ESSAYS ON INTERNATIONAL TRADE AND EDUCATION

A Dissertation in Economics
by
James Key

© 2013 James Key

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

December 2013
The dissertation of James Key was reviewed and approved* by the following:

Kala Krishna  
Professor of Economics  
Dissertation Advisor, Chair of Committee

Stephen Yeaple  
Associate Professor of Economics

Paul Grieco  
Assistant Professor of Economics

John Moran  
Associate Professor of Health Policy and Administration

Vijay Krishna  
Distinguished Professor of Economics  
Director of Graduate Studies, Department of Economics

*Signatures are on file in the Graduate School.
Abstract

In the first chapter, I develop a model in which firms vary by both labor productivity and the quality of their output, with inputs of heterogeneous quality required in the production process. The model shows how input choice is affected by a firm’s exogenously determined characteristics, and how the distribution of firms responds to changes in factor prices. In doing so, this paper extends the model of Sutton (2007), by allowing heterogeneous inputs, and a stark difference between the models is seen for low wage economies. Further, this paper examines the determinants of correlation between usage of high quality inputs and productivity, finding that unlike Kugler and Verhoogen (2012), a positive correlation need not exist for a given economy.

The second chapter develops a simple model that explains some patterns in the export behavior of firms. For example, in a standard gravity model, even when controlling for all the usual variables, the rank of GDP of a country is positively and significantly related to inward trade flows, but not outward flows. We also see that firms which expand rapidly into new markets are more likely to survive than firms which expand slowly. I assume that the demand for a firm’s product is positively correlated across markets, and that the uncertainty is reduced through entry into markets. As a result, the expansion behavior of a firm is driven not only by the expected profits net of entry costs, but also by the information value of an additional observation. Allowing for learning can lead to considerably different outcomes for markets that have similar fundamental properties, through both the volume of entry and the types of firms that enter. One consequence of this is that a reduction in trade barriers typically increases entry in some markets, but not others.

In the final chapter, I develop a model that explains how students approach the multiple choice exams for university entrance in Turkey. I estimate this model using simulated method of moments to recover the parameters. This allows us
to perform counterfactual experiments regarding the effectiveness of the system allocating students to universities. While I find substantive differences between the different gender in terms of exam behavior, these differences have very little impact on resulting allocation of students to universities.
# Table of Contents

List of Figures vii

List of Tables xi

Chapter 1
   Capabilities and Input Choice
      1.1 Introduction .............................................. 1
      1.2 The Model ................................................ 2
         1.2.1 Demand ............................................. 2
         1.2.2 Production ......................................... 3
         1.2.3 Aggregation ......................................... 4
         1.2.4 Firm Entry and Exit ................................. 4
         1.2.5 Equilibrium Analysis ............................... 6
      1.3 Heterogeneous Inputs .......................... 8
         1.3.1 Input Choice ....................................... 9
         1.3.2 Inputs and Productivity .......................... 10
         1.3.3 Elasticity of costs ............................... 12
      1.4 Trade in Intermediate Inputs ........ 13
      1.5 Conclusion ............................................. 15

Chapter 2
   Priors and Posteriors: Implications for Exporters 20
   2.1 Introduction .............................................. 20
   2.2 The Model ................................................ 24
      2.2.1 Demand ............................................. 24
      2.2.2 Production ......................................... 25
      2.2.3 Product Appeal Updating .......................... 27
# List of Figures

1.1 The cutoff cost $k^*$ represented in capability space ........................................ 6  
1.2 The change in capabilities of active firms ......................................................... 7  
1.3 Iso-input curves in $(c,u)$ space ................................................................. 10  
1.4 Cost elasticity with respect to $w$ ................................................................. 13  
1.5 Capability Thresholds ....................................................................................... 14  
1.6 Input schedule for the home country before and after opening to input trade, and the foreign country ................................................................. 15  
1.7 Cost elasticities w.r.t. $w$ when inputs are homogeneous ............................... 19  

2.1 Differ by one variable ......................................................................................... 29  
2.2 Differ by two variables ....................................................................................... 29  
2.3 Observations affect entry decisions .................................................................. 33  
2.4 Exit Thresholds in productivity/demand space .................................................. 34  
2.5 Expansion within the EU, given number of markets previously entered ........ 51  
2.6 Propensity to expand within the EU, given number of markets previously entered .................................................................................................................. 51  
2.7 World expansion, given number of markets previously entered ....................... 52  
2.8 Propensity to expand, given number of markets previously entered ............... 52  
2.9 Expansion behavior and firm age ...................................................................... 53  
2.10 Destination choice within the EU, conditional on experience ......................... 53  
2.11 World wide destination choice, conditional on experience .............................. 54  
2.12 Likelihood of entry and survival within the EU, for new firms expanding to a single market ................................................................. 54  
2.13 Likelihood of entry and survival within the EU, for new firms expanding to two markets .................................................................................................................. 55  
2.14 Likelihood of entry and survival within the EU, for new firms expanding to three markets .................................................................................................................. 55  
2.15 Likelihood of entry and survival within the EU, for firms in the 2001 cohort .................................................................................................................. 56  
2.16 Correlation between market GDP rank and various entry ranks ................... 56
3.28 Score distribution from data vs simulated distribution resulting from estimates ......................................................... 106
3.29 Score distribution from data vs simulated distribution resulting from estimates ......................................................... 106
3.30 Score distribution from data vs simulated distribution resulting from estimates ......................................................... 107
3.31 Score distribution from data vs simulated distribution resulting from estimates ......................................................... 107
3.32 Score distribution from data vs simulated distribution resulting from estimates ......................................................... 108
3.33 Score distribution from data vs simulated distribution resulting from estimates ......................................................... 108
3.34 Distributions of social science ability, as estimated by the model ................................................................. 109
3.35 Distributions of Turkish ability, as estimated by the model ................................................................. 109
3.36 Distributions of social science ability, as estimated by the model ................................................................. 110
3.37 Distributions of Turkish ability, as estimated by the model ................................................................. 110
3.38 Distributions of social science ability, as estimated by the model ................................................................. 111
3.39 Distributions of Turkish ability, as estimated by the model ................................................................. 111
3.40 The relationship between the proportion of males and ÖSS-SOZ score quantiles from the counterfactual exercises ......................................................... 112
3.41 The relationship between the proportion of first time takers and ÖSS-SOZ score quantiles from the counterfactual exercises ......................................................... 112
3.42 The relationship between social science ability and ÖSS-SOZ score quantiles from the counterfactual exercises ......................................................... 113
3.43 The relationship between Turkish ability and ÖSS-SOZ score quantiles from the counterfactual exercises ......................................................... 113
3.44 Score distribution from data vs simulated distribution from two type extended model ......................................................... 116
3.45 Score distribution from data vs simulated distribution from two type extended model ......................................................... 116
3.46 Score distribution from data vs simulated distribution from two type extended model ......................................................... 117
3.47 Score distribution from data vs simulated distribution from two type extended model ......................................................... 117
3.48 Score distribution from data vs simulated distribution from two type extended model ......................................................... 118
3.49 Score distribution from data vs simulated distribution from two type extended model ......................................................... 118
3.50 Score distribution from data vs simulated distribution from two type extended model ......................................................... 119
3.5.1 Score distribution from data vs simulated distribution from two type extended model  

119
List of Tables

2.1 Estimation results: Gravity ........................................ 39
2.2 Estimation results: Survival Probit ................................. 42
2.3 Estimation results: Expand Probit ................................. 44
2.4 Estimation results: Expansion ...................................... 45
2.5 Estimation results: Entry Correlation ............................ 48
2.6 Firm/destination survival over time, EU, conditional on firm cohort and year of entry into the destination market ....................... 57
2.7 Ordinal Logit Regression: Expansion ............................. 57

3.1 Test Weights .......................................................... 86
3.2 Summary Statistics ..................................................... 86
3.3 Question outcomes for various parameter values: probabilities of skipping, being correct, being incorrect, and the average points per question ................................................................. 101
3.4 Cutoff Estimates for First Time Takers ............................ 103
3.5 Cutoff Estimates for Second Time Takers ........................ 103
3.6 Ability Distribution Parameters for First Time Takers ............ 104
3.7 Ability Distribution Parameters for Second Time Takers ........ 104
3.8 First Time - Low Type Cutoffs ..................................... 114
3.9 Second Time - Low Type Cutoffs ................................... 114
3.10 First Time - Share of Low Type Students .......................... 114
3.11 Second Time - Share of Low Type Students ....................... 115
Capabilities and Input Choice

1.1 Introduction

This paper analyzes the distribution of firms within a 2 dimensional space consisting of quality and labor productivity. As in Sutton (2007), to which this paper is closely related, production requires intermediate inputs. Although firm heterogeneity and production is of a similar form, this paper is embedded within the Melitz (2003) framework, instead of Cournot competition. In addition, instead of a single homogenous input, this paper will allow for input heterogeneity. Unlike recent work such as Goldberg et al. (2010) that has focused on input heterogeneity in terms of varieties, this paper focuses on quality differences among inputs.

In terms of inputs, this paper is similar to Kugler and Verhoogen (2012), where input quality and the firm’s quality aptitude are complementary in raising the quality of the firm’s output. However in that paper it is assumed that firms are heterogeneous in one dimension, termed “capability” (analogous to the “calibre” of Hallak and Sivadasan (2009)), which both labor productivity and quality aptitude are increasing in. This paper does not assume such a relationship between labor productivity and quality aptitude. In their recent paper, Kugler and Verhoogen (2012), the authors show that use of high quality inputs is associated with measures of productivity - the best firms choose the best inputs. This is predicted by their earlier theoretical framework.
Recent work (Feng et al. (2012)) has shown that industries with a substantial quality component have responded greatly to increased availability of high quality inputs, more so than industries that vary primarily through productivity.

In this paper I will show that when allowing a more general form for the relationship between the quality of a firm’s output and its labor productivity, the association between usage of high quality inputs and productivity need not be positive. The correlation can be positive or negative, depending on factor prices and the distribution of firms. Moreover, what appeared to be noise in their model, as productivity and input quality are increasing in the “capability” of the firm, may in fact be explained by firm heterogeneity.

It can also be shown that high wage economies react can react in a very different fashion than low wage countries. In particular, the reaction of the distribution of firms to shocks in factor prices can be opposite.

The rest of the paper is organized as follows: in section 2 I develop the basic model keeping inputs homogenous. In section 3 the model is extended to allow for heterogeneous inputs. In section 4 an extension to a general equilibrium model is briefly discussed. Section 5 concludes.

1.2 The Model

1.2.1 Demand

The preferences of the representative consumer are given by a C.E.S. utility function over a continuum of goods indexed by $\omega$:

$$U = \left[ \int_{\omega \in \Omega} s(\omega)^{\frac{\sigma - 1}{\sigma}} d\omega \right]^\frac{\sigma}{\sigma - 1}$$  \hspace{1cm} (1.1)

where $s(\omega) = x(\omega)q(\omega)$ is equal to the number of services consumed of variety $\omega$, where $x(\omega)$ is the physical amount of units consumed, and each unit is of quality $q(\omega)$. $\Omega$ represents the set of available goods. The measure of $\Omega$ represents the mass of available goods. Goods are substitutes; the elasticity of substitution is $\sigma > 1$. As in Dixit and Stiglitz (1977), there exists an associated aggregate good
such that $Q = U$ with a price given by:

$$P = \left[ \int_{\omega \in \Omega} p(\omega)^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}}$$  \hfill (1.2)

Note that $p(\omega)$ is equal to the price per service of variety $\omega$, not necessarily the price per output of variety $\omega$. The representative consumer has income $I$, it follows that optimal consumption decisions can be expressed as:

$$q(\omega) = \frac{Ip(\omega)^{-\sigma}}{P^{1-\sigma}}$$  \hfill (1.3)

### 1.2.2 Production

There exists a continuum of firms, each capable of producing a different variety $\omega$. The production technology entails a single unit of input being augmented by labor into a single unit of output. Firms differ by both the labor requirement as well as the firm's innate aptitude for quality, the latter determining the quality of output jointly with the type of input used.

The production function exhibits constant marginal cost, with a fixed overhead cost. All firms have the same overhead cost $f_o$, measured in terms of labor. The labor requirement for a unit of output varies across firms and is denoted by $c$. In addition, each unit of output requires a single unit of input, whose cost is normalized to 1. A single unit of output provides $u$ services, we allow $u$ to vary across firms. The pair $(u, c)$ is referred to as the capability of the firm, is fixed for any firm $\omega$, and is invariant over time. has a joint distribution function $G(u, c)$ with density $g(u, c)$. The marginal cost of a service for a firm with output quality $u$ and unit labor requirement $c$ is given by:

$$k(u, c) = \frac{cw + 1}{u}$$  \hfill (1.4)

Each firm faces a residual demand curve with constant elasticity $\sigma$, yielding a constant markup pricing rule:

$$p(u, c) = \frac{\sigma}{\sigma - 1} \frac{cw + 1}{u}$$  \hfill (1.5)
Firm profit is given by:

$$\pi(u, c) = \frac{I}{P^{1-\sigma}} p(u, c)^{-\sigma} \left[ p(u, c) - k(u, c) \right] - f_o$$  \hspace{1cm} (1.6)

Which, under optimal pricing, is equal to:

$$\pi(u, c) = \frac{I}{\sigma P^{1-\sigma}} \left[ \frac{\sigma}{\sigma - 1} \right]^{1-\sigma} k(u, c)^{1-\sigma} - f_o$$  \hspace{1cm} (1.7)

### 1.2.3 Aggregation

The equilibrium is characterized by a mass $M$ of firms and a distribution $\mu(u, c)$ of capabilities for active firms. The price index $P$ can now be expressed as:

$$P = \left[ \int_c \int_u p(u, c)^{1-\sigma} M \mu(u, c) dudc \right]^{\frac{1}{1-\sigma}}$$  \hspace{1cm} (1.8)

Using the pricing rule (5), this can be written as $P = M^{1/(1-\sigma)} \frac{\sigma}{\sigma - 1} \tilde{k}$, where

$$\tilde{k} = \left[ \int_c \int_u k(u, c)^{1-\sigma} \mu(u, c) dudc \right]^{\frac{1}{1-\sigma}}$$  \hspace{1cm} (1.9)

where $\tilde{k}$ is the weighted average cost per service of active firms.

### 1.2.4 Firm Entry and Exit

There exists an unlimited pool of potential entrants into the market. Firms are ex ante identical. Entry requires a fixed entry cost $f_e > 0$ measured in the same unit as inputs and operating costs, which is sunk after entry. Upon entry, firms draw their capability - the parameter pair $(u, c)$ drawn from a distribution $G(u, c)$. Upon entry and realization of it’s capability, a firm may choose to either exit, incurring no further costs, or remain in the industry. Firms that remain in the industry face a constant probability $\delta$ each period of a shock that causes immediate exit.

Since each firm’s capability does not change over time, the optimal profit per period will not change over time in a steady state. Therefore a firm that has a capability yielding negative profits would never remain active. Conversely, firms whose capabilities enable them to make positive profits will remain active, earning
positive profits, until they receive the bad shock and exit.

Since the capability \((u, c)\) enters the profit function only through the cost per service \(k(u, c)\), the profit function can be expressed as \(\hat{\pi}(k(u, c)) = \pi(u, c)\). Each firm’s value function can be written as:

\[
v(k) = \max \left\{ 0, \sum_{t=0}^{\infty} (1 - \delta)^t \hat{\pi}(k) \right\} = \max \left\{ 0, \frac{1}{\delta} \hat{\pi}(k) \right\}
\]

Let \(k^* = \sup \{ k : v(k) > 0 \} \) define the highest cost per service that an active firm will possess. Thus, firms with a capability draw yielding \(k(u, c) > k^*\) will choose to exit immediately. Firms with \(k(u, c) \leq k^*\) will remain active, earning a per period profit of \(\hat{\pi}(k)\), until they receive the bad draw. In a steady state, \(\hat{\pi}(k^*) = 0\).

Since firm exit due to the bad shock is unrelated to cost \(k(u, c)\), the equilibrium productivity distribution will not be affected by the exit process. Let \(H(k)\) denote the distribution of costs determined by the initial capability draw. We can now express the distribution of capabilities among active firms as:

\[
\mu(u, c) = \begin{cases} 
\frac{g(u, c)}{H(k^*)} & \text{if } k(u, c) \leq k^* \\
0 & \text{otherwise}
\end{cases}
\]

where

\[
H(k) = \int_c \int_{u = \frac{cw + 1}{k}} 1dG(u, c)
\]

The aggregate cost level can be described as a function of the cutoff cost level:

\[
\bar{k} = \left[ \frac{1}{H(k^*)} \int_c \int_{u = \frac{cw + 1}{k}} \left( \frac{cw + 1}{u} \right)^{1-\sigma} g(u, c) dudc \right]^{\frac{1}{1-\sigma}}
\]

Similarly the aggregate price level can be written as

\[
P = M^{1/(1-\sigma)} \frac{\sigma}{\sigma - 1} \left[ \frac{1}{H(k^*)} \int_c \int_{u = \frac{cw + 1}{k}} \left( \frac{cw + 1}{u} \right)^{1-\sigma} g(u, c) dudc \right]^{\frac{1}{1-\sigma}}
\]

As in Melitz (2003) it is possible to use the zero cutoff profit condition, obtaining \(1\)this is since \(f_o > 0\), therefore the set \(\{ k : \hat{\pi}(k) < 0 \}\) is non-empty.
the following relation between cutoff cost $k^*$ and average profits among active firms:

$$\bar{\pi} = f_o \left[ \left( \frac{k^*}{k(k^*)} \right)^{\sigma-1} - 1 \right]$$

(1.15)

The free entry condition leads to:

$$\bar{\pi} = \frac{\delta f_e}{H(k^*)}$$

(1.16)

Through this, it can be shown that there exists a unique cutoff cost $k^*$ and $\bar{\pi}$. The mass of firms $M$ is obtained by setting $\pi(k^*) = 0$.

### 1.2.5 Equilibrium Analysis

Consider an equilibrium that results from a wage equal to $w_0 > 0$. The equilibrium can be characterized by the cutoff cost and the aggregate price level. Denote these by $k_0^*$ and $P_0$. Firms that have a cost $k < k_0^*$ are active, setting their prices optimally at $\frac{\sigma}{\sigma-1}k$. The following diagram describes the set of active firms:

![Figure 1.1: The cutoff cost $k^*$ represented in capability space](image)

Firms that receive a capability draw above the cutoff line $u = \frac{cw_0 + 1}{k_0^*}$ are active. Suppose that the market wage falls to $w_1 < w_0$. This rotates the locus downwards, allowing firms that previously had high unit labor requirements, relative to
the quality of the output, to have a sufficiently low cost per service. At the same
time as these firms are entering the market, the costs of existing firms have been
lowered since their labor requirement is relatively cheaper. Both of these effects
cause the aggregate price level to decrease. In equilibrium, the set of active firms
shifts in the following way:

\[ \frac{cw_0 + 1}{k_0^*} \]
\[ \frac{cw_1 + 1}{k_1^*} \]

Figure 1.2: The change in capabilities of active firms

\[ \frac{1}{k_1^*} \]
\[ \frac{1}{k_0^*} \]

\[ c \]

\[ u \]

\(^2\)Of course those firms with \( c = 0 \) do not see their costs lowered.
The maximum cost $k^*$ must fall. Suppose not: $k_0^* \leq k_1^*$. This implies that the aggregate price level has risen. But now all firms that were active will still be active, and will have higher profits. Additionally, the set of firms able to make positive profits has expanded. Therefore the free entry condition has been violated - the expected profit that a prospective entrant observes has increased. Therefore this cannot be an equilibrium with free entry.

Also, there will exist firms that were not active in the equilibrium under $w_0$ that will become active. If not, we know that the change in the price level will be greater than the change in the cost for all active firms, thus we have fewer firms active, all earning lower profits than before. This again violates free entry.

### 1.3 Heterogeneous Inputs

We now introduce heterogeneous inputs into the model. There is a perfectly competitive input sector, with no fixed costs of production, producing inputs of quality $i \in [0, 1]$ using capital and labor with the following CRS production function:\footnote{The positive association between capital intensity and quality is motivated by Flam and Helpman (1987) - rich countries tend to be better at producing high quality output}

$$F_i(K, L) = \frac{L^{1-i}}{(1-i)^{1-i}} \frac{K^i}{i^i}$$  \hspace{1cm} (1.17)

In equilibrium, the price at which inputs are be offered will be equal to the unit cost of production. Therefore:

$$p(i) = w^{1-i}r^i$$  \hspace{1cm} (1.18)

With the cost of capital normalized to 1, this becomes $p(i) = w^{1-i}$.

As before, production of one unit of output requires a single input. However the services per output, i.e. the quality of the final output, will be determined jointly by the firms quality capability $u$ and the quality of the input $i$. These variables are assumed to be complementary in raising the quality of the output. As such, a CES function is used:

$$\text{Output Quality} = \left[\frac{1}{2}u^\rho + \frac{1}{2}(1 + i)^\rho \right]^{1/\rho}$$  \hspace{1cm} (1.19)
where $\rho < 0$.

Firms choose the quality of the input in order to minimize the cost per unit of output:

$$k(u, c) = \min_{i \in [0, 1]} \frac{cw + p(i)}{[\frac{1}{2}u^\rho + \frac{1}{2}(1 + i)^\rho]^{1/\rho}}$$ (1.20)

For $w \geq r = 1$ this problem is trivial. It follows that $p(i)$ is decreasing in $i$ yet the amount of services per unit of output is increasing in $i$, hence all firms will choose the highest quality input, $i = 1$. In this case there is no meaningful input heterogeneity - in equilibrium all firms must choose the same input. As a result the case where $w < 1$ will be the focus of the remainder of the paper.

For $w < 1$ the input price function is increasing and convex over $[0, 1]$. Also, the CES output quality function is increasing and concave for $i \in [0, 1]$. The objective function is therefore convex. It follows that there exists a unique solution to the minimization problem. Denote the cost minimizing input as $i_w(u, c)$. For an interior solution, the following first order condition must be satisfied:

$$-\ln w = \frac{(cw^i + 1)(1 + i)^{\rho - 1}}{w^\rho + (1 + i)^\rho}$$ (1.21)

### 1.3.1 Input Choice

The following figure illustrates input choice for firms for a given level of $w$.

As can be seen in the figure, the optimal choice of input quality $i$ is increasing in both the firm’s capability for quality $u$ and the unit labor requirement $c$ (in the interior).

Since $u$ and $i$ are assumed to be complements in raising the quality of the output, it is not surprising that firms with a higher draw for $u$ choose a higher quality input. As $u$ increases, the firm’s ability to utilize the better inputs improves, so they are able to pay more for the input.

Input choice is also increasing in $c$, albeit for a different reason. Compare two firms, with $u_1 = u_2$ but $c_1 < c_2$. For any given input choice the cost of output will be higher for firm 2 than for firm 1. However the input share of total costs would be smaller for firm 2 for a given input choice. Therefore for firm 2 the cost of output is less sensitive to input choice than for firm 1 (in terms of elasticity,
not absolute changes in cost). However there is no difference regarding the effect of \( i \) on the quality of the output. Since the firm is trying to minimize the cost per service provided, firm 2 should choose a higher level of \( i \).

While both \( u \) and \( c \) raise the level of \( i \) chosen optimally, a higher \( u \) raises productivity while a higher \( c \) lowers productivity. Hence it is incorrect to assert that high productivity firms choose high quality inputs in general. The correlation could be positive, negative or non existent depending on the distribution of \((u, c)\). A firm that chooses a high quality input might do so because it is productive or because it is unproductive.

An increase in wages will lead to higher levels of \( i \) being chosen: although the cost of all inputs \( i \in [0, 1) \) increases, \( \frac{dp(i)}{di} \) falls. In addition, an effect similar to that of \( c \) on \( i \) occurs: labor costs rise, reducing the importance of input costs.\(^4\) First order conditions dictate that a higher quality input will be chosen.

### 1.3.2 Inputs and Productivity

In the previous section, it was established that costs of production, expressed as \( k \), the cost per service, has an ambiguous relationship with the input chosen by a firm. As the labor requirement \( c \) increases, a higher input quality is chosen and

\(^4\)Although input costs may rise, they rise less than wages, therefore less than labor costs.
costs increase: a positive relationship. As the firm’s quality aptitude $u$ increases, a higher input quality is chosen but costs decrease: a negative relationship. It follows that for the same parameters of production consumption, there exists some $G(u, c)$ that can generate any relationship between input choice and productivity. However monotonic relationships will only be seen for a very narrow class of distributions.\footnote{In particular, it is required that for all $(u, c)$ such that $g(u, c) > 0$, for any open ball centered at $u,c$ there exists some element $(u^*, c^*)$ with $g(u^*, c^*) = 0$}

More generally, the correlation, if any, will be less than perfect. The sign of the correlation will be increasing in the variability among firms with respect to $u$, and decreasing with respect to labor productivity. Hence in economies where heterogeneity of labor requirements is small, most of the observable heterogeneity (input choices and productivity) is due to differences across $u$, and a positive relationship between input choice and productivity would be expected.

Similarly, for a given $G(u, c)$ the correlation between input choice and productivity is increasing in $w$.\footnote{Note that as $w \to 1$, all firms choose $i = 1$ and there is not variation in input choice, hence no correlation} It can be shown that the slope of the iso-input curve at some $(u, c)$ is increasing in $w$. As the curve becomes flatter, an increase in input choice is increasingly more likely to have resulted from a higher $u$ than from a higher $c$. As a result the correlation between input choice and productivity should increase.

See attached diagrams for illustrations of the endogeneity of correlation.

The correlation is important when trade in intermediate inputs is considered. For a low wage country, trade in intermediate input will likely make high quality input more affordable. If there is a positive correlation between input choice and productivity, liberalization will benefit the most productive firms, increasing their size. At the same time, the least productive firms will face additional competition due to the liberalization not lowering the relevant input prices, while increasing the competition from the largest domestic firms. A negative correlation would predict the opposite. Therefore intermediate trade liberalization can have very different effects on the distribution of firms, depending on the sign of correlation.
1.3.3 Elasticity of costs

In this section I will discuss the patterns of cost elasticity with respect to w. In the homogeneous inputs case, the only effect of w on cost was through the labor unit requirement c. As a result firms that had high labor productivity had little to gain from a decrease in wages. When heterogenous inputs enter the production function, wages are able to indirectly affect costs.

As inputs are produced using labor and capital, the cost of labor affects the price of inputs. For low quality inputs this effect is considerable. Since low quality inputs are relatively labor intensive, \( \partial p(i)/\partial w \) is decreasing in i. In particular, the lowest quality input \( i = 0 \) is produced using only labor, \( p(0) = w \). Firms that choose this input are those with low draws for c and u, the firms that previously would have had little to gain from cheaper labor. In fact these firms now have the greatest elasticity of costs with respect to wages: their costs are linear in wages.

\[
\varepsilon = \frac{cw + (1 - i(u, c))w^{1-i}}{cw + w^{1-i}}
\] (1.22)

Firms that choose \( i = 0 \) have \( \varepsilon = 1 \). As in the homogeneous case, the high c firms still gain from cheaper labor, \( \varepsilon \to 1 \) as \( c \to \infty \) (although such firms are unlikely to be active). It is the firms that choose high quality inputs, but do not have high labor costs, that have the least to gain from cheaper labor. The following diagram illustrates the high elasticity of firms that choose low input, due to low draws of u or c. See attached diagram for elasticities with homogeneous inputs.

Consider a steady state where there exist active firms that choose the lowest quality inputs \( i = 0 \). Suppose that w falls. All firms that chose \( i = 0 \) will still be active in the new steady state. Furthermore, there exists some \( c^* > 0 \) such that \( \forall c \in [0, c^*), \) the minimum level of u required to be active is strictly lower than it was in the old steady state.

This is in contrast to the result in the homogeneous input case that the minimum u for low values of c increases. However if there are no active firms that choose \( i = 0 \) this result does not hold: firms that choose \( i > 0 \) are not guaranteed to gain the most from a decrease in wages and may see the price level fall more than their costs.

It follows that, once heterogeneous inputs are allowed for, high w economies
not only appear different, but will also respond in different ways when compared to low income countries. As the following figure demonstrates, the change in the distribution of active firms from as wages fall from 0.2 to 0.1 is very different to that of a fall from 0.8 to 0.7.

While a decrease in \( w \) at a relatively high wage had an effect similar to rotating the locus clockwise, a decrease in \( w \) at a relatively low wage had an effect similar to rotating the locus counter-clockwise.

### 1.4 Trade in Intermediate Inputs

Consider a small, low wage, closed economy (country H) that opens up to trade in intermediate inputs with a large economy with (country F) with \( w_F > w_H \) and \( r_F < r_H = 1 \) (with \( w < r \) for both economies). For low quality inputs, the home price, \( p_H(i) \) will be less than the foreign price, \( p_F(i) \), as the home country has a comparative advantage in the labor intensive, low quality input. Conversely, the foreign country has an advantage in the high quality inputs as a result of its lower capital cost.
Upon opening up to intermediate trade, we can immediately note two things. Firstly, the home country will export low quality inputs while importing high quality inputs. Secondly, the wage in the home country must increase as a result of the increased demand for labor. As seen in Figure 1.6, input qualities to the left of the intersection of (Home New) and (Foreign) are exported, those to the right are imported. The price schedule that firms face is given by the minimum of the two lines.

We can also see that firms will shift to higher quality inputs: the slope of the price of the home produced input has fallen (due to the higher wage), while the slope of the foreign produced input has fallen (due to higher value of $w/r$ abroad. At the same time, there is a discontinuity in the slope of the price of inputs available in the home country, at the input quality that home stops producing at. So many firms will undergo a discrete jump in the quality of their inputs as a result of the more affordable high quality imported intermediates.

Moreover, it is straightforward to see which firms will gain from the trade liberalization, when grouping firms in terms of quality. The cost of low quality inputs has increased, while the cost of high quality inputs has fallen. So it is the
firms that chose high quality inputs who have gained in terms of changes in costs.

However, when grouping firms by profitability or market share (cost per service provided) we cannot say which group will benefit from the liberalization, as it depends on the the relationship between input quality and profitability. It is possible for the best firms to choose high quality inputs when firms vary primarily by the quality capability; however if firms vary mainly by per unit labor productivity, the worst firms might gain the most.

As a result, while opening up to intermediate input trade has a predictable effect on input choice throughout the industry, the effects on the composition of the industry depend very much on the distribution of capabilities throughout the industry.

### 1.5 Conclusion

Compared to a model of capabilities where inputs are homogeneous, allowing for heterogeneity is better able to account for the observation that low wage economies tend to produce low quality output. This is due to both the decrease in the quality of inputs used by firms, and the fact that firms producing low quality output can benefit the most from a decrease in wages. Furthermore, the effects of changes in
factor prices can vary greatly within a model with heterogeneous inputs.

Allowing a more general form for capabilities than Kugler and Verhoogen (2012) has also shown that the correlation between input choice and productivity need not be positive. While it may have been positive for Colombia, it may be negative for a country with a lower cost of labor, or for a country where firms have greater variation in labor productivity.
Figure 1.7: Cost elasticities w.r.t. $w$ when inputs are homogeneous
Chapter 2

Priors and Posteriors: Implications for Exporters

2.1 Introduction

What drives the entry of firms into foreign markets? To understand trade flows, and to better predict the implications of trade policies, this question needs to be answered. Models incorporating firm level heterogeneity such as Melitz (2003) suggest that firms face sunk costs when entering a market for the first time, leading to high productivity firms entering a market, while less productive firms self-select to refrain from entry.

However, recent empirical research suggests that firms exhibit substantial heterogeneity in market outcomes that plant level productivity alone does not account for. Cherkashin et al. (2010) find that although most firms satisfy the market hierarchy as implied by the Melitz model, a substantial (in terms of output) number of firms serve difficult markets while remaining absent from easier ones. Hu et al. (2012) show that market specific demand is a much better predictor of export behavior than productivity.

In addition, the amount of exit from markets observed is hard to explain in a model where profitability is known before entry into foreign markets. Eaton et al. (2008) find that, typically, two thirds of new exporters in Colombia do not survive past their first year, suggesting that “weeding out” occurs in the first period.
Similar patterns are seen when considering survival of Chinese exporters in new markets: a large amount of exit in the first year, followed by quite moderate exit.\footnote{See Table 2.6, showing survival of Chinese exporters is destinations, by cohort of firm/year of entry.} Moreover, this weeding out isn’t specific to new exporters. Experienced exporters in new markets exhibit a similar pattern of dramatic exit in the first year: first year exporters have a failure rate of 66-70 percent, while firms with a year of export experience have a failure rate of 58-60 percent when entering a new market.

While failure appears to be the most likely outcome of entry, some new exporters will continue to enter new markets, at an (initially) increasing rate. This suggests that the acquisition of experience encourages the firms to be more aggressive in their expansion choice, entering more markets at the same time.\footnote{Tables 2.7 and 2.8 show that as firms enter more markets, the likelihood that they will continue to expand their set of destinations increases, as does the conditional speed of expansion. Similarly when the set of markets is restricted to the EU (Tables 2.5 and 2.9), although naturally both decrease as the number of attractive markets remaining falls.}

In order to reconcile the high chance of failure with the high costs of entry as estimated by Das et al. (2007), I consider a model where firms learn through entry into markets. If the sunk costs of entry are high, it would be expected that the chance of exit is low: firms that are productive enough to overcome the entry barrier should not be expected to exit almost immediately. As opposed to learning resulting in productivity improvements, as in papers such as Bai et al. (2012), or in reduced fixed costs, such as in Sheard (2011), firms instead learn about their demand. Although demand in each market is uncertain prior to entry, if it is allowed to be correlated across markets then entry will provide valuable information regarding the distribution of outcomes in other markets. As a result, firms enter markets knowing that the outcome will impact their expansion decisions in the future, justifying the large initial cost.\footnote{Here the decision is which markets to enter, however outcomes could affect choices for markets already entered. Eaton et al. (2012) for example let outcomes guide the search for new customers within a market.}

Unlike Jovanovic (1982), and Arkolakis and Papageorgiou (2010), where firms learn about profitability over time, firms learn through observing market specific demand draws, without any variation over time within a market.\footnote{While the time aspect of learning finds empirical support, explaining the relationship between age, output and exit, it severely complicates the problem in a multiple market setup.} As such, it is the
experience of a firm, in terms of markets entered, that governs the beliefs regarding expected returns from entry. While age and experience do not convey real benefits to productivity, they provide a different type of advantage: a more accurate belief regarding profitability. Firms will typically only enter difficult markets when they have enough experience to convince themselves that they can earn a sufficiently high profit.

In addition, belief updating will typically encourage firms to be slow in their expansion, in order to make more informed decisions. When expanding, firms choose between sequential entry and simultaneous entry. Sequential, rather than simultaneous, entry allows for more accumulation of information and “better” entry decisions, at the cost of delaying profit flows from entry (discounting occurs). As a result only the highly profitable firms would be expected to enter multiple markets simultaneously, whereas others will either enter sequentially or will cease entering further markets. It follows that firms which enter multiple markets simultaneously are those that have a very high belief regarding profitability, so should be more likely to survive, and will be more likely to expand later on.\(^5\)

As shown in Eaton et al. (2011), the number of firms selling to a market increases with the size of the destination market, so that more firms choose to enter the larger markets than the smaller markets.\(^6\) While such a pattern is consistent with standard models of trade, introducing learning has strong implications for the relationship between market size and trade flows, while also accounting for the number of firms exporting to each market.\(^7\)

While the order in which firms expand into new markets is often ignored, this paper shows that a market rank is highly important with respect to entry, exit and survival of firms. Holding market characteristics equal, markets that are typically

\(^5\)As seen in Tables 2.12, 2.13 and 2.14, there is a positive relationship among new exporters to the EU between survival and speed of expansion.

\(^6\)Alternatively, the the survival rate of firms could be much higher in the larger markets, however the magnitude of the variation rules this out as the primary cause.

\(^7\)When looking at exporters to the EU we observe that more firms go to the larger markets, shown in Table 2.15. In addition, we see that new firms overwhelmingly choose to enter the larger markets first, such as Germany and the United Kingdom, and move down the ladder to smaller markets as they acquire more experience (Table 2.10). As could be expected, the smaller markets see lower survival rates. However, it seems that the effect of size on survival is tempered somewhat by the experience of entrants: while smaller markets are in general harder to survive, the firms which choose to enter typically do so after gaining a deeper understanding of their demand.
entered into earlier will have greater entry, greater exit rates and will have more firms active.

The empirical section focuses on exports to the EU for three reasons. First, it avoids the issue of differences in trade policy which could lead to substantially different exporting decisions, as seen in Kee and Krishna (2008), and Demidova et al. (2012). Second, focusing on a set of markets with relatively similar preferences makes modeling choices more plausible: outcomes in a market need to be relevant to other markets. Finally, the number of markets is such that estimation could be computationally feasible.

This paper is closely related to Albornoz et al. (2012) and Nguyen (2012). The former features a model of learning regarding the export costs of a firm, where heterogeneous domestic firms consider either simultaneous or sequential entry into foreign markets. This paper extends Albornoz et al by allowing more than two destination markets, and by extension more than two periods, and by relating the differences between markets to patterns of firm dynamics. At the same time, assumptions regarding these differences are imposed on the model in order be applicable to exports from China to the EU. Nguyen (2012), similar to this paper, features a theoretical model in which firms learn through observing market specific demand shocks, while imposing symmetry on markets. This paper relaxes symmetry, whilst imposing structure on how markets differ in order to maintain tractability.

Timoshenko (2012) features product appeal learning in an environment featuring correlation of demand shocks for a firm across products and time within the same destination, but not between destinations. This accounts for product switching within markets; new exporters respond to fluctuations in demand to a greater extent than old firms.

As in Cherkashin et al. (2010), the effect of liberalization across markets is analyzed. The learning aspect of the model allows for positive spill-overs across markets. When one market reduces trade barriers, other markets will tend to see increased entry: learning and experience facilitates entry. However the hierarchical nature of the model implies that liberalization has the potential to harm nearby markets, as a change in the position in the order of expansion can greatly reduce entry.
Section 2 presents the general model, section 3 examines the model in a partial equilibrium setting, section 4 presents reduced form evidence, section 5 presents an extension, section 6 looks at the policy implications, and section 7 concludes.

2.2 The Model

The economy consists of the firm’s home country and \( N - 1 \) foreign economies. Time is discrete. In each period firms make decisions regarding market entry based upon their beliefs regarding their own profitability as in Melitz (2003). Firms then receive a draw of a market specific demand shock in markets which they enter. Firms then choose to either exit immediately, or export to a market, depending on demand, productivity and costs of production.

2.2.1 Demand

Within each market there exists a unit measure of identical consumers, endowed with a constant elasticity of substitution (CES) utility function over varieties \( \omega \), as in Dixit and Stiglitz (1977). The utility of the representative individual in market \( n \) is given by:

\[
U_n = \left[ \int_{\omega \in \Omega_n} (x_n(\omega)q_n(\omega))^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}} \tag{2.1}
\]

where \( q_n(\omega) \) denotes consumption of firm \( \omega \)'s variety, \( x_n(\omega) \) represents variety \( \omega \)'s market specific product appeal and \( \Omega_n \) denotes the set of firms that are currently supplying market \( n \). The elasticity of substitution is \( \sigma > 1 \).

The associated aggregate price index for a given market \( n \) is given by:

\[
P_n = \left[ \int_{\omega \in \Omega_n} \left( \frac{p_n(\omega)}{x_n(\omega)} \right)^{\frac{1}{1-\sigma}} d\omega \right]^{\frac{1}{1-\sigma}} \tag{2.2}
\]

where \( p_n(\omega) \) is the price that firm \( \omega \) sets in market \( n \).

The representative consumer in market \( n \) has income \( Y_n \), it follows that optimal consumption decisions can be expressed as:

\[
q_n(\omega) = Y_n \frac{p_n(\omega)^{-\sigma}}{x_n(\omega)^{1-\sigma}P_n^{1-\sigma}} \tag{2.3}
\]
To see how this equation is obtained, consider a model of services, where $x$ is interpreted as the services provided by a nominal unit of output. The price of a service from firm $\omega$ is $p(\omega)/x(\omega)$, so that the demand for services, $x(\omega)q(\omega)$, follows the standard CES demand function. Rearranging the equation gives the demand for units of output. The product appeal $x_n(\omega)$ can be interpreted as the number of consumers that are interested in the product: increasing the appeal simply increases demand for any given price set by the firm.

\subsection*{2.2.2 Production}

In order to enter the industry, each firm must pay an entry cost $f$. Once this cost has been incurred, the firm is able to produce a unique variety $\omega$. While each firm has the potential to sell in any of the $N$ markets, it must first pay a sunk cost in order to enter any particular market. The cost of entry into market $n$ is denoted as $f_{ne}$. In addition, there are fixed operation costs of $f_{no}$ per period for each market $n$. Firms discount future earnings by a discount factor $\delta < 1$.

Firms are heterogeneous in terms of productivity: a firm can produce $\varphi$ units of output using a single unit of labor, the cost of which is normalized to unity in the home market. A firm’s productivity $\varphi$ is drawn from a distribution $H(\varphi)$ and is invariant across time and markets.\footnote{\textit{H}(\cdot) only has positive support for $\varphi > 0$.}

In addition to heterogeneous productivity, firms differ by the appeal of their variety in the various markets.\footnote{While product appeal and productivity may be isomorphic with respect to profits within a market, the implications for expectations are not.} Firms must first enter a market in order to learn the market specific product appeal. However, product appeals are correlated across markets for a given firm: as they enter more markets they have a better understanding of their profitability in markets they have yet to enter.

In essence, there are two terms which affect the profitability of a firm - one on the supply side and one on the demand side. Both are needed in order to capture key features of the data, for example why some firms serve difficult markets but not easier ones (Kee and Krishna (2008)). However it is only necessary for one to be imperfectly correlated across markets, providing dynamics that result from learning about profitability. As opposed to demand, productivity is not market
Mechanically, this is achieved by letting product appeal depend on two variables. Each firm is endowed with $\alpha(\omega)$. This firm specific component is identical across markets and time, and captures how appealing the variety is in general. The second component, $\beta_n(\omega)$, is the market $n$ specific demand shock that the firm producing variety $\omega$ draws. These combine to give the firm a product appeal in market $n$ of $x_n(\omega) = \alpha(\omega)\beta_n(\omega)$. While firms observe product appeals $x_n$ they do not directly observe the components $\alpha$ and $\beta_n$. Firms particularly care about $\alpha$ as it provides information on the expected outcomes in markets which it has yet to enter.

Unlike Jovanovic (1982), I assume that there is no variation in product appeal over time. As a result there is no additional information gained by a firm while remaining for subsequent periods in a particular market: everything is learned in the first period.\footnote{The combination of correlation across markets and variation over time would significantly increase the complication of the entry/exit problem, thus I abstract from this aspect of learning for tractability.}

The distribution of $\alpha$ draws is log-normal with parameters $\mu_\alpha$ and $\sigma_\alpha$, denoted $F_0(\alpha)$, and is known by all prospective firms. Similarly, $\beta_n$ is distributed log-normally\footnote{This parametrization ensures a closed conjugate prior; other possibilities include the gamma distribution with known shape parameter $\alpha$.} with $\mu_\beta = 0$ and $\sigma_\beta$, with draws independently and identically distributed across markets.\footnote{It is conceivable that these shocks are not iid: perhaps some markets have a lower variance, or a particular pair of markets have highly correlated preferences.} Denote the distribution function of $\beta$ as $G(\beta)$.

At the start of each period, each firm observes its productivity $\varphi$, which markets it has previously entered, the product appeals for each of those markets, and which markets it has exited from. The firm then considers entry into available markets (those which it has not yet entered) based on its beliefs over $\alpha$.\footnote{Beliefs are formed rationally.} After paying the sunk costs of entry, the firm receives appeal draws for its newly entered markets. Production (or exit) then occurs. Finally, firms update their beliefs regarding $\alpha$, based on the history of product appeal draws which they have observed.

For the purposes of entry decisions, the state of the firm is the vector of $(\varphi, A, \{x_i | i \in A\})$ where $A$ is the set of markets that have been entered previ-
ously. Note that the particular market where each product appeal \( x \) is observed is irrelevant when considering expansion to different markets.\(^{14}\)

If a firm is active in a market, it will maximize profits by offering a price equal to:

\[
p = \frac{\sigma}{\sigma - 1} \left( 1 - \frac{1}{\varphi} \right)
\]

(2.4)
as the wage has been normalized to one.

Accordingly, it will receive a per period profit equal to:

\[
\pi(x_n, \varphi, P_n) = \left( \frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \frac{Y_n}{\sigma P_n^{1-\sigma}} (\varphi x_n)^{\sigma-1} - f_{no}
\]

(2.5)

As there is no variation in either the firm’s productivity or the level of competition over time, the firm will exit immediately if \( \pi(x, \varphi, P) < 0 \), otherwise it will continue. This can be expressed as an optimal exit rule: exit if \( x\varphi < A \), where:\(^{15}\)

\[
A = \left( f_{no} \frac{\sigma P_n^{1-\sigma}}{Y_n} \left( \frac{\sigma}{\sigma - 1} \right)^{\sigma-1} \right)^{-1}
\]

(2.6)

### 2.2.3 Product Appeal Updating

At the start of each period, each firm holds a belief over its core demand parameter \( \alpha \). Denote the belief of firm \( \omega \) as \( F_\omega(\cdot) \). This belief is generated through Bayesian updating of the initial belief, \( F_0(\cdot) \) (identical across firms), based upon observations of the demands, and therefore market specific product appeals, in the various markets it has entered.

Recall that the distribution of market specific shocks is \( G(\beta) \). Suppose that a product appeal of \( x \) is observed. According to Bayes’ rule, \( P(\alpha|x) = \frac{P(x, \alpha)}{P(x)} \). The updated pdf of \( \alpha \), \( f^{\text{new}}_\omega(\alpha) \), is given by:

\[
f^{\text{new}}_\omega(\alpha) = \frac{g(\frac{x}{\alpha}) f_\omega(\alpha)}{\int_{a=0}^{\infty} g(\frac{x}{a}) f_\omega(a) da} = \frac{P(x, \alpha | \text{beliefs})}{P(x | \text{beliefs})}
\]

(2.7)

where \( G'(\beta) = g(\beta) \).

\(^{14}\)The history of product appeal draws can be substituted by the belief regarding \( \alpha \).

\(^{15}\)The exact decomposition of \( x\varphi \) is not relevant to the firm’s decision regarding whether or not to exit an market in which it has already entered - only the product of the three unknown terms \( \alpha\beta\varphi \) matters.
This process continues for all observed product appeal draws. As such, we can describe the belief of a firm that has observed \( n \) product appeals from \( n \) distinct markets as \( F(\alpha|n, Y) \), where \( Y = \sum \log(x_i) \).\(^{16}\) This belief is log normally distributed with the following parameters:

\[
\mu_{\alpha}^{\text{new}} = \frac{Y\sigma_{\alpha}^2 + \mu_{\alpha}\sigma_{\beta}^2}{n\sigma_{\alpha}^2 + \sigma_{\beta}^2} \quad (2.8)
\]

\[
\sigma_{\alpha}^{\text{new}} = \sqrt{\frac{\sigma_{\alpha}^2 \sigma_{\beta}^2}{n\sigma_{\alpha}^2 + \sigma_{\beta}^2}} \quad (2.9)
\]

As expected, as firms observe more draws their beliefs get tighter, converging to the true value of \( \alpha \).

Note that there are no spillovers across firms: a particular firm’s experiences in a market are of no relevance to other firms. This rules out, for example, a successful exporter of a particular product to a market acting as a pioneer, being followed by many firms.\(^{17}\)

### 2.3 Partial Equilibrium

There are three factors that affect the decision to enter a particular market. First is the expected profit flow from the new market, given the beliefs that the firm holds over \( \alpha \) and \( \varphi \). Second, the expected value of an additional product appeal draw, again given beliefs that the firm holds over \( \alpha \) and \( \varphi \). Such information is valuable as it provides information relevant to the entry decision in future periods. Finally, it may be optimal to delay the entry decision, waiting until outcomes from other markets are realized.\(^{18}\)

\(^{16}\)Note that this is imperfectly correlated with average profits.

\(^{17}\)Freund and Pierola (2010) look at Peruvian exporters, finding that experimentation in new products is rare and typically performed by large experienced exporters. Iacovone and Javorcik (2010) similarly find that export discoveries are rare and quickly imitated using Mexican trade data. Bolton and Harris (1999) provides a theoretical model that examines the strategic effects of experimentation in an environment where players learn from others.

\(^{18}\)While a firm may delay into a particular market, i.e. wait until they have acquired more information, the model does not allow for a firm to make no entry in one period, then enter next period. As there is no gain in information, and hence no change in beliefs, the firm should either have entered a period earlier, or not entered at all.
There is however only one factor that determines whether a firm should exit a market - the maximized per period profit, denoted \( \pi_n \). Since the entry cost is a sunk cost, if the combination of productivity and demand is sufficient to overcome the fixed operation costs, the firm will continue to operate. As product appeal is time invariant, there is no additional information gained by remaining active. Therefore the optimal decision is to exit immediately if the value of \( x \varphi \) is less than the cutoff given by equation 6, otherwise to remain active in the market.

In this section I will describe the firm’s export decision in a restricted version of the model, under partial equilibrium. In particular, I make the assumption that markets are identical in all dimensions except for one:

**Assumption 1.** Markets vary by only one dimension: either sunk entry costs \( f_{ne} \), operating costs \( f_{no} \), or market potential \( \frac{Y_n}{F_n^{-\sigma}} \).

This allows a convenient ordering of markets:

**Lemma 2.3.1.** There exists a unique ordering of markets such that for all possible beliefs \( F(\cdot) \) and all levels of productivity \( \varphi \), \( i < j \) implies \( E(\max(\frac{\pi_i}{1-\delta},0) - f_{ie}|F_\omega(\cdot),\varphi) > E(\max(\frac{\pi_j}{1-\delta},0) - f_{je}|F_\omega(\cdot),\varphi) \)

The following diagrams showing profit (net of entry costs) as a function of realized \( x \), for a fixed value of \( \varphi \), illustrate this:

![Figure 2.1: Differ by one variable](image1)

![Figure 2.2: Differ by two variables](image2)

The figure on the left illustrates the difference in profits between two markets which are identical except for market potential \( \frac{Y_n}{F_n^{-\sigma}} \). Regardless of the realization of \( x \), profits are (weakly) higher in the market with the larger size (the dashed line). As beliefs must afford support for all positive values of \( x \), the firm expects
higher profits in the larger market. The value of productivity $\varphi$ does not affect this relationship, therefore all firms must expect higher profits from the larger market. Similarly for when markets vary by entry costs or operating costs.

However, if markets vary by multiple variables, as they do in the figure on the right, then a consistent ranking is no longer guaranteed. A market with low operating costs yet small market potential might be attractive relative to the alternative (high operating costs, high market potential), for a subset of firms, but certainly not all. As illustrated, there exists a range of product appeals $x$ (given productivity $\varphi$) for which the dashed line (low cost, small) gives higher profits, whereas the best firms will certainly prefer the higher potential of the alternative.

The assumption regarding how markets differ ensures that all firms would have the same ordering of the net present value of the flow of profits (inclusive of entry costs, in expectation) of the markets. Otherwise it would be possible for firms with different productivities/beliefs to have different orderings of markets.

The assumption that $\beta_n$ is identically and independently distributed ensures that the information gained from a market is treated the same as information gained from another market. For example, spending a single period in a given market is expected to give as much information regarding $\alpha$ as a single period in any other market. Therefore entry order can be restricted as follows:

**Proposition 2.3.2.** If a firm enters market $j$, it must have entered all markets $i$ such that $i < j$.

The ordering restriction allows us to restrict the order of entry, which allows a much simpler problem for the firm. Instead of choosing a set of markets to enter, the firm can be modeled as simply choosing how many markets to enter. In addition to reducing the choice set, the state space for the value function is greatly reduced. As a result, conventional estimation approaches may be feasible, instead of approaches such as Morales et al. (2011), using moment inequalities.

Suppose that a firm with productivity $\varphi$ has entered markets $\{1, \ldots, n\}$ and has observed a sum of log product appeals of $X$. The value to this firm of the options

---

19 While it is plausible that certain markets may yield more information regarding $\alpha$, due to a lower variance of the $\beta$ draws, that is beyond the current scope of this paper. It also dramatically increases the complexity of solving the problem, as the state depends on all demand draws, not the sum.
to enter markets \( \{n+1, \ldots, N\} \) is denoted \( V(n, X, \varphi) \). Note that \( \forall X, \varphi \), the option value \( V(N, X, \varphi) = 0 \), as there are no markets remaining.

Denote the expected flow of profits that this firm will receive from entering the next \( q \) markets as:

\[
W(n, X, \varphi, q) = \sum_{i=n+1}^{n+q} E[\max(\frac{\pi_i}{1-\delta}, 0)] - f_{ie}
\]

The problem that a firm that has entered markets \( \{1, \ldots, n\} \) faces is given by:

\[
q^* = \arg\max_{q \in \{0, \ldots, N-n\}} W(n, X, \varphi, q) + \delta E[V(n + q, X + \sum_{i=n+1}^{n+q} x_i, \varphi)]
\]

Hence, the value of a firm (with respect to available markets) is:

\[
V(n, X, \varphi) = W(n, X, \varphi, q^*) + \delta E[V(n + q^*, X + \sum_{i=n+1}^{n+q^*} x_i, \varphi)]
\]

**Proposition 2.3.3.** A solution to the firm’s entry and exit problem exists.

**Proposition 2.3.4.** The optimal number of markets entered into simultaneously is increasing in \( X \), the sum of log product appeals.

### 2.3.1 Variation in Entry Costs

Suppose that markets differ only in terms of sunk entry costs \( f_{ne} \), while fixed costs of operation, size, and price indices are constant across markets.

Note that the productivity/demand threshold for exit is the same for each market in this setup: there exists a unique \( A > 0 \) such that for each market, entrants with \( x\varphi \geq A \) will remain active, otherwise exiting. Moreover, conditional on entry, all markets have the same relationship between product appeal, productivity, and per period profit. However, the net present value of entry includes the sunk entry cost, which varies across markets, hence assumption 1 is satisfied. Without loss of generality, the markets are ordered by \( f_{ne} \), \( i < j \) implies that \( f_{ie} < f_{je} \). By
proposition 1, firms will enter low \( n \) markets first. Denote the market with the lowest sunk cost of entry as the home market.

As price levels and fixed operating costs are identical across markets, for a given distribution of firms, exit rates should be the same across markets. However the distribution of entering firms will vary across markets. For example, firms that enter market \( N \) expect to obtain a discounted profit stream greater than the sunk cost of entry \( f_{Ne} \), whereas the required expected discounted profit stream for all markets \( n < N \) is strictly less than \( f_{Ne} \).

In addition, as \( n \) increases, the value of an additional product appeal draw will tend to decrease. First, there are fewer markets for which the information is relevant. Second, the additional draw provides less additional information (refinement of beliefs) than preceding draws. This further supports the notion that firms entering high \( n \) markets have a more favorable belief regarding \( F_\omega(\alpha) \).

The following diagram, Figure 2.3, looks at the problem of a firm that has entered a single market, with two markets available. It can be seen that for low realizations of product appeal in the first market, the firm will optimally choose to not expand. However as realizations improve the posterior belief regarding \( \alpha \) will enable the firm to expand into other markets; as observed demand increases the firm first expands into the second market (retaining the option to enter the third market subsequently), then it becomes optimal to simultaneously enter both remaining markets.

**Proposition 2.3.5.** If markets vary only by entry costs, exit rates are decreasing in the sunk entry cost \( f_{ne} \).

The model predicts that, if markets only differ by sunk entry costs, the markets entered into later on should have higher survival rates among entrants than markets entered earlier on. However, that is not what we observe with Chinese firms exporting to the EU. As shown in Figure 2.15, exit rates are in fact increasing as we move down the ladder.
Figure 2.3: Observations affect entry decisions
2.3.2 Variation in Market Size

Suppose instead that markets vary by market size $Y_n$ (or, less restrictively, market potential $\frac{Y_n}{P_n}$). Assumption 1 is satisfied; firms enter the larger markets first. This is consistent with what is observed with Chinese firms exporting to the EU: entry order is highly correlated\(^{20}\) with market size, as is total entry.\(^{21}\)

The relationship between entry order and exit rates is more flexible than when markets differ by entry costs. From equation 6, the cutoff, below which firms will exit, is decreasing as market size increases; it follows that the cutoff increases as we move from the markets entered early on to the markets entered later. As shown in Figure 2.4, for a given level of productivity, the set of values of demand $x$ which support survival are decreasing between the large market (threshold locus denoted by $M_1$) and the small market (threshold locus denoted by $M_2$). Thus the direct effect of market size implies that smaller markets feature higher exit rates.

\[ \text{Figure 2.4: Exit Thresholds in productivity/demand space} \]

However, similar to the previous section, the distribution of firms in terms of productivity and demand improves as we move down the ladder: the distributions of firms entering the smaller markets first order stochastically dominate those of the larger markets. As a consequence exit rates may in fact decrease as market size

\(^{20}\)Although not perfectly - not all firms choose the same market as their first. However as shown in Figure 2.10, a hierarchy still exists.

\(^{21}\)As seen in Figure 2.16, the rank of market size, by GDP, is highly correlated with various measures of entry rank. Included are ranks by the destination choice of new entrants, for different speeds of expansion in the first year, and the total entry by the 2001 cohort over the 6 years they are observed. In addition, the rank by the average level of experience of entrants, in terms of markets previously entered, is included.
decreases; there is a discrete change in the distribution regardless of the market size, while the change in the cutoff may be very small if the market sizes are similar.

As such, this setup allows for both the negative trend seen, as well as the non-monotonicity: if two markets are similar the selection effect may dominate, otherwise the direct effect may dominate.

2.3.3 Market Rank Effects

The model, with learning occurring as markets are entered, suggests that, ceteris paribus, the rank of the market is important. Even if two markets are virtually identical, slight differences will lead to different positions in the hierarchy of markets. Due to the sequential nature of expansion, the rank will play a large role in determining entry, exit, and survival.

Consider two markets, identical except for market potential. Let market \( A \) have a size of \( y \), while market \( B \) has a size of \( y - \varepsilon \), so that firms prefer \( A \) to \( B \). Note that \( \forall \varepsilon > 0 \), \( A \) is ranked ahead of \( B \), i.e. firms that enter \( B \) must have entered \( A \). As we take the limit of \( \varepsilon \to 0 \), \( A \) is identical to \( B \) except in rank.

However the patterns of entry and exit will be different; the difference in rank plays an important role in determining entry when learning is allowed. While all firms that enter \( B \) have entered \( A \), by Proposition 1, it is not true that every firm that enters \( A \) will also enter \( B \). Firms face a different value from entering \( B \), due to the information being less beneficial. Not only is the information applicable to fewer markets, but firms would usually have received more draws when considering entering \( B \) rather than \( A \); the marginal refinement in beliefs would be lower. In addition, firms that enter \( A \) may simply observe a low product appeal, leading to a cessation of expansion.

While fewer firms enter \( B \) than \( A \), the distribution of firms is better for \( B \): the firms that perform poorly in \( A \) do not continue to \( B \). Thus the firms which enter \( B \) are less numerous, but better, than those which enter \( A \). As a result the slightly larger market has more entry, more firms serving the market, and higher exit rates; while the two markets are fundamentally very similar, the difference in rank interacts with learning to produce significantly different outcomes.
2.4 Evidence

The implications of the model are now tested.

2.4.1 Data

Two sets of data are utilized: the Centre d’Etudes Prospectives et d’Informations Internationales (CEPII) gravity dataset and the Chinese Customs dataset.

2.4.1.1 Trade Data

The CEPII gravity dataset provides bilateral trade flows for all pairs of countries for the period 1948 to 2006. I restrict attention to trade flows during 2001. Data is restricted to observations where trade flows are non-negative.\textsuperscript{22}

2.4.1.2 Customs Data

The Chinese Customs dataset includes all shipments from China during the period 2000 to 2006. We observe the date, destination, 8-digit HS code, quantity, and value for each shipment. Importantly, we observe the firm ID, so that we can track the entry and exit behavior of firms over time. Shipments are aggregated to yearly data, processing trade\textsuperscript{23,24} and trade intermediaries are removed. For the purposes of analysis, I treat each firm-product combination as being a different entity. This rules out interactions between products within the same firm: any information obtained through entry into a market is not applicable across products. As a result, products within a firm are modeled to act as an independent firm.\textsuperscript{25}

Sheveleva (2012), a related paper, takes the opposite approach: outcomes for one product are correlated with outcomes in related products.

\textsuperscript{22}The dataset is available at http://www.cepii.fr/anglaisgraph/bdd/gravity/col_regfile09.zip. The description can be found at http://www.cepii.fr/anglaisgraph/bdd/gravity.htm

\textsuperscript{23}Processing firms are responsible for about one-third of Chinese exports and considerably less productive, as shown by Dai et al. (2011), consistent with the comparison between domestic firms and exporters found by Lu (2010).

\textsuperscript{24}Processing firms could be expected to have similar dynamics, for example instead of learning about demand, potential customers learn about the firm’s reliability, quality etc, (see LeCates (2011)) so that firms may procure a contract (enter a market) in order to gain customers. However we cannot observe the recipient of a shipment, and there may be multiple customers being served within a country.

\textsuperscript{25}I do allow clustering by firms however.
2.4.2 Gravity Equation

The learning model suggests that the rank of the destination market should have an impact on the trade flows, even when controlling for other market parameters, such as market size. However, the rank of the origin country is not predicted to be relevant.\textsuperscript{26} The following table (Table 1) gives the results of gravity equation regressions, both with and without terms controlling for the market size rank of both the origin and destination.\textsuperscript{27} Robust standard errors are used.

Trade flows from the origin to the destination are regressed on all usual gravity model terms, including the log of GDP of both the origin and the destination, the log of the bilateral distance, and a number of dummy variables indicating a shared border, a historical colonial relationship, a shared official language and a shared common currency. GATT membership and regional trade agreements are also controlled for, as are per capita output levels and population sizes. The equation to be estimated is:

\[
\ln \text{Flow}_{od} = \beta_1 \ln Y_o + \beta_2 \ln Y_d + \beta_3 \ln d_{od} + \beta_4 I(\text{contig.}) + \\
+ \beta_4 \ln \text{Rank}_d + \beta_5 \ln \text{Rank}_o + \gamma X_{od} + \epsilon_{od}
\]

(2.13)

where \(X_{od}\) includes various aforementioned indication variables.

Markets are ranked by GDP (among trading partners), a low value implies a large GDP. The effect of rank is predicted to be more important for markets that are entered early on for two reasons: the information gained from entry is applicable to more markets (as there are more markets remaining), and the clarity provided by an additional draw diminishes as firms have more draws. As such, the log of the rank of the destination market (rank is in GDP terms) is used in the regression.

As usual, we see that market size is highly significant for both the origin and the destination, with coefficients close to unity when not accounting for size ranks. The dummy variables are positive (as would be expected), and significant, with

\textsuperscript{26}While larger markets would have larger inflows, potentially increasing productivity, output and outflows, this effect will be captured by market size measured by GDP.

\textsuperscript{27}The log of the rank is used, as rank is predicted to be less important as we look at markets which are entered later on.
the exception of the common currency variable.

As predicted by the model, even when controlling for GDP, the rank of the destination market is important, especially among the largest markets. The coefficient is negative, highly significant, and quantitatively important. Moving from the second largest market to the largest market increases incoming trade flows by around fifty percent, while moving from the tenth to the ninth increases flows by ten percent.

As the destination market size rank is correlated with market size, it is natural to find that it is significant. Possible reasons could include non-linearity of the relationship, imperfect measurement of output etc. However, if the strong relationship is simply an artifact, the result of simply being a moment of market size, one would expect the origin market size rank to have a similar relationship with trade flows. However this is not observed, the coefficient is both small and insignificant. Hence this effect of rank is unique to the destination market, consistent with the model.

\footnote{This result is robust to other specifications, such as fixed effects models using panel data.}
\footnote{The coefficient and significance level is similar when ignoring flows from the largest markets such as USA, Japan, Germany, UK, France and China.}
\footnote{With the exception of specification (2), a very basic specification.}
\footnote{Significant at the 10\% level in specification 4.}
Table 2.1: Estimation results: Gravity

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(gdp) orig.</td>
<td>1.181***</td>
<td>1.104***</td>
<td>1.170***</td>
<td>1.138***</td>
<td>1.132***</td>
<td>1.145***</td>
</tr>
<tr>
<td></td>
<td>(156.9)</td>
<td>(44.58)</td>
<td>(150.4)</td>
<td>(45.74)</td>
<td>(102.6)</td>
<td>(45.51)</td>
</tr>
<tr>
<td>log(gdp) dest.</td>
<td>0.919***</td>
<td>0.596***</td>
<td>0.916***</td>
<td>0.635***</td>
<td>0.878***</td>
<td>0.630***</td>
</tr>
<tr>
<td></td>
<td>(126.5)</td>
<td>(27.61)</td>
<td>(124.4)</td>
<td>(29.84)</td>
<td>(84.51)</td>
<td>(29.25)</td>
</tr>
<tr>
<td>log(Distance)</td>
<td>-1.390***</td>
<td>-1.403***</td>
<td>-1.228***</td>
<td>-1.252***</td>
<td>-1.23***</td>
<td>-1.245***</td>
</tr>
<tr>
<td></td>
<td>(-65.43)</td>
<td>(-67.09)</td>
<td>(-52.13)</td>
<td>(-53.39)</td>
<td>(-51.61)</td>
<td>(-52.60)</td>
</tr>
<tr>
<td>Contiguous</td>
<td>1.232***</td>
<td>1.210***</td>
<td>0.862***</td>
<td>0.882***</td>
<td>0.915***</td>
<td>0.908***</td>
</tr>
<tr>
<td>Colonial Hist.</td>
<td>1.065***</td>
<td>0.889***</td>
<td>1.056***</td>
<td>0.900***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(11.92)</td>
<td>(9.888)</td>
<td>(11.82)</td>
<td>(10.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Lang.</td>
<td>0.736***</td>
<td>0.691***</td>
<td>0.717***</td>
<td>0.693***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(15.22)</td>
<td>(14.36)</td>
<td>(14.79)</td>
<td>(14.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GATT orig.</td>
<td>0.332***</td>
<td>0.344***</td>
<td>0.290***</td>
<td>0.300***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.532)</td>
<td>(6.804)</td>
<td>(5.652)</td>
<td>(5.859)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GATT dest.</td>
<td>0.123***</td>
<td>0.145***</td>
<td>0.0927***</td>
<td>0.146***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.730)</td>
<td>(3.241)</td>
<td>(2.018)</td>
<td>(3.186)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTA</td>
<td>0.865***</td>
<td>0.826***</td>
<td>0.856***</td>
<td>0.826***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(14.27)</td>
<td>(13.77)</td>
<td>(13.93)</td>
<td>(13.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Curr.</td>
<td>0.161</td>
<td>0.000858</td>
<td>0.124</td>
<td>0.00171</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.232)</td>
<td>(0.00663)</td>
<td>(0.948)</td>
<td>(0.0131)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP per cap dest.</td>
<td>1.28e-05***</td>
<td>-2.23e-08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.538)</td>
<td>(-0.00888)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP per cap orig.</td>
<td>1.05e-05***</td>
<td>9.58e-06***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.193)</td>
<td>(4.468)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pop. orig.</td>
<td>0.000526***</td>
<td>0.000485***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.816)</td>
<td>(6.186)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pop. dest.</td>
<td>0.000373***</td>
<td>-0.000191</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.032)</td>
<td>(-1.484)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(size rank) dest.</td>
<td>-0.775***</td>
<td>-0.673***</td>
<td></td>
<td></td>
<td>-0.696***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-17.04)</td>
<td>(-15.01)</td>
<td></td>
<td></td>
<td>(-13.79)</td>
<td></td>
</tr>
<tr>
<td>log(size rank) orig.</td>
<td>-0.190***</td>
<td>-0.084*</td>
<td></td>
<td></td>
<td>0.0127</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-3.934)</td>
<td>(-1.722)</td>
<td></td>
<td></td>
<td>(0.242)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-8.647***</td>
<td>-0.732</td>
<td>-10.48***</td>
<td>-4.174***</td>
<td>-9.900***</td>
<td>-4.587***</td>
</tr>
<tr>
<td></td>
<td>(-39.29)</td>
<td>(-1.215)</td>
<td>(-46.89)</td>
<td>(-6.915)</td>
<td>(-40.78)</td>
<td>(-7.206)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.636</td>
<td>0.641</td>
<td>0.647</td>
<td>0.651</td>
<td>0.650</td>
<td>0.653</td>
</tr>
</tbody>
</table>

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1
2.4.3 Firm Dynamics

Using the Chinese Customs data, I examine the determinants of survival, in addition to expansion behavior.

2.4.3.1 Survival

The probability that a firm is successful is examined. Here, success is measured by surviving in a market: profitable firms survive, continuing to export to a market, while other firms fail to survive. I define survival to be continuing to export to a market in the year subsequent to entry, in accordance with the observation that most exit occurs between the first and second years.

Table 2 below gives the results of a probit regression of survival for all new firm/product/destination observations. Age is the firm’s age in years since the first export observation, experience is the number of markets entered previously, and expansion is the number of markets entered into this period, i.e. simultaneously (recall that a period is a calendar year). Relative price is the price that the firm sets, relative to other firms in the market (market in this sense is the destination/product combination, at the HS8 level). Sales history is the average of first year sales in markets entered previously (relative to other firms in each market). Taking the CES demand structure, given price and sales it is possible to back out the demand coefficient - given the relative price has a firm sold more or less than a firm with an average demand shock. Survival history is the share of markets entered previously that have been successful i.e. the firm continues to export in the year subsequent to entry. Errors are clustered by firm, allowing for possible firm level effects such as capital constraints.

The age of the firm is first seen to be mildly negative (and significant) consistent with older firms being less profitable, conditional on experience, since good firms tend to expand quickly. However adding more regressors describing the state of the firm causes the effect to be insignificant. This is consistent with the model - the value function, and therefore decisions/outcomes, do not explicitly depend on the age of the firm - they depend on beliefs, which are only correlated with the age (conditional on experience).

As firms enter more markets they are able to form a more accurate belief
regarding profitability/demand; this is reflected in the coefficient on experience. Firms with more experience are less likely to fail, consistent with firms that possess more information being able to avoid entry when the chance of success is low (ex-post).

The model implies that simultaneous entry occurs when a firm believes that the expected profits from entry are high enough to justify entering with limited information. Therefore, the number of markets expanded to in a given year should be associated with a higher chance of survival. The probit regression shows this to be the case: firms that choose to rapidly expand are much more likely to survive in the destination market, at the rate of ten percent for each market entered simultaneously.

Sales history appears to be insignificant in specifications 2 and 3; sales are a noisy measure of product appeal, which is what matters in the model. Demand history is positive and significant - firms which have observed greater demand are more likely to survive. Finally, the share of markets entered in which survival is observed is positive, significant and reasonably large: firms that have survived previously are more likely to survive subsequent market entries.
Table 2.2: Estimation results: Survival Probit

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Survive</th>
<th>(2) Survive</th>
<th>(3) Survive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.0325**</td>
<td>-0.0311**</td>
<td>-0.000205</td>
</tr>
<tr>
<td></td>
<td>(-2.180)</td>
<td>(-2.101)</td>
<td>(-0.0165)</td>
</tr>
<tr>
<td>Experience</td>
<td>0.0874***</td>
<td>0.0871***</td>
<td>0.0714***</td>
</tr>
<tr>
<td></td>
<td>(10.71)</td>
<td>(10.70)</td>
<td>(9.377)</td>
</tr>
<tr>
<td>Experience^2</td>
<td>-0.00105</td>
<td>-0.00122</td>
<td>-0.000919</td>
</tr>
<tr>
<td></td>
<td>(-1.373)</td>
<td>(-1.587)</td>
<td>(-1.221)</td>
</tr>
<tr>
<td>Expansion</td>
<td>0.142***</td>
<td>0.143***</td>
<td>0.126***</td>
</tr>
<tr>
<td></td>
<td>(23.63)</td>
<td>(23.73)</td>
<td>(20.88)</td>
</tr>
<tr>
<td>Relative Price</td>
<td>0.00982***</td>
<td>0.00480*</td>
<td>0.00319</td>
</tr>
<tr>
<td></td>
<td>(3.535)</td>
<td>(1.829)</td>
<td>(1.240)</td>
</tr>
<tr>
<td>Sales History</td>
<td>0.0174***</td>
<td>-0.00138</td>
<td>-0.00575*</td>
</tr>
<tr>
<td></td>
<td>(6.366)</td>
<td>(-0.432)</td>
<td>(-1.828)</td>
</tr>
<tr>
<td>Demand History</td>
<td>0.0510***</td>
<td>0.0392***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(10.48)</td>
<td>(8.715)</td>
<td></td>
</tr>
<tr>
<td>Survival History</td>
<td></td>
<td></td>
<td>0.351***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(19.13)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.817***</td>
<td>-0.868***</td>
<td>-1.116***</td>
</tr>
<tr>
<td></td>
<td>(-3.457)</td>
<td>(-3.678)</td>
<td>(-4.662)</td>
</tr>
<tr>
<td>HS2 Dummies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Market Dummies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>165,821</td>
<td>165,821</td>
<td>165,821</td>
</tr>
</tbody>
</table>

Robust z-statistics in parentheses

*** p < 0.01, ** p < 0.05, * p < 0.1

2.4.3.2 Expansion

Attention is now paid to the expansion behavior: what determines the likelihood that a firm will expand, and, if so, how aggressively i.e. how many markets would they enter in the same period. Restricting the sample to firm/product combinations that expanded to an EU market in the previous year, Table 3 gives the results of a probit of Expand on various regressors, where Expand is equal to unity if the firm/product continues to expand within the EU, entering a new destination, subsequent to entering a new market the previous period.

We see the expected relationship with experience: initially more experience increases the likelihood of expansion, however as firms move down the ladder the remaining markets are less attractive (and fewer in number) so that they become
less likely to continue to expand.

The effect of age is significantly negative. While this is not entirely unexpected, in the model age should have no impact on choices, conditional on the state of the firm. However, if some aspect of the firm’s beliefs is not controlled for, for example productivity, a negative coefficient could persist as the productive firms would be expected to expand quickly, and continue to expand. Note that in the previous regression the term `expansion` served to control for the firm’s beliefs. However, including the expansion of the previous year reduces the effect of age substantially.

As before, units sold is positively correlated with outcomes, while the more coarse measure of success, survival, is highly correlated with a firm continuing to expand.

Conditional on expansion, firms choose how many markets to enter into simultaneously. Running a simple regression of `expansion` on the same variables, a similar pattern to the previous regression is seen in Table 4. The effect of experience is at first positive, then negative: initially more experience allows the firm to have more accurately know their expected demand shocks, however as experience increases the number of attractive markets diminishes. The effect of age is, as before, negative. The effect of last year’s expansion is positive, implying that expansion speed has some degree of persistence. Survival strongly predicts expansion speed, consistent with the model. We do however see that firms with lower demand tend to expand faster, conditional on expansion. A possible explanation for this is that such firms would likely have a higher productivity, given that they are choosing to expand despite low demand. As productivity is persistent across markets, learning about demand may not be as important as firms which would rely on high demand in order to generate profits.

Results of a ordinal logit regression are presented in Table 7, with similar findings.
Table 2.3: Estimation results: Expand Probit

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Expand</th>
<th>(2) Expand</th>
<th>(3) Expand</th>
<th>(4) Expand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.0924***</td>
<td>-0.0913***</td>
<td>-0.0918***</td>
<td>-0.0179**</td>
</tr>
<tr>
<td></td>
<td>(-13.89)</td>
<td>(-13.67)</td>
<td>(-13.81)</td>
<td>(-2.489)</td>
</tr>
<tr>
<td>Experience</td>
<td>0.504***</td>
<td>0.504***</td>
<td>0.443***</td>
<td>0.352***</td>
</tr>
<tr>
<td></td>
<td>(95.71)</td>
<td>(95.84)</td>
<td>(88.62)</td>
<td>(52.48)</td>
</tr>
<tr>
<td>Experience$^2$</td>
<td>-0.0360***</td>
<td>-0.0361***</td>
<td>-0.0343***</td>
<td>-0.0299***</td>
</tr>
<tr>
<td></td>
<td>(-72.92)</td>
<td>(-73.44)</td>
<td>(-72.24)</td>
<td>(-58.08)</td>
</tr>
<tr>
<td>Sales History</td>
<td>0.0258***</td>
<td>0.0108***</td>
<td>0.00138</td>
<td>0.00148</td>
</tr>
<tr>
<td></td>
<td>(18.29)</td>
<td>(6.414)</td>
<td>(0.789)</td>
<td>(0.849)</td>
</tr>
<tr>
<td>Demand History</td>
<td>0.0510***</td>
<td>0.00374</td>
<td>0.00278</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(18.26)</td>
<td>(1.465)</td>
<td>(1.085)</td>
<td></td>
</tr>
<tr>
<td>Survival Share</td>
<td>1.065***</td>
<td>1.068***</td>
<td>1.068***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(115.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prev. Expansion</td>
<td>0.0805***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(17.08)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.545***</td>
<td>-1.597***</td>
<td>-1.960***</td>
<td>-1.960***</td>
</tr>
<tr>
<td></td>
<td>(-16.87)</td>
<td>(-17.43)</td>
<td>(-18.65)</td>
<td>(-18.42)</td>
</tr>
<tr>
<td>HS2 Dummy</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>534,567</td>
<td>534,567</td>
<td>534,567</td>
<td>534,567</td>
</tr>
</tbody>
</table>

Robust z-statistics in parentheses

*** p < 0.01, ** p < 0.05, * p < 0.1
Table 2.4: Estimation results: Expansion

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Expansion</th>
<th>Expansion</th>
<th>Expansion</th>
<th>Expansion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.190***</td>
<td>-0.190***</td>
<td>-0.156***</td>
<td>-0.0931***</td>
</tr>
<tr>
<td></td>
<td>(-25.11)</td>
<td>(-25.17)</td>
<td>(-21.12)</td>
<td>(-11.56)</td>
</tr>
<tr>
<td>Experience</td>
<td>0.244***</td>
<td>0.244***</td>
<td>0.212***</td>
<td>0.139***</td>
</tr>
<tr>
<td></td>
<td>(40.64)</td>
<td>(40.63)</td>
<td>(35.22)</td>
<td>(19.82)</td>
</tr>
<tr>
<td>Experience^2</td>
<td>-0.0216***</td>
<td>-0.0216***</td>
<td>-0.0205***</td>
<td>-0.0171***</td>
</tr>
<tr>
<td></td>
<td>(-42.86)</td>
<td>(-42.87)</td>
<td>(-41.27)</td>
<td>(-32.82)</td>
</tr>
<tr>
<td>Sales History</td>
<td>0.0134***</td>
<td>0.0132***</td>
<td>0.00425</td>
<td>0.00445*</td>
</tr>
<tr>
<td></td>
<td>(5.632)</td>
<td>(4.830)</td>
<td>(1.602)</td>
<td>(1.686)</td>
</tr>
<tr>
<td>Demand History</td>
<td></td>
<td>0.000512</td>
<td>-0.0202***</td>
<td>-0.0212***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.151)</td>
<td>(-6.098)</td>
<td>(-6.432)</td>
</tr>
<tr>
<td>Survival Share</td>
<td></td>
<td></td>
<td>0.603***</td>
<td>0.602***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(51.58)</td>
<td>(51.65)</td>
</tr>
<tr>
<td>Prev. Expansion</td>
<td></td>
<td></td>
<td></td>
<td>0.0618***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(14.97)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.435***</td>
<td>1.434***</td>
<td>1.085***</td>
<td>1.087***</td>
</tr>
<tr>
<td></td>
<td>(9.012)</td>
<td>(9.008)</td>
<td>(7.734)</td>
<td>(7.759)</td>
</tr>
<tr>
<td>HS2 Dummy</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>112,405</td>
<td>112,405</td>
<td>112,405</td>
<td>112,405</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.037</td>
<td>0.037</td>
<td>0.076</td>
<td>0.078</td>
</tr>
</tbody>
</table>

Robust t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1
2.5 Extension: Differing Correlations

In this section I relax the assumption that the correlation of demand shocks is the same for any pair of markets, explain the implications, and present supporting reduced form evidence. While this would complicate the model to a great extent, significantly increasing the difficulty of the firm’s problem, there are some clear implications.

It is be natural to expect that the correlation between outcomes in developed economies is greater than that between a developed economy and a developing economy, or perhaps between two culturally similar countries. It follows that the information gained after entry will be more relevant to certain markets than to others.\textsuperscript{32}

Consider a new firm that will enter a particular market, and is considering whether to simultaneously enter other markets. What is the relationship between the correlation of outcomes between two markets and the decision to enter simultaneously?

Suppose the firm is going to enter market \( A \). As the correlation between \( A \) and a prospective market \( B \) approaches zero, the benefit of delaying entry into \( B \) approaches zero: information obtained through entry into \( A \) becomes irrelevant to the decision to enter \( B \). Since the outcomes are not correlated, the outcome in market \( A \) does not affect the firm’s beliefs regarding \( B \). On the other hand, suppose that the markets are in fact highly correlated. In that case, the information from \( A \) is quite valuable, it follows that the benefit of delaying entry into \( B \) is high; the outcome in market \( A \) is very informative of what the firm can expect from market \( B \).

This implies that when a pair of markets are highly correlated in terms of outcomes, the likelihood of simultaneous entry should be low for new exporters.\textsuperscript{33} In order to test for this I regress the correlation of entry among new exporters\textsuperscript{34}

---

\textsuperscript{32} Recent work by Defever et al. (2011) shows that firms tend to enter markets that are close (in terms of geography and culture) to their current destinations.

\textsuperscript{33} This would not hold in general for experience exporters. Entry into two closely related markets may occur simultaneously as firms have already entered a third closely related market.

\textsuperscript{34} As previously, a firm is interpreted as a firm-HS8 combination, simultaneous entry is defined as the same calendar year.
for each market pair in the EU (excluding Luxembourg\textsuperscript{35}) on various regressors from the CEPII gravity dataset, the difference in GDP per capita levels (the ratio of the highest of the pair to the lowest of the pair), and the absolute difference in GDP rank. In addition I include correlations of outcomes across markets: survival and fitted demand shifter, by both new firms and all firms.

For each country pair, I find the correlation of entry into the new markets among new exporters: observing for each firm a pair of entry indicators, in \{0, 1\} × \{0, 1\} space. If the pairwise correlation of these indicators is high, firms that enter one of the pair in their first year tend to enter the other in the same year.

Table 2.5 shows when the outcomes are highly correlated, the likelihood of simultaneous entry is lower: the correlation of entry among new firms is lower. A standard deviation increase in the outcome correlation decreases the correlation of entry by \(0.19 - 0.27\) standard deviations. This is something that a traditional model would have difficulty explaining: if outcomes are correlated so should entry. It is also seen the a shared official language is positively associated with simultaneous entry, perhaps due to interaction with the costs of entry.

\textsuperscript{35}There are too few firms that enter Luxembourg and other markets in the first year
Table 2.5: Estimation results: Entry Correlation

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contiguous</td>
<td>-0.00372</td>
<td>0.00230</td>
<td>0.000594</td>
<td>-0.00122</td>
<td>-7.71e-05</td>
</tr>
<tr>
<td></td>
<td>(-0.350)</td>
<td>(0.211)</td>
<td>(0.0548)</td>
<td>(-0.112)</td>
<td>(-0.00723)</td>
</tr>
<tr>
<td>Common Lang.</td>
<td>0.0367***</td>
<td>0.0323***</td>
<td>0.0330***</td>
<td>0.0331***</td>
<td>0.0300***</td>
</tr>
<tr>
<td></td>
<td>(3.909)</td>
<td>(3.456)</td>
<td>(3.569)</td>
<td>(3.425)</td>
<td>(3.114)</td>
</tr>
<tr>
<td>log(Distance)</td>
<td>0.00512</td>
<td>0.00324</td>
<td>0.00356</td>
<td>0.00667</td>
<td>0.00745</td>
</tr>
<tr>
<td></td>
<td>(0.717)</td>
<td>(0.447)</td>
<td>(0.490)</td>
<td>(0.937)</td>
<td>(1.067)</td>
</tr>
<tr>
<td>GDP per cap. Diff.</td>
<td>0.00369</td>
<td>-0.00120</td>
<td>-3.32e-05</td>
<td>-0.00161</td>
<td>-0.00207</td>
</tr>
<tr>
<td></td>
<td>(0.690)</td>
<td>(-0.227)</td>
<td>(-0.00615)</td>
<td>(-0.288)</td>
<td>(-0.388)</td>
</tr>
<tr>
<td>Rank Diff.</td>
<td>-0.000232</td>
<td>-0.00117*</td>
<td>-0.00109</td>
<td>-0.000721</td>
<td>-0.000654</td>
</tr>
<tr>
<td></td>
<td>(-0.304)</td>
<td>(-1.697)</td>
<td>(-1.524)</td>
<td>(-0.936)</td>
<td>(-0.883)</td>
</tr>
<tr>
<td>Surv. Corr. (All)</td>
<td>-0.197***</td>
<td>-0.149***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-3.600)</td>
<td>(-3.105)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surv. Corr. (New)</td>
<td>-0.197***</td>
<td>-0.149***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-3.600)</td>
<td>(-3.105)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demand Corr. (All)</td>
<td>-0.0406***</td>
<td>-0.0413***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-3.895)</td>
<td>(-3.995)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demand Corr. (New)</td>
<td>-0.0406***</td>
<td>-0.0413***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-3.895)</td>
<td>(-3.995)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.0273</td>
<td>0.0526</td>
<td>0.0656</td>
<td>-0.0124</td>
<td>-0.0119</td>
</tr>
<tr>
<td></td>
<td>(-0.536)</td>
<td>(1.001)</td>
<td>(1.159)</td>
<td>(-0.244)</td>
<td>(-0.239)</td>
</tr>
<tr>
<td>Observations</td>
<td>182</td>
<td>182</td>
<td>182</td>
<td>182</td>
<td>182</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.062</td>
<td>0.119</td>
<td>0.095</td>
<td>0.091</td>
<td>0.106</td>
</tr>
</tbody>
</table>

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1
2.6 Policy Implications

The model implies that an attempt to encourage entry into a market will have delayed effects for the target market. Consider a drop in the sunk entry cost of market \( n \) at time \( t = 1 \). Firms that have entered markets \( \{1, \ldots, n-1\} \) will either immediately enter market \( n \), or will never enter market \( n \); there will be an immediate effect. Similarly, firms which have not yet entered market \( n-1 \) are still able to enter market \( n \) by expanding simultaneously: there will be an increased likelihood that they expand to market \( n \) (and all remaining markets \( m \) such that \( m < n \)) immediately.

However, that is not the only impact of increasing the attractiveness: the value function is increased for all levels of experience less than \( n \). The increased profitability in market \( n \) will drive entry into lower ranked markets: firms are more willing to incur entry costs with the possibility of expansion into market \( n \). This will, with time, increase the amount of entry into market \( n \). However, these effects are delayed: it will take \( n-1 \) periods for the full impact to be realized.

Additionally, for small changes, liberalization in one market increases entry for all other markets. As before, increasing the attractiveness of market \( n \) will increase entry into preceding markets. Similarly, there would be more entry into markets \( n+1, \ldots, N \) - there will be more firms that have expanded to \( n \) markets, therefore more firms considering entry into \( n+1 \), and so on. The time taken for the full effect to be realized depends on the rank of the market: it will take \( m-1 \) periods for the effect on market \( m \) to be fully realized.

While small changes would generate spillovers for all markets, instead of gains at the expense of others (similar to Cherkashin et al. (2010)), a large change can have a different outcome. Consider a large increase in the attractiveness of market \( n \), e.g. an increase in market size. If the change is large enough, it may improve the rank of market \( n \), at the expense of market \( n-1 \). As a result of the change in rank, entry into market \( n \) would rise significantly, while market \( n-1 \) would see entry, and trade flows, drop significantly. Therefore, development within a country may have negative impacts on neighbouring countries - the developing country could increase its rank at the expense of a spatially\(^{36}\) close country.

\(^{36}\)Although spatially in the model would be in \((f_{ne}, f_{no}, \frac{\lambda}{\mu - \sigma})\) space, countries that are in...
2.7 Conclusions

This model shows that, with correlated profitability across markets, beliefs, and updating, the decision to enter a market is not made in isolation. The information gained from entry has important implications regarding entry into other markets, and gives rise to unique behavior.

Reduced form results show support for the model, finding that past experiences are vital in guiding firms towards their expansion decisions. Firms which perform well tend to expand quickly, forgoing the benefit of information due to discounting, which corresponds to a positive relationship between expansion speed and outcomes within a market.

The relationship between markets is found to be important: a market’s rank relative to competing markets is important in determining incoming trade flows, especially for countries with large market shares.
2.8 Appendix

Figure 2.5: Expansion within the EU, given number of markets previously entered

Figure 2.6: Propensity to expand within the EU, given number of markets previously entered
Figure 2.7: World expansion, given number of markets previously entered

Figure 2.8: Propensity to expand, given number of markets previously entered
Figure 2.9: Expansion behavior and firm age

Figure 2.10: Destination choice within the EU, conditional on experience
Figure 2.11: World wide destination choice, conditional on experience

Figure 2.12: Likelihood of entry and survival within the EU, for new firms expanding to a single market
Figure 2.13: Likelihood of entry and survival within the EU, for new firms expanding to two markets

Figure 2.14: Likelihood of entry and survival within the EU, for new firms expanding to three markets
Figure 2.15: Likelihood of entry and survival within the EU, for firms in the 2001 cohort

Figure 2.16: Correlation between market GDP rank and various entry ranks
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs.</td>
<td>$n_t/n_{t-1}$</td>
<td>Obs.</td>
</tr>
<tr>
<td>Entry</td>
<td>65334</td>
<td>$\cdot$</td>
<td>81755</td>
</tr>
<tr>
<td>Survive 1 yr</td>
<td>19465</td>
<td>0.298</td>
<td>27607</td>
</tr>
<tr>
<td>Survive 2 yr</td>
<td>12726</td>
<td>0.654</td>
<td>19292</td>
</tr>
<tr>
<td>Survive 3 yr</td>
<td>10096</td>
<td>0.793</td>
<td>15313</td>
</tr>
<tr>
<td>Survive 4 yr</td>
<td>8221</td>
<td>0.814</td>
<td>12145</td>
</tr>
<tr>
<td>Survive 5 yr</td>
<td>6704</td>
<td>0.815</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.6: Firm/destination survival over time, EU, conditional on firm cohort and year of entry into the destination market

Table 2.7: Ordinal Logit Regression: Expansion

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Expansion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.0546***</td>
</tr>
<tr>
<td></td>
<td>(-4.476)</td>
</tr>
<tr>
<td>Experience</td>
<td>0.646***</td>
</tr>
<tr>
<td></td>
<td>(61.94)</td>
</tr>
<tr>
<td>Experience^2</td>
<td>-0.0568***</td>
</tr>
<tr>
<td></td>
<td>(-66.26)</td>
</tr>
<tr>
<td>Sales History</td>
<td>0.00323</td>
</tr>
<tr>
<td></td>
<td>(1.032)</td>
</tr>
<tr>
<td>Demand History</td>
<td>-0.00378</td>
</tr>
<tr>
<td></td>
<td>(-0.879)</td>
</tr>
<tr>
<td>Survival History</td>
<td>1.937***</td>
</tr>
<tr>
<td></td>
<td>(113.1)</td>
</tr>
<tr>
<td>Prev. Expansion</td>
<td>0.144***</td>
</tr>
<tr>
<td></td>
<td>(21.33)</td>
</tr>
<tr>
<td>HS2 Dummy</td>
<td>Yes</td>
</tr>
<tr>
<td>Cutoffs</td>
<td></td>
</tr>
<tr>
<td>Cut1</td>
<td>3.414***</td>
</tr>
<tr>
<td></td>
<td>(18.85)</td>
</tr>
<tr>
<td>Cut2</td>
<td>4.626***</td>
</tr>
<tr>
<td></td>
<td>(25.55)</td>
</tr>
<tr>
<td>Cut3</td>
<td>5.562***</td>
</tr>
<tr>
<td></td>
<td>(30.69)</td>
</tr>
<tr>
<td>Cut4</td>
<td>6.396***</td>
</tr>
<tr>
<td></td>
<td>(35.28)</td>
</tr>
<tr>
<td>Cut5</td>
<td>7.184***</td>
</tr>
<tr>
<td></td>
<td>(39.48)</td>
</tr>
<tr>
<td>Cut6</td>
<td>7.943***</td>
</tr>
<tr>
<td></td>
<td>(43.51)</td>
</tr>
<tr>
<td>Cut7</td>
<td>8.740***</td>
</tr>
<tr>
<td></td>
<td>(47.31)</td>
</tr>
<tr>
<td>Cut8</td>
<td>9.539***</td>
</tr>
<tr>
<td></td>
<td>(50.40)</td>
</tr>
<tr>
<td>Cut9</td>
<td>10.46***</td>
</tr>
<tr>
<td></td>
<td>(52.51)</td>
</tr>
<tr>
<td>Cut10</td>
<td>11.46***</td>
</tr>
<tr>
<td></td>
<td>(50.57)</td>
</tr>
<tr>
<td>Cut11</td>
<td>12.86***</td>
</tr>
<tr>
<td></td>
<td>(41.74)</td>
</tr>
<tr>
<td>Cut12</td>
<td>14.24***</td>
</tr>
<tr>
<td></td>
<td>(26.79)</td>
</tr>
<tr>
<td>Cut13</td>
<td>15.63***</td>
</tr>
<tr>
<td></td>
<td>(15.39)</td>
</tr>
</tbody>
</table>

Observations 534,567

Robust t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1
Test Taking Behavior in Turkish ÖSS Exams

3.1 Introduction

The multiple choice test structure is commonly used to evaluate the knowledge of candidates in a wide variety of situations. Its main advantages are to allow more broad evaluation of candidate’s knowledge in a short time, it is easy to grade, and there is no subjective affect of the grader in the evaluation. Because of these properties, it is preferred in both high and low stake exams in many countries. Examples include the University Entrance exam in India, Turkey, Japan, China, and the SAT exam in US. A disadvantage of such exams is that candidates may attempt to guess the answer without having any knowledge on it. In other exam types, such as short answer based exams, such uneducated responses are unlikely to reap any benefit. As a response to this problem, test administers may apply negative marking for wrong answers. Grading methods in multiple choice tests may be designed in such a way that the expected score from randomly guessing a question is equal to the expected score from skipping the question. This grading method is fair only under the assumption that candidates either know the answer, or they do not. However, if they have partial knowledge about the question, the candidate’s decision to guess/attempt or skip the question will not only depend on their knowledge, but also on their degree of risk aversion. This problem may
undermine the validity and the fairness of test scores, reducing the efficacy of the testing mechanism.

The multiple choice test results are used to allocate students to colleges, to measure effectiveness of schools, teachers, or to allocate an open position.\(^1\) Baker et al. (2010) criticize the use of test results of students to evaluate the value-added of teachers and schools, among other reasons, because of the measurement error that will be generated by random guessing. Baldiga (2013) shows in an experimental setting that conditional on students’ knowledge of the test material, those who skip more questions tend to perform worse. In light of this finding, if certain groups are favored in multiple choice tests, analysis made by using these scores will be upward biased towards these groups.

In this paper we model and estimate students’ multiple choice test taking behavior in order to understand the effects which characteristics of students have on test scores by using data from Turkish University Entrance Exam (ÖSS). The ÖSS is a highly competitive, centralized examination that is held once a year.\(^2\) In Turkey, college admission depends only on the score obtained from the ÖSS, and the high school GPA.\(^3\) However, the ÖSS score has the highest weight.\(^4\) In the ÖSS, for each correct answer the student gains one point, and for each wrong answer 0.25 points are deducted, while no points are awarded/deducted for skipping a question. As it is a very competitive and high stakes exam, student expend significant time and effort to get prepared for this exam. Therefore we assume that they are aware of the scoring method. These properties of the ÖSS exam provide us with a convenient environment in which to investigate the exam taking strategies of students.

Psychology and education literature have long been interested in multiple choice tests to characterize optimal test designs that generate fair results, with a valid measurement method. Burgos (2004) investigates the score correction methods that will award partial knowledge by using prospect theory. They model the behavior of a representative agent, so they assume away the heterogeneity in risk

---

\(^1\)In Turkey, public sector jobs are allocated according to the score obtained in a multiple choice central exam, called KPSS.

\(^2\)Every year only one third of the students are assigned to colleges.

\(^3\)This GPA is normalized at the school year level

\(^4\)ÖSS score has at least 75% weight in scores used to allocate students to colleges.
aversion and ability of agents. Similarly Bernardo (1998) analyzes the decision problem of students in a multiple choice exam to derive a fair grading rule. Espinosa and Gardeazabal (2010) model the students’ optimal behavior in a multiple choice exam to find the optimal penalty that will increase the validity of the test, i.e., increase the correlation between students knowledge and the test score by simulating their model under distributional assumptions on students ability, difficulty of questions and risk aversion. Any of these papers attempt to test their results empirically. Espinosa and Gardeazabal (2005) tests students’ rationality with an experiment by using equivalent scoring rules, but while one punishes the wrong answer, and the other one awards skipping. They find evidence that students are expected utility maximizers.

Risk attitudes of students are an important factor in the decision to attempt whenever there is uncertainty associated with the outcome. In the literature, females are shown to be more risk averse than males (see Eckel and Grossman (2008)). To test the hypothesis that female students skip more question than males since they are more risk averse, Ben-Shakhar and Sinai (1991) investigates test taking strategies of students in Hadassah and PET tests in Israel and find that females do, in fact, tend to skip more questions. Baldiga (2013) explores the gender differences in skipping/guessing behavior when students are uncertain about answers. They conduct an experiment to disentangle whether a gender gap in the tendency to skip questions exists, and if so, whether this gap is driven by differential confidence in knowledge of the material, differences in risk preferences, or differential responses to high pressure testing environment. Tannenbaum (2012) also investigates the effect of gender differences in risk aversion on multiple choice test results. Tannenbaum (2012) shows that female students are more risk averse, so they skip more often than male students. He finds that risk aversion is able to account for the 40% of the gender differences in performance in multiple choice exams. Tannenbaum (2012) is the closest paper to ours. However our conclusions conflict somewhat: we allow risk aversion to differ between various groups (gender, predicted scores) whereas he estimates a unique risk aversion for each gender.

We contribute to the literature by using a structural model to estimate students’ exam taking behavior. We investigate the effects of different characteristics of stu-

---

5Essentially testing for the Framing Effect
students on their exam performance. Particularly, we investigate students’ behavior according to their gender, and experience in the exam. Our model allows us to run counterfactual analysis to investigate the predicted change in score distributions in a variety of situations.

In the next section, we present an overview of the data and testing environment. The particular patterns of the multiple choice tests are discussed in more detail in section three. In the fourth section, the model is presented. Section five details the estimation strategy with results in section six. Sections seven and eight contain counterfactual experiments and a discussion of an extension, respectively. Section nine concludes.

3.2 Background and Data

In Turkey, college admission is based on an annual, nationwide, central university entrance exam governed by the Student Selection and Placement Center (ÖSYM). High school seniors and graduated students can take the exam. There is no restriction on retaking, i.e., students are allowed to take the exam repeatedly over the years. However they are not allowed to carry over their scores: the score obtained in a year can be used only in that year. All departments, with the exception of those that requires special talent (such as art schools) accept students based on a weighted average score of university entrance exam and high school grade point average.

The university entrance exam is held once a year all over the country at the same time. It is a multiple choice exam with four tests, Turkish, social science, math, and science. Students are given 180 minutes for 180 questions. Each test includes 45 questions, and each question has 5 possible answers. Students get one point for each correct answer, and they lose 0.25 points for each wrong answer. If they skip the question, they receive 0 points. Students’ raw test scores are calculated by deducting $\frac{1}{4}$ of the number of incorrect answers from the number of correct answers. The university entrance exam is a paper-based exam. All students receive the same questions, and they do not receive any feedback on whether their answer is correct or not during the exam.

Students choose one of the Science, Turkish-Math, Social Science, or Language
tracks at the beginning of high school.⁶ Students’ university entrance exam scores
are calculated as a weighted average of their raw scores in each test.

Table 3.1 shows the test weights according to each track. For the social science
track students, the Turkish and social science tests have the highest weight, while
math and science have a relatively low effect on the ÖSS-SÖZ score.⁷

Students are required to pass threshold of 105 points to submit preferences
(submit an application) to 2-year college programs, while they need 120 points to
apply to 4-year college programs. Students’ allocation score (Y-ÖSS) is calculated
based on their high school and exam performance as follows

\[
Y_{\text{ÖSS},X_i} = \text{ÖSS}_X_i + \alpha \text{AOBP}_X_i
\]

where \(X \in \{\text{SAY, SÖZ, EA, DIL}\}\), and \(\alpha\) is a constant that changes according to
the student’s track, preferred department and whether the student was placed
(accepted) into a regular program in the previous year or not. ÖSYM publishes
the lists of departments open to students’ according to their tracks. When students
choose a program from this list, \(\alpha\) will be 0.5, while if it is outside the open list, \(\alpha\)
will be 0.2. If the student has graduated from a vocational high school, and prefers
a department that is compatible with his high school field, \(\alpha\) will be 0.65. If the
student was placed in a regular university program in previous year, the student
is punished and \(\alpha\) will be equal to either 0.25, 0.1, or 0.375. For those students,
the \(\alpha\) coefficient is equal to half of the regular \(\alpha\). This punishment structure gives
students an incentive to stay in their track, and to accept a position when offered.

The data used in this study comes from multiple sources. Our main source of
data is the institutional data of the 2002 university entrance exam takers. This
data set includes students’ raw test scores in each test, weighted test scores, high
school, track, high school GPA, gender, and number of previous attempts.

The second source of data is the 2002 university entrance exam candidate survey.
This survey is filled by all students while they are making their application for
this exam. This data set has information on students’ family income, education

---

⁶For more detail on track choice in high school in Turkey see Ş. P. Akyol et al. (2013)
⁷In the calculation of ÖSS scores, firstly raw scores in each field are normalized to mean 50
and standard deviation 10 by using mean and standard deviation of scores in the corresponding
field. Then these normalized scores are multiplied by the weights presented in Table 3.1.
level, and expenditure on preparation.

We received a random sample of around 40,000 students from each track (Social Science, Turkish-Math, Science). In this study we focus on the social science track students (from both regular high schools and Anatolian high schools). This is due to the unique patterns seen in these test that enable identification of key parameters, patterns which do not exist in the science and math tests. This arises due to the style of questions being asked. As a result we focus on the students for whom these tests matter the most: the social science track students. Accordingly, we must exclude science high schools, as they are exclusively science track.

Table 3.10 presents the summary statistics.

We observe substantial differences in test outcomes between first time takers and second time takers. Figure 3.1 shows that second time takers achieve much higher scores, for the entirety of the distribution. An aim of this paper will be to examine whether or not this is due to the second time group having a better distribution of ability compared to first time takers, or if is simply due to a change in test taking behavior, for example their willingness to guess when uncertain.

In addition, we see difference between male and female students’ scores in Figures 3.2 and 3.3. Males tend to have a wider distribution, with more mass on the upper and lower tails. It is notable that the difference on the upper tail is more pronounced for second time takers; first time male test takers differ from their female counterparts mainly by having more students with low scores.

---

8In particular, math and science exams have many questions where the student must solve the problem to find the answer: either the student successfully solves the question or they fail to solve the question and have no information regarding what the correct answer is likely to be.

9We do not separate learning and selection effects in this paper
3.3 Multiple Choice Exam Scores

In this section we examine students’ scores in each section of the ÖSS exam: the Turkish section, social science section, math section, and science section. Recall that each section of the exam has 45 questions. For each question, there are five possible answers; answering correctly gains the student a single point, skipping the question (not giving an answer) gives zero points, but attempting the question and answering incorrectly results in a loss of a quarter point.

The scoring structure results in each multiple of 0.25 between -11.25 and 45 (with the exception of certain numbers above 42) being a possible outcome of an exam section. For example, attempting all questions and being incorrect with each question results in a score of \(-\frac{45}{4} = -11.25\), while getting everything correct nets the student 45 points.

First, Figures 3.7, 3.8, 3.9, and 3.10 show the distribution of scores in the social science and Turkish portions of the exam, for first time takers and second time takers in the social science track. These histograms use a bin width equal to 1.\(^{10}\) It can be seen that the distributions are roughly bell shaped, as would be expected. As noted in previous work (see Frisancho et al. (2012)), the distributions appear to shift to the right as we move from first time takers to second time takers.

Next, we show the histograms of scores of social science track students in the math and science portions of the exam in Figures 3.11, 3.12, 3.13, and 3.14. Two facts immediately stand out in these diagrams. First, there are a lot of scores in the zero region. Secondly, even if the large spike at zero is removed, the distribution of scores place almost all of the weight at very low scores, relative to those seen in the Turkish and social science sections of the exam.

There are two explanations for both of the two preceding facts. First, the math and science test scores have relatively little weight in the ÖSS score of social science track students. These scores are multiplied by 0.4, whereas social science and Turkish scores are multiplied by 1.8 - a substantial difference. As a result, students have little incentive to expend time and effort on these questions during the exam. Accordingly, many students opt to not even look at these sections,\(^{10}\)While the data is not rounded to the nearest integer, we do this so that the reader may see the overall distribution of scores - as will be seen shortly, this can be difficult to see when bins are not used.
resulting in many observations of zero scores. Furthermore, students are explicitly told that if they are social science track students, they should spend less time in math and science.\textsuperscript{11} If they do attempt questions, it will be relatively few, with less than advantageous results. A second reason is that these students are not well prepared in math and science - since the 9th grade they have been engaged in the social science track curriculum, meanwhile their study efforts would optimally be directed towards the exams that matter most: Turkish and social science.

A score can correspond to a single outcome (by outcome we mean the number of correct, wrong and skipped questions), multiple outcomes, or none at all. It is clear that there is only one way that a student could obtain $-11.25$ or 45; a score of 42.5 could only have arisen through attempting all questions, getting 43 questions correct and 2 incorrect. A score of 40 has two possible origins: 40 correct and 5 skips, or 41 correct and 4 incorrect. It is impossible to achieve a score of 42.25: the student must have at least 43 questions correct, and at least 3 questions skipped, which is not possible given there are only 45 questions.

There are 46 particular scores that are highly relevant: those which correspond to attempting all questions. These are spaced 1.25 points apart, starting at -11.25, finishing at 45 points.

Examining the distributions of actual scores (not placed into bins of width one), we can see some striking patterns. Figures 3.15 through 3.18 show the distributions for the social science and Turkish portions of the exam, with very prominent spikes along the distribution. It is no coincidence that the spikes appear evenly placed; they correspond to scores which could be achieved while attempting all questions. This will be very important in identifying the behavior, and implies that many students are skipping very infrequently. Additionally, there is no discernable pattern as to which students obtain such scores - these spikes remain present when examining the distributions of different groups, for example high income female first time takers with low GPAs.\textsuperscript{12}

Math and science score distributions do not exhibit this behavior, most students

\textsuperscript{11}In the exam booklet there is a note before the social science/Turkish part of the exam that says: "If you want higher score in OSS-SOZ, it may be better for you to spend more than 90 minutes on verbal part of the exam."

\textsuperscript{12}It is possible that there is some unobserved heterogeneity that determines which students lie on these spikes.
obtain a score of zero. There are no other apparent patterns to be gleaned from disaggregating the distribution, apart from a small spike at 1 in three of the four curves.\textsuperscript{13}

3.4 Model

We model the test taking procedure as follows. When a student approaches a question, they observe a signal for each of the five possible answers. The vector of signals for the question is then transformed into belief. This belief is the likelihood that each answer is in fact the correct answer. The student then decides whether or not to answer the question, and if so, which answer to choose.

We model the test taking procedure as if each test took place separately - we do not allow for outcomes in one section of the test to have any bearing on other sections.\textsuperscript{14} In addition, each question is approached simultaneously, so that outcomes (or beliefs regarding outcomes) of one question have no impact on other questions.

Signals for each of the five answers depend on whether or not the answer is actually the correct answer, and are drawn as follows:

- Incorrect answers - draw a signal from a distribution $G$, where $G$ is pareto with support $[A_I, \infty)$ and shape parameter $\beta > 1$.

- Correct answer - draw a signal from a distribution $F$, where $F$ is pareto with support $[A_C, \infty)$ and shape parameter equal to $\alpha > 1$.

Assumption 2. Both $F$ and $G$ have support $[A, \infty)$, where $A > 0$.

Suppose that the student observes five signals, given by the following vector:

$$X = (x_1, x_2, x_3, x_4, x_5) \quad (3.1)$$

where $x_i$ is the signal that the student receives when examining answer $i$. What

\textsuperscript{13}This could correspond to students only answering the easiest question in the exam.

\textsuperscript{14}We are explicitly ignoring time restrictions, whereby a quick performance in one section of the exam might afford the student additional time in another section, allowing the student to more carefully examine each question.
then is the student’s belief regarding the likelihood that each answer is correct? Using Bayes’ rule, the probability that answer $i$ is correct can be expressed as:

$\text{Prob(Answer } i \text{ is correct} | X) = \frac{\text{Prob}(X | \text{Answer } i \text{ is correct}) \times 0.2}{\text{Prob}(X)}$ (3.2)

Expressing the numerator in terms of the densities of the two distributions, $F$ and $G$, for the case where $i = 1$:

$\text{Prob}(X | \text{Answer 1 is correct}) = \alpha A^\alpha \frac{\beta A^\beta}{x_1^{\alpha+1} x_2^{\beta+1} x_3^{\alpha+1} x_4^{\beta+1} x_5^{\beta+1}}$ (3.3)

In essence, the density of $F(\cdot)$ at $x_1$ (as this is conditional on 1 being correct) multiplied by the product of the density of $G(\cdot)$ at the other signals.

It follows, by substituting equation 3.3 into equation 3.2, that the probability that answer $i$ is correct, conditional on $X$, can be expressed as:

$\text{Prob}(i \text{ is correct} | X) = \frac{\alpha A^\alpha \prod_{j \neq i}^5 \beta A^\beta}{\sum_{m=1}^5 \left( \frac{\alpha A^\alpha}{x_m^{\alpha+1}} \prod_{n \neq m} \frac{\beta A^\beta}{x_n^{\beta+1}} \right)}$ (3.4)

where $i, j, m, n \in \{1, ..., 5\}$.

This can be further simplified to:

$\text{Prob}(i \text{ is correct} | X) = \frac{1}{x_i^{\gamma+1}} \prod_{j \neq i} x_j^{\gamma+1}$ (3.5)

Letting $\gamma = \beta - \alpha$, so that $\frac{1}{x_i^{\gamma+1}} = \frac{1}{x_i^{\beta+1}} x_i^\gamma$, the expression further simplifies to:

$\text{Prob}(i \text{ is correct} | X) = \frac{x_i^\gamma}{\sum_{m=1}^5 x_m^\gamma}$ (3.6)

Note that the sum of beliefs for each of the five answers adds up to unity. Without loss of generality, we assume that $\beta \geq \alpha$, which leads to positive relationship between the value of the signal, and the likelihood that the answer is correct.\textsuperscript{15,16}

\textsuperscript{15}A higher shape parameter for a Pareto distribution shifts probability mass to the left.

\textsuperscript{16}If a student were to draw from distributions with $\beta < \alpha$, smaller signals would be associated with the correct answer.
The following propositions show that the model is insensitive to certain parameters:

**Proposition 3.4.1.** The outcome of the model is the same for all $A > 0$

**Proposition 3.4.2.** The outcome of the model is the same for all $(\alpha, \beta)$ that satisfies $\frac{A}{\alpha} = k \geq 1$

Accordingly, we can, without loss of generality, take $A = 1$ for all students, and take $\alpha = 1$ for all students. As a result, the structure of a student’s signals can be represented by the shape parameter of the incorrect answer: $\beta$. A higher value of $\beta$ draws the mass of the distribution towards the minimum, $A = 1$, allowing the student to more clearly separate the incorrect signals from the signal given by the correct answer. In other words, higher $\beta$ students are what would be referred to as high ability students.\(^{17}\)

The effect of a higher $\beta$ on test outcomes can be decomposed into three effects. First, the correct answer has a higher probability of generating the highest signal. Increasing $\beta$ shifts the CDF of the incorrect answers’ signals to the left, and the student’s best guess (the answer with the highest signal) will be correct more often. Secondly, when the correct answer actually gives the highest signal, the probability with which the student believes that it comes from the correct answer increases as the weighted sum of the incorrect signals decreases. If the first answer is the correct answer, lowering $\sum_{i=2}^{5} x_i^* \gamma_i$ increases the student’s belief that answer 1 is correct.

Finally, there is a subtle effect of $\beta$ on tests. Students with high ability, i.e. a high value of $\beta$, will be more confident in their choices. Even with the same signals, as we increase $\beta$, the student’s belief that the highest signal comes from the correct answer increases. This is formally stated below:

**Lemma 3.4.3.** Suppose there are two students: one with ability parameter $\beta = b_1$ and the other with ability parameter $\beta = b_2 > b_1$. Suppose that the two students receive identical signals $X$ for a question. Let $x_{\max} = \max\{x_1, ..., x_5\}$. The student with the higher value of $\beta$ has a higher belief that $x_{\max}$ is drawn from the correct answer.

\(^{17}\)Signal distributions for a student with ability $\beta = 3$ are shown in Figure 3.23
Once students have observed signals for each of the five possible answers to the question, they are faced with six possible alternatives: either choosing one of the five answers, or skipping the question. Skipping the question does not affect their ÖSS score, answering correctly increases the score by 1, while answering incorrectly decreases the score by 0.25 points. Note that the expected value of a random guess is $0.2 \times 1 - 0.8 \times 0.25 = 0$.

If a student were to choose an answer, they would choose the one which was most likely to be correct. A slightly higher score is clearly preferred. In this model, the answer which is most likely to be correct is the one with the highest value of $x_i$. Also, this answer trivially has a probability of being correct (conditional on observed signals and the student’s ability) greater than or equal to twenty percent.

However, the relationship between ÖSS score and utility need not be linear. It is reasonable to suggest that there may be a degree of risk aversion present, both student’s general attitudes towards risk and the structure of the relationship between ÖSS score and university admission. On the other hand, certain areas could well exhibit risk loving behavior: students must score above 120 in order to be qualified to make a preference submission to a four year college program.

As such, we stipulate that students have a cutoff for the belief. If the student believes that the best answer (highest signal) has a probability of being correct greater than the cutoff, they will attempt the question, choosing the best answer. However, if all answers have a probability lower than this cutoff, then they will skip the question. This cutoff lies in the interval $[0.2, 1]$.

A higher value for the cutoff implies a higher degree of risk aversion, while a cutoff of 0.2 would be supported by risk neutral/risk loving preferences.

Consider a student with ability parameter $\beta$ and attempt threshold $c \in (0.2, 1)$. From these two parameters, we are able to calculate the probability that they would skip a given question, the probability of answering correctly, and the probability of answering incorrectly.

In order to answer a question, with answer $n$, the signal drawn for answer $n$, $x_n$, must satisfy two conditions. First, it must be the highest signal. Second, it must be high enough, given the other signals, so that the belief is greater than the cutoff. 

---

18 As there will always exist an answer with probability of being correct greater than or equal to 0.2, we do not consider cutoffs below 0.2, as they would result in the same behavior: always attempting the question, never skipping.
the cutoff required to attempt the question. We define the following function as the minimum signal $x_n$ required to attempt with the $n^{th}$ answer, given the other signals:

$$K(\{x_i\}_{i \neq n}) = \max \left( \max\{x_i\}_{i \neq n}, \left( \frac{c}{1 - c} \left( \sum_{i \neq n} x_i^\gamma \right)^{1/\gamma} \right) \right) \quad (3.7)$$

Suppose that answer number 1 is the correct answer. The chance that answer number 2 is selected by the student, that is, provided as the answer, is:

$$\int_{x_5 = A}^{\infty} \int_{x_4 = A}^{\infty} \int_{x_3 = A}^{\infty} \int_{x_2 = K(x_1, x_3, x_4, x_5)}^{\infty} 1dG(x_2)dF(x_1)dG(x_3)dG(x_4)dG(x_5) \quad (3.8)$$

So that the chance of the student submits an incorrect answer is the value of the above equation multiplied by the four possible incorrect answers. Similarly, the probability that the student submits a correct answer (in this case, answer number 1) is:

$$\int_{x_5 = A}^{\infty} \int_{x_4 = A}^{\infty} \int_{x_3 = A}^{\infty} \int_{x_2 = K(x_1, x_3, x_4, x_5)}^{\infty} 1dF(x_1)dG(x_2)dG(x_3)dG(x_4)dG(x_5) \quad (3.9)$$

The probability that the student skips the question can be obtained similarly, by finding for each answer the probability that it gives the highest signal, yet is below the threshold to attempt.

These lead to three functions that describe the probabilities of each of the three possible outcomes of a question, conditional on student ability $\beta$, and cutoff $c$:

$$\text{Prob}(\text{Correct}) = P_C(\beta, c) \quad (3.10)$$
$$\text{Prob}(\text{Wrong}) = P_W(\beta, c) \quad (3.11)$$
$$\text{Prob}(\text{Skip}) = P_S(\beta, c) \quad (3.12)$$

where $P_S(\cdot) = 1 - P_C(\cdot) - P_W(\cdot)$. Table 3.3 provides these, in addition to

---

19 A diagram showing choices conditional on signal observations for a simplified two answer setup is shown in Figure 3.24
the average points earned per question, for various parameter values. Consistent with the literature, as the probability to skip increases, the average points per question decreases (for a fixed ability).

In each exam, the student faces 45 questions, with signals and outcomes independent across all questions in the exam. From this, we can find the probability that the student attempts \( x \in \{0, ..., 45\} \) questions, skipping \( 45 - x \) questions:

\[
\text{Prob(Answer } x \text{ questions)} = \binom{45}{x} (P_C + P_W)^x (P_S)^{45 - x} \tag{3.13}
\]

Conditional on answering \( x \) questions, the probability that \( y \in \{0, ..., x\} \) questions are answered correctly is:

\[
\text{Prob(Answer } y \text{ of } x \text{ questions correctly)} = \binom{x}{y} \left(\frac{P_C}{P_C + P_W}\right)^y \left(\frac{P_W}{P_C + P_W}\right)^{y-x} \tag{3.14}
\]

A student that attempts \( x \) questions, correctly answering \( y \) questions, achieves a score in that exam of:

\[
\text{Score}(x, y) = y - \frac{(x - y)}{4} \tag{3.15}
\]

Accordingly, we can find the probability that a student with ability \( \beta \) and cutoff \( c \) obtains a score of \( s \), through all of the possible methods of obtaining such a score. Thus we obtain a mapping from \((\beta, c)\) to the distribution of scores:

\[
\text{Prob(Score } = s \text{)} = M(\beta, c, s) \tag{3.16}
\]

### 3.5 Estimation Strategy

In our model, students’ scores depend on students’ ability \((\beta)\) and risk aversion cutoff, \(c\). In our data set we observe student’s scores, but not the decomposition.

---

\(^{20}\)A \( \beta \) of 3 is later found to be approximately median.

\(^{21}\)Of course the average points per question attempted increases.

\(^{22}\)For example, the probability of obtaining a score of 40 is the probability of (40 correct, 0 incorrect, 5 skipped) added to the probability of (41 correct, 4 incorrect, 0 skipped)
In this section we use our model to estimate the distribution of ability, beta, and risk aversion cutoff.

Estimation of the parameters of interest, student ability $\beta$ and risk aversion cutoff $c$, is conducted jointly for each exam, gender, and attempt number. In addition, we recognize that the relationship between ÖSS-SOZ score and utility is not necessarily constant throughout the range of scores: the degree of risk aversion may be different. In particular, we could expect that students anticipating low scores would be considerably less risk averse, since scores below a cutoff result in the same outcome: an inability to submit preferences/apply to universities. This would result in a jump in the payoff function as students cross the cutoff score.

As a result, while we allow cutoffs to vary by gender, attempt number, and subject area, we also allow cutoffs to depend on the interval in which the student’s predicted ÖSS-SOZ score lies, for example 120-130. To accomplish this, we first regress ÖSS-SOZ on GPA (adjusted for school)\(^{23}\), education level of both parents, monthly income of parents, and preparation on the four subject areas. We use the results of the regression to derive fitted values of ÖSS-SOZ, predicted exam scores given observable characteristics, for each student. This estimation is conducted separately for each gender/attempt number.

We divide students into groups, according to gender, attempt number, and the range into which their predicted ÖSS-SOZ score lies: (0, 90), [90, 100), [100, 110), [110, 120), [120, 130), [130, 140), and [140, $\infty$). For each group, we examine the two subjects separately.\(^{24}\) While these intervals may not contain equal numbers of students, it will allow us to make comparisons across genders and attempt numbers. For each group, we take the cutoff $c$, and the distribution of ability $\beta$ within the group. The ability of each student is given by $1 + e^X$, where $X$ is distributed normally with mean $\beta_X$ and variance $\sigma_X^2$. This ensures that each student has an ability greater than 1, and results in a log normal shaped distribution (shifted 1 unit to the right). We then generate a sample of $n$ students, characterized by their cutoff $c$ and ability $\beta_i$.

\(^{23}\)To adjust for school quality, we adjust the GPA of student within a school based on the performance of the school in the exam. We observe normalize GPA for each students, which is able to be converted to a ranking within the school. As we observe the mean and variance of exam scores for each school, we can easily convert the GPA to a measure that reflects the quality of the school.

\(^{24}\)The only tests of interest are Turkish and social science
As we now have \((\beta, c)\) for each student, we are able to find the distribution of scores \(s\) for each student. That is, if a student has characteristics \((\beta, c)\), what is the probability the student obtains each of the 220 possible scores for the test in question. We compare this to the actual distribution observed in the data, by aggregating the students. Specifically, we look at moments related to the intensity of the spikes, and the shape of the distribution. More specifically, the difference between the mass of students with scores corresponding to attempting all questions (i.e. 45, 43.75,...) and the mass of students with scores corresponding to skipping a single question (i.e. 44, 42.75,...) captures in the intensity of the spikes. For example, \(0.4 - 0.3 = 0.1\). If the spikes are very prominent, this difference will be large; if they are non-existent, this difference will be minimal. In addition, we use the mean of scores, and the variances of scores.

\[
\left(\frac{1}{N} \sum_{n=1}^{N} g(\beta, c)\right)'^t \left(\frac{1}{N} \sum_{n=1}^{N} g(\beta, c)\right) \tag{3.17}
\]

Accordingly, the estimates of the cutoff \(c\) and ability distribution parameters \(\mu, \sigma^2\) for each group will be the GMM estimator, given by:

\[
\hat{c}, \hat{\mu}, \hat{\sigma^2} = \arg \min \left(\frac{1}{N} \sum_{n=1}^{N} g(\beta(\mu, \sigma), c)\right)' \left(\frac{1}{N} \sum_{n=1}^{N} g(\beta(\mu, \sigma), c)\right) \tag{3.18}
\]

With the identity matrix used as the weighting matrix, we obtain an estimate of the cutoff of each group that is consistent and asymptotically normal. Applying the two step procedure,\(^{25}\) this estimate is used to generate a weighting matrix. Using the new weighting matrix, the procedure is repeated, and an efficient, consistent, and asymptotically normal estimate is obtained.

### 3.5.1 Identification

Identification of the cutoff beliefs is achieved through matching the intensity of the spikes. For example, if students are risk averse then they will tend to skip. Thus, at low values of \(c\), students will have a very low probability of skipping a question:

\(^{25}\)See Hansen (1982)
it is unlikely that the answer with the highest signal has a low enough probability of being correct to be below the cutoff. As a result, almost all of the probability mass of a given student’s distribution will be located on scores corresponding to attempting all questions. As the cutoff increases, students become more and more likely to skip some questions, resulting in more mass lying on scores unreachable by attempting all questions (i.e. some questions must be skipped), while the spikes still remain prominent. Increasing the cutoff further results in enough skipping activity so that spike cannot be seen.

This is illustrated in figure 3.25, where the score distribution for a student (with a fixed, approximately median, ability of $\beta = 3$) is shown for various cutoff levels. A cutoff of $c = 0.225$ gives virtually all of the mass to the attempt all scores. As the cutoff increases to 0.3, the spikes all but disappear.

The relationship between the intensity of the spikes and the cutoff is not constant. For a fixed cutoff $c$, as we increase ability, the intensity of the spikes increases. While low ability students might have a high chance of having a highest belief below the cutoff, it becomes increasingly rare as we move to the high ability students.

The distribution of the ability of a group of students is identified by distribution of scores. An increase in the mean parameter $\beta_X$ moves the score distribution to the right, increasing the mean, while an increase in the variance parameter $\sigma_X^2$ increases the variance of the score distribution. This is due to strong relationship between ability and exam scores.

### 3.6 Results

Tables 3.4 and 3.5 contain the estimates of the belief cutoff, below which a student will skip a question, for the various groups,\(^{26}\) in addition to the standard errors of the estimates.

Two facts are apparent. While males and females have roughly similar cutoffs, males tend to have lower risk aversion cutoffs, especially for high scoring students. This is even more so among second time takers. This is in line with the literature

\(^{26}\)Estimates for the second time takers in the ÖSS score range less than 90 are not obtained due to insufficient observations.
- males are acting in a less risk averse manner. Secondly, the cutoff rises systematically in the score range below 120. This matches what we know about the payoff structure. For low scores, students should be much less risk averse since any score below 105 will not allow the student to submit preferences for any school, and any score below 120 will not permit the student to submit preferences for four year college programs. Above 120, the cutoff remains relatively high: a lower score could see the student forced to attend a much less desirable institution.

Also noteworthy is the observation that the cutoffs tend to decrease among high scoring students between first time takers and second time takers, whereas they tend to increase among the low scoring students.

Figures 3.26 through 3.33 show the simulated distributions compared to observed distributions for the various groups. While the estimation procedure was designed only to match subgroups of the sample, the entire simulated distribution fits the data relatively well, with some exceptions: it systematically under-predicts scores which correspond to skipping multiple questions. In addition, the skipping behavior is underestimated among low scoring students - this is likely due to such students correctly anticipating their low expected score and acting accordingly, whereas in the estimation many of these are restricted to acting with a (high) cutoff corresponding to their fitted score.

Estimates of the parameters governing the distribution of ability for each group are presented in Tables 3.6 and 3.7. Recall that ability is parameterised as $1 + e^X$, where $X \sim N(\mu, \sigma^2)$. The mean and variance of $X$ in each group are presented.

As predicted, groups that are predicted to have high exam scores have much better distributions of ability for both Turkish and social science. However, there is significant variance in the distributions, reflective of the fact that the fitted score is an imperfect measure of overall student ability. We see that females tend to have higher ability in Turkish, but lower ability in social science, when compared to males in the corresponding group. This implies that males tend to have a comparative advantage in social science.

In addition, we observe that males tend to have higher variance in their distribution of ability. In fact, the variance is greater for all groups. This has two interpretations. First, the distribution of abilities is more dispersed among males,

\footnote{This issue is addressed in more detail in a later section.}
which is also implies by the distribution of exam scores. There is another possibility, that the fitted ÖSS score is not as accurate for males. There could be a number of causes for this, such as GPA being a poor predictor of exam scores, etc.

Looking at the distributions of ability across the various groups, we see similar patterns. As shown in figures 3.34 and 3.35, second time takers have higher abilities across the distribution, for both social science and Turkish. However, we can conclude definitively why this is the case.

There are two possibilities for this difference in ability: selection and learning. It is possible that students tend to learn between their first and second attempts, so that they increase their ability in the social science and Turkish sections of the exam. However, it is also possible that the perceived change in the distribution is simply due to a selection effect. Consider the students who choose to continue. They could very well be different from students who choose not to retake the exam. It could be that the best students do not retake the exam: they are admitted into a university program and so have no reason to take the exam. On the other side of the distribution, the worst students may have very little incentive to return, as they have almost no chance of meeting the threshold required to apply to programs. If the second effect is important, selection could very well result in a better distribution of student abilities among second time takers.

In figures 3.36 and 3.37, we see how the genders differ the first time they take the exam. The lower portion of the social science ability distribution is indistinguishable, however males have a considerably better distribution for the top portion, compared to females. This is not the case with the Turkish portion - females are much better for all points in the distribution. This provides an interpretation of the observed differences in ÖSS-SOZ scores. Males are overall worse at Turkish, but the best males make up for it in social science.

The second time takers exhibit a similar pattern in figures 3.38 and 3.39; however the advantage of males amongst the top social science students is more pronounced, and their disadvantage in Turkish decreases at the high end. As seen in the ÖSS-SOZ score distributions 28, the comparison between male and female low ability students is similar to that of first time takers; whereas the high ability males gain a more favorable distribution (relative to females) in the second attempt.

28 See figures 3.2 and 3.3
3.7 Counterfactuals

Having recovered the parameters regarding the risk aversion exhibited by students in the multiple choice tests, in addition to estimates regarding ability (as measured by $\beta$ for each subject, the parameter in the Pareto distribution that governs dispersion), we are now able to perform counterfactual experiments.

In these experiments, we will compare outcomes of a number of testing regimes, and student behaviors. For example, how would exam outcomes differ if all students attempted (answered) every question, as would happen if the penalty for answering incorrectly were removed? This is relevant because it is fully feasible to change the testing regime, and there is the possibility that the regime has an effect on the outcomes: males and females, first and second time takers, act differently.

The objects of interest in these experiments are as follows:

- The relationship between ability in social science ($\beta_{SS}$) and social science section exam score percentile
- The relationship between ability in Turkish ($\beta_{T}$) and social science section exam score percentile
- The relationship between gender and social science section exam score percentile
- The relationship between attempt number and social science section exam score percentile

The first two objects are clearly important. The ÖSS exam system is an allocation mechanism, presumably designed to give the students with high abilities access to the most desirable institutes of higher education. However the exam score is an imperfect measure of student ability. The students who score in the top one percent are not necessarily those in the top one percent as measured by ability. The testing regime, and restrictions on behavior, may affect the dispersion of possible scores for students, affecting how precisely the system identifies the best students.

The third relation of interest, gender vs exam scores, is also important. It is recognized in the literature that males are less risk averse than females in test
situations (see Eckel and Grossman (2008)). In addition, we have found that females tend to have higher thresholds for attempting to answer a question, i.e. they are more risk averse. Since the testing regime in question is forcing students to accept an element of risk when choosing to answer a question, the preferences regarding risk affect the distribution of final exam scores. This may tend to favor male test takers, leading in essence to a systemic bias in the testing procedure. In addition, the distributions of abilities are considerably different across gender; the regime may end up attenuating these differences.

Finally, we also see different observed attitudes to risk across attempt number, as well as vastly different distributions of abilities. As a result, the regime may be able to influence the proportion of students that are first time takers in a given final exam score percentile. While this paper does not go into much detail on the topic, the costs/benefits of delaying entry into university are likely to be important.

The seven possible regimes used in this counterfactual experiment are:

1. The baseline model, as estimated in the previous section

2. Preferences of males and females are switched, so that the cutoff used by a male $i^{th}$ time taker in exam subject $j$, with ÖSS-SOZ score interval $k$ is switched with that used by a female $i^{th}$ time taker in exam subject $j$, with ÖSS-SOZ score interval $k$, and vice versa

3. Preferences of first and second time takers are switched, so that the cutoff used by a male 1$^{st}$ time taker in exam subject $j$, with ÖSS-SOZ score interval $k$ is switched with that used by a male 2$^{nd}$ time taker in exam subject $j$, with ÖSS-SOZ score interval $k$, and vice versa

4. All students attempt all questions. This is equivalent to assuming that all students are risk neutral/loving $^{29}$

5. Each question has only 4 answers to choose from, with the penalty for an incorrect answer adjusted accordingly

$^{29}$This is also more or less identical to removing the penalty for answering incorrectly - both cause rational students to answer every question. Scores would however need to be rescaled to reflect the absence of such a penalty: instead of ranging from $-11.25$ to $45$, they would range from $0$ to $45$. 
6. The penalty for answering incorrectly is increased from 0.25 points to 0.5 points.

7. The penalty for answering incorrectly is increased from 0.25 points to 1 point.

While the second is clearly eliciting the gender effect on outcomes, and the third the effect of experience (through test behavior), the fourth counterfactual seeks to examine the effect of having penalties for incorrect answers, as opposed to the simple, standard approach of a single point for each question answered correctly.

The fifth requires more explanation. In the default regime, there are five questions, with a single point for correct answers and a quarter point lost for incorrect answers. This results in an expected gain of zero from a random guess; accordingly, we set the penalty equal to one third of a point in the four answer scenario, resulting in a random guess having an expected gain of zero.

As a result, the cutoffs for attempting a question must be different. To convert cutoffs from the five answer case, we first assume a CARA utility function, and solve for the risk aversion parameter that generates a given cutoff. This is repeated for each student. We then convert the risk aversion parameter to a cutoff in the four answer case.\textsuperscript{30}

The sixth counterfactual is designed to elicit more skipping from students, in order to increase the impact that differences regarding risk preference have on exam outcomes. The seventh continues, increasing the penalty even further. Similarly to the four answer counterfactual, new cutoffs are obtained for both counterfactuals.

For each of the seven possible regimes, we find the resulting distributions of scores for the entire sample of students, and segment scores into bins of five percent.\textsuperscript{31,32} For each of the twenty bins, ranging from the lowest scores to the highest, we find four objects of interest: share of males, share of first time takers, average social science ability and average Turkish ability.\textsuperscript{33}

Figures 3.40 through 3.43 shows how the four objects of interest differ across the different regimes. As expected, given their greater exam score variance, the

\textsuperscript{30}For example, a cutoff of 0.240 in the five answer case implies risk aversion coefficient of 0.383 (CARA utility), which results in a cutoff of 0.300 in the four answer case.

\textsuperscript{31}Five percent of the number of students.

\textsuperscript{32}The rationale behind segmenting into percentiles, not scores, is to see the effects on the resulting allocation of students to university programs.

\textsuperscript{33}The average of log(\(\beta\)) is used for each subject.
male fraction is u-shaped, as shown in figure 3.40. However, there is no discernable pattern in the differences between the seven regimes throughout most of the range. While there are some small differences around the median, all seven are fairly similar. We do see some small differences in the top performing students: “Attempt All” gives a lower Male fraction (the more abundant Males are seemingly disadvantaged by the added variance in scores) whereas the “Higher Penalty” regime has a very slightly higher Male fraction, i.e. the small difference in risk aversion begins to have a small effect.

This insensitivity is even more apparent when examining the other graphs. Second time takers are dominant in the high exam scores percentiles, as would be expected, with the seven curves lying on top of each other in figure 3.41.

Figures 3.42 and 3.43 have higher ability students in the higher score percentiles, with no differences across the six cases featuring risk aversion. We do however see that the “Attempt All” regime gives slightly lower average abilities amongst the highest performing students, consistent with penalties and risk aversion allowing an improved separation of students by abilities.

Although the seven regimes can lead to considerable different score distributions, the relationship between gender, attempt number and ability, and exam score percentiles is relatively invariant. The reasoning for this is relatively straightforward. While there may be differences in attitudes to risk, and resulting test taking behavior, the implications of these differences happen to be rather small due to the characteristics of situations when these differences are relevant. While a difference in the cutoff of 0.23 versus 0.25 may be considerable given that the risk neutral cutoff is 0.2, and implies considerably different attitudes to risk, the effect on scores is small for two reasons. Firstly, there is a relatively low chance that a student has a belief lying between 0.23 and 0.25 for a given question. Secondly, if the belief does lay in that region, the expected gain from answering (and hence that from having a low cutoff) is at most 0.0625 points. Even when the penalty is raised, leading to more skipping behavior, the total effect on allocations is minor.

However, a degree of caution should be applied when applying this result to other tests with different students. Here, the lack of an effect is the result of a relatively low degree of risk aversion overall, in addition to an exam where students are able to be confident enough to answer a vast majority of the time. While there
is no obvious reason why the first might be particular to this group of students, it is very reasonable to suggest that the second depends very much on the style of the exam, questions asked and so on.

3.8 Extension

While this paper has focused on the Social Science track students, and in particular on the social science and Turkish tests, a variant of the model can be constructed to be applied to the science and math exams.

If one were to apply the model to the science and math exams (for the science track students) one would find a very high belief threshold for answering questions. This is due to the absence of spikes throughout most of the distribution. However this is not likely to be the case: instead of science students being extremely risk averse, it is simply much more reasonable to posit that the style of questions asked is inherently different.

For example, consider a math question involving geometry, where the student is asked to find the value of a particular angle. There are five possible answers provided, but it would not be advisable strategy to look at each answer and contemplate the degree to which it is likely to satisfy the implied system of equations. Rather, the student would simply solve for $x$, see that her solution is found in the list of possible answers, and choose that as her answer. Or, they would fail to solve the question, and realize that each of the answers has a twenty percent probability of being correct, and so they would skip the question.

Clearly, the situation described is not well modeled with the distributional assumptions of the baseline model. Instead, it is reasonable for there to be some proportion of the questions in an exam modeled as in the standard model, while the other questions could have a bimodal distribution for the signal of the correct answer. If the student is not able to solve the question, the signal for the correct answer will be found close to the lower mode, while if they solve the question, the signal will be close to the upper mode. Higher ability students would have a

---

34 Exceptional probability mass being placed on scores corresponding to attempting all questions

35 With the exception of scores in the 40-45 range
greater probability mass around the higher mode, i.e. they are more likely to solve the question.

Such an extension would explain why we see so much skipping behavior, even among high scoring students. While in the standard case a high ability student is able to be confident enough most of the time, with the bimodal case they would in essence usually skip a question unless they solve the question: their belief about the likelihood of being correct is either very low or very high.

In addition to enabling analysis of the science track exams, such an extension will improve the fit of the model in the social science track exams. While the model does a reasonable job in matching the patterns seen in the score distributions, it is not capable of matching certain areas: namely the scores which could only be obtained by skipping four questions. The reasoning for this is that skipping four questions is highly unlikely in the model for all but the lowest ability students. As a result, the model under predicts the likelihood of such scores.

The simple explanation is that there are some questions in the social science and Turkish exams that fall into the category of the bimodal distribution of the correct answer. Either students know the correct answer, or they have no idea. While such questions are evidently not the majority, there are enough to impact the resulting distributions.

A rudimentary method for allowing for such behavior is to allow a fraction of students to have a much higher propensity to skip, while the remainder of the students act as before, with a reasonable, low cutoff. While this clearly isn’t completely in line with the extended model, it will allow more skipping whilst retaining the spikes, allowing a better fit, and enable interpretation of the results.

As before, for each group (for example, female first time students in the ÖSS-SOZ score range 120-130, Turkish exam) we find the cutoff (for the low type cutoff) below which they skip the question. In addition, we find the proportion of students that act as if they are very risk averse (taken as a cutoff of 0.50) for each group. As before, the first term is captured by the prominence of the scores corresponding to attempting all questions. Meanwhile, the second term is identified by how sharply

---

36 For example they all but rule out one or more incorrect answers
37 These correspond to scores that are 0.25 points below the spikes; for example 38.75 could be obtained by getting 40 correct and 5 wrong, 38.5 could only be obtained by getting 39 right, 2 wrong and 4 skipped.
the probability decreases as we move from scores corresponding to skipping a single question, to those corresponding to skipping two questions, and so on.

Estimates of the parameters are shown in tables 3.8, 3.9, 3.10, and 3.11. The estimates of the cutoffs (for the low type students) follow similar patterns to the original model, and are relatively similar, with the exception of the average students.

The proportion of high type students versus percentile follows a u-shaped pattern. This explains why the estimates for the average students have changed more than other groups: there was significant skipping behavior which the basic model attempted to account for. It is also what would be expected by the model. High ability students should be able to "solve" the bimodal type questions much more frequently than the average ability students, resulting in a low amount of skipping. This would result in the basic model (where the proportion of low type students is equal to one) matching relatively well. Average ability students on the other hand would be unable to solve the questions considerably more frequently, which results in this exercise finding a small probability of the low cutoff type students. Students with low ability are perhaps likely to attempt questions even if they have no information regarding the answer - unless their score gets above 105 points they cannot get into any university/college.

The resulting distribution also provides a much better fit, compared to the basic model, as seen in figures 3.44 through 3.51. In the lens of the extended model this would be due to the presence of a few questions that would only be answered (with any degree of certainty) when solved.

An additional dimension through which the model can be extended is the simultaneous nature of the exam. Time is scarce for many students, it follows that spending additional time on one section of the exam reduces the time available for other sections of the exam. Conversely, students with a high ability in one section of the exam will be able to finish that section relatively briskly, and will have more time available to ponder the other section.

This constraint could be addressed by a model in which a student possesses two ability terms - one for social science and the other for turkish - and the ability to distinguish correct from incorrect depends on both inherent ability, and the amount of time allocated to the question. The ability to distinguish is increasing in both,
however the two could be said to be substitutes. The student then optimally decides how much time to assign each section of the exam. Preliminary evidence tends to support this: the correlation between estimated Turkish and social science ability parameters tends to be higher for low ability students - consistent with the notion that high ability students are not affected by time constraint as much as low ability students.

Estimation would proceed similarly to the simple model, except instead of solving for abilities in each section separately, they would be estimated jointly, with a much more complicated objective function. In addition, the relationship between ability, time per question, and likelihood of choosing the right answer would need to be estimated.

\[ \beta_E = \left[ \beta \sigma^{\frac{1}{\sigma}} + t \sigma^{-1} \right] \sigma^{\frac{1}{\sigma}}. \]

Here, \( \beta \) would be the true ability of the student, while \( t \) is the time allocated to the question.

---

38 For feasibility, it would need to be assumed that the student looks at all questions in the social science and Turkish sections, and none of the science/math questions, which is not too far from reality for the majority of students.

39 If a student is time constrained, a higher ability in Turkish will afford them more time in social science, allowing them to perform better in the social science.

40 A candidate would be a CES function, where the effective ability of a student, \( \beta_e \), takes the form \( \beta_E = \left[ \beta^{\frac{1}{\sigma}} + t^{\frac{1}{\sigma}} \right] \sigma^{\frac{1}{\sigma}}. \) Here, \( \beta \) would be the true ability of the student, while \( t \) is the time allocated to the question.
3.9 Conclusions

We observe significant differences in the way that males and females approach the exam, with females acting in a more risk averse manner. While this theoretically leads to a disadvantage in tests which impose a penalty (resulting in zero expected return from a random guess), we find that differences have very little bearing on aggregate outcomes.
3.10 Appendix

Table 3.1: Test Weights

<table>
<thead>
<tr>
<th>Track</th>
<th>Math</th>
<th>Science</th>
<th>Turkish</th>
<th>Social Science</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science Track (ÖSS-SAY)</td>
<td>1.8</td>
<td>1.8</td>
<td>0.4</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>Social Science Track (ÖSS-SÖZ)</td>
<td>0.4</td>
<td>0.4</td>
<td>1.8</td>
<td>1.8</td>
<td>0</td>
</tr>
<tr>
<td>Turkish-Math Track (ÖSS-EA)</td>
<td>0.8</td>
<td>0.4</td>
<td>0.8</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>Language Track (ÖSS-DIL)</td>
<td>0</td>
<td>0</td>
<td>0.4</td>
<td>0.4</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Table 3.2: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std.Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>37650</td>
<td>0.567</td>
<td>0.496</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ÖSS-SÖZ</td>
<td>37650</td>
<td>116.737</td>
<td>18.912</td>
<td>0</td>
<td>161.166</td>
</tr>
<tr>
<td>Normalized High School GPA</td>
<td>37650</td>
<td>47.550</td>
<td>7.825</td>
<td>30</td>
<td>80</td>
</tr>
<tr>
<td>Raw Turkish Score</td>
<td>37650</td>
<td>23.983</td>
<td>9.965</td>
<td>-10</td>
<td>45</td>
</tr>
<tr>
<td>Raw Social Science Score</td>
<td>37650</td>
<td>19.229</td>
<td>10.397</td>
<td>-8.75</td>
<td>45</td>
</tr>
<tr>
<td>Raw Math Score</td>
<td>37650</td>
<td>1.853</td>
<td>4.061</td>
<td>-9</td>
<td>42.5</td>
</tr>
<tr>
<td>Raw Science Score</td>
<td>37650</td>
<td>0.197</td>
<td>1.296</td>
<td>-8.75</td>
<td>25.75</td>
</tr>
</tbody>
</table>

Education level of Dad

<table>
<thead>
<tr>
<th>Level</th>
<th>Obs</th>
<th>Mean</th>
<th>Std.Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary or less</td>
<td>37650</td>
<td>0.557</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Middle/High school</td>
<td>37650</td>
<td>0.281</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2-year higher education</td>
<td>37650</td>
<td>0.028</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>College/Master/Phd</td>
<td>37650</td>
<td>0.050</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Missing</td>
<td>37650</td>
<td>0.085</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Income Level

<table>
<thead>
<tr>
<th>Level</th>
<th>Obs</th>
<th>Mean</th>
<th>Std.Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;250 TL</td>
<td>37015</td>
<td>0.459</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>250-500 TL</td>
<td>37015</td>
<td>0.393</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>500-750 TL</td>
<td>37015</td>
<td>0.095</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>750-1000 TL</td>
<td>37015</td>
<td>0.030</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1000-1500 TL</td>
<td>37015</td>
<td>0.012</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1500-2000 TL</td>
<td>37015</td>
<td>0.006</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>&gt;2000 TL</td>
<td>37015</td>
<td>0.005</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std.Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Attempts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st attempt</td>
<td>35470</td>
<td>0.261</td>
<td>0.261</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>35470</td>
<td>0.250</td>
<td>0.250</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3rd attempt</td>
<td>35470</td>
<td>0.242</td>
<td>0.242</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4th attempt</td>
<td>35470</td>
<td>0.154</td>
<td>0.154</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5th attempt</td>
<td>35470</td>
<td>0.092</td>
<td>0.092</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Prep School Expenditure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No prep school</td>
<td>37577</td>
<td>0.262</td>
<td>0.262</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Scholarship</td>
<td>37577</td>
<td>0.008</td>
<td>0.008</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1000 TL</td>
<td>37577</td>
<td>0.223</td>
<td>0.223</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1000-2000 TL</td>
<td>37577</td>
<td>0.095</td>
<td>0.095</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2000 TL</td>
<td>37577</td>
<td>0.034</td>
<td>0.034</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Missing</td>
<td>37577</td>
<td>0.378</td>
<td>0.378</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Time Spend in Preparation in 11th Grade</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkish Preparation</td>
<td>37650</td>
<td>0.439</td>
<td>0.874</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Social Science Preparation</td>
<td>37650</td>
<td>0.465</td>
<td>0.905</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Math Preparation</td>
<td>37650</td>
<td>0.301</td>
<td>0.653</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Science Preparation</td>
<td>37650</td>
<td>0.125</td>
<td>0.393</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>
Figure 3.1: Distribution of OSS-SOZ scores, first vs second time takers

Figure 3.2: Distribution of OSS-SOZ scores, first time takers, male vs female
Figure 3.3: Distribution of OSS-SOZ scores, second time takers, male vs female

Figure 3.4: Density of OSS-SOZ scores, first vs second time takers
Figure 3.5: Density of OSS-SOZ scores, first time takers, male vs female

Figure 3.6: Density of OSS-SOZ scores, second time takers, male vs female
Figure 3.7: Histogram of scores from social science test, first time takers, bins of width one

Figure 3.8: Histogram of scores from social science test, second time takers, bins of width one
Figure 3.9: Histogram of scores from Turkish test, first time takers, bins of width one

Figure 3.10: Histogram of scores from Turkish test, second time takers, bins of width one
Figure 3.11: Histogram of scores from math test, first time takers, bins of width one

Figure 3.12: Histogram of scores from math test, second time takers, bins of width one
Figure 3.13: Histogram of scores from science test, first time takers, bins of width one

Figure 3.14: Histogram of scores from science test, second time takers, bins of width one
Figure 3.15: Distribution of scores from social science test, first time takers

Figure 3.16: Distribution of scores from social science test, second time takers
Figure 3.17: Distribution of scores from Turkish test, first time takers

Figure 3.18: Distribution of scores from Turkish test, second time takers
Figure 3.19: Distribution of scores from math test, first time takers

Figure 3.20: Distribution of scores from math test, second time takers
Figure 3.21: Distribution of scores from science test, first time takers

Figure 3.22: Distribution of scores from science test, second time takers
Figure 3.23: Distributions of signals for a student with $\beta = 3$, approximately median
Figure 3.24: Action conditional on signals for a simple two answer model (parameter values: $\beta = 3$ and cutoff = 0.55)
Table 3.3: Question outcomes for various parameter values: probabilities of skipping, being correct, being incorrect, and the average points per question

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>Cutoff</th>
<th>Prob(S)</th>
<th>Prob(C)</th>
<th>Prob(I)</th>
<th>PPQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.2</td>
<td>0</td>
<td>0.405</td>
<td>0.595</td>
<td>0.257</td>
</tr>
<tr>
<td>2</td>
<td>0.225</td>
<td>0.012</td>
<td>0.403</td>
<td>0.585</td>
<td>0.257</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.085</td>
<td>0.386</td>
<td>0.529</td>
<td>0.254</td>
</tr>
<tr>
<td>2</td>
<td>0.275</td>
<td>0.192</td>
<td>0.359</td>
<td>0.449</td>
<td>0.247</td>
</tr>
<tr>
<td>2</td>
<td>0.3</td>
<td>0.303</td>
<td>0.328</td>
<td>0.370</td>
<td>0.235</td>
</tr>
<tr>
<td>2</td>
<td>0.325</td>
<td>0.403</td>
<td>0.297</td>
<td>0.300</td>
<td>0.222</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>0</td>
<td>0.535</td>
<td>0.465</td>
<td>0.419</td>
</tr>
<tr>
<td>3</td>
<td>0.225</td>
<td>0.003</td>
<td>0.534</td>
<td>0.463</td>
<td>0.419</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>0.030</td>
<td>0.528</td>
<td>0.442</td>
<td>0.418</td>
</tr>
<tr>
<td>3</td>
<td>0.275</td>
<td>0.081</td>
<td>0.515</td>
<td>0.404</td>
<td>0.414</td>
</tr>
<tr>
<td>3</td>
<td>0.3</td>
<td>0.143</td>
<td>0.498</td>
<td>0.360</td>
<td>0.408</td>
</tr>
<tr>
<td>3</td>
<td>0.325</td>
<td>0.208</td>
<td>0.478</td>
<td>0.315</td>
<td>0.399</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
<td>0</td>
<td>0.619</td>
<td>0.381</td>
<td>0.524</td>
</tr>
<tr>
<td>4</td>
<td>0.225</td>
<td>0.001</td>
<td>0.619</td>
<td>0.380</td>
<td>0.524</td>
</tr>
<tr>
<td>4</td>
<td>0.25</td>
<td>0.017</td>
<td>0.616</td>
<td>0.368</td>
<td>0.524</td>
</tr>
<tr>
<td>4</td>
<td>0.275</td>
<td>0.049</td>
<td>0.608</td>
<td>0.344</td>
<td>0.522</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
<td>0.091</td>
<td>0.596</td>
<td>0.314</td>
<td>0.517</td>
</tr>
<tr>
<td>4</td>
<td>0.325</td>
<td>0.137</td>
<td>0.581</td>
<td>0.281</td>
<td>0.511</td>
</tr>
</tbody>
</table>
Figure 3.25: Distribution of scores resulting from various cutoff levels
### Table 3.4: Cutoff Estimates for First Time Takers

<table>
<thead>
<tr>
<th></th>
<th>&lt;90</th>
<th>90-100</th>
<th>100-110</th>
<th>110-120</th>
<th>120-130</th>
<th>130-140</th>
<th>&gt;140</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSS</td>
<td>c</td>
<td>0.2078</td>
<td>0.2186</td>
<td>0.2280</td>
<td>0.2413</td>
<td>0.2501</td>
<td>0.2585</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.0004</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0005</td>
<td>0.0005</td>
</tr>
<tr>
<td>MSS</td>
<td>c</td>
<td>0.2128</td>
<td>0.2212</td>
<td>0.2305</td>
<td>0.2416</td>
<td>0.2488</td>
<td>0.2520</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.0010</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0007</td>
<td>0.0008</td>
</tr>
<tr>
<td>FT</td>
<td>c</td>
<td>0.2186</td>
<td>0.2300</td>
<td>0.2374</td>
<td>0.2509</td>
<td>0.2669</td>
<td>0.2728</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.0006</td>
<td>0.0008</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.0005</td>
<td>0.0006</td>
</tr>
<tr>
<td>MT</td>
<td>c</td>
<td>0.2143</td>
<td>0.2259</td>
<td>0.2351</td>
<td>0.2494</td>
<td>0.2621</td>
<td>0.2680</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.0006</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0007</td>
<td>0.0006</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

### Table 3.5: Cutoff Estimates for Second Time Takers

<table>
<thead>
<tr>
<th></th>
<th>90-100</th>
<th>100-110</th>
<th>110-120</th>
<th>120-130</th>
<th>130-140</th>
<th>&gt;140</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSS</td>
<td>c</td>
<td>0.2200</td>
<td>0.2319</td>
<td>0.2427</td>
<td>0.2483</td>
<td>0.2550</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.0007</td>
<td>0.0009</td>
<td>0.0008</td>
<td>0.0007</td>
<td>0.0006</td>
</tr>
<tr>
<td>MSS</td>
<td>c</td>
<td>0.2394</td>
<td>0.2328</td>
<td>0.2380</td>
<td>0.2421</td>
<td>0.2472</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.0042</td>
<td>0.0011</td>
<td>0.0009</td>
<td>0.0010</td>
<td>0.0010</td>
</tr>
<tr>
<td>FT</td>
<td>c</td>
<td>0.2321</td>
<td>0.2441</td>
<td>0.2525</td>
<td>0.2610</td>
<td>0.2678</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.0007</td>
<td>0.0009</td>
<td>0.0008</td>
<td>0.0007</td>
<td>0.0007</td>
</tr>
<tr>
<td>MT</td>
<td>c</td>
<td>0.2339</td>
<td>0.2377</td>
<td>0.2464</td>
<td>0.2529</td>
<td>0.2554</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.0009</td>
<td>0.0008</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0012</td>
</tr>
</tbody>
</table>
### Table 3.6: Ability Distribution Parameters for First Time Takers

<table>
<thead>
<tr>
<th></th>
<th>&lt;90</th>
<th>90-100</th>
<th>100-110</th>
<th>110-120</th>
<th>120-130</th>
<th>130-140</th>
<th>&gt;140</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1SS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\mu)</td>
<td>-1.40</td>
<td>-0.74</td>
<td>-0.11</td>
<td>0.47</td>
<td>1.03</td>
<td>1.47</td>
<td>1.91</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>0.90</td>
<td>0.86</td>
<td>0.69</td>
<td>0.68</td>
<td>0.67</td>
<td>0.56</td>
<td>0.58</td>
</tr>
<tr>
<td>M1SS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\mu)</td>
<td>-1.50</td>
<td>-0.68</td>
<td>0.04</td>
<td>0.66</td>
<td>1.22</td>
<td>1.74</td>
<td>2.15</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>1.23</td>
<td>0.97</td>
<td>0.89</td>
<td>0.86</td>
<td>0.75</td>
<td>0.79</td>
<td>0.70</td>
</tr>
<tr>
<td>F1T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\mu)</td>
<td>-0.42</td>
<td>0.06</td>
<td>0.57</td>
<td>1.09</td>
<td>1.66</td>
<td>2.24</td>
<td>2.94</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>0.78</td>
<td>0.77</td>
<td>0.68</td>
<td>0.70</td>
<td>0.67</td>
<td>0.69</td>
<td>0.87</td>
</tr>
<tr>
<td>M1T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\mu)</td>
<td>-0.80</td>
<td>-0.17</td>
<td>0.38</td>
<td>0.90</td>
<td>1.42</td>
<td>2.05</td>
<td>2.54</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>0.95</td>
<td>0.85</td>
<td>0.79</td>
<td>0.78</td>
<td>0.71</td>
<td>0.82</td>
<td>0.99</td>
</tr>
</tbody>
</table>

### Table 3.7: Ability Distribution Parameters for Second Time Takers

<table>
<thead>
<tr>
<th></th>
<th>90-100</th>
<th>100-110</th>
<th>110-120</th>
<th>120-130</th>
<th>130-140</th>
<th>&gt;140</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1SS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\mu)</td>
<td>-0.46</td>
<td>-0.15</td>
<td>0.42</td>
<td>0.98</td>
<td>1.42</td>
<td>1.71</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>0.88</td>
<td>0.95</td>
<td>0.89</td>
<td>0.80</td>
<td>0.68</td>
<td>0.55</td>
</tr>
<tr>
<td>M1SS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\mu)</td>
<td>-0.54</td>
<td>0.01</td>
<td>0.59</td>
<td>1.23</td>
<td>1.74</td>
<td>1.95</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>1.22</td>
<td>1.20</td>
<td>1.17</td>
<td>1.06</td>
<td>0.94</td>
<td>0.67</td>
</tr>
<tr>
<td>F1T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\mu)</td>
<td>0.36</td>
<td>0.63</td>
<td>1.14</td>
<td>1.66</td>
<td>2.18</td>
<td>2.32</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>0.78</td>
<td>0.85</td>
<td>0.88</td>
<td>0.79</td>
<td>0.72</td>
<td>0.69</td>
</tr>
<tr>
<td>M1T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\mu)</td>
<td>0.05</td>
<td>0.42</td>
<td>0.92</td>
<td>1.49</td>
<td>1.97</td>
<td>2.26</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>0.99</td>
<td>1.05</td>
<td>1.03</td>
<td>1.01</td>
<td>0.97</td>
<td>0.70</td>
</tr>
</tbody>
</table>
Figure 3.26: Score distribution from data vs simulated distribution resulting from estimates

Figure 3.27: Score distribution from data vs simulated distribution resulting from estimates
Figure 3.28: Score distribution from data vs simulated distribution resulting from estimates

Figure 3.29: Score distribution from data vs simulated distribution resulting from estimates
Figure 3.30: Score distribution from data vs simulated distribution resulting from estimates

Figure 3.31: Score distribution from data vs simulated distribution resulting from estimates
Figure 3.32: Score distribution from data vs simulated distribution resulting from estimates

Figure 3.33: Score distribution from data vs simulated distribution resulting from estimates
Figure 3.34: Distributions of social science ability, as estimated by the model

Figure 3.35: Distributions of Turkish ability, as estimated by the model
Density of SS Ability

Figure 3.36: Distributions of social science ability, as estimated by the model

Density of Turkish Ability

Figure 3.37: Distributions of Turkish ability, as estimated by the model
Figure 3.38: Distributions of social science ability, as estimated by the model

Figure 3.39: Distributions of Turkish ability, as estimated by the model
Figure 3.40: The relationship between the proportion of males and ÖSS-SOZ score quantiles from the counterfactual exercises

Figure 3.41: The relationship between the proportion of first time takers and ÖSS-SOZ score quantiles from the counterfactual exercises
Figure 3.42: The relationship between social science ability and ÖSS-SOZ score quantiles from the counterfactual exercises.

Figure 3.43: The relationship between Turkish ability and ÖSS-SOZ score quantiles from the counterfactual exercises.
Table 3.8: First Time - Low Type Cutoffs

<table>
<thead>
<tr>
<th></th>
<th>&lt; 90</th>
<th>90-100</th>
<th>100-110</th>
<th>110-120</th>
<th>120-130</th>
<th>130-140</th>
<th>&gt; 140</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1SS</td>
<td>0.204</td>
<td>0.217</td>
<td>0.224</td>
<td>0.235</td>
<td>0.250</td>
<td>0.251</td>
<td>0.259</td>
</tr>
<tr>
<td>M1SS</td>
<td>0.207</td>
<td>0.216</td>
<td>0.228</td>
<td>0.238</td>
<td>0.247</td>
<td>0.253</td>
<td>0.256</td>
</tr>
<tr>
<td>F1T</td>
<td>0.213</td>
<td>0.223</td>
<td>0.232</td>
<td>0.244</td>
<td>0.255</td>
<td>0.278</td>
<td>0.268</td>
</tr>
<tr>
<td>M1T</td>
<td>0.211</td>
<td>0.221</td>
<td>0.228</td>
<td>0.248</td>
<td>0.256</td>
<td>0.258</td>
<td>0.259</td>
</tr>
</tbody>
</table>

Table 3.9: Second Time - Low Type Cutoffs

<table>
<thead>
<tr>
<th></th>
<th>&lt; 90</th>
<th>90-100</th>
<th>100-110</th>
<th>110-120</th>
<th>120-130</th>
<th>130-140</th>
<th>&gt; 140</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2SS</td>
<td>0.209</td>
<td>0.216</td>
<td>0.224</td>
<td>0.237</td>
<td>0.251</td>
<td>0.255</td>
<td>0.258</td>
</tr>
<tr>
<td>M2SS</td>
<td>0.208</td>
<td>0.216</td>
<td>0.224</td>
<td>0.236</td>
<td>0.242</td>
<td>0.244</td>
<td>0.245</td>
</tr>
<tr>
<td>F2T</td>
<td>0.218</td>
<td>0.226</td>
<td>0.235</td>
<td>0.244</td>
<td>0.262</td>
<td>0.265</td>
<td>0.264</td>
</tr>
<tr>
<td>M2T</td>
<td>0.213</td>
<td>0.224</td>
<td>0.231</td>
<td>0.238</td>
<td>0.248</td>
<td>0.254</td>
<td>0.258</td>
</tr>
</tbody>
</table>

Table 3.10: First Time - Share of Low Type Students

<table>
<thead>
<tr>
<th></th>
<th>&lt; 90</th>
<th>90-100</th>
<th>100-110</th>
<th>110-120</th>
<th>120-130</th>
<th>130-140</th>
<th>&gt; 140</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1SS</td>
<td>0.835</td>
<td>0.825</td>
<td>0.420</td>
<td>0.280</td>
<td>0.596</td>
<td>0.778</td>
<td>1.000</td>
</tr>
<tr>
<td>M1SS</td>
<td>0.854</td>
<td>0.710</td>
<td>0.579</td>
<td>0.667</td>
<td>0.596</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>F1T</td>
<td>0.844</td>
<td>0.687</td>
<td>0.500</td>
<td>0.553</td>
<td>0.473</td>
<td>0.918</td>
<td>1.000</td>
</tr>
<tr>
<td>M1T</td>
<td>0.905</td>
<td>0.581</td>
<td>0.432</td>
<td>0.523</td>
<td>0.566</td>
<td>0.737</td>
<td>0.996</td>
</tr>
</tbody>
</table>
Table 3.11: Second Time - Share of Low Type Students

<table>
<thead>
<tr>
<th></th>
<th>&lt; 90</th>
<th>90-100</th>
<th>100-110</th>
<th>110-120</th>
<th>120-130</th>
<th>130-140</th>
<th>&gt; 140</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2SS</td>
<td>0.852</td>
<td>0.625</td>
<td>0.319</td>
<td>0.367</td>
<td>0.553</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>M2SS</td>
<td>0.816</td>
<td>0.731</td>
<td>0.610</td>
<td>0.732</td>
<td>0.810</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>F2T</td>
<td>0.828</td>
<td>0.706</td>
<td>0.580</td>
<td>0.474</td>
<td>0.748</td>
<td>0.936</td>
<td>1.000</td>
</tr>
<tr>
<td>M2T</td>
<td>0.877</td>
<td>0.713</td>
<td>0.520</td>
<td>0.372</td>
<td>0.653</td>
<td>0.892</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Figure 3.44: Score distribution from data vs simulated distribution from two type extended model

Figure 3.45: Score distribution from data vs simulated distribution from two type extended model
Figure 3.46: Score distribution from data vs simulated distribution from two type extended model

Figure 3.47: Score distribution from data vs simulated distribution from two type extended model
Figure 3.48: Score distribution from data vs simulated distribution from two type extended model

Figure 3.49: Score distribution from data vs simulated distribution from two type extended model
Figure 3.50: Score distribution from data vs simulated distribution from two type extended model

Figure 3.51: Score distribution from data vs simulated distribution from two type extended model
Appendix A

Chapter Two Proofs

Proof of Lemma 2.3.1

Suppose that markets differ by sunk entry costs. Order the markets so that \( i < j \) implies that \( f_{i e} < f_{j e} \). As markets are identical in other dimensions, it holds that \( \forall i, j, \pi_i = \pi_j \), for any realization of productivity and demand. It follows that:

\[
\max(\frac{\pi_i(x, \varphi)}{1 - \delta}, 0) - f_{i e} > \max(\frac{\pi_j(x, \varphi)}{1 - \delta}, 0) - f_{j e}.
\] (A.1)

As this relation holds regardless of \( x, \varphi \),

\[
E(\max(\frac{\pi_i(x, \varphi)}{1 - \delta}, 0) - f_{i e} | F_\omega(\cdot), \varphi) > E(\max(\frac{\pi_j(x, \varphi)}{1 - \delta}, 0) - f_{j e} | F_\omega(\cdot), \varphi) \tag{A.2}
\]

Similarly with operating costs: if \( f_{i o} < f_{j o} \), \( \pi_i(x, \varphi) \geq \pi_j(x, \varphi) \). This inequality is strict for values of \( (x, \varphi) \) such that \( \pi_i(x, \varphi) > 0 \). As \( \varphi > 0 \) and, for all rational beliefs, \( x \) has full support over \( \mathbb{R}^+ \), the result follows.

Similarly with effective market size \( Y \), \( \frac{Y_i}{\bar{p}_i - \sigma} > \frac{Y_j}{\bar{p}_j - \sigma} \) leads to \( \pi_i(x, \varphi) \geq \pi_j(x, \varphi) \).

Proof of Proposition 2.3.2

A firm has entered \( n \in [0, N - 2] \) markets. Without loss of generality, it has entered the first \( n \) markets. Let \( i < j \).

If the firm enters market \( i \) it receives:
\[ v_i = E[\max(\frac{\pi_i}{1-\delta}, 0) - f_{ie} + \delta \text{Value}(j, j+1, ..., N|\text{beliefs}, \varphi)] \quad (A.3) \]

where \( \text{Value}(A) \) is a function that gives the expected value of the option to enter markets in the set \( A \), in the order listed.

If the firm enters market \( j \) it receives:

\[ v_j = E[\max(\frac{\pi_j}{1-\delta}, 0) - f_{je} + \delta \text{Value}(i, j+1, ..., N|\text{beliefs}, \varphi)] \quad (A.4) \]

Suppose that, subsequent to entering \( j \), the firm has a correspondence between beliefs and entry in the next period, i.e. if observed demand is in the interval \([0, A)\) then do not expand, if it is in \([A, B)\) then expand to the next market, \([B, C)\) expand to the next two etc (with \(0 < A < B < C \) and so on).

Now, compare \( v_j \) to \( v_i^* \), where expansion is restricted to follow the same rule as in \( v_j \). As the observed demand is independent of the market (in expectation) we can simplify \( v_i^* - v_j \) as:

\[ (\max(\frac{\pi_i}{1-\delta}, 0) - f_{ie}) - (\max(\frac{\pi_j}{1-\delta}, 0) - f_{je}) - \delta \text{Prob}(\text{demand} \geq A) (\max(\frac{\pi_j}{1-\delta}, 0) - f_{je}) - (\max(\frac{\pi_i}{1-\delta}, 0) - f_{ie}) \quad (A.5) \]

which is strictly greater than zero, as \( \delta < 1 \). As this expansion rule is feasible, \( v_i > v_i^* > v_j \).

So the firm will never choose to switch any two markets in the expansion order.

Proof of Proposition 2.3.3

\( V(N, Y) = 0, \forall Y \), as there are no markets available to enter. It follows that \( V(N-1, Y) \) is well defined, as \( W(\cdot) \) is well defined and finite under the distributional assumptions. By induction, \( V(n, Y) \) is well defined and finite, for all \( n \leq N \). As there is a finite choice set to the problem, a solution exists.

Proof of Proposition 2.3.4

Suppose that \( n = N - 1 \), i.e. there is a single market remaining. In this case the entry decision is simply to enter the final market if \( W(n, X, \varphi, 1) \geq 0 \). Since
$W(\cdot)$ is increasing in $X$, this holds.

Suppose that $n = N - 2$, i.e. there are two markets remaining. The difference between expanding at all and not is $V(\cdot, 1) - V(\cdot, 0)$, which is increasing in $X$, as both $W$ and $V$ are increasing in $X$.

The difference between expanding to one market and both markets is increasing in $X$, as the profit forgone due to delay increases while the probability of information dissuading entry decreases.

Similarly for when there are more than two markets remaining.

Proof of Proposition 4

The joint distribution of productivity/$\alpha$ for firms entering market $n$ first order stochastically dominates that of firms entering market $m < n$. As a result, exit rates are lower in market $n$ than in market $m < n$. 
Proof of Proposition 3.4.1

Replace A with cA. Now, for any \( P \in [0, 1) \), the signal generated by the correct answer, \( F^{-1}(P) \), will be \( c \) times as large. Similarly for signals generated for the incorrect answers. Compare this to a situation where the student arbitrary decides to inflate all signals by \( c \). This will clearly have no impact on the decisions/outcome probabilities of a rational agent, but mirrors what would be seen when replacing A by \( cA \). □

Proof of Proposition 3.4.2

Let the student transform all signal vectors \( X \) to \( Y \), such that \( y_i = (x_i)^\alpha \). The correct answer now has a signal distribution of:

\[
F(y_i) = 1 - \frac{B}{y_i} \tag{B.1}
\]

where \( B = A^\alpha \). Similarly, the incorrect answers’ signals have the following distribution:

\[
G(y_i) = 1 - \left( \frac{B}{y_i} \right)^{\beta/\alpha} \tag{B.2}
\]

So this re-scaling of the signals preserves all of information contained in the original signals, and the resulting signals have a distribution identical to one where the correct answers come from a Pareto distribution with the shape parameter equal to one, and the incorrect answers have shape parameter equal to \( \beta/\alpha \). As shown in the earlier proposition, the scale parameter is irrelevant. □

Proof of Lemma 3.4.3
The belief is given by $\frac{x_{\text{max}}}{\sum_{m=1}^{5} x_\gamma}$. Taking logs, and differentiating with respect to $\gamma$, yields the following expression:

$$\frac{d \log(\text{Belief})}{d\gamma} = \log x_{\text{max}} - \frac{x_1^\gamma \log x_1 + x_2^\gamma \log x_2 + x_3^\gamma \log x_3 + x_4^\gamma \log x_4 + x_5^\gamma \log x_5}{x_1^\gamma + x_2^\gamma + x_3^\gamma + x_4^\gamma + x_5^\gamma}$$

(B.3)

Since $\log x_{\text{max}} \geq \log x_i$, and $x_i > 0$,

$$\frac{d \text{Belief}}{d\gamma} \geq 0$$

(B.4)

With the inequality strict unless $x_1 = x_2 = x_3 = x_4 = x_5$. Since $\gamma \equiv \beta - \alpha$, the student with the highest value of $\beta$ has the strongest belief ($\alpha = 1$ for both students). □
Bibliography


Vita

James Key

James Key was born in New Zealand in 1983. He received a Bachelor of Science and a Bachelor of Commerce and Administration from Victoria University of Wellington in 2005. The next year he received a Honours degree in Commerce and Administration from Victoria University of Wellington, and enrolled in the Economics department at the Pennsylvania State University in 2007. In 2013 he completed his studies in the fields of international trade, development, and education.