive than work experience in a rural area.

By including a return to white-collar experience in the blue-collar sector, the specification allows for the full or partial transferability of occupation-specific experience acquired in the white-collar sector. In addition, utility in the blue-collar sector depends on entry^2, which is the one-time (non-wage) cost of entering the blue-collar sector if the individual was not working in the blue-collar sector in the previous period ($k_{t-1} \neq 2$). This non-wage entry cost captures the idea that there may be search costs in finding a blue-collar offer or starting to work in the blue-collar sector (transportation, clothing, etc.).

The wage function also includes a return to continuing to work in the blue-collar sector ($k_{t-1} = 2$) which, along with the entry cost function, is designed to capture the persistence of career choices across time periods (see Keane and Wolpin, 1997). Also, including a return to staying in the same sector is equivalent to incorporating a human capital depreciation effect, whereby workers lose part of their human capital if they drop out of the sector for at least one period. The log wage offer is also subject to a linear stochastic component $\varepsilon^2_{\text{city}_t}$, which is uncorrelated across city status and time.

White-Collar Sector Utility ($k_t = 3$).

$$u^3_{\text{city}_t} = \text{wage}^3_{\text{city}_t}(X_t, \text{type}, \text{afqt}, \ k_{t-1}, \text{cityexp}_t, \ \varepsilon^3_{\text{city}_t}) + \text{entry}^3(X_t, \ k_{t-1}), \tag{4}$$

where each component is defined analogously to the utility components in the blue-collar sector, although the parameters differ for each occupation and location.

3.2.2. City utility. The single-period utility associated with city status city_t is a function of the person's type, afqt, and career choice k_t.

$$u(\text{city}_t) = \begin{cases} u\text{city}(\text{type}, \text{afqt}, k_t) & \text{if} \quad \text{city}_t = 0 \\ 0 & \text{if} \quad \text{city}_t = 1, \end{cases} \tag{5}$$

where the net utility of living within a city is normalized to 0.

3.2.3. Moving costs. Moving from the city to the rural area or back is costly. Therefore, moving costs are dependent on *age* and career choice k_t:

$$\text{Moving costs} = \begin{cases} \text{movecosts}(\text{city}_t, \text{age}, \ k_t) & \text{if} \quad \text{city}_t \neq \text{city}_{t-1} \\ 0 & \text{if} \quad \text{city}_t = \text{city}_{t-1}, \end{cases} \tag{6}$$

where moving costs are incurred only if the person chooses to move, and the cost of moving from the city to the rural area is allowed to be different than the cost of moving in the reverse direction. This specification allows for age and certain career choices to affect whether people are more or less mobile between residential locations.

3.2.4. Structure of the career and city shocks. In each period t, an individual receives eight separate stochastic shocks: one in each of the four career sectors within the city and the rural areas. These shocks were depicted in equations (1)–(4) by $\varepsilon^k_{\text{city}_t}$ where $k = 0, 1, 2, 3$ and $\text{city}_t = 0, 1$. All of these shocks are presumed to be normally distributed with mean zero and mutually serially independent over time (Keane and Wolpin, 1994). In addition, the blue-collar and white-collar wage shocks are allowed to be correlated contemporaneously within city status. The remaining shocks are independently distributed.[7]

7. Empirically, it is feasible to make all the shocks contemporaneously correlated, but this would require estimating 36 elements of the Cholesky decomposition matrix. Most of these correlations did not appear to be significant, so for the sake of parsimony, only the occupational wage shocks were estimated to be correlated within locations, which resulted in estimating 10 elements of the Cholesky matrix (as depicted in the Appendix).

3.2.5. Job offer function. Individuals who are in the "home" sector are allowed to re-enter the workforce in the next period with a certain probability. The probability of receiving an offer to work in sector k is represented by π_t^k, where $k=2$ for the blue-collar sector and $k=3$ for the white-collar sector. Individuals who are "home" are not required to accept the offer, but they cannot choose to enter either occupation in period t without receiving a contemporaneous offer in the associated occupation.

3.2.6. Type probabilities for men. Each individual is assumed to be one of three discrete types corresponding to three mass points in a non-parametric distribution of permanent unobserved heterogeneity (Heckman and Singer, 1984). The probability of being a certain type (type $= 1, 2, 3$) of individual is modelled as a tri-variate logit:

$$\pi^{\text{type}} = \pi^{\text{type}}(\text{afqt}, \text{family}),\ \text{type} = 1, 2, 3, \tag{7}$$

where the probability of being a certain type depends on the individual's afqt score and family background, all of which are assumed to be exogenously determined for each individual at the age of 16. This formulation allows for the possibility that individuals could be a "high" type even if they come from a disadvantaged background or a "low" type if they come from a good background.

3.2.7. Objective function. At each age from 16 ($t=1$) to 35 ($t=T$), the individual chooses his career sector and location of residence in order to maximize the expected present discounted value of lifetime utility. Let Ω_t represent the relevant information set with which the individual enters period t. Ω_t includes the individual's history of career decisions (denoted by X_{t-1}), city status (city_{t-1}), city work experience (cityexp_{t-1}), the individual's $type$, afqt score, and family background. Given this set of relevant information, the one-period utility associated with any combination of city status and career choice is denoted by $U(\text{city}_t, k_t \mid \Omega_t)$, and is determined by equations (1)–(6) above:

$$U(\text{city}_t, k_t \mid \Omega_t) = u^k_{\text{city}_t} + u(\text{city}_t) + \text{movecosts}(\text{city}_t,\ k_t). \tag{8}$$

This specification demonstrates the interaction between career and city status: historical career choices and residential decisions affect the wage offers in each sector in both the city and the rural area, and work experience acquired in the city may increase the productivity of workers after they move away from the city. Career decisions also affect the mobility of individuals by affecting their moving costs. Furthermore, the structure of the model explicitly allows for moving decisions to be correlated with good or bad shocks in all career sectors in both the current and alternative location of residence. Finally, it is worth noting that city status affects the entire wage function, not just the intercepts as is commonly specified in the literature. In this manner, the model endogenizes the dynamics of working or not working, going to school, occupational choice, and city–rural status.

The potential choice set in period t is given by the Cartesian product of the four career sectors multiplied by the two city status categories. If the individual chooses to be in the home sector in period $t-1$, his feasible choice set may be restricted in period t if he does not get an offer to work in the blue or white-collar sectors. We denote the choice of element j in his feasible choice set in period t as $d_t^j = 1$ ($j = 1, \ldots, J_t$), and the utility associated with that choice as U_t^j (specified in equation (8)). The individual's objective function in each period t is then represented as

$$V_t(\Omega_\tau) = \max_{\{d_t^j\}} E\left[\sum_{\tau=t}^{T}\sum_{j=1}^{J_\tau} \delta^{T-\tau} U_\tau^j d_\tau^j \mid \Omega_\tau\right], \tag{9}$$

where δ is the discount factor (fixed at 0·95) and E is the expectation operator taken over the joint distribution of stochastic shocks ($\varepsilon^k_{\text{city}_t}$ where $k = 1, 2, 3, 4$ and $\text{city}_t = 0, 1$) as well as the distribution of job offer probabilities (π_t^1 and π_t^2). The solution to this problem in each period yields the optimal stream of joint career and location decisions over time.

3.3. *Model solution and estimation*

3.3.1. Solution. The model does not have an analytic solution and therefore is solved numerically using backward recursion starting from a terminal age T. The maximization problem in equation (9) can be re-written as the maximization over the value functions of the available set of joint location and career states j ($j = 1, \dots, J_t$) at time t, denoted as $V_t^j(\Omega_\tau)$, which satisfy the Bellman (1957) equation:

$$V_t(\Omega_t) = \max\left[V_t^1(\Omega_t), \dots, V_t^J(\Omega_t)\right]$$

$$V_t^j(\Omega_t) = U_t^j + \delta E\left[V_{t+1}(\Omega_{t+1}) \mid (d_t^j = 1),\ \ \Omega_t\right]. \tag{10}$$

Therefore, given any set of parameters, solving the model consists of simulating all of the stochastic components at each point in the state space (every possible combination of historical decisions over city status and career alternatives for every type and afqt score up to period t), and using backwards recursion to calculate U_t^j and $E\left[V_{t+1}(\Omega_{t+1})\right]$ (see Keane and Wolpin, 1994, 1997). The latter term is called the Emax_{t+1} function for convenience. At each iteration in the estimation, 30 draws of the entire set of stochastic components were taken according to the current set of parameters to recursively estimate the Emax_t at every point in the state space. The value of Emax_T for the terminal period T is parameterized as a function of the individual's afqt as well as his historical and terminal state choices: $\text{Emax}_T(X_T, k_T, \text{afqt})$.

3.3.2. Estimation. To estimate the model, the numerical solution of the dynamic programming problem described in the previous section is nested within an algorithm that maximizes a likelihood function. The likelihood function is constructed by simulating a set of choice histories and matching the choice histories with observed choices in the data. The estimation strategy deals effectively with the problem of unobserved initial conditions (see Heckman, 1981) and state variables. These problems can be quite severe when constructing city status and employment histories from NLSY data, since the relevant information is frequently missing for some respondents in various years, especially after 1994 when the NLSY survey was conducted every other year. Therefore, this technique uses all the information available in the data, without having to worry about constructing the complete career and residence history for each respondent.[8]

The estimation algorithm is based on simulating the complete career and residence histories of a set of artificial agents ($n = 1, \dots, N$). Given a set of parameter values, the simulation for agent n is performed as follows:

8. The estimation uses techniques developed by a long list of papers. A partial list includes Heckman (1981), Heckman and Singer (1984), Rust (1987), Hotz and Miller (1988), Hotz, Miller, Sanders and Smith (1994), Keane and Wolpin (1994, 1997, 2001), and Stern (1997).

1. Draw agent n's parental background (family), initial location status (city_0), and afqt according to the actual proportions in the data (and according to the actual correlations of these variables to each other in the data—which were estimated by logits and multivariate logits outside the estimation algorithm using the NLSY sample).
2. Using the simulated background variables (family, city_0, and afqt), draw the agent's type ($\text{type} = 1, 2, 3$) as specified in equation (7).
3. Draw from all the stochastic elements in the model ($\varepsilon^k_{\text{city}_t}$ where $k = 1, 2, 3, 4$ and $\text{city}_t = 0, 1$), including the job offer probabilities (π^1_t and π^2_t), to determine the value of the available career offers in the city and rural areas (conditional on n's afqt and type).
4. According to the agent's type, afqt, and realizations of the stochastic elements in step (3), the agent evaluates each career option in the city and rural area using equation (10) and the Emax_{t+1} for each option (which was already constructed at the current parameterization as described above). The agent chooses the option with the highest expected value over the current set of J options: $\max\left[V^1_t(\Omega_t), \ldots, V^J_t(\Omega_t)\right]$.
5. The state variables (X_t, cityexp_t, k_t, city_t) are updated according to the choice in step (4).
6. Repeat steps (3)–(5) until $t = T$.

Doing this N times produces N artificial agents with a complete set of career and residential outcomes over T periods ($N = 80{,}000$ in the actual estimation). The likelihood function is then built using a frequency simulator, although as Lerman and Manski (1981) point out, the probability that the entire career and residential choices of a simulated agent (including wages) matches someone in the data is infinitely small. Instead, the likelihood is built on the simulated frequencies of period-by-period choices within type and afqt score.

In particular, there are eight possible joint career and residential choices (four career choices in both the city and rural area) observed for every individual in the data at time t. Denote the probability that individual i, given his exogenously determined type and afqt score, chooses the joint residence–career state j ($j = 1, \ldots, 8$) and earns log wage w_{ijt} at time t:

$$\begin{aligned} P^{it}_{\text{type,afqt}}(\text{choice}_{it} = j, \text{lnwage} = w_{ijt} \mid \text{type, afqt}) &= \text{prob}(\text{choice}_{it} \\ = j \mid \text{type, afqt}) * \varphi\left(\left(\frac{w_{ijt} - \mu_{jt}}{\sigma_{jt}}\right) \mid \text{choice}_{it} = j, \text{type, afqt}\right), \end{aligned} \tag{11}$$

where $\varphi(\cdot)$ is a standard normal p.d.f. and μ_{jt} and σ_{jt} are the mean and S.D. of log wages conditional on choosing the joint residence–career state j at time t and given the individual's type and afqt. The first component can be estimated by the proportion of simulated agents with the same type and afqt who choose j at time t. The mean and variance of the conditional wage distribution can be estimated similarly by computing the mean and variance of log wages for simulated agents with the same type and afqt who choose j in time t. Substituting these estimates into equation (11) yields the simulated probability $\widehat{P}^{it}_{\text{type,afqt}}$. The probability of the sequence of person i's choices and wages over time, given type and afqt, is then estimated as

$$\text{prob}(\text{choices}_i \mid \text{type, afqt}) = \prod_{t=1}^{T} \widehat{P}^{it}_{\text{type,afqt}}. \tag{12}$$

If the person's type were known, equation (12) would be the likelihood contribution of person i. However, a person's type is unknown, but the probability of being a certain type is estimated as a function of observable initial characteristics (family and afqt). Therefore, the likelihood contribution for person i is simply the weighted sum of equation (12) over the three possible types using the probability of person i being each type as the weight:

$$L_i = \sum_{\text{type}=1}^{3} \text{prob}(\text{choices}_i \mid \text{type}, \text{afqt}) * \pi^{\text{type}}(i = \text{type} \mid \text{family}, \text{afqt}), \qquad (13)$$

where π^{type} is defined above in equation (7) as the probability of person i being a certain type (type $= 1, 2, 3$). Thus, we have derived the contribution of each person in the data to the likelihood function conditional on his afqt score and family background characteristics.[9] It is important to note that the likelihood contribution of each person is composed only of elements when there is a recorded choice for person i in the data. In periods where person i does not report a choice or wage, this element is simply dropped in the process of building the likelihood.[10]

3.3.3. Identification. The parameters of the model are identified by trying to fit the career outcomes and locational choices of individuals over time and variation across individuals in the patterns of their historical choices and wages.[11] In this manner, this paper follows the line of empirical research on dynamic models (Eckstein and Wolpin, 1989; Keane and Wolpin, 1997, 2001; and many others), which views any decision or outcome after the age of 16 as endogenous. Therefore, any decision or outcome after the age of 16 is rendered unsuitable as an exclusion restriction that could help identify the parameters of the choice equations. However, although exclusion restrictions would be useful, they are not necessary with panel data to control for the potential correlation between an individual's unobserved ability and sector choices. In a cross-sectional framework, we cannot observe the same person in two different states; so it is especially important to have some exogenous variable which affects the choice equation in order to control for unobserved ability or tastes that cause a person to choose one sector over the other. With panel data on the choices of individuals over time, we often observe the same person in two different sectors, and this allows us to control for unobserved ability with a simple fixed-effects strategy.

Therefore, to understand the value-added of the paper, it is best to compare the identification strategy of our model to a simple fixed-effects model. In a fixed-effects specification, the unobserved fixed effect is identified from variation in wages across people (*i.e.* people who tend to earn persistently high unexplained wages get a high fixed effect, and those that earn little get a low fixed effect). The structural model identifies the unobserved "types" in a similar way: those that earn a lot will be estimated to be a "high" type and those that earn little will likely be one of the "lower" types.

After controlling for the unobserved ability of each person, a fixed-effects model identifies the city wage premiums from changes over time in the wages and locational choices within each

9. In practice, the simulated likelihood function was optimized using a simplex algorithm since the likelihood function is not smooth. S.E. were estimated as follows. Let g_i be the vector of derivatives of the log contribution to the log likelihood of person i with respect to the set of parameters $\theta : g_i = \frac{\partial \ln L_i}{\partial \theta}$. This derivative was approximated by taking small steps in the estimated parameters $\theta : \widehat{g}_i = \frac{\ln L_i(\theta) - \ln L_i(\theta + h)}{h}$ where $h = \theta * 10^{-2}$. The covariance matrix is then estimated by $\left(\sum_i \widehat{g}_i \widehat{g}_i'\right)^{-1}$.

10. In theory, we could match person i's choices in each period t to simulated moments conditional on person i's previous choices. However, this would entail throwing out any observation in the sample where the previous period's choice is unobserved. As a consequence, the sample size would be dramatically reduced, in particular, because the NLSY was sampled only every other year from 1994 onward. Thus, if we conditioned on the previous period's choice, the sample used in the analysis would contain few people beyond the age of 29.

11. Formal identification of the 131 parameters of the model stems from fitting the 640 moments of career-location choices over time (four career choices in both the city and rural area, 20 time periods, for each of the four AFQT levels) and the 640 moments of wage outcomes over time (the mean and variance of wages in the blue-collar and white-collar sectors in both the city and rural area, 20 time periods, for each of the four AFQT levels).

individual. For example, a fixed-effect model will identify a positive city wage premium by the increase (decrease) in wages of workers who move to (leave) the city. In the structural model, the effect of being in the city is also identified by the change in a person's wage as he moves across locations. However, the identification of the city wage premium in a typical fixed-effect model is spurious if moving across locations is non-random and correlated with other decisions and outcomes for the individual, which may be producing the change in wages regardless of the move. As seen in Section 2, the decision to move across locations is correlated with other outcomes of the individual, such as occupational choices and wages. Therefore, the main advantage of the structural model is that it accounts for the endogeneity between location and career decisions, which is likely to induce a spurious correlation between wages and location decisions in a fixed-effects model.

The structural model identifies the parameters of the decision-making process by trying to fit not only the observed wages and choices of individuals in the data, but also by fitting all the evidence in the data about why individuals are moving or not across locations. This, in turn, adds considerable identification power to the model since the range of parameters that can fit choices across individuals and within individuals over time is very restricted. For example, let us suppose that workers who move to the city appear to receive a wage increase. To the extent that the wage increase is due to a higher wage profile for similar "high" type individuals regardless of whether they move or not, the model will attribute this to a steeper wage profile for this unobserved type rather than to the effect of living in a city. To the extent that this wage increase is correlated with a switch in occupations, which may produce a wage change regardless of moving locations, the model will have to produce parameters which cause individuals of a similar "type" to simultaneously move to the city and switch occupations and determine how much of the wage increase is due to either decision by comparing their wages to similar workers who do not move but change occupations and to workers who change locations, but do not change occupations. If movers to the city experience a wage increase which appears in the data to be temporal, the structural model will produce estimates whereby workers in the rural area are enticed to move to the city by a positive shock in the city wage offer function, which appears in the data to disappear after arriving. If, however, the variance of wage shocks produces too many people moving into a particular sector to take advantage of positive shocks, the model will correct for this by identifying the appropriate level of moving costs, which produces the closest match to the actual sector proportions.

These are just a few of the possible scenarios which place restrictions on the range of parameters that are able to fit the patterns of behaviour in the data. Furthermore, all of these scenarios have different implications for the observed moments in the data, which the likelihood is trying to fit. For example, if higher ability blue-collar workers move to the city to become white-collar workers in their late twenties, this process will not only affect the proportion of people choosing each occupation and location, but will also impact the mean and variance of wages in each choice, since these workers will likely move from the top of the blue-collar wage distribution to the middle or bottom of the white-collar distribution. As such, the estimation procedure searches for the parameters which are the most consistent with all the variation in the data across and within individuals about why certain people move and others do not move, and how these decisions affect the observed moments in the data.

Finally, it is important to note that estimating a dynamic model like the present one requires many functional form assumptions that are not necessary in a simple fixed-effects model. However, the structural approach has two major advantages which should be emphasized: (i) as stated above, the fixed-effect model has to assume that individuals move locations in random ways that are uncorrelated with other career outcomes; and (ii) a fixed-effect model has to assume that individuals move across occupations in random ways, since one of the goals of the paper is to estimate the wage functions for the white-collar and blue-collar sectors separately. Both of these assumptions are clearly refuted in Section 2. In particular, Tables 4 and 5 showed that locational switches are correlated with occupational status and occupational switches—thus underscoring the need to consider career and locational decisions as endogenous outcomes. So, in many respects, the structure imposed by the model is less restrictive than the refuted assumptions of a simple fixed-effects approach.

4. RESULTS

4.1. *Fit of the model*

Before discussing the results, I first present the fit of the model. Figure 1(a)–(d) shows the fraction of all men who choose each location and career choice as reported in the NLSY sample and after simulating the model with the final parameter estimates. Overall, the model produces patterns of career choices in both the city and the rural area which are very similar to those reported in the NLSY sample. In particular, the model captures the sharp drop in schooling and the growth of the white-collar sector over time relative to the blue-collar sector. The hypothesis that the sector proportions from the model estimates are equal to those in the actual data is not rejected by a chi-square goodness-of-fit test at the 10% significance level.[12] The model also picks up the dominance of the blue-collar sector over the white-collar sector in the rural area. Figure 2(a) and (b) display a close fit of the model to blue-collar and white-collar wages in both locations as well, in particular during the prime ages of working. Although not shown, the model captures very closely all of the differences across AFQT levels in career patterns, locational decisions, and wages. In particular, the model correctly predicts that AFQT scores are positively related to schooling, working in white-collar sector and higher wages in both the city and the rural area. Therefore, the model captures not only the career and location patterns for all men over time, but also matches the cross-sectional variation in these patterns across men with different AFQT scores.

Since the likelihood function is built to fit the wages and proportions of men who choose each state in each period, it is reassuring, but perhaps not surprising, that the model fits the wages and sector proportions quite well. A good test of the model, therefore, is to see how the model fits moments which the likelihood function is not explicitly trying to match. The model appears to do this quite well by closely fitting the transition rates to and from each sector and location as depicted in Table 6. The hypothesis that the transition cells from the model estimates are equal to those in the actual data is not rejected by a chi-square goodness-of-fit test at the 10% significance level.[13] The fact that the estimation procedure is able to fit moments that the likelihood is not explicitly trying to fit should increase our confidence in the model's predictions. However, it is left to the reader to decide whether the overall fit is good enough to draw credible inferences from the results presented in the rest of the paper.

4.2. *Discussion of the estimates*

There are too many (131 to be exact) coefficients to discuss individually, but several patterns emerge in the estimates presented in the Appendix. First, there are three types of men who are generally characterized by an increasing quality of parental background (parental education and whether the person lived with both parents at age 14). Type 3 men clearly come from the best family circumstances, while Types 1 and 2 come from more disadvantaged backgrounds. As such,

12. The test statistic is 87·57 while the critical value at the 10% significance level is 96·58.
13. The test statistic is 68·94 while the critical value at the 10% significance level is 78·86.

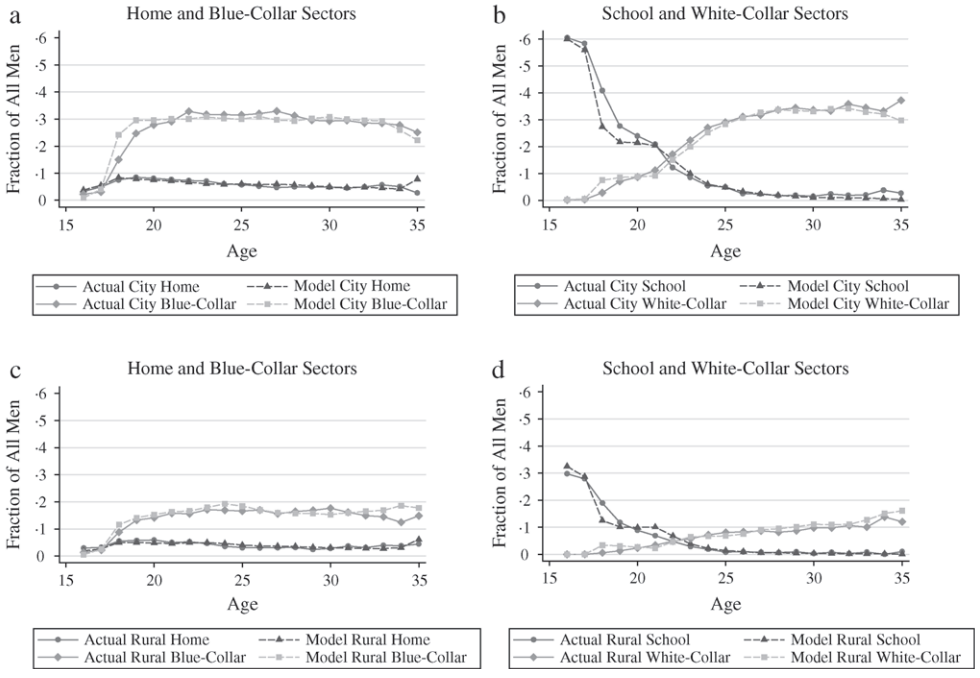

FIGURE 1

Model fit of sector choices in the city: (a) home and blue-collar sectors and (b) school and white-collar sectors and in the rural area: (c) in home and blue-collar sectors and (d) school and white-collar sectors

Type 3 men typically earn higher wages in both the white and blue-collar sectors, while Types 1 and 2 lag behind in that order. All three types of men are found working in each occupation in the city and the rural area. However, Type 1 men tend to work disproportionately in white-collar jobs in the rural area, while Type 2 men are often found in blue-collar jobs in the city, and Type 3 men work predominantly in white-collar jobs in the city. After controlling for the self-selection of these unobserved "types" across locations and occupations, the estimates display very little differences in the returns to education between the city and the rural area.

4.3. *The selection-corrected city wage premium*

In order to estimate the true city wage premium, we need to be clear about what we are trying to estimate. It is well known that the observed gap in wages between workers in the city vs. the rural area cannot be interpreted as the effect of being in a city. The reason for this is straightforward: workers in the city are likely to have fundamentally different characteristics from workers in the rural area. For example, workers in the city are likely to be different than workers in the rural area in terms of their family backgrounds, aptitudes, tastes, investments in education, labour force participation, and occupational choices over time. These differences are the result of differences in the characteristics of people who grow up in the city vs. the rural area (the initial conditions), and the self-selection of people over time into the sector and location that best suits them according to their tastes and abilities.[14]

14. For example, only 60% of individuals in the lower quartile of the AFQT distribution lived in a city at age 17, while 70% of those in the upper quartile lived in the city at age 17. Also, only 59% of individuals whose mothers dropped

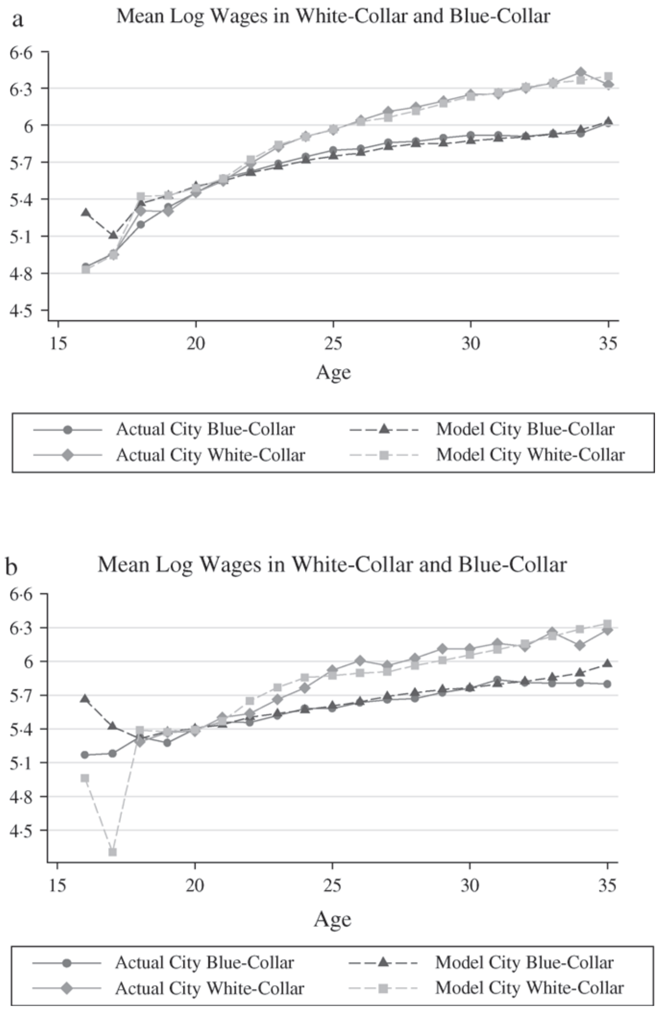

FIGURE 2

Model fit of wages (a) in the city and (b) in the rural area (mean log wages in white collar and blue collar)

Ideally, we would like to control for the differences in the types of people found in each location by comparing the wages of the same person in both the city and rural area at the same time. Since it is impossible to observe the same person under two different scenarios at the same time, the next best option would be to generate two comparable samples of individuals and randomly put one sample in the city and the other in the rural area. This is also not practical, since real people cannot be manipulated in this manner. However, after estimating the model, it is possible to perform this kind of a thought experiment by simulating the model after manipulating the parameters in a certain way.

To see how this is done, it is important to remember that the model is designed to account for differences in the initial conditions of people across locations and the subsequent endogenous

out of high school lived in a city at age 17, while this number increases to 77% for those with mothers who completed college.

TABLE 6

Model fit of the transition matrix

	City							
	Home		School		Blue collar		White collar	
City								
Home	0·41		0·11		0·29		0·12	
		0·39		0·16		0·28		0·13
School	0·08		0·59		0·12		0·15	
		0·06		0·55		0·16		0·15
Blue	0·05		0·02		0·75		0·13	
collar		0·06		0·02		0·77		0·08
White	0·03		0·03		0·14		0·75	
collar		0·03		0·02		0·09		0·79
Rural								
Home	0·05		0·02		0·04		0·03	
		0·03		0·02		0·02		0·01
School	0·01		0·09		0·02		0·04	
		0·02		0·15		0·01		0·01
Blue	0·0		0·00		0·07		0·02	
collar		0·02		0·01		0·08		0·01
White	0·0		0·01		0·02		0·13	
collar		0·02		0·01		0·02		0·17

	Rural							
	Home		School		Blue collar		White collar	
City								
Home	0·03		0·01		0·02		0·01	
		0·02		0·01		0·01		0·00
School	0·01		0·03		0·01		0·01	
		0·01		0·06		0·00		0·00
Blue	0·00		0·00		0·03		0·01	
collar		0·02		0·00		0·05		0·00
White	0·00		0·00		0·01		0·03	
collar		0·01		0·00		0·01		0·06
Rural								
Home	0·46		0·07		0·27		0·05	
		0·40		0·11		0·28		0·13
School	0·08		0·55		0·13		0·08	
		0·06		0·49		0·15		0·11
Blue	0·06		0·01		0·75		0·08	
collar		0·06		0·01		0·78		0·04
White	0·02		0·03		0·18		0·61	
collar		0·04		0·01		0·08		0·66

Notes: The number in the upper left part of each row and column represents the fraction of men in the sample drawn from the NLSY who chose the row in period t and the column in period $t+1$. The number in the lower right part of each row and column represents the same transition calculated from simulations using the estimated model parameters.

decisions over their career and location over time. Individuals in the model self-select themselves into the sector and location which maximizes their lifetime utility in each period. However, after estimating the model, the model can be simulated under the condition that individuals are not allowed to choose their optimal career sector and location. By denying the power of free will to the simulated agents, we no longer have the problem that the sample of workers are not comparable across locations and occupations due to the self-selection of workers into their optimal career-location choice. Furthermore, to overcome the problem that individuals differ in their initial conditions across locations (family background, AFQT, etc.), we can deny the power of free will in the simulation after randomly distributing the simulated agents across locations in the initial period (*i.e.* a random initial assignment).

Therefore, to estimate the true city wage premium for each occupation, the thought experiment that we are trying to produce is the following: (i) create a sample of individuals at the age of 16 that grew up in the city that are comparable in terms of their characteristics to a different sample of individuals that grew up in the rural area; and (ii) deny everyone the power of free will by randomly choosing the occupation and location for each individual instead of letting him choose his optimal choice in each period. In this scenario, we could compare the wages of workers in the blue-collar sector in the city to the wages of blue-collar workers in the rural area, and conclude that the wage gap stems from the difference in location rather than differences in abilities and other characteristics of workers across locations. Similarly, the city wage premium for white-collar workers is estimated by comparing the simulated wages of white-collar workers across the two locations. In this manner, we can estimate the city wage premium in each occupation after controlling for the non-random initial allocation of individuals between cities and rural areas, the selection of workers into the workforce vs. not working, selection into white-collar or blue-collar work, selection into schooling, endogenous moving between locations, and the dynamic interaction of all these endogenous decisions over time.

However, in order to present the estimated city wage premiums for each occupation as described in the thought experiment above, we cannot simply look at one or two coefficients on a dummy variable for living in a city, as if we ran a typical wage regression. A simple wage regression with a dummy variable for living in a city implicitly assumes that being in a city affects only the intercept of the wage function, while we are allowing it to affect the slope and the intercept of each sector's wage function. Evidence for both an intercept and a slope effect was presented in Section 2 and also in Glaeser and Mare (2001). Section 2 and Glaeser and Mare (2001) also discuss why distinguishing between the two effects allows us to rule in or out certain explanations for the existence of a city wage premium, if one exists at all. As a result, we allow the entire wage function to differ between both locations for each occupation. Therefore, presenting our estimates of the true city wage premium will require us to show how the estimated city wage premium for each occupation changes over time with accumulated experience.

As a benchmark case, Figure 3(a) and (b) present the city wage gap over time for both occupations after simulating the dynamic programming model using the final estimates, denoted by "model" in the graphs. These estimates are influenced by the non-random initial placement of individuals across locations and the self-selection of individuals into their optimal career and location choice over time. At age 30, the wage gap according to the model is 10·8% in the blue-collar sector and 17·5% in the white-collar sector. However, these gaps are likely to be biased estimates of the true effect of living in the city due to the correlation between personal characteristics and locational choice. It is worth noting, though, that the direction of the bias is not predictable in advance, since the self-selection process will tend to attract workers into the location and occupation that best suits their tastes and talents. This process, therefore, will typically drive up the observed wages in all locations and occupations in comparison to randomly distributing workers into occupations and locations.[15] If this process drives up the average blue-collar wage in both the city and rural area at the same rate, then the observed city

15. For the purposes of this discussion, I explain the case where there is "positive" selection into each location and occupation. This is not necessarily true in reality, although the basic point about the unknown direction of the bias still holds true if one sector displays negative selection. For lengthy discussions of all the possible directions of selection, see Roy (1951), Heckman and Sedlacek (1985), and Gould (2002).

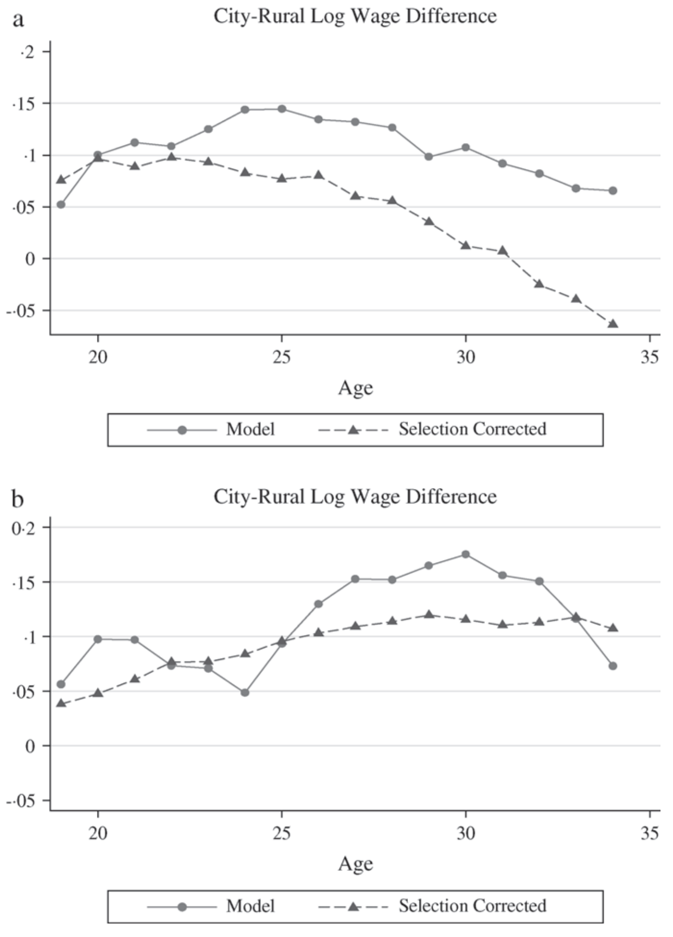

FIGURE 3

City wage premium in (a) blue collar and (b) white collar (city–rural log wage difference)

wage gap will reflect the true city wage premium. If, however, this process drives up the average wage in one location more than the other, then the observed city wage gap will be biased up or down in relation to the true city wage premium. Therefore, the direction of the bias is unknown in advance.

We now try to estimate the true city wage premium by controlling for all of the sources of selection bias by simulating the model under the thought experiment described above. First, we randomly sort simulated agents into the city or rural area in the initial period without any regard to their parental background. Second, agents are denied the right of free will by forcing everyone to complete high school (regardless of whether this maximizes their utility). Then, agents are randomly sorted into an occupation within their location and forced to stay there until the end of the sample period. That is, the power of choice is completely removed from the individual. However, the results from this simulation should be interpreted as the effect of randomizing a typical worker given the current labour market equilibrium (*i.e.* skill prices within each location and occupation). If the decisions of all workers were actually manipulated, this would have general equilibrium effects on skill prices within each location and occupation.

The results for this thought experiment are displayed in Figure 3(a) and (b) and are denoted by "Selection Corrected".[16] For the white-collar sector, the city wage premium from the thought experiment is generally lower than the wage gap estimated from the full model, which allowed for free will in every period. For example, the estimated wage gap from the full model at age 30 was 17·5% in the white-collar sector, while falling to 13·0% in the current thought experiment. Although the experiment produces a city wage gap in white-collar work that is about 34% lower, the gap is still positive and increasing slowly with white-collar experience. That is, a randomly placed person into the white-collar sector does earn more in the city vs. the rural area.

Figure 3(a), however, shows very different results for the blue-collar sector. The estimated wage premium from the full model at age 30 was 10·8%, but falls dramatically to 2·4% in the current thought experiment. In addition, the city wage premium from the thought experiment appears to be declining with experience, which tracks the actual trend in the data.[17] Overall, the results indicate that a given blue-collar worker earns basically the same wage in a city vs. a rural area.

Figure 4(a) and (b) breaks down the results by AFQT level. For the blue-collar sector in Figure 4(a), there are two general results. First, the city wage premium is inversely related to AFQT score—those with the lowest AFQT scores have higher city wage premiums, while those in the upper half of the AFQT distribution have zero or even a negative city wage premium by the age of 30. Second, as seen in Figure 3(a) for the overall population, the city wage premium in blue-collar work is declining with age for all levels of AFQT scores. The question remains, however, as to why low AFQT workers have a true city wage premium for blue-collar work at all and why they do not all move to the city to take advantage of it? A possible explanation may be related to the finding that even though a city wage premium exists for these workers, it disappears over time. In addition, low AFQT workers, who are disproportionately in rural areas during their childhood, may not move to the city to capitalize on the city wage premium in blue-collar work due to tastes for living in the rural area, high moving costs, or a combination of both.

Figure 4(b) presents the city wage premiums in the white-collar sector for each AFQT level, and the results are exactly reversed from those for blue-collar work in Figure 4(a). First, the city wage premium is increasing with AFQT score. Second, the city wage premium is increasing with age. The finding that the city wage premium is largest for workers with high AFQT scores could be an important explanation for why high-ability individuals tend to be white-collar workers in the city.

Overall, this section shows that the endogenous selection of workers into occupations and locations can almost completely explain why blue-collar workers earn more in the city vs. the rural area, but only explains roughly a third of the observed wage gap between white-collar workers across locations. Although the results vary according to AFQT level, the overall picture suggests that a given worker is paid more for white-collar work in the city, but similarly between locations for blue-collar work.

4.4. *Is there a return to city work experience in the rural area?*

Perhaps the most interesting parameters of the model are the returns to city work experience in the wage functions in the rural sector. The estimated coefficients are 0·003 for blue-collar

16. Technically, it is important to note that these simulations are a product of the estimated parameters, and therefore, there is a variance to all of the estimates. However, the S.E. are very small for most of the parameters, so we follow most of the literature by not reporting error bands on the following simulations.

17. The observed blue-collar city wage gap in the data is 17% at age 26, 8% at age 31, and 10% at age 32. The city wage premium for blue-collar work in the experiment does not have to follow the trend in the data, but since the actual observed gap also declines, it should not be surprising that the estimated true city wage premium follows suit.

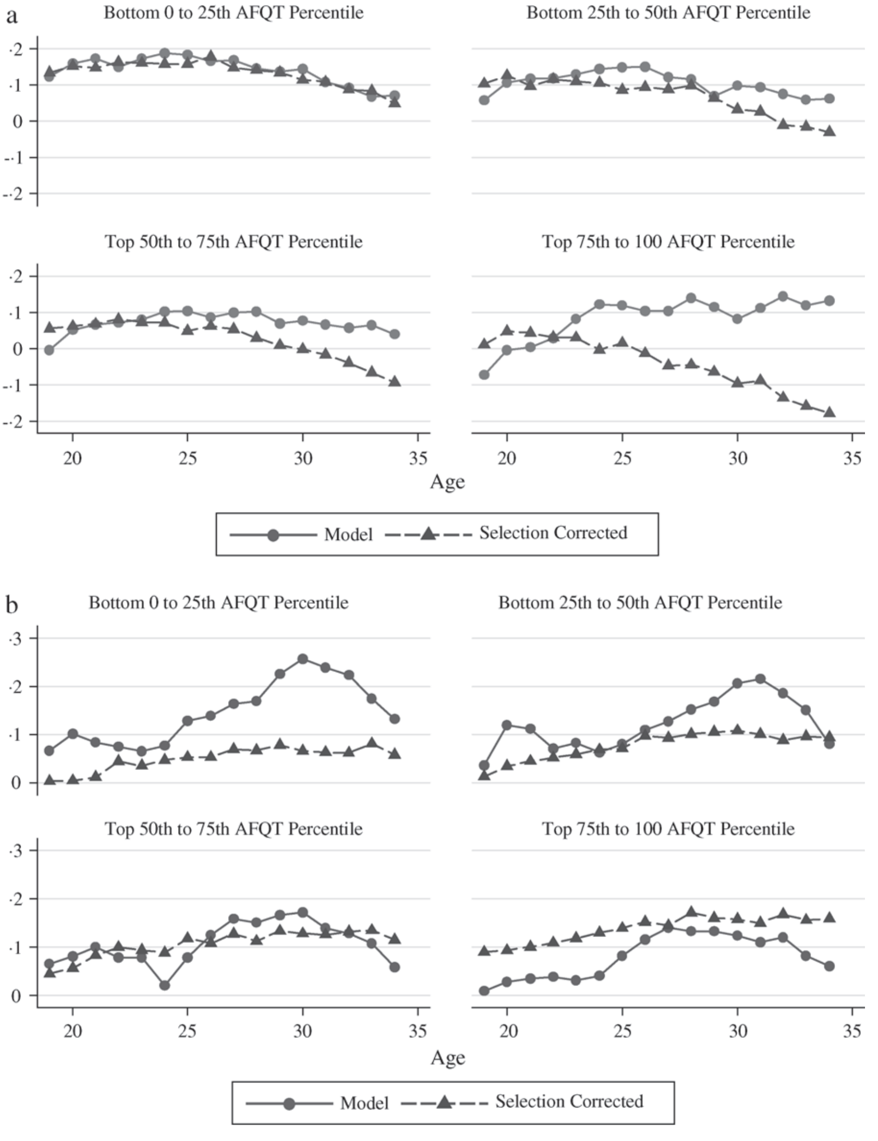

FIGURE 4

City wage premium (a) in blue collar and (b) in white collar

work and 0·017 for white-collar work in the rural sector. These estimates show that work experience acquired in the city is worth more than work experience acquired in the rural area for white-collar work in the rural area, but not for blue-collar work. For example, five years of city work experience increases the wage of a white-collar worker in the rural area by 8·5% compared to a similar worker with the same total level of work experience which was acquired entirely in the rural area. This result suggests that the positive city wage premium seen in Figure 3(b) for white-collar work is likely due to the idea that cities make workers more productive vs. competing theories which predict that human capital acquired in the city should not be rewarded more in the rural area.

5. CONCLUSION

This paper estimates the city wage premium in each occupation, after controlling for the non-random allocation of individuals across locations in their youth and the complete history of endogenous choices over location and career outcomes over time. The results indicate that the observed wage gap for blue-collar work is almost entirely explained by differences in the types of workers across locations. In contrast, differences in the types of workers across locations explain only 34% of the observed gap in wages for the white-collar sector at the age of 30. Therefore, a given worker does appear to be paid more for white-collar work in the city vs. the rural area, and these gains are the largest for high-ability individuals. In contrast, we find no evidence of a true city wage premium for blue-collar work. The estimates also suggest that human capital gains acquired from working in the city vs. the rural area are transferable to the rural area for white-collar work, but not for blue-collar work.

These results shed light on the importance of several competing explanations for the observed gap in wages between workers in cities vs. rural areas. The idea that firms in cities pay higher wages because workers are inherently better is true, but mostly for the blue-collar sector. For the white-collar sector, firms seem to pay more in cities for the same worker. Although there are several reasons for why a white-collar worker is paid more in a city, the evidence is most supportive of the idea that cities make white-collar workers more productive, since experience acquired in the city is rewarded favourably in the rural area. Glaeser and Mare (2001) also interpreted their results as evidence for higher productivity levels in cities, but they did not break their results down by occupational sector, and therefore, did not identify different results between the two occupations. In addition, their empirical strategy relied on the assumption that changing locations is random conditional on the individual's fixed effect. The evidence presented in this paper shows that moving locations is correlated with wages, occupational choices, and occupational switching. Therefore, the estimates in this paper are the first to explicitly account for these correlations by modelling the joint decision over occupation and location.

Explaining why white-collar workers, and not blue-collar workers, earn more in cities is not straightforward, and the model cannot identify the precise mechanism driving the different results for each occupation. One possibility is that human capital externalities in cities are more important for white-collar work vs. blue-collar work, which may be due to differences in the way workers interact and learn from one another. However, this possibility should be investigated along with other explanations in future work.

APPENDIX

TABLE A.1

Model estimates

	Log weekly wage functions			
	City		Rural	
	Blue collar	White collar	Blue collar	White collar
Type 1 intercept	4·774	4·678	5·002	4·663
	(0·0059)	(0·0047)	(0·0065)	(0·0056)
Type 1 experience	0·079	0·071	0·064	0·042
	(0·0011)	(0·0035)	(0·0011)	(0·0021)
Type 1 experience squared/100	−0·198	−0·317	−0·234	−0·047
	(0·0120)	(0·0701)	(0·0133)	(0·0137)

TABLE A.1—continued

Type 2 intercept	4·991	4·569	4·815	4·526
	(0·0067)	(0·0077)	(0·0075)	(0·0067)
Type 2 experience	0·066	0·076	0·081	0·0706
	(0·0015)	(0·0022)	(0·0013)	(0·0048)
Type 2 experience squared/100	−0·391	−0·270	−0·207	−0·325
	(0·0136)	(0·0318)	(0·0141)	(0·0636)
Type 3 intercept	5·187	5·015	5·007	4·859
	(0·0067)	(0·0060)	(0·0070)	(0·0066)
Type 3 experience	0·068	0·097	0·102	0·082
	(0·0013)	(0·0009)	(0·0005)	(0·0021)
Type 3 experience squared/100	−0·350	−0·336	−0·249	−0·283
	(0·0227)	(0·0107)	(0·016)	(0·0215)
HS graduate	0·025	0·106	0·038	0·146
	(0·0039)	(0·0091)	(0·0062)	(0·0134)
College graduate	0·137	0·254	0·106	0·265
	(0·0149)	(0·0088)	(0·0190)	(0·017)
Education	0·032	0·062	0·024	0·060
	(0·0025)	(0·0015)	(0·0030)	(0·0026)
Standardized AFQT	−0·007	0·0367	0·061	0·017
	(0·0021)	(0·0031)	(0·0036)	(0·0033)
City experience	0·018	−0·0001	0·003	0·017
	(0·0014)	(0·0001)	(0·0007)	(0·0015)
Under 18 dummy	−0·308	−0·416	0·038	−0·775
	(0·045)	(0·0916)	(0·0283)	(0·3371)
Experience in the other occupation	0·0323	0·052	0·026	0·045
	(0·0018)	(0·0012)	(0·0017)	(0·0024)
Experience squared in other occupation/100	−0·308	−0·1204	−0·031	−0·232
	(0·0450)	(0·0117)	(0·0136)	(0·0238)
Worked in same occupation in previous period	0·156	0·149	0·150	0·151
	(0·0114)	(0·0115)	(0·0086)	(0·0155)

Probability functions for types of men
(multivariate logit)

	Type 2	Type 3
Intercept	−0·233	−1·537
	(0·0331)	(0·0719)
Standardized AFQT	0·125	0·232
	(0·0275)	(0·037)
Lived with both parents at age 14	0·251	1·003
	(0·0507)	(0·0788)
Mother high-school graduate	0·350	0·558
	(0·0658)	(0·066)
Mother college graduate	0·551	1·202
	(0·1334)	(0·1428)
Father high-school graduate	0·250	0·402
	(0·0610)	(0·0615)
Father college graduate	0·338	0·870
	(0·078)	(0·1020)
Estimated percentage of men of each type	40·1%....	33·5%

Job offer functions from the home sector
(logit functions)

	Blue collar	White collar
Intercept	−0·339	−1·053
	(0·0495)	(0·0835)

TABLE A.1—continued

Current period non-pecuniary utilities	
Home sector (log utility)	
Type 1 male intercept	4·862
	(0·0159)
Type 2 male intercept	4·883
	(0·0123)
Type 3 male intercept	4·915
	(0·0163)
Age	0·005
	(0·0007)
Age squared/100	−0·056
	(0·0064)
Standardized AFQT	−0·009
	(0·0029)
School sector	
Intercept	−16·476
	(2·434)
Under 18	146·075
	(26·3893)
High school	590·711
	(79·0841)
College	−108·659
	(5·4994)
Post-college	182·113
	(13·125)
Living in the rural area	
Type 1 intercept	87·911
	(2·7001)
Type 2 intercept	−46·855
	(2·953)
Type 3 intercept	−24·968
	(2·1421)
Standardized AFQT	−8·133
	(0·9097)
Current period costs	
Moving costs	
Intercept (from city to rural)	−11·8078
	(3·7095)
Intercept (from rural to city)	−19·259
	(4·9018)
Age 18 or under	−72·965
	(13·214)
Age	−11·660
	(0·4595)
Age squared/100	−345·319
	(29·8358)
School in previous period	−227·223
	(9·0836)
Entry costs into sectors	
High school	−206·393
	(31·0543)
College	−357·325
	(12·8692)
White collar	−163·281
	(5·8299)
Blue collar	−134·459
	(5·9347)

TABLE A.1—continued

Terminal value Emax function	
Working in blue collar	71·816 (16·4739)
Working in white collar	98·679 (19·0094)
Education	269·051 (13·3697)
Blue-collar experience	166·112 (4·1140)
White-collar experience	191·640 (4·5008)
Education ∗ AFQT	170·078 (4·9459)
Blue collar Experience ∗ AFQT	42·884 (3·1315)
White collar Experience ∗ AFQT	72·371 (3·0655)

Cholesky decomposition matrix of shocks

	City				Rural			
	Log home utility	School utility	Log blue-collar wage	Log white-collar wage	Log home utility	School utility	Log blue-collar wage	Log white-collar wage
City								
Log home utility	0·948 (0·011)							
School utility		287·511 (5·671)						
Log blue-collar wage			0·491 (0·004)					
Log white-collar wage			−0·016 (0·004)	0·396 (0·003)				
Rural								
Log home utility					0·923 (0·010)			
School utility						232·604 (5·152)		
Log blue-collar wage							0·478 (0·004)	
Log white-collar wage							−0·036 (0·008)	0·401 (0·006)

TABLE A.1—continued

Initial conditions of parental background		
Lived with both parents at age 14 (logit function)		
Intercept	1·624	
	(0·063)	
AFQT	0·350	
	(0·075)	
	HS graduate	College graduate
Mother's education (multivariate logit)		
Intercept	0·952	−1·007
	(0·134)	(0·224)
AFQT	0·810	1·678
	(0·076)	(0·122)
Lived with both parents at 14	0·304	0·264
	(0·145)	(0·240)
Father's education (multivariate logit)		
Intercept	−0·790	−3·301
	(0·166)	(0·341)
AFQT	0·427	1·114
	(0·078)	(0·106)
Lived with both parents at 14	0·235	0·605
	(0·153)	(0·229)
Mother HS graduate	1·780	2·696
	(0·131)	(0·294)
Mother college graduate	2·532	5·520
	(0·420)	(0·481)
Living in a city at age 16 (logit)		
Intercept	0·382	
	(0·236)	
AFQT	0·006	
	(0·103)	
Lived with both parents at 14	−0·168	
	(0·202)	
Mother HS graduate	−0·003	
	0·202	
Mother college graduate	0·090	
	(0·336)	
Father HS graduate	0·570	
	(0·193)	
Father college graduate	0·757	
	(0·275)	

Notes: Estimated standard errors are in parentheses. Numbers that appear without S.E. are derived from other estimated parameters (with S.E.) or from the simulated data. All parameter estimates were estimated within the maximization of the likelihood function except for the initial conditions for parental background, which were estimated separately outside the model in order to start the simulation procedure with simulated agents who have similar family background characteristics to the sample observed in the National Longitudinal Survey of Youth data (the logits for these characteristics are designed to produce the same proportion of agents with each background characteristic and the correlations between those characteristics). AFQT, armed forces qualification test; HS, high school.

Acknowledgements. I thank Simcha Srebnik for critical computer support, as well as Charles Manski, Michael Keane, and Daniele Paserman, the editor, and two anonymous referees for valuable comments.

REFERENCES

BELLMAN, R. (1957) *Dynamic Programming* (Princeton, NJ: Princeton University Press).

BORJAS, G. J. (1987), "Self-Selection and the Earnings of Immigrants", *American Economic Review*, **77**, 531–553.

ECKSTEIN, Z. and WOLPIN, K. I. (1989), "Dynamic Labour Force Participation of Married Women and Endogenous Work Experience", *Review of Economic Studies*, **56**, 375–390.

GLAESER, E. L. (1999), "Learning in Cities", *Journal of Urban Economics*, **46**, 254–277.

GLAESER, E. L. and MARE, D. C. (2001), "Cities and Skills", *Journal of Labor Economics*, **19** (2), 316–342.

GOULD, E. D. (2002), "Rising Wage Inequality, Comparative Advantage, and the Growing Importance of General Skills in the United States", *Journal of Labor Economics*, **20** (1), 105–147

HAM, J. C., LI, X. and REAGAN, P. B. (2001), "Matching and Selection Estimates of the Effect of Migration on Wages for Young Men" (Ohio State Working Paper, August 2001).

HECKMAN, J. J. (1981), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence", in C. F. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications* (Cambridge, MA: MIT Press) 179–195.

HECKMAN, J. J. and SEDLACEK, G. L. (1985), "Heterogeneity, Aggregation and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market", *Journal of Political Economy*, **93**, 1077–1125.

HECKMAN, J. J. and SINGER, B. (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, **2**, 271–320.

HOTZ, V. J. and MILLER, R. A. (1988), "An Empirical Analysis of Life Cycle Fertility and Female Labor Supply", *Econometrica*, **56**, 91–118.

HOTZ, V. J., MILLER, R. A., SANDERS, S. and SMITH, J. (1994), "A Simulation Estimator for Dynamic Models of Discrete Choice", *Review of Economic Studies*, **61** (2), 265–289.

KEANE, M. P. and WOLPIN, K. I. (1994), "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence", *Review of Economics and Statistics*, **17**, 648–672.

KEANE, M. P. and WOLPIN, K. I. (1997), "The Career Decisions of Young Men", *Journal of Political Economy*, **105**, 473–522.

KEANE, M. P. and WOLPIN, K. I. (2001), "The Effect of Borrowing Constraints on Educational Attainment", *International Economic Review*, **42** (4), 1051–1103.

LERMAN, S. R. and MANSKI, C. R. (1981), "On the Use of Simulated Frequencies to Approximate Choice Probabilities", in C. F. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications* (Cambridge, MA: MIT Press) 305–319.

LUCAS, R. E., Jr. (2001), "Externalities and Cities", *Review of Economic Dynamics*, **4** (2), 245–274.

RAUCH, J. E. (1993), "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities", *Journal of Urban Economics*, **34** (3), 380–400.

ROBACK, J. (1982), "Wages, Rents, and the Quality of Life", *Journal of Political Economy*, **90**, 1257–1278.

ROY, A. D. (1951), "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers*, **3**, 135–146.

RUST, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Howard Zurcher", *Econometrica*, **55** (5), 999–1033.

STERN, S. (1997), "Simulation-Based Estimation", *Journal of Economic Literature*, **35** (4), 2006–2039.

TABER, C. R. (2001), "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?", *Review of Economic Studies*, **68** (3), 665–691.

WHEATON, W. C. and LEWIS, M. J. (2002), "Urban Wages and Labor Market Agglomeration", *Journal of Urban Economics*, **51** (3), 542–562.