The Caloric Costs of Culture: Evidence from Indian Migrants

David Atkin

Online Appendix

A Cultural preferences for food in India

In this appendix, I review the literature on cultural food preferences in India and present a finding that motivates the main analysis in the paper. I provide suggestive evidence that, despite the low levels of caloric intake and high child-malnutrition rates detailed in Section 2.2, households seem to be making nutritionally-suboptimal food choices in part due to cultural preferences for certain foods.

The first thing to note is that almost all households could have purchased a bundle providing the recommended caloric norms for a smaller outlay than they actually spent on food. For example, if households reallocated all their food spending to the item in their consumption bundle that provided the most calories per Rupee, the percentage of households below the recommended caloric norms would fall from 64.9 percent to 9.5 percent. However, this bundle is likely to be deficient in other nutrients such as proteins and vitamins, as well as highly monotonous. Therefore, the fact that some households are not purchasing the most calorically-efficient bundle does not imply that they are making nutritionally-suboptimal food choices based on cultural considerations. Households may simply be maximizing nutrition and a taste for variety.

A review of the literature on cultural food preferences in India provides important background to my paper and motivates a more convincing counterfactual.

A.1 Review of the literature on cultural food preferences in India

It will be important for my analysis that there are many different food cultures within India, and that these food cultures differ across states. The field of nutritional anthropology has identified many different food cultures across religious, caste and ethnolinguistic groups within India.

Chakravarti (1974) combines fieldwork and survey evidence to categorize many dimensions of food culture in India, while Harris (1985) and Nair (1987) and Simoons (1994) focus narrowly on animal consumption. It is here, in the attitude towards animal products, where cultural food preferences vary most dramatically across India. Of the major religious groups in India, Jains and Buddhists are generally vegetarian due to a belief in non-violence towards animals. Christians, Sikhs and Muslims eat animal products although the latter will not eat pork. For the Hindu majority that comprises over 80 percent of the population, adherence to vegetarianism differs by caste.[39] Typically, members of the Brahmin (priest) and Vaishya (trading) castes are vegetarian while members of the Kshatriya (warrior) and Kayasthas (service) castes are non-vegetarian. Lower caste households vary in attitudes towards meat eating, with some groups even consuming beef, a taboo food for most Hindus.

[39] There are several justifications for Hindu vegetarian practices: the principle of non-violence towards animals (Simoons, 1994, p. 6), and the purifying "satvic" properties of cereals and most vegetables (Khare, 1992, p. 208).

However, this categorization by caste masks substantial regional heterogeneity. For example, Hindus of all castes eat meat in the parts of the far north states of Himchal Pradesh, Uttar Pradesh and Jammu and Kashmir. Brahmin Hindus in West Bengal consume both fish and goat but not chicken which is taboo. Meanwhile, Brahmins in Assam further east do eat chicken, as well as fish and goat. Chakravarti (1974) argues that it is difficult to provide a single explanation for this diversity of attitudes and hypothesizes that interactions with neighboring cultures, local geography and adherence to different Hindu deities all play roles.[40]

There are also differences in the acceptance of non-meat items. Vegetarians in some parts of India consume eggs but others, such as Gujaratis who follow Swaminarayan, will not. In addition, many North Indians avoid eggs in summer believing them to be a "hot" food that can harm the body by raising its temperature. Conversely, "cold" foods such as citrus fruits, are avoided in winter when they may cause colds and pneumonia (Chakravarti, 1974; Pool, 1987). In general, balancing "hot" and "cold" foods is believed to be necessary for the body's well-being (except for pregnant women for whom hot foods are harmful and cold foods beneficial). However, as Nag (1994) catalogs in detail, which precise foods are classified as hot and which foods cold varies dramatically across regions. For example groundnut is perceived as hot in Tamil Nadu but cold in Gujarat. There is also substantial regional diversity in beliefs regarding milk consumption, even within areas where bovines were historically used for agricultural production (Simoons, 1970).[41]

Attitudes towards vegetables also vary considerably across India. High caste and strict Hindus in the north of India will refrain from eating plants in the Allium genus like onions and garlic which are thought to overexcite passions (an aversion shared by some Jains and Buddhists), while in the south of India onion is often permitted but not garlic (Behura, 1962; Simoons, 1998). There is similar variation in the consumption of mushrooms as some high caste Hindus consider them unclean since they grow in dung. Simoons (1998) even highlights regional variations in attitudes towards eating potatoes, salt and the legume urd.

A.2 Rice and wheat consumption across India

I now turn to cereals, the category of food consumption most relevant for nutrition. Rice and wheat are the two dominant carbohydrates in India, together accounting for an enormous 65.5 percent of total caloric consumption in my sample. The NSS data provides strong empirical evidence of regional differences in preferences for rice and wheat. The upper panel of Appendix Figure A.1 plots the relative consumption of rice and wheat calories against the relative price of rice and wheat calories for the 22,148 villages where I observe both rice and wheat purchases.[42] Although the relationship between relative price and relative quantity is downward sloping, for most observed price ratios there are some

[40] For example, the goddess Varahi is particularly revered in West Bengal and is depicted holding a fish, potentially explaining why Brahmins eat fish in that state.

[41] Of course, unlike every other example in this section, there is a strong genetic component to variation in milk consumption since some adult populations are lactose intolerant. In my empirical work, I will not be able to separate genetic variation in lactose tolerance across states from variation in cultural preferences for milk across states.

[42] I exclude 2698 villages where I observe no purchases of rice or wheat and so do not know the relative price. These villages would likely strengthen my findings since they possess extreme rice/wheat preferences. Relative consumption is the share of rice calories in total rice and wheat calories consumed by sample households in the village. The relative price is the log of the ratio of median unit values (i.e. expenditures divided by calories) across purchasing households.

villages consuming relatively large quantities of rice and others relatively large quantities of wheat. This observation applies both when rice is a relatively cheap calorie source compared to wheat and when it is relatively expensive.

Regional preferences for rice and wheat in India are well documented in the nutritional anthropology literature. As highlighted in the introduction, even in 300 A.D. inhabitants of Northwest India were known for their special liking of wheat while inhabitants of East India were known for their love of rice (Prakash, 1961). Chakravarti (1974) classifies modern India into three divisions based on food habits for cereals: rice consuming areas in east and along the south and south-west coastline of India, bread consuming areas in the north and northwest of the country, and rice and bread consuming areas in the center of the country.

As an illustrative example, the lower panel of Appendix Figure A.1 highlights two states with substantial price overlap, Kerala in the south and Punjab in the north. Keralites consumed thirteen times more rice than wheat and Punjabis ten times more wheat than rice at similar relative prices (the Kerala fixed effect in a regression of rice budget shares on relative prices is a massive 0.80).

Do these regional preferences for rice and wheat in India mean that households are consuming fewer calories than they could if they only cared about nutrition and dietary variety? To answer this question, I perform the following counterfactual. For every household, I calculate the amount it would cost to purchase a bundle with the same total quantity of calories they currently obtain from rice and wheat but swapping the quantities of rice and wheat.[43] For 42 percent of households, this bundle would cost less than they currently spend on rice and wheat (these are the households in the north-east and south-west quadrants of the upper panel of Figure A.1, with 94 percent of the switchers in the north-east quadrant). For this subset of households, I calculate the hypothetical caloric intake if the household allocated the cost savings to the cheaper of the two foods. Averaging over switching and non-switching households, hypothetical caloric intake is 6.2 percent higher than actual intake. This gap provides a measure of the number of calories households forgo in order to accommodate their preferences for rice or wheat.

This counterfactual solves several problems associated with the earlier counterfactual. First, by switching the quantities allocated to rice and wheat, the amount of dietary variety is not declining (in the sense that every household still obtains the same number of calories from the less-consumed food). Second, these two foods provide a similar number of calories per Rupee and contain a similar set of nutrients.[44] Therefore, switching between rice and wheat does not alter household nutrition along other dimensions.

Although informative, some of the switching households may prefer rice or wheat for idiosyncratic reasons, rather than because of the regional preferences which have a cultural dimension to them. For example, if I observe one household in rice-loving Kerala eating more wheat than rice despite rice being cheaper in that state, I would not want to conclude that this behavior is due to cultural preferences. In

[43] For households that don't consume both rice and wheat, I use the village median price as the price for the unpurchased cereal. As before, I restrict attention to the villages where I observe both rice and wheat purchases.

[44] In fact wheat is slightly more nutritious than rice, particularly if the rice is polished. As the caloric gains in the counterfactual come almost entirely from households switching rice for wheat, nutrition is also likely to improve along other dimensions. The average unit value for rice was 0.95 Rupees per 1000 calories in 1983 and 1.09 in 1987/88. Equivalent numbers for wheat were 0.66 and 0.75. According to the NSS, a kilogram of rice contains 3460 calories, 75 grams of protein and 5 grams of fat. Equivalent numbers for wheat were 3410, 121 and 17. Wheat has slightly more fiber and vitamins, particularly folic acid.

order to highlight the cultural dimension, I classify states as having either "rice-loving" or "wheat-loving" cultures. Atkin (2013) shows that, over many generations, habit formation leads to regional preferences for the foods the region is relatively well-suited to produce. Therefore, I utilize measures of land suitability from the Global Agro-Ecological Zones project (GAEZ) to proxy for each state's rice/wheat culture. The GAEZ data report the suitability of each state in India for growing both rice and wheat on a scale of 0 (not suitable) to 1 (very high suitability).[45] Consistent with the example highlighted above, Punjab is better suited to wheat production and Kerala to rice (the difference between the crop-suitability score for rice and wheat is 0.29 in Kerala and -0.13 in Punjab). Reassuringly for my proxy, land suitability strongly predicts relative consumption of rice and wheat across states, even after conditioning on relative prices.[46]

I label states as possessing a rice-loving culture if their GAEZ suitability measures are higher for rice than wheat, and a wheat-loving culture otherwise. With these labels, I redo the counterfactual but only allow two types of consumer to switch their consumption patterns: households in states with a rice-loving culture but where wheat is a cheaper calorie source than rice, and households in states with a wheat-loving culture but where rice is cheaper than wheat. Therefore, I would no longer switch the rice and wheat consumption of the household in Kerala mentioned above that eats more wheat than rice. In this new counterfactual, Indian households are forgoing 6.1 percent higher caloric intake in order to accommodate their cultural preferences.[47] Appendix Figure A.2 plots the full distribution of caloric intake under the counterfactual, as well as a histogram of the caloric gains generated. If the foregone calories were realized, the percentage of households consuming less than the recommended caloric norms would fall by 7.3 percent. Of course, if many households did switch across rice and wheat, local prices would change and the actual gain in caloric intake would likely be smaller.

This last exercise provides suggestive evidence that households in India make food choices based on cultural preferences that are suboptimal from a nutritional sense. However, since every village has a different vector of prices, some of these consumption patterns may be rationalized by complex substitution patterns. For example, if coconuts are stronger complements with rice than with wheat, the relatively high consumption of rice in the south of India may be due to cheap coconuts rather than a preference for rice. In the the analysis presented in the main text, I present a methodology that sidesteps this issue by considering all 169 foods and by focusing on consumption differences between households that have different cultural preferences yet face the same prices.

References

Atkin, David. 2013. "Trade, Tastes and Nutrition in India." *The American Economic Review*, 103(5): 1629–1663.

[45] I use is the "crop suitability index" for rain-fed agriculture using intermediate input usage. For rice, I take the maximum of the two index values for wetland and dryland rice cultivation. The GAEZ website www.iiasa.ac.at/Research/LUC/GAEZv3.0/ contains further details. Data are available for 28 of the 31 states.

[46] A simple OLS regression of the household rice budget share on both village-level relative rice prices and state-level relative suitability for growing rice yields a coefficient of 0.97 with a t-value of 149.41.

[47] Rice suitability is higher than wheat suitability for 23 of the 28 states and 39.5 percent of households switch in this counterfactual. Alternatively, I classify a state as rice-loving if the state fixed-effect in a regression of rice budget shares on relative prices is greater than 0.5. Using this measure, 25 out of 31 states are rice-loving. Under this definition, Indian households forgo 6.0 percent higher caloric intake in aggregate and 37.7 percent of households switch.

Behura, NK. 1962. "Meals and Food Habits in Rural India." *Bulletin of the Anthropological Survey of India*, 11(4): 111–137.

Chakravarti, AK. 1974. "Regional Preference for Food: Some Aspects of Food Habit Patterns in India." *The Canadian Geographer/Le Géographe canadien*, 18(4): 395–410.

Harris, Marvin. 1985. *Good to Eat: Riddles of Food and Culture.* Simon and Schuster.

Khare, R.S. 1992. "Annambrahman: Cultural Models, Meanings, and Aesthetics of Hindu Food." In *The Eternal Food: Gastronomic Ideas and Experiences of Hindus and Buddhists*. , ed. R.S. Khare, 201–220. SUNY Press.

Nag, Moni. 1994. "Beliefs and Practices about Food during Pregnancy: Implications for Maternal Nutrition." *Economic and Political Weekly*, 29(37): pp. 2427–2438.

Nair, K Narayanan. 1987. "Animal Protein Consumption and the Sacred Cow Complex in India." In *Food and Evolution: Toward a Theory of Human Food Habits*. , ed. Marvin Harris and Eric B. Ross, 445–454. Temple University Press.

Pool, R. 1987. "Hot and Cold as an Explanatory Model: The Example of Bharuch District in Gujarat, India." *Social Sciences and Medicine*, 25(4): 389–399.

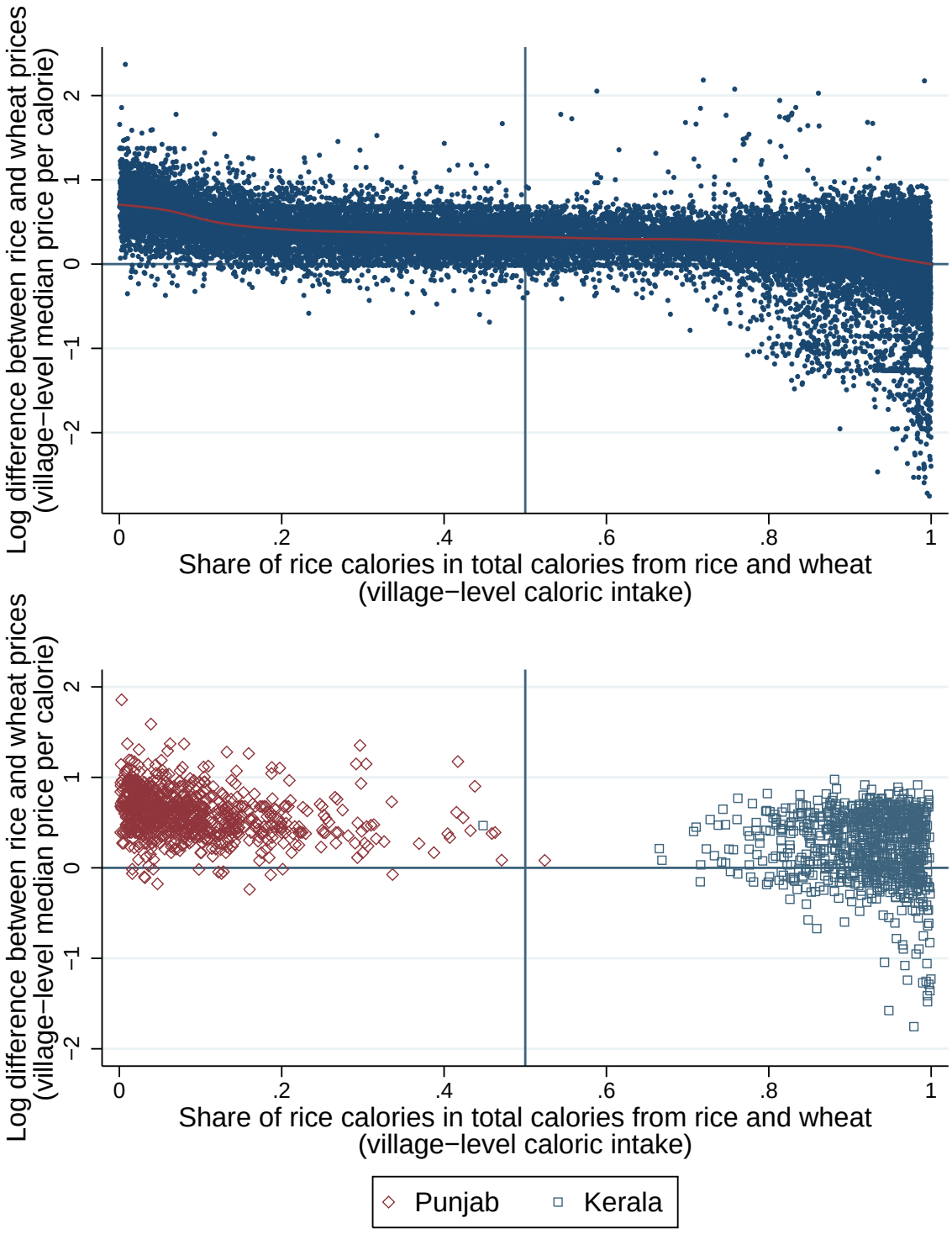
Prakash, Om. 1961. *Food and Drinks in Ancient India: from Earliest Times to c. 1200 AD.* Munshi Ram Manohar Lal. p. 1–33.

Simoons, Frederick J. 1970. "The Traditional Limits of Milking and Milk Use in Southern Asia." *Anthropos*, 65(3/4): 547–593.

Simoons, Frederick J. 1994. *Eat Not This Flesh: Food Avoidances from Prehistory to the Present.* University of Wisconsin Press.

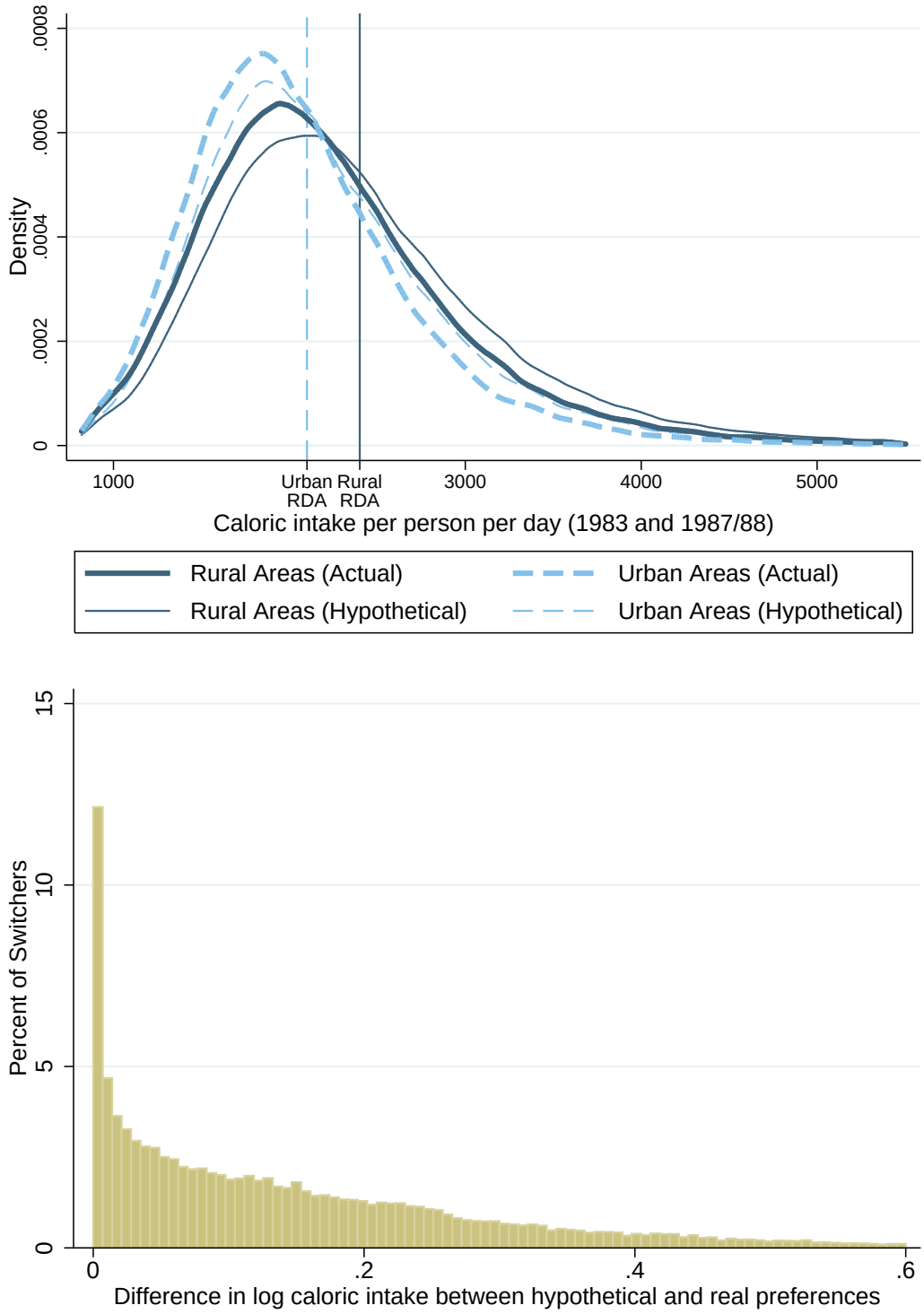
Simoons, Frederick J. 1998. *Plants of Life, Plants of Death.* University of Wisconsin Press.

Figure A.1: Regional consumption patterns for rice and wheat



Note: The upper panel plots the relative price of rice and wheat calories against the share of rice calories in the total calories from rice and wheat for the 22,148 villages where I observe both rice and wheat purchases in the two survey rounds. The lower panel highlights the observations for villages in Kerala and Punjab.

Figure A.2: Caloric intake under actual and hypothetical preferences for rice and wheat



Note: The upper panel displays the counterfactual distribution of caloric intake if households switched the quantities purchased of rice and wheat where advantageous, spending any cost savings on the cheaper of the two foods. The distribution includes all households in the 22,148 villages where I observe both rice and wheat purchases in the two survey rounds. The lower panel displays a histogram of the caloric gains available to the switchers in the counterfactual exercise.

B Examining the bias from endogenous controls for food expenditure

In Section 4.1 of the main text, I instrument the food expenditure polynomial with a polynomial in non-food expenditure. In this appendix, I provide more details on the claim that the true value of the coefficient on the migration dummy is bounded (or at least very likely to be bounded) between the instrumented and uninstrumented estimates.

Consider a simplified version of the uninstrumented regression where I include only one control variable, a control for log food expenditure:

$$\ln calories_i = \alpha + \beta migrant_i + \pi \ln foodexp_i + \varepsilon_i. \tag{8}$$

The endogeneity concern is that a positive shock to the demand for calories will raise both the error term and food expenditure, thereby biasing the estimate of β , the coefficient of interest (i.e. the concern is that $Cov(\ln foodexp_i, \varepsilon_i) > 0$). The bias on β can be written in terms of variances and covariance between the $migrant_i$ dummy and $\ln foodexp_i$:

$$\underset{n\to\infty}{plim}\hat{\beta} - \beta = \frac{Var(\ln foodexp_i)Cov(migrant_i, \varepsilon_i) - Cov(\ln foodexp_i, migrant_i)Cov(\ln foodexp_i, \varepsilon_i)}{Var(\ln foodexp_i)Var(migrant_i) - Cov(\ln foodexp_i, migrant_i)^2}$$

or

$$\underset{n\to\infty}{plim}\hat{\beta} - \beta = -\frac{Cov(\ln foodexp_i, migrant_i)Cov(\ln foodexp_i, \varepsilon_i)}{Var(\ln foodexp_i)Var(migrant_i) - Cov(\ln foodexp_i, migrant_i)^2}$$

since $migrant_i$ is (assumed) exogenous to the error term.

If instead I instrument $\ln foodexp_i$ with $\ln nonfoodexp_i$, in the first stage I would obtain the fitted values $\ln \widehat{foodexp}_i$ from the following regression:

$$\ln foodexp_i = \delta + \gamma_1 migrant_i + \gamma_2 \ln nonfoodexp_i + u_i$$

and the bias on the IV estimate of β, β_{IV}, can be written as:

$$\underset{n\to\infty}{plim}\hat{\beta_{IV}} - \beta_{IV} = -\frac{Cov(\ln \widehat{foodexp}_i, migrant_i)Cov(\ln \widehat{foodexp}_i, \varepsilon_i)}{Var(\ln \widehat{foodexp}_i)Var(migrant_i) - Cov(\ln \widehat{foodexp}_i, migrant_i)^2}.$$

A sufficient condition for the claim that β lies between $\hat{\beta}$ and $\hat{\beta_{IV}}$ to be true is that one of the biases is positive and the other is negative. The key insight is that $Cov(\ln foodexp_i, \varepsilon_i)$ and $Cov(\ln nonfoodexp_i, \varepsilon_i)$ likely take the opposite sign. This is because, as long as income does not respond to shocks to the demand for calories and there are no savings, an increase in food expenditure must be matched by a decrease in non-food expenditure. Therefore, since $Cov(\ln \widehat{foodexp}_i, \varepsilon_i) = Cov(\hat{\gamma}_2 \ln nonfoodexp_i, \varepsilon_i)$, the sufficient condition above is satisfied if:

$$sign\left(\frac{Var(\ln \widehat{foodexp}_i)Var(migrant_i)}{Cov(\ln \widehat{foodexp}_i, migrant_i)} - Cov(\ln \widehat{foodexp}_i, migrant_i)\right) \neq$$
$$sign\left(-\frac{Var(\ln foodexp_i)Var(migrant_i)}{Cov(\ln foodexp_i, migrant_i)} + Cov(\ln foodexp_i, migrant_i)\right)$$

or, equivalently:

$$sign\left(\frac{1}{Corr(\ln \widehat{foodexp}_i, migrant_i)} - Corr(\ln \widehat{foodexp}_i, migrant_i)\right) \neq sign\left(-\frac{1}{Corr(\ln foodexp_i, migrant_i)} + Corr(\ln foodexp_i, migrant_i)\right).$$

As the correlation coefficient is bounded between 0 and 1, this expression will clearly be satisfied as long as:

$$sign\left(Cov(\ln \widehat{foodexp}_i, migrant_i)\right) = sign\left(Cov(\ln foodexp_i, migrant_i)\right)$$

or, equivalently:

$$sign(\hat{\gamma}_1 Var(migrant_i) + \hat{\gamma}_2 Cov(\ln nonfoodexp_i, migrant_i)) = sign\left(Cov(\ln foodexp_i, migrant_i)\right)$$

(abstracting from the collinear case where either correlation is 1).

In general, a bounding argument of this sort is likely to work if two conditions are satisfied: 1) the covariance between the instrument and the error term is of opposite sign to the covariance between the endogenous variable and the error, and 2) the covariance between the instrument and the exogenous variable is of the same sign as the covariance between the endogenous variable and the exogenous variable if the instrument is positively related to the endogenous variable, and the opposite sign if the instrument is negatively related. If $\hat{\gamma}_1 = 0$, these two conditions are sufficient. As typically the instrument has a stronger relationship with the endogenous variable than the exogenous variable does, $\hat{\gamma}_1$ will often be small and so these two conditions will likely be sufficient.

In the case considered in the paper the sign condition is satisfied: food and non-food expenditures are positively correlated conditional on exogenous variables, $\hat{\gamma}_2 = 0.24 > 0$; the coefficient on the migrant dummy in the first stage is small and positive, $\hat{\gamma} = 0.020$; and both food and non-food expenditures are positively correlated with migrant status, $Corr(\ln foodexp_i, migrant_i)) = 0.151$ and $Corr(\ln nonfoodexp_i, migrant_i)) = 0.157$. Therefore, the true coefficient on the migrant dummy should lie between the uninstrumented and instrumented estimates (with the proviso that the argument above abstracts away from both the quadratic and cubic terms in log food expenditure and the additional control variables in equation 1 which, under extreme covariance structures, could overturn this conclusion).

C Additional Tables

Table C.1: Proportion of migrant households by migration route

Destination State: State Name	State No.	Origin State: 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	Total	State Size
A&N Islands	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.2	0.0
Andhra Pradesh	2	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	1.5	0.1	0.0	0.1	0.8	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.1	0.0	0.9	0.0	0.0	0.0	4.4	8.2
Arunachal	3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1
Assam	4	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.5	2.2
Bihar	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.1	0.0	0.0	0.0	0.3	0.8	1.8	10.1
Chandigarh	6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.1	0.0	0.9	0.1
Dadra & N Haveli	7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Delhi	8	0.0	0.1	0.0	0.1	0.3	0.0	0.0	0.0	0.0	0.1	1.8	0.1	0.1	0.0	0.1	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.5	0.0	0.1	0.1	4.1	0.2	8.8	0.9
Goa, Daman & Diu	9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.2
Gujarat	10	0.0	0.1	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.3	0.8	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.7	0.0	0.0	0.0	0.5	0.1	3.2	4.7
Haryana	11	0.0	0.0	0.0	0.0	0.2	0.1	0.0	0.8	0.0	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.4	2.1	0.0	0.0	0.0	1.7	0.0	6.9	2.0
Himachal Pradesh	12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.6
Jamu & Kashmir	13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.9
Karnataka	14	0.0	1.5	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.5	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	2.0	0.0	0.1	0.0	6.4	5.4
Kerala	15	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.0	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	1.8	3.4
Lakshdweep	16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Madhya Pradesh	17	0.0	0.2	0.0	0.0	0.3	0.0	0.0	0.1	0.0	0.2	0.1	0.0	0.0	0.0	0.1	0.0	0.0	1.6	0.0	0.0	0.0	0.0	0.1	0.0	0.2	1.3	0.0	0.0	0.0	2.6	0.2	7.4	8.0
Maharashtra	18	0.0	0.9	0.0	0.1	0.1	0.0	0.0	0.1	0.3	2.5	0.1	0.1	0.1	2.9	0.3	0.0	1.5	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.2	0.5	0.0	0.3	0.0	1.9	0.2	12.4	8.9
Manipur	19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
Meghalaya	20	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2
Mizoram	21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
Nagaland	22	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Orissa	23	0.0	0.4	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	1.0	3.4	4.0
Pondicherry	24	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.8	0.1
Punjab	25	0.0	0.0	0.0	0.1	0.3	0.1	0.0	0.2	0.0	0.1	1.6	0.6	0.2	0.0	0.0	0.0	0.2	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.6	0.0	0.0	0.0	1.3	0.1	5.7	2.6
Rajasthan	26	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.4	0.0	0.7	1.8	0.2	0.1	0.1	0.1	0.0	1.7	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.1	0.0	0.0	1.5	0.2	8.1	5.0
Sikkim	27	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0
TamilNadu	28	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.8	1.3	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.1	0.0	0.0	0.0	0.0	0.1	4.5	7.2
Tripura	29	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3
Uttar Pradesh	30	0.0	0.1	0.0	0.2	2.0	0.1	0.0	1.2	0.0	0.2	1.0	0.2	0.1	0.1	0.0	0.0	2.0	0.7	0.0	0.0	0.0	0.0	0.1	0.0	0.8	1.3	0.0	0.1	0.2	0.0	0.8	11.5	16.5
West Bengal	31	0.0	0.2	0.1	0.5	5.4	0.0	0.0	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.2	0.0	0.1	0.1	1.4	0.0	9.5	8.2
Total		0.2	4.9	0.2	1.8	10.1	0.4	0.0	3.2	0.4	4.4	6.9	1.7	1.0	6.1	3.0	0.0	7.1	6.9	0.2	0.3	0.1	0.3	1.8	0.5	5.1	7.8	0.2	5.1	0.5	15.8	4.0	100	100
State Size (NSS sample)		0.0	8.2	0.1	2.2	10.1	0.1	0.0	0.9	0.2	4.7	2.0	0.6	0.9	5.4	3.4	0.0	8.0	8.9	0.2	0.2	0.1	0.0	4.0	0.1	2.6	5.0	0.0	7.2	0.3	16.5	8.2	100	

Note: Each cell of the table shows the proportion of all migrants households that moved from a particular origin state (the columns) to a particular destination state (the rows). The State Size row and column reports the proportion of all households currently living in each state at the time of the surveys. All proportions use the household weights provided by the NSS.

Table C.2: Comparing the meals served and meals eaten out by migrants and non-migrants

Dependent variable:	(1) (Meals out)/ (meals at home)	(2) (Meals out)/ (meals at home)	(3) (Meals served to others)/ (meals at home)	(4) (Meals served to others)/ (meals at home)
$migrant_i$	0.00883	-0.000700	-0.000865	0.00264
	(0.00854)	(0.00715)	(0.00523)	(0.00816)
$migrant_i$		0.0250		-0.00956
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$		(0.0198)		(0.00808)
Observations	232,697	231,774	231,281	230,362
Within R-squared	0.007	0.007	0.027	0.027
Food Expenditure Controls	Yes	Yes	Yes	Yes
Demographics/Household Type Controls	Yes	Yes	Yes	Yes
Village-Round FE	Yes	Yes	Yes	Yes

Note: Dependent variable for columns 1-2 is the ratio of meals consumed by household members outside the home to those consumed by household members within the home. Dependent variable for columns 3-4 is the ratio of meals served to others (either guests or employees) to those meals consumed by household members within the home. Independent variables are $migrant_i$, a dummy for whether the household head or their spouse is an interstate migrant, and $migrant_i$ interacted with with an indicator variable that takes the value of 1 if a 1 Rupee reference bundle from the migrant's origin-state provides fewer calories than a 1 Rupee reference bundle from the migrant's destination village both priced at destination-village prices. All specifications include village-round fixed effects and flexible survey-round-specific controls for household size, demographics and type as well as a third-order polynomial in log per-capita food expenditure. All regressions are weighted using household weights and the standard errors are two-way clustered at the village-round and origin-state level. * signifies significance at the 10 percent level, ** at the 5 percent level and *** at the 1 percent level.

Table C.3: Comparing the caloric intake of migrants and non-migrants: heterogenous responses

Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)
	Log caloric intake (per person per day)					
Specification:	Round heterogeneity	Round heterogeneity	Sector heterogeneity	Sector heterogeneity	Route heterogeneity	Route heterogeneity
$migrant_i \times round38$	-0.0143***	-0.00002				
	(0.00389)	(0.00491)				
$migrant_i \times round43$	-0.0173***	-0.0108**				
	(0.00389)	(0.00498)				
$migrant_i \times round38$		-0.0387***				
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$		(0.00806)				
$migrant_i \times round43$		-0.0161**				
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$		(0.00789)				
$migrant_i \times rural_o\text{-}rural_d$			-0.00608	0.0129*		
			(0.00438)	(0.00673)		
$migrant_i \times rural_o\text{-}urban_d$			-0.00155	0.00945*		
			(0.00574)	(0.00529)		
$migrant_i \times urban_o\text{-}rural_d$			-0.0271***	-0.00781		
			(0.00682)	(0.0104)		
$migrant_i \times urban_o\text{-}urban_d$			-0.0362***	-0.0313***		
			(0.00639)	(0.00691)		
$migrant_i \times rural_o\text{-}rural_d$				-0.0347***		
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$				(0.00915)		
$migrant_i \times rural_o\text{-}urban_d$				-0.0773***		
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$				(0.0295)		
$migrant_i \times urban_o\text{-}rural_d$				-0.0318**		
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$				(0.0147)		
$migrant_i \times urban_o\text{-}urban_d$				-0.0364**		
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$				(0.0159)		
$migrant_i \times rice_o\text{-}rice_d$					-0.0191***	-0.00992**
					(0.00379)	(0.00473)
$migrant_i \times rice_o\text{-}wheat_d$					-0.0155**	-0.0104
					(0.00772)	(0.0127)
$migrant_i \times wheat_o\text{-}rice_d$					-0.00948	-0.00273
					(0.00607)	(0.00785)
$migrant_i \times wheat_o\text{-}wheat_d$					-0.0136***	0.00124
					(0.00470)	(0.00579)
$migrant_i \times rice_o\text{-}rice_d$						-0.0269***
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$						(0.00777)
$migrant_i \times rice_o\text{-}wheat_d$						-0.00979
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$						(0.0172)
$migrant_i \times wheat_o\text{-}rice_d$						-0.0216*
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$						(0.0122)
$migrant_i \times wheat_o\text{-}wheat_d$						-0.0374***
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$						(0.0112)
Observations	235,126	234,157	235,126	234,157	235,126	234,157
Within R-squared	0.732	0.731	0.732	0.732	0.732	0.731
Food Expenditure Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics/Household Type Controls	Yes	Yes	Yes	Yes	Yes	Yes
Village-Round FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: Dependent variable for all columns is log caloric intake per person per day. Independent variables are $migrant_i$, a dummy for whether the household head or their spouse is an interstate migrant, and $migrant_i$ interacted with with an indicator variable that takes the value of 1 if a 1 Rupee reference bundle from the migrant's origin-state provides fewer calories than a 1 Rupee reference bundle from the migrant's destination village both priced at destination-village prices. These independent variables are further interacted with dummy variables for survey round (columns 1-2), the origin and destination sectors of the migration route (columns 3-4), and the predominant cereal of the origin and destination states of the migration route based on whether the state-average rice or wheat budget share is larger (columns 5-6). All specifications include village-round fixed effects and flexible survey-round-specific controls for household size, demographics and type as well as a third-order polynomial in log per-capita food expenditure. All regressions are weighted using household weights and the standard errors are two-way clustered at the village-round and origin-state level. * signifies significance at the 10 percent level, ** at the 5 percent level and *** at the 1 percent level.

Table C.4: Comparing the anthropometrics of migrants and non-migrants

	(1)	(2)	(3)	(4)	(5)	(6)
Sample:	Children aged 0-5					
Dependent variable:	Weight-for-age Z score			Height-for-age Z score		
Specification:	Baseline	Household level	Unwinsorized	Baseline	Household level	Unwinsorized
migrant $_i$	-0.220**	-0.232**	-0.282**	-0.222*	-0.167	-0.325
	(0.103)	(0.105)	(0.129)	(0.122)	(0.130)	(0.214)
Observations	35,211	23,806	35,211	35,025	23,679	35,025
Within R-squared	0.035	0.056	0.027	0.019	0.030	0.013
Total Expenditure Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics/Household Type Controls	Yes	Yes	Yes	Yes	Yes	Yes
Village-Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Person Level and Weights	Yes	No	Yes	Yes	No	Yes
Household Level and Weights	No	Yes	No	No	Yes	No
Winsorized at +/- 5 standard deviations	Yes	Yes	No	Yes	Yes	No

	(7)	(8)	(9)	(10)	(11)	(12)
Sample:	Children aged 6-11			Adults age 12+		
Dependent variable:	Weight-for-age Z score	Height-for-age Z score	Body mass index	Body mass index (kg/m^2)		
Specification:	Baseline	Baseline	Baseline	Baseline	Household level	Unwinsorized
migrant $_i$	0.0670	0.0626	0.0130	-0.148	-0.172	-0.254*
	(0.0913)	(0.0904)	(0.243)	(0.0993)	(0.106)	(0.144)
Observations	27,537	32,867	39,025	136,244	74,345	136,244
Within R-squared	0.029	0.020	0.008	0.115	0.043	0.014
Total Expenditure Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics/Household Type Controls	Yes	Yes	Yes	Yes	Yes	Yes
Age and Sex Controls	No	No	Yes	Yes	Yes	Yes
Village-Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Person Level and Weights	Yes	Yes	Yes	Yes	No	Yes
Household Level and Weights	No	No	No	No	Yes	No
Winsorized at +/- 5 standard deviations	Yes	Yes	Yes	Yes	Yes	No

Note: Data come from the 2005 and 2011 rounds of the India Human Development Survey that surveyed 41,554 households in 1503 villages and 971 urban blocks in 2005, and 42,152 households in 1420 villages and 1042 urban blocks in 2011. Anthropometric data only collected from children age 0-5, 8-11 and women under 49 in 2005 round. Dependent variable for columns 1-3 is the child weight-for-age Z scores and dependent variable in columns 4-6 is child height-for-age Z scores, both using the 2006 WHO child growth standards for all children aged 0 to 5. Columns 7-8 repeat the baseline analysis for older children aged 6-11. Dependent variable in columns 9-12 is the body mass index (kg/m^2) for both children aged 6-11 and adults aged 12 and older. Independent variable is $migrant_i$, a dummy for whether the household reports having migrated from another state. All specifications include village-round fixed effects and flexible survey-round-specific controls for household size and demographics (the proportion of household members that fall into five sex-specific age brackets; 0-4, 5-9, 10-14, 15-55 and over 55), household type (11 occupation categories), as well as a third-order polynomial in log per-capita expenditure. Columns 9-12 are not age and sex adjusted Z scores and so also include a full set of age fixed effects for every ages below 18 and a single dummy for age 18 and above, each interacted with both gender and round. As with the NSS data, village denotes a village in rural areas and urban blocks in urban areas. Regressions in columns 1, 3-4, 6-10 and 12 are at the person level and weighted using representative person weights. Regressions in columns 2, 5 and 11 are at the household level, averaging the anthropometric measures across children if multiple children are present in the household, and use representative household weights. Standard errors are clustered at the village-round level. The absence at the time of writing of linking files between the two survey rounds of the panel precludes cross-round clustering at the household level. * signifies significance at the 10 percent level, ** at the 5 percent level and *** at the 1 percent level.

Table C.5: Comparing bundles of migrants and non-migrants: subpopulations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable: Correlation $corr_{io_v}$ of household i budget shares with the average budget shares of the origin states o_v of migrants in village (169 foods)							
Sample:	Baseline Specification	Caloric shortfall	Caloric shortfall >250	Caloric shortfall >500	Below median per capita expenditure	Bottom quartile per capita expenditure	Below median per capita food expend.	Bottom quartile per capita food expend.
$migrant_i^{o_v}$	0.0226***	0.0213***	0.0175***	0.0133***	0.00646**	0.00142	0.00900**	0.00312
	(0.00290)	(0.00352)	(0.00325)	(0.00432)	(0.00314)	(0.00694)	(0.00392)	(0.00688)
Observations	108,743	50,925	34,651	19,179	21,179	7,326	23,165	8,445
Within R-squared	0.098	0.105	0.112	0.125	0.090	0.100	0.102	0.104
Food Expenditure Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics/Household Type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village-o_v-Round FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Dependent variable: Correlation $corr_{io_v}$ of household i budget shares with the average budget shares of the origin states o_v of migrants in village (169 foods)							
Sample:	Children under 5 in household	Children under 5 and below median food expend.	Children under 16 in household	Children under 16 and below median food expend.	Migrated 5 or more years ago	Migrated 10 or more years ago	Migrated 20 or more years ago	Literate
$migrant_i^{o_v}$	0.0183***	0.0114**	0.0191***	0.00875**	0.0216***	0.0209***	0.0187***	0.0257***
	(0.00359)	(0.00469)	(0.00273)	(0.00407)	(0.00275)	(0.00283)	(0.00385)	(0.00320)
Observations	41,718	12,628	76,074	20,407	99,383	93,185	82,590	76,684
Within R-squared	0.073	0.113	0.066	0.093	0.093	0.093	0.092	0.125
Food Expenditure Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics/Household Type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village-o_v-Round FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Dependent variable is the correlation between household *i*'s vector of food budget shares and the vector of mean budget shares for non-migrant households in state o_v, where state o_v is the origin-state of a migrant in the household's village. The independent variable is $migrant_i^{o_v}$ (a dummy for whether the household head or their spouse is an interstate migrant from o_v). All specifications include village-o_v-round fixed effects and flexible survey-round-specific controls for household size, demographics and type as well as a third-order polynomial in log per-capita food expenditure. The column headings denote the various subpopulations on which the regressions are run. Caloric shortfall defined relative to the 2400 calorie urban/2100 rural calorie norms. All regressions are weighted using household weights and the standard errors are two-way clustered at the village-round and origin-state level. * signifies significance at the 10 percent level, ** at the 5 percent level and *** at the 1 percent level.

Table C.6: Comparing cultivation patterns of migrants and non-migrants

	(1)	(2)	(3)
	Dependent variable: Household cultivates crop g (binary variable)		
Sample:	All 43rd Round Households	All 43rd Round Cultivating Households	All 43rd Round Migrant Households
Proportion of origin-state households cultivating g	0.0378**	0.0667	0.0357
	(0.0179)	(0.0409)	(0.0352)
Observations	937,496	422,768	59,600
Within R-squared	0.041	0.008	0.021
Food Expenditure Controls	Yes	Yes	Yes
Demographics/Household Type	Yes	Yes	Yes
Crop-Village FE	Yes	Yes	Yes
Number of Crops	8	8	8

Note: Dependent variable is a dummy variable that takes the value of 1 if household i cultivates crop g. The independent variable is the proportion of households in the origin-state of household i who cultivate crop g (the current state for non migrants). The cultivation question was only asked in round 43 and only about 8 crops; rice, wheat, jowar, bajra, maize, barley, ragi and gram. Hence, there are 8 observations per household. All specifications include crop-village fixed effects and flexible controls for household size, demographics and type as well as a third-order polynomial in log per-capita food expenditure. The column headings denote the various subpopulations on which the regressions are run. Cultivating households are households which report cultivating at least one of the 8 crops. All regressions are weighted using household weights and the standard errors are two-way clustered at the village and origin-state level. * signifies significance at the 10 percent level, ** at the 5 percent level and *** at the 1 percent level.

Table C.7: Comparing the caloric intake of migrants and non-migrants across migration routes: subpopulations and alternative nutrition measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable: Log caloric intake (per person per day)							
Sample:	Baseline Specification	Caloric shortfall	Caloric shortfall >250	Caloric shortfall >500	Below median per capita expenditure	Bottom quartile per capita expenditure	Below median per capita food expend.	Bottom quartile per capita food expend.
$migrant_i$	-0.00564	-0.00485	0.00236	-0.000726	0.00400	0.0273	0.000374	0.0245
	(0.00423)	(0.00543)	(0.00625)	(0.00796)	(0.00757)	(0.0195)	(0.00922)	(0.0176)
$migrant_i$	-0.0267***	-0.0270***	-0.0359***	-0.0304***	-0.0232**	-0.0537***	-0.0234***	-0.0456**
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$	(0.00574)	(0.00697)	(0.00820)	(0.0101)	(0.00907)	(0.0190)	(0.00880)	(0.0191)
Observations	234,157	130,770	95,911	59,457	86,877	36,390	86,378	36,869
Within R-squared	0.731	0.640	0.627	0.622	0.674	0.689	0.643	0.633

	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Dependent variable: Log caloric intake (per person per day)							
Sample:	Children under 5 in household	Children under 5 and below median food expend.	Children under 16 in household	Children under 16 and below median food expend.	Migrated 5 or more years ago	Migrated 10 or more years ago	Migrated 20 or more years ago	Literate
$migrant_i$	0.00221	0.00412	-0.00618	0.000972	-0.00166	0.00380	0.0140***	-0.0116**
	(0.00599)	(0.0116)	(0.00398)	(0.00845)	(0.00401)	(0.00384)	(0.00453)	(0.00500)
$migrant_i$	-0.0350***	-0.0285**	-0.0263***	-0.0229**	-0.0248***	-0.0258***	-0.0306***	-0.0196***
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$	(0.00782)	(0.0128)	(0.00571)	(0.00904)	(0.00600)	(0.00606)	(0.00874)	(0.00751)
Observations	103,325	47,784	177,897	75,507	230,547	227,894	223,220	128,026
Within R-squared	0.699	0.629	0.713	0.637	0.732	0.732	0.732	0.731

	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Dependent Variable:	Caloric shortfall	Caloric shortfall (binary)	Caloric shortfall >250 (binary)	Caloric shortfall >500 (binary)	Caloric shortfall	Caloric shortfall (binary)	Caloric shortfall >250 (binary)	Caloric shortfall >500 (binary)
Sample:	All households				Children under 5 in household			
$migrant_i$	4.373	-0.00335	0.00951	0.00673	-2.491	-0.00421	0.00669	-0.00588
	(4.989)	(0.00502)	(0.00661)	(0.00662)	(7.364)	(0.00881)	(0.0108)	(0.00943)
$migrant_i$	24.35***	0.0173**	0.00932	0.0197***	36.26***	0.00693	0.0102	0.0408***
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$	(6.446)	(0.00835)	(0.0105)	(0.00753)	(10.89)	(0.0145)	(0.0182)	(0.0115)
Observations	234,157	234,157	234,157	234,157	103,325	103,325	103,325	103,325
Within R-squared	0.635	0.397	0.401	0.379	0.627	0.347	0.365	0.360

	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)
Dependent Variable:	Caloric shortfall	Caloric shortfall (binary)	Caloric shortfall >250 (binary)	Caloric shortfall >500 (binary)	Log protein intake (gm. per person per day)	Log fat intake (gm. per person per day)	Log meals per household member per day	Obtain two square meals a day (binary)
Sample:	Children under 16 in household				All households			38th round only
$migrant_i$	4.047	-0.00176	0.00874	0.00534	-0.00346	-0.00298	0.00202	0.00380
	(4.917)	(0.00709)	(0.00754)	(0.00669)	(0.00447)	(0.00484)	(0.00254)	(0.00522)
$migrant_i$	24.12***	0.0161	0.00723	0.0181**	-0.0327***	0.00626	-0.00851**	-0.00960
$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$	(7.374)	(0.0109)	(0.0116)	(0.00873)	(0.00686)	(0.00659)	(0.00428)	(0.00747)
Observations	177,897	177,897	177,897	177,897	234,155	234,152	233,872	109,259
Within R-squared	0.630	0.367	0.380	0.367	0.682	0.745	0.043	0.162

Note: Dependent variable for columns 1-16 is log caloric intake per person per day. Dependent variable for columns 17-32 is detailed in column heading and described in Section 4.2. Independent variables are $migrant_i$, a dummy for whether the household head or their spouse is an interstate migrant, and $migrant_i$ interacted with with an indicator variable that takes the value of 1 if a 1 Rupee reference bundle from the migrant's origin-state provides fewer calories than a 1 Rupee reference bundle from the migrant's destination village both priced at destination-village prices. All specifications include village-round fixed effects and flexible survey-round-specific controls for household size, demographics and type as well as a third-order polynomial in log per-capita food expenditure. The column headings for columns 1-16 and the subheadings for columns 17-32 denote the various subpopulations on which the regressions are run. Caloric shortfall defined relative to the 2400 calorie urban/2100 rural calorie norms. All regressions are weighted using household weights and the standard errors are two-way clustered at the village-round and origin-state level. * signifies significance at the 10 percent level, ** at the 5 percent level and *** at the 1 percent level.

Table C.8: Results broken-down by whether husband or wife is a migrant

Full sample of households		Variable interacted with migrant-structure dummies	Only head is migrant	Only spouse is migrant	Both migrants	No spouse	Observations	Within R-squared
(1)	Dependent variable: Correlation $corr_{io_v}$	$migrant_i^{o_v}$	0.00982**	0.00715***	0.0416***	0.0331***		
			(0.00459)	(0.00267)	(0.00454)	(0.00820)	108,743	0.100
(2)	Dependent variable: Log caloric intake	$migrant_i$	-0.0186***	-0.0103***	-0.0230***	-0.0141**		
			(0.00479)	(0.00323)	(0.00570)	(0.00619)	235,126	0.732
(3)	Dependent variable: Log caloric intake	$migrant_i$	-0.00778	-0.00243	-0.00927	-0.000687		
			(0.00609)	(0.00434)	(0.00674)	(0.00665)		
		$migrant_i$	-0.0252***	-0.0162***	-0.0630***	-0.0358***		
		$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$	(0.00922)	(0.00568)	(0.0145)	(0.0116)	234,155	0.732
Full sample of households		Variable interacted with migrant-structure dummies	Only head is migrant and no mother in house	Only spouse is migrant	Both migrants or migrant head with mother in house	No spouse	Observations	Within R-squared
(4)	Dependent variable: Correlation $corr_{io_v}$	$migrant_i^{o_v}$	0.0123**	0.00713***	0.0388***	0.0330***		
			(0.00483)	(0.00267)	(0.00486)	(0.00819)	108,743	0.100
(5)	Dependent variable: Log caloric intake	$migrant_i$	-0.0141***	-0.0103***	-0.0244***	-0.0142**		
			(0.00476)	(0.00323)	(0.00549)	(0.00617)	235,126	0.732
(6)	Dependent variable: Log caloric intake	$migrant_i$	-0.00403	-0.00244	-0.0102	-0.000702		
			(0.00594)	(0.00434)	(0.00663)	(0.00664)		
		$migrant_i$	-0.0225***	-0.0162***	-0.0618***	-0.0357***		
		$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$	(0.00826)	(0.00567)	(0.0136)	(0.0116)	234,155	0.732

Note: Panels 1-3 repeat the regressions shown in Tables 3, 5 and 7 but interacting every instance of a migrant dummy variable with indicator variables for four mutually-exclusive categories of within-household migrant structure: only the head is a migrant, only the spouse is a migrant, both head and spouse are migrants, and there is no spouse. Each panel comprises one regression. Panel 1 includes village-o_v-round fixed effects, and Panels 2 and 3 include village-round fixed effects. Panels 4-6 repeat the same exercise but replace the only-head indicator with an indicator for households where only the head is a migrant and his mother does not live in the house (with households where the mother is present placed in the both category). All panels include flexible survey-round-specific controls for household size, demographics and type as well as a third-order polynomial in log per-capita food expenditure and a dummy for households where the head has no spouse. All regressions are weighted using household weights and the standard errors are two-way clustered at the village-round and origin-state level. * signifies significance at the 10 percent level, ** at the 5 percent level and *** at the 1 percent level.

Table C.9: Results broken-down by within-household migrant structure: origin-region interactions

Full sample of households		Migrant dummy interacted with migrant-structure and origin-region dummies	Only head is migrant	Only spouse is migrant	Both migrants	No spouse	Observations	Within R-squared
(1)	Dependent variable: Log caloric intake	$migrant_i$	-0.0270***	-0.00209	-0.0247**	-0.0217**		
			(0.00691)	(0.00653)	(0.00996)	(0.0110)		
		$migrant_i$	0.00654	-0.0206*	-0.0484***	-0.0580***		
		$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$	(0.0257)	(0.0115)	(0.0157)	(0.0198)		
		$migrant_i \times Center_o$	0.0194*	-0.0108	0.0175	0.0118		
			(0.0118)	(0.0126)	(0.0143)	(0.0123)		
		$migrant_i \times Center_o$	-0.0365	0.00973	-0.00305	0.0401		
		$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$	(0.0308)	(0.0182)	(0.0325)	(0.0284)		
		$migrant_i \times North_o$	0.0266***	0.00537	0.0237*	0.0393***		
			(0.00990)	(0.00932)	(0.0122)	(0.0112)		
		$migrant_i \times North_o$	-0.0537*	-0.00208	-0.0448	0.0104		
		$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$	(0.0291)	(0.0146)	(0.0275)	(0.0254)		
		$migrant_i \times North\ East_o$	0.0707***	0.0106	-0.00388	-0.0463		
			(0.0221)	(0.0198)	(0.0233)	(0.0316)		
		$migrant_i \times North\ East_o$	-0.0507	0.0366	0.0764*	0.197***		
		$\times \mathbf{1}[\ln K(\boldsymbol{bshare}_i^o, \boldsymbol{P}_i^v) < \ln K(\boldsymbol{bshare}_i^v, \boldsymbol{P}_i^v)]$	(0.0391)	(0.0253)	(0.0392)	(0.0333)	234,155	0.732

Note: Panel 1 repeats the specification in Panel 3 of Table C.8 but further interacting every instance of a migrant dummy variable with indicator variables for the origin-region of the migrant. There are three origin-region interactions, Center, North and North East, with the omitted category being the South of India. Panel 1 comprises one regression and includes village-round fixed effects, flexible survey-round-specific controls for household size, demographics and type as well as a third-order polynomial in log per-capita food expenditure and a dummy for households where the head has no spouse. Regression weighted using household weights and the standard errors are two-way clustered at the village-round and origin-state level. * signifies significance at the 10 percent level, ** at the 5 percent level and *** at the 1 percent level.