Expense ratios of North American mutual funds

Karen Ruckman   *John Molson School of Business, Concordia University*

Abstract. The average expense ratio paid by Canadian mutual fund investors is 50% higher than that paid in the United States. This discrepancy is commonly thought to exist because Canadian funds do not take advantage of economies of scale and have less competition. A monopolistic competition framework is used to develop a model for the mutual fund industry. By allowing each fund to have different attributes, the model permits funds to charge different expense ratios in equilibrium and is found to strongly fit the North American mutual fund market. Empirical analysis indicates that these two common explanations and measurable fund attributes account for 24% of the discrepancy. JEL Classification: L11, L13 and G15

Les ratios de dépenses des fonds mutuels nord-américains. Le taux moyen de dépenses payées par les investisseurs canadiens dans les fonds mutuels sont de 50% plus élevées que celles qu'on paie aux Etats-Unis. Cet écart est attribué d'habitude au fait que les fonds canadiens ne tirent pas profit des économies d'échelle et qu'il y a moins de concurrence au Canada. On utilise un modèle de concurrence monopolistique pour analyser l'industrie des fonds mutuels. En permettant à chaque fond d'avoir certains attributs, le modèle permet aux fonds de charger des taux de dépenses différents en équilibre. Il semble que cela corresponde aux caractéristiques du marché des fonds mutuels américains. Une analyse empirique montre que les deux explications usuelles et les attributs mesurables des fonds expliquent 24% de l'écart.

1. Introduction

Investments in the Canadian mutual fund market by the end of 2001 totalled $CDN 450 billion or 84% of the amount of money invested in guaranteed

investment certificates, chequing and savings accounts combined (source: Investor Economics). The average Canadian management expense ratio (MER), the fee paid by mutual fund investors, is consistently over 50% higher than its American counterpart. If Canadian investors paid the same average MER as U.S. investors, they would have saved over $CDN 4 billion in 2001. In this paper why MERs are higher in Canada than in the United States is investigated.

The two most common explanations for the discrepancy in MERs are that Canadian funds do not take advantage of economies of scale and have fewer rival funds. The monopolistic competition structure employed in this paper has the advantage of accounting for these two explanations. An industry of hetero-geneous but substitutable products is a hallmark of monopolistic competition and this feature is demonstrated in the mutual fund industry by the extent to which mutual fund companies advertise the distinctions of their managers' ability, fund orientation and different fee structures. Canadian regulations do not allow U.S. residents to purchase Canadian-owned mutual funds or U.S.-owned mutual funds to be sold to Canadians unless they are registered with a provincial commission. These regulations essentially segregate the Canadian and U.S. mutual fund markets, and this allows the model developed to be estimated on the pool of funds from both countries while controlling for the differences between the two markets. An estimation of the model determines the extent to which these two common explanations account for the difference in MERs and reveals the mark-up in Canadian MERs that remains unexplained.

In traditional monopolistic competition models used in empirical studies, firms are assumed to be fully symmetric, with every firm charging the same price for its differentiated product. In this paper we relax this symmetry assumption and, in doing so, allow consumers to choose funds according to their tastes for the differing fund characteristics and funds to charge different prices, or MERs, in equilibrium. The MER pricing decision is a function of the fund's individual characteristics, the competition faced by a fund and the size of the fund, the latter two reflecting the two common explanations for the MER discrepancy.

To determine the extent to which the model fits the data, the actual number of funds in the markets is compared with those predicted by the model. The actual values are found to be in the predicted region and justify estimating the MER price equation and the associated demand equation. The regression results show that differences in fund sizes, intensity of competition and measurable fund attributes account for about 24% of the Canadian MER mark-up. The Canadian MERs are even higher than the monopolistic competition model would predict, which suggests that the mark-up is due to either monopoly power beyond that allowed by monopolistic competition or a difference in accepted distribution practices. In addition, using the U.S. market as a base, the number of Canadian mutual funds predicted by the model can be calculated

using the estimated parameters. The model's predicted number of Canadian funds is very close to the actual number of funds and is shown to be closer than the simpler form of monopolistic competition that does not allow for heterogeneous funds.

This paper differs from the existing literature because we attempt to explain mutual fund fees by utilizing a monopolistic competition model. Previous authors have recognized the importance of economies of scale and other fund-specific variables in explaining mutual fund performance.[1] Recently, Carhart (1997) concludes that the three fund characteristics, MERs, transaction costs, and turnover rates, explain almost all the persistence in returns. Grinblatt and Titman (1994) suggest that turnover rate is significantly related to the ability of the fund managers to earn abnormal returns. There are also a number of studies in which the determinants of mutual fund fees are examined in a purely empirical context.[2] Trzcinka and Zweig (1990) and Malhotra and McLeod (1997) conclude that funds with 12b-1 fees have larger MERs. Malhotra and McLeod (1997) also determine that economies of scale have a significant role in mutual funds fees. We also offers an extension to the traditional monopolistic competition model that allows for varying firm characteristics that can easily be applied to other industries.

This paper is organized as follows. In section 2 we provide the background information about the U.S. and Canadian mutual fund markets with a discussion of the common explanations for the difference in MERs. In section 3 we develop the consumer choice theory and the monopolistic competition model in the context of this industry and examine how closely the model fits the data. In section 4 we estimate the MER pricing equation and the demand equation developed in the model and use the estimated parameters to examine the existing number of funds in Canada. In section 5 we conclude the analysis and offer explanation for the unexplained Canadian mark-up.

2. Background information

The mutual fund data are obtained from Morningstar Inc. for the United States and PALTrak Inc. for Canada. It is a cross-section as of 31 January 1999. These tracking services provide a large amount of fund-level data for all mutual funds offered in their respective countries. The PALTrak service has a good coverage of the Canadian industry. The usable data from PALTrak (about 10% of the observations were dropped, owing to missing observations)

---

1 There is also a set of studies linking the form of fund manager compensation with fund performance. Significant contributors include: Brown, Harlow, and Starks (1996) and Berkowitz and Kotowitz (1993).
2 Tufano and Sevick (1997) connect board of director characteristics with fund fees and determine that smaller boards and boards with a larger fraction of independent members approve lower fees.

TABLE 1
North American equity mutual funds

Country	Total assets ($USbil)	Number of funds	Mean fund size ($USmil)	Mean MER (%)	Mean 3-year annualized return (%)
United States	2,926	5,749	509.1	1.50	13.92
Canada	203	1,455	140.0	2.29	9.28
U.S.:Canada ratio	14.41	3.95	3.64	0.66	1.50

total about 85% of industry assets reported by Investment Funds Institute of Canada (IFIC). The IFIC is a regulatory institute whose membership list includes 97% of Canadian mutual funds. Bond and money market (maturity less than one year) mutual funds are not monitored closely by the U.S. data source, Morningstar Inc. Morningstar reports that there was $US 772 billion invested in bond and money market funds in the U.S. but the Investment Company Institute (ICI), the national association of U.S. mutual funds, claims that there was $US 2,182 billion. It is supposed that these short-term funds are not reported by the companies that own them because they are used as a temporary investment in between other investment opportunities for company's current investors and does not necessitate exposure in the Morningstar dataset. There is an incentive to expose equity funds, on the other hand, since investors that are not current clients regularly purchase them. The sum of U.S. equity fund assets in our dataset is 88% of the sum claimed by ICI. Again, about 10% of the U.S. observations were omitted, owing to missing observations. The analysis in this paper will focus only on equity funds, and it can be assumed that 'funds' in this paper will always refer to equity funds. Table 1 reports some introductory statistics.

The total value of the American mutual fund industry is over 14 times larger than that of Canada. There are almost four times as many funds in the U.S. than in Canada, and the average fund size in the United States is about three and a half times larger than in Canada. The most remarkable difference between the two markets is the mean MER. Both countries define a MER as the ratio of a fund's total costs to total assets and report it as a percentage. It is the sum of management fees (these are company profits and fees to fund managers if contracted out), operating expenses, distribution fees and commissions to sellers. MER is regarded as the price of a mutual fund even though it is paid on continual basis by netting-out returns. The mean Canadian MER is about 50% higher than the mean U.S. MER.

It is instructive at this point to divide the funds into mutually exclusive categories according to their objective and orientation. For a detailed discussion of the fundtype categorization, see the data appendix. Dividing the funds into fundtypes is informative because there are a number of different types of funds associated with very different attributes. For instance, mutual funds that

invest in large capitalization stocks will have significantly lower MERs than funds that invest in emerging market equities because of smaller research expense and transaction costs. If Canada had a relatively small proportion of funds in lower MER fundtypes, the difference in the over-all average MER could be simply explained.

Sometimes the mean of a variable can be a misleading statistic, and it is helpful to view the distribution of that variable for the whole population. Figure 1 displays the distribution of MERs for all fundtypes for both Canada and the United States. The upper-left graph demonstrates that the mean is not a misleading indicator of the whole distribution, and, as the other graphs confirm, every fundtype has the Canadian distribution of MERs at a higher level than that of the United States. The lower-right graph is not for a fundtype but rather for a characteristic of funds. Index funds are found in every fundtype, and they mimic movements of specific stocks, industries, or various other financial indicators. They are often considered to be cheap to produce, as is evident by their lower range of MERs.

Table 2 displays the different fundtypes along with their descriptive statistics. For every fundtype, the mean U.S. MER is smaller than the mean Canadian MER and the U.S. mean fund size is larger than the Canadian mean fund size. There is a large variation of mean fund size differences ranging from about twice as large (Balanced funds) to five and a half times larger (Blend funds). The proportion of a country's assets in a given fundtype allows us to determine the predominance of certain types of equity funds found in the country. For instance, Canada has a higher proportion of foreign oriented funds (24%) compared with that of the United States (15%). This may be because the U.S. economy has a more diversified industrial structure. Also, the United States has a higher proportion of Growth funds and Canada has relatively more Balanced funds.

The mean market share of funds in each category is the inverse of the number of funds. This accounts for the difference in fundtype size, while it proxies the degree of competition. Canada's Growth funds on average have a large market share within that category (4.17%), while U.S. Growth funds have on average a very small market share (0.14%). The difference between average market shares in Canada and the United States for other fundtypes is less dramatic but always positive, indicating there is less competition in Canada in every fundtype.

The fundamental question is: *Why are the MERs higher in Canada?* Experts in the Canadian mutual fund industry often attribute the high Canadian MERs to the following five factors.[3] Each explanation is followed with a critical analysis. Regardless of the critique, each explanation will be accounted for in the empirical examination.

---

3 John Kaszel, director of research at the IFIC, was especially helpful in determining these five factors.
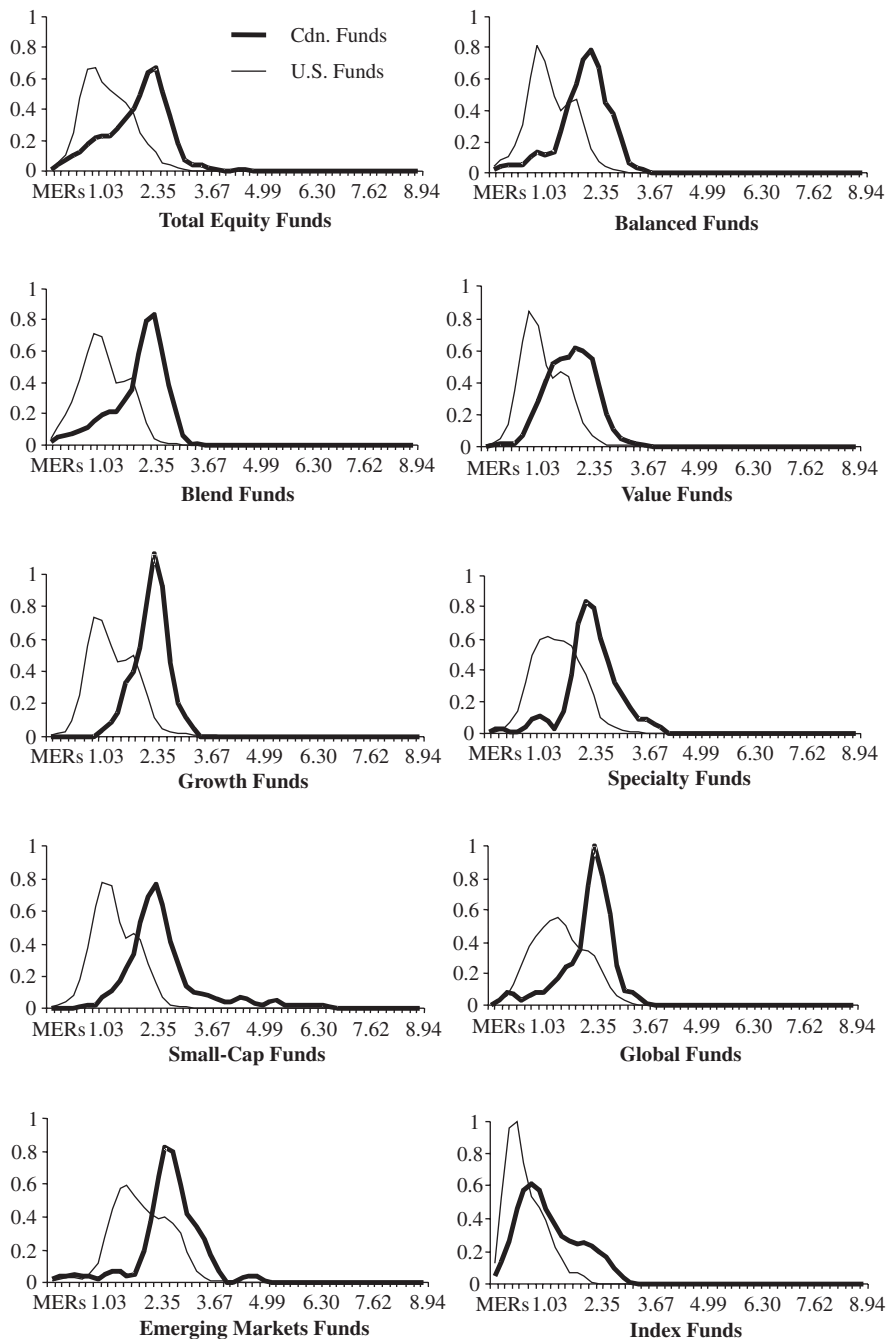
FIGURE 1   Distribution of MERs

TABLE 2
Descriptive statistics, by fundtype

Fundtype		Country	Mean MER (%)	Mean fund size ($USmil)	Mean 3-year return (%)	Fundtype's share of country's total market (%)	Mean market share of fund in fundtype (1/n)%
Domestic:	Balanced	U.S.	1.33	450	14.61	9.87	0.155
		Canada	2.16	203	9.77	24.52	0.406
	Blend	U.S.	1.26	949	21.95	30.51	0.106
		Canada	2.15	168	9.64	19.41	0.423
	Value	U.S.	1.36	848	17.66	19.22	0.150
		Canada	2.03	240	13.13	12.13	0.970
	Growth	U.S.	1.49	668	22.19	16.47	0.138
		Canada	2.43	195	13.95	2.30	4.166
	Specialty	U.S.	1.65	239	11.02	4.40	0.185
		Canada	2.50	57	−8.39	2.12	1.315
	Small-cap	U.S.	1.53	206	12.84	4.75	0.148
		Canada	2.77	78	4.53	3.62	1.052
Foreign:	Global	U.S.	1.66	302	6.59	14.30	0.072
		Canada	2.39	125	7.77	24.27	0.252
	Emerging	U.S.	2.13	78	−10.82	0.48	0.549
		Canada	2.82	28	−7.82	0.70	1.923
U.S.-invested Canadian Funds		Canada	2.15	98	22.26	11.03	0.437

a) *There are a larger number of index funds in the United States that tradition- ally have much lower MERs that push down the mean MER.*

There is a larger number of index funds in the United States (163 compared with 60 in Canada) but the ratio of index funds is not substantially different than the ratio of the total number of funds. However, index funds comprise 1.51% of the total assets of the Canadian market and 7.04% of the U.S. market. This difference may explain part of the difference in MERs and will be controlled for in the regression analysis.

b) *Higher marketing costs in Canada caused by typical load structure.*

Many mutual funds are sold with either a front or rear load. A front load requires that the consumer pay a commission directly to the broker when the fund is purchased. Rear loads (or 'deferred' loads) are paid to the company when the consumer withdraws her investment. They usually decline over time, so that long-term investors are not subjected to the fee. When a rear-loaded fund is sold, the mutual fund company pays a commission to the investment adviser and part of the MER is used to fund the initial commission. In fact, Dellva and Olson (1998) conclude that funds sold with front-end loads generally have lower expenses and this is reflected in a lower expense ratio.

If there is a higher proportion of rear-load funds in Canada, then this would partially explain the difference in MERs. It is difficult to determine this, as Canadian mutual funds sold with an option of front or rear loads do not have to disclose what proportion was sold with each. However, anecdotal evidence indicates that rear-loaded funds are outselling front-loaded funds in Canada by a factor of four to one (*Financial Post* 1998b).

c) Higher costs in Canada due to higher trailer fees.

Trailer fees are paid by the mutual fund company to the investment adviser for continued customer service. All funds, whether they are sold with a rear or front or without a load, pay trailer fees if sold by a third party. In the United States this is the 12b-1 fee and is named for the SEC rule that created and permits it. Trailer fees on load funds are usually between 25 and 50 basis points, but it is not possible to compare the average fees between the two countries because the Canadian information is not readily available. Canadian trailer fees are required to be disclosed on prospectuses, thanks to recent efforts by Glorianne Stromberg (1998), but the coverage by tracking agencies is not comprehensive.[4] The average trailer fee in Canada is believed to be about twice that in the United States (see *Financial Post* 1998a). However, even if the trailer fees in Canada were readily available, the amount of a fund that had been sold with a certain trailer fee and load combination would also have to be disclosed to properly determine the average trailer fee in Canada.

The reason for the difference between the size of trailer fees in Canada and the 12b-1 fees in the United States is assumed to be institutional. For example, one possible reason for the difference is that U.S. mutual funds are sold through independent advisers more than they are in Canada. Independent advisers do not directly benefit from a trailer fee and would not have the incentive to suggest a client to purchase shares in a mutual fund because it has a particular trailer fee or load structure. This may lead to a lower average 12b-1 fee in the United States. Unfortunately, there is no data that indicate the proportion of Canadian mutual fund consumers that utilize independent advisers.

d) The larger average fund size in the U.S. leads to economies of scale.

Economies of scale play a role in mutual funds through recapturing fixed costs and also reduction in marginal costs in large volumes. If total costs are assessed every time a fund's assets are traded, then costs associated

---

4 Not all funds give their associated trailer fees to tracking agencies. They are found embedded in fund prospectuses but what with over 7000 funds in the dataset, the labour needed to compile them is beyond what we have available. Even if this was possible, Canadian funds are sold with differing load and trailer options, and it would be impossible to determine how much of the fund was sold with one trailer and how much with another.

with trading assets are considered to be marginal costs because they are related to fund size. Larger mutual funds can perform block trading at lower prices, and their size may allow them to command lower transaction charges. Commission fees paid to the broker are also marginal costs, which can also be negotiated down, owing to an increase in bargaining power on the part of the mutual fund company because of fund reputation and success. The reduction of these marginal costs due to size leads to economies of scale. Research expenses are fixed costs, because they are not related to fund size and yet are expended continually. Research and other operating costs are more easily recaptured with larger fund size, also leading to economies of scale.

e) The average size of families of funds in Canada is smaller than in the U.S., which does not allow Canada to take advantage of economies of scope.

If a mutual fund company offers various different funds, they are often grouped together as a 'family.' Families may be subject to economies of scope because industrial research may serve more than one fund. A family may become more efficient at researching one type of fund, given that they are knowledgeable in other fundtypes. There are 578 fund families in the United States compared with 183 in Canada, or a United States to Canada ratio of 3.15. There are more fund families in Canada relative to the size of the market, thus leading to higher average family costs. There are an average of 16 funds in American fund families and 10 in Canadian families. The absolute size of fund families in Canada is smaller than that of the United States and does not take advantage of economies of scope.

Even though in the previous section the impact of loads and trailers on MERs was introduced separately, a better discussion would talk about the combination of MER/load/trailer together. Many funds are offered to customers with a choice of three combinations of MER and loads (and trailers, although trailers are not emphasized to mutual funds consumers). The typical pattern is either (i) front load/low trailer/low MER, (ii) rear load/high trailer/high MER, or (iii) no load or small rear load/high trailer/high MER. It is not possible to take into consideration each mutual fund's individual combination, and because of deficiencies in the data that will be discussed in the regression results section, in the empirical part of this paper we will keep these factors separate when attempting to determine their contribution to the difference in MERs between the United States and Canada.

It has been established that mutual funds in Canada have higher MERs. This begs the question: Why, then, do Canadians buy Canadian-owned mutual funds, at all? The answer to this involves a discussion of legalities. Each province has a securities commission that oversees securities including mutual funds. To sell mutual funds in any province, a fund must be registered with the commission in that province. The IFIC has regulations that prohibit U.S.

residents from purchasing Canadian mutual funds. These regulations suggest that the Canadian and American mutual fund markets are segregated. That is, only Canadians can buy Canadian-owned mutual funds and U.S.-owned mutual funds can only be sold to Canadians if they have an affiliate registered in a Canadian province.

There are, however, three ways Canadian consumers can bypass the law to buy U.S.-owned mutual funds. First, there is nothing to prevent a Canadian from going directly to a U.S. fund to make a purchase. This is commonly referred to as an unsolicited sale. Each fund organization then decides to accept or reject such a request. U.S. mutual fund companies that have a Canadian affiliate automatically reject these requests, but other U.S. companies without affiliates are recently joining suit. A possible reason for this trend is an increasing concern over antagonizing Canadian authorities, especially if they are considering entering the Canadian market in the future. Second, a consumer can set up an account with an American broker, who then buys the U.S.-owned fund. Third, it is possible to buy funds through the Internet, even though each province's securities commission requires that foreign companies that sell securities to people even through the Internet must be registered with the commission in that province. Clearly, each of these loopholes is difficult to monitor. There is another institutional reason why the markets are virtually segregated. Canadian investors must not invest more than 30% of their investment income into non-Canadian based investments to claim retirement savings plan (RSP) tax benefits. This is usually enough incentive for the majority of Canadian mutual fund investors to purchase Canadian-based mutual funds.

Precise data are not available on how much of the U.S. mutual fund market is owned by Canadians. The Investment Company Institute, a U.S. organization, reports the total sales of U.S.-based funds to Canadians, but their report may not include all U.S. funds. It claims that in 1990 Canadians purchased $US 28 million in U.S. funds, comprising 0.02% of the U.S. market and, converting to Canadian dollars, amounts to 0.13% of the Canadian market. These fractions have been steadily declining since the 1960s. Because these figures are miniscule, their effects will be ignored and the two markets will be regarded as segregated.

To summarize, the mutual fund industry has many heterogeneous but substitutable goods, each with fund-specific attributes. This affords each fund some market power, and this is shown in funds charging different MERs. The funds may be subject to economies of scale, and the prices charged may also depend on the degree of competition. In the following section we will link these assumptions with a monopolistic competition framework to model the mutual fund industry. Then, using the fact that the Canadian and American markets are virtually segregated, the attributes of the two markets can be compared to test how closely the model fits the data.

3. Monopolistic competition model

The basic monopolistic competition framework has a large number of firms producing and selling goods that are close substitutes, allowing each firm a degree of monopoly over the sale of its own product. The other basic premise of the model is that each firm faces a demand that has firms selling less the greater are the number of firms in the market and selling more the higher are the prices charged by its rivals. In equilibrium, a firm's average cost depends on the size of the market and the number of firms in the industry: the more firms there are in the industry, the higher is a firm's average cost. Even though there is free entry in the market, there can be a short-run equilibrium involving profits. The long-run equilibrium, however, has each firm earning zero economic profits, charging a price equal to average cost and, compared with perfect competition, operating at excess capacity.

Much of empirical work based on monopolistic competition models is based on the assumption that products differ in a qualitative manner. Fortunately, mutual funds have many observable and measurable characteristics. The incorporation of these attributes into the model causes fund demand to depend not only on market size and number of rivals but also on these fund-specific characteristics. In addition, most empirical monopolistic competition models impose symmetry on firms for simplicity; in this paper, however, the inclusion of fund-specific attributes permits funds to charge different MERs and earn non-zero profits in equilibrium. At the time of entry, it is assumed that the response of the market to the particular characteristics of the fund is not known, and hence funds in each category face the same expected return prior to entry. It follows that at the free-entry equilibrium, expected profits are driven to zero.

3.1. Consumer's problem
A consumer's mutual fund investment decision in this model follows a two-stage process. In the first stage, the investor decides what proportion of her income is available for mutual fund investment to optimally allocate in the various fundtypes (i.e., 20% in growth funds, 30% in value funds and 50% in balanced funds) based on expected fundtype return and variance of fundtype returns.[5] In the second stage, one fund in each fundtype is chosen based on the fund's MER and individual attributes and the consumer beliefs about the value of the fund's attributes in predicting future returns.

The benefits of a two-stage decision problem are three-fold. First, it incorporates the essential concept of risk (variance of returns). Without the sequential

---

5 The allocation of investment income to alternatives (stocks, bonds, bond funds, gambling, etc.) is done in stage 1 simultaneously, since she allots income to mutual fund categories. We model only the within fundtype choice for mutual funds and assume that the other investment choices are solved implicitly.

decision problem, risk would have to be incorporated directly into the demand function. Unfortunately, there is no ideal risk measure because many funds are not old enough to warrant an accurate risk assessment over time. The first stage of the consumer decision allows for risk between fundtypes and explains why consumers allocate income to fundtypes with returns that do not vary together. That is, they allocate income to fundtypes with low covariance of returns. Secondly, the 2-stage decision process justifies comparing funds with other funds only in its own fundtype. Without it, funds would be compared (in terms of MER and individual attributes) with all other funds across fundtypes. This is important because it would not be appropriate or useful to compare the returns from two funds in such different fundtypes as growth funds and balanced funds with the same average return. The balanced fund could have a return lower than the average, suggesting that it is an unattractive fund, even though its return may be a higher than the average balanced fund. Lastly, this two-stage decision process reflects the advice presented in the personal investment literature (see, among many others, Gadsden 1998).

A consumer's mutual fund utility is increasing in fund returns, decreasing in variance of returns and is subject to constant relative risk aversion. This implies that the proportions invested in the different fundtypes are invariant to different levels of initial wealth. When a consumer is deciding what proportion of her available mutual fund income to efficiently allocate to the different fundtypes, she observes the average expected return and expected risk (variance of returns) for each fundtype. Funds within a given fundtype have the same risk[6] and risk varies between fundtypes.

A consumer derives utility from a fund's future returns but also from a low MER and other fund characteristics such as whether it is sold with a load. The consumer does not know the future returns of any fund and must use observable characteristics, including the history of a fund's returns, as predictors of future returns. The utility consumer k receives from fund i in fundtype f, μ_{if}^k, is equal to the average return in fundtype f net of fundtype average MER, r_f, plus a variable that reflects the fund's observable characteristics, a_{if}^k.

$$\mu_{if}^k = r_f + a_{if}^k, \text{ where } a_{if}^k = f\left(P_{if} - \overline{P}_f, A_{if}, \varepsilon_{if}^k\right). \tag{1}$$

---

6 This assumption is primarily made because of the limited time series (three years) with which to calculate fund-level variances. I have experimented with adding the variance of three-year fund-level returns as an exogenous variable in the demand equation. It had a positive, but not statistically significant estimated coefficients. This is probably attributable to the imprecision of the three-year measure.

Consumers value funds with MERs, P_{if}, below the average fundtype MER, $\overline{P}_f$, $(P_{if} - \overline{P}_f < 0)$. The variable A_{if} reflects fund i's attributes, and consumers value higher values of this variable. The term ε_{if}^k reflects consumer k's beliefs about the value of the fund i's different attributes.

The fund attributes, A_{if}, are a linear summation of six characteristics: the age of the fund, whether it has a load, the difference of the fund's past return from the fundtype's mean past return, if is an index fund, the number of associated family members, and whether it is RRSP/IRA-eligible. There may be other fund attributes that consumers value, but these six are observable and measurable and will be used in the empirical section to follow. Empirically, each of the six attributes is differenced from the mean fundtype value for the attribute and, therefore, the expected value within a fundtype is zero. This is a within fundtype monotonic transformation that does not compromise the interpretations of the final results.

Consumers have different beliefs about the ranking of the fund characteristics in terms of their ability to predict future net returns. For instance, some consumers rank the difference of a fund's past return from its fundtype's mean past return as the highest of the six attributes to indicate future returns. Other consumers may believe that the age of a fund correlates with higher future returns because of greater experience and rank this attribute among the highest of the six. Some consumers believe many family members is a good indicator of future returns because of positive economies of scope.

Consumer k's utility of her mutual fund investment (R_k) is

$$
U(R_k) = \ln\left\{ \sum_f t_f^k \left[\sum_{i=1}^{n_f} \mu_{if}^k I_{if}^k \right] - \frac{1}{2} \sum_f \sum_g t_f^k t_g^k C_{fg} \right\}^7,
\tag{2}
$$

where $I_{if}^k =$ an indicator function that equals 1 if fund i chosen by consumer k; 0 otherwise

$t_f^k =$ the fraction of mutual fund investment income allocated to fundtype f (non-negative) by consumer k

$C_{fg} =$ the co-variance of fundtype f and g's average expected net returns.

---

7 The expected utility function is additively separable and builds upon Hey (1979, 49) and Borch (1968, 50). Their investigation is a one-stage decision of the optimal fraction of income to allocate to individual stocks. The expected utility above adds to their work by changing the optimal fraction of mutual fund investment income from an allocation across stocks to an allocation across fundtypes and allowing for different attributes of the funds within their fundtypes in the second-stage choice of which funds to buy. It is necessary to incorporate fund-specific characteristics because the first stage of the decision process observes only average fundtype expected risk and returns and does not reflect the fact that consumers expect the fund they invest in will have a return different from the average fundtype return.

Recall that the consumer chooses only one fund per fundtype;[8] thus, the term inside the square brackets in equation (2) will equal the expected utility (μ_{if}^k) for the chosen fund for each fundtype. This structure reflects that consumer utility depends only on the funds chosen for investment. In fact, the term in the square bracket is the object to be maximized in the second stage of the consumer's problem. The negative sign in front of the second of the two terms inside the logarithm function in equation (2) indicates that investment in every fundtype is not worthwhile. If there is a high covariance of returns between two fundtypes, the consumer's utility will decrease if she invests in both. Not all fundtypes have a covariance of mean returns. This explains why consumers invest in a few fundtypes, particularly those who have no or small co-variation of mean returns (i.e., Balanced and Growth).

In the second stage, consumer k chooses $\{I_{if}^k\}$ to maximize $\sum_{i=1}^{n_f} \mu_{if}^k I_{if}^k$ for every fundtype f subject to the constraint the only one fund may be chosen per fundtype. A consumer uses her beliefs about the value of fund attributes (ε_{if}^k) to choose the fund in each fundtype that will give her the maximum utility. Because this belief function is implicit, we cannot solve for the set of indicator functions $\{I_{if}^k\}$ but can determine the probability of each fund's being chosen. The probability function is the expected value of the indicator function.

The probability of a fund's being chosen depends on the number of funds in the fundtype and some function that reflects the perceived value of its attributes. The probability of fund i's being chosen by consumer k is

$$\Pr_{if}^k = \frac{1}{n_f} + a_{if}^k. \tag{3}$$

An increase in the number of funds in a fundtype decreases the probability of a fund's being chosen, regardless of its attributes. A candidate functional form for a_{if}^k [9] is

$$a_{if}^k = -b\left(P_{if} - \overline{P}_f\right) + A_{if} + \varepsilon_{if}^k, \tag{4}$$

---

8 In reality, a consumer usually allocates her money among a few types of equity funds, but chooses more than only one fund in each of those categories. The theory accommodates a consumer choosing to buy two (or more) funds in a given fundtype by viewing the funds as being bought by two (or more) consumers with identical tastes. It is possible, using linear combinations of the six fund attributes, to find two (or more) funds in a fundtype that can give a consumer equal expected utility. The consumer is, therefore, indifferent between these two funds, and so buys one of these funds while the identical consumer buys the other.

9 There are two restrictions on a_{if}^k that must be satisfied. First, the sum of a_{if}^k over all funds in a fundtype is zero, since the sum of probabilities of being chosen over all funds in a fundtype must equal one ($\sum_{i=1}^{n_f} \Pr_{if}^k = 1$). Second, a_{if}^k must be bounded by $\{-1/n_f, 1 - 1/n_f\}$, so that the probabilities of any fund's being chosen by any consumer are always between zero and one.

where ε_{if}^k are independently and identically distributed uniformly within the bounds of

$$\left\{ -\left[\frac{1}{n_f} - b\left(P_{if} - \overline{P}_f\right) + A_{if} \right], \quad \left[\frac{1}{n_f} - b\left(P_{if} - \overline{P}_f\right) + A_{if} \right] \right\}.$$

This function reflects the criteria that consumers value higher evaluations of fund attributes (higher values of A_{if}) and lower MERs (low $P_{if} - \overline{P}_f$). There are other possible candidates, but this function will be used for simplicity. The empirical section will not distinguish between candidates.

The first stage of the consumer problem, allocating investment income to the different fundtypes, is solved implicitly. Implicitly maximizing expected utility with respect to t_f^k's subject to $\sum_f t_j^k = 1$ for consumer k yields a set of optimally chosen fractions of mutual fund investment income: $\{t_f^{k^*}\}$. Given this optimally chosen set, the consumer chooses one fund from each category in the second stage.

The demand for fund i in fundtype f, X_{if}, is the sum of all the efficiently allocated mutual fund investment income of the consumers that purchased fund i. There are m consumers who have identical incomes, y and t_f^k is uncorrelated with I_{if}^k.

$$X_{if} = \sum_{k=1}^{m} y t_f^{k^*} I_{if}^k. \tag{5}$$

The composite consumer acts as the typical consumer does in the aggregate. If the average proportion of mutual fund investment income allocated to fundtype f is t_f, the average expected demand for fund i in fundtype f is

$$E[X_{if}] = y t_f \sum_{k=1}^{m} E[I_{if}^k] = m y t_f \mathrm{Pr}_{if}, \tag{6}$$

where $\mathrm{Pr}_{if} = \frac{1}{m} \sum_{k=1}^{m} \mathrm{Pr}_{if}^k$ is the probability of fund i's being chosen by the composite consumer. The term $m y t_f$ is the total spending on fundtype f and is defined as S_f. The final form for average expected demand for fund i is

$$E[X_{if}] = S_f \left[\frac{1}{n_f} - b\left(P_{if} - \overline{P}_f\right) + A_{if} \right]. \tag{7}$$

3.2. Fund's problem
Each fund has a total cost function with a quadratic relationship in asset size,

$$TC_{if} = F_{if} + cX_{if} + \frac{d}{2}X_{if}^2, \tag{8}$$

where F_{if} is fund i's fixed cost, c is the linear portion of total cost, d is meant to reflect possible efficiencies of scale[10] and X_{if} is fund i's total assets. If total cost

---

10 The parameter d is not restricted to be negative. It can be positive as long as average cost is decreasing with size ($d < 2F_{if}/X_{if}^2$).

is assessed every time the fund changes assets, then the fixed cost component reflects the research and operational costs during that period. The linear portion of total cost function, cX_{if}, reflects costs that are related to fund size such as trading costs and trailer/12b-1 fees paid to brokers. The quadratic portion of total cost, $\frac{d}{2}X_{if}^2$, reflects decreases in trading costs due to block trading and other efficiencies or inefficiencies of scale.

Each fund[11] maximizes expected profits

$$E[\pi_{if}] = P_{if}E[X_{if}] - \left(F_{if} + cE[X_{if}] + \frac{d}{2}E[X_{if}^2]\right) \tag{9}$$

with respect to MER, assuming that the change in their own MER will not affect the average fundtype MER. The first-order condition rearranges to

$$P_{if} = \frac{1}{2}\left[dE[X_{if}] + b\frac{1}{n_f} + bA_{if} + \overline{P}_f + c\right]. \tag{10}$$

Fund owners do not know their profits before they enter the market, although they know their own attributes. There is free entry, and entry will occur until total industry expected profits are zero.

Suppose that the market size increases by some δ-tuple amount ($S_1 = \delta S_0$). If n_0 is the number of funds when the market size is S_0 and n_1 is for S_1 and because fixed costs do not change with market size, the fixed costs can be equated for the two different market sizes. To calculate simple comparative statics, the quadratic portion of total costs will be ignored for the time being ($d = 0$). This restriction will be relaxed below, when the equations are estimated and the comparative statics are revisited. Appendix A shows that the equilibrium relationship reduces to

$$\frac{1}{n_1} = \frac{\sigma^2}{8} + \sqrt{\frac{(\sigma^2)^2}{64} + \frac{\sigma^2}{4\delta n_0} + \frac{1}{\delta n_0^2}}. \tag{11}$$

One can see that, without a variance in fund-specific characteristics ($\sigma^2 = 0$), a doubling of market size increases the number of funds by exactly the square root of that amount (about 1.4 times). If fund heterogeneity is allowed ($\sigma^2 > 0$), the effect on the number of funds is an increase of more than the square root of the increase in the market size. This is because consumers prefer variety and also because as the number of funds increase, the average size of funds decrease ($\overline{X}_f = S_f/n_f$).

---

11 The term 'fund' instead of 'firm' has been carefully used to emphasize that this is the profit maximization of an individual fund, not the firm, which may offer many funds. The model does not explicitly take account of interaction among funds that one firm offers (economies of scope). However, the empirical section does control for funds' family size and allows some inference about the relationship between family size, MER, and the associated demand.

TABLE 3
Evidence for the monopolistic competition model

Fundtype		Ratio of market sizes $(S_1/S_0 = \delta)$	Square root of ratio of market sizes $(\sqrt{\delta})$	Ratio of number of funds (n_1/n_0)
Domestic:	Balanced	5.8	2.4	2.6
	Blend	22.6	4.8	4.0
	Value	22.8	4.8	6.4
	Growth	102.9	10.1	30.1
	Specialty	29.8	5.5	7.1
	Small Cap.	18.9	4.3	7.1
Foreign:	Global	8.5	2.9	3.5
	Emerging	9.9	3.1	3.5
All Funds		14.1	3.8	4.8

NOTE: All figures are United States to Canada ratios.

One way to determine how closely the model fits the data is to compare the actual number of funds with the predictions from the model. The results of this exercise are shown in table 3. The first column displays the eight mutually exclusive categories of equity fundtypes that both countries have in common. The second column shows the ratio of United States to Canada total value of each fundtype or the number of times larger the U.S. fundtype is. The next column displays the square root of the ratio of market sizes, and the last column shows the actual ratio of number of funds. Recall that the model outlined specifies that if there are no individual characteristics included; then, if a market increases by a certain amount, the number of funds should increase by the square root of that amount. If, however, individual characteristics are included, the number of funds should increase by more than the square root of the inflation of market size. The model predicts that the figures in the right most column (ratio of the actual number of funds) should be greater than the figures in the column second from the right (square root of the ratio of market sizes).

The table clearly shows that for all fundtypes, the ratios of number of funds are all in the predicted range. This is evidence that the framework outlined closely fits the North American mutual fund market. The number of funds in the Balanced, Blend, Global, and Emerging fundtypes are predicted very closely (actuals are within 20% of predictions). This suggests that individual attributes are not as important in these fundtypes.

4. Regression results

The monopolistic competition model has been shown to be an acceptable representation of the mutual fund industry in North America. Using its

framework, it remains to be shown to what extent the different attributes of the funds and fundtypes and markets can explain the difference in MERs. We discuss the MER pricing and expected demand equations in detail and then issues with their estimation before presenting the results.

Recall that a fund in a given fundtype sets its MER taking into consideration its average expected demand, the competition it faces within its fundtype, the fund's individual characteristics, the average MER in the fundtype and some marginal cost. The average expected demand will be proxied by the current assets of the fund. The average fundtype MER variable is separated into an average MER for each fundtype in the U.S. and Canadian mark-up component. The estimated coefficient on the Canadian dummy proves the most interesting, since it reflects the average mark-up on Canadian funds after accounting for all other variables. The estimated coefficient on the Canadian dummy measures the average Canadian mark-up due to both marginal and non-marginal cost differences, because it indicates the average difference between the Canadian and American MERs across fundtypes.

Recall from equation (7), a fund faces average expected demand that is a function of the number of funds in and size of its fundtype, its MER relative to the average MER in its fundtype and the fund's individual characteristics. The significance of the estimated attribute coefficients will indicate whether it has an effect on demand. It is assumed that there are unobservable elements of demand not explicitly defined in the expected demand equation. The second stage of the consumer problem involves the consumer's choosing one fund from all the other funds in its category. Therefore, the demand faced by an individual fund is relative to the other funds in its category and should be measured as its share of its fundtype's total assets. We use the fund's share of its fundtype market size to estimate the theoretical average expected demand equation by scaling the equation by fundtype size.

One might have already noticed an endogeneity between the MER pricing and expected demand equations. The MER pricing equation regresses MERs on assets and the expected demand equation regresses expected market share (assets over fundtype size) on MERs differenced from fundtype average MER. Although the competition effect $(1/n_f)$ is determined simultaneously with fundtype size, the short time period of this model ('snapshot') ensures that exogenous shocks, such as an increase in demand for funds not captured by the control variables, do not have time to work their way through the system to cause an increase in entry and hence a decrease in $1/n_f$.

A solution for the endogeneity is to instrument for the endogenously determined variables using the exogenous variables not included in each equation. There are three exogenous fund attributes that will be included in both the MER pricing and the expected demand equations, because they are believed to have an effect on both the cost of a mutual fund and the demand. These are the fund's age, a dummy for load, and the number of associated family members.

For the MER pricing equation, the excluded exogenous variables are those variables included in the expected demand equation: three-year annualized fund return differenced from its fundtype average return and a dummy that indicates whether a fund is eligible for a retirement plan. Both of these variables do not necessarily affect the cost of a fund but most certainly affect demand for the fund. Likewise, for the expected demand equation, the excluded exogenous variable is the exogenous variable found only in the MER pricing equation: an index dummy. Being an index fund directly affects the cost of a fund, and consumers value the lower MER of the fund more than the fact it is an index fund.

Table 4 displays the instrumental variables (IV) estimation of the theoretical MER pricing equation. The validity of the instruments is determined by observing two statistics: the first stage F-statistic (Staiger and Stock, 1997) and the correlation between the instrumental variables and the endogenous variable. The first stage F-statistic tests the hypothesis that the instruments do not enter the first stage regression. In this case, the hypothesis that the estimated coefficients on the instruments are zero is strongly rejected. Although the pairwise correlations between the two instruments and the endogenous variable appear to be small, they are both found to have strong significance. The instruments are determined to be adequate to allow inference on the estimation. The Hausman test indicates that ordinary least squares is an inconsistent estimator for this equation.

The estimation shows that a larger sized fund will have a lower MER because of economies of scale. Malhotra and McLeod (1997) find that a 100% increase in the size of the mean U.S. equity fund for 1993 would decrease its MER by 7.3 basis points. Noting that the data set used in our paper includes Canadian funds, we find that a doubling of the mean fund's assets would decrease its MER by more than 22 basis points. Malhotra and McLeod (1997) do not include the inverse of number of funds (the variable accounting for different competition in different fundtypes between countries) and, as table 5 uncovers, this accounts for part of the discrepancy. There is reason to believe that because there are some very large funds that skew the distribution of asset size, the asset variable would be more accurately defined in logarithmic form. The results are shown in appendix C and have a similar structure to those of Malhotra and McLeod's study. The inference on all variables is robust and the effect of an increase in fund size on MER is smaller. A doubling of the mean fund size would decrease its MER by 12 basis points, bringing the economies of scale effect closer to Malhotra and McLeod's findings.

The estimated coefficient for the age variable weakly indicates that an older fund will have a higher MER. This result confirms a conclusion that Tufano and Sevick (1997, 347) make. They find that older funds tend to charge higher MERs. As explained earlier, funds with loads tend to have inflated MERs and the strong positive sign of the coefficient corroborates the story. Funds with

TABLE 4
IV Estimation of the MER Pricing Equation

Theoretical equation to be estimated: $P_{if} = \frac{1}{2}\left[dE[X_{if}] + b\,\frac{1}{n_f} + bA_{if} + \overline{P}_f + c\right]$

Dependent variable: MER
Endogenous variable: $E[X_{if}]$
Instrumental variables: return − average return (*Correlation with endogenous var.*: 0.084*)
   dummy for **RRSP/IRA** (*Correlation with endogenous var.*: 0.037*)
1st & 2nd stage F-statistics: 1st: $F(9,4177) = 47.56^{*}$ 2nd: $F(8,4178) = 189.39^{*}$
Hausman test: $X^2(8) = 20.92^{*}$

Variable in model	Effect capturing	Definition	Estimated coefficient	Standard error	Significance
$E[X_{if}]$	Economies of scale	assets	−0.2253	0.0460	*
$1/n_f$	Competition effect	inverse of number of funds in fundtype	7.0093	4.7826	
A_{if}	Individual attributes	⎧ log(age)	0.0255	0.0558	
		⎨ dummy: 1 if has a load	0.4933	0.0301	*
		⎩ log(number of family members)	−0.0171	0.0153	*
		dummy: 1 if index fund	−0.3013	0.1145	*
$\overline{P}_f$	⎧ Mean U.S. MER	mean U.S. MER for fundtype	0.5420	0.0938	*
	⎨ Unexplained	dummy: 1 if Canadian fund	0.6496	0.0494	*
	⎩ Cdn mark-up				
Constant			0.7811	0.1478	*

* Indicates significance below 1%.

TABLE 5
Decomposition of MER difference

Variable in model	Definition	Estimated coefficient	Cdn. Average	U.S. average	Est. coef.* (Cdn.avg − U.S.avg)	% of difference explained
$E[X_{if}]$	assets	−0.2253	0.2305	0.7704	0.1216	14.26
$1/n_f$	inverse of number of funds in fundtype	7.0093	0.0062	0.0014	0.0341	4.00
A_{if}	log(age)	0.0255	0.6497	0.4589	0.0049	0.57
	dummy: 1 if has a load	0.4933	0.0360	−0.0069	0.0212	2.48
	log(number of family members)	−0.0171	−0.6392	0.1211	0.0130	1.52
	dummy: 1 if index fund	−0.3013	−0.0273	−0.0075	0.0060	0.70
$\overline{P}_f$	mean U.S. MER for fundtype	0.5420	1.5068	1.5020	0.0026	0.31
	dummy: 1 if Canadian fund	0.6496	1	0	0.6496	76.16
Constant		0.7811	1	1	0	0.00
			sum		0.8530	100.00

many family members appear to have lower MERs because of economies of scope. As expected, index funds have lower MERs because they require less research expense.

The regression shows that even after accounting for the smaller average fund size of the Canadian companies and difference in competition, there is still a significant unexplained average mark-up in Canadian equity MERs. The unexplained mark-up is estimated to be an average of 65 basis points or 76% of the Canadian mark-up. Table 5 decomposes the difference in the Canada and U.S. MERs using the estimated parameters of the model to determine the percentage of the difference in MERs explained by each variable.

Economies of scale contributes to the difference in MERs between the two countries. It accounts for more than 14% of the difference in MERs. Appendix C shows the same decomposition for the log-linear model. Economies of scale play a similar role in the log-linear model. The difference in competition between the two markets plays a surprisingly small role in accounting for the difference in MERs. It accounts for 4% of the difference in MERs in the linear regression and almost 10% in the log-linear estimation. In both cases, however, the unexplained Canadian mark-up is significant. Recall that to measure the amount of the mark-up due to only marginal cost differences, the random error term (c_i) would be separated into a fund-specific random component and a Canadian dummy instead of separating the mean fundtype MER into the U.S. mean fundtype MER and a Canadian dummy. A decomposition similar to table 5 shows that the 85% of the difference that is unexplained can be divided into about one-third, owing to differences in marginal costs, and two-thirds because of differences in other costs. As a check that this estimation does

not affect the overall results but is simply a method of separating out the type of unexplained cost differences, the sum of the estimated coefficients[*] (Canadian avg. – U.S. avg.) found in the column second from the right adds up the identical number found in table 5.

Table 6 displays the estimates for the expected demand equation developed in the model (for an alternate log-linear specification estimation, see appendix C). Again, the first stage *F*-statistic and the correlation between the endogenous variable and the instrument determine that the instrument is adequate to allow inference on the estimation. The two most important variables in this regression are estimated in the predicted direction. The results strongly indicate that the more funds in a fundtype (a decrease in $1/n_f$), the smaller market share each fund in that fundtype will have. The estimation predicts that the higher is a fund's MER over its fundtype's average MER, the lower will be its market share in that fundtype. The estimated significances suggest that the competition effect is stronger than the effect of an above average MER in determining market share. The equation was also estimated with the difference between a fund's MER and the weighted average MER in its fundtype instead of the above variable. The reasoning is that a fund might be more influenced by the MERs of larger funds in its fundtype than smaller funds. The results were robust with the estimated coefficient on that variable being virtually identical.

Older funds are strongly found to have larger market shares. This could be because of confidence instilled in consumers or simply because they have had a longer time to build a clientele. Load funds also tend to have a larger market share, suggesting that consumers are somewhat attracted to loaded funds, although this is not a strong result. As expected, funds with better than average returns have larger market shares. Funds that are associated with larger families are predicted to have larger market shares. This is most likely because mutual fund consumers may look to purchase funds that allow them to switch costlessly among funds in the same family. Being eligible for a registered retirement savings plan (RRSP) in Canada or individual retirement account (IRA) in the United States appears to increase market share, but not significantly. This effect is reduced by combining the Canadian and American datasets. In the United States, IRA-eligible funds tend to be distributed evenly among equity fundtypes with fund families making the decision whether to allow the eligibility; whereas in Canada, RRSP-eligibility is strictly due to foreign content of the fund and, therefore, is mostly restricted to the domestic fundtypes. Since the demand equation is fundtype specific, it is almost impossible to extract the effect on Canadian demand of RRSP-eligibility. If the demand is regressed only on the U.S. data, the IRA-eligibility variable is strongly, positively significant.

The estimated parameters can now be used to determine more precisely how well the monopolistic competition model fits the North American mutual fund

TABLE 6
IV Estimation of the Expected Demand Equation

Theoretical equation to be estimated: $\frac{E[X_{if}]}{S_f} = 1/n_f - b(P_{if} - \overline{P}_f) + A_{if}$

Dependent variable:	fundtype share
Endogenous variable:	$P_{if} - \overline{P}_f$
Instrumental variable:	dummy for index (*Correlation with endogenous var.:* −0.191*)
1st & 2nd stage F-statistics:	1st: *F*(7,4179) = 268.75* 2nd: *F*(7,4179) = 110.70*
Hausman test:	$X^2(7) = 0.14$

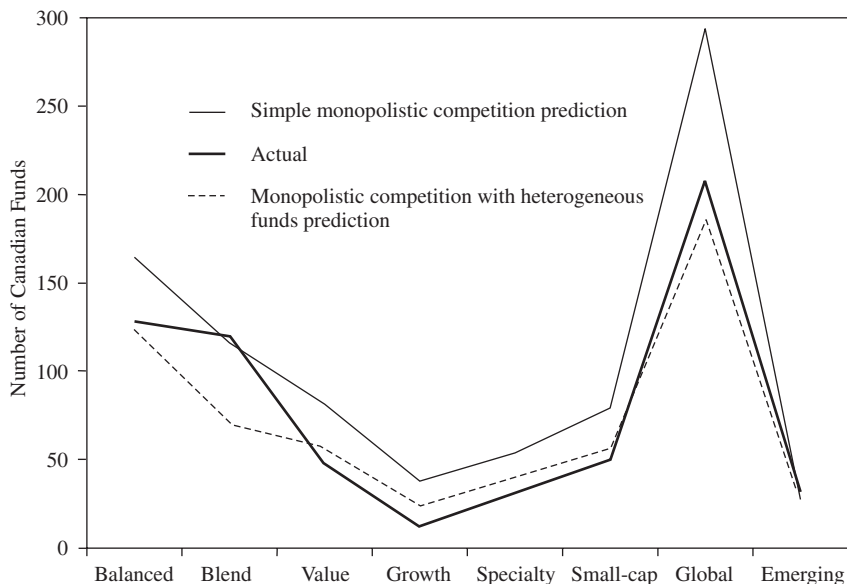| Variable in model	Effect capturing	Definition	Estimated coefficient	Standard error	Significance
$1/n_f$	Competition effect	inv. of number of funds in fundtype	1.4930	0.0639	*
$P_{if} - \overline{P}_f$	Pricing above average	MER − mean MER in fundtype	−0.1896	0.2808	*
		log(age)	0.3055	0.0736	*
		dummy: 1 if has a load	0.1069	0.1477	
A_{if}	Individual attributes	return − average return	0.0183	0.0043	*
		log(number of family members)	0.0402	0.0257	
		dummy: 1 if RRSP or IRA eligible	0.0775	0.0911	
Constant			−0.1439	0.0365	*

* Indicates significance below 1%.

FIGURE 2   Number of actual and predicted Canadian funds

market. Equation (12) is a version of equation (11) without the linear para-
meter of marginal costs, d, restricted to be zero.

$$\frac{1}{n_1^2} + \frac{\sigma^2}{(2 + bd\delta S_0)^2}\frac{1}{n_1} - \frac{2 + bdS_0}{\delta(2 + bd\delta S_0)}\left[\frac{1}{n_0^2} + \frac{\sigma^2}{(2 + bdS_0)^2 n_0}\right] = 0 \qquad (12)$$

Equation (12) is a quadratic equation in $1/n_1$ that can be solved using
the quadratic formula when it is noted that the quantities in equation (12)
are either observed or can be recovered from the estimated regressions. We will
view S_0 as the U.S. mutual fund market size, n_0 as the number of funds
in the U.S. market and $\delta = S_1/S_0$ as the amount that the Canadian market
size is smaller than that of the United States. Figure 2 compares the predicted
number of funds according to the version of monopolistic competition
developed in this paper with the actual number of Canadian mutual funds
and a simple version of monopolistic competition without heterogenous
fund attributes. The eight fundtypes are listed in the same order as they
were in table 3, and the bold line indicates the actual number of mutual
funds found in Canada. The thin solid line is the predicted number of
Canadian funds using a simple form of monopolistic competition without the
heterogenous fund addition developed in this paper (equation (7) with $A_{if} = 0$),
and the dotted line is the prediction using the model developed in this paper.

Figure 2 shows that the addition of heterogeneous funds does not adversely
change the prediction of the number of Canadian funds, except in the case of

Blend funds. Allowing for heterogeneity among funds brings the predicted number of funds very close to the actual number of Canadian funds for the Balanced, Value, Small-Cap and Global fundtypes. The sum of the squared difference between the predicted and the actual number of funds over the eight fundtypes for the simple model is 11,871 while for the model with heterogeneous funds it is 3,278. The version of monopolistic competition developed in this paper provides a closer fit to the North American mutual fund market than does a simple version of monopolistic competition with homogeneous funds.

5. Conclusion

In this paper we investigate the factors that determine MERs in North American mutual funds in an attempt to explain the mark-up in Canadian MERs. It is commonly believed that Canadian MERs are higher because Canadian funds are, on average, smaller and there are fewer rival funds. A monopolistic competition framework was used to model the mutual fund market because it directly addresses the issues of economies of scale and degree of competition. The framework was further developed to incorporate a distinctive consumer choice theory that allows funds to charge different MERs in equilibrium. The model fits the North American data well, and its estimation determines the extent to which the two common explanations account for the discrepancy in MERs. The difference in fund sizes, degree of competition, and measurable fund attributes are determined to explain about 24% of the Canadian mark-up.

There are other factors that may be influencing the MER difference. Trailer fees in Canada are anecdotally thought to be twice as large as 12b-1 fees in the United States. Trailer/12b-1 fees are a marginal cost and can account for only up to one third of the difference in MERs. A significant difference in costs, such as labour costs, would influence a difference in MERs between the two markets. We have no reason to think these costs are significantly different, but this point remains uninvestigated, because there are no available data on cost factors. Lastly, we anecdotally know that Canadian investors buy rear-loaded funds four times as often as front-loaded funds, whereas U.S. investors buy them equally as often. This remains a feasible but as of yet immeasurable reason for part of the difference in MERs.

Data appendix

Each mutual fund claims to have a specific investment objective usually outlined in the prospectus and incorporated into the name of the fund. This creates a problem if fund managers deviate from the stated objective to the point that the fund would be appropriately listed as another fundtype altogether. Morningstar and Paltrak attempt to account for this behaviour by

TABLE A1
Detailed description of fundtypes

Fundtype			Country	Tracking organization's categories included in fundtype
1	Domestic:	Balanced	U.S. Canada	Domestic Hybrid CdnSAA, CndTAA
2		Blend	U.S. Canada	Large Blend, Mid-Cap Blend Divers, LCDivr
3		Value	U.S. Canada	Large Value, Mid-Cap Value Divdnd, LCValu, Value
4		Growth	U.S. Canada	Large Growth, Mid-Cap Growth Growth, LCGrwt
5		Specialty	U.S. Canada	Convertibles, Spec.-Communication, Spec.-Financial, Spec.-Health, Spec.-Natural Res, Spec.-Precious Metal, Spec.-Real Estate, Spec.-Technology, Spec.-Unaligned, Spec.-Utilities CdnRE, Consum, Currncy, FinSer, Other, PrMetl, Resorc
6		Small-cap	U.S. Canada	Small Blend, Small Growth, Small Value LSVC, SCDivr, SCGrwt, SCValu
7	Foreign:	Global	U.S. Canada	Diversified Pacific Asia Stock, Europe Stock, Foreign Stock, Inter'l Hybrid, Pacific Asia ex-Japan Stock, World Stock China, Europe, GblSAA, GblTAA, German, GlPrMt, GlobRE, Global, India, IntMkt, IntlEq, Japan, Korean, NrthAm, PacRim
8		Emerging	U.S. Canada	Diversified Emerging Mkt, Latin America Stock Amrcas, Emerg, Latin
9	U.S.-invested Canadian Funds		Canada	GlobST, USDivers, USLgCap, USSmCap

NOTE: For precise definitions of tracking organization's categories, please see Morningstar for U.S. categories and Paltrak for Canadian categories.

allocating funds to categories based on portfolio statistics and composition over the previous three years. However, Brown and Goetzmann (1997) find that the categories used by mutual fund tracking organizations are a poor characterization of fund returns. They conduct an analysis of mutual fund categories using past returns to determine a natural grouping of funds that has some predictive power in explaining the future cross-sectional dispersion in fund returns. The dataset used in this paper is a compromise. We group the categories given in Morningstar and Paltrak using the criteria found in Brown and Goetzmann (1997). Table A1 describes in detail the composition of the fundtypes used in this paper.

Appendix A: Fund's problem in detail

Fund profit (π_{if}) is total revenue less the associated total cost:

$$E[\pi_{if}] = (P_{if} - c)\, E[X_{if}] - \frac{d}{2} E[X_{if}^2] - F_{if}. \tag{A1}$$

Using the definition of the second moment $\left(E[X_{if}^2] = \mathrm{VAR}[X_{if}] + (E[X_{if}])^2\right)$, the expected profit and $\mathrm{VAR}[X_{if}] = \left(S_f E[X_{if}] + (E[X_{if}])^2\right) M_f$ (see appendix B), where $M_f = (1/S_f^2) \sum_k \left(yt_f^k\right)^2$, this relationship can be utilized to yield the expect profit function:

$$E[\pi_{if}] = \left(P_{if} - c - \frac{d}{2} S_f M_f\right) E[X_{if}] - \frac{d}{2}(M_f + 1)\left(E[X_{if}]\right)^2 - F_{if}. \tag{A2}$$

Each fund maximizes expected profits with respect to MER (price), assuming that t3heir MER change will not affect the average fundtype MER. The substitution of the expected demand found in equation (7) into equation (A2) yields the first-order condition (rearranged):

$$P_{if} = \frac{1 + bd S_f(M_f + 1)}{b(2 + bd S_f(M_f + 1))} \left[\frac{1}{n_f} + b\overline{P}_f + A_{if}\right] + \frac{c + \frac{d}{2} S_f M_f}{2 + bd S_f(M_f + 1)}. \tag{A3}$$

The definition of average expected demand derived in equation (7) can be substituted into the terms inside the square brackets ofequation (A3). Then, by rearranging terms so MER is only found on the left-hand side of the equation, equation (A3) can be manipulated to lead to the following simple relationship between MER and average expected demand:

$$P_{if} = \frac{1}{b S_f} [X_{if}] + d(M_f + 1)\, E[X_{if}] + c + \frac{d}{2} S_f M_f. \tag{A4}$$

Another simple substitution of equation (7) in for the first average expected demand in equation (A4) leads to the following form of the MER pricing decision:

$$P_{if} = \frac{1}{2}\left[d(M_f + 1)\, E[X_{if}] + b\frac{1}{n_f} + bA_{if} + \overline{P}_f + c + \frac{d}{2} S_f M_f\right]. \tag{A5}$$

Equation (10) in the text differs from equation (A5) by the assumption that $M_f = 0$. The term M_f is the summation over all consumers of the square of each consumer's share of fundtype assets and is expected to be an extremely small number. In 1999 the IFIC claims that their members have \$CDN 389.7 billion assets and 45.8 million unitholder accounts. On average, each Canadian unitholder's share is 1.923×10^{-8} of total assets. For the same year, the ICI claims 226.8 million unitholder accounts for \$US 6,846 billion assets.

On average, U.S. unitholder shares are 4.409 10^{-9} of total assets. This is a simplifying assumption but not a necessary one. Equation (A5) could be used for the basis of the regressions in table 4. The term will be retained in the derivations in the rest of the appendix to display the theory in its unrestricted form.

Fund owners do not know their profits before they enter the market, although they know their own attributes. There is free entry, and entry will occur until total industry expected profits are zero:

$$E[\pi_f] = \sum_{i=1}^{n_f} E[\pi_{if}] = 0. \tag{A6}$$

The MER pricing equation found in equation (A4) can be substituted into the expected profits function in equation (A2) to yield the following equilibrium condition:

$$\sum_{i=1}^{n_f} E[\pi_{if}] = \left(\frac{1}{bS_f} + \frac{d}{2}(M_f + 1)\right) \sum_{i=1}^{n_f} (E[X_{if}])^2 - \sum_{i=1}^{n_f} F_{if} = 0. \tag{A7}$$

Average expected demand can be characterized as a function of only fund attributes by noticing that the MERs of any two funds i and j in the same fundtype are related to each other by the differences in their individual characteristics (use equation (A3)) and that if one fund is fixed, the average MER can be defined as a function of the MER of only that fund and the difference between its characteristics and the average characteristic in that fundtype ($\bar{A}_f$). We already know that the average characteristic in a fundtype is zero ($\bar{A}_f = E[A_{if}] = 0$), and this leads to the following form of the average expected demand function.

$$E[X_{if}] = S_f \left[\frac{1}{n_f} + \frac{1}{2 + bdS_f(M_f + 1)} A_{if}\right]. \tag{A8}$$

The equilibrium relationship becomes

$$\sum_{i=1}^{n_f} E[\pi_{if}] = \left(\frac{1}{bS_f} + \frac{d}{2}(M_f + 1)\right) S_f^2 \left[\frac{1}{n_f} + \frac{\sigma^2}{(2 + bdS_f(M_f + 1))^2}\right] - n_f F_f = 0, \tag{A9}$$

where σ^2 is the variance of fund attributes ($\text{VAR}[A_{if}] = \sigma^2$) and the average fixed cost for a fundtype f is F_f, regardless of the number of funds.

If n_0 is the number of funds when the market size is S_0 and n_1 is for S_1, the corresponding relationship is

$$F_f = \left(\frac{1}{bS_0} + \frac{d}{2}(M_0 + 1)\right) S_0^2 \left(\frac{1}{n_0^2} + \frac{\sigma^2}{(2 + bdS_0)^2 n_0}\right)$$

$$F_f = \left(\frac{1}{bS_1} + \frac{d}{2}(M_1 + 1)\right) S_1^2 \left(\frac{1}{n_1^2} + \frac{\sigma^2}{(2 + bdS_1)^2 n_1}\right).$$

(A10)

These two relationships are equated in equation (12) in the text and also in equation (11) with the linear portion of marginal cost, d, restricted to be zero. Both equations in the text assume that a fundtype increase occurs through an equal increase of income among all consumers:

$$\left(M_1 = \sum_k \left(\frac{yt^k}{S_1}\right)^2 = \sum_k \left(\frac{\delta yt^k}{\delta S_0}\right)^2 = \sum_k \left(\frac{yt^k}{S_0}\right)^2 = M_0\right)$$

and that $M_0 = M_1 = 0$.

Appendix B: Deriving the variance of demand

Recall that demand (equation 8) is defined as

$$X_{if} = \sum_{k=1}^{m} yt_f^{k*} I_{if}^k \tag{B1}$$

and that its expected value is

$$E[X_{if}] = S_f \mathrm{Pr}_{if}. \tag{B2}$$

The demand function is a summation over the random variable I_{if}^k. This leads to the following definition for the variance of X_{if}:

$$\mathrm{VAR}[X_{if}] = \sum_{k=1}^{m} \left(yt^k\right)^2 \mathrm{VAR}[I_{if}^k]. \tag{B3}$$

The definition of variance for a discrete random variable, X, is

$$\mathrm{VAR}[X] = \sum_i \left(x_i - E[X_i]\right)^2 f(x_i), \tag{B4}$$

where x_i are the values of X and $f(x_i)$ is the probability density function of X. Using this definition, we can define the variance for the indicator function as follows:

$$\mathrm{VAR}[I_{if}^k] = (0 - \mathrm{Pr}_{if})^2 (1 - \mathrm{Pr}_{if}) + (1 - \mathrm{Pr}_{if})^2 \mathrm{Pr}_{if} = \mathrm{Pr}_{if} - \mathrm{Pr}_{if}^2. \tag{B5}$$

This can be substituted back into the variance of demand function to yield:

$$\text{VAR}\,[X_{if}] = \left(\text{Pr}_{if} + \text{Pr}_{if}^2\right) \sum_{k=1}^{m} \left(yt^k\right)^2. \tag{B6}$$

A simple of substitution of the definition for expected demand leads to the following relationship:

$$\text{VAR}\,[X_{if}] = \left(S_f E[X_{if}] + \left(E[X_{if}]\right)^2\right) M_f,$$

where

$$M_f = \sum_k \left(\frac{yt_f^k}{S_f}\right)^2. \tag{B7}$$

Appendix C: Estimation of the model in log-linear functional form

TABLE C1
Instrumental variables estimation of the log-linear pricing equation

Dependent variable:	MER
Endogenous variable:	$E[X_{if}]$
Instrumental variables:	return − average return (*Correlation with endogenous var.*: 0.218*)
	dummy for RRSP/IRA (*Correlation with endogenous var.*: 0.038*)
1st & 2nd stage F-statistics:	1st: $F(9,4177) = 246.71^*$ 2nd: $F(8,4178) = 430.94^*$
Hausman test:	$X^2(8) = 46.19^*$

Variable in model	Effect capturing	Definition	Estimated coefficient	Standard error	Significance
$E[X_{if}]$	Economies of scale	log(assets)	−0.1746	0.0154	*
$1/n_f$	Competition effect	log(inv. of number of funds in fundtype)	0.0628	0.0186	*
A_{if}	Individual attributes	log(age)	0.0014	0.0241	
		dummy: 1 if has a load	0.4263	0.0216	*
		log(number of family members)	−0.0063	0.0091	
		dummy: 1 if index fund	−0.4884	0.0631	*
$\overline{P}_f$	Mean U.S. MER	mean U.S. MER for fundtype	0.4601	0.0575	*
	Unexplained Cdn mark-up	dummy: 1 if Canadian fund	0.6271	0.0354	*
Constant			0.7619	0.1499	*

* Indicates significance below 1%.

TABLE C2
Decomposition of MER difference using log-linear pricing equation

Variable in model	Definition	Estimated coefficient	Cdn average	U.S. average	Est. coef.* (Cdn avg – U.S. avg)	% of difference explained
$E[X_{if}]$	log(assets)	−0.1746	−2.9995	−2.3942	0.1056	12.38
$1/n_f$	log(inv. of number of funds in fundtype)	0.0628	−5.3428	−6.6965	0.0850	9.96
A_{if}	log(age)	0.0014	0.6497	0.4589	0.0003	0.03
	dummy: 1 if has a load	0.4263	0.0360	−0.0069	0.0183	2.14
	log(number of family members)	−0.0063	−0.6392	0.1211	0.0048	0.56
	dummy: 1 if index fund	−0.4884	−0.0273	−0.0075	0.0097	1.14
$\overline{P}_f$	mean U.S. MER for fundtype	0.4601	1.5068	1.5020	0.0022	0.26
	dummy: 1 if Canadian fund	0.6271	1	0	0.6271	73.51
Constant		0.7619	1	1	0	0.00
				sum	0.8530	100.00

NOTE: Country averages are only for funds used in regression.

TABLE C3
Instrumental variables estimation of the log-linear demand equation

Dependent variable: logged fundtype share
Endogenous variable: $P_{if} - \overline{P}_f$
Instrumental variable: dummy for index (*Correlation with endogenous var.*: −0.191*)
1st & 2nd stage F-statistics: 1st: $F(7,4179) = 269.00^*$ 2nd: $F(7,4179) = 488.09^*$
Hausman test: $X^2(7) = 0.03$

Variable in model	Effect capturing	Definition	Estimated coefficient	Standard error	Significance
$1/n_f$	Competition effect	log(inv. of number of funds in fundtype)	1.4654	0.0391	*
$P_{if} - \overline{P}_f$	Pricing above average	MER – mean MER in fundtype	−0.7816	0.3535	**
A_{if}	Individual attributes	log(age)	1.1732	0.0932	*
		dummy: 1 if has a load	−0.3224	0.1847	
		return – average return	0.0536	0.0054	*
		log(number of family members)	0.3144	0.0317	*
		dummy: 1 if RRSP or IRA eligible	0.3403	0.1155	*
Constant			−1.1969	0.0891	*

* Indicates significance below 1%; ** below 5%.

References

Borch, K. (1968) *The Economics of Uncertainty* (Princeton, NJ: Princeton University Press)

Berkowitz, M., and Y. Kotowitz (1993) 'Incentives and efficiency in the market for management services: a study of Canadian mutual funds,' *Canadian Journal of Economics* 26, 850–66

Brown, K., W. Harlow, and L. Starks (1996) 'Of tournaments and temptations: an analysis of managerial incentives in the mutual fund industry,' *Journal of Finance* 51, 85–110

Brown, S., and W. Goetzmann (1997) 'Mutual fund styles,' *Journal of Financial Economics* 43, 373–99

Carhart, M. (1997) 'On persistence in mutual fund performance' in *Journal of Finance* 52, 57–82

Dellva, W., and G. Olson (1998) 'The relationship between mutual fund fees and expenses and their effects on performance,' *Financial Review* 33, 85–103

Financial Post (1998a) 'Trailer-free rebate could lead to lower fund MERs,' 28 March
— (1998b) 'What's in an MER? More than just management,' 11 April

Gadsden, S. (1998) *The Canadian Mutual Funds Handbook* (Toronto: McGraw-Hill Ryerson)

Grinblatt, M., and S. Titman (1994) 'A study of monthly mutual fund returns and performance evaluation techniques,' *Journal of Financial and Quantitative Analysis* 29, 419–44

Hensher, D., and L. Johnson (1981) *Applied Discrete-Choice Modeling* (New York: John Wiley)

Hey, J. (1979) *Uncertainty in Microeconomics* (New York: New York Press)

Investment Company Institute (1991) Mutual Fund Fact Book (New York)

Krugman, P. and M. Obstfeld (1997) *International Economics: Theory and Policy* (Massachusetts: Addison-Wesley)

Malhotra, D., and R. McLeod (1997) 'An empirical analysis of mutual fund expenses,' *Journal of Financial Research* 20, 175–90

Markowitz, H. (1970) *Portfolio Selection: Efficient Diversification of Investments* (Yale University Press)

Quirin, G. (1969) *A Study of the Canadian Mutual Funds Industry* (Canadian Mutual Funds Association)

Staiger, D. and J. Stock (1997) 'Instrumental variables regression with weak instruments,' *Econometrica* 65, 557–86

Stromberg, G. (1998) *Investment Funds in Canada and Consumer Protection, Strategies for the Millennium* (Ottawa: Industry Canada)

Trzcinka, C., and R. Zweig (1990) *An Economic Analysis of the Cost and Benefits of S.E.C. Rule 12b-1* (New York: Salomon Brothers Center for the Study of Financial Institutions)

Tufano, P., and M. Sevick (1997) 'Board structure and fee-setting in the U.S. mutual fund industry,' *Journal of Financial Economics* 46, 321–55.

Wooldridge, J. (2000) *Introductory Econometrics: A Modern Approach* (South-Western College Publishing)