Performance-based Labeling

Robert L. Hicks

1 Introduction

Many see product labeling as a way for information to impact markets for goods that have a negative social or environmental impact. So long as consumers value not only the good itself but also how the good was produced, it is argued, then a labeling scheme that gives consumers information on the production processes and methods (PPM) will fundamentally alter the market towards greener or socially responsible methods of production. Consumers who value these attributes will be willing to pay more for labeled products. This price premium will provide incentives for producers to choose PPM to mitigate environmental or social problems.[1] Invoking this line of argument, Bass, Markopoulos, and Grah (2000) contend that labeling is "at the heart of many of today's greatest economic, social, environmental, and political challenges, which involve getting the tradeoffs right for sustainable development".

Theoretical studies have attempted to evaluate these claims by investigating the conditions under which eco and social labeling programs can, in fact, "get the tradeoffs right" by allowing consumers to differentiate products according to its associated environmental or social impact. For example, Sedjo and Swaddle (2002) and Basu et al. (2004) investigate the viability of labels and standards in a general equilibrium context for eco-labeled forest products and socio-labels guaranteeing a product was produced without the use of child labor, respectively. In both of these models, equilibrium is based on the price premium an eco or socially conscious consumer is willing to pay to attain a labeled product and the relative costs to the producer of meeting PPM standards.[2] A higher willingness to pay on the part of consumers is seen as a reward by producers in the south and

[1] Example of eco-labeling programs includes the dolphin-safe label in the U.S. canned tuna market, the Nordic Swan, and the Blue Angel in Germany. Social labels include the RUGMARK child-labor free rug label began in Germany and in use in the United States, and the FLO and Transfair fair trade label for coffee and other fair trade products.

[2] Basu, Chau, and Grote (2004) model the actions of producers in the north and south, consumers in the north, and the household labor decisions including the

will tend to shift producers toward the eco or socially preferred method of production.[3]

Given the importance of the northern consumer's willingness to pay for the success of labeling programs, empirical studies of the demand for labeled products have shown the existence of premiums for numerous products ranging from canned tuna to organic textiles (Teisl et al. 2002; Nimon and Beghin 1999; Bjorner et al. 2004). However, a closer look at many labeling programs (and producers' opinions about labeling) shows that green PPMs have not been widely adopted and remain a small market segment for most products (Auld et al. 2001; Baharuddin and Simul 1994; Irland and Waffle 2002). This is occurring even while consumers' stated support for eco-labeled products is on the rise. Because of the theoretical importance of the existence of a price premium and the mixed results in the empirical literature concerning the size of the price premium, I investigate consumer preferences for an expanded range of attributes associated with labeled products and show that one explanation for the relatively low willingness to pay for labeled products can be attributed to consumers' lack of information about the performance of labeling programs.

1.1 Performance Labeling

Given the modest price premia found in many studies in the empirical literature, I investigate a labeling approach that goes beyond the traditional labeling paradigm of informing consumers about a good's PPM. Consu-

decision to employ child labor. Their model contains numerous testable hypotheses concerning credibility, price premia, relative production costs in the north and south, as well as the role of trade policy for influencing child labor policy. To fully appreciate the impact of performance labeling as presented in this paper, such a general equilibrium approach should be undertaken.

[3] Basu et al. (2004) also discuss label credibility and the monitoring and enforcement of production standards as important determinants of the overall shift in production and associated welfare impacts in the south. A theoretical model by Brown (2001) shows that most of the premia associated with child-free product labels will be captured by the producers in the south and not adult laborers, making households worse off and that labeling credibility will suffer because of false labeling. She concludes "children are found to benefit only if consumers pay an additional amount that can be contributed to a child welfare fund" or bids adult wages in the south to a sufficient level to allow southern households to avoid child labor. Basu (1999) offers a summary of the child labor issue and discusses household production models coupled with a production sector for explaining the child labor decision.

mers may derive value from knowing that the production related to their purchased product met the PPM standard. However, it is also likely that they might be interested in the overall performance of the labeling program as to how successful it has been in meeting the overall goal set forth by the certifying agency. How the consumer may value the performance of a labeled product is ambiguous. On the one hand, consumers may not be willing to pay for a labeled product if the program is making no appreciable difference to the overall problem; while, on the other hand, a well-performing program may be able to capture higher consumer willingness to pay.[4] The performance of a labeling program is collectively defined (e.g. the overall impact of a child-free label) and depends on how the mar-ket for labeled versus non-labeled products work. Compared to a tradi-tional label, in which consumers have no information on a labeling program's overall performance, the performance label offers the consu-mers more information and, perhaps, will increase the price premium asso-ciated with labeled products.

Consider an example from social labeling, namely fair trade labeling of coffee. The goal of the fair trade program is to pay growers an adequate price per pound in order to guarantee the livelihoods of coffee growers in the developing south. However, when purchasing fair trade coffee, the consumer in the north only knows that the product bears a label guaranteeing a grower a minimum price for their coffee plus a predetermined social price premium.[5] Setting aside the important issue of label credibility, the current fair trade label informs the consumer about how her *one-time* purchase of coffee impacted growers. The consumer derives a private benefit from the personal satisfaction of knowing that her purchase ensured fair wages to farmers. However, the purchase guarantees nothing about meeting the objectives of the labeling program. Important issues like program sustainability, the economic benefits to farmers, and how many farmers participate in this program are not conveyed under traditional labeling programs. To know the full impact of a labeling program, the buyer must also know if the labeling program is meeting the overall goal of the program, a public good determined by collective choice. An individual's one-time purchase of fair trade coffee provides benefits back to growers, and likely helps in the support of a larger goal related to the labeling program. However, the public benefits - what a purchaser believes

[4] It is also possible that consumers armed with more information on the amount of a public good collectively provided by the labeling program may freeride on the purchases of others.

[5] The label guarantees the FLO minimum price of $1.21 per pound and pay a social premium of $.05 per pound (Murray et al. 2003 p. 6).

she is contributing to a public good like livelihoods of farmers in the south - are simply unknown to consumers under the current labeling regime making product differentiation across performance attributes impossible.

1.2 Fair Trade Coffee

In this paper, I investigate the impact of including performance attributes on consumer willingness to pay for fair trade coffee.[6,7] I do this because (1) coffee is the fair trade product with the longest history and largest sales volume (James 2000), (2) consumers are used to seeing and evaluating fair trade coffee in the marketplace, and (3) performance metrics are readily identifiable and already measured by fair trade organizations such as Transfair USA.

Following the collapse of the International Coffee Agreement in 1989, real coffee prices fell precipitously to their lowest level in nearly a century while additional countries began producing coffee (e.g. Vietnam). During this time, producers' share of coffee revenues dropped by thirty-five percent as coffee supply increased. In response, the fair trade movement began a labeling campaign aimed at informing consumers that growers received a "fair price" for their product (Transfair USA) and programs were instituted to "facilitate a wider distribution of benefits to small growers" (Taylor 2004). Consumers in the United States and Europe routinely make choices over coffee products that are fair-trade labeled and not. Fair trade coffee in 2003, accounted for only 1% of the world coffee market, yet represented over one-half million growers in the developing

[6] Consider performance labeling in an eco-labeling context. The tuna-dolphin eco-label exhibits significant private and public benefits. A consumer purchasing the eco-labeled product is assured that her purchase of tuna in no way involved the intentional encirclement, capture, or harm to dolphins in the Eastern Tropical Pacific Ocean. That is, consumers benefit from knowing that whatever the status of dolphin stocks in the ocean, their purchasing behavior did not have direct negative impacts on the stocks. It is also possible that over and above these private benefits, consumers may value dolphin stocks directly. That is, their willingness to pay (WTP) for labeled products might vary significantly as a function of dolphin stocks levels. Purchasing the eco-labeled tuna product pro-vides a public good to society (through the preservation of dolphin stocks) even if others in society do not buy dolphin safe tuna. Collectively, consumer's buy-ing the eco-labeled product determine some level of environmental quality that benefits everyone in society.

[7] For an excellent summary of fair trade coffee and Forest Stewardship Council (FSC) certified timber, see Taylor (2004).

south. In the United States, the fair trade market currently accounts for over 4% of the specialty coffee market and nearly 2% overall (Transfair USA 2005). Finally, fair trade coffee certifying agencies routinely collect performance indicators on the overall achievements of their coffee labeling programs. For example, Transfair reports "Coffee Producer Performance" as the "Additional Farmer Income Generated by Fair Trade in the U.S." and shows that additional revenues have climbed to over twenty five million dollars in 2004. Given farmer participation levels in fair trade programs, rough calculations reveal that farmers can expect to receive no less than almost $70 per year in additional revenues from participating in the program.[8]

Additionally, consumers may also want information about the performance of a labeling program as a further check on label credibility (beyond that of the certifying agency). For example, a recent Wall Street Journal article (Stecklow and White 2004) revealed that only a small portion of the fair trade markup is actually going to coffee growers. Consumers may be quite worried about label veracity - can they believe that the social or environmental claims being made on the label are being delivered? A label that not only informs about the PPM of the product but also relates the performance of the label may in some ways alleviate consumer concerns about whether their price premium is being used to increase producer compensation. For the case of fair trade coffee, a performance metric specifying the increased revenues accruing to program participants would allow consumers to differentiate coffees described in the aforementioned article, where a large portion of the price premium paid by consumers are being captured by the supply chain.

Given that certifying agencies collect and report program performance data (e.g., increases in revenues going to growers, the number of growers enrolled in local fair trade cooperatives), and that such a performance based-labeling initiative could be instituted, several empirical questions need to be addressed to assess the impact of performance-based labeling on the price premium, including (1) does reporting performance as part of the fair trade label always lead to higher consumer willingness to pay for labeled coffee as compared to traditionally labeled coffees, (2) when evaluating a traditional fair trade label, do consumers have an *a priori* belief about program performance, and (3) are performance-based labels always preferred to traditional fair trade labels. Using a stated preference choice experiment, I tackle each of these questions.

---

[8] It is likely that this figure is a substantial underestimate of revenue increase per farmer, since the number of participating farmers is reported over a five year period, rather than yearly.

In the following section I outline the contingent valuation literature on valuing public goods because similar issues of information provision and the importance of defining the good being valued are central to the contingent valuation methodology. Additionally, I offer a brief introduction to stated preference techniques for measuring consumer preferences for product attributes. Section three outlines the choice experiment including data collection and experimental design. The fourth section details the econometric approach for testing a number of hypotheses concerning fair trade labels, including the importance of certifying agency, and price premia for performance-based fair trade labeling. I conclude with a brief summary of findings and the potential policy implications of performance-based labeling.

2 Literature

A well-known finding in the contingent valuation literature on valuing public goods is the importance of the amount and quality of information for consumer willingness to pay (WTP) for public goods (Mitchell and Carson 1989; Hoehn and Randall 2002). The consumer wants information about the provision rule and the level of public good being purchased in the political market.[9] In the market for labeled goods, information matters in many of the same ways. The consumer buying the product wants to know if it meets the PPM standard, and how the overall market level is impacting the public good. A label that only informs as to the PPM of the product will likely be perceived by consumers to be a very different product to one that meets both the PPM requirement *and* informs the consumer as to the performance of public goods provision.

While more information does inform consumer choice, in a real market place, consumers do not have significant time to devote to studying product labels. In the contingent valuation context, where a large public project is often described, it may be reasonable to assume that voters in a political market would be willing to spend significant amounts of time studying pro-ject information. In a market setting, it is not likely that the average consumer will devote the same amount of time for studying label content.

[9] The payment vehicle is also important for contingent valuation experiments. In the eco-labeling setting, the payment vehicle is less important since the consumer buys the green attributes of products through market transactions.

In this study, I employ stated preference techniques to assess how consumers value the public and private components associated with social labeling programs (the technique is termed Stated Preference Discrete Choice (SPDC)). The technique is summarized in Louviere et al. (2000), and has been applied in numerous studies of recreational demand (Hicks 2002); Deshazo and Fermo 2002; Adamowicz et al. 1994) and eco-labeled products (Gudmundssen and Wessels 2000; O'Brien and Teisl 2004).[10] Like contingent valuation, SPDC techniques applied to eco-labeling yield information about preferences by analyzing choices over hypothetical labeled products. Further, SPDC considers a product as a bundle of attributes. Using experimental design techniques, respondents are given product comparisons that are optimal in the sense that they require the respondent to make tradeoffs across the different product characteristics attributes simultaneously.

Additionally, new policy-relevant attributes can be examined; for example, respondents are asked to consider a product under the existing labeing program and one with performance based labeling. Like contingent valuation, SPDC is based upon hypothetical rather than real behavior. There is a growing body of literature comparing revealed and stated preference methods showing that for many cases, parameter esti-ates across revealed and stated preference data are statistically equivalent (Swait et al. 1994, Adamowicz et al. 1994). These tests are seen as validity checks for the SPDC method so that policy guidance resulting from the SPDC model will be relevant for real-world application.

3 Data and Experimental Design

To investigate how information impacts WTP for eco-labeled products, I conduct a split sample experiment. The first sample of respondents are asked to evaluate products based upon the traditional labeling programs informing of production practices and methods only, while the second sample evaluate labeled products with additional information on label performance. In both samples, the PPM of the products and all other aspects of the survey design are identical. Consequently, I am able to isolate the impact of performance-based labeling on the WTP for labeled products.

[10] The seafood labeling study of Gudmundssen and Wessels (2000) examines a specific form of performance-based labeling. In their product choice experiment, consumers evaluate products that are either sustainable or not.

I investigate WTP for fair trade labeled coffee because it exhibits significant public goods properties. Buyers of fair trade products are purchasing something like a welfare assistance program in a foreign country for a select group of program participants. The performance of the public good provision due to the label is something that can be enjoyed by everyone in society whether the individual purchases the labeled good or not. Early on in the survey, respondents read the following statement on fair trade products[11]:

Advocates argue that Fair Trade certified products ensure that farmers, workers, and artisans are paid a fair price for their products or labor, don't use child labor or forced labor, have healthy and safe working conditions, use sustainable and environmentally friendly production methods, and have long-term and direct relationships with buyers. Others feel that fair trade is discriminatory against growers and countries that don't have the resources to institute a Fair Trade program.

This statement was purposely worded to convey to respondents, that there are potential up and downsides related to fair trade programs. Participating growers potentially benefit from participation. However, respondents were also informed that for non-participants there might be potential downsides from a fair trade program. I include both perspectives on fair trade because of the need for a balanced survey instrument that give respondents a concise description of the many facets of fair trade, and to lay the groundwork for the performance metrics introduced later in the survey. These metrics are intentionally designed to focus the respondents on the performance of the labeling program for program participants only. Further, respondents are asked several questions about their knowledge of and purchasing habits for fair trade products.

Using a blocked experimental design, I construct two SPDC experiments. Table 1 lists the attributes and levels for each experiment. Note that aside from the two performance metrics, the levels and attributes of the two experiments are identical[12]. Before responding to the choice comparisons shown in Figures 1 and 2 (Appendix), respondents read:

In this section we would like for you to imagine that you are in your favorite campus coffee shop and are looking to purchase a cup of coffee. There are three different brands available for you to purchase. We will ask you to repeat the brand choice several times. Please assume that the brand attributes are identical except for price and any information given on the

[11] The survey is available from the author.

[12] Kenya was dropped from the performance labeling experiment because country of origin effects were found to be small in the traditional labeling experiment and dropping one country of origin increased the design efficiency.

labels. For example, please assume that product quality is the same across the three different brands. If the fair trade or organic label is blank, then there is no information regarding whether that product meets standards or not.

Table 1. Experimental designs

Variable	Traditional label	Performance-based label
Price (Labeled)	{$2.25,$2.50,$2.75,$3.25}	{$2.25,$2.50,$2.75,$3.25}
Price (Non-labeled)	{$1.50,$1.75,$2.00}	{$1.50,$1.75,$2.00}
Country of Origin	{Brazil, Costa Rica, Kenya, Colombia}	{Brazil, Costa Rica, Colombia}
Organic (for non-labeled coffee)	{Yes, No}	{Yes, No}
Certifying Agency	{USDA, Consumer's Union, Coffee Grower's Association}	{USDA, Consumer's Union, Coffee Grower's Association}
Increased Revenue	No Information	{10%, 25%, 50%}
Increased Participation	No Information	{3%, 20%, 40%}

Blocked experimental design techniques were used to select the fifteen sets of 5 questions that maximize respondent tradeoffs across coffees. Although the levels and attributes of the two experiments are identical (except for performance information), the actual levels of the attributes chosen by the experimental design algorithm differ by question, block, and experiment. For each of the two experiments, respondents are randomly assigned to one of the fifteen blocks.

Respondents consisted of students taking large introductory classes (in Economics and Environmental Studies classes) at the College of William and Mary during the fall of 2005. For each treatment, respondents were evenly divided across the economics and environmental studies classes. The survey was filled out during class time and was handed out at the beginning of class. The performance-based experimental design includes two additional attributes. Because the focus of the study was to investigate performance attributes I allocated twice as many respondents to the performance-based design.

4 Model

Consider a consumer faced with a choice over several products. Some of the products are labeled and some are not. The choice problem for the consumer is to choose the best product given preferences and available alternatives. First, consider the choice problem presented in Figure 1, where no performance-based information is given. Let the consumer's indirect utility function for option i be written as

$$V(P_i,\mathbf{C}_i,\mathbf{A}_i,O_i,\varepsilon_i) = \alpha P_i + \mathbf{C}_i'\beta + \mathbf{A}_i'\delta + \omega O_i + \varepsilon_i \quad (1)$$

where

P_i = Price of Coffee i

$\mathbf{C}_i$ = 1 x k vector of dummy variables indicating country of origin

$\mathbf{A}_i$ = 1 x j vector of dummy variables indicating certifying agency if labeled

O_i = dummy variable = 1 if organic but not fair trade labeled

ε_i = unobserved (by the researcher) error term associated with alternative i

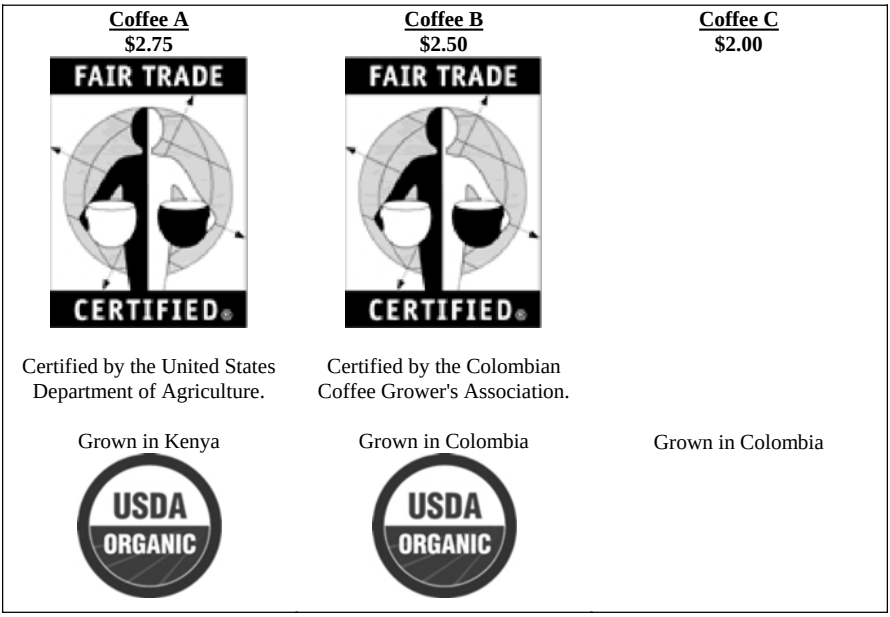

Fig. 1. Traditional label experiment

Notice, that since the label informs about the PPM of the product, the consumer derives some benefit from purchasing the green product (so long as any $\delta>0$). The consumer's choice problem is to choose the product i that maximizes their utility over the choice occasion

$$v_i = \max\left[V(P_s, \mathbf{C}_s, \mathbf{A}_s, O_s, \varepsilon_s) \forall s \in S\right] \tag{2}$$

Assuming that the error terms are distributed as GEV I, then the probability of observing the choice of product i can be written as

$$\text{Prob}_i\left(P, C, \mathbf{A}, O; (\alpha, \beta, \delta, \omega)\right) = \frac{e^{\alpha P_i + C_i'\beta + \mathbf{A}_i'\delta + \omega O_i}}{\sum_s e^{\alpha P_s + C_s'\beta + \mathbf{A}_s'\delta + \omega O_s}} \tag{3}$$

Now consider a consumer that faces the choice problem of Figure 2. The consumer is informed of the performance of the labeled product beyond the description of the PPM. Using the performance data, the consumer can gauge how the labeled product is impacting some larger public good through the collective actions of participants in the market. Rewrite the consumer's indirect utility function as

$$V(P_i,\mathbf{C}_i,\mathbf{A}_i,L_i,O_i,\mathbf{G}_i,\varepsilon_i) = \alpha P_i + \mathbf{C}_i'\beta + \mathbf{A}_i'\delta + \theta L_i + \mathbf{G}_i'\phi + \omega O_i + \varepsilon_i \quad (4)$$

where the definitions of equation (1) are still operative and $\mathbf{G}_i$ is a vector of performance metrics associated with the public good provided by the labeling program. Consumers will choose the optimal product as in equation (2) and from the researcher's perspective, the probability of choosing product i can be written as

$$\text{Prob}_i\left(P,C,\mathbf{A},\mathbf{G},O;(\alpha,\beta,\delta,\omega)\right) = \frac{e^{\alpha P_i + C_i'\beta + \mathbf{A}_i'\delta + \mathbf{G}_i'\phi + \omega O_i}}{\sum_s e^{\alpha P_s + C_s'\beta + \mathbf{A}_s'\delta + \mathbf{G}_k'\phi + \omega O_s}} \quad (5)$$

Comparing equations (3) and (5) reveal the similarities of the choice problem faced by individuals. In both cases, they gain some benefit associated with consuming a good that has been produced with a certified PPM. However, as equation (4) makes clear, consumers are also hypothesized to value the performance of the labeling program with the addition of the term $G_i'\phi$.

The vector δ is capturing several effects. First, it is capturing the effect of certifier credibility and label veracity. Products with more well known and trusted certifiers will likely be preferred to those having either no certification or fly-by-night certification, *ceteris paribus.* Additionally, the consumer may attach the private benefits from purchasing a labeled product and knowing that their purchase had positive impacts on the related public good. If δ is indeed capturing only these effects, then the estimate across the two experiments should be roughly equal given a sufficient sample size. However, it may also be the case that consumers attach priors about a labeling program's performance to the certification agency parameter when performance data is absent. If this is indeed happening, then it is likely that the parameter on certifying agency will play a much larger role under the traditional label than for performance labeled products.

Econometrically, these competing hypotheses can be tested by jointly estimating both models and restricting parameters to be equal across common elements of the choice problem-$\{\alpha,\beta,\omega\}$. This approach assumes, for common data elements, that respondents evaluate information (and make economic tradeoffs) in the same way across the two choice problems. Most importantly, when parameters are restricted across models, the performance-based model simplifies the traditional model when the performance of the labeled product is zero. An alternative estimation strategy freely estimates each set of parameters, and implicitly allows respondents to react differently to labels and information when choosing products.

Since the restricted model is nested within the model where both sets of parameters are freely estimated, we can test whether consumers do in fact value certification veracity and private benefits the same (δ vectors are equal across choice experiments), if they have priors over program performance (δ vectors are not equal) or if they base their purchasing decisions solely on the certification and private benefits associated with the label (ϕ vector is not significant).[13,14]

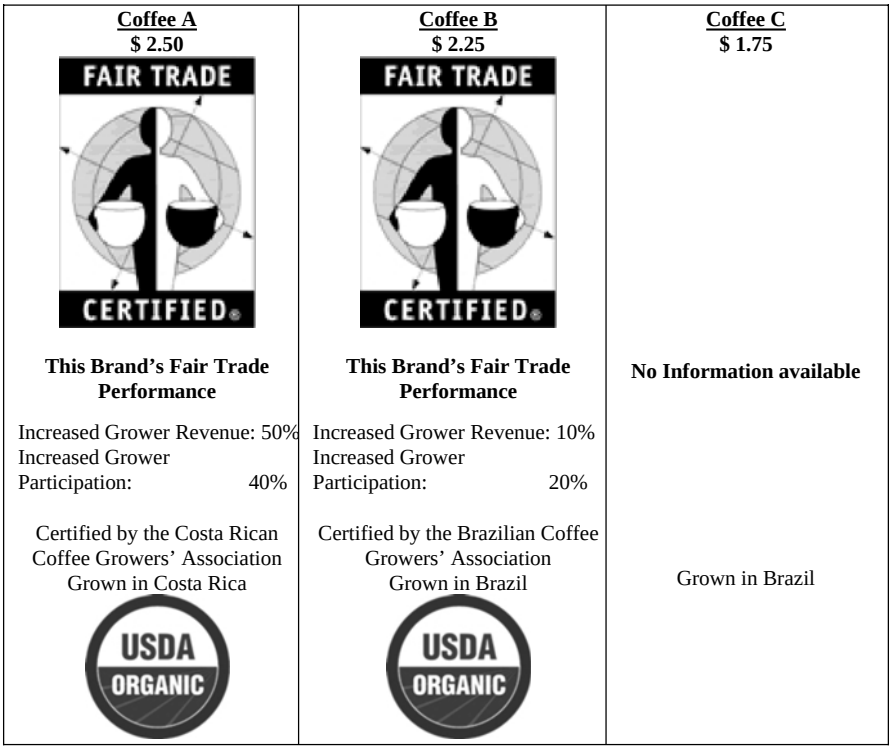

Fig. 2. Performance-based label experiment

[13] Previous research has shown that when information is missing, consumers often look for proxies from other attributes of the product or from knowledge about closely related brands (Ross and Creyer 1992; Johnson and Levin 1985; and Ford and Smith 1987).

[14] Unfortunately, my experimental design did not allow me to disentangle the private benefits and certifying agency effects associated with δ. In order to do so, respondents would need to evaluate a subset of labeled products having no certifying agency information.

Using econometric methods proposed by Louviere et al. (2000), it is possible to exploit the differences in equations (3) and (5) to test whether parameter homogeneity holds and therefore to completely isolate how performance impacts valuation of the labeled product. To do this, we estimate two models: a model where common parameters across the two experiments are restricted to be equal, and an unrestricted model where parameters are freely estimated across the two experiments. Define the joint set of parameters from the traditional labeled (denoted by t) and the performance labeled (denoted by p) programs to be estimated as $\psi = \{\alpha^t, \beta^t, \delta^t, \omega^t, \alpha^p, \beta^p, \delta^p, \phi^p, \omega^p, \lambda\}$, where λ is the relative scale parameter that calibrates the restricted parameter estimates to account for error structure differences across the models (see Louviere et al. 2000 for a detailed discussion of the scale parameter).

The likelihood function for the joint model is given by

$$\begin{aligned} L(\psi) = & \sum_{n \in t} \sum_{s \in S^t} y_{sn} \ln\left[\text{Prob}_{is}\left(P, C, A, O, \left(\lambda\alpha^t, \lambda\beta^t, \lambda\delta^t, \lambda\omega^t\right)\right)\right] \\ & + \sum_{n \in p} \sum_{s \in Sp} y_{sn} \ln\left[\text{Prob}_{is}\left(P, C, A, O, G, \left(\alpha^p, \beta^p, \delta^p, \lambda\omega^p, \phi^p\right)\right)\right] \end{aligned} \tag{6}$$

where $y_{sn} = 1$ if respondent n chooses product s. The restricted model can be estimated by setting $\{\alpha^t = \alpha^p, \beta^t = \beta^p, \delta^t = \delta^p, \omega^t = \omega^p\}$. To freely estimate both set of parameters, equation (6) is estimated with only one restriction, λ=1.[15]

In stated preference studies, where respondents are given all information necessary to make a product choice, Louviere et al. (2000) argue that the error term in the model is capturing the difficulty in assessing and choosing a product. The relative scale parameter (λ) provides a way of measuring the difficulty (commonly referred to as the cognitive burden) of the two experiments (Deshazo and Fermo 2002; Mazzotta and Opaluch 1995; Holmes and Boyle 2005). Given our parameterization of the model, an estimate of λ greater than one reveals that the variance of the error term in the traditional model is smaller than the performance-based model. As the variance of the error term for a given GEV model increases, the unobservable elements of the choice increasingly dominates the discrete choice comparison. Since all information relevant for choice is included in the

[15] This is equivalent to separately estimating the traditional and performance-based models.

survey instrument for SPDC experiments, increased dominance of ε indicates increased cognitive burden.

5 Results

Table 2 presents the results from the jointly estimated model and the unrestricted performance-based and traditional labeling model. The results across all three columns in the table reveal striking similarities: in each model a higher price decreases the likelihood of purchasing a given coffee product, and consumers tended to be more willing to purchase USDA certified products. Only in the joint model are country of origin coefficients (relative to Colombia) positive and significant at the five percent level. The organic coefficient on the non-fair trade coffee was not positive or significant (except for the joint model), indicating that consumers choosing the non-fair trade labeled coffee are not more likely to choose a non-fair trade labeled product if it is organic.

The certification agency and private benefit effects of a labeled coffee is always positive and significant for nearly all certifying agencies in both the traditional and performance-based label. The marginal value of a label ensuring fair trade PPM $\left(-\delta/\alpha\right)$ is worth {\$.83,\$1.41,\$.99} for the traditional model (for the Consumer's Union, the USDA, and a foreign country Growers' Association, respectively) and only {\$.17,\$.62,\$.22} for the performance based label. These results suggest that the consumer who is evaluating a traditional labeled product bundles with that label some priors concerning the performance of the labeling program.[16] An analogous explanation is that consumers' who are evaluating performance-based labels are able to evaluate label veracity via the performance data rather than proxying with certifying agency. The performance of the fair trade product was found to significantly increase the likelihood of purchasing a fair trade labeled coffee. Higher performing products are preferred to lower performing products. Both poverty reduction and the level of grower participation had similar effects on the likelihood of choosing the product.

[16] An anonymous reviewer conjectured that the respondent may proxy performance with other information on the label such as country of origin when evaluating a traditional label. We tested this conjecture, by estimating a model where country of origin parameters were unrestricted in the choice model. Our results show that country of origin effects were not significantly different across the models.

Table 2. Estimation results

	Parameter	Joint Model	Traditional Model	Performance Model
	Price (α)	-1.0263**	-1.1535**	-1.2413**
		(-7.840)	(-4.608)	(-9.597)
Certifying Agency	Consumer Union (δ)	.3552**	.9596**	.2152
		(2.724)	(2.660)	1.532
	USDA (δ)	.9762**	1.6315**	.7708**
		(8.388)	(5.600)	(5.747)
	Grower's Association (δ)	.5614**	1.1412**	.3373**
		(5.301)	(4.805)	(2.454)
Country of Origin	Brazil (β)	.2582**	.3383	.2624*
		(2.270)	(1.573)	(1.876)
	Kenya (β)	.0713	.1061	N/A
		(.394)	(.498)	
	Costa Rica (β)	.2137**	.1220	.2362*
		(1.979)	(.590)	(1.824)
Label Attributes	Organic (ω)	-.2643**	.0025	-.1981
		(-2.119)	(.9870)	(-1.286)
	Poverty (ϕ)	1.2127**	N/A	1.9152**
		(4.158)		(6.070)
	Participation (ϕ)	1.3360**	N/A	1.8763**
		(4.212)		(5.638)
	Relative Scale (λ)	1.0285**	N/A	N/A
		(4.711)		
	Mean Log Likelihood	-1.00737	-1.01240	-0.986745
	N	1270	448	822

Using the joint model I test whether consumers trade-off coffee product attributes in the same way. To estimate the joint model, I restrict the parameters on price, certifying agency, and country of origin across the traditional and performance-based labeling products. I can then investigate consumer priors about the performance of the labeling program. If parameter homogeneity holds (that the restrictions are appropriate) then consumers in both experiments value the certification and private benefit effects in similar ways across the two experiments, and the addition of performance attributes to a traditional label merely increases consumer WTP for labeled products over and above these benefits. Results indicate that parameter homogeneity is rejected using standard log-likelihood ratio tests. This provides evidence that consumers evaluating a traditionally labeled product are willing to pay significantly more for the labeled product than might be expected based on the certification and private

benefits effects alone. There is evidence that consumers have priors about the program performance even when no information on performance is provided. Another interesting finding from the joint model is that cognitive burden associated with the performance label seems to be relatively similar to the traditionally labeled product (since $\lambda \approx 1$). Following the interpretation of relative scale parameters, the addition of two additional attributes describing the labeling programs performance makes the choice problem no more difficult than under a traditional label. Contrary to other studies, my results do not show significant increases in cognitive burden when adding label information and provides some evidence that consumers who have no information about performance make guesses as to how effective labeling programs are.[17]

The models can also be used to examine price premiums (or WTP) for eco-labeled products over and above what would have paid for an identical yet not labeled product.[18] Using standard formula for WTP multinomial logit discrete choice models (Hanemann 1999), I show WTP for the traditional label (denoted by the horizontal line) and the performance-based label (assuming a 10% increase in grower participa-tion) for varying levels of increased revenues going to the grower. Notice that in Figure 3 there is a critical value of performance beyond which higher performance increases WTP relative to the traditional label. If the goal of the labeling program is to incentivize grower PPM due to higher consumer WTP, then this finding suggests that a new labeling programs may benefit from starting with a traditional PPM labeling scheme until performance improves beyond a threshold level. As performance rises, the switch to performance-based labeling could begin.[19]

---

[17] Scammon (1997), Roe et al. (1999), and Bei and Widows (1999) explore the issue of quantity of information and its effect on cognitive burden. Of these studies, only Bei and Widows (1999) find that increased information actually improves response efficiency, especially for experienced consumers.

[18] To calculate WTP I compare two coffees: one coffee is not labeled, is priced at $1, and is grown in Colombia, while the other coffee is priced at $1, is labeled, and is certified by the coffee grower's association in the country of origin, also Colombia. To calculate the WTP for performance labeled coffee, I compare the identical coffees except that participation rate increases are 10% and we allow changes in grower revenue to vary for Figure 3. Of course, parameter vectors differ according to the type of label.

[19] An anonymous referee points out that if such a rule were institutionalized, then the rule will likely become internalized in consumer expectations. If this is the case, then a lack of performance data on a label will be a clear signal to consumers that performance criteria are not being met.

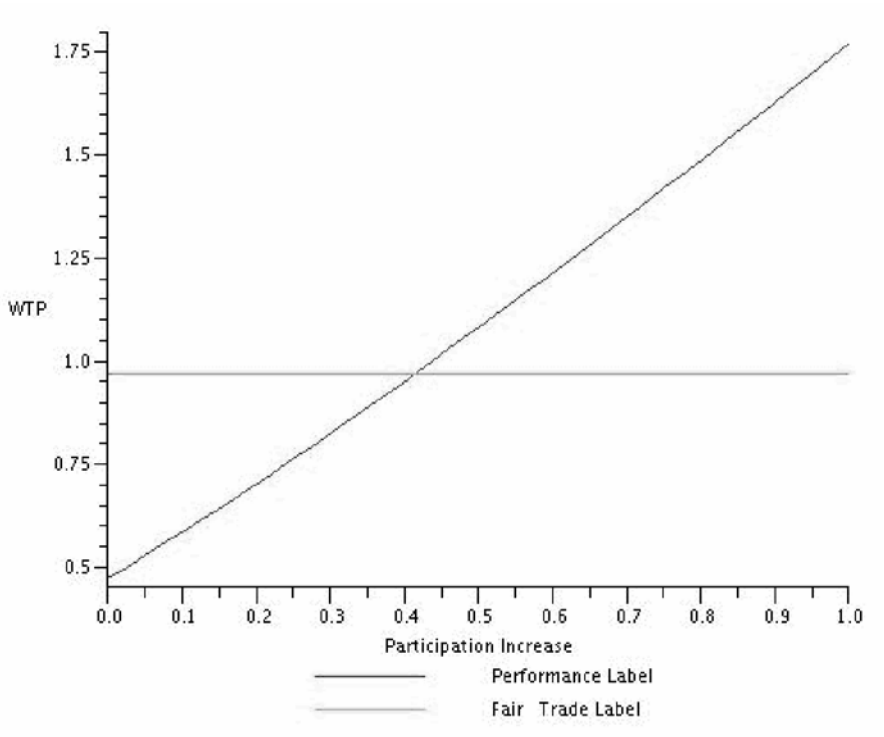

Fig. 3. Price premia for labeled coffee (assumes 10 % increase in grower participation)

6 Conclusion

In this paper, I argue that eco and social labeling schemes as currently implemented leave a lot to the imagination when it comes to consumer preferences for labeled products. Consumers who are interested in more than merely the private benefits associated with purchasing a labeled product are left guessing as to the overall impact of the labeling program on the related public good for the vast majority of labeling programs found around the world today. If consumers' WTP for labeled products is a function of both the overall provision of public goods provided by a labeling program and the private benefits from choosing a labeled product, then the information conveyed by today's labeling schemes may be woefully inadequate from a consumer's standpoint. In this paper, I examine the issue of private and public goods benefits related to a labeling program.

Using two stated preference experiments, I investigate how consumers react to two very closely related purchases of labeled products. In the first experiment, I offer the consumer several coffee products some labeled and some not. The label in the first experiment merely informs the consumer that the production related to the product they purchased did not have negative socio or environmental impacts and is intentionally designed to mimic the majority of labeled products on the market today. In the second experiment, I introduce more information related to the label. In addition to assuring consumers that the product meets socially responsible production standards, the second experiments informs consumers of public goods provided by the labeling program by including on the label, performance metrics of fair trade coffee programs - measured by grower involvement and increased grower revenue.

The results show that consumers are willing to pay more for a higher performing labeled product and provides evidence that consumers' value both public and private benefits from labeled products. Additionally, the econometric specification allows a test of what consumers believe to be the performance of labeling programs when the information is absent from the label. The results show that people probably do have priors over label performance, and, further, labels with more information do not place more cognitive burden on respondents.

Practically speaking, implementation of performance labeling programs does increase the information requirements for certifying agencies, and the results show that poorly performing programs will not receive as high a price premium than a better performing program. Therefore, some care should be taken when starting a new labeling program where poor performance is predictable. The use of performance-based labels does provide further incentives to producers. Since the performance of a labeling program depends on the actions of a number of producers, the label effectively builds a collective reputation of the producer (as indicated by performance). The implications of such a program in a general equilibrium sense is beyond the scope of this paper but is being pursued in other research.[20]

[20] It should be noted that the theoretical literature on child-labor free labeling provides interesting hints about the general equilibrium implications of performance-based labeling. Consider the model of Basu et al. (2004), where the developed country production is child-free, and it is competing with labeled and unlabeled products from developing countries. The northern product could be considered a performance-based product having perfect performance and it is competing with traditionally labeled products. We thank an anonymous referee for this insight.

References

Adamowicz WL, Louviere J, Williams M (1994) Combining revealed and stated preference methods for valuing environmental amenities. Journal of Environmental Economics and Management 26:271-292

Auld GB, Cashore B, Newsome D (2001) A look at forest certification through the eyes of the United States wood and paper producers. Paper presented at the Global Initiative and Public Policies: First International Conference on Private Forestry in the 21st Century, Atlanta Georgia, 25-27 March

Baharuddin HJ, Simula M (1994) Certification schemes for all timber and timber products. International Tropical Timber Association, Yokohama

Bass T, Markopoules R, Grah G (2001) Certification's impacts on forests, stakeholders and supply chains: instruments for sustainable private sector forestry services. International Institute for Environment and Development, London

Basu AK, Chau NH, Grote U (2006) Guaranteed manufactured without child labor: the economics of consumer boycotts, social labeling and trade sanctions. Review of Development Economics, forthcoming.

Basu AK (1999) Child labor: cause, consequence, and cure, with remarks on international labor standards. Journal of Economic Literature 37:1083-1119

Bei LT, Widows R (1999) Product knowledge and product involvement as moderators of the effects of information on purchase decisions: a case study using the perfect information frontier approach. Journal of Consumer Affairs 33:165-186

Ben-Akiva M, Morikawa T (1990) Estimation of switching models from revealed preferences and stated intentions. Transportation research 24:485-496

Bjorner TB, Hansen LG, Russell CS (2004) Environmental labeling and consumers' choice - an empirical analysis of the effect of the Nordic swan. Journal of Environmental Economics and Management: forthcoming

Boxall PC, Adamowicz WL, Swait J, Williams M, Louviere J (1996) A comparison of stated preference methods for environmental valuation. Ecological Economics 18:243-253

Brown D (1999) Can consumer product labels deter foreign child labor exploitation? Department of Economics Working Paper 19, Tufts University, Department of Economics, Medford

DeShazo JR, Fermo G (2002) Designing choice sets for stated preference methods: the effects of complexity on choice consistency. Journal of Environmental Economics and Management 44:123-143

Ford GT, Smith RA (1987) Inferential beliefs in consumer valuations: an assessment of alternative processing strategies. Journal of Consumer Research 14:363-371

Gudmundssen E, Wessells CR (2000) Ecolabeling seafood for sustainable production: implications for fisheries management. Marine Resource Economics 15:97-113

Hanemann WM (1999) Welfare analysis with discrete choice models. In: Herriges JA, Kling CL (eds) Valuing recreation and the environment. Edward Elgar Publishers, Northampton

Hicks RL (2002) Stated preference methods for environmental management: recreational summer flounder angling in the northeastern United States. National Marine Fisheries Service, Siver springs, http://rlhick.people.wm.edu/Working_Papers/Stated%20Preference%20Final%20Report_April%2016.pdf

Hoehn, JP, Randall A (2002) The effect of resource quality information on resource injury perception and contingent values. Resource and Energy Economics 24:13-31

Holmes TP, Boyle KJ (2005) Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions. Land Economics 81:114-126

Irland LC (2002) The elusive green premium. Paper presented at the Sustainable Wood Supply through Market Based Incentives Workshop, Orono, 17 May

James D (2000) Justice and Java: coffee in a fair trade market. NACLA Report to the Americas 34:11-42

Johnson RD, Levin IP (1985) More than meets the eye: the effects of missing information on purchasing evaluation. Journal of Consumer Research 12:169-177

Louviere JJ, Hensher DA, Swait JD (2000) Stated choice methods: analysis and applications with a contribution by Adamowicz W, Cambridge. Cambridge University Press, New York and Melbourne

Mazzotta J, Opaluch JJ (1995) Decision making when choices are complex: a test of Heiner's hypothesis. Land Economics 71:500-515

Mitchell RC, Carson, RT (1989) Using surveys to value public goods: the contingent valuation method resources for the future. RFF Press, Washington DC

Murray DL, Reynolds T, Taylor PL (2003) One cup at a time: poverty alleviation and fair trade coffee in Latin America. The Ford Foundation, New York

Nimon W, Beghin JC (1999) Are eco-labels valuable? Evidence from the apparel industry. American Journal of Agricultural Economics 81:801-811

O'Brien KA, Teisl MF (2004) Eco-information and its effect on consumer values for environmentally certified forest products. Journal of Forest Economics 10:75-96

Roe B, Levey AS, Derby BM (1999) The impact of health claims on consumer search and product evaluation outcomes: results from FDA experimental data. Journal of Public Policy and Management 18:89-105

Ross W, Creyer EH (1992) Making inferences about missing information: the effects of existing information. Journal of Consumer Research 19:14-25

Scammon DL (1977) Information load and consumers. Journal of Consumer Research 4:148-155

Sedjo RA, Swallow SK (2002) Voluntary eco-labeling and the price premium. Land Economics 78:272-284

Stecklow S, White E (2004) At what price virtue? At some retailers, 'fair trade' carries a very high cost stores charge big markups on goods intended to help farmers in poor countries. The Wall Street Journal, June 8:A1

Swait, J, Adamowicz W (1996) The effect of choice environment and task demands on consumer behavior: discriminating between contribution and confusion. Department of Rural Economy, University of Alberta

Swait J, Louviere JJ, Williams M (1994) A sequential approach to exploiting the combined strengths of SP and RP data: application to freight shipper choice. Transportation 21:135-152

Taylor PL (2005) In the market but not of it: fair trade coffee and forest stewardship council certification as market-based social change. World Development 33:129-147

Teisl MF, Roe B, Hicks RL (2002) Can eco-labels tune a market? Evidence from dolphin-safe labeling. Journal of Environmental Economics and Management 43:339-359

TransFair USA (2005) 2005 fair trade facts and figures. TransFair USA, Oakland

USEPA (1993) Status report on the use of environmental labels worldwide. United States Environmental Protection Agency, Washington DC

USEPA (1994) Determinants of effectiveness of environmental certification and labeling programs office of pollution prevention and toxics. United States Environmental Protection Agency, Washington, DC

Waffle R (1997) Forest certification: who profits? Green certification is not needed for sustained forestry. Wood and Wood Products, September:97-101