Occupations and Import Competition: Evidence from Denmark

Sharon Traiberman

ONLINE APPENDIX

A Table and Figure Appendix (For Online Publication Only)

A.1 Sructural Model Parameters

Table A.1: Time Averaged Skill Prices Across Occupations

Occupation	Manufacturing	Other Services	FIRE	Health & Educ.
Managers	0.553	0.474	0.452	
	(0.023)	(0.013)	(0.020)	
Science Professional	0.627	0.470	0.526	
	(0.017)	(0.017)	(0.015)	
Science Assc. Professional	0.462	0.456	0.385	
	(0.011)	(0.014)	(0.013)	
Other Assc. Professional	0.394	0.360	0.418	
	(0.009)	(0.005)	(0.008)	
Clerks	0.285	0.282	0.205	0.183
	(0.007)	(0.004)	(0.006)	(0.006)
Agriculture	0.459			
	(0.023)			
Building Trades	0.346			
	(0.006)			
Metal Trades	0.455	0.438		
	(0.008)	(0.009)		
Other Crafts	0.329			
	(0.011)			
Plant Operator	0.321			
	(0.014)			
Machine Operator	0.399			
	(0.005)			
Drivers	0.503	0.373		
	(0.020)	(0.010)		
Laborers	0.405	0.255		
	(0.008)	(0.006)		
Other Professional		0.262	0.238	
		(0.006)	(0.007)	
Personal Workers		0.159		0.233
		(0.003)		(0.002)
Retail Workers		0.154		
		(0.003)		
Elementary Occupations		0.150		
		(0.002)		
Customer Service			0.356	
			(0.007)	
Health Professional				0.462
				(0.017)
Teachers				0.213
				(0.003)
Health Assc. Professional				0.340
				(0.005)
Teaching Assc. Professional				0.312
				(0.004)

Notes: Skill prices time-averaged for clarity. Units are relative to unconditional mean income (normalized to 1). Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

Table A.2: Income Regression Coefficients

(a) Manufacturing

	β_{Ten}	β_{Age}	β_{Age^2} [a]	$\beta_{Type:1}$	$\beta_{Type:2}$	$\beta_{Type:3}$	$\beta_{Type:4}$	$\beta_{Type:5}$	σ^2
Managers	0.035	0.042		-1.248	-0.123	-0.906	-0.070	-0.706	0.147
	(0.001)	(0.006)		(0.032)	(0.006)	(0.025)	(0.004)	(0.031)	(0.002)
Science Prof.	0.058	0.033				-0.442	-0.033	-0.266	0.147
	(0.001)	(0.002)				(0.022)	(0.004)	(0.013)	(0.001)
Science Assc. Prof.	0.041	0.046		-0.953	-0.163	-0.526	-0.116	-0.485	0.126
	(0.001)	(0.004)		(0.031)	(0.005)	(0.005)	(0.003)	(0.021)	(0.001)
Other Assc. Prof.	0.064	0.003	-1.000	-0.679	-0.123	-0.464	-0.074	-0.346	0.156
	(0.001)	(0.011)	(0.255)	(0.023)	(0.004)	(0.008)	(0.003)	(0.011)	(0.001)
Clerks	0.063	0.014	-1.000	-0.867	-0.174	-0.515	-0.107	-0.610	0.157
	(0.001)	(0.001)	(0.000)	(0.016)	(0.006)	(0.009)	(0.006)	(0.022)	(0.001)
Agriculture	0.009	0.034		-1.105	-0.053	-0.941	0.033	-1.172	0.200
	(0.001)	(0.002)		(0.022)	(0.017)	(0.022)	(0.018)	(0.035)	(0.002)
Building Trades	0.040	0.044		-0.711	-0.046	-0.624			0.236
	(0.001)	(0.001)		(0.008)	(0.003)	(0.004)			(0.001)
Metal Trades	0.042	0.037		-0.588	-0.014	-0.456	0.052	-0.884	0.179
	(0.001)	(0.001)		(0.014)	(0.006)	(0.009)	(0.005)	(0.028)	(0.001)
Other Crafts	0.055	0.047		-0.774	-0.167	-0.574			0.206
	(0.001)	(0.020)		(0.014)	(0.004)	(0.009)			(0.002)
Plant Operator	0.044	0.003	-1.000	-0.573	0.021	-0.285	0.146	-0.748	0.160
	(0.001)	(0.008)	(0.196)	(0.022)	(0.006)	(0.013)	(0.007)	(0.022)	(0.002)
Machine Operator	0.049	0.038		-0.683	-0.112	-0.461			0.187
	(0.000)	(0.001)		(0.006)	(0.002)	(0.004)			(0.001)
Drivers	0.055	0.032		-0.682	-0.044	-0.501			0.159
	(0.001)	(0.002)		(0.022)	(0.004)	(0.023)			(0.002)
Laborers	0.059	0.036		-0.754	-0.063	-0.501			0.194
	(0.001)	(0.001)		(0.009)	(0.002)	(0.006)			(0.001)

(b) Services

	β_{Ten}	β_{Age}	β_{Age^2} [a]	$\beta_{Type:1}$	$\beta_{Type:2}$	$\beta_{Type:3}$	$\beta_{Type:4}$	$\beta_{Type:5}$	σ^2
Managers	0.032	0.044		-1.215	-0.067	-0.913	-0.034	-1.028	0.177
	(0.001)	(0.001)		(0.024)	(0.005)	(0.021)	(0.004)	(0.034)	(0.002)
Science Prof.	0.055	0.040		-1.394	-0.001	-1.154	0.047	-0.985	0.153
	(0.001)	(0.002)		(0.017)	(0.006)	(0.032)	(0.005)	(0.027)	(0.002)
Other Prof.	0.049	0.017	-1.000	-1.131	-0.019	-0.930	0.005	-0.570	0.176
	(0.001)	(0.001)	(0.000)	(0.013)	(0.003)	(0.023)	(0.002)	(0.007)	(0.001)
Science Assc. Prof.	0.037	0.046		-1.063	-0.093	-0.688	-0.082	-0.863	0.141
	(0.001)	(0.021)		(0.019)	(0.004)	(0.017)	(0.004)	(0.027)	(0.001)
Other Assc. Prof.	0.042	0.002	-1.000	-0.864		-0.464	0.033	-0.525	0.155
	(0.001)	(0.019)	(0.421)	(0.011)		(0.005)	(0.002)	(0.008)	(0.001)
Clerks	0.043	0.008	-1.000	-0.953	-0.041	-0.525	-0.003	-0.992	0.143
	(0.000)	(0.001)	(0.000)	(0.008)	(0.004)	(0.005)	(0.004)	(0.008)	(0.001)
Personal Services	0.091	0.019	-1.000	-0.873	0.017	-0.601	0.126	-0.879	0.231
	(0.001)	(0.001)	(0.000)	(0.008)	(0.007)	(0.009)	(0.007)	(0.011)	(0.001)
Retail Occs.	0.045	0.034	-1.000	-0.971	-0.181	-0.567			0.234
	(0.001)	(0.001)	(0.000)	(0.006)	(0.004)	(0.004)			(0.001)
Metal Trades	0.040	0.042		-0.856	-0.043	-0.723			0.146
	(0.001)	(0.001)		(0.020)	(0.004)	(0.018)			(0.001)
Drivers	0.056	0.044		-0.719	-0.065	-0.454			0.204
	(0.001)	(0.016)		(0.012)	(0.003)	(0.007)			(0.001)
Elementary Occs.	0.039	0.018	-1.000	-0.772	-0.034	-0.521	0.128	-0.850	0.215
	(0.001)	(0.001)	(0.000)	(0.007)	(0.006)	(0.007)	(0.006)	(0.009)	(0.001)
Laborers	0.083	0.007	-1.000	-0.928	-0.058	-0.552			0.237
	(0.001)	(0.001)	(0.000)	(0.011)	(0.003)	(0.011)			(0.002)

(c) FIRE Industries

	β_{Ten}	β_{Age}	β_{Age^2} [a]	$\beta_{Type:1}$	$\beta_{Type:2}$	$\beta_{Type:3}$	$\beta_{Type:4}$	$\beta_{Type:5}$	σ^2
Managers	0.026	0.040		-1.431	-0.091	-1.295	-0.012	-1.165	0.128
	(0.001)	(0.002)		(0.023)	(0.006)	(0.020)	(0.004)	(0.022)	(0.001)
Science Prof.	0.052	0.039		-1.576	-0.030	-0.412	0.009	-0.334	0.180
	(0.001)	(0.001)		(0.023)	(0.005)	(0.018)	(0.003)	(0.011)	(0.001)
Other Prof.	0.052	0.028	-1.000	-1.049	-0.064	-0.399	0.025	-0.540	0.166
	(0.001)	(0.002)	(0.000)	(0.036)	(0.005)	(0.025)	(0.003)	(0.019)	(0.002)
Science Assc. Prof.	0.050	0.006	-1.000	-0.964	-0.072	-0.525	-0.064	-0.457	0.165
	(0.001)	(0.001)	(0.000)	(0.021)	(0.005)	(0.010)	(0.004)	(0.013)	(0.001)
Other Assc. Prof.	0.043	0.048		-0.902	-0.107	-0.567	-0.037	-0.508	0.156
	(0.001)	(0.022)		(0.016)	(0.003)	(0.010)	(0.003)	(0.014)	(0.001)
Clerks	0.067	0.024	-1.000	-0.981	-0.141	-0.568	-0.050	-0.899	0.206
	(0.001)	(0.001)	(0.000)	(0.013)	(0.005)	(0.010)	(0.005)	(0.013)	(0.002)
Customer Service	0.029	0.045		-0.919	-0.086	-0.485			0.135
	(0.001)	(0.001)		(0.008)	(0.003)	(0.005)			(0.001)

(d) Health & Education

	β_{Ten}	β_{Age}	β_{Age^2} [a]	$\beta_{Type:1}$	$\beta_{Type:2}$	$\beta_{Type:3}$	$\beta_{Type:4}$	$\beta_{Type:5}$	σ^2
Health Prof.	0.029	0.039						-0.463	0.186
	(0.001)	(0.002)						(0.006)	(0.001)
Teachers	0.044	0.018	-1.000	-1.271	-0.104	-1.097	-0.021	-0.842	0.127
	(0.001)	(0.001)	(0.000)	(0.006)	(0.003)	(0.012)	(0.002)	(0.007)	(0.001)
Health Assc. Prof.	0.022	0.042		-1.153	-0.218	-0.599	0.013	-0.501	0.139
	(0.001)	(0.001)		(0.006)	(0.005)	(0.011)	(0.003)	(0.005)	(0.001)
Teaching Assc. Prof.	0.036	0.038		-1.203	-0.082	-1.216	0.005	-0.515	0.142
	(0.001)	(0.001)		(0.004)	(0.003)	(0.005)	(0.003)	(0.005)	(0.001)
Clerks	0.065	0.006	-1.000	-0.767	0.004	-0.345	0.096	-0.871	0.154
	(0.001)	(0.009)	(0.196)	(0.018)	(0.008)	(0.009)	(0.008)	(0.018)	(0.001)
Personal Services	0.056	0.042		-0.807	-0.041	-0.413	0.005	-0.755	0.153
	(0.001)	(0.000)		(0.004)	(0.003)	(0.004)	(0.003)	(0.009)	(0.001)

[a] Presented $\times 10^3$ for clarity.

[b] Coefficients from a log-linear Mincer regression of wages on worker attributes. Types refer to estimates of unobservable heterogeneity across workers. Low skilled workers are either of type 1 or 2; Medium skilled workers are of type 3 or 4; High skilled workers are of type 5 or 6. Type 6 coefficients all normalized to 0. Skill prices are in table A.1. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

Table A.3: Mobility Productivity Parameters

Age	0.008
	(0.000)
	[0.001]
$\text{Age}^2(\times 1000)$	0.000
	(0.000)
	[0.000]
Type 1	0.027
	(0.010)
	[0.016]
Type 2	0.120
	(0.009)
	[0.016]
Type 3	0.043
	(0.009)
	[0.014]
Type 4	0.150
	(0.009)
	[0.018]
Type 5	-0.018
	(0.008)
	[0.013]

Notes: Coefficients from a log-linear inverse productivity function. Types refer to estimates of unobservable heterogeneity across workers. Low skilled workers are either of type 1 or 2; Medium skilled workers are of type 3 or 4; High skilled workers are of type 5 or 6. Type 6 coefficients all normalized to 0. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

Table A.4: Mobility Cost (Γ) Parameters

	Up Tasking	Down Tasking
Constant	0.723	
	(0.009)	
	[0.016]	
Occ. Dummy	0.142	
	(0.003)	
	[0.002]	
Sec. Dummy	0.258	
	(0.005)	
	[0.007]	
Task 1	0.017	0.021
	(0.007)	(0.008)
	[0.013]	[0.015]
Task 2	0.035	0.018
	(0.005)	(0.005)
	[0.009]	[0.008]
Task 3	0.040	0.036
	(0.028)	(0.028)
	[0.049]	[0.041]
Task 4	0.110	−0.101
	(0.027)	(0.029)
	[0.054]	[0.060]
Task 5	0.126	−0.037
	(0.011)	(0.013)
	[0.019]	[0.024]
Task 6	−0.071	0.111
	(0.040)	(0.037)
	[0.067]	[0.060]
Task 7	0.076	0.024
	(0.031)	(0.030)
	[0.061]	[0.055]
Task 8	0.207	−0.198
	(0.034)	(0.045)
	[0.053]	[0.071]
Task 9	0.043	0.124
	(0.033)	(0.033)
	[0.041]	[0.044]
Task 10	−0.054	0.013
	(0.057)	(0.057)
	[0.109]	[0.109]

Notes: Coefficients from a log-linear cost function featuring a constant, a dummy for switching occupations, a dummy for switching sectors, and coefficients for moving in task space. The cost function is naturally scaled by the variance of shocks, ρ, and results for the constant are not presented adjusted. The first column presents the coefficients for moving up in task space and second column presents coefficients for moving down. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

Table A.5: Non-Pecuniary Benefits (η) Parameters

Occupation	Manufacturing	Other Services	FIRE	Health & Educ.
Managers	1.761	1.496	1.992	
	(0.094)	(0.087)	(0.109)	
	[0.325]	[0.280]	[0.333]	
Science Professional	1.513	1.723	1.409	
	(0.116)	(0.103)	(0.110)	
	[0.345]	[0.340]	[0.298]	
Science Assc. Professional	1.438	1.528	1.623	
	(0.091)	(0.086)	(0.094)	
	[0.328]	[0.296]	[0.312]	
Other Assc. Professional	1.429	1.645	1.596	
	(0.090)	(0.093)	(0.095)	
	[0.294]	[0.296]	[0.327]	
Clerks	1.565	1.751	1.609	2.223
	(0.083)	(0.087)	(0.087)	(0.100)
	[0.291]	[0.277]	[0.291]	[0.303]
Agriculture	1.471			
	(0.082)			
	[0.271]			
Building Trades	1.468			
	(0.084)			
	[0.285]			
Metal Trades	1.573	1.746		
	(0.089)	(0.086)		
	[0.281]	[0.302]		
Other Crafts	1.321			
	(0.082)			
	[0.274]			
Plant Operator	0.847			
	(0.080)			
	[0.279]			
Machine Operator	1.714			
	(0.095)			
	[0.314]			
Drivers	1.489	1.580		
	(0.084)	(0.094)		
	[0.265]	[0.301]		
Laborers	1.810	1.647		
	(0.096)	(0.092)		
	[0.299]	[0.304]		
Other Professional		1.534	1.435	
		(0.097)	(0.104)	
		[0.289]	[0.326]	
Personal Workers		1.903		2.118
		(0.093)		(0.092)
		[0.299]		[0.303]
Retail Workers		1.682		
		(0.086)		
		[0.277]		
Elementary Occupations		2.104		
		(0.094)		
		[0.304]		
Customer Service			1.392	
			(0.082)	
			[0.282]	
Health Professional				1.795
				(0.118)
				[0.346]
Teachers				2.159
				(0.105)
				[0.303]
Health Assc. Professional				1.895
				(0.106)
				[0.314]
Teaching Assc. Professional				1.665
				(0.090)
				[0.279]

Notes: Non-pecuniary benefits to each occupation and sector cell. Blanks occur because not all occupations are present in all sectors. The Health & Education sector reflects public sector and does not include things like R&D. Data appendix contains list of industry codes in each sector. Units are proportional the unconditional sample mean income. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

Table A.6: Non-Employment (u) Parameters

Age	-0.020
	(0.003)
	[0.007]
$\text{Age}^2 (\times 1000)$	0.700
	(0.075)
	[0.239]
Type 1	1.400
	(0.071)
	[0.223]
Type 2	0.486
	(0.035)
	[0.072]
Type 3	1.380
	(0.063)
	[0.202]
Type 4	-0.433
	(0.045)
	[0.084]
Type 5	1.199
	(0.066)
	[0.214]

Notes: Coefficients from a quadratic specification for the virtual value of non-employment. Types refer to estimates of unobservable heterogeneity across workers. Low skilled workers are either of type 1 or 2; Medium skilled workers are of type 3 or 4; High skilled workers are of type 5 or 6. Type 6 coefficients all normalized to 0. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

A.2 Data Appendix Tables

Table A.7: Comparing Raw Data and Sample Frame

	Raw Data			Sample		
	Mean	Q25	Q75	Mean	Q25	Q75
Income (1000s)	230.80	151.38	229.72	93.04		
Age	40.94	10.45	41.08	10.08		
Low Skilled (%)	33.57		28.57			
Med. Skilled (%)	42.65		46.28			
High Skilled (%)	21.46		25.14			
Manufacturing (%)	28.08		25.66			
Services (%)	35.24		32.72			
FIRE (%)	12.06		10.36			
Health & Educ. (%)	24.62		31.26			
Employment (%)	83.15		91.51			
Observations	33502593		18513301			

Notes: All variables pooled across units and time periods. Raw data refers to collected data on all units in IDAS from 1996-2007 between 23 and 59. Sample data includes imputed, cleaned and dropped data. See data appendix for complete description of sample creation. Income is in 1000s of DKK deflated by the 2000 Danish CPI. Low skilled mean completion of short-cycle or less education; medium-skilled implies vocational or medium-cycle education; high skilled means university and higher. See the data appendix for a mapping of NACE 1.1 industries to sectors.

Table A.8: Sector Level Summary Statistics

Age	40.55	40.79	40.51	42.05
Tenure	3.17	2.92	2.51	3.76
Low Skilled (%)	33.65	33.75	19.09	18.31
Med. Skilled (%)	59.42	51.10	51.57	30.24
High Skilled (%)	6.92	15.15	29.34	51.45
Observations	4346990	5542706	1754626	5296437

Notes: Pooled across time periods. Averages are taken with respect to individuals in the estimation sample, conditional on employment.

Table A.9: Transition Matrix Across Sectors

	Man.	Services	FIRE	H&E
Manufacturing	93.11	4.67	1.18	1.04
Services	3.64	89.84	2.51	4.01
FIRE	2.36	6.26	89.16	2.22
H & E	0.56	3.39	1.00	95.05

Notes: Pooled across time periods. Averages are taken with respect to individuals in the estimation sample, conditional on employment.

Table A.10: Occupation Aggregation Breakdown

(a) Manufacturing

Occupation	DISCO 3 Title	% Represented
Managers	Other Department Managers	35.94
	Production And Operations Department Managers	31.52
	General Managers	16.26
	Directors And Chief Executives	12.76
Science Professional	Architects, Engineers And Related Profs.	80.17
	Computing Profs.	12.53
Science Assc. Professional	Physical And Engineering Science Technicians	80.13
	Computer Associate Profs.	8.45
	Safety And Quality Inspectors	7.49
Other Assc. Professional	Finance And Sales Associate Profs.	47.41
	Administrative Associate Profs.	40.27
	Business Services Agents And Trade Brokers	6.73
Clerks	Secretaries And Keyboard-Operating Clerks	53.96
	Material-Recording And Transport Clerks	21.62
	Numerical Clerks	10.81
	Other Office Clerks	8.56
Agriculture	Market Gardeners And Crop Growers	40.94
	Market-Oriented Crop And Animal Producers	26.10
	Agricultural, Fishery And Related Laborers	14.43
	Market-Oriented Animal Producers And Related Workers	9.69
Building Trades	Building Frame And Related Trades Workers	45.92
	Building Finishers And Related Trades Workers	38.09
	Painters, Building Structure Cleaners And Related Trades Workers	13.47
Metal Trades	Blacksmiths, Tool-Makers And Related Trades Workers	32.68
	Metal Moulders, Welders,..., And Related Trades	29.19
	Machinery Mechanics And Fitters	21.27
	Electrical And Electronic Equipment Mechanics And Fitters	13.04
Other Crafts	Food Processing And Related Trades Workers	34.45
	Printing And Related Trades Workers	22.63
	Wood Treaters, Cabinet-Makers And Related Trades Workers	20.58
	Textile, Garment And Related Trades Workers	6.35
	Precision Workers In Metal And Related Materials	6.11
Plant Operator	Chemical-Processing-Plant Operators	34.57
	Metal-Processing-Plant Operators	16.99
	Power-Production And Related Plant Operators	9.12
	Wood-Processing- And Papermaking-Plant Operators	6.79
	Glass, Ceramics And Related Plant Operators	6.08
	Mining- And Mineral-Processing Plant Operators	5.50
Machine Operator	Food And Related Products Machine Operators	26.69
	Assemblers	22.03
	Metal- And Mineral-Products Machine Operators	11.97
	Other Machine Operators And Assemblers	10.21
	Rubber- And Plastic-Products Machine Operators	6.97
	Printing-, Binding- And Paper-Products Machine Operators	6.57
	Wood-Products Machine Operators	5.37
Drivers	Agricultural And Other Mobile-Plant Operators	48.67
	Motor-Vehicle Drivers	47.79
Laborers	Mining And Construction Laborers	55.19
	Manufacturing Laborers	30.57
	Transport Laborers And Freight Handlers	11.09

(b) Services

Occupation	DISCO 3 Title	% Represented
Managers	Other Department Managers	35.53
	Production And Operations Department Managers	28.46
	General Managers	18.86
	Directors And Chief Executives	10.84
Science Professional	Architects, Engineers And Related Profs.	65.29
	Computing Profs.	27.59
Other Professional	Business Profs.	32.57
	Social Science And Related Profs.	23.96
	Writers And Creative Or Performing Artists	17.49
	Legal Profs.	10.77
	Archivists, Librarians, And Related Information Profs.	9.72
Science Assc. Professional	Physical And Engineering Science Technicians	53.90
	Computer Associate Profs.	22.24
	Ship And Aircraft Controllers And Technicians	10.58
	Optical And Electronic Equipment Operators	8.03
Other Assc. Professional	Administrative Associate Profs.	44.25
	Finance And Sales Associate Profs.	29.90
	Customs, Tax And Related Government Associate Profs.	9.77
	Business Services Agents And Trade Brokers	6.29
Clerks	Secretaries And Keyboard-Operating Clerks	57.54
	Library, Mail And Related Clerks	19.69
	Material-Recording And Transport Clerks	10.64
	Other Office Clerks	6.31
Personal Workers	Protective Services Workers	34.66
	Housekeeping And Restaurant Services Workers	25.38
	Personal Care And Related Workers	19.31
	Travel Attendants And Related Workers	9.30
	Other Personal Services Workers	5.48
Retail Workers	Shop Salespersons And Demonstrators	96.36
Metal Trades	Machinery Mechanics And Fitters	65.01
	Electrical And Electronic Equipment Mechanics And Fitters	20.34
	Blacksmiths, Tool-Makers And Related Trades Workers	7.29
	Metal Moulders, Welders,..., And Related Trades	6.49
Drivers	Motor-Vehicle Drivers	74.38
	Agricultural And Other Mobile-Plant Operators	12.98
	Locomotive-Engine Drivers And Related Workers	9.25
Elementary Occupations	Domestic And Related Helpers, Cleaners And Launderers	70.74
	Building Caretakers, Window And Related Cleaners	17.81
Laborers	Transport Laborers And Freight Handlers	72.56
	Manufacturing Laborers	15.62
	Mining And Construction Laborers	8.98

(c) FIRE Industries

Occupation	DISCO 3 Title	% Represented
Managers	Production And Operations Department Managers	57.47
	General Managers	22.92
	Other Department Managers	12.03
Science Professional	Architects, Engineers And Related Profs.	47.94
	Computing Profs.	44.25
Other Professional	Business Profs.	75.04
	Legal Profs.	9.24
	Social Science And Related Profs.	7.44
	Writers And Creative Or Performing Artists	5.88
Science Assc. Professional	Physical And Engineering Science Technicians	43.81
	Computer Associate Profs.	39.76
	Optical And Electronic Equipment Operators	8.90
	Safety And Quality Inspectors	5.69
Other Assc. Professional	Finance And Sales Associate Profs.	55.81
	Administrative Associate Profs.	34.56
	Business Services Agents And Trade Brokers	6.36
Clerks	Secretaries And Keyboard-Operating Clerks	52.82
	Numerical Clerks	27.04
	Other Office Clerks	9.59
Customer Service	Cashiers, Tellers And Related Clerks	56.72
	Client Information Clerks	41.73

(d) Health & Education

Occupation	DISCO 3 Title	% Represented
Health Professional	Health Profs. (Except Nursing)	45.40
	Nursing And Midwifery Profs.	43.24
	Life Science Profs.	10.51
Teachers	Primary And Preprimary Education Teaching Profs.	61.02
	Secondary Education Teaching Profs.	17.32
	Other Teaching Profs.	10.28
	College, University And Higher Education Teaching Profs.	8.09
Health Assc. Professional	Nursing And Midwifery Associate Profs.	57.33
	Modern Health Associate Profs. (Except Nursing)	27.36
	Life Science Technicians And Related Associate Profs.	14.78
Teaching Assc. Professional	Preprimary Education Teaching Associate Profs.	60.47
	Special Education Teaching Associate Profs.	33.03
	Primary Education Teaching Associate Profs.	5.61
Clerks	Secretaries And Keyboard-Operating Clerks	88.00
	Other Office Clerks	5.20
Personal Workers	Personal Care And Related Workers	94.89

This presents the DISCO 3 codes aggregated into each occupation-sector pair. For clarity, only occupations with over 5% representation in the aggregation are shown. Full list available upon request. Data appendix completes description of aggregation and relabeling of occupations.

Table A.11: Industry-Sector Breakdown

(a) Manufacturing

NACE 1.1 Title	% Represented
Construction	19.58
Manufacture of food products and beverages	11.80
Manufacture of machinery and equipment n.e.c.	11.33
Manufacture of fabricated metal products, except machinery and equipment	6.29
Manufacture of furniture; manufacturing n.e.c.	4.30
Manufacture of chemicals and chemical products	4.17
Manufacture of electrical machinery and apparatus n.e.c.	3.42
Manufacture of rubber and plastic products	3.33
Publishing, printing and reproduction of recorded media	3.25
Manufacture of other non-metallic mineral products	3.13
Sewage and refuse disposal, sanitation and similar activities	3.05
Manufacture of medical, precision and optical instruments, watches and clocks	2.47
Manufacture of wood and of products of wood and cork	2.36
Health and social work	1.87
Electricity, gas, steam and hot water supply	1.81
Manufacture of other transport equipment	1.77
Manufacture of radio, television and communication equipment and apparatus	1.72
Manufacture of pulp, paper and paper products	1.45
Manufacture of basic metals	1.35
Wholesale trade and commission trade, except of motor vehicles and motorcycles	1.35
Manufacture of motor vehicles, trailers and semi-trailers	1.15
Manufacture of textiles	1.10
Other business activities	1.05

(b) Services

NACE 1.1 Title	% Represented
Public administration and defence; compulsory social security	20.75
Wholesale trade and commission trade, except of motor vehicles and motorcycles	14.65
Retail trade, except of motor vehicles and motorcycles	9.58
Health and social work	8.05
Post and telecommunications	6.83
Land transport; transport via pipelines	5.13
Education	4.88
Other business activities	4.41
Recreational, cultural and sporting activities	4.35
Sale, maintenance and repair of motor vehicles and motorcycles	4.05
Supporting and auxiliary transport activities; activities of travel agencies	3.09
Hotels and restaurants	2.88
Activities of membership organizations n.e.c.	2.66
Air transport	1.29
Real estate activities	1.16
Other service activities	1.05

(c) FIRE Industries

NACE 1.1 Title	% Represented
Other business activities	29.97
Financial intermediation, except insurance and pension funding	25.89
Computer and related activities	9.59
Health and social work	8.20
Insurance and pension funding, except compulsory social security	6.29
Education	4.68
Real estate activities	3.46
Research and development	2.99
Activities auxiliary to financial intermediation	1.31
Post and telecommunications	1.29
Supporting and auxiliary transport activities; activities of travel agencies	1.14

(d) Health & Education

NACE 1.1 Title	% Represented
Health and social work	65.68
Education	28.62
Retail trade, except of motor vehicles and motorcycles	1.13

Presents all NACE 1.1 industries at the 2 digit levels within defined broad sectors. Percentages calculated by worker headcount representation. Only industries with at least 1% of total industry employees reflected. Full list available upon request from author. Overlap occurs due to reclassifying occupations due to small cell problems. See data appendix for formal description of sector construction.

Table A.12: Description of Tasks

	% Var	Top 5 Chars.	Weight	Bottom 5 Chars	Weight
Task 1	39.06	Writing	11.67	Manual Dexterity	−9.74
		Written Expression	11.61	Extent Flexibility	−9.45
		Speaking	11.55	Handling and Moving Objects	−9.42
		Reading Comprehension	11.55	Multilimb Coordination	−9.40
		Written Comprehension	11.53	Static Strength	−9.29
Task 2	17.09	Operation Monitoring	14.58	Clerical	−2.58
		Quality Control Analysis	14.39	Perform for/Work Directly w/ Public	−2.28
		Inspect Equipment, Structures, Materials	14.26	Fine Arts	−2.14
		Physics	14.15	Service Orientation	−2.00
		Visualization	14.04	Customer and Personal Service	−1.85
Task 3	6.74	Assisting and Caring for Others	20.73	Programming	−14.00
		Therapy and Counseling	19.73	Engineering and Technology	−13.97
		Perform for/Work Directly w/ Public	18.00	Design	−13.19
		Psychology	17.25	Mathematics	−13.07
		Medicine and Dentistry	15.56	Interacting With Computers	−12.91
Task 4	3.50	Sales and Marketing	24.28	Documenting/Recording Information	−16.21
		Economics and Accounting	20.40	Perceptual Speed	−14.87
		Selling or Influencing Others	19.28	Selective Attention	−14.37
		Administration and Management	18.96	Medicine and Dentistry	−13.96
		Building and Construction	18.90	Flexibility of Closure	−13.24
Task 5	2.47	Fine Arts	20.44	Night Vision	−21.54
		Training and Teaching Others	16.10	Peripheral Vision	−20.61
		Thinking Creatively	14.41	Transportation	−19.27
		Education and Training	14.23	Glare Sensitivity	−19.14
		Chemistry	13.97	Spatial Orientation	−19.02
Task 6	2.28	Geography	27.30	Management of Financial Resources	−15.00
		History and Archeology	24.15	Resolve Conflicts/Negotiate w/ Others	−13.45
		Physics	19.04	Monitoring and Controlling Resources	−13.30
		Biology	18.52	Selling or Influencing Others	−13.19
		Telecommunications	16.91	Service Orientation	−13.12
Task 7	2.01	Customer and Personal Service	31.41	Thinking Creatively	−14.54
		Clerical	21.59	Originality	−12.92
		Economics and Accounting	18.79	Learning Strategies	−12.61
		Perform for/Work Directly w/ Public	17.01	Peripheral Vision	−12.32
		Sales and Marketing	16.26	Instructing	−11.52
Task 8	1.91	Fine Arts	20.73	Evaluate Information to Determine Compliance	−19.80
		Visualization	18.25	Processing Information	−17.23
		Sales and Marketing	17.82	Training and Teaching Others	−16.13
		Originality	16.54	Developing and Building Teams	−15.93
		Visual Color Discrimination	16.01	Communicating w/ Supervisors, Peers, etc.	−15.02
Task 9	1.61	Installation	18.39	Food Production	−29.27
		Repairing	16.82	Mathematics	−24.27
		Thinking Creatively	16.16	Number Facility	−23.49
		Equipment Maintenance	16.14	Biology	−23.33
		Repair/Maintain Electronic Equipment	14.99	Chemistry	−22.85
Task 10	1.22	Perform for/Work Directly w/ Public	23.24	Monitoring	−16.61
		Communicate w/ Persons Outside Organization	19.46	Auditory Attention	−15.91
		Selling or Influencing Others	19.00	Production and Processing	−14.68
		Updating and Using Relevant Knowledge	16.11	Selective Attention	−14.12
		Extent Flexibility	13.99	Public Safety and Security	−13.07

Source: ONET Database. Describes survey questions assigned to each task after PCA performed. Variance explained refers to total variance in survey responses accounted for by each component ("task"). Characteristics presented are the top (bottom) most positively (negatively) weighted survey questions for each component. All weights multiplied by 100. The full list is available upon request. Some survey question descriptions shortened for clarity.

A.3 Additional Results

Table A.13: Comparative Advantage Across Occupations

		Type					
Sector	Occupation	1	2	3	4	5	6
Man.	Managers	1	1	1	1	1	1
	Science Prof.			1.43	1.04	1.44	1
	Science Assc. Prof.	1.34	0.96	1.40	0.96	1.16	1
	Other Assc. Prof.	1.72	1.00	1.45	1.00	1.30	1
	Clerks	1.37	0.95	1.41	0.97	0.97	1
	Agriculture	1.11	1.09	0.90	1.13	0.53	1
	Building Trades	1.54	1.00	1.12	1		
	Metal Trades	1.91	1.12	1.48	1.14	0.71	1
	Other Crafts	1.43	0.89	1.18	1		
	Plant Operator	2.03	1.16	1.76	1.25	0.82	1
	Machine Operator	1.61	0.94	1.36	1		
	Drivers	1.59	1.01	1.29	1		
	Laborers	1.45	0.99	1.28	1		
Serv.	Managers	1.01	1.06	1.02	1.04	0.72	1
	Science Prof.	0.83	1.13	0.77	1.12	0.76	1
	Other Prof.	1.08	1.11	0.90	1.08	1.06	1
	Science Assc. Prof.	1.16	1.04	1.31	1.01	0.70	1
	Other Assc. Prof.	1.39	1.13	1.47	1.11	1.09	1
	Clerks	1.26	1.09	1.44	1.08	0.64	1
	Personal Services	1.38	1.17	1.25	1.24	0.75	1
	Retail Occs.	1.18	0.87	1.23	1		
	Metal Trades	1.30	1.00	1.03	1		
	Drivers	1.53	0.99	1.37	1		
	Elementary Occs.	1.60	1.10	1.38	1.24	0.76	1
	Laborers	1.23	0.99	1.21	1		
FIRE	Managers	0.79	1.03	0.58	1.06	0.58	1
	Science Prof.	0.65	1.10	1.54	1.08	1.38	1
	Other Prof.	1.17	1.06	1.61	1.10	1.10	1
	Science Assc. Prof.	1.31	1.05	1.37	1.01	1.20	1
	Other Assc. Prof.	1.35	1.01	1.34	1.03	1.12	1
	Clerks	1.21	0.99	1.31	1.03	0.72	1
	Customer Service	1.24	0.96	1.34	1		
H & E	Health Prof.					1.19	1
	Teachers	0.92	1.02	0.72	1.05	0.77	1
	Health Assc. Prof.	1.10	0.91	1.30	1.09	1.13	1
	Teaching Assc. Prof.	1.04	1.04	0.68	1.09	1.10	1
	Clerks	1.58	1.13	1.61	1.19	0.76	1
	Personal Services	1.53	1.11	1.53	1.10	0.81	1

Notes: Plots the comparative advantage of type k in occupation o relative to type 6 (High absolute advantage + college educated) in occupation 1 (Manufacturing Managers).

Table A.14: Distribution of Lifetime Earnings Differentials (w/ η)

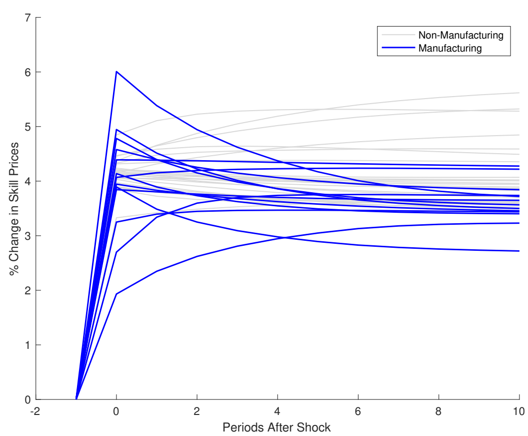
	Mean	Std. Dev	Q5	Q50	Q95	% < 0
Low [L]	1.39	3.92	−0.22	0.82	6.60	5.36
Low [H]	2.10	1.89	0.94	1.87	4.57	3.12
Med [L]	1.97	3.78	−0.52	1.25	7.82	5.89
Med [H]	2.02	1.18	1.43	1.93	3.01	1.88
High [L]	2.73	4.45	−0.10	2.05	8.21	5.28
High [H]	2.26	1.90	1.64	2.04	3.93	2.24
Total	1.92	2.63	0.62	1.85	4.71	3.50

Notes: Tables report the (100x) log difference in discounted total earnings across individuals. Results are based on simulating 100,000 individuals from the initial cohort under both the equilibrium with and without changes in trade prices. The same shocks are used in both simulations. Discounted at $\beta = .96$.

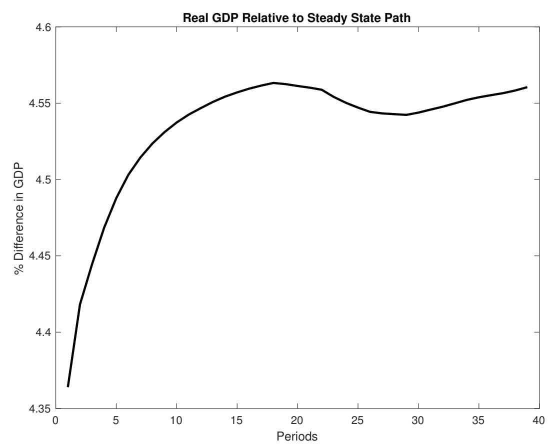
A.4 Figures

Figure 1: Dynamic Effects of an Import Price Shock on Real Prices

(a) Real Skill Price Dynamics



(b) Total Real Income



B Estimation Details

B.1 Model Solution and Derivation of the Regression Equation

Here I summarize in more detail the solution to the dynamic programming problem that leads to an estimating equation. First some notation is in order. Define the *inclusive value* of state ω, o at time t as:

$$D_t(\omega, o) = \sum_{o' \in \mathcal{O}} \exp\left[u_t(o, o', \omega)/\rho + \beta E_t V_{t+1}\left(T(\omega, o, o'), o'\right)/\rho\right]$$

Then, as shown in Rust (1987) , if moving cost shocks are GEV(1) the integrated value function (conditional on aggregate shocks having been realized) can be written as:

$$V_t(o, \omega) = \gamma\rho - \rho \log\left(D_t(o, \omega)\right)$$

Dividing through by ρ, subtracting $\gamma/(1-\beta)$ and plugging in the definition of the inclusive value gives rise to the following Bellman equation:

$$\tilde{V}_t(\omega, o) = \log \sum_{o' \in \mathcal{O}} \exp\left\{\frac{-C(\omega, o, o') + \eta_{o'} + w_{o't} E h_{o'}(o, \omega)}{\rho} + \beta \tilde{E} V_{t+1}\left(T\left(o, o', \omega\right), o'\right)\right\} \tag{B.1}$$

where $\tilde{V}$ is $V/\rho - \gamma/(1-\beta)$. Incidentally, this demonstrates that one *cannot* separately identify a coefficient in front of w and the variance of the shocks (a familiar result from nonlinear binary regression models—the variance of the shock scales coefficients). Equation (B.1) can be solved via Bellman Function Iteration for steady state or given any path of wages and a process on forecasting aggregate shocks. Moreover, transition rates take the following logit form:

$$\log \pi_t(\omega, o, o') = u_t(o, o', \omega) + \beta E_t V_{t+1}\left(T(\omega, o, o'), o'\right) - \log D_t(\omega, o) \tag{B.2}$$

Combining equations (B.1) and (B.2) yields the following:

$$\tilde{V}_t(\omega, o) = \gamma + \frac{u_t(o, o', \omega)}{\rho} + \beta E_t \tilde{V}_{t+1}\left(T(\omega, o, o'), o'\right) - \log \pi_t(\omega, o, o')$$

where I have combined all flow payoff terms into u_t. This formula can be plugged into the formula for transition probabilities iterating for finitely many periods to arrive at the equation from the main text. Iterating forward once more yields,

$$\begin{aligned}
\tilde{V}_t(\omega, o) =& \frac{u_t(o, o', \omega)}{\rho} \\
&+ \beta E_t\left(\frac{u_{t+1}(o', o'', \omega)}{\rho} + \beta E_{t+1} \tilde{V}_{t+2}\left(T(\omega', o', o''), o''\right) - \log \pi_t(\omega', o', o'')\right) \\
&- \log \pi_t(\omega, o, o')
\end{aligned}$$

Rearranging one can write this as,

$$\log \pi_t(\omega, o, o') + \beta E_t \log \pi_{t+1}(\omega, o', o'') = \frac{u_t(o, o', \omega) + \beta E_t u_{t+1}(o', o'')}{\rho} + \beta^2 E_t V_{t+2}(T(\omega', o', o''), o'') - \tilde{V}_t(\omega, o)$$

Finally, consider substituting $E_t \log \pi_{t+1}(\omega, o', o'')$ with the *realization* of the transition rate and an expectational error. As long as workers have rational expectations, this forecast error will be uncorrelated with any time t variables. This leads to the final equation:

$$\log \pi_t(\omega, o, o') + \beta \log \pi_{t+1}(\omega, o', o'') = \frac{u_t(o, o', \omega) + \beta E_t u_{t+1}(o', o'')}{\rho} + \beta^2 E_t V_{t+2}(T(\omega', o', o''), o'') - \tilde{V}_t(\omega, o) + \zeta_{o', o'', t+1} \tag{B.3}$$

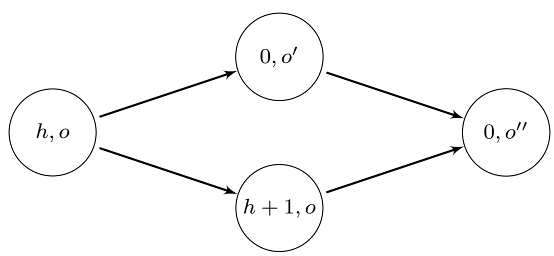
where $\zeta_{o',o'',t} = \beta \log \pi_{t+1}(\omega, o', o'') - \beta E_t \log \pi_{t+1}(\omega, o', o'')$. There are two expectational terms left, we will take care of these after we have set up the differences and differences between workers.

Next notice that in the model switching occupations is a *renewal action*—if two workers in different occupations with identical age, skills and comparative advantage, both switch into the same third occupation, then they will look identical in all dimensions. Formally, consider two workers, 1 and 2 that are identical in age, a, skill and comparative advantage, s, but differ in human capital, *ten*, and their most recent occupation. Then $(\omega_1, o) = (a, s, ten_{o_1}, o_1)$ where the subscript on tenure is because tenure is occupation-specific. Similarly, $(\omega_2, o) = (a, s, ten_{o_2}, o_2)$. If o'' is such that $o'' \neq o_1$ and $o'' \neq o_2$ then we have:

$$T(\omega_1, o_1, o'') = T(\omega_2, o_2, o'')$$

This equality above hinges on the fact that human capital fully depreciates after a switch is made, but can easily be weakened to allow for partial depreciation of human capital *as long as the process on depreciation is known in advance.* Such an extension is explored in Appendix F. Notice that for workers 1 and 2, $\beta^2 E_t V_{t+2}(T(\omega_1, o_1, o''), o'') = \beta^2 E_t V_{t+2}(T(\omega_2, o_2, o''), o'')$.

Further suppose that workers 1 and 2 both were in occupation o_1 at time t, that worker 2 went into occupation o_2 at $t+1$ and both enter o'' at $t+2$. Moreover, that each worker has the same state, ω, at time t. This situation is illustrated in the diagram below:



The final regression equation starts from subtracting equation (B.3) evaluated for worker 2 from (B.3) evaluated for worker 1:

$$\begin{aligned}\log \frac{\pi_t(\omega, o_1, o_2)}{\pi_t(\omega_2, o_1, o_1)} + \beta \log \frac{\pi_{t+1}(\omega, o_2, o'')}{\pi_{t+1}(\omega_1, o_1, o'')} =& \frac{[u_t(o_1, o_2, \omega) + \beta E_t u_{t+1}(o_2, o'', \omega_2)] - [u_t(o_1, o_1, \omega) + \beta E_t u_{t+1}(o_1, o'', \omega_1)]}{\rho} \\ &+ \zeta_{o_2,o'',t+1} - \zeta_{o_1,o'',t+1}\end{aligned}$$

where both $\beta^2 E_t V_{t+2}(\cdot, o'')$ and $V_t(\omega_1, o_1)$ cancel out. This leaves only utility parameters and forecast errors on the right hand side of the equation. Turning to the difference in u_t first:

$$u_t(o_1, o_2, \omega) - u_t(o_1, o_1, \omega) = f(\omega)C(o_1, o_2) + \eta_{o_2} - \eta_{o_1} + \frac{1}{\rho}\left(w_{o_2,t} E_\varsigma(H_{o_2}(\omega, \varsigma)) - w_{o_1,t} E_\varsigma(H_{o_1}(\omega, \varsigma))\right)$$

For the next period:

$$\begin{aligned}E_t\left(u_{t+1}(o_1, o_2, \omega) - u_{t+1}(o_1, o_1, \omega)\right) =& E_t\bigg(f(\omega_2)C(o_2, o'') + \eta_{o''} + w_{t+1,o''} E_\varsigma((T(\omega_2, o_2, o''), o'')) \\ &- f(\omega_1)C(o_1, o'') + \eta_{o''} + w_{t+1,o''} E_\varsigma((T(\omega_1, o_1, o''), o''))\bigg) \\ =& f(\omega_2)C(o_2, o'') - f(\omega_1)C(o_1, o'')\end{aligned}$$

Since switching is a renewal action, and the workers are otherwise identical, the only difference in the flow utility terms is the difference in switching costs for the workers. This is a constant, and so the expectation term can be dropped. Combining these into the first equation one has the final estimating equation:

$$\log \frac{\pi_t(\omega, o_1, o_2)}{\pi_t(\omega_2, o_1, o_1)} + \beta \log \frac{\pi_{t+1}(\omega, o_2, o'')}{\pi_{t+1}(\omega_1, o_1, o'')} = \frac{f(\omega)C(o_1, o_2) + \beta[f(\omega_2)C(o_2, o'') - f(\omega_1)C(o_1, o'')]}{\rho} + \frac{\eta_{o_2} - \eta_{o_1}}{\rho} + \frac{1}{\rho}\left(w_{o_2,t}E_\varsigma(H_{o_2}(\omega, \varsigma)) - w_{o_1,t}E_\varsigma(H_{o_1}(\omega, \varsigma))\right) + \zeta_{o_2,o'',t+1} - \zeta_{o_1,o'',t+1} + m_{o_2,o_1,o'',t,t+1} \tag{B.4}$$

where the new term, $m_{o_2,o_1,o'',t,t+1}$, reflects measurement error in the $\log \pi$ terms. From the first stage, described in the next subsection, one can recover the parameters of H, including the parameters governing ζ. Hence, in the second stage the income differences are known. The exception is the value of non-employment, which I estimate as a quadratic function of age, and a linear function of skill. Thus, the parameters estimated in the above regression are the parameters of C, η, ρ and $w_{Unemp.}$. Notice that only *differences* in η are identified and there is no constant. Thus, I normalize $\eta_{Unemp.} = 0$ and assume that entering non-employment is costless (the cost is absorbed by the value of non-employment).

Since time t skill prices are orthogonal to the $t+1$ forecast error, this can be estimated with non-linear least squares. If instead w_t were forecasted *before* decisions were made, there would be additional forecast error that *would* be correlated with contemporaneous variables. This would call for an instrumental variables strategy. In Appendix H I show that there is not much sensitivity to estimates if one instruments skill prices with lagged skill prices. In principle, this equation could be constructed for every possible ω—however there are over 75 million possible combinations of state variables. Below I describe how I choose which subset of the state space to use.

The error term in the regression is the difference in forecast errors. These forecast errors will likely be correlated across different sets of occupations. For example, any pair of initial occupations that end in o'' will all have correlated expectational errors. In order to deal with this I cluster standard errors by period. In appendix G I modify the estimator slightly to construct pseudo-histories for each individual and use an individual-level version of (B.4) to construct data that is not clustered across individuals within a time period. Using this alternative estimator, errors become larger, but most point estimates do not appreciably change.

B.2 Likelihood Setup for First Stage

Before detailing the procedure I review the likelihood function without unobserved heterogeneity, arising from the timing and distributional assumptions discussed in the main text. In this section I index an occupation-sector pair by o and employment status by e where $e = 1$ means employed. An observation for individual i at time t is a wage, w_{it}, employment situation, $(o, e)_{it}$, and controls X_{it} (polynomials in age, tenure and skill level). Wages in the unemployed state do no matter as I assume there is no unobserved randomness in that state. This implies that the likelihood for individual i is given by,

$$\mathcal{L}_i = \mathcal{F}(w_{iT}, ..., w_{i0}, (o, e)_{iT}, ..., (o, e)_{i0}, X_{iT}, ..., X_{i0})$$

From the paper, the following assumptions give rise to a first order Markov structure on the worker's career path: (1) that cost shocks are iid and (2) that tenure resets upon switching occupations. Moreover, the assumption that wage shocks are unobserved at the time that an occupation is chosen implies that the likelihood is separable in the wage and occupational decision. Thus the likelihood can be factored as follows:

$$\mathcal{L}_i = f(w_{iT}|(o,e)_{iT}, X_{iT})\pi((o,e)_{i,T}|(o,e)_{i,T-1}, X_{iT}) \times \cdots \times f(w_{i1}|(o,e)_{i1}, X_{i1})\pi((o,e)_{i,1}|(o,e)_{i,0}, X_{i1})\pi((o,e)_{i0}|X_{i0})\pi(X_{i0})$$

where $\pi((o,e)_{i0}|X_{i0})\pi(X_{i0})$ is the probability of observing the initial state. From the log normal assumption on wages one has that $f(w_{it}|o_{it}, X_{it}) = \phi\left(\frac{w_{it} - \beta_{ot}X_{it}}{\sigma_o}\right)$ where β are the occupation specific Mincer coefficients (the time subscript reflecting year fixed effects) and σ is the occupation specific standard deviation of the wage shocks. π is determined from the model and is generally infeasible to calculate. Nevertheless, the wage parameters can be estimated separately in a first stage. This fact help guides how one adds unobserved heterogeneity into the model.

Suppose that workers' unobservable type is an $|\mathcal{O}|+1$-dimensional vector where each of the first $|\mathcal{O}|$ elements are a workers' productivity shifter in each occupation (comparative advantage) while the final element is a workers' inverse moving productivity shifter. Following Heckman and Singer (1984), I assume that θ is drawn from a finite distribution, $\mathbf{Q}_\theta$. Moreover, I index a worker's type by k and define $K = |\mathbf{Q}_\theta|$. The likelihood becomes,

$$\mathcal{L}_i = \sum_{k=1}^{K} f(w_{iT}|(o,e)_{iT}, X_{iT}, k)\pi((o,e)_{i,T}|(o,e)_{i,T-1}, X_{iT}, k) \\ \times \cdots \times f(w_{i1}|(o,e)_{i1}, X_{i1}, k)\pi((o,e)_{i,1}|(o,e)_{i,0}, X_{i1}, k)\pi((o,e)_{i0}|X_{i0}, k)\pi(X_{i0}|k)\pi(k)$$

This is not log additively separable, but the EM algorithm can be employed in order to estimate the model parameters. Before specifying things further I introduce some notation. Define: $\mathcal{H}_{is}$ to be the history of data for worker i up to time s; $q_{ik} = P(k|\mathcal{H}_{iT})$, the probability that worker i is of type k given their full history; $\Xi = (\beta_{mincer}, \beta_\pi)$ to be the set of all parameters to be estimated. In this case, the objective function becomes:

$$J = \max \frac{1}{NT} \sum_{i=1}^{N} \sum_{k=1}^{K} q_{ik} \times \\ \left(\log \pi((o, e_{i0}|X_{i0}, k; \Xi) + \log \pi(X_{i0}|k; \Xi) + \sum_{t=1}^{T} [e_{it} \log f\left(w_{it}|\omega_{it}, o_{it}, k; \Xi\right) + \log \pi\left(\omega_{it}, (o,e)_{it}|\mathcal{H}_{it-1}, k; \Xi\right)] \right)$$

where the first two terms in the summation account for the initial state of the worker and the sum term reflects the log likelihood of a particular history given a type.

In the standard EM algorithm one could now maximize this likelihood by iterating over a guess of q_{ik} and the parameters of the model (in a manner I will detail below). However, this is infeasible as π is generated by the solution to a complicated and non-stationary dynamic problem. Moreover, the initial conditions require integrating over many career paths for each combination of observed initial values. The next subsection discusses how I use Arcidiacono and Miller (2011) in order to construct an approximation to the true likelihood that can be used to back out the parameters of the wage equation as well as the distribution of types.

B.3 Approximate Likelihood for First Stage

In order to maximize the above objective without solving the model, I approximate the likelihood by using the observed transition matrix across states in order to estimate π. That is to say, I maximize the objective:

$$\hat{J} = \max \frac{1}{NT} \sum_{i=1}^{N} \sum_{k=1}^{K} q_{ik} \times \\ \left(\log \hat{\pi}((o, e_{i0}|X_{i0}, k; \Xi) + \log \hat{\pi}(X_{i0}|k; \Xi) + \sum_{t=1}^{T} [e_{it} \log f\left(w_{it}|\omega_{it}, o_{it}, k; \Xi\right) + \log \hat{\pi}\left(\omega_{it}, (o,e)_{it}|\mathcal{H}_{it-1}, k; \Xi\right)] \right)$$

where now $\hat{\pi}$ are estimated from the data.

Given q_{ik}, solving for the parameters of the income equation is straightforward: the normal structure implies a weighted least squares regression. In particular, one needs to stack the vector of income and covariates K times and include $K-1$ dummies for each type. Then the q_{ik} play the role of regression weights.

For a particular guess of q_{ik}, solving for $\hat{\pi}$ amounts to a weighted average of transitions. Suppose that there were only a small number of finite state transitions. Then one could solve for the probability of moving from state s_1 to s_2 at time t as,

$$\hat{\pi}(s_2|s_1) = \frac{\sum_{i=1}^{N} q_{ik}\mathbf{1}(s_{it} = s_2)\mathbf{1}(s_{it-1} = s_1)}{\sum_{i=1}^{N} q_{ik}\mathbf{1}(s_{it-1} = s_1)}$$

which is just the weighted average of transitions.[1] In practice, the bin estimator above only works for states which are relatively parsimonious and a transition matrix across states that is very dense. In my setting this is not the case. While I observe many occupation to occupation transitions, I also include skill, age, and tenure as state variables which greatly enlarge the state space. In this case one must smooth out the transition probabilities.[2] In order to do this I specify a linear probability model for each skill-type-year-occupation-pair that has age, tenure and previous employment status (e) as covariates. This setup allows for complete flexibility of the transition rates along the skill, year, type and occupation pair dimension and imposes a polynomial structure on the continuous variables.[3] This requires over 1000 regressions to be performed, which is why I use an LPM framework instead of a logit or other non-linear setup. For a discussion of some issues and limitations see the final subsection on computational and implementation issues.

In order to perform the E step of the EM algorithm, one needs to be able to calculate the likelihood of a particular unobserved state given a full history. From Bayes' rule one will have,

$$P(k|\mathcal{H}_{iT}) = \frac{\left[\prod_{t=1}^{T} f(w_{it})\pi((o,e)_{it}|(o,e)_{t-1}, X_{it-1})\right] \times \pi(k|X_{i0})\pi(X_{i0})}{\sum_{k'}\left[\prod_{t=1}^{T} f(w_{it})\pi((o,e)_{it}|(o,e)_{t-1}, X_{it-1})\right] \times \pi(k'|X_{i0})\pi(X_{i0})}$$

While the term $\pi(X_{i0})$ drops out, one needs to deal with the workers' initial state. I estimate $\pi(k|X_{i0})$ by regressing q_{ik} on the initial state. I do this separately for each initial employment-status-occupation-skill-year and include a cubic in age, a cubic in tenure and an interaction term in the regression.[4]

B.4 Modified EM Algorithm for First Stage

The above subsection discussed how, given a set of weights q_{ik}, one could solve for the maximized likelihood. In this subsection I write out the full algorithm and discuss its initialization. The algorithm is given as:

1. Begin with a guess of $\pi^{(1)}\left(o_{it}|o_{it-1}, X_{it-1}\right)$, $\Xi^{(1)}$, and $\pi^{(1)}(k|X_{i0}, o_{i0})$
2. Generate an update of weights:

$$q_{ik}^{(j)} = \frac{\left[\prod_{t=1}^{T} f\left(w_{it}|\omega_{it}, o_{it}, k; \Xi^{(j-1)}\right)\pi\left(\omega_{it}, o_{it}|\mathcal{H}_{it-1}, k; \Xi^{(j-1)}\right)\right]\pi^{j-1}(k|X_{i0}, o_{i0})}{\sum_{k'}\left[\prod_{t=1}^{T} f\left(w_{it}|\omega_{it}, o_{it}, k'; \Xi^{(j-1)}\right)\pi\left(\omega_{it}, o_{it}|\mathcal{H}_{it-1}, k'; \Xi^{(j-1)}\right)\right]\pi^{j-1}(k'|X_{i0}, o_{i0})}$$

[1] This discussion derives from Arcidiacono and Miller (2011) and the above formula is equation (5.9) of their paper.

[2] Examples of other papers that confronted similar issues are Ransom (2018) (who used a logit regression) and Scott (2014) (who used a Laplace prior to deal with small bins).

[3] For the actual LPM, I allow for a cubic in age, a cubic in tenure, a cross-product in age and tenure, a dummy for being previously non-employed and an interaction between the dummy and a full cubic in age. I found the model began to become overfitted if I allowed for more interaction terms (partially because tenure is 0 for many workers in non-employment). Moreover I bound the probabilities between $1 - 10^{-6}$ and 10^{-6} to keep the likelihood well behaved numerically.

[4] This is for cases where the number of observations in the cell is greater than or equal to 25, for the cases where an initial cell is less than 25 people I simply use the mean of q within the bin and include no regressors. These cases account for .15% of workers.

3. Update $\pi^{(j)}(k|X_{i0}, o_{i0}) = \frac{\pi^{(j)}(k, o_{i0}|X_{i0})}{\pi(o_{i0}|X_{i0})}$ by regressing $q_{ik}^{(j)}$ on a polynomial of X_{i0} for each occupation-k pair.
4. Update $\Xi^{(j)}$ by maximizing $\hat{J}$
5. Iterate on steps 2-4 until $\hat{J}$ stops growing to some tolerance.

To operationalize the procedure one needs an initial guess of model parameters and type distribution. I follow the suggestion of AM and set $\Xi^{(1)}$ to a perturbed set of parameters around the results for a single type. As a point on defining convergence, I divide by N instead of NT since the panel is not balanced and I use a tolerance of 10^{-6}. Because of occasionally very small or even negative values of LPM, I found that the likelihood was numerically unstable at lower tolerances. Figure 2 plots the convergence of the likelihood.

B.5 Computational and Implementation Issues for First Stage

While the above outlines the algorithm I employ as well as several details of implementation there are two final issues that need to be highlighted: missing data and small bins. These points are minor in the sense that they relate very precisely to my particular implementation of the above algorithm. On the other hand, they do mean that I am not maximizing the likelihood exactly as written above so I describe them for the sake of completeness.

First, for a great many individuals, the occupation is unknown.[5] While in principle one could integrate out missing data by focusing on observed trajectories, I cannot do this. This is because with the modified likelihood, I would need to estimate approximate transitions in and out of various possible combinations of missing data which would put overly burdensome demands on the data in terms of bin size. And so, I ignore missing observations, creating holes in workers' histories. When I look at the data, it does not appear that missing observations are correlated strongly with any demographic characteristic. However, as discussed in my data section in regards to imputation, missing data is likely to occur when workers switch occupations. Unfortunately, this potential source of bias runs throughout my estimation.

Second, while the size of the data allows me to handle a large state space, there are still several very small transitions. In particular, some skill-year-occupation-pair bins are so small that one cannot estimate the transition probabilities without overfitting (in some cases, there are more variables to fit than observations). For these small cases, I drop the transitions from the estimation, treating them essentially the same as missing data. I use a cutoff of twenty five observations in a bin before estimating an LPM. I have experimented with cutoffs in the range of twenty to thirty five and found little sensitivity (flexibility of the LPM appears to be much more important). This cutoff leaves about 99% of all observations in the likelihood and about 91% of switches.

B.6 Construction of Variables for Second Stage

Before turning to the actual regressions in the second stage I discuss construction of the left and right hand side variables. As a reminder, the left hand side in the estimation is given by:

$$LHS = \log \frac{\pi(o'|o, X)}{\pi(o|o, X)} + \beta \log \frac{\pi(o''|o', X')}{\pi(o|o, \tilde{X}')}$$

which says that the left hand side is the discounted log difference in observing career trajectory $o \to o' \to o''$ and $o \to o \to o''$, given the same initial set of covariates, X. The reason for the notation X' and $\tilde{X}'$ reflects that age and tenure change for workers and these changes depend on the career path. Notice that this assumes that a workers' unobserved type is known. This is why the EM algorithm is performed in the first stage: one can extract π given the type as a solution to a weighted least squares regression where the weights are exactly the probability of each individual being of a certain type. Thus, given a choice of β, one can construct the left hand side with parameters already estimated in the first stage.

[5]In the data appendix I discuss more precisely how and when imputation occurs and what observations are dropped.

Due to the high dimensional state space and panel length there are several hundred million possible combinations of the above formula. In order to deal with this problem computationally I pre-specify a grid over the deterministic part of the sate space. In particular, I focus on age beginning at 28 increasing in increments of 5, tenure increasing in increments of 2, and copy this grid for all skill types. This tremendously lowers the state space by several orders of magnitude. Then for this grid I calculate all possible transition paths given the observed pair-wise transitions observed in the data in a given year (i.e., I use every possible $\hat{\pi}$ from the first stage). This leaves on the order of 4.5 million observations, a substantially more tractable number. Due to the choice of grid and the fit of the LPM, some probabilities are estimated at boundaries. I do not remove these observations for two reasons: (1) this would be inconsistent with the first stage estimation where in fact the LPM probabilities were bounded; and (2) as discussed in Artuç and McLaren (2015), ignoring low probability events can bias results. This is because if a switch is observed in one period and not another due to a change in wages, this is useful information. Ignoring such data points will artificially lower measured worker responses to wage differentials.[6]

The right hand side variables are the costs and income associated with different occupational choices. The construction of occupational characteristics occurs in the main text, so I do not discuss it here. For the income, much like the left hand side, the type of the agent is assumed to known. Once again, the wages of certain types and the distribution of types can be recovered from the first stage. Thus, the second stage of the estimation uses the predicted income for workers in a particular state coming from the first stage wage equation.

B.7 Regression Details for Second Stage

As a reminder of the main estimating equation I run regressions of the form:

$$\begin{aligned}\log\frac{\pi(o'|o,X)}{\pi(o|o,X)}+\beta\log\frac{\pi(o''|o',X')}{\pi(o|o,\tilde{X}')} =& -\tilde{C}(o,o',o'',X,X',\tilde{X};\Gamma)\\ &+[\eta_o'-\eta_o]+\frac{1}{\rho}\left[w_{o't}g_o'(X)-w_{ot}g_o(X)\right]+\varepsilon_{oo'tX}\end{aligned}$$

where the first term reflects moving costs, the second bracketed term is the difference in compensating differentials across occupations, the third bracketed term is the difference in incomes across occupations conditional on year and state, and the final term is the error term, which is the $t+1$ forecast error on the difference in continuation values for occupation o' and o. Breaking out the cost term:

$$\tilde{C}(o,o',o'',X,X',\tilde{X};\Gamma)=f(X;\Gamma_f)C(o,o';\Gamma_C)+\beta[f(X';\Gamma_f)C(o',o'';\Gamma_C)-f(\tilde{X}';\Gamma_f)C(o,o'';\Gamma_X)]$$

where f and C are non-linear functions of observables described in the main text. There are several special cases of the above that are important for understanding the regressions. First of all, I normalize the value of non-employment for 25 year old, low skilled workers to 0 (it's clear that one occupational choice must be normalized as I can only identify differences in the values of different states). Second, I normalize the cost of entering non-employment to be zero.

Since there are too many points of support along the state space to construct all combinations of the above regression I use a grid approach. In particular,

1. I use all observed pairs of occupational moves to construct counterfactual one-shot deviation transitions. For example, if I observe transitions from occupation A to B, from A to C and from B to C then I construct the counterfactual probabilities of going from A to B to C and then from A to A to C.

2. Step 1 only defines the possible paths, in step 2 I define a grid across individual states by specifying: (1) a subset of ages starting at 28 and moving up by 5s until 53; (2) repeating this for all possible

[6] In a robustness check I estimate the model using only non-bounded estimates. The estimated coefficient on wages decreases from approximately 1.2 to .4, as expected. Despite the seemingly large magnitude, these numbers actually change the simulation results little. This is because switching is low at the baseline so jumps in $1/\rho$ have moderate effects on the overall change in switching elasticity.

skill groups and types ; and (3) a subset of possible initial tenures, $1, 3, 5$.

This leaves me with on the order 4.5 million points in the regression, which is a manageable number from a memory perspective.

B.8 Standard Errors

I estimate the standard errors for the model's parameters via block bootstrap. However, I do this in two ways. In the first, I draw, with replacement, entire individual career paths (so that I am clustering by individual). I use 100 bootstrap samples to construct standard errors. Specifically, I draw uniformly from the pool of individuals. This risks over sampling from years with larger populations (i.e., later in the sample), but this should not necessarily bias the standard errors. This procedure yields strictly more conservative standard errors than those implied by treating first stage parameters as estimated and treating the second stage estimate as a regression, with the number of periods going to infinity. Moreover, this properly accounts for the fact that standard errors are based on the population of individuals and time periods going to ∞ and *not* the number of occupations, which is treated as constant. Note that this second point also implies that there is no incidental parameters problem in estimating the occupation fixed effects. In practice, first stage estimates are very precise. In light of the sample size, that little measurement error remains may not be surprising. However, this cross-sectional variation in N ignores that second stage forecast errors will be correlated across individuals. This motivates the T bootstrapping discussed next.

The block bootstrapping above relies on the idea that T is large. While $T = 12$ is a good sized panel for *identification* purposes[7] and such a panel length is not atypical in structural work (e.g., Das et al. (2007)), it may underestimate standard errors to ignore the fixed time dimension. To this end, I also block bootstrap in the time dimension in order to assess this issue.[8] As resampling both individuals and years is difficult, I focus on just resampling years, and treat the first stage parameters as estimated. Ignoring first stage error does not seem problematic as first stage parameters are very tightly estimated—and this produces very tight standard errors when coupled with the procedure in the previous paragraph. Specifically, I draw, with replacement full sets of switches uniformly from the set of integers $\{1, ..., 10\}$ and then take $t, t+1, t+2$ from the data. I experimented with block length, and a length of 1 produced the most conservative standard errors. In Appendix G I discuss a third strategy based on Altuğ and Miller (1998). This relies on, essentially, weighting the estimating regression by the empirical distribution of switchers. This weighting, paired with a particular resampling strategy, yields $\sqrt{NT}$ consistent estimates. This produces tighter standard errors than bootstrapping on t, and changes the parameter estimates slightly. /

B.9 Jackknife Bias Correction

One parameter of interest, the semi-elasticity of transitions with respect to wages, $1/\rho$, is identified predominantly from time variation, rather than cross-sectional variation[9]. In simulations, quantitative and qualitative predictions do not change substantially with differences with changes in ρ, unless these changes are large. Nevertheless, it plays an important role in governing the speed of adjustment to shocks in the model and so its identification merits careful discussion. Given the short length of the panel, one must worry not only about suitably adjusting standard errors but also the risk of short panel bias (see Arellano (2003) p. 85). Indeed, in simple Monte Carlo exercises, there is the potential for negative bias in the estimation of this parameter with short panels. While this bias tends to evaporate quickly in simulations, it would be helpful to have a sense of the potential for bias in the actual data. I do this relying on the split sample jackknife technique introduced and discussed in Dhaene and Jochmans (2015). The key is that in the second stage estimation, ρ, is essentially estimated as if the data is linear panel data. To that end, the finite T bias has a

[7] See, e.g., Honoré and Tamer (2006) for a discussion in which $T > 3$ is enough for identification in a closely related class of models.

[8] I am particularly indebted to Bo Honore and Elena Manresa for discussions on this issue.

[9] As discussed in the main text, differences in transition rates with respect to tenure also identify this parameter

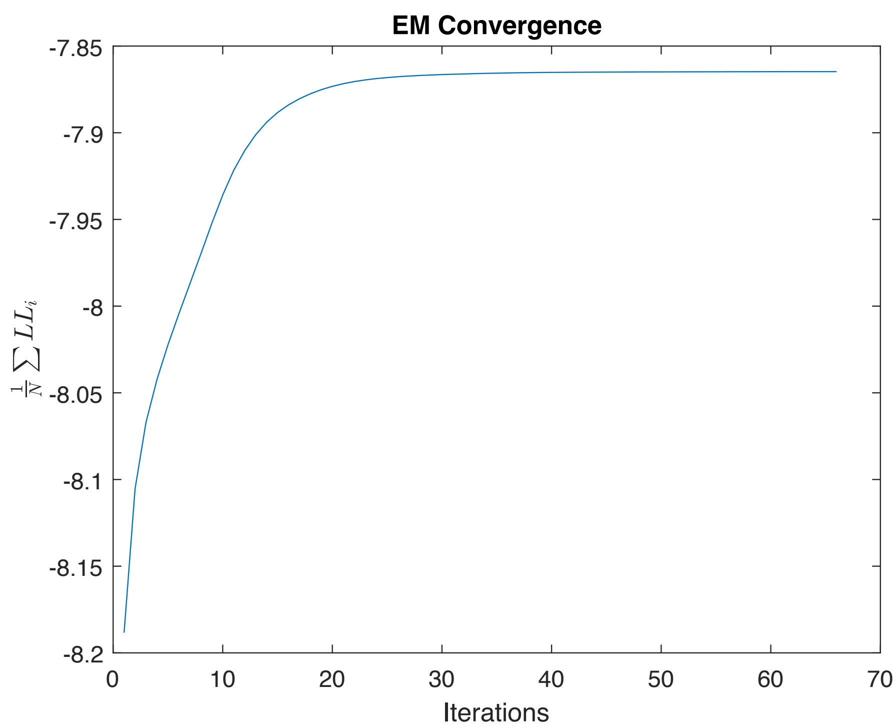
well known form, $c_0/T + O(T^{-2})$ for some time-invariant constant, c_0. Given this formula, a consistent and bias corrected estimator of ρ can be given by:

$$\hat{\rho}^{JK} = 2\hat{\rho}_0 - \frac{\hat{\rho}_{P1} + \hat{\rho}_{P2}}{2}$$

where $\hat{\rho}_0$ is the original estimate, $\hat{\rho}_{P1}$ is the estimate from the first half of the panel and $\hat{\rho}_{P2}$ is the estimate on the second half of the panel[10]. The value for this corrected $1/\rho$ is 1.9, while the original estimate is 1.5. The corrected number is larger, as should be the case—however the uncertainty is large. Moreover, despite being quite a bit larger in proportional terms, the difference in estimates is small in terms of importance for counterfactuals. Thus, the panel length seems to be sufficient for identifying the parameter of interest. Appendix G contains a third estimate for $1/\rho$ based on a different weighting of the regression equations that is $\sqrt{N}$ consistent. This estimator has $1/\rho = 2.2$, which is quite close to the corrected estimate in this appendix. Appendix H also contains a linear approximation to the non-linear model and uses both an OLS and IV approach to estimate the parameters. In these cases, $1/\rho$ ranges from 2.01 to 2.06. These different specifications, done with different assumptions in mind, all point to a fairly robust estimate of $1/\rho$ that is between 1.5 and 2.5. Numbers in this range are likely to lead to similar counterfactuals, which can also help explain why the data may not be able to distinguish between values of $1/\rho$ in that range.s

B.10 Appendix B Figures

Figure 2: EM Algorithm Convergence



[10]I construct pseudo-initial conditions based on $T/2 - 1$ for the second half panel.

C Details on Counterfactuals

In this section I detail more precisely how I perform counterfactuals. All counterfactual analysis must begin from a specified initial equilibrium. I either use the observed distribution of workers and prices in Denmark in 1996 or I use the steady state distribution of workers and prices that arises from simulating the model forward assuming perfect foresight starting from the 1996 equilibrium. In the paper I specify which initial condition for various counterfactuals.

C.1 Updating Labor Supply

1. Given initial conditions $\{L_{ot-1}\}$ and a guess of wages $\{w_{ot}\}$ solve the worker's problem to determine $\left\{H^S_{ot}\right\}$.

The details of the labor supply problem are in the main text, but this procedure reduces to iterating on a Bellman equation.

C.2 Updating Labor Demand

As solving the firms' problem is less central to the paper I discuss many details in this appendix. First I review the basic set up, then explain how I update the system.

C.2.1 Technology and Prices

There is a representative firm within each industry using a Cobb-Douglas production function in labor, capital and intermediates:

$$Y_i^D = \tilde{z}_i K_i^{\beta_{ik}} \left(\prod_o L_{io}^{\gamma_{io}}\right)^{\beta_{iL}} \left(\prod_j M_{ij}^{\nu_{ij}}\right)^{\beta_{iM}}$$

All subsets of coefficients sum to 1. For tradable goods, there is an aggregate good given by:

$$Y_i = \left(\left(\tilde{A} Y_i^D\right)^{\rho_i} + \left(Y_i^F\right)^{\rho_i}\right)^{\frac{1}{\rho_i}}$$

and $\sigma_i = \frac{\rho_i}{\rho_i - 1}$ is the elasticity of substitution.

Assuming perfect competition, the technology above gives rise to the following expenditure system:

$$E_i = \left(\alpha_i \left(W + p_k K\right) + \sum_j \beta_{jM} \nu_{ji} E_j\right) \frac{A\left(P_i^D\right)^{1-\sigma}}{A\left(P_i^D\right)^{1-\sigma} + \left(P_i^F\right)^{1-\sigma}} + D_i^F \left(P_i^D\right)^{1-\sigma}$$

and the following price system:

$$\log P_i^D = -\log z_i + \beta_{ik} \log\left(p_k\right) + \beta_{iL} \sum_o \gamma_{io} \log\left(w_o\right) + \beta_{iM} \sum_j \nu_{ij} \log\left(A\left(P_i^D\right)^{1-\sigma} + \left(P_i^F\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$$

From a change in *nominal* factor prices or technology, one can solve for expenditures and output prices as well as the aggregate price index.

C.2.2 The System in Changes

The above problem can be recast in terms of *relative prices of domestic and foreign* goods. These are observed in the data from data on expenditure and elasticities of substitution, while data on relative prices between sectors is not.

To begin, rewrite the expenditure system in terms of relative prices between foreign and domestic:

$$E_i = \left(\alpha_i \left(W + p_k K\right) + \sum_j \beta_{jM} \nu_{ji} E_j\right) \frac{r_i^{1-\sigma}}{r_i^{1-\sigma} + 1} + X_i r_i^{1-\sigma}$$

where

$$X_i = \frac{D_i^F \left(P_i^F\right)^{1-\sigma}}{A}$$

and

$$r_i = \frac{A P_i^D}{P_i^F}$$

Notice that for nontradable goods $r_i = 0$ so that the expenditure system *does not* depend on the *levels* of prices or productivity in the non-tradable sector.

To keep this intuition working in the small open economy case, consider rewriting the price system as follows:

$$\begin{aligned}
\log P_i^D &= -\log z_i + \beta_{ik} \log(p_k) + \beta_{iL} \sum_o \gamma_{io} \log(w_o) + \beta_{iM} \sum_j \nu_{ij} \log \left(A \left(P_j^D\right)^{1-\sigma_j} + \left(P_j^F\right)^{1-\sigma_j}\right)^{\frac{1}{1-\sigma_j}} \\
&= -\log z_i + \beta_{ik} \log(p_k) + \beta_{iL} \sum_o \gamma_{io} \log(w_o) + \beta_{iM} \sum_{j \in \mathcal{T}} \nu_{ij} \left[\log \left(r_i^{1-\sigma_j} + 1\right)^{\frac{1}{1-\sigma_j}} + \log P_i^F\right] + \beta_{iM} \sum_{j \in \mathcal{NT}} \nu_{ij} \log P_i \\
&= -\log z_i + \beta_{iM} \sum_{j \in \mathcal{T}} \nu_{ij} \log P_i^F + \beta_{ik} \log(p_k) + \beta_{iL} \sum_o \gamma_{io} \log(w_o) + \beta_{iM} \sum_{j \in \mathcal{T}} \frac{\nu_{ij}}{1-\sigma_j} \left[\log \left(1 + r_i^{1-\sigma_j}\right)\right] + \beta_{iM} \sum_{j \in \mathcal{NT}} \nu_{ij} \log P_i
\end{aligned}$$

Now for *tradable* goods subtract the foreign price,

$$\log r_i = -\log z_i + \sum_j \left(\beta_{iM} \nu_{ij} - \delta_{ij}\right) \log P_i^F + \beta_{ik} \log(p_k) + \beta_{iL} \sum_o \gamma_{io} \log(w_o) + \beta_{iM} \sum_{j \in \mathcal{T}} \frac{\nu_{ij}}{1-\sigma_j} \left[\log \left(1 + r_i^{1-\sigma_j}\right)\right] + \beta_{iM} \sum_{j \in \mathcal{NT}} \nu_{ij} \log P_i$$

where δ is the Kronecker delta function.

Ordering the industries by tradable and non-tradable, and dropping the logarithms as understood, and first difference leads to the following matrix system:

$$\Delta \begin{bmatrix} \log r_i \\ \cdots \\ \log P_i \end{bmatrix} = \underbrace{-\Delta \log z + \begin{bmatrix} B^M_{\mathcal{T},\mathcal{T}} - I_{|\mathcal{T}|} \\ \cdots \\ B^M_{\mathcal{NT},\mathcal{T}} \end{bmatrix} \Delta \log P_i^F + B^K \log \Delta p_k}_{\text{Exogeneous}} + \underbrace{B^L \Delta \log w + \begin{bmatrix} B^M_{\mathcal{T},\mathcal{T}} \\ \cdots \\ B^M_{\mathcal{NT},\mathcal{T}} \end{bmatrix} \Delta \log \tilde{r}_i + \begin{bmatrix} B^M_{\mathcal{T},\mathcal{NT}} \\ \cdots \\ B^M_{\mathcal{NT},\mathcal{NT}} \end{bmatrix} \Delta \log P_i}_{\text{Endogeneous}}$$

This system is at the heart of the simulation algorithm because with knowledge of r_{it-1} and a guess of Δw one can solve for r_{it} without needing to solve for the level of any other prices.

This is a non-linear system of equations. However, linearizing $\tilde{r}_{is}$ around r_{it} yields the approximate linear solution:

$$\Delta \begin{bmatrix} \log r_i \\ \cdots \\ \log P_i \end{bmatrix} = \left(I_{N \times N} - \begin{bmatrix} B^M_{\mathcal{T},\mathcal{T}} \circ \frac{r_t^{1-\sigma}}{1+r_t^{1-\sigma}} & B^M_{\mathcal{T},\mathcal{NT}} \\ \cdots & \cdots \\ B^M_{\mathcal{NT},\mathcal{T}} \circ \frac{r_t^{1-\sigma}}{1+r_t^{1-\sigma}} & B^M_{\mathcal{NT},\mathcal{NT}} \end{bmatrix}\right)^{-1} \left(\underbrace{-\Delta \log z + \begin{bmatrix} B^M_{\mathcal{T},\mathcal{T}} - I_{|\mathcal{T}|} \\ \cdots \\ B^M_{\mathcal{NT},\mathcal{T}} \end{bmatrix} \Delta \log P_i^F + B^K \Delta \log p_k}_{\text{Exogeneous}} + B^L \underbrace{\Delta \log w}_{\text{Current Guess}}\right)$$

where $\circ$ refers to multiplying the $(i,j)^{th}$ element of B by the j^{th} element of $\frac{r^{-\sigma}}{1+r^{1-\sigma}}$. From here one can move back to the expenditure system *in levels* using the following updating rule for export expenditure:

$$\Delta \log X_{it+1} = \Delta \log D_{it+1}^F + (1-\sigma) \Delta \log P_{it+1}^F - \Delta \log A_{t+1}$$

Summarizing this algorithm:

1. Given $\left\{r_{t-1}, w_{t-1}, \Delta P^F\right\}$ and a guess of w_t, solve the system of equations for $\{\Delta r_t\}$
2. Construct $r_{it} = r_{it-1} + \Delta r_i$
3. Adjust export demand by $X_{it} = X_{i,t-1}\left(\frac{P^F_{it}}{P^F_{i,t-1}}\right)^{1-\sigma}$
4. Solve the expenditure system for $\{E_i\}$
5. Construct labor and capital demand from the expenditure system:

$$H^D_{ot} = \frac{\sum_i \beta_{iL}\gamma_{io}E_{it}}{w_{ot}}$$

with an analogous formula for capital.

C.2.3 Choice of Numeraire

As this is a general equilibrium model, I need to make one normalization in order to pin down a numeraire. I treat the foreign demand shifter as fixed. To see that this is the case consider a one-sector, two-country version of the trading environment. In this case, expenditure at home is given by,

$$\begin{aligned} E^D &= (W^D + \beta_M E^D)\frac{A^D(P^D)^{1-\sigma}}{A^D(P^D)^{1-\sigma} + (\tau_{DF}P^F)^{1-\sigma)}} + (W^F + \beta_M E^F)\frac{(\tau_{FD}P^D)^{1-\sigma}}{(\tau_{FD}P^D)^{1-\sigma} + (P^F)^{1-\sigma)}} \\ &= (W^D + \beta_M E^D)\frac{A^D(P^D)^{1-\sigma}}{A^D(P^D)^{1-\sigma} + (\tau_{DF}P^F)^{1-\sigma)}} + \underbrace{\frac{(W^F + \beta_M E^F)(\tau_{FD})^{1-\sigma}}{(\tau_{FD}P^D)^{1-\sigma} + (P^F)^{1-\sigma)}}}_{D_{it}}(P^D)^{1-\sigma} \end{aligned}$$

Here one can see both the small open economy assumption and choice of numeraire. First, the small open economy assumption is that Denmark is sufficiently small that no foreign variables respond to it. Second, when updating X_{it} I adjust only by the change in (nominal) foreign prices, which implies that D_{it} is constant.

C.3 Final Algorithm for Supply and Demand

In spite of the above discussions, solving the model itself boils down to a shooting algorithm. In particular one first needs information on initial relative prices of goods and wage *levels*, $\{r_{t-1}, w_{t-1}\}$ as well as information on *changes* in exogenous variables $\left\{\Delta P^F, \Delta z\right\}$.Next, given a current guess of $w_t^{(j)}$ and information on $\{r_{t-1}, w_{t-1}\}$ as well as any changes in exogenous variables:

1. From exogenous changes in foreign prices and the current guess of wages and capital prices, use perfect competition and cost minimization to solve for the implied path of prices and continuation values.
2. Given a guess of the path of continuation values, solve for the equilibrium path of wages clearing markets in each period and in each market.
3. Update wages using a non-linear Gauss-Seidel updating rule:

$$w^{(j+1)} = \lambda w^{(j)} + (1-\lambda)w^{New}$$

I find the above algorithm to be reasonably stable given a conservative λ between .85 and .9. I initialize from a myopic guess, where in each period workers treat the current wage as that which will continue forever. This allows one to solve the model period-by-period. On a 2.7 GhZ Intel i7 processor, the Matlab code takes approximately 12 hours from start to finish.

D Data Appendix

D.1 Occupations and Occupational Characteristics

I use administrative data on workers in Denmark in order to estimate the worker supply model. In this section I describe the main data sets, the construction of occupation codes and cleaning, as well as the definition of income and other variables used. The last section discusses construction of occupational characteristics.

Description of Datasets

The extract of the Danish LEED which I rely on, called IDA, contains information on the position of each worker in *November* of the calendar year it is collected. This dataset contains information on the employment status of the worker, the worker's firm, the worker's total annual income from their November employer, hours worker (which are imputed), the worker's current level of education, age, as well as other demographic variables that I do not use. In addition, it contains information on the worker's occupation recorded at the ISCO four digit level and a quality flag that tells me the source of information for the occupational code.

In addition to IDAS, I use two datasets on establishments in Denmark, FIRE and FIRM. These contain standard information on the balance sheet of firms. The reason I use both datasets is that they contain information on the industry code of plants and firms in the economy. Whenever possible I rely on plant-level information in order to assign a worker to a particular industry. When this is not feasible I use firm level data.

For trade data, I use the Danish customs data, UHDI. This database contains different aggregations of the data. The extract on which I rely contains price and quantity data at the country-product-year level for each firm.

In addition to the administrative data I use several publicly available datasets: (1) the Danish National accounts; (2) the COMTRADE database as cleaned by CEPII (BACE); (3) Data on the Danish CPI from Statistics Denmark; (4) Data on the LIBOR rate from the FRED database. These datasets are all used in the calibration of the labor demand side and there is detailed information in that section.

Data Cleaning

Occupations - I use the two digit code for each four digit code in the data. In order to avoid imputed and low quality data, I follow the recommendation of Statistics Denmark and only use those codes coming from administrative surveys or pensions funds and not imputation or lagged information.[11] In addition to the use of Stats DK's quality flags I perform the following imputations: (1) if a worker is employed in the same occupation at $t-1$ and $t+1$ but has missing data at t, I impute the occupation as that at $t-1$; (2) if a worker switches firms between t and $t+1$ and there is data on occupation in $t+2$, I impute any missing occupational codes at $t+1$. The first imputation, if anything, biases me against finding large transition probabilities. The second imputation is done because a large amount of switching occurs through firms and often occupational codes are slow to adjust. If occupational codes change when in non-employment, I reset them to the occupational code observed in the last employed spell.

There are two changes to the ISCO's definitions: first, I combine all managerial occupations (10-13) into one code. This is because the paper is focused on horizontal, not upward mobility; second, I drop legislators and those employed in military service.

There are some occupations that are insufficiently observed in certain skill groups to accurately estimate transition rates. When this occurs, I have aggregated two digit occupations into the more dominant occupation (e.g., combining codes 51 and 52 into code 51). This predominantly affected high skilled workers who have at least the equivalent of an American bachelor's or master's and are, predictably, rarely working in low-skilled occupations. The table below summarizes these changes:

[11] This corresponds to quality codes 1,2,4, and 10. For details see:
http://www.dst.dk/da/Statistik/dokumentation/Times/personindkomst/discotyp.aspx

Small Occupation Code	Skill Groups Affected	Assigned to Code
92 (Ag. Laborers)	All	61 (Skilled Ag. Workers)
21 (Science Profs.)	Low, Medium	22 (Science Assoc. Profs.)
41 (Office Clerks)	High	42 (Customer Service Clerks)
51 (Personal Service Work)	High	52 (Salespersons)
71, 74 (Building/Crafts Workers)	High	72 (Machine Workers)
82, 83 (Drives, Plant Operators)	High	81 (Machinists)
93 (Manufacturing Laborer)	High	91 (Elementary Occupations)

Finally, some occupations that are very small in a certain sector (e.g., teachers in manufacturing) were moved across sectors because of the same small bins issue. Table (15) in the online table appendix in the appendix contains the full list of occupations that can be found in each sector.

Industry - I use NACE 1.1 four-digit codes for industries. These are assigned to plants, which are in turn matched to workers. I aggregate NACE 1.1 codes into 4 sectors. Manufacturing includes construction, agriculture and utilities and corresponds to NACE 1 2-digit codes 0-45; FIRE refers to NACE 1 2-digit codes 64-74; Public Services refer to NACE 1 2-digit codes 75, 80, 85-90; Other Services contains all remaining codes. Table (16) contains a complete description of this aggregation, including the shares of each industry in each sector.

Skill - IDA reports a worker's completed type of education. I group workers with high school or short cycle education in low skilled workers, workers with vocational or medium cycle education into middle skilled workers, and those with a bachelor's or more into high skilled workers. If workers education is ever reported as lower than in a previous year, I impute the higher education. If education is ever missing, I use the year before education if available, if that is unavailable, I use the next year reported education. The short-cycle education in Denmark is one to two years of some vocational training, some apprenticeship work and qualifies students for either continuing education or for basic occupations in fields as varied as IT to healthcare. The medium-cycle education system in Denmark is most similar to an American associate's degree but slightly more expansive. It is geared towards vocational training, but the breadth of topics includes traditional vocational careers, as well as STEM jobs, business, and teaching. In general, the Danish population has been steadily increasing in educational attainment over time, however there are still substantial workers who enter associate's and vocational type degrees. The second panel of 3 plots trends in these variables. There are secular declines in the share of workers with short cycle or lower education, and rapid growth in attainment of a university degree. These trends are uncorrelated with cyclical changes in the employment to population ratio, and reflect long run changes in the economy.

Occupational Tenure I construct occupational tenure for early cohorts using data from 1990 - 1996. Tenure is censored above at 6 years in order to maintain an accurate measure across cohorts. I construct tenure at the occupation-sector pair level.

Income - I focus on income as a measure of earnings, as this is the most well measured earnings concept in the Danish register data, and hours are often imputed. IDA reports total salary based income in November of each year, based on a worker's primary and any second occupation. I do not observe the secondary occupation, and would be unable to distribute the earnings. Hence, salaried income is fully attributed to the reported occupation.

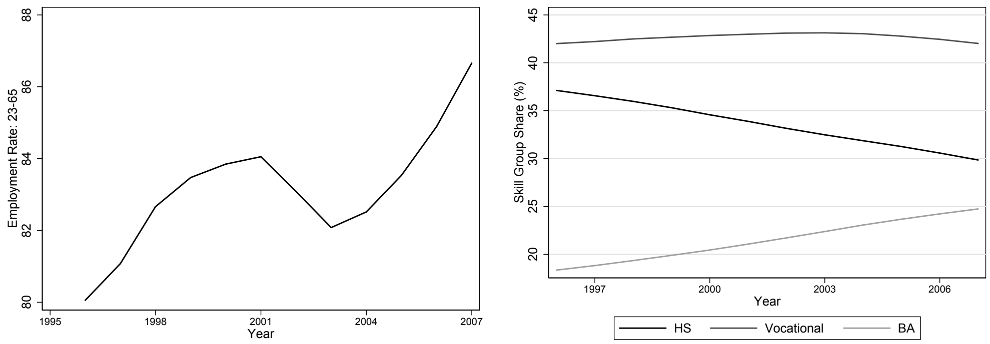
Employment - I count a worker as employed in t if they are employed in November of t. The first panel of Table 3 plots aggregate employment to population ratio using this definition. The ratio is about 82% on average. This is fairly close to aggregate figures from Statistics Denmark[12] (not pictured), which average 78% over this interval. Moreover, the trends, including the drop and recovery in the early 2000s, match up almost exactly.

In order to see the outcome of this data cleaning, tables A.7 and A.8 in the data appendix contain summary statistics in the raw sample of individuals and in the final sample frame. Employment is higher in the sample frame, as is the share of workers in the health and education sectors. The former occurs if a worker

[12]Figures on employment, unemployment, and not-in-labor-force can be downloaded from https://www.dst.dk/en/Statistik/emner/arbejde-indkomst-og-formue/beskaeftigelse

Figure 3: Employment and Educational Trends in Denmark

(a) Aggregate Employment (b) Trends in Educational Attainment



is removed because they are non-employed for the full duration of the sample. The latter occurs because higher quality income and occupational information is available in careers in the health and education sector. This is likely a result of these reflecting public sector occupations. There is somewhat underrepresentation of low skilled workers. This also likely reflects reporting quality and the fact that some low skilled workers may have been non-employed for the duration of the sample.[13] Otherwise the sample frame and raw data look very similar in terms of demographics and mean income.

Construction of Characteristics

In order to construct tasks I use the ONet database. This database contains comprehensive survey questions on various aspects of jobs. I use the "Work Activities," "Skills," "Abilities," and "Knowledge" surveys. These surveys contain 161 questions on 983 occupations at the SOC 2000 level. The questions are on a scale of 1-5. I follow Firpo et al. (2011) in combining importance and level scales into a composite index. Then I use quantiles in each survey response and treat the answers as cardinal.

In order to move from the large number of survey questions to a manageable number I use PCA, which is the solution to a factor model. In particular, I treat the Onet database occupation measures as signals generated by a set of underlying fundamental tasks; i.e., indexing occupations by o, fundamental tasks by t and observed questions by n leads to the following factor model for occupational characteristics:

$$x_{no} = \lambda_n' F_o + \varepsilon_{no}$$

where λ_n is an $T \times 1$ coefficient specific to data series n that maps the $1 \times T$ vector of tasks F_o to x.

After estimation, I have 10 characteristics for 983 occupation in the SOC 2000 classification. Notice that the actual tasks are not unique as factor models are only defined up to a normalization. For example, if one estimates factor models by PCA the implicit normalization is that $F'F/O = I_O$ where I is the identity. The assumption that tasks are uncorrelated is arbitrary but useful, especially given that I will use these to price distance in task space.

To aggregate this to the ISCO four digit classification I first match the SOC 2000 codes to the United States Census codes. Then I use the US employment shares to calculate weights. Because I wish for the characteristics to be time-invariant, I use the weights by pooling over the years 1999 - 2003, reweighing each year by its population size. This assumes that the US employment shares for the more disaggregated occupations are roughly similar to the Danish shares. Without more disaggregated data it is impossible to check whether this is accurate. However, within an ISCO code, the variance between characteristics is not

[13] This does not mean that these workers are involuntarily unemployed. Some of them may be parents who are at home, workers on disability, or workers in graduate school or some kind of other education for the full period.

large; on average the standard deviation of the characteristics within a code is 40% the standard deviation across all codes. If no census weights are available, I simply take the equally weighted mean over occupations within an ISCO 4 digit code. Once the four digit codes are mapped to, I use the occupational weights in Denmark to aggregate to two digit codes. To insure time-invariance I use the same pooling procedure as I did for the census codes.

D.2 Labor Demand/Calibration Details

In order to close the model for counterfactuals one must solve for the parameters governing production, consumption, capital and trade. The table below summarizes in detail the set of variables and the data source from which I draw the information. Because the algorithm relies on solving for variables in differences (see technical appendix), no data on initial relative prices, initial relative productivity, or home bias terms are required. In the remainder of this section I describe in some detail how I use the databases to calibrate various parameters.

Symbol	Size	Meaning	Data Sources
Production			
α	$\lvert\mathcal{I}\rvert \times 1$	Consumer Demand	Natl. Accounts
$B_{\mathcal{T},\mathcal{T}}$	$\lvert\mathcal{T}\rvert \times \lvert\mathcal{T}\rvert$	IO Matrix from Tradables to Tradables	Natl. Accounts
$B_{\mathcal{T},N\mathcal{T}}$	$\lvert\mathcal{T}\rvert \times \lvert\mathcal{NT}\rvert$	IO Matrix from Tradables to Nonradables	Natl. Accounts
$B_{\mathcal{NT},\mathcal{T}}$	$\lvert N\mathcal{T}\rvert \times \lvert\mathcal{T}\rvert$	IO Matrix from Nontradables to Tradables	Natl. Accounts
$B_{\mathcal{NT},\mathcal{NT}}$	$\lvert\mathcal{NT}\rvert \times \lvert\mathcal{NT}\rvert$	IO Matrix from Nontradables to Nontradables	Natl. Accounts
B_K	$\lvert\mathcal{I}\rvert \times 1$	Capital input coefficients	Natl. Accounts
B_L	$\lvert\mathcal{I}\rvert \times \lvert\mathcal{O}\rvert$	Labor input coefficients	IDAS
$\{\sigma_i\}_{i=1}^{\mathcal{T}}$	$\lvert\mathcal{T}\rvert \times 1$	Substitution elasticities	Broda and Weinstein (2006)
Prices			
r_0	$\lvert\mathcal{T}\rvert \times 1$	Initial relative prices	
$\left\{\Delta P_t^F\right\}_{t=1}^T$	$\lvert\mathcal{T}\rvert \times 1 \times T$	Path of foreign prices	UHDI/Comtrade
$\left\{\Delta P_t^{NT}\right\}_{t=1}^T$	$\lvert\mathcal{NT}\rvert \times 1 \times T$	Path of nontradable prices	Natl. Accounts
$\{\Delta r_{Kt}\}_{t=1}^T$	$\lvert\mathcal{I}\rvert \times 1 \times T$	Path of capital prices	LIBOR
$\{\Delta z_t\}_{t=1}^T$	$\lvert\mathcal{I}\rvert \times 1 \times T$	Path of industry productivities	Estimated
Income			
$\{r_{Kt}K_t\}_{t=1}^T$	$T \times 1$	Capital Income	Natl. Accounts
$\{X_{it}\}_{t=1}^T$	$\lvert\mathcal{T}\rvert \times 1 \times T$	Foreign Demand Shifter	Natl. Accounts
$\left\{\Delta D_{it}^F/A_{it}\right\}_{t=1}^T$	$\lvert\mathcal{T}\rvert \times 1 \times T$	Change in Relative Foreign Demand	Natl. Accounts

Production and Utility Function Parameters

In order to calculate the Cobb-Douglas coefficients in the production function I use expenditure shares from the Danish national accounts. Index industries by i, j and time with t. I use the time averaged shares leading to the following number:

$$\beta_{ij} = \frac{1}{T}\sum_{t=1}^{T}\frac{E_{ijt}}{E_{it}}$$

At the level of aggregation used in the analysis (NACE 1.1 2-digit level), there is very little time variation in these parameters over the time frame I consider. The R^2 of the expenditure share on i, j fixed effects is over .95.

From the national accounts I extract the capital share (defined as gross surplus) and the labor share in a similar fashion. To go from the labor share to the occupation coefficients I use the observed income shares in the data. Hence,

$$\beta_{oi}^{H} = \beta_{iL} \times \frac{\sum_{m \in \{o \cap i\}} y_m}{\sum_{m \in \{i\}} y_m}$$

where m indexes workers and y_m is worker income.

To calibrate consumer utility parameters I use final expenditure net of exports:

$$\alpha_i = \frac{1}{T} \sum_{t=1}^{T} \frac{E_{it} + M_{it} - \left(\sum_j E_{jit}\right) - X_{it}}{I_{it}}$$

I.e., the demand parameter is domestic absorption net of intermediates over total income (labor, capital, etc.).

Relative Prices and Productivity

To calibrate the path of relative prices (including the initial price) I use the share of expenditures and information on σ:

$$r_{it} = \left(\frac{E_{it}^{D}}{E_{it}^{F}}\right)^{1-\sigma_i}$$

In the baseline counterfactual, I set the Armington elasticity, σ_i, to 4 in all industries. This estimate comes from Simonovska and Waugh (2014). To calibrate changes in foreign prices I use the method and substitution elasticity estimates of Broda and Weinstein (2006), who build on the seminal work of Feenstra (1994). The CES price index for tradable goods can be constructed year-to-year by decomposing price changes into the *intensive margin* and the *extensive margin.* The former refers to products that are purchased in both years and only prices and quantities change. For this subset of goods, called the common set and denoted by I, the price index can be constructed using Sato-Vartia weights:

$$P_{t+1}^{Int}/P_{t}^{Int} = \prod_{i \in I} (p_{i,t+1}/p_{i,t})^{\frac{s_{i,t}-s_{i,t-1}}{\log(s_{i,t})-\log(s_{i,t-1})} \times \left(\sum_{j \in I} \frac{s_{j,t}-s_{j,t-1}}{\log(s_{j,t})-\log(s_{j,t-1})}\right)^{-1}}$$

where s is the share of purchases from within the common set of goods. To correct for the adjustment margin, let λ_{t+1} be the share of total expenditure in $t+1$ on varieties newly available in $t+1$ and let λ_t be the share of total expenditure in t on varieties that are no longer purchased in $t+1$. I.e., these are new and destroyed varieties of goods, respectively. In this case, the total price index is given by:

$$P_{t+1}/P_t = P_{t+1}^{Int}/P_{t}^{Int} \times (\lambda_{t+1}/\lambda_t)^{1/(\sigma-1)}$$

Notice, the elasticity of substitution, σ, only matters for understanding the impact of new varieties. This term is most important in times of trade liberalization as many new import relationships are formed, and old relationships destroyed. National accounts often ignore these changes (assuming prices are infinite in one or the other period), which leads to an upward bias in price indices whenever variety growth occurs. In my implementation, I use customs data aggregated to the country-product level to construct a variety-corrected CES index for each industry. I construct a similar index from Comtrade. In a last step, I use information from the national accounts, which are biased upwards as they ignore variety gains. I use the geometric mean of these three indices. This is an *ad hoc* attempt to smooth the time path of foreign price changes.[14]

To calibrate changes in capital prices I use data from the LIBOR on interest rates. As the depreciation rate, I use a value of $\delta = .1$ for all industries.

[14]For most products all three of these series are highly correlated. However, over time and for a handful of goods, the units in which certain goods were recorded changed. These time periods make the indices I construct from customs data volatile. The national accounts uses a government defined price index that appears to smooth over these issues.

Finally, given all of the parameters estimated thus far, changes in TFP can be backed out as a residual from the price system implied by Cobb-Douglas production and perfect competition:

$$\Delta \log z = \begin{bmatrix} B^M_{\mathcal{T},\mathcal{T}} - I_{|\mathcal{T}|} \\ \cdots \\ B^M_{\mathcal{NT},\mathcal{T}} \end{bmatrix} \Delta \log P_i^F + B^K \Delta \log p_k + B^L \Delta \log w$$
$$+ \begin{bmatrix} B^M_{\mathcal{T},\mathcal{T}} \circ \frac{r_t^{1-\sigma}}{1+r_t^{1-\sigma}} - I_{|\mathcal{T}|} & B^M_{\mathcal{T},\mathcal{NT}} \\ \cdots & \cdots \\ B^M_{\mathcal{NT},\mathcal{T}} \circ \frac{r_t^{1-\sigma}}{1+r_t^{1-\sigma}} & B^M_{\mathcal{NT},\mathcal{NT}} - I_{|\mathcal{NT}|} \end{bmatrix} \Delta \begin{bmatrix} \log r_i \\ \cdots \\ \log P_i \end{bmatrix}$$

The estimation of z is sensitive to the choice of δ and r_t. Nevertheless, given my assumptions on these variables, which are standard in the literature, I estimate an industry weighted TFP growth for Denmark over the period 1995 to 2005 of -5%. However, this is driven largely by a handful of outliers, including the energy sector which I do not use in the analysis. Removing these outliers leave a total change in TFP of nearly 0 ($< -1\%$). In the same period, the OECD estimates a roughly 0 change in TFP over the period, with slightly sub -.5% annual growth from 1997 to 2001 and slightly above .5% annual growth from 2001 to 2005. That these numbers are close suggest the calibration is reasonable.

Non-Labor Income

In order to calculate capital stock, I use data on aggregate gross surplus. In the baseline counterfactuals I use the aggregate stock of capital in 1996. For export demand, it's shown in the technical appendix that a sufficient statistic is a combination of true export demand, home bias, and the domestic price level. The exact statistic can be calculated from data on the level of exports, relative prices and σ. It is given by the formula:

$$X_{it} = \frac{\text{Exports}_{it}}{r_{it}^{(1-\sigma)}}$$

This completes the calibration.

E ACM Estimator Details

E.1 Review of ACM

For the set-up in ACM, once again let i index individuals, t index time, and o index occupation-sector pairs. The assumptions in ACM are similar to mine with a few key differences. Below are the four main assumptions in my model restated with equivalents in ACM:

1. **Timing and Information Sets**: Workers choose their occupations for $t+1$ in period t. They can only forecast wages at time $t+1$ using information at time t. However, they know the value of their moving cost shocks at time t.
2. **Errors on Human Capital**: There are no idiosyncratic shocks to human capital and workers are homogenous conditional on their occupation.
3. **Human Capital Accumulation:** Workers do not accumulate human capital in the model.
4. **GEV Shocks:** Moving cost shocks following a Gumbel distribution with scale parameter ρ.

From assumptions (1) - (4), the Bellman equation for a worker in occupation o at time t is given by:

$$V_t(o) = w_t(o) + \max_{o'} \{-C(o, o') + \rho\varepsilon_{o'} + \beta EV_{t+1}(o')\} \tag{E.1}$$

Using the properties of the logistic distribution, one can solve for the probability of switching occupations at time t (for a new occupation at time $t+1$) by:

$$\pi_t(o'|o) = \frac{\exp\left(-\frac{C(o,o')}{\rho} + \frac{\beta}{\rho} EV_{t+1}(o')\right)}{\sum_{o''} \exp\left(-\frac{C(o,o'')}{\rho} + \frac{\beta}{\rho} EV_{t+1}(o'')\right)} \tag{E.2}$$

Moreover, the value of an occupation is given by:

$$V_t(o) = w_t(o) + \rho\gamma - C(o, o'') + \beta E_t V_{t+1}(o'') - \rho \log \pi_t(o''|o) \tag{E.3}$$

Note that this formula will hold for any choice of o''. Taking the ratio of equation E.2 at o and o' yields:

$$\log \frac{\pi_t(o'|o)}{\pi_t(o|o)} = -\frac{C(o,o')}{\rho} + \frac{\beta}{\rho} E_t \left(V_{t+1}(o') - V_{t+1}(o)\right)$$

Plugging in E.7 for both $V_t(o)$ and $V_t(o')$ and choosing $o'' = o'$ for both yields:

$$\log \frac{\pi_t(o'|o)}{\pi_t(o|o)} = -\frac{C(o,o')}{\rho} + \frac{\beta}{\rho} E_t \left(w_{t+1}(o') - \rho \log \pi_{t+1}(o'|o') - w_{t+1}(o) + C(o,o') + \rho \log \pi_{t+1}(o'|o)\right)$$

Rearranging the above, and plugging in realizations and expectational error for the expected values yields the estimating equation:

$$\log \frac{\pi_t(o'|o)}{\pi_t(o|o)} = -\frac{(1-\beta)C(o,o')}{\rho} + \frac{\beta}{\rho}(w_{t+1}(o') - w_{t+1}(o)) + \beta \log \frac{\pi_{t+1}(o'|o)}{\pi_{t+1}(o'|o')} + \zeta_{oo't+1} \tag{E.4}$$

This regression has a similar flavor to the main text's with two differences. First, the "renewal action" is simply switching occupations from o to o'. There is no need for a third occupation because there is no human capital accumulation. Nevertheless, the identification is similar: one regresses wage differentials on flows with a particular control function for dynamic selection. The second difference is that, because of the timing assumption on choices, the coefficient on wages is scaled by β and wages will be correlated with the expectational error. I return to the latter issue in the implementation section.

E.2 Adding Observable Heterogeneity

Adding time invariant skills to ACM is trivial: one simply calculates transition rates separately by skill group. One can then pool by skill group, or estimate separate parameters. I choose to pool by skill group, as I do in the main text.

Adding age, which increases over time, is more difficult. ACM recommend, in lieu of a life-cycle model, approximating aging through a stochastic process, as in many macro models. This allows them to keep the simple structure of their estimator. In particular, suppose that age can be divided into a small number of bins and that workers belong to an age category indexed by $a = 1, ..., A$. At any moment in time there is a probability λ_a of moving from group a to $a+1$, with $\lambda_A = 0$. The worker's problem is now:

$$V_t(o,a) = w_t(o,a) + \max_{o'} \{-C(o,o',a) + \rho\varepsilon_{o'} + \beta E_t\left((1-\lambda)_a V_{t+1}(o',a) + \lambda_a V_{t+1}(o',a+1)\right)\} \tag{E.5}$$

Let $\tilde{V}(o,a) = ((1-\lambda)_a V_{t+1}(o',a) + \lambda_a V_{t+1}(o',a+1))$. Then the probability of switching occupations is only slightly modified from before:

$$\pi_t(o'|o,a) = \frac{\exp\left(-\frac{C(o,o')}{\rho} + \frac{\beta}{\rho} E\tilde{V}_{t+1}(o')\right)}{\sum_{o''}\exp\left(-\frac{C(o,o'')}{\rho} + \frac{\beta}{\rho} E\tilde{V}_{t+1}(o'')\right)} \tag{E.6}$$

And the solution to the value function is also given by:

$$V_t(o,a) = w_t(o,a) + \rho\gamma - C(o,o'') + \beta E_t \tilde{V}_{t+1}(o'') - \rho \log \pi_t(o''|o,a) \tag{E.7}$$

Now similarly defining $\tilde{w}(o) = \lambda_a w(o,a+1) + (1-\lambda_a) w(o,a)$, one rearrange equations E.6 and E.5 to yield a new estimating equation:

$$\begin{aligned}\log\frac{\pi_t(o'|o)}{\pi_t(o|o)} =& -\frac{(1-\beta)C(o,o')}{\rho} + \frac{\beta}{\rho}(\tilde{w}_{t+1}(o') - \tilde{w}_{t+1}(o)) \\ &+ \beta\left[\lambda_a \log\frac{\pi_{t+1}(o'|o,a+1)}{\pi_{t+1}(o'|o',a+1)} + (1-\lambda_a)\log\frac{\pi_{t+1}(o'|o,a)}{\pi_{t+1}(o'|o',a)}\right] + \zeta_{oo't+1}\end{aligned} \tag{E.8}$$

This slightly modified version of the ACM regression equation thus incorporates heterogeneity in age, and can easily done separately by skill group.

E.3 Details of ACM Implementation

In order to implement the ACM estimator, I need to specify instruments for wage differentials and specify the cost function C. Like ACM, for wages and flows at $t+1$ I use wages and flows at $t-1$. Twice lagged differentials tend to be more well behaved instruments and avoid yielding negative estimates of ρ. I fix β at .96, so I do not actually estimate this value. Instead, I run an instrumental variables regression of flows on wage differentials using twice lagged wage differentials and twice lagged flows as instruments. Since there are more instruments than variables, I can use two-stage GMM. I implement this in Stata using the `ivregress gmm` command and bootstrap standard errors, clustering over time periods and ignoring uncertainty in the estimation of transition rates and mean income.

In their original paper, ACM use a variety of specifications of C in their estimation. I find that including a simple constant, C, behaves very poorly and leads to negative values of ρ. This seems to be driven by a few aberrant observations, but is concerning. To that end, I improve the estimation (in terms both of sensible estimates and precision) by including simple occupational controls. In particular, I include an exit and entry cost, as in ACM and their follow, Artuç and McLaren (2015). I also include occupational characteristics. The addition of characteristics has little effect on point estimates but tends to decrease standard errors. Thus, the final specification includes wage differentials, fixed effects for entry occupation, for exit occupation, characteristics of each occupation and uses as instruments, lagged differentials and lagged flows. As a minor point, I use lagged wage differentials, not wages separately; and I use lagged values of the log difference in flows that are on the right hand side of E.4, not the flows separately.

When implementing the ACM estimator with observable heterogeneity, I split the sample by the three skill groups in the population and I bin workers into the sets, 23-35, 35-50, and 50-60. Notice that λ_a can be calculated directly from the data, as I observe workers aging. In this specification I use the same instruments as before properly transformed. That is to say, the instruments are twice lagged values of $\tilde{w}$ and twice lagged values of the function of flow probabilities in the second line of equation E.8. Age and skill are directly controlled for in the regression, and are treated as part of the switching cost. As in the main text, young workers in the highest skill group serve as the base group.

E.4 Review of AM

The model in Artuç and McLaren (2015) (henceforth, AM) is identical to that of ACM, and this section focuses on the homogenous worker case. In the implementation (discussed in the sequel), I estimate separate parameters (except for ρ) by skill type, and ignore age. The first key equation in AM's method is to combine equations (E.2) with (E.7). As an intermediating step, note that we can rearrange (E.7) so as to relate it to the choice probability denominator:

$$V_t(o) = w_t(o) + \rho \log\left(\sum_{o'}\left(\frac{-C(o,o') + \beta EV_{t+1}(o')}{\rho}\right)\right)$$

Hence, the choice probabilities can be rewritten as:

$$\pi_t(o'|o) = \exp\left(\frac{-C(o,o') + \beta E_t V_{t+1}(o') - [V_t(o) - w_t(o)]}{\rho}\right)$$

Finally, by multiplying the transition probability by the total number of workers in o at time t, one has the *level* of switchers at time t:

$$L_t(o'|o) = \exp\left(\frac{-C(o,o')}{\rho} + \frac{\beta E_t V_{t+1}(o')}{\rho} - \frac{[V_t(o) - w_t(o) - \log L_t(o)]}{\rho}\right) \tag{E.9}$$

Am rely on a two step procedure inspired by Silva and Tenreyro (2006)'s work on the gravity model. In the first stage, the estimate equation E.9 using Pseudo Poisson Maximum Likelihood (PPML) by saturating unobservables with fixed effects and parametrizing the cost function. Specifically, in the first stage one estimates the following model:

$$L_t(o'|o) = \exp\left(\frac{-C(o,o')}{\rho} + \lambda_{o't} + \alpha_{ot}\right) + \epsilon_{oo't} \tag{E.10}$$

In the second stage, one uses the structure of the model to construct a regression equation with no unobservables. In particular, AM point out the following relationship:

$$\lambda_{o,t-1} + \beta\alpha_{ot} - \beta L_{ot} = \frac{\beta}{\rho} w_t(o) + \zeta_t(o)$$

where ζ is the same expectational error that appears in the main text and in ACM. Due to the timing assumption, wages and expectational errors will be correlated, and so AM instrument with twice lagged wages. Collecting the left hand side parameters into a single term, ϕ_{ot} and using a hat to denote first stage estimates yields the final second stage estimating equation:

$$\hat{\phi}_{ot} = \frac{\beta}{\rho} w_t(o) + \zeta_{ot} \tag{E.11}$$

E.5 Details of AM Implementation

In implementing the AM estimator I add three components to E.11 that follow AM's implementation: (1) workers are differentiated by skill (but not age); (2) the first stage is estimated cross-section by cross-section, hence C is time varying; (3) there are non-pecuniary benefits to each occupation,[15] that follow a trend. The cost function is modeled as in AM: there is a fixed cost of switching sectors and/or occupations. There is also an interaction term that can be positive (suggesting convexity in the costs of switching along both margins) or negative (suggesting concavity in the costs of switching along both margins). Adding these ingredients and parametrizing the cost function yields the following estimating equations that largely mirror AM:

$$L_{st}(o'|o) = \exp\left(C_{Sec}\mathbf{1}\{Sec\} + C_{Occ}\mathbf{1}\{Occ\} + C_{Both}\mathbf{1}\{Sec\} \times \mathbf{1}\{Occ\} + \lambda_{o'st} + \alpha_{ost}\right) + \epsilon_{oo'st} \tag{E.12}$$

$$\hat{\phi}_{ot} = \zeta_{st} + \eta_o + \delta_o t + \frac{\beta}{\rho} w_t(o) + \zeta_{ot} \tag{E.13}$$

where s indexes skill groups. In order to estimate the first stage, I follow the advice of Santos Silva and Tenreyro (2011) and use Stata's iterated, reweighted least-squares algorithm from the `glm` command, as the usual poisson estimator can perform badly (in a numerical sense) with many fixed effects. Standard errors are calculated by block bootstrapping and clustered at the yearly level. In order to bootstrap, blocks of years (the estimation requires data from $t-2, t-1$ and t) are drawn with replacement $T-2$ times.

[15]In AM, these benefits also differ by skill group. I found that this number of fixed effects dramatically increased the imprecision on ρ (indeed, $\rho < 0$ appeared in some specifications), and so I treat these as constant across skill groups — this aligns with the main text. Nevertheless, these estimates (along with estimates that keep C constant, ignore δ and aggregate workers, are all available upon request.

F Sectoral Human Capital Extension

F.1 Modeling and Identifying Sectoral Human Capital

In the model of the main text, occupational-specific human capital is non-transferable across sectors, even in very similar occupations. While this keeps the focus on occupational human capital, it is also restrictive. In this appendix I outline how one can incorporate sector and occupation specific human capital and present the estimation of these parameters. Crucially, the model can be amended in a way that preserves finite dependence while allowing for different types of human capital.

In order to maintain finite dependence and allow for sectoral human capital, I treat sector human capital symmetrical with occupational human capital but keep track of these two states separately. Unlike in the main text, let o refer specifically to *occupations*, let $E \in \{0,1\}$ refer to employment status and let s refer specifically to sectors. Now we make occupation triples explicit: $(1, o, s)$ refers to a worker employed in sector s and occupation o; $(0, o, s)$ refers to a worker that is not-employed who most recently worked in occupation o and sector s. Notationally, let $\exp^o$ refer to experience in a particular 2-digit occupation and $\exp^s$ refer to experience in a particular sector. With this in hand, I modify Assumption 3 from the main text as follows:

Assumption 3′ *Sector specific experience resets upon switching sectors and similarly for occupational specific experience. However, sectoral (occupational) experience is fully transferred under any transition that preserves sectors (occupations). Mathematically, for workers going from any state to an employed state, sectoral human capital evolves according to:*

$$exp_t^s = \begin{cases} exp_{t-1}^s + 1 & \text{if } s_t = s_{t-1} \\ 0 & \text{if } s_t \neq s_{t-1} \end{cases}$$

while occupation specific human capital accumulates as follows:

$$exp_t^o = \begin{cases} exp_{t-1}^o + 1 & \text{if } o_t = o_{t-1} \\ 0 & \text{if } o_t \neq o_{t-1} \end{cases}$$

For workers going from any state to a non-employed state, sectoral human capital evolves according to:

$$exp_t^s = \begin{cases} exp_{t-1}^s & \text{if } E_{t-1} = 1 \\ 0 & \text{if } E_{t-1} = 0 \end{cases}$$

while occupation specific human capital accumulates as follows:

$$exp_t^o = \begin{cases} exp_{t-1}^o & \text{if } E_{t-1} = 1 \\ 0 & \text{if } E_{t-1} = 0 \end{cases}$$

i.e., one keeps their accumulated experience for the first year of non-employment but loses it if they are not re-employed after a period.

This is a similar set up to that in the main text. There is one point requiring discussion: how to define an "occupation" across sectors. For example, are drivers in services and in manufacturing truly performing the same occupation? I proceed under the seemingly reasonable assumption that occupational capital is transferable across sectors within two digit occupational codes. However, two possibilities exist. First, transferring sectors can destroy all specific human capital. In this scenario, there would be a "manufacturing" capital as well as a capital for each occupation within manufacturing, and similar for other sectors. Second, transferring sectors can only destroy sector-specific human capital if one's occupation remains the same. In this scenario, managers who move from manufacturing to services would keep their managerial experience. From the perspective of a model, these are arbitrary decisions, while from the perspective of data they are related to decisions over aggregation. The model can be extended in either direction. Indeed, if a modeler had foreknowledge of which occupations had knowledge that could be moved across sectors and those that did not (e.g., managers versus drivers) they could specify any arbitrary system of allowed spillovers. The table below outlines how capital accumulates starting from an initial point (s, o) to different combinations of sectors and occupations both in the case that occupational human capital does and does not move across sectors.

$(s, o) \rightarrow$	(s, o)	(s', o)	(s, o')	(s', o')
Case I	$exp_s + 1, exp_o + 1$	$0, 0$	$exp_s + 1, 0$	$0, 0$
Case II	$exp_s + 1, exp_o + 1$	$0, exp_o + 1$	$exp_s + 1, 0$	$0, 0$

Regardless of the rules governing human capital accumulation, workers that switch either occupations or sectors or both see some of the specific human capital destroyed (while their *general* human capital continues to accumulate). In terms of the model's solution, the only change here is that ω, the worker's state, now contains *both* occupational and sectoral experience. The key equations governing worker's value functions and decision rules remain the same:

$$\begin{aligned}\tilde{V}_t(\omega, (o, s)) =& \gamma + \frac{u_t((o, s), (o, s)', \omega)}{\rho} + \beta E_t \tilde{V}_{t+1}\left(T(\omega, (o, s), (o, s)'), (o, s)'\right) \\ & - \log \pi_t(\omega, (o, s), (o, s)')\end{aligned} \tag{F.1}$$

$$\begin{aligned}\log \pi_t(\omega, (o, s), (o, s)') =& u_t((o, s), (o, s)', \omega) + \beta E_t V_{t+1}\left(T(\omega, (o, s), (o, s)'), (o, s)'\right) \\ & - \log D_t(\omega, (o, s))\end{aligned} \tag{F.2}$$

where, as in the main text, D is the inclusive value of state $(\omega, o, s)_t$ and u is the flow payoff. The only difference from the equations in the main text is that I have made explicit that o is actually an occupation-sector pair, (o, s) (while suppressing E to economize on space).

In analyzing the expressions above, the first thing to note is that these changes imply nothing for the first stage estimation of wage parameters, switching probabilities and unobservables via the EM algorithm. The exclusion restriction in this first stage is that the shock to wages occurs after selection has occurred so that the expected wage in period t only depends on contemporaneous observables, lagged observables, and the time-invariant unobservable. Since this holds regardless of the elements of ω, the first stage is exactly as in the main text. Threats to separately identifying returns to sectoral versus occupational human capital remain if there is insufficient movement of workers and insufficient occupations within and across sectors. As an extreme example, if there is a sector-occupation pair that is distinct from all other occupations and sectors, one could never separately identify sectoral and occupational human capital parameters since these are perfectly multicollinear. Similarly, if there are many occupations within a sector, but no one switches across these occupations, then returns to occupational and sectoral experience cannot be separately identified as observed experience growth in either would be perfectly multicollinear. In practice, I have no such narrow occupation-sector pairs nor is there insufficient movement in the data.

With the first stage estimated, one can turn to the second stage estimation, which identifies the level and variance of switching costs. Before proceeding, it will be useful to rewrite the cost function being estimated:

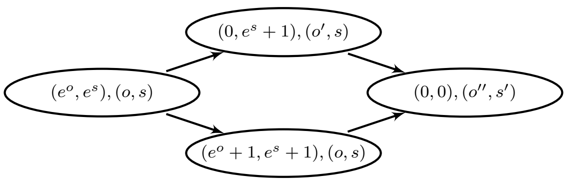
$$C((o, s), (o', s')) = \exp\left\{\Gamma_0 + \sum_{v \in \mathcal{V}} |v_{o,s} - v_{o',s'}| \Gamma_v^{+/-} + \Gamma_o D_{o \neq o'} + \Gamma_d D_{s \neq s'}\right\} \tag{F.3}$$

where the v terms reflect changes in occupational attributes, while Γ_o and Γ_s are constants for switching occupations and sectors respectively. As shown in the main text, the constant term measures any complementarities in switching both occupations and sectors simultaneously. Variation in occupation-sector pair characteristics will identify the loadings on different tasks. However, what identifies the constant terms? In order to separately identify the cost of switching occupations only, sectors only, and both one needs to construct career paths featuring these kinds of movements that nevertheless ensure finite dependence holds.[16]

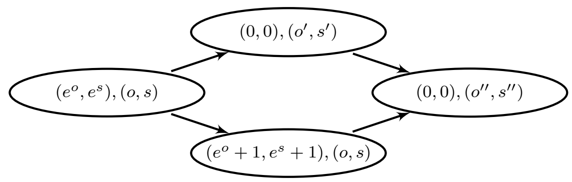
[16]One can always lean on functional form restrictions. Indeed, I posit a log-linear cost function instead

From table (1), one can derive several career paths for workers that fulfill this identification need and the finite dependence assumption. In particular, if workers begin in the same occupation sector pair at time t, with identical observables (including experience), then finite dependence will hold if at $t+2$ one of the following holds: (1) both workers switch into an occupation and a sector they did not work in at $t+1$ and one worker switches either sectors, occupations or both at $t+1$; (2) both workers stay in their initial sector in all periods, switch into a new occupation at $t+2$ and one worker switched occupations at t; (3) both workers stay in their initial occupation in all periods, switch into a new sector at $t+2$ and one worker also is in a different sector at t. I will call this a full-switch, a within-sector switch, and a within-occupation switch. The figure below summarizes the types of transitions that can occur:

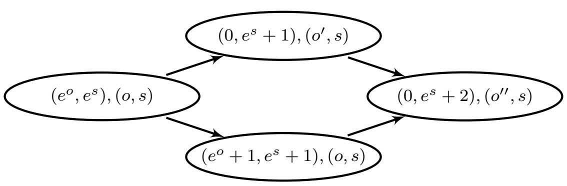
(a) Full Switch Example 1



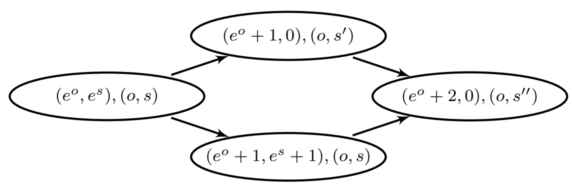
(b) Full Switch Example 2



(c) Within Sector



(d) Within Occupation



An immediate drawback to this method is the need to observe workers moving across sectors frequently *and* these sectors must be different. One can naturally ask how often this occurs. Approximately 5.4% of workers work in at least 3 sectors. This increases to 6.2% of workers who are observed for at least three periods. However, this number drops to a tiny fraction when one conditions on keeping the same occupation throughout. In fact, out of 20,000 possible paths observed in the data, less than 300 fulfill this requirement. Partially, this is a function of having many occupations and few sectors. Nevertheless, this is in sharp contrast to the the number of workers who work in at least three occupations over the 12 year period under study. Over 13.5% of people in the sample work in at least three occupations, and this climbs to 16% for workers present for at least three periods. Moreover, there is a great deal of within-sector occupational mobility, as discussed in the main text. Hence, while switching sectors is not uncommon, experiencing many sectoral switches and keeping one's occupation is decidedly rarer. Hence, an overly flexibly cost function may not be identified and will introduce noise into estimation. To this end, I report results with and without the interaction term. The latter specification does not preclude all interaction between switching sectors versus occupations; indeed, the type-specific constants and the component of costs related to characteristics dominates in all settings. Importantly, characteristics still differ by occupation-sector pair. This is true even if the occupation code or sector code does not change.[17]

of arbitrarily estimating many fixed effects. Nevertheless, the point holds in a non-parametric setting and I use all possible paths for identification.

[17] As a reminder, occupation-sector characteristics are determined by the composition of 3 digit occupations in each cell, as these compositions change so will the characteristics, even for the same occupational code. For example, "drivers" in manufacturing and services have different characteristics because most of the drivers in the former operate heavy trucks or machinery, while most in the latter operate light trucks or simply cars.

This section demonstrates how easily the finite dependence framework can be adapted to accommodate multiple kinds of human capital. The literature on worker adjustment (including movement across firms) often abstracts from different kinds of specific human capital. Nevertheless, it is clear that specific experiences matter a great deal in explaining cross-sectional variation in incomes, the time path of income for an individual, and thus also the reluctance of workers to abandon a sinking ship. The methods above are thus a promising avenue to incorporate specific human capital in other contexts, such as models that allow for workers to switch across both firms and industries or firms and occupations.

Finally, the above discussion also demonstrates some of the limitations of the finite dependence framework. In particular, I treat human capital as *observed*. The assumption that human capital resets to 0 is not necessary, but the value taken after a switch must be known and there must be a series of moves that workers take that reset their value functions.[18] This rules out an unobserved level of human capital that accumulates with experience and depreciates at an unknown rate with switching. With a sufficiently long panel, one could allow for a finite number of human capital levels per occupation and then estimate a transition matrix across these unobservables for each type. Such an extension of the first-stage EM algorithm was discussed in appendix B with many more details in AM. Nevertheless, for $|\mathcal{O}|$ occupations, K types and H levels of human capital, this requires estimating $|\mathcal{O}| \times K \times H$ parameters of a transition matrix, as well as the actual returns to being in each state. These numbers may be reasonable if the choice set is small, but in my setting this would amount to approximately 400 parameters. One promising alternative may be to avoid modeling comparative advantage separately per occupation and instead model a low-dimensional hidden Markov structure governing comparative advantage. For example, there could be m hidden states and occupation-sector pairs pay different returns on each state. This is the strategy pursued by Dvorkin (2017).

F.2 Results

While all the parameters have been re-estimated, in this subsection I focus on the parameters that change the most across specifications. The EM algorithm suggests that within each skill group there is a type that has absolute advantage over the other type, and I continue to use "high" and "low" type to refer to this distinction. The correlation between the comparative advantage vectors across this specification and that in the original model is .9975[19], and so there is not much difference in unobserved comparative advantage between these two specifications. However, things differ in returns to tenure, the sensitivity to wage differentials, and in the magnitude of moving costs.

Turning first to the returns to tenure, table F.2 contains the coefficients from the wage regressions allowing for multiple kinds of tenure. Unsurprisingly, the returns to occupational and sectoral tenure estimated separately tend to be lower than the occupation-sector returns to tenure estimated in the original specification. As I now allow for partial transferability of human capital, the coefficients on age (capturing general labor market experience) also tend to decrease. In a taxonomy of models studying labor market adjustments in the trade literature, to the best of my knowledge, all papers have either ignored human capital accumulation or allowed only for sectoral human capital accumulation.[20]. Hence, the specification here nests most of the extant work. The coefficients on all kinds of human capital are on the same order of magnitude, suggesting that intra-sectoral human capital accumulation is non-negligible.

Concretely, the median return to occupation specific human capital is 1.8% per year, and ranges from essentially 0 to 4.5% per year. To put these numbers in perspective, the average manufacturing wage in Denmark in the sample (in 2015 USD) is approximately 58 thousand dollars. Hence, a manufacturing worker earning the mean return on their experience would lose 5 thousand dollars in income if a shock

[18] In Arcidiacono and Miller (2018), Arcidiacono and Miller propose an extension of finite dependence based on weighting paths that does not strictly require a renewal action. Nevertheless, one always needs to know the transition on states and be able to observe the states (perhaps after a first stage).

[19] This correlation is done unconditionally and can be contaminated by differences in means across groups. The correlation in comparative advantage across occupations conditional on type varies between .936 and .996

[20] This is not true of all labor market adjustment models. However, as international trade tends to study reallocation across industries, this has been the dominant margin studied in that field

pushed them into a new occupation, even if the new occupation pays similarly per skill unit and the worker's comparative advantage were constant. While these results showcase the importance of modeling occupations narrowly, they also highlight that human capital appears to be transferable, especially across occupations within a sector.[21] While the main text focuses on the drastic situation where human capital is specific to an occupation-sector pair, understanding more deeply the multifaceted nature of workers' human capital is clearly an important avenue of research for anyone interested in understanding the short run burden of reallocation.

The changes in the estimated returns to human capital have immediate ramifications for the second stage estimates of the cost function. All relevant structural parameter estimates can be found in tables F.3-F.7 in section F.3, below. Table F.3 shows the estimates for $1/\rho$ and the average value of C/ρ when allowing for partially transferable human capital in the unrestricted and restricted cases discussed above. Moreover, tables F.5 and F.4 contain the estimates for the cost function parameters. The most striking difference in estimates is in $1/\rho$ which climbs from 1.5 to between 2.1 and 2.8 in the new specification. To understand why, recall that $1/\rho$ is identified off of workers' sensitivity to wage differentials after properly controlling for selection. Hence, an increase in $1/\rho$ suggests more sensitivity to occupational and sectoral wage differentials. If tenure was non-transferable or the returns to tenure were low, then richer specifications of human capital would have limited impact on 1ρ, as there would be limited changes in wage differentials across specifications. However, the increase in $1/\rho$ suggests that when properly controlling for human capital, workers are even more responsive to wage differentials than previously measured.

The implications of a richer specification of human capital on estimated moving costs are ambiguous. On the one hand, transferable human capital shrinks wage differences across occupations and sectors, which helps explains the low rate of switching observed empirically. However, the higher sensitivity to wage differentials also suggests a more acute response to the aforementioned smaller differences. Overall, the effect on C/ρ is quite small, but due to the change in ρ the overall switching costs smaller (on average, 79% of the original costs). Hence, more flexibility in human capital accumulation tends to lower the fixed costs of switching occupations to 3 years of income or less. This is relatively reasonable number, especially in light of the very large costs implied by models that do not control for human capital accumulation and other heterogeneity across workers. Nevertheless, while different than the numbers in the main text, the change in moving costs is not nearly as large when including transferable human capital as the change from simply including human capital in the model.

The non-pecuniary factors and value of non-employment shift in terms of levels–in particular, they are closer to 0 in these specification. Partially this reflects the fact that the variance in shocks is smaller. Scaling the values of η by ρ would bring them closer together. Nevertheless, the values of η are smaller when human capital can be transferred. Recall that the normalized state is non-employment, so η being closer to 0 suggests that non-employment is not as bad a state. Understood this way, the shrinking value of η is intuitive: if human capital can be transferred, then workers that switch occupations have higher incomes than they would if they could not keep their human capital; this means the model can more easily rationalize workers switching occupations more than they switch into non-employment without appealing to unobservables. The relative values of η, however, do not change much. The correlation coefficient between the original and new values of η is .925. This suggests that the importance of transferable human capital is not in pinning down the relative sizes of occupations, but in the attractiveness of non-employment relative to employment.

[21]This model does not strictly nest the model in the main text as there are no returns to occupation-sector pairs measured. Income regressions using person-occupation FE, in lieu of the full structural model, suggest that when all terms are allowed transferability is not rejected but neither is occupation-sector pair specific human capital. Indeed, in a regression with coefficients common across occupations and sectors, the returns to pair, occupation and sector human capital are, respectively, 1.2, 1.4 and 3.5%.

F.3 Tables

Table F.1: Time Averaged Skill Prices Across Occupations

Occupation	Manufacturing	Other Services	FIRE	Health & Educ.
Managers	0.469	0.354	0.365	
	(0.023)	(0.011)	(0.018)	
Science Professional	0.643	0.420	0.570	
	(0.020)	(0.016)	(0.018)	
Science Assc. Professional	0.444	0.426	0.409	
	(0.013)	(0.014)	(0.015)	
Other Assc. Professional	0.404	0.320	0.382	
	(0.011)	(0.005)	(0.008)	
Clerks	0.288	0.254	0.216	0.178
	(0.009)	(0.004)	(0.007)	(0.006)
Agriculture	0.449			
	(0.021)			
Building Trades	0.303			
	(0.006)			
Metal Trades	0.410	0.380		
	(0.008)	(0.009)		
Other Crafts	0.292			
	(0.011)			
Plant Operator	0.303			
	(0.014)			
Machine Operator	0.363			
	(0.005)			
Drivers	0.419	0.320		
	(0.019)	(0.009)		
Laborers	0.359	0.202		
	(0.008)	(0.005)		
Other Professional		0.246	0.248	
		(0.006)	(0.008)	
Personal Workers		0.119		0.219
		(0.003)		(0.002)
Retail Workers		0.126		
		(0.003)		
Elementary Occupations		0.143		
		(0.002)		
Customer Service			0.350	
			(0.008)	
Health Professional				0.428
				(0.016)
Teachers				0.211
				(0.003)
Health Assc. Professional				0.328
				(0.005)
Teaching Assc. Professional				0.295
				(0.004)

Notes: Skill prices time-averaged for clarity. Units are relative to unconditional mean income (normalized to 1). Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

Table F.2: Income Regression Coefficients

(a) Manufacturing

	$\beta_{Ten,Occ}$	$\beta_{Ten,Sec}$	β_{Age}	β_{Age^2} [a]	$\beta_{Type:1}$	$\beta_{Type:2}$	$\beta_{Type:3}$	$\beta_{Type:4}$	$\beta_{Type:5}$	σ^2
Managers	0.009	0.055	0.039	-0.000	-1.270	-0.149	-0.913	-0.107	-0.581	0.145
	(0.001)	(0.001)	(0.002)	(0.000)	(0.038)	(0.006)	(0.027)	(0.004)	(0.031)	(0.002)
Science Prof.	0.014	0.053	0.027	-0.000			-0.623	-0.085	-0.266	0.142
	(0.001)	(0.001)	(0.001)	(0.000)			(0.038)	(0.004)	(0.013)	(0.001)
Science Assc. Prof.	0.017	0.050	0.041	-0.000	-0.978	-0.177	-0.525	-0.128	-0.424	0.125
	(0.001)	(0.001)	(0.001)	(0.000)	(0.046)	(0.006)	(0.007)	(0.003)	(0.019)	(0.001)
Other Assc. Prof.	0.021	0.059	0.041	-0.000	-0.758	-0.148	-0.525	-0.100	-0.362	0.154
	(0.001)	(0.001)	(0.001)	(0.000)	(0.022)	(0.004)	(0.010)	(0.003)	(0.013)	(0.001)
Clerks	-0.000	0.082	0.050	-0.001	-0.980	-0.189	-0.548	-0.125	-0.641	0.154
	(0.001)	(0.001)	(0.001)	(0.000)	(0.018)	(0.007)	(0.011)	(0.006)	(0.029)	(0.001)
Agriculture	-0.001	0.030	0.028	-0.000	-1.217	-0.078	-1.038	0.010	-1.282	0.203
	(0.002)	(0.002)	(0.002)	(0.000)	(0.028)	(0.018)	(0.027)	(0.018)	(0.045)	(0.003)
Building Trades	0.026	0.040	0.044	-0.000	-0.751	-0.055	-0.667			0.243
	(0.001)	(0.001)	(0.001)	(0.000)	(0.012)	(0.003)	(0.006)			(0.001)
Metal Trades	0.024	0.046	0.033	-0.000	-0.663	-0.052	-0.501	0.025	-0.939	0.180
	(0.001)	(0.001)	(0.001)	(0.000)	(0.015)	(0.007)	(0.011)	(0.006)	(0.030)	(0.001)
Other Crafts	0.040	0.038	0.046	-0.000	-0.818	-0.172	-0.613			0.214
	(0.001)	(0.002)	(0.002)	(0.000)	(0.017)	(0.004)	(0.013)			(0.002)
Plant Operator	0.029	0.044	0.047	-0.001	-0.604	-0.010	-0.310	0.112	-0.820	0.169
	(0.001)	(0.002)	(0.002)	(0.000)	(0.026)	(0.007)	(0.016)	(0.006)	(0.027)	(0.002)
Machine Operator	0.028	0.051	0.035	-0.000	-0.716	-0.108	-0.460			0.199
	(0.001)	(0.001)	(0.001)	(0.000)	(0.010)	(0.002)	(0.005)			(0.001)
Drivers	0.027	0.053	0.031	-0.000	-0.738	-0.045	-0.534			0.165
	(0.001)	(0.001)	(0.002)	(0.000)	(0.031)	(0.004)	(0.030)			(0.002)
Laborers	0.033	0.058	0.032	-0.000	-0.806	-0.063	-0.523			0.204
	(0.001)	(0.001)	(0.001)	(0.000)	(0.013)	(0.002)	(0.009)			(0.001)

(b) Services

	$\beta_{Ten,Occ}$	$\beta_{Ten,Sec}$	β_{Age}	β_{Age^2} [a]	$\beta_{Type:1}$	$\beta_{Type:2}$	$\beta_{Type:3}$	$\beta_{Type:4}$	$\beta_{Type:5}$	σ^2
Managers	0.008	0.052	0.049	-0.000	-1.263	-0.113	-0.941	-0.080	-0.908	0.176
	(0.001)	(0.001)	(0.001)	(0.000)	(0.023)	(0.005)	(0.026)	(0.004)	(0.043)	(0.002)
Science Prof.	0.010	0.059	0.039	-0.000	-1.462	-0.027	-1.128	0.009	-0.884	0.156
	(0.001)	(0.001)	(0.002)	(0.000)	(0.017)	(0.006)	(0.045)	(0.005)	(0.032)	(0.002)
Other Prof.	0.025	0.042	0.062	-0.001	-1.218	-0.048	-0.986	-0.034	-0.553	0.183
	(0.001)	(0.001)	(0.001)	(0.000)	(0.015)	(0.003)	(0.032)	(0.002)	(0.007)	(0.001)
Science Assc. Prof.	0.011	0.051	0.044	-0.000	-1.143	-0.119	-0.600	-0.092	-0.943	0.140
	(0.001)	(0.001)	(0.001)	(0.000)	(0.023)	(0.005)	(0.020)	(0.004)	(0.031)	(0.001)
Other Assc. Prof.	0.024	0.041	0.048	-0.001	-0.982	-0.037	-0.527	-0.002	-0.512	0.156
	(0.000)	(0.000)	(0.001)	(0.000)	(0.011)	(0.003)	(0.008)	(0.003)	(0.010)	(0.001)
Clerks	0.010	0.063	0.050	-0.001	-1.082	-0.066	-0.525	-0.025	-1.077	0.140
	(0.000)	(0.001)	(0.001)	(0.000)	(0.010)	(0.005)	(0.007)	(0.004)	(0.010)	(0.001)
Personal Services	0.013	0.098	0.069	-0.001	-0.957	0.008	-0.625	0.152	-0.963	0.250
	(0.001)	(0.001)	(0.001)	(0.000)	(0.010)	(0.008)	(0.011)	(0.008)	(0.011)	(0.001)
Retail Occs.	0.023	0.056	0.082	-0.001	-1.012	-0.179	-0.567			0.251
	(0.001)	(0.001)	(0.001)	(0.000)	(0.007)	(0.003)	(0.004)			(0.001)
Metal Trades	0.002	0.055	0.039	-0.000	-0.921	-0.061	-0.766			0.143
	(0.001)	(0.001)	(0.001)	(0.000)	(0.036)	(0.005)	(0.026)			(0.001)
Drivers	0.030	0.057	0.043	-0.000	-0.782	-0.062	-0.452			0.210
	(0.001)	(0.001)	(0.001)	(0.000)	(0.015)	(0.003)	(0.010)			(0.001)
Elementary Occs.	0.019	0.034	0.068	-0.001	-0.836	-0.038	-0.555	0.131	-0.914	0.237
	(0.001)	(0.001)	(0.001)	(0.000)	(0.007)	(0.006)	(0.008)	(0.006)	(0.009)	(0.001)
Laborers	0.040	0.073	0.055	-0.001	-1.004	-0.055	-0.579			0.256
	(0.001)	(0.001)	(0.001)	(0.000)	(0.014)	(0.003)	(0.015)			(0.002)

(c) FIRE Industries

	$\beta_{Ten,Occ}$	$\beta_{Ten,Sec}$	β_{Age}	β_{Age^2} [a]	$\beta_{Type:1}$	$\beta_{Type:2}$	$\beta_{Type:3}$	$\beta_{Type:4}$	$\beta_{Type:5}$	σ^2
Managers	-0.047	0.069	0.045	-0.000	-1.467	-0.107	-1.377	-0.049	-1.166	0.133
	(0.002)	(0.001)	(0.002)	(0.000)	(0.026)	(0.006)	(0.027)	(0.004)	(0.027)	(0.002)
Science Prof.	0.017	0.052	0.031	-0.000	-1.726	-0.061	-0.433	-0.023	-0.299	0.176
	(0.001)	(0.001)	(0.001)	(0.000)	(0.020)	(0.004)	(0.021)	(0.003)	(0.012)	(0.002)
Other Prof.	0.013	0.082	0.062	-0.001	-1.088	-0.126	-0.432	-0.084	-0.443	0.163
	(0.001)	(0.001)	(0.002)	(0.000)	(0.043)	(0.005)	(0.030)	(0.003)	(0.019)	(0.002)
Science Assc. Prof.	-0.002	0.074	0.042	-0.000	-1.008	-0.096	-0.524	-0.067	-0.425	0.166
	(0.001)	(0.001)	(0.002)	(0.000)	(0.026)	(0.005)	(0.010)	(0.004)	(0.014)	(0.001)
Other Assc. Prof.	0.011	0.077	0.041	-0.000	-0.885	-0.139	-0.522	-0.092	-0.391	0.154
	(0.001)	(0.001)	(0.001)	(0.000)	(0.023)	(0.003)	(0.008)	(0.003)	(0.013)	(0.001)
Clerks	0.001	0.115	0.054	-0.001	-1.008	-0.145	-0.489	-0.078	-0.863	0.207
	(0.001)	(0.001)	(0.001)	(0.000)	(0.016)	(0.006)	(0.011)	(0.005)	(0.016)	(0.002)
Customer Service	-0.011	0.061	0.040	-0.000	-0.965	-0.068	-0.484			0.140
	(0.001)	(0.001)	(0.001)	(0.000)	(0.011)	(0.003)	(0.006)			(0.001)

(d) Health & Education

	$\beta_{Ten,Occ}$	$\beta_{Ten,Sec}$	β_{Age}	β_{Age^2} [a]	$\beta_{Type:1}$	$\beta_{Type:2}$	$\beta_{Type:3}$	$\beta_{Type:4}$	$\beta_{Type:5}$	σ^2
Health Prof.	0.037	-0.020	0.046	-0.000					-0.456	0.190
	(0.001)	(0.001)	(0.002)	(0.000)					(0.006)	(0.001)
Teachers	0.046	0.002	0.065	-0.001	-1.357	-0.111	-1.212	-0.028	-0.899	0.137
	(0.001)	(0.001)	(0.001)	(0.000)	(0.006)	(0.003)	(0.015)	(0.003)	(0.008)	(0.001)
Health Assc. Prof.	0.017	0.014	0.042	-0.000	-1.183	-0.221	-0.613	0.012	-0.512	0.146
	(0.001)	(0.001)	(0.001)	(0.000)	(0.007)	(0.005)	(0.014)	(0.003)	(0.006)	(0.001)
Teaching Assc. Prof.	0.032	0.012	0.040	-0.000	-1.226	-0.089	-1.236	0.017	-0.533	0.152
	(0.001)	(0.001)	(0.001)	(0.000)	(0.004)	(0.003)	(0.006)	(0.002)	(0.006)	(0.001)
Clerks	0.026	0.059	0.051	-0.001	-0.848	0.011	-0.395	0.106	-0.925	0.178
	(0.001)	(0.001)	(0.001)	(0.000)	(0.016)	(0.007)	(0.011)	(0.007)	(0.020)	(0.001)
Personal Services	0.042	0.029	0.043	-0.000	-0.847	-0.017	-0.423	0.027	-0.836	0.170
	(0.001)	(0.000)	(0.000)	(0.000)	(0.005)	(0.004)	(0.005)	(0.004)	(0.010)	(0.000)

[a] Presented $\times 10^3$ for clarity.

[b] Coefficients from a log-linear Mincer regression of wages on worker attributes. Types refer to estimates of unobservable heterogeneity across workers. Low skilled workers are either of type 1 or 2; Medium skilled workers are of type 3 or 4; High skilled workers are of type 5 or 6. Type 6 coefficients all normalized to 0. Skill prices are in table A.1. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

Table F.3: Switching Elasticity and Switching Costs with Transferable HC

	With Constant	No Constant
$1/\rho$	2.146	2.784
	(0.057)	(0.056)
Mean C/ρ	5.511	4.938
	(0.046)	(0.044)

Notes: Results from regressing transitions rates on wage differentials. Mean C/ρ refers to simple mean across all cells of cost matrix with no adjustment for observed transition rates. Standard errors, in parentheses, calculated from 100 block bootstrap samples.

Table F.4: Mobility Productivity Parameters with Transferable HC

	With Constant	No Constant
Age	0.010	0.016
	(0.000)	(0.000)
Age^2($\times$1000)	-0.036	-0.202
	(0.011)	(0.012)
Type 1	-0.013	0.047
	(0.009)	(0.010)
Type 2	0.075	0.130
	(0.008)	(0.009)
Type 3	0.001	0.059
	(0.007)	(0.008)
Type 4	0.086	0.139
	(0.008)	(0.008)
Type 5	-0.060	-0.003
	(0.011)	(0.012)

Notes: Coefficients from a log-linear inverse productivity function. Types refer to estimates of unobservable heterogeneity across workers. Low skilled workers are either of type 1 or 2; Medium skilled workers are of type 3 or 4; High skilled workers are of type 5 or 6. Type 6 coefficients all normalized to 0. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

Table F.5: Mobility Cost (Γ) Parameters with Transferable HC

	With Constant		Without Constant	
	Up Tasking	Down Tasking	Up Tasking	Down Tasking
Constant	0.763			
	(0.012)			
Occ. Dummy	0.081		0.088	
	(0.004)		(0.004)	
Sec. Dummy	0.631		1.279	
	(0.008)		(0.011)	
Task 1	−0.006	0.031	−0.007	0.032
	(0.002)	(0.002)	(0.002)	(0.002)
Task 2	0.008	0.028	0.009	0.028
	(0.002)	(0.002)	(0.002)	(0.002)
Task 3	0.120	−0.134	0.123	−0.137
	(0.003)	(0.006)	(0.002)	(0.005)
Task 4	0.035	−0.017	0.037	−0.017
	(0.006)	(0.006)	(0.006)	(0.006)
Task 5	0.006	0.039	0.007	0.040
	(0.005)	(0.005)	(0.005)	(0.005)
Task 6	−0.043	0.063	−0.047	0.065
	(0.006)	(0.006)	(0.006)	(0.006)
Task 7	−0.015	0.061	−0.003	0.058
	(0.016)	(0.014)	(0.015)	(0.013)
Task 8	−0.082	0.114	−0.084	0.125
	(0.013)	(0.011)	(0.012)	(0.010)
Task 9	0.127	−0.018	0.146	−0.029
	(0.012)	(0.009)	(0.011)	(0.009)
Task 10	−0.049	0.039	−0.054	0.041
	(0.013)	(0.014)	(0.013)	(0.014)

Notes: Coefficients from a log-linear cost function featuring a constant, a dummy for switching occupations, a dummy for switching sectors, and coefficients for moving in task space. The cost function is naturally scaled by the variance of shocks, ρ, and results for the constant are not presented adjusted. The first column presents the coefficients for moving up in task space and second column presents coefficients for moving down. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

Table F.6: Non-Pecuniary Benefits (η) Parameters with Transferable HC

	With Constant				No Constant			
Occupation	Manufacturing	Other Services	FIRE	Health & Educ.	Manufacturing	Other Services	FIRE	Health & Educ.
Managers	0.878	0.749	1.112		0.421	0.326	0.634	
	(0.076)	(0.067)	(0.072)		(0.047)	(0.040)	(0.043)	
Science Pro	0.655	0.986	0.780		0.194	0.511	0.289	
	(0.072)	(0.075)	(0.080)		(0.045)	(0.046)	(0.049)	
Science Assc. Pro	0.723	0.903	0.825		0.280	0.430	0.357	
	(0.071)	(0.072)	(0.075)		(0.042)	(0.043)	(0.045)	
Other Assc. Pro	0.703	0.965	0.828		0.267	0.485	0.376	
	(0.067)	(0.070)	(0.069)		(0.040)	(0.040)	(0.040)	
Clerks	0.739	1.065	0.893	1.014	0.341	0.607	0.473	0.573
	(0.064)	(0.067)	(0.062)	(0.068)	(0.039)	(0.038)	(0.037)	(0.041)
Agriculture	0.951				0.532			
	(0.070)				(0.043)			
Building Trades	0.949				0.500			
	(0.070)				(0.041)			
Metal Trades	0.816	0.900			0.383	0.470		
	(0.066)	(0.066)			(0.039)	(0.040)		
Other Crafts	0.764				0.360			
	(0.066)				(0.041)			
Plant Op	0.482				0.108			
	(0.072)				(0.048)			
Machine Op	1.090				0.603			
	(0.072)				(0.042)			
Drivers	0.699	0.935			0.302	0.483		
	(0.074)	(0.072)			(0.047)	(0.043)		
Laborers	1.017	1.024			0.557	0.573		
	(0.073)	(0.069)			(0.044)	(0.041)		
Other Pro		0.933	0.744			0.447	0.298	
		(0.071)	(0.078)			(0.041)	(0.049)	
Personal Workers		1.079		1.173		0.625		0.710
		(0.066)		(0.069)		(0.038)		(0.040)
Retail Workers		1.027				0.580		
		(0.068)				(0.040)		
Elementary Occs		1.482				0.939		
		(0.077)				(0.044)		
Customer Service			0.859				0.431	
			(0.063)				(0.037)	
Health Pro				1.042				0.513
				(0.090)				(0.056)
Teachers				1.265				0.744
				(0.074)				(0.043)
Health Assc. Pro				1.198				0.698
				(0.075)				(0.044)
Teaching Assc. Pro				1.046				0.600
				(0.069)				(0.041)

Notes: Non-pecuniary benefits to each occupation and sector cell. Blanks occur because not all occupations are present in all sectors. The Health & Education sector reflects public sector and does not include things like R&D. Data appendix contains list of industry codes in each sector. Units are proportional the unconditional sample mean income. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

Table F.7: Non-Employment (u) Parameters with Transferable HC

	With Constant	No Constant
Age	-0.012	0.000
	(0.004)	(0.003)
$\text{Age}^2(\times 1000)$	0.000	0.000
	(0.000)	(0.000)
Type 1	0.646	0.331
	(0.048)	(0.029)
Type 2	0.543	0.225
	(0.086)	(0.063)
Type 3	0.871	0.383
	(0.097)	(0.068)
Type 4	-0.038	-0.201
	(0.085)	(0.063)
Type 5	0.670	0.221
	(0.097)	(0.068)

Notes: Coefficients from a quadratic specification for the virtual value of non-employment. Types refer to estimates of unobservable heterogeneity across workers. Low skilled workers are either of type 1 or 2; Medium skilled workers are of type 3 or 4; High skilled workers are of type 5 or 6. Type 6 coefficients all normalized to 0. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

G Inference with Aggregate Shocks

This appendix presents an alternative estimator, based off of Altuğ and Miller (1998)'s work, that allows for $\sqrt{N}$ consistent estimation of the second stage structural parameters in the presence of aggregate shocks, where N is the number of individuals. In the first subsection I review the problem with forecast errors, and list out key assumptions on the time series of processes of skill prices that will be necessary for the estimator to work. After that I introduce the estimator, present the results and contrast them with the estimator from the main text. In the final section I describe the connection between the Altuğ & Miller-inspired estimator and that from the main text. Both estimators are based on regressing differentials in flow payoffs on conditional choice probabilities, but place different weights on different career paths.

G.1 Aggregate Shocks

Recall that for a worker in state ω at time t, the aggregate state is the vector of skill prices $\{w_{jt}\}_{j\in\mathcal{O}}$. This implies that workers are forecasting all future skill prices. This is a complicated forecasting problem as there are 40 prices in the model. A strength of the CCP estimator is that it does not require one to fully parametrize how beliefs are formed in the model. As a reminder of how this is accomplished, the main text's estimating equation relies on an occupation-triple transition $j \to j'$ and a second transition, $j' \to j''$, conditional on an initial state, ω. With such transitions in mind, one can form a non-linear regression equation of the form:

$$Y_{\omega jj'j''t} = \tilde{C}(j, j', j'', \omega; \Gamma) + 1/\rho(w_{\omega j't} - w_{\omega jt}) + u_{\omega jj'j''t+1} \tag{G.1}$$

which is a rewriting of equation (17) from the main text. Here Y is a function of estimated choice probabilities, $\tilde{C}$ is a function of switching costs, and u is the regression error–containing both measurement error, m, and forecast errors, ζ. While u is not correlated with any contemporaneous variables, one faces the problem that the u's are correlated across ω in the same period because all workers are forecasting the same set of wages. While the above estimator is consistent, inference that ignores within-period correlation in u may lead to smaller than justified confidence bands. For example, block bootstrapping individuals will appropriately deal with measurement error in choice probabilities, but not with within-period correlation in forecast errors. One solution would be to bootstrap on time, thus clustering by period in the second stage.[22] But this procedure may yield poor results if the number of clusters (i.e., the panel length) is small.

The estimator introduced in this appendix is based on a resampling scheme that fixes workers observed choices in each period t but draws an arbitrary future aggregate state for each worker. The key insight, outlined more formally below, is that as long as workers have rational expectations, *any* realized aggregate shock in future periods can be used if one knows the conditional choice probabilities associated with moving onward from that shock. There is no reason to use the actual realized shock at $t+1$. However, operationalizing this requires placing assumptions on the time series process that governs aggregate shocks:

Assumption 1 *The stochastic process governing wage differentials across occupations follows a stationary, first order Markov process. In particular,*

$$E_t(V_{t+1}(\omega, o) - V_{t+1}(\omega, o')|w_t, w_{t-q}, \omega_t, \omega_{t-q}) = E(V_{t+1}(\omega, o) - V_{t+1}(\omega, o')|w_t, \omega_t) \tag{G.2}$$

for any choice of q.

The right hand side of (G.2) both drops the dependence of the expectation on calendar time after controlling for covariates, and also removes any lags in the individual or aggregate state. This is slightly weaker than the assumption that wages themselves follow a stationary Markov process, as only wage differentials enter the estimation. In particular, non-stationary, or more backward looking, aggregate productivity shocks that affect all wages *uniformly* are allowed. What matters is that economy-wide shocks are such that the difference in wages (in levels) still follow a stationary Markov process.

[22] In the main text, I include these standard errors in brackets for the second stage parameters.

By way of comparison, the the CCP estimator presented in the main text *can identify* time-invariant model primitives even in the presence of non-stationary shocks.[23] This is because the renewal actions framework only requires forecasting one period forward.[24] Nevertheless, as will become apparent, the estimator below mandates stationarity.

G.2 An Estimator Based on Randomizing Aggregate Shocks

Instead of considering a particular state, ω, consider the panel data on individuals over time. In lieu of ω, I change notation now to index workers by i with the understanding that (i, t) maps to a particular state. Recall that the first stage estimator attached to each worker a probability of being an unobserved type, k, and so in the second stage, the distribution of unobserved heterogeneity is treated as observed. Hence, the data vector for individual, i in period t is given by their (1) current occupation, (2) current state variable (age education, type, tenure), (3) lagged occupation, (4) lagged state, and (5) a weight attached to being type k. There are k observations for each individual i, t, each weighted by the probability of being that type.

Since identification only depends on switchers, we can consider only the subset of data such that the current and lagged occupation are different. From assumption 1, the relevant wage for workers is not their *realized* income, but their expectation of the income uncontaminated by on-the-job productivity shocks. Hence, for each individual I construct their expected income at the time of their decision. I can also construct this income for any potential choice that the worker *could have made*, including the choice to stay in their current occupation. Hence, for each individual i at time t I construct the conditional choice probability regression as before:

$$\log(\pi_{ij'jt}/\pi_{ijjt}) = f(\omega_{it})C_{jj'} + 1/\rho(w_{ij't} - w_{ijt}) + \beta E_t[V_{j't'}(\omega_{it'}) - V_{jt'}(\omega_{it'})]$$

where j is the worker's lagged occupation, j' is their current occupation, and t' refers to the next period. I do not use $t + 1$ to avoid confusing the next period with calendar time. As before, the problem with this regression is that the difference in continuation values is unobserved. However, we may use the fact that V is a function of choice probabilities to rewrite this as,

$$\log(\pi_{ij'jt}/\pi_{ijjt}) = \tilde{C}(\omega_{it}, j, j', j''; \Gamma) + 1/\rho(w_{ij't} - w_{ijt}) + \beta E_t \left[\log\left(\frac{\pi_{j'j''t'}}{\pi_{jj''t'}}\right)\right] \tag{G.3}$$

which bears close resemblance to equation (16) in the text. Altuğ and Miller's insight is that one does *not* need to use the realized shocks at $t+1$ when substituting in for the expectational term. Instead, any realized aggregate shock in the subsequent period can be used as long as the choice probabilities associated with such an outcome are known. In particular, *if* one could observe the realized CCPs at calendar time $t + 1$ if the aggregate shock led to the prices observed at $s \neq t + 1$ then one could plug those into equation (G.3) to yield an estimable equation:

$$\log(\pi_{ij'jt}/\pi_{ijjt}) = \tilde{C}(\omega_{it}, j, j', j''; \Gamma) + 1/\rho(w_{ij't} - w_{ijt}) + \beta \log\left(\frac{\pi_{j'j''s;t}}{\pi_{jj''s;t}}\right) + u_{istjj'j''} \tag{G.4}$$

which plugs in π as if the aggregate state had progressed from where it was at time t to where it was in period s. The dependence on t on the π terms on the RHS is to draw attention to the fact that these are

[23] An example of a primitive that is time varying and thus could not be identified would be aggregate shocks to the non-pecuniary value of occupations. If the values changed then one would need to assume either (1) that the process governing their change was known and deterministic (as Dix-Carneiro (2014) does with technological change) or (2) followed a stationary process, independent of any other shocks or policy changes. If one made either of these assumptions, such shocks could also be identified from the estimator presented here.

[24] Compare this to full solution methods which requires the forecasting process on wages for any time $t+S$ to be estimable from the observed time series. For a discussion of non-stationarity and finite dependence that covers far more ground than the one-step-ahead, single agent dynamic problem in this paper, see Arcidiacono and Miller (2018).

CCPs one would observe if the aggregate state moved on the path $t \to s$. In order to exploit this insight, notice that since the model has been estimated at every possible t and ω, many possible aggregate states *have been observed*. However, for rational expectations to work when setting t' to some arbitrary date, s, in lieu of $t+1$, the crucial Markov and stationarity assumptions come into play. To see why, realize that plugging in a value of s for t' can only be done if one knows the CCPs with aggregate shocks $t' = s$. This is tantamount to the CCPs being independent of the previous aggregate state's realization, and so there being no dependence on t conditional on s in estimating CCPs. This is best understood with an example. Fix the period to be $t = 5$ and ask what would happen if instead of the world unfolding at $t = 6$ as it did, it instead looked exactly like it did at $t = 3$. If the process governing states is stationary Markov, then the world has essentially reset itself to $t = 3$ and workers act as they would act at the observed $t = 3$. On the other hand, if this were not the case, than observing the shock at $t = 3$ later in the sample may be an aberrant shock and the worker's forecast of the future may very well be different. For example, if there are occupational specific trends in the wages then observing a particular deviation from that trend would *not* lead to the same set of choices probabilities as when the shock seemed like a smaller deviation from trend. For the rest of this appendix, I will treat shocks as stationarity in order to do inference robust to aggregate shocks.

The final estimating equation comes from plugging in realized values of π into equation (G.3) and using the Markov assumption to drop the dependence on t in the estimated CCPs for aggregate state associated to s:

$$Y_{ijj'j''tt'} = \tilde{C}(j, j', j'', \omega_{it}; \Gamma) + 1/\rho(w_{ij't} - w_{ijt}) + u_{ijj'j''t'} \tag{G.5}$$

Armed with the fact that we can sample any aggregate shock for t', we can implement the estimator. Two choices have to be made: for an individual, (i, t), the econometrician must choose a renewal action j'' and a shock for the next period, s'. The first choice is novel to the environment of renewal actions. There are in principle many choices of j'' and one could construct many synthetic histories for individual (i, t). The actual choice is irrelevant as long as it fulfills the renewal action conditions. I use the most popular option for j' to switch to *unconditionally* to be j''.[25] The choice of s is where the next insight from Altuğ & Miller comes into play. They point out that if one draws s randomly for each individual then the aggregate shocks in regression (G.5) will be uncorrelated across individuals.[26] In particular, consider a random variable t'_{it} that is uniform on the integer set[27] $\{1, ..., T\}$. Then for an individual indexed by i, t, j, j', j'' the aggregate shock in the synthetic next period will be,

$$\sum_{s=1}^{T} u_{\omega(i),s,j,j',j''} \delta_{s=t'} \tag{G.6}$$

where δ is an indicator function and u collects all flow payoff terms. If t' is drawn iid across individuals then these indicator functions will be independent across individuals and so the aggregate shocks will not be correlated across i. In order to implement the estimator, after fixed j'' and drawing an s' for each individual, I use non-linear least squares to estimate the structural parameters of the model.

G.3 Results

The results of the aggregate estimator can be found in tables (G.1)-(G.5) at the end of this appendix. There are a few differences between the main text's estimator and this new estimator. Three, in particular, stand

[25] This could leave some correlation across individuals, as even though $E(\zeta_{tj''}\zeta_{sj''}) = 0$, the set from which t and s are drawn is small. In principle, one could draw j'' as well; This was not pursued to save on computational time, and to keep the subsequent discussion about weighting as clear as possible.

[26] They estimate π nonparametrically as a function of wages and construct artificial histories for workers by perturbing w_{it} with iid Gaussian shocks. This would be infeasible in my case because I am not computing a single wage and a single value of leisure. However, I also do not need to simulate histories as far into the future as they do, so I just draw from the discrete set of observed outcomes.

[27] There is no reason for the drawing scheme to be uniform. However, it is hard to estimate what the drawing scheme would be. In Altuğ & Miller, they have a similar free parameter in determining the variance on simulated aggregate shocks. In practice, one should pick a sampling scheme outside the model.

out.

First, as expected, the standard errors are much larger using this estimator. For most parameters, they are roughly one order of magnitude larger than the standard errors reported from block bootstrapping individuals, but quite close to those that come from bootstrapping over time. The scale parameters on shocks remains significant, as do most of the parameters of the cost function. The cost function parameters are jointly significantly different from zero.

Second, the estimated value of $1/\rho$ is 2.2 in this estimate, while it is closer 1.5 in the main text. This is likely driven by the estimator's re-weighting of regression equations, discussed in the next section. This value suggests that workers are slightly more responsive to shocks than in the main text. The number is not large enough to drastically alter the speed of adjustment to shocks; however, a smaller value of ρ increases the ratio of net to gross flows in any counterfactual.

Third, the level of most variables—including costs and the estimated non-pecuniary values of each occupation—are substantially lower (negative in the case of non-pecuniary values) than in the main text. In both this estimator and in the main text, the normalization of all parameters is the non-employment state of high skilled, high absolute-advantage workers. Thus, the interpretation of this level shift is that the non-employment state of highly educated workers is seen as giving high utility relative to employment. This seems unlikely and may reflect imprecision in the AM style estimator for these parameters. Only 7% of the occupational switchers are high absolute advantage, college educated workers and this group has the smallest non-employment rate. One can see that this is truly a shift in the level of costs and other variables, versus a change in relative terms, by looking at the correlation between non-pecuniary values in the AM estimator and in the main text's estimator. The correlation coefficient is .68, and the actual values of η in either specification are plotted against each other in figure 5.

To interpret the parameters in relative terms, one can renormalize the parameters for each type, by subtracting off the value of non-employment. Then, for each type, the non-pecuniary values reflect the value of being employed in occupation o versus being non-employed, and the non-pecuniary value plus the mean income (still normalized to be mean 1) is the average value of employment. Then, one can calculate the average moving cost as a fraction of the value of being employed for each type. This exercise suggests that the costs are approximately 70% of their estimated value in the main text. Thus, the estimated costs are smaller as a fraction of employment, but are similar in magnitudes.

Ultimately, the AM estimator and the estimator of the main text differ on three dimensions. First, the standard errors are larger, suggesting that adjusting for aggregate shocks does increase uncertainty about parameter estimates. Second, the value of non-employment for high absolute advantage, college graduates, is drastically different across the two estimators–suggesting that this is an imprecisely estimated parameter. Third, workers are estimated to be more responsive to shocks in the AM estimator, and smaller moving costs are needed to rationalize worker flows. Nevertheless, most parameter estimates, especially for the remaining 5 types, are broadly similar and even the wider error bands still suggest that the key elasticity, $1/\rho$, is statistically significant, and at somewhere between 1.5 and 2.5. Indeed, the main methodological takeaway is only strengthened: workers are more responsive to fundamentals than suggested by more aggregated models. In the final subsection, a link between the AM estimator and that in the main text is established, which more clearly fleshes out the source of changes in parameter estimates.

G.4 Comparison to the Main Text's Estimator

In order to understand the connection between the estimator above and that in the main text, it is best to ignore the random sampling of t' and instead just treat it as t. The actual redrawing of t' had little impact on the *point estimates* of the second estimator, but impacted the size of the confidence bands. Hence, this random drawing was not important, empirically, from an identification standpoint. Instead, as will be shown, the key difference in the estimators is the weighting assigned to different sequences of choices.

To make things concrete, given a state, ω and a series of occupations, j, j', j'', one may construct for any t the following equation for the residual that combines terms from equation (G.1):

$$u_{\omega j j' j'' t}(\theta) = Y_{\omega j j' j'' t} - g(X_{\omega,j,j',j'',t}; \theta)$$

which is the term that appears in equation (18). Now X contains all covariates in the cost function and income differentials, and Y is a function of CCPs. The main text estimator solves the problem:

$$\min_{\theta} \sum_{\omega} \sum_{j} \sum_{j' \neq j} \sum_{j'' \neq j, j'} \sum_{t} s_{\omega j j' j'' t} u_{\omega j j' j'' t}(\theta)^2 \tag{G.7}$$

where s is the weight assigned to the possibility of observing a particular state and path of occupations. Once the CCPs and flow utility primitives are estimated for every possible such combination, one can use every possible such combination (even those unobserved) in the estimation. In the main text, s is either 0 or 1, depending on whether a point is on a pre-specified grid[28] The estimator in this appendix constructs the same residual for each individual:

$$u_{ijj'j''t}(\theta) = Y_{ijj'j''t} - g(X_{i,j,j',j'',t}; \theta)$$

which is the same as the above but for each i instead of each ω. However, from the first stage estimation, we have already binned all individuals into states, ω. Moreover, an explicit, single, choice of j'' was made for each j and j', rather than using all possible outcomes. Hence we may write the above as,

$$u_{\omega(i)jj'j''(j,j')t}(\theta) = Y_{\omega(i)jj'j''(j,j')t} - g(X_{\omega(i),j,j',j''(j,j'),t}; \theta)$$

which makes explicit that j'' is a function of j and j' and that ω is now a function of i. Summing over individuals and minimizing the squared errors leads to the the objective function for the new estimator:

$$\min_{\theta} \sum_{i} \sum_{j} \sum_{j' \neq j} \sum_{t} u^2_{\omega(i)jj'j''(j,j')t}(\theta)$$

Notice that here there is no weighting, all individuals and periods enter the objective symmetrically. Collapsing by ω we have,

$$\min_{\theta} \sum_{\omega} \sum_{j} \sum_{j' \neq j} \sum_{t} N_{\omega,j,j',t} u_{\omega jj'j''(j,j')t}(\theta)^2 \tag{G.8}$$

where now $N_{\omega,j,j',t}$ is the measure of individuals (possibly 0) in state ω with transitions j to j' in period t. Comparing equations (G.7) and (G.8), the latter places *more* weight on histories and states that are more common in the data. This difference in weighting can account for much of the difference in the point estimates between the two estimators. The benefits of the weighting scheme above is clear, as it allows for inference in the presence of aggregate shocks. On the other hand, it heavily downweights events that are observed less frequently. For example, unemployment spells are rare, and only 7% of switches are into and out of unemployment (vis-a-vis permanent exit from employment). This is even lower for high skilled individuals, as discussed above. As seen in the preceding section, the virtual value of non-employment accounts for the biggest differences between the estimator in the main text and in this appendix, and is likely due to the difference in weights across estimators.

[28] Recall that I used a grid to save on computational power. The grid use all possible values of j, j', j'', as well as all possible types and period, but used an evenly spaced subset of ages and tenure.

G.5 Tables

Table G.1: Switching Elasticity and Switching Cost

	With Constant
$1/\rho$	2.224
	(0.099)
Mean C/ρ	2.055
	(0.094)

Notes: Results from regressing transitions rates on wage differentials. Mean C/ρ refers to simple mean across all cells of cost matrix with no adjustment for observed transition rates. Standard errors, in parentheses, calculated from 100 block bootstrap samples.

Table G.2: Mobility Productivity Parameters

	With Constant
Age	0.007
	(0.002)
$\text{Age}^2 (\times 1000)$	-0.326
	(0.066)
Type 1	0.345
	(0.057)
Type 2	0.599
	(0.049)
Type 3	0.341
	(0.053)
Type 4	0.472
	(0.058)
Type 5	0.174
	(0.058)

Notes: Coefficients from a log-linear inverse productivity function. Types refer to estimates of unobservable heterogeneity across workers. Low skilled workers are either of type 1 or 2; Medium skilled workers are of type 3 or 4; High skilled workers are of type 5 or 6. Type 6 coefficients all normalized to 0. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

Table G.3: Mobility Cost (Γ) Parameters

	Up Tasking	Down Tasking
Constant	-2.227	
	(0.799)	
Occ. Dummy	0.310	
	(0.013)	
Sec. Dummy	2.086	
	(0.788)	
Task 1	0.065	-0.028
	(0.007)	(0.013)
Task 2	0.077	0.082
	(0.011)	(0.010)
Task 3	0.079	0.011
	(0.022)	(0.022)
Task 4	0.153	-0.225
	(0.024)	(0.033)
Task 5	0.151	-0.096
	(0.026)	(0.032)
Task 6	-0.062	0.056
	(0.027)	(0.032)
Task 7	0.061	0.068
	(0.052)	(0.032)
Task 8	0.087	-0.189
	(0.026)	(0.036)
Task 9	0.130	0.046
	(0.023)	(0.035)
Task 10	0.021	0.057
	(0.032)	(0.044)

Notes: Coefficients from a log-linear cost function featuring a constant, a dummy for switching occupations, a dummy for switching sectors, and coefficients for moving in task space. The cost function is naturally scaled by the variance of shocks, ρ, and results for the constant are not presented adjusted. The first column presents the coefficients for moving up in task space and second column presents coefficients for moving down. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

Table G.4: Non-Pecuniary Benefits (η) Parameters

	Sector			
Occupation	Manufacturing	Other Services	FIRE	Health & Educ.
Managers	−2.159	−2.137	−1.842	
	(0.051)	(0.053)	(0.051)	
Science Professional	−2.391	−2.241	−2.155	
	(0.059)	(0.055)	(0.047)	
Science Assc. Professional	−2.210	−2.139	−2.119	
	(0.053)	(0.054)	(0.053)	
Other Assc. Professional	−2.141	−1.878	−1.902	
	(0.055)	(0.040)	(0.042)	
Clerks	−2.084	−1.832	−1.766	−2.288
	(0.054)	(0.046)	(0.052)	(0.066)
Agriculture	−2.000			
	(0.066)			
Building Trades	−1.922			
	(0.046)			
Metal Trades	−1.907	−1.938		
	(0.046)	(0.063)		
Other Crafts	−2.180			
	(0.059)			
Plant Operator	−2.500			
	(0.076)			
Machine Operator	−1.646			
	(0.038)			
Drivers	−2.003	−1.889		
	(0.063)	(0.049)		
Laborers	−1.751	−1.826		
	(0.042)	(0.051)		
Other Professional		−2.181	−2.483	
		(0.058)	(0.071)	
Personal Workers		−1.862		−1.251
		(0.052)		(0.031)
Retail Workers		−1.569		
		(0.039)		
Elementary Occupations		−1.308		
		(0.035)		
Customer Service			−1.823	
			(0.058)	
Health Professional				−2.103
				(0.079)
Teachers				−1.741
				(0.042)
Health Assc. Professional				−1.729
				(0.052)
Teaching Assc. Professional				−1.641
				(0.051)

Notes: Non-pecuniary benefits to each occupation and sector cell. Blanks occur because not all occupations are present in all sectors. The Health & Education sector reflects public sector and does not include things like R&D. Data appendix contains list of industry codes in each sector. Units are proportional the unconditional sample mean income. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

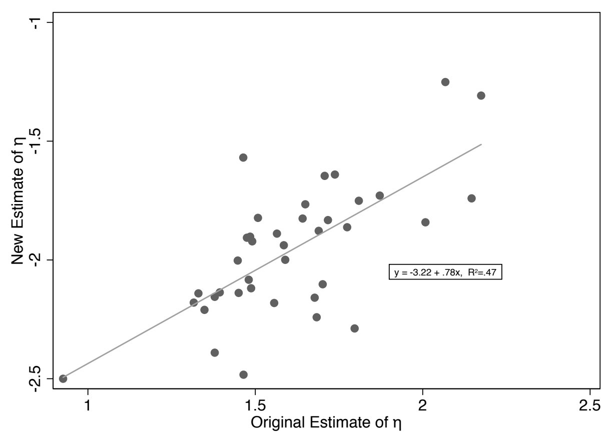
Table G.5: Non-Employment (u) Parameters

Age	0.013
	(0.004)
Age^2($\times$1000)	0.000
	(0.000)
Type 1	-0.684
	(0.026)
Type 2	-0.365
	(0.065)
Type 3	-0.936
	(0.063)
Type 4	-0.141
	(0.064)
Type 5	-1.036
	(0.071)

Notes: Coefficients from a quadratic specification for the virtual value of non-employment. Types refer to estimates of unobservable heterogeneity across workers. Low skilled workers are either of type 1 or 2; Medium skilled workers are of type 3 or 4; High skilled workers are of type 5 or 6. Type 6 coefficients all normalized to 0. Standard errors are in parentheses and based on 100 block-bootstrap samples of the underlying sample.

G.6 Figures

Figure 5: Relative and Level Changes in η Across Specifications

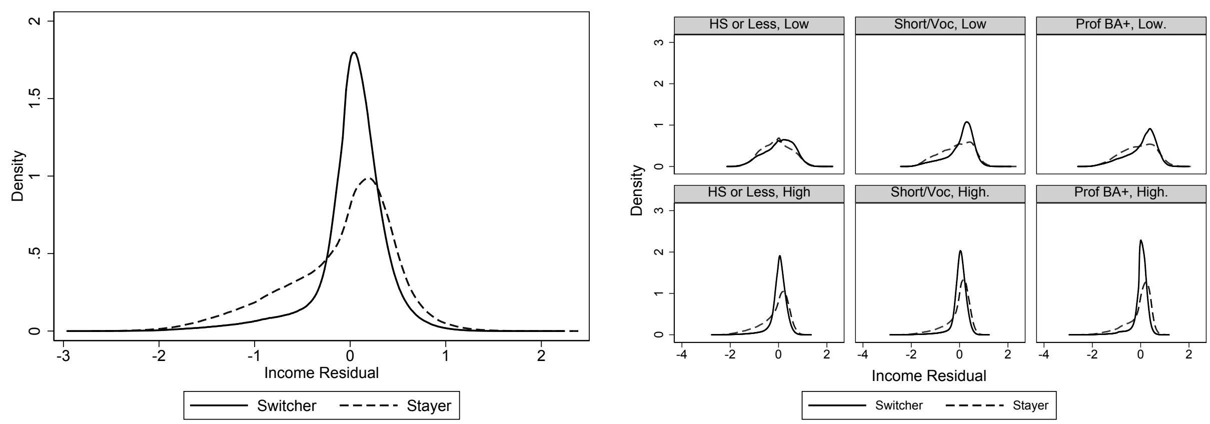


H Bias Under Different Information Assumptions

As noted by Dickstein and Morales (2018), assumptions about the information set of individuals can have complicated effects on parameter estimates in a discrete choice model. Per assumption 1 in the main text, workers have complete information on skill prices at time t when deciding where to work at t, while they have no information on their idiosyncratic productivity in each occupation, ς. The goal of this appendix is to explore the potential for bias if this assumption is violated. To do so I keep the basic structure on ς; in particular that $\varsigma \sim \mathcal{N}(0, I)$ and is independent across individuals and time. However, I consider two different assumptions on what the worker knows. At one extreme I consider the case that the worker knows ς at the time of making their decision. At the other extreme, I consider the case that the worker has no information on ς or skill prices, and can only forecast with $t-1$ information.

Figure 6: Distribution of Log Income Residuals

(a) Dispersion Across Occupations (b) Dispersion Across Sectors



Before examining the potential magnitude for bias, one can ask what diagnostic tests could be run to analyze the plausibility of the timing assumptions made in the model. One implication of the assumption on information is that the distribution of income residuals between occupation switchers and occupation stayers ought to be the same.[29] Figures 6 plots these residuals.[30] The first panel is the unconditional distribution, while the second panel is the distribution plotted for each type. The distribution of residuals for both switchers and stayers have similar peaks and are both fat tailed, but the distribution for stayers has more mass to the left of the peak.[31] This is consistent with workers having some information about ς. In particular, workers do not appear to switch if it implies a negative residual and some stayers eat negative

[29]This is not quite true: the variance of income shocks can vary by occupations and the fraction of switchers and stayers varies by occupation. Nevertheless, the distributions should be close. Moreover, the figures and discussion proceed almost identically if I scale residuals by estimated standard deviation. In this case, the residuals are truly $\mathcal{N}(0, 1)$.

[30]Another reason these residuals may differ if the assumptions on transferability of human capital are wrong. A regression of the residuals for switchers on the Mahalanobis distance between occupations has a statistically significant positive coefficient on distance. A *positive* coefficient is not consistent with mismeasured human capital, but is consistent with workers having more information: they only move to far away occupations when they get a good shock. However, the R^2 is low, and occupational distance explains only .2% of the variation in residuals.

[31]When scaled by σ, the residuals have kurtosis in excess of 3, suggesting another source of misspecification may be in assuming Gaussian errors. However, the computational benefits afforded by the Gaussian structure are hard to match with more fat tailed distributions.

residuals, suggesting frictions to switching. The distributions conditional on type look even more similar between stayers and switchers, suggesting that the wage parameters may vary by type more flexibly than I allowed for. Nevertheless, the evidence here militates for an investigation of the bias that arises from workers understanding their income residuals.

H.1 Locally Full Information

In order to pin down the ramifications of the timing assumption on wages, consider a model where workers observe their vector of productivity shocks in an occupation, ς. To analyze this world I make two simplifying abstractions: (1) wages are constant over time (or that there is perfect foresight with constancy eventually); (2) workers are homogeneous. In such a situation, the worker's problem is:

$$v(o, \varepsilon, \varsigma) = \max_{o'} -\tilde{C}_{oo'} + \rho\varepsilon_{o'} + \tilde{w}_{o'} e^{\sigma_{o'}\varsigma_{o'}} + \beta\tilde{V}(o') \tag{H.1}$$

where $\sigma_{o'}$ is the variance of the shocks and $\varsigma_{o'}$ is a standard Gaussian. I have also collected occupation fixed effects directly into the cost function. In this case $\tilde{V}$ refers to the expectation over the two idiosyncratic shocks. The model's solution *conditional* on a realization of ς is just as in the main text. However, now to aggregate across individuals one needs to integrate out the wage shock:

$$\begin{aligned}\tilde{V}(o) =& \rho\gamma + \rho \int \log\left(\sum_{o''} \exp\left(\frac{-\tilde{C}_{oo''} + \tilde{w}_{o''} e^{\sigma_{o''}\varsigma_{o''}} + \beta\tilde{V}(o'')}{\rho}\right)\right) dF(\varsigma)\\ \pi(o'|o) =& \int \frac{\exp\left(\frac{-\tilde{C}_{oo'} + \tilde{w}_{o'} e^{\sigma_{o'}\varsigma_{o'}} + \beta\tilde{V}(o')}{\rho}\right)}{\sum_{o''} \exp\left(\frac{-\tilde{C}_{oo'} + \tilde{w}_{o''} e^{\sigma_{o''}\varsigma_{o''}} + \beta\tilde{V}(o'')}{\rho}\right)} dF(\varsigma)\end{aligned}$$

In order to economize on notation, I will momentarily make the following substitutions:

$$\begin{aligned}C. =& \tilde{C}./\rho\\ V =& \tilde{V}/\rho - \gamma/(1-\beta)\\ w'_o =& \tilde{w}'_o/\rho\end{aligned}$$

Moreover, it will be cleaner to collect all flow payoff terms into a single term, $u_{o'o}(\varsigma)$. Having made these substitutions, the worker's value function and transition probability can be written as:

$$V(o;\sigma) = \int \log\left(\sum_{o''} \exp\left(u_{o''o}(\varsigma) + \beta V(o'')\right)\right) dF(\varsigma) \tag{H.2}$$

$$\pi(o'|o;\sigma) = \int \frac{\exp\left(u_{o'o}(\varsigma) + \beta V(o)\right)}{\sum_{o''} \exp\left(u_{o''o}(\varsigma) + \beta V(o'')\right)} dF(\varsigma) \tag{H.3}$$

If one could bring the integral inside of these expressions, then equations (H.2) and (H.3), would be the Bellman equation and formula on transition probabilities from the main text. However, in order to make progress in the current setting, I carry out a second order expansion around $\boldsymbol{\sigma} = 0$. Hence, this can be thought of as analyzing bias when σ is local to 0, or a local to 0 approximation to full information. First, I evaluate the functions at $\boldsymbol{\sigma}$, then take the first derivatives and evaluate at $\boldsymbol{\sigma} = 0$ and finish with the second derivative. To streamline analysis, it makes sense to focus on the derivative of the flow probability conditional on ς:

$$\pi(o'|o, \varsigma) = \frac{\exp\left(u_{o'o}(\varsigma) + \beta V(o)\right)}{\sum_{o''} \exp\left(u_{o''o}(\varsigma) + \beta V(o'')\right)}$$

The first step of the expansion is evaluation at 0:

$$u_{o'o}(\varsigma;\sigma=0) = -C_{o'o} + w_{o'} \tag{H.4}$$

$$V(o;\sigma=0) = \log\left(\sum_{o''}\exp\left(u_{o''o} + \beta V(o'')\right)\right) \tag{H.5}$$

$$\pi(o'|o;\sigma=0) = \frac{\exp\left(u_{o'o} + \beta V(o')\right)}{\sum_{o''}\exp\left(u_{o''o} + \beta V(o'')\right)} \tag{H.6}$$

In words, setting $\sigma = 0$ completely removes the shock from the aggregation. This is now almost exactly the set of equations in the main text; the key difference is that this removes even the expectation of the wage shock from the flow utility.

H.1.1 First Derivatives

Turning to first derivatives, I begin with the derivative of the flow:

$$\frac{\partial u_{o''o}}{\partial \sigma_{o'}} = \begin{cases} w_{o'}e^{\sigma_{o'}\varsigma_{o'}}\varsigma_{o'} & \text{if } o''=o' \\ 0 & \text{else} \end{cases} \tag{H.7}$$

For the Bellman equation, we use the Leibniz rule to bring the derivative inside the integral and define the derivative implicitly:

$$\begin{aligned}
\frac{\partial V(o)}{\partial \sigma_{o'}} &= \int\sum_{o''}\pi(o''|o,\varsigma)\times\left[\frac{\partial u_{o''o}}{\partial\sigma_{o'}} + \beta\frac{\partial V(o'')}{\partial\sigma_{o'}}\right]dF(\varsigma) \\
&= \sum_{o''}\int\pi(o''|o,\varsigma)\frac{\partial u_{o''o}}{\partial\sigma_{o'}}dF(\varsigma) + \beta\sum_{o''}\frac{\partial V(o'')}{\partial\sigma_{o'}}\int\pi(o''|o,\varsigma)dF(\varsigma) \\
&= \int\pi(o'|o,\varsigma)\frac{\partial u_{o'o}}{\partial\sigma_{o'}}dF(\varsigma) + \beta\sum_{o''}\frac{\partial V(o'')}{\partial\sigma_{o'}}\int\pi(o''|o,\varsigma)dF(\varsigma)
\end{aligned} \tag{H.8}$$

where the second line uses the fact that $V(o)$ does not depend on ς and so the derivative w/r/t σ is a constant, and the third line uses (H.7) to remove all off diagonal terms. Finally, for the conditional probability function:

$$\frac{\partial\pi(o''|o,\varsigma)}{\partial\sigma_{o'}} = \pi(o''|o,\varsigma)\times\left[\frac{\partial u_{o''o}}{\partial\sigma_{o'}} + \beta\frac{\partial V(o'')}{\partial\sigma_{o'}}\right] - \pi(o''|o,\varsigma)\times\left(\sum_{o'''}\pi(o'''|o,\varsigma)\times\left[\frac{\partial u_{o'''o}}{\partial\sigma_{o'}} + \beta\frac{\partial V(o''')}{\partial\sigma_{o'}}\right]\right) \tag{H.9}$$

While these expressions appear complicated, they simplify substantially when evaluated at $\boldsymbol{\sigma} = 0$. Starting with equation (H.7) :

$$\left.\frac{\partial u_{o''o}}{\partial\sigma_{o'}}\right|_{\sigma=0} = \begin{cases} w_{o'}\varsigma_{o'} & \text{if } o''=o' \\ 0 & \text{else} \end{cases}$$

An immediate implication of the above result is:

$$E\left(\left.\frac{\partial u_{o''o}}{\partial\sigma_{o'}}\right|_{\sigma=0}\right) = 0 \ \ \forall o'' \tag{H.10}$$

We will exploit the result in (H.10) extensively, along with the fact that equation (H.6) implies that the transition probabilities can be pulled out of all integrals. Turning to evaluation of (H.8):

$$\left.\frac{\partial V(o)}{\partial \sigma_{o'}}\right|_{\sigma=0} = \pi(o'|o)E\left(\frac{\partial u_{o''o}}{\partial \sigma_{o'}}\right) + \beta \sum_{o''} \frac{\partial V(o'')}{\partial \sigma_{o'}} \pi(o''|o)$$
$$= \beta \sum_{o''} \pi(o''|o) \times \frac{\partial V(o'')}{\partial \sigma_{o'}}$$

In matrix form, $D_{\sigma_{o'}} V = \beta \Pi D_{\sigma_{o'}} V$ where Π is the transition matrix on states. Since the row sums of $\beta\Pi$ are strictly bounded by 1, there is no unit eigenvalue to this matrix and the only solution is the next key result:

$$\left.\frac{\partial V(o)}{\partial \sigma_{o'}}\right|_{\sigma=0} = 0 \quad \forall o, o' \tag{H.11}$$

Next, turning to the conditional probability:

$$\left.\frac{\partial \pi(o''|o,\varsigma)}{\partial \sigma_{o'}}\right|_{\sigma=0} = \pi(o''|o,\varsigma) \times \frac{\partial u_{o''o}}{\partial \sigma_{o'}} - \pi(o''|o,\varsigma)\pi(o'|o,\varsigma) \times \frac{\partial u_{o'o}}{\partial \sigma_{o'}}$$

which uses the derivative of u and (H.5). Plugging in (H.7) once more yields the next result:

$$\left.\frac{\partial \pi(o''|o,\varsigma)}{\partial \sigma_{o'}}\right|_{\sigma=0} = \begin{cases} \pi(o'|o)(1-\pi(o'|o))\frac{\partial u_{o'o}}{\partial \sigma_{o'}} & \text{if } o''=o' \\ -\pi(o'|o)\pi(o''|o)\frac{\partial u_{o'o}}{\partial \sigma_{o'}} & \text{else} \end{cases} \tag{H.12}$$

Notice that the derivative of the transition probabilities is a function of transition probabilities and derivatives of flow payoffs—this fractal property of the derivative will be particularly useful in taking second derivatives. Continuing, integrating over this function yields the derivative of the probability of moving:

$$\left.\frac{\partial \pi(o''|o)}{\partial \sigma_{o'}}\right|_{\sigma=0} = 0 \quad \forall o', o'' \tag{H.13}$$

These first set of results simply point out that since the unobserved productivity shock is mean 0, there is no first order bias from the worker knowing the value of this shock.

H.1.2 Second Derivatives

Now we can turn to second and cross-derivatives. First for flow payoffs,

$$\frac{\partial^2 u_{o''o}}{\partial \sigma_{o'}^2} = \begin{cases} w_{o'} e^{\sigma_{o'}\varsigma_{o'}} \varsigma_{o'}^2 & \text{if } o''=o' \\ 0 & \text{else} \end{cases} \tag{H.14}$$

$$\frac{\partial u_{o''o}}{\partial \sigma_{o'} \partial \sigma_{o'''}} = 0 \quad \forall o''' \neq o' \tag{H.15}$$

Next for the Bellman equation, we can apply the chain rule to (H.8):

$$\frac{\partial^2 V(o)}{\partial \sigma_{o'}^2} = \int \left[\frac{\partial \pi(o'|o,\varsigma)}{\partial \sigma_{o'}} \frac{\partial u_{o'o}}{\partial \sigma_{o'}} + \pi(o'|o,\varsigma) \frac{\partial^2 u_{o'o}}{\partial \sigma_{o'}^2}\right] dF(\varsigma) + \beta \sum_{o''} \left[\frac{\partial^2 V(o'')}{\partial \sigma_{o'}^2} \pi(o''|o) + \frac{\partial V(o'')}{\partial \sigma_{o'}} \frac{\partial \pi(o''|o)}{\partial \sigma_{o'}}\right]$$

Since many of the above terms will either be equal to zero or integrate to zero when evaluated $\sigma = 0$, it will be notationally convenient to keep the expression above in terms of derivatives of π. With this in mind and rearranging things slightly:

$$\begin{aligned}\frac{\partial^2 V(o)}{\partial \sigma_{o'}^2} &= \int \frac{\partial \pi(o'|o,\varsigma)}{\partial \sigma_{o'}} \frac{\partial u_{o'o}}{\partial \sigma_{o'}} dF(\varsigma) + \int \pi(o'|o,\varsigma) \frac{\partial^2 u_{o'o}}{\partial \sigma_{o'}^2} dF(\varsigma) \\ &+ \beta \sum_{o''} \frac{\partial^2 V(o'')}{\partial \sigma_{o'}^2} \pi(o''|o) + \beta \sum_{o''} \frac{\partial V(o'')}{\partial \sigma_{o'}} \frac{\partial \pi(o''|o)}{\partial \sigma_{o'}}\end{aligned} \tag{H.16}$$

For the cross derivative we can perform a similar operation:

$$\begin{aligned}\frac{\partial V(o)}{\partial \sigma_{o'''} \partial \sigma_{o'}} &= \int \frac{\partial \pi(o'|o,\varsigma)}{\partial \sigma_{o'''}} \frac{\partial u_{o'o}}{\partial \sigma_{o'}} dF(\varsigma) + \int \pi(o'|o,\varsigma) \frac{\partial u_{o'o}}{\partial \sigma_{o'} \partial \sigma_{o'''}} dF(\varsigma) \\ &+ \beta \sum_{o''} \frac{\partial V(o'')}{\partial \sigma_{o'} \partial \sigma_{o'''}} \pi(o''|o) + \beta \sum_{o''} \frac{\partial V(o'')}{\partial \sigma_{o'}} \frac{\partial \pi(o''|o)}{\partial \sigma_{o'''}}\end{aligned} \tag{H.17}$$

Turning to the conditional probability function:

$$\begin{aligned}\frac{\partial^2 \pi(o''|o,\varsigma)}{\partial \sigma_{o'}^2} =& \frac{\partial \pi(o''|o,\varsigma)}{\partial \sigma_{o'}} \times \left[\frac{\partial u_{o''o}}{\partial \sigma_{o'}} + \beta \frac{\partial V(o'')}{\partial \sigma_{o'}}\right] + \pi(o''|o,\varsigma) \times \left[\frac{\partial^2 u_{o''o}}{\partial \sigma_{o'}^2} + \beta \frac{\partial^2 V(o'')}{\partial \sigma_{o'}^2}\right] \\ &- \frac{\partial \pi(o''|o,\varsigma)}{\partial \sigma_{o'}} \times \left(\sum_{o'''} \pi(o'''|o,\varsigma) \times \left[\frac{\partial u_{o'''o}}{\partial \sigma_{o'}} + \beta \frac{\partial V(o''')}{\partial \sigma_{o'}}\right]\right) \\ &- \pi(o''|o,\varsigma) \times \left(\sum_{o'''} \left[\frac{\partial \pi(o'''|o,\varsigma)}{\partial \sigma_{o'}} \times \left[\frac{\partial u_{o'''o}}{\partial \sigma_{o'}} + \beta \frac{\partial V(o''')}{\partial \sigma_{o'}}\right] + \pi(o'''|o,\varsigma) \times \left[\frac{\partial^2 u_{o'''o}}{\partial \sigma_{o'}^2} + \beta \frac{\partial^2 V(o''')}{\partial \sigma_{o'}^2}\right]\right]\right)\end{aligned} \tag{H.18}$$

With these equations in hand we can turn to evaluating these expressions at zero. First, we list some useful integrals of derivatives of u:

$$\begin{aligned}E\left(\left.\frac{\partial^2 u_{o''o}}{\partial \sigma_{o'}^2}\right|_{\sigma=0}\right) &= \begin{cases} w_{o'} & \text{if } o'' = o' \\ 0 & \text{else}\end{cases} \\ E\left(\left.\frac{\partial u_{o''o}}{\partial \sigma_{o'} \partial \sigma_{o'''}}\right|_{\sigma=0}\right) &= 0 \quad \forall o''' \neq o' \\ E\left(\left(\left.\frac{\partial u_{o''o}}{\partial \sigma_{o'}}\right|_{\sigma=0}\right)^2\right) &= \begin{cases} w_{o'}^2 & \text{if } o'' = o' \\ 0 & \text{else}\end{cases} \\ E\left(\left.\frac{\partial u_{o''o}}{\partial \sigma_{o'}} \frac{\partial u_{o''o}}{\partial \sigma_{o'''}}\right|_{\sigma=0}\right) &= 0 \quad \forall o''', o'\end{aligned}$$

where the last line follows from an assumption of independence across ε. Turning to the Bellman equation, from (H.11) and (H.12) we can rewrite (H.16) as:

$$\begin{aligned}\frac{\partial^2 V(o)}{\partial \sigma_{o'}^2} =& \pi(o'|o)(1-\pi(o'|o)) \int \left(\frac{\partial u_{o'o}}{\partial \sigma_{o'}}\right)^2 dF(\varsigma) + \pi(o'|o) \int \frac{\partial^2 u_{o'o}}{\partial \sigma_{o'}^2} dF(\varsigma) \\ &+ \beta \sum_{o''} \frac{\partial^2 V(o'')}{\partial \sigma_{o'}^2} \pi(o''|o)\end{aligned}$$

Now using the properties of the u listed above:

$$\frac{\partial^2 V(o)}{\partial \sigma_{o'}^2} = \pi(o'|o)(1-\pi(o'|o)) w_{o'}^2 + \pi(o'|o) w_{o'} + \beta \sum_{o''} \frac{\partial^2 V(o'')}{\partial \sigma_{o'}^2} \pi(o''|o) \tag{H.19}$$

This can be written in matrix form as $D_{\sigma_o}^2 V = b(\Pi, w_{o'}) + \beta \Pi D_{\sigma_o}^2 V$ where b can be inferred from (H.19).

Alternatively, $D^2_{\sigma_{o'}}V = (I-\beta\Pi)^{-1}b(\Pi, w_{o'})$. Performing similar substitutions for the cross-derivative yields:

$$\begin{aligned}\left.\frac{\partial V(o)}{\partial\sigma_{o'''}\partial\sigma_{o'}}\right|_{\sigma_{o'}=0} =& -\pi(o'|o)\pi(o'''|o)\int\frac{\partial u_{o'''o}}{\partial\sigma_{o'''}}\frac{\partial u_{o'o}}{\partial\sigma_{o'}}dF(\varsigma)+\beta\sum_{o''}\frac{\partial V(o'')}{\partial\sigma_{o'}\partial\sigma_{o'''}}\pi(o''|o)\\ =&\beta\sum_{o''}\frac{\partial V(o'')}{\partial\sigma_{o'}\partial\sigma_{o'''}}\pi(o''|o)\end{aligned}$$

where the last line follows from the rules on integrals of derivatives of u. Thus, for the same reason as before, the only solution to this equation is that the cross-derivatives are zero:

$$\left.\frac{\partial V(o)}{\partial\sigma_{o'''}\partial\sigma_{o'}}\right|_{\sigma=0}=0\quad\forall o', o''' \tag{H.20}$$

With some work, the second derivative of π can be simplified as follows:

$$\begin{aligned}\left.\frac{\partial^2\pi(o''|o,\varsigma)}{\partial\sigma^2_{o'}}\right|_{\sigma=0} =&\frac{\partial\pi(o''|o,\varsigma)}{\partial\sigma_{o'}}\times\frac{\partial u_{o''o}}{\partial\sigma_{o'}}+\pi(o''|o)\times\left[\frac{\partial^2 u_{o''o}}{\partial\sigma^2_{o'}}+\beta\frac{\partial^2 V(o'')}{\partial\sigma^2_{o'}}\right]-\frac{\partial\pi(o''|o,\varsigma)}{\partial\sigma_{o'}}\pi(o'|o)\times\frac{\partial u_{o'o}}{\partial\sigma_{o'}}\\ &-\pi(o''|o)\times\frac{\partial\pi(o'|o,\varsigma)}{\partial\sigma_{o'}}\times\frac{\partial u_{o'o}}{\partial\sigma_{o'}}-\pi(o''|o)\times\pi(o'|o)\times\frac{\partial^2 u_{o'o}}{\partial\sigma^2_{o'}}\\ &-\beta\pi(o''|o)\times\left(\sum_{o'''}\pi(o'''|o)\times\frac{\partial^2 V(o''')}{\partial\sigma^2_{o'}}\right)\end{aligned}$$

Since many of the partials of u are zero, it will be easier to focus on the cases $o'=o''$ and $o'\neq o''$ separately. First, if $o'=o''$ the above reduces as follows:

$$\begin{aligned}\frac{\partial^2\pi(o'|o,\varsigma)}{\partial\sigma^2_{o'}} =&\pi(o'|o)(1-\pi(o'|o))(1-2\pi(o'|o))\times\left(\frac{\partial u_{o'o}}{\partial\sigma_{o'}}\right)^2+\pi(o'|o)(1-\pi(o'|o))\times\frac{\partial^2 u_{o'o}}{\partial\sigma^2_{o'}}\\ &+\pi(o'|o)^2(1-\pi(o'|o))w^2_{o'}+\pi(o'|o)^2 w_{o'}\\ &+\pi(o'|o)\times\left(\beta\frac{\partial^2 V(o')}{\partial\sigma^2_{o'}}-\frac{\partial^2 V(o)}{\partial\sigma^2_{o'}}\right)\end{aligned}$$

Where I have exploited (H.19) to substitute out many terms for the derivative of the continuation value of $V(o)$. Integrating and rearranging,

$$\frac{\partial^2\pi(o'|o)}{\partial\sigma^2_{o'}}=\pi(o'|o)(1-\pi(o'|o))^2 w^2_{o'}+\pi(o'|o)w_{o'}+\pi(o'|o)\times\left(\beta\frac{\partial^2 V(o')}{\partial\sigma^2_{o'}}-\frac{\partial^2 V(o)}{\partial\sigma^2_{o'}}\right)$$

For the case that $o'\neq o''$:

$$\begin{aligned}\left.\frac{\partial^2\pi(o''|o,\varsigma)}{\partial\sigma^2_{o'}}\right|_{\sigma=0} =&-\pi(o''|o)\pi(o'|o)(1-2\pi(o'|o))\times\left(\frac{\partial u_{o'o}}{\partial\sigma_{o'}}\right)^2-\pi(o''|o)\times\pi(o'|o)\times\frac{\partial^2 u_{o'o}}{\partial\sigma^2_{o'}}\\ &+\pi(o''|o)\pi(o'|o)(1-\pi(o'|o))w^2_{o'}+\pi(o''|o)\pi(o'|o)w_{o'}\\ &+\pi(o''|o)\left(\beta\times\frac{\partial^2 V(o'')}{\partial\sigma^2_{o'}}-\frac{\partial^2 V(o)}{\partial\sigma^2_{o'}}\right)\end{aligned}$$

Performing a similar series of substitutions and integrating:

$$\left.\frac{\partial^2\pi(o''|o)}{\partial\sigma^2_{o'}}\right|_{\sigma=0}=\pi(o''|o)\pi(o'|o)^2 w^2_{o'}+\pi(o''|o)\left(\beta\times\frac{\partial^2 V(o'')}{\partial\sigma^2_{o'}}-\frac{\partial^2 V(o)}{\partial\sigma^2_{o'}}\right)$$

Combining the above leads to the final result and key ingredient to the second order expansion:

$$\left.\frac{\partial^2\pi(o''|o)}{\partial\sigma_{o'}^2}\right|_{\sigma=0} = \begin{cases}\pi(o'|o)(1-\pi(o'|o))^2 w_{o'}^2 + \pi(o'|o)w_{o'} + \pi(o'|o)\times\left(\beta\frac{\partial^2 V(o')}{\partial\sigma_{o'}^2} - \frac{\partial^2 V(o)}{\partial\sigma_{o'}^2}\right) & \text{if } o' = o'' \\ \pi(o''|o)\pi(o'|o)^2 w_{o'}^2 + \pi(o''|o)\left(\beta\times\frac{\partial^2 V(o'')}{\partial\sigma_{o'}^2} - \frac{\partial^2 V(o)}{\partial\sigma_{o'}^2}\right) & \text{else}\end{cases} \tag{H.21}$$

As a final point before moving on to the approximation, the cross-derivative of the conditional probability function, $\partial\pi(o''|o,\varsigma)/\partial\sigma_{o'}\partial\sigma_{o'''}$, will contain terms that are functions of transition probabilities multiplying cross-derivatives of u and V, as well terms multiplying $\partial u_{o''o}/\partial\sigma_{o'''}$ and $\partial u_{o''o}/\partial\sigma_{o'}$. It's already been established that such terms are either identically 0 or integrate to 0. Hence,

$$\left.\frac{\partial\pi(o''|o)}{\partial\sigma_{o'''}\partial\sigma_{o'}}\right|_{\sigma=0} = 0$$

H.1.3 A Second Order Approximation

The regression equation in the main text uses the logarithm of the transition probabilities. Hence, one can use equation (H.21) to approximate the regression equation. First taking derivatives of the log probabilities:

$$\frac{\partial\log(\pi(o''|o))}{\partial\sigma_{o'}} = \frac{1}{\pi(o''|o)}\frac{\partial\pi(o''|o)}{\partial\sigma_{o'}\partial\sigma_{o'''}} - \frac{1}{\pi(o''|o)^2}\frac{\partial\pi(o''|o)}{\partial\sigma_{o'}}\frac{\partial\pi(o''|o)}{\partial\sigma_{o'''}}$$

Per the work above, nearly all of the terms in this expression are zero. Summing over $\boldsymbol{\sigma}$ component-wise leads to the final expression:

$$\left.\log\pi(o'|o;\sigma)\right|_{\boldsymbol{\sigma}=0} \approx \log\pi(o'|o) + \frac{1}{2}\sum_{o''}\frac{1}{\pi(o'|o)}\frac{\partial^2\pi(o'|o)}{\partial\sigma_{o''}^2}\sigma_{o''}^2 \tag{H.22}$$

There are four such expressions that go into the final estimating equation. The first two are as follows:

$$\begin{aligned}\log\pi(o'|o;\sigma) \approx& \log\pi(o'|o) + \left[(1-\pi(o'|o))^2 w_{o'}^2 + w_{o'} + \left(\beta\frac{\partial^2 V(o')}{\partial\sigma_{o'}^2} - \frac{\partial^2 V(o)}{\partial\sigma_{o'}^2}\right)\right]\frac{\sigma_{o'}^2}{2} \\ &+ \left[\pi(o|o)^2 w_o^2 + \left(\beta\times\frac{\partial^2 V(o')}{\partial\sigma_o^2} - \frac{\partial^2 V(o)}{\partial\sigma_o^2}\right)\right]\frac{\sigma_o^2}{2} + \frac{1}{2}\sum_{o''\neq o',o}\frac{1}{\pi(o'|o)}\frac{\partial^2\pi(o'|o)}{\partial\sigma_{o''}^2}\sigma_{o''}^2 \\ \log\pi(o|o;\sigma) \approx& \log\pi(o|o) + \left[(1-\pi(o|o))^2 w_o^2 + w_o + \left(\beta\frac{\partial^2 V(o)}{\partial\sigma_o^2} - \frac{\partial^2 V(o)}{\partial\sigma_o^2}\right)\right]\frac{\sigma_o^2}{2} \\ &+ \left[\pi(o'|o)^2 w_{o'}^2 + \left(\beta\times\frac{\partial^2 V(o)}{\partial\sigma_{o'}^2} - \frac{\partial^2 V(o)}{\partial\sigma_{o'}^2}\right)\right]\frac{\sigma_{o'}^2}{2} + \frac{1}{2}\sum_{o''\neq o',o}\frac{1}{\pi(o|o)}\frac{\partial^2\pi(o|o)}{\partial\sigma_{o''}^2}\sigma_{o''}^2\end{aligned}$$

Expanding out the remaining terms in the summation leads to expressions of the following sort:

$$\frac{1}{\pi(o'|o)}\frac{\partial^2\pi(o'|o)}{\partial\sigma_{o''}^2} = \pi(o''|o)^2 w_{o''}^2 + \left(\beta\times\frac{\partial^2 V(o')}{\partial\sigma_{o''}^2} - \frac{\partial^2 V(o)}{\partial\sigma_{o''}^2}\right)$$

Notice that only one term on the right hand side depends on o'. Hence, when subtracting the expressions above most terms will cancel out. In particular, combining the above leads to the approximation for the first part of the regression equation:

$$\begin{aligned}\log\left(\frac{\pi(o'|o;\sigma)}{\pi(o|o;\sigma)}\right) \approx& \log\left(\frac{\pi(o'|o)}{\pi(o|o)}\right) + \left[(1-2\pi(o'|o))w_{o'}^2 + w_{o'}\right]\frac{\sigma_{o'}^2}{2} \\ &- \left[(1-2\pi(o|o))w_o^2 + w_o\right]\frac{\sigma_o^2}{2} + \beta\sum_{o'''}\left(\frac{\partial^2 V(o')}{\partial\sigma_{o'''}^2} - \frac{\partial^2 V(o)}{\partial\sigma_{o'''}^2}\right)\frac{\sigma_{o'''}^2}{2}\end{aligned}$$

For the second set of expressions:

$$\log \pi(o''|o';\sigma) \approx \log \pi(o''|o') + \sum_{o'''} \left[\pi(o'''|o')^2 w_{o'''}^2 + \beta \frac{\partial^2 V(o'')}{\partial \sigma_{o'''}^2} - \frac{\partial^2 V(o')}{\partial \sigma_{o'''}^2} \right] \frac{\sigma_{o'''}^2}{2}$$
$$\log \pi(o''|o;\sigma) \approx \log \pi(o''|o) + \sum_{o'''} \left[\pi(o'''|o)^2 w_{o'''}^2 + \beta \frac{\partial^2 V(o'')}{\partial \sigma_{o'''}^2} - \frac{\partial^2 V(o)}{\partial \sigma_{o'''}^2} \right] \frac{\sigma_{o'''}^2}{2}$$

Subtracting:

$$\log\left(\frac{\pi(o''|o';\sigma)}{\pi(o''|o;\sigma)}\right) = \log\left(\frac{\pi(o''|o')}{\pi(o''|o)}\right) + \sum_{o'''} \left[\pi(o'''|o')^2 - \pi(o'''|o)^2\right] \frac{(w_{o'''}\sigma_{o'''})^2}{2} - \sum_{o'''} \left[\frac{\partial^2 V(o')}{\partial \sigma_{o'''}^2} - \frac{\partial^2 V(o)}{\partial \sigma_{o'''}^2} \right] \frac{\sigma_{o'''}^2}{2}$$

This leads to the following combined approximation for the regression equation:

$$\begin{aligned}
\log\left[\left(\frac{\pi(o'|o;\sigma)}{\pi(o|o;\sigma)}\right)\left(\frac{\pi(o''|o';\sigma)}{\pi(o''|o;\sigma)}\right)^{\beta}\right] \approx{} & \log\left[\left(\frac{\pi(o'|o)}{\pi(o|o)}\right)\left(\frac{\pi(o''|o')}{\pi(o''|o)}\right)^{\beta}\right] \\
& + \left[(1-2\pi(o'|o))w_{o'}^2 + w_{o'}\right]\frac{\sigma_{o'}^2}{2} - \left[(1-2\pi(o|o))w_o^2 + w_o\right]\frac{\sigma_o^2}{2} \\
& + \beta\sum_{o'''}\left[\pi(o'''|o')^2 - \pi(o'''|o)^2\right]\frac{(w_{o'''}\sigma_{o'''})^2}{2}
\end{aligned}$$

The first term on the right hand side is *almost* the regression equation except that it has $\sigma = 0$, whereas I evaluate $E(we^{\varsigma\sigma}) = we^{\sigma^2/2}$. However, to finalize things, plug in the main regression equation, rearrange, and recall that the second order approximation of the expectation is given by $we^{\sigma^2/2} \approx w(1+\sigma^2/2)$:

$$\begin{aligned}
\log\left[\left(\frac{\pi(o'|o;\sigma)}{\pi(o|o;\sigma)}\right)\left(\frac{\pi(o''|o';\sigma)}{\pi(o''|o;\sigma)}\right)^{\beta}\right] \approx{} & -\left[C(o',o) + \beta[C(o'',o' - C(o'',o)] + w_{o'}(1+\sigma_{o'}^2/2) - w_o(1+\sigma_o^2/2)\right. \\
& + \left[(1-2\pi(o'|o))w_{o'}^2\right]\frac{\sigma_{o'}^2}{2} - \left[(1-2\pi(o|o))w_o^2\right]\frac{\sigma_o^2}{2} \\
& + \beta\sum_{o'''}\left[\pi(o'''|o')^2 - \pi(o'''|o)^2\right]\frac{(w_{o'''}\sigma_{o'''})^2}{2} \\
& \approx -[C(o,o') + \beta[C(o',o'' - C(o,o'')] + w_{o'}e^{\sigma_{o'}^2/2} - w_o e^{\sigma_o^2/2} \\
& + (\frac{1}{2} - \pi(o'|o))(w_{o'}\sigma_{o'})^2 - (\frac{1}{2} - \pi(o|o))(w_o\sigma_o)^2 \\
& + \frac{\beta}{2}\sum_{o'''}\left[\pi(o'''|o')^2 - \pi(o'''|o)^2\right](w_{o'''}\sigma_{o'''})^2
\end{aligned}$$

Finally, we can plug back in the substitutions that removed ρ to yield the final approximation:

$$\begin{aligned}
\log\left[\left(\frac{\pi(o'|o;\sigma)}{\pi(o|o;\sigma)}\right)\left(\frac{\pi(o''|o';\sigma)}{\pi(o''|o;\sigma)}\right)^{\beta}\right] \approx{} & -\frac{[C(o,o') + \beta[C(o',o'' - C(o,o'')]}{\rho} + \frac{\left(w_{o'}e^{\sigma_{o'}^2/2} - w_o e^{\sigma_o^2/2}\right)}{\rho} \\
& + \left(\frac{1}{2} - \pi(o'|o)\right)\left(\frac{w_{o'}\sigma_{o'}}{\rho}\right)^2 - \left(\frac{1}{2} - \pi(o|o)\right)\left(\frac{w_o\sigma_o}{\rho}\right)^2 - \\
& + \frac{\beta}{2}\sum_{o'''}\left[\pi(o'''|o')^2 - \pi(o'''|o)^2\right]\left(\frac{w_{o'''}\sigma_{o'''}}{\rho}\right)^2
\end{aligned} \tag{H.23}$$

H.1.4 Implications for Bias

Equation (H.23) shows that, to a second order, the omitted variable bias from a full information timing assumption can be broken down into two terms–one that augments the wage differentials, and an extra dynamic term. The sign of the bias is complicated and will depend on the correlation structure between the level of wages, their variance, and switching probabilities in a non-linear way. The size of the bias is more tractable. In particular, the size of the bias is mediated by the ratio of the variance in wages, σ, to the variance on moving shocks, ρ. Unfortunately, putting numbers on the bias is difficult as estimating the model when ς is observed is infeasible. However, the parameters estimated under the the model where ς is unobserved suggests that $(\sigma/\rho)^2$ will be on the order of 10^{-1}. This is because $\rho \in (.4, .7)$ depending on the estimate, while $\sigma_o < .25\ \forall o$. All of the probability terms will similarly be on the order of 10^{-1}, as will $\beta/2$. Hence, the bias terms will be on the order of 10^{-2} to 10^{-4}, while the income differentials in the regression will be on the order of 10^0. This suggests that the bias may not be large.

The above discussion is not a statement about Roy models in general, but a conclusion that depends on the particular data at hand. If the residual wage variance were larger than the variance on moving cost shocks, then the squared terms in (H.23) would be of the same or a *larger* order of magnitude as the level of income differentials. In this case, the potential for bias might be substantial. This could be especially problematic if the bias generates a smaller value of $1/\rho$, implying an overestimate of the shocks on moving cost and overestimating the costs of switching. The intuition behind this thinking is that, in the case where the worker had full information, whichever set of shocks has a larger variance ought to drive most worker reallocation. In the setting under study, there is evidence that residual wage variance is smaller than the variance in moving cost shocks. This appears to be true in nearly all specifications, weightings and aggregations.

To explore the potential bias further, I estimate a version of the model under the static equilibrium assumptions above. In particular, I use the time average of income and transition probabilities across years, ignoring worker heterogeneity. In this case, it is clear that the non-pecuniary benefits in the model would be multicollinear with wages as there is no time variation. Instead, we can identify ρ from cross-sectional variation using a constant and occupational distance to measure C.[32] The estimated value of $1/\rho$ in this exercise is .38, which is much smaller than the model that includes heterogeneity, both observed and unobserved. Nevertheless, this is surprisingly close to the estimator that allows for time variation, as well as the ACM estimator (see the next section). Such a small value of $1/\rho$ implies a much larger variance in shocks than the estimator from the main text. At first glance this may suggest this exercise will understate the concerns for bias, but σ is also larger in this situation.

With this simplified model calibrated, I can construct the potential bias term. Notice that this bias term is, at best, an approximation since neither ρ nor σ are properly estimated under model misspecification. Nevertheless, the orders of magnitude may be informative. I focus predominantly on the first bias term as it is clearly going to dominate the second term, and is also much easier to compute. The variance of the bias term, and its covariance with income differentials is, as predicted, two orders of magnitude smaller than the variance of the wage differentials. To interpret these numbers I construct the implied bias using an approximate omitted variable bias formula, which ignores the second order terms, occupational characteristics, and the fact that the skill prices are also incorrectly estimated if the model is misspecified. In particular, denoting income differentials by Δw and the first order terms by ΔFOT, I compute the following bias term:

$$bias \approx \frac{Cov(\Delta w, \Delta FOT)}{Var(\Delta w)}$$

To get a sense of how the bias changes under different true values of σ and ρ, table H.1 shows the estimated bias, as a percent of the true $1/\rho$, when I multiply each element of the estimated σ by $2/3, 1$ and $3/2$ and do similarly for $1/\rho$.

[32] I find that controlling for occupational characteristics is very important in disciplining ρ.

Table H.1: Potential Bias Under Full Information Assumption

	$\frac{2}{3}\times\sigma$	$1\times\sigma$	$\frac{3}{2}\times\sigma$
$\frac{2}{3}\times\frac{1}{\rho}$	0.53	1.21	2.72
$1\times\frac{1}{\rho}$	0.81	1.81	4.08
$\frac{3}{2}\times\frac{1}{\rho}$	1.21	2.72	6.12

The implied bias is positive in all cases. At the estimated values of $1/\rho$ and σ, the bias is 1.81%. Setting the variance on moving cost shocks lower (i.e., adjusting $1/\rho$ upwards), increases the bias. Nevertheless, the bias remains small as long as σ is being estimated relatively close to the truth. As expected, the bias becomes most severe precisely when the true moving cost shocks have small variance and the income shocks have large variance. This is depicted in the bottom right of the table, when the implied variance on moving cost shocks is 2/3 of its true value and the variance on income shocks is 3/2 of its estimated value. In this situation, the bias is more than 5%, which is large but remains modest. Importantly, this is more suggestive evidence that, at least in this simplified setting and given the data at hand, the bias stemming from the model's information assumption versus a full information assumption is not very large. This is consistent with the figure above since, despite some evidence of the information assumption failing, the distribution of income residuals still overlap a great deal.

H.2 No Information

Another stark assumption that one could make is that workers' do not know the skill prices when making their decision. This is similar to ACM except I still assume that workers make their decision in the same period in which they are paid. This set up is decidedly simpler than that above, and requires no complicated expansions. In particular, the aggregation still proceeds as before but the estimating equation *ought* to be:

$$\begin{aligned}\log\left[\left(\frac{\pi_t(o'|o,\omega_t)}{\pi_t(o|o,\omega_t)}\right)\left(\frac{\pi_{t+1}(o''|o',\omega'_{t+1})}{\pi_{t+1}(o''|o,\omega'_{t+1})}\right)^{\beta}\right] =& -C(o,o',o'',\omega_t,\omega'_{t+1},\omega_{t+1}) \\ &+\frac{1}{\rho}E_{t-1}(w_{o't}h(o',\omega_t)-w_{ot}h(o,\omega_t))+\zeta_{t+1}\end{aligned} \tag{H.24}$$

Rearranging we can write this as a new forecast error and realization:

$$\begin{aligned}\log\left[\left(\frac{\pi_t(o'|o,\omega_t)}{\pi_t(o|o,\omega_t)}\right)\left(\frac{\pi_{t+1}(o''|o',\omega'_{t+1})}{\pi_{t+1}(o''|o,\omega'_{t+1})}\right)^{\beta}\right] =& -C(o,o',o'',\omega_t,\omega'_{t+1},\omega_{t+1})+\frac{1}{\rho}(w_{o't}h(o',\omega_t)-w_{ot}h(o,\omega_t)) \\ &+E_{t-1}(w_{o't}h(o',\omega_t)-w_{ot}h(o,\omega_t))-(w_{o't}h(o',\omega_t)-w_{ot}h(o,\omega_t)) \\ &+\zeta_{t+1}\end{aligned} \tag{H.25}$$

Letting $-e_t$ be the forecast error, the above setting can be modeled as a case of classical measurement error. In particular,

$$(w_{o't}h(o',\omega_t)-w_{ot}h(o,\omega_t)) = E_{t-1}(w_{o't}h(o',\omega_t)-w_{ot}h(o,\omega_t))+e_t$$

where, as long as agents are rational, e_t is orthogonal to the unobserved, true, regressor. Hence, if the worker knows less than I assume, there ought to be attenuation bias in the slope term. This should also enlarge the constant term. In other words, if worker's have less information, the costs of switching will be overestimated. The economic intuition behind this result is that workers will appear to respond less to wage differentials (as they are mismeasured) and this can only be rationalized with large switching costs. Without unobserved heterogeneity, and thus no first stage, one could employ an IV estimator in the spirit of ACM to estimate the model. In particular, lagged wage differentials would be a valid instrument for wage differentials.

Since the second stage estimator is non-linear, to explore the importance of the assumption on information I estimate an augmented version of the model that is linear in the cost function terms (i.e., the constants, occupation characteristics, and worker characteristics). The results of these regressions are presented in the first two columns of table H.2, below.

Table H.2: Comparing OLS and IV Estimates of $1/\rho$

Estimation Method				
OLS	IV	ACM-OLS	ACM-IV	ACM-IVC
2.04	2.06	0.13	0.49	0.46
[0.08]	[0.07]	[0.23]	[0.39]	[0.21]

This augmented second stage yields an estimate of $1/\rho = 2.04$, which is close to the original estimate and well within the range of values estimated in the main text and the appendices. The IV estimator in this setup yields $1/\rho = 2.06$, which *is* larger, but close to the original.[33] This suggests that worker's have good information on skill prices in the period in which they move. This is not to say there is no variability in skill prices over time. Indeed, an AR(1) of income differentials on lagged differentials has an R^2 of .47. The point is that workers are as responsive to observed wages as to an instrument that is valid under either timing assumption.

This is different from the result in ACM, who found that instrumenting changed point estimates considerably. In particular, they find that instrumenting increased the point estimate on $1/\rho$ from .22 to .35. Implementing the ACM estimator on aggregated Danish data[34], leads to a similarly large disparity between OLS and IV. This can be seen in the last three columns of H.2. The third column is the estimate of $1/\rho$ under OLS in the ACM setup; the fourth column uses an IV estimator based on lags; the final column uses an IV estimator and includes occupational characteristics as control variables. Standard errors, calculated by bootstrapping on time, ignoring measurement error in transition rates, are in brackets below the estimates. While the OLS and IV estimator are far apart the error bands are large. This is likely a result of having a relatively short panel compared to ACM. Nevertheless, the OLS and IV estimates are contained in each other's confidence bands. Hence, while this data replicates the findings in ACM on aggregated data, using more data on individual characteristics seems, in this dataset at least, to alleviate some of the concerns over timing and information.

H.3 Remarks

The calculations performed above are approximate as they ignore heterogeneity and are local to zero approximations of the model in the main text. Nevertheless, all of the theoretical and numerical exercises suggest that the potential for bias coming from my timing assumptions versus two extreme alternatives is small. In the case that workers know *more* than I assume, it seems that this extra information (in the form of wage shocks) is not as important as the information contained in the moving cost shocks. In the case that workers know *less* than I assume, it seems that they are "close" to having the information I assume in the sense of small forecast errors. Combining these two observations leads to the conclusion that workers seem to have a very good idea of their wages when they are making their occupational choice if all they know is the current skill prices—these skill prices are both forecasted well and there is little residual variance afterward.

While these exercises ought to assuage much concern over the potential for bias, there are some potentially important situations that are not covered in the above. For example, workers may be heterogeneous in their ability to forecast based on their characteristics. Under such a situation, there would be a biased measure of ρ for every possible worker's state and I am measuring an average (potentially not even population-weighted) of these biased estimates. Because for some groups this might be over- or under- estimated, it is unfortunately hard to know where this average would lay. Another case that is not considered is when agents have a noisy signal of their income shocks that is correlated across occupations. For example, it might

[33] While the confidence intervals around the point estimates overlap, the IV estimator was larger than the OLS estimator in most bootstrap replications, thus the bias was positive and statistically significant. Nevertheless, it is economically small.

[34] In particular, I use twice lagged differentials and twice lagged switching probabilities as instruments for contemporaneous differentials. To be as close to ACM as possible in this exercise, I use only two-way occupation fixed effects, and not occupational characteristics. However, including these controls has little effect. The estimate for ρ is close to 2.1, which is surprisingly close to ACM's estimate from a long panel in the US.

be that workers have better information about occupations similar to their own occupation than about very different occupations. As a third example, workers may have incomplete information about their own state variables; for example, they may not know their comparative advantage in occupations they have never tried. Understanding the role of workers' information set is beyond the scope of this paper, and requires different methods than those I've pursued. Nevertheless, different assumptions on information sets that are amenable to analysis do not seem problematic; while future research, perhaps using moment inequalities or other methods, may be able to shed light on the importance of more flexible modeling of workers' information.

I Anticipation Counterfactual

I.1 Setup

In this counterfactual, the setup is exactly as described in the main text. In particular, workers face the same dynamic choice problem as before, labor supply is the same, and the model is solved with workers having perfect foresight of the economy's evolution in response to news of any future foreign price changes is revealed. However, instead of an unexpected shock at $t = 0$, the workers have perfect foresight that the shock will occur at $t = 10$. The model begins from an initial steady state, so news of the lower trade costs is still unanticipated. This setup gives workers time to adjust in preparation of the shock's onset. The actual shock is the same as that in the main text, and so I do not review the details here.

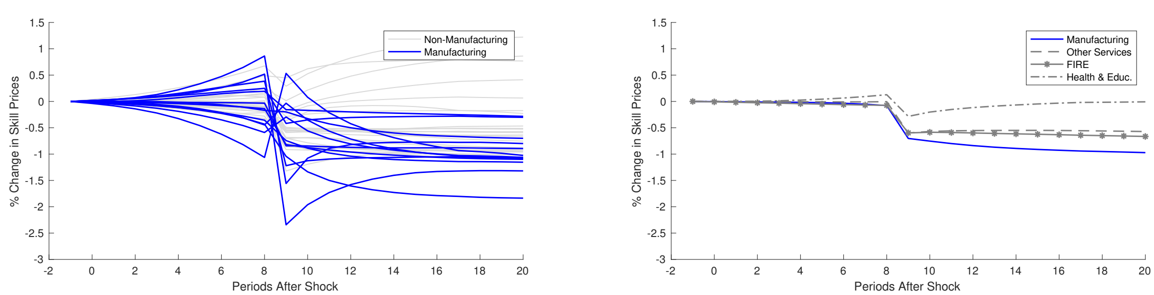
I.2 Impact on Workers with Anticipation

In discussing the impact on workers who have foreknowledge of changes in trade prices, I focus on differences between the results in the main text and in this counterfactual. Previewing the results, the importance of occupations remains a salient feature in all counterfactuals. And swings in skill prices are muted, they are still large at impact—suggesting it is the size of the foreign price change, and not its being unexpected that matters. A novel part of the new analysis is that there are now substantial differences in outcomes along the life cycle. Young workers are more able to adapt in anticipation of the shock, leading to larger gains in income.

Figure 7 reproduces the graph of the evolution of skill prices in the economy in the new counterfactual. In the first panel, which plots changes in skill prices across occupations, two distinctions stand out from the original counterfactual. First, the effect on nominal skill prices at impact is smaller, as one would expect if workers adjust in advance. In the main text, the maximum change in skill prices relative to a world without changes in import prices was 1.5%, and the minimum was -2.75%. However, when workers can adjust in anticipation, the maximum increase is only .5% while the largest drop is less than 2.5%. The reason for these muted effects is explained by the second key difference from the original model: there are anticipatory effects on skill prices. Since workers are aware that there will be a large change in skill prices at $t = 10$, some workers begin leaving occupations predestined for a drop in demand early (or choose not to enter into them). This leads to a temporary *increase* in skill prices for remaining workers before a drop after import prices decline. A mirrored phenomenon takes place in other occupations, as workers begin to enter sectors in anticipation of increased demand.

Figure 7: Dynamic Effects of an Import Price Shock on Nominal Skill Prices–Different Timing

(a) Dispersion Across Occupations (b) Dispersion Across Sectors



Adjustment costs—both fixed and in terms of human capital loss—are key in explaining these patterns in skill price adjustments. Workers who receive favorable switching cost shocks before a change in labor demand takes place are keen to exploit their good fortune. Other workers who do not receive sufficiently favorable shocks remain in their occupation. In a model without switching costs, workers would never adjust occupations before a shock took place as the continuation values of all occupations would be equalized, even if compensating differentials could sustain skill price dispersion.

In the second panel of Figure 7, one sees that qualitatively similar phenomena to those describe above occur with sectoral average skill prices. There are slight changes in skill prices before the price change takes place, in anticipation, followed by larger changes in skill prices when the price change occurs. Comparing each panel of the figure leads to the same conclusion as in the main text: the dispersion in skill price changes *within* sectors, is much larger than across sectors. In fact, these within versus between differences are even clearer when workers can anticipate price changes. For example, focusing on manufacturing in panel (b), the change in skill prices from anticipation effects is almost *zero* for the average skill price in manufacturing. However, from panel (a), it is actually clear that there are large changes in skill prices even before import competition ramps up. This difference reflects that much reallocation, both before and after the foreign price change, occurs within manufacturing.

Table I.1: Percent of Income Variance Explained

	Levels	Differences		
		Announcement	Short Run	Long Run
Occs/Sectors	65.6	94.6	60.1	47.5
Occupations Only	58.0	76.6	52.6	40.1
Sectors Only	16.4	11.3	13.9	4.6

Notes: Income variance is calculated across all individuals present, weighting by the distribution of workers in the equilibrium *with trade*. Residual variance reflects worker demographics within each economic unit. Short run is defined as the period at impact, long run is based on a steady state that is assumed to be reached at 40 periods.

In order to quantify the dispersion alluded to above, Table I.1 recreates Table 8 from the main text. There is one additional column, besides the short run and long run effects, the table contains information on the variance decomposition when news of the price change is announced. The major findings do not change from that in the main text. In particular, occupations have more explanatory power than sectors, especially in the short run. However, echoing the discussion above, the new column reveals that one's occupation and sector in the economy also matter a great deal for how their income changes *before* changes in foreign prices even occur. In fact, at announcement, one's occupation is even more important for explaining variation in income than after trade prices adjust. While interesting, it mostly buttresses the key point that much of the reallocation taking place occurs as workers rotate through different tasks, even within the same broad part of the economy.

Finally, one can quantify outcomes along different dimensions of worker heterogeneity. Here one must take a stand on which set of workers to consider—workers at the time of announcement, workers alive at the time of the price change, or another period. I look at workers alive at the time of the announcement, but focus on changes in income *after* foreign prices change. This is to make the changes in earnings more comparable to the table in the main text, with the caveat that this ignores workers preemptive actions.

Turning to heterogeneity along the skill distribution, Table I.2 plots the distribution of lifetime earnings for different sets of skills and unobserved ability.

The patterns are broadly the same as in the main text. The majority of workers who lose in terms of lifetime income are workers with a medium level of education, or workers who are of low absolute advantage. On average, 7% of workers see decreases in lifetime earnings—while larger than the original counterfactual, this masks the fact that many workers make *more* in income before the foreign price change takes place by reallocating preemptively. The median worker sees an increase in earnings of 3.55%, close to the 4.14% in the original counterfactual. Including the η terms, and giving them an interpretation as part of compensation,

Table I.2: Distributional Impact of Trade on Workers After Shock

(a) Distribution of Lifetime Earnings Differentials

	Mean	Std. Dev	Q5	Q50	Q95	% < 0
Low [L]	4.95	9.15	−0.24	4.46	16.65	5.75
Low [H]	3.58	5.18	−0.23	3.38	9.01	7.15
Med [L]	4.59	9.34	−1.73	3.94	17.40	7.69
Med [H]	3.40	4.54	−0.19	3.38	5.87	6.59
High [L]	4.34	8.57	−1.10	3.50	16.67	9.64
High [H]	3.55	4.49	−0.14	3.34	8.74	7.10
Total	3.94	6.72	−0.23	3.55	10.92	6.74

(b) Distribution of Lifetime Earnings Differentials (w/ η)

	Mean	Std. Dev	Q5	Q50	Q95	% < 0
Low [L]	1.29	4.61	−0.30	0.78	6.97	6.36
Low [H]	1.75	1.85	−0.03	1.65	4.19	5.78
Med [L]	1.77	4.08	−0.38	1.16	7.96	7.26
Med [H]	1.66	1.32	−0.02	1.72	2.64	5.11
High [L]	2.23	5.41	−0.54	1.61	6.80	9.64
High [H]	1.82	1.91	−0.07	1.79	3.54	6.02
Total	1.63	2.93	−0.05	1.55	4.50	5.89

Notes: Tables report the (100x) log difference in discounted total earnings across individuals. Results are based on simulating 100,000 individuals from the initial cohort under both the equilibrium with and without changes in trade prices. The same shocks are used in both simulations. Discounted at $\beta = .96$.

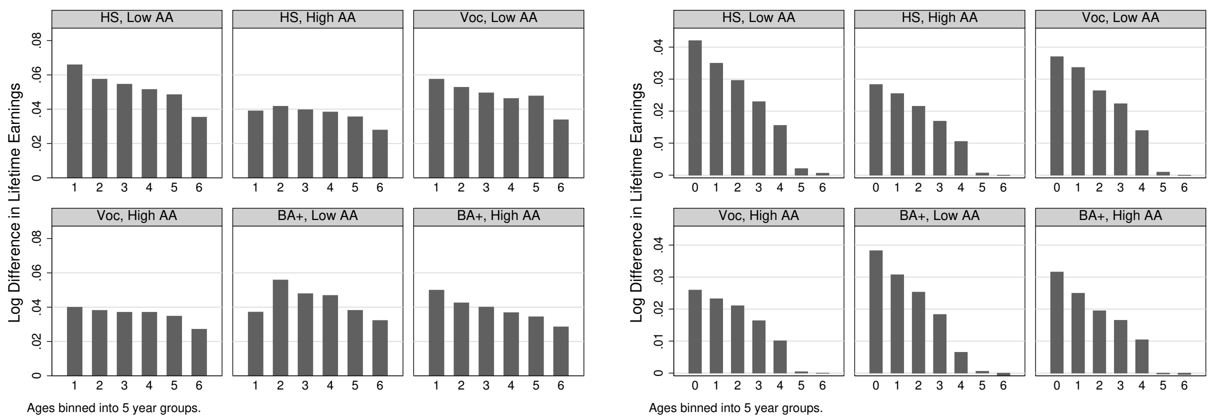
leads to smaller gains, but leads to a smaller fraction of workers losing in terms of compensation.

Thus far, there have been only small quantitative differences between a situation where workers do not foresee increased globalization and in which they do. Qualitatively, key patterns remain the same: intra-industry reallocation plays a much larger role in determining distributional outcomes than across sectors, and losses, especially large losses, appear concentrated on low absolute advantage workers with at least some education. However, there are large differences along the life cycle of workers. Intuitively, workers can anticipate the coming changes in foreign prices, but only the young are likely to face low enough adjustment costs to react. To illustrate these differences, Figure 8 recreates Figure 6 from the main text, and plots mean changes in income for different age groups. The first panel plots these differences for the cohort alive at announcement, but only plots income after the trade liberalization, while the second panel plots these differences for the full life span of workers.

In both panels, there is a downward slope in average changes in earnings across age groups. This is particularly pronounced for workers alive at the time of the announcement. The relationship is partially mechanical in the second panel, especially for the oldest workers, as younger workers are in the labor pool for a longer time after trade prices fall. However, this is not entirely mechanical, which is clear as some older workers actually see small income declines on average. Even focusing on workers only after import prices change, one can see a sharper decline in earnings growth along age than the decline in the main text's counterfactual. Both situations reflect the same phenomenon that was stated above: when there is an anticipatory reaction, it exacerbates the effects on older workers who are not positioned to react.

Figure 8: Distributional Impacts Across Age Groups

(a) Workers After Foreign Price Change (b) Worker After Announcement



Notes: Figures report the (100x) log difference in discounted total earnings across individuals. Results are based on simulating 100,000 individuals from the initial cohort under both the equilibrium with and without changes in trade prices. The same shocks are used in both simulations. Discounted at $\beta = .96$.

I.3 Conclusion

In this counterfactual, workers have ten periods to preemptively respond to news that there will be a large shift in prices. The impacts of trade liberalization across occupations remain large, despite the fact that there is clear evidence that works react in anticipation. Indeed, one's occupation at the time that an upcoming surge in import competition is announced is even more important in explaining how workers react than at the actual onset of the foreign price change. In relative terms, sectoral differences actually become much more muted as a substantial amount of preemptive reallocation ultimately occurs within sectors.

Altering how much workers can anticipate foreign price changes does not seem to matter for the distributional consequences either across one's initial occupation or across the skill distribution. However, there are clear differences in how younger versus older workers are able to incorporate this information. In particular, young workers can shift around, and stand to reap large gains from trade. Older workers, in this new settings, are seemingly hurt by the ability of the young to act; their income growth is lower, on average, than in the main text's counterfactual and is even negative for the average older worker alive at the announcement for information. These impacts across the age distribution are a clear avenue for future research. In particular, understanding better how younger versus older workers are able to exploit the slow revelation of changes in globalization may yield more precise estimates of the winners and losers from trade, and insights into why certain trade shocks seem to have large effects in the reduced form, while others do not.

References

Altuğ, Sumru and Robert A. Miller, "The Effect of Work Experience on Female Wages and Labour Supply," *Review of Economic Studies*, January 1998, *65* (1), 45–85.

Arcidiacono, Peter and Robert A. Miller, "CCP Estimation of Dynamic Models with Unobserved Heterogeneity," *Econometrica*, November 2011, *79* (6), 1823–1868.

— **and** — , "Nonstationary Dynamic Models with Finite Dependence," Mimeo 2018.

Arellano, Manuel, *Panel Data Econometrics*, Oxford University Press, 2003.

Artuç, Erhan and John McLaren, "Trade Policy and Wage Inequality: A Structural Analysis with Occupational and Sectoral Mobility," *Journal of International Economics*, 2015, *97* (2), 278–294.

Broda, Christian and David E. Weinstein, "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, May 2006, *121* (2), 541–585.

Das, Sanghamitra, Mark J. Roberts, and James R. Tybout, "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," *Econometrica*, 2007, *75* (3), 837–873.

Dhaene, Geert and Koen Jochmans, "Split-panel Jackknife Estimation of Fixed-effect Models," *The Review of Economic Studies*, 2015, *82* (3), 991–1030.

Dickstein, Michael J and Eduardo Morales, "What do Exporters Know?*," *The Quarterly Journal of Economics*, 2018, *133* (4), 1753–1801.

Dix-Carneiro, Rafael, "Trade Liberalization and Labor Market Dynamics," *Econometrica*, 2014, *82* (3), 825–885.

Feenstra, Robert C, "New Product Varieties and the Measurement of International Prices," *American Economic Review*, March 1994, *84* (1), 157–77.

Firpo, Sergio, Nicole M Fortin, and Thomas Lemieux, "Occupational tasks and changes in the wage structure," 2011.

Heckman, James and Burton Singer, "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 1984, *52* (2), 271–320.

Honoré, Bo E. and Elie Tamer, "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 2006, *74* (3), 611–629.

Ransom, Tyler, "Labor Market Frictions and Moving Costs of the Employed and Unemployed," *Mimeo*, 2018.

Rust, John, "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, September 1987, *55* (5), 999–1033.

Scott, Paul, "Dynamic Discrete Choice Estimation of Agricultural Land Use," *Mimeo*, 2014.

Silva, J. M. C. Santos and Silvana Tenreyro, "The Log of Gravity," *The Review of Economics and Statistics*, November 2006, *88* (4), 641–658.

Silva, J. M. C. Santos and Silvana Tenreyro, "poisson: Some convergence issues," *Stata Journal*, June 2011, *11* (2), 215–225.

Simonovska, Ina and Michael E. Waugh, "The elasticity of trade: Estimates and evidence," *Journal of International Economics*, 2014, *92* (1), 34–50.