Trade, Tastes and Nutrition in India

David Atkin

Online Appendix

B Model Extensions

B.1 An Analytically Tractable Model with Iso-Elastic Utility

For analytical tractability, I choose three simple functional forms for preferences, habit formation and technology that satisfy the assumptions of my model. These simple functional forms allow the no-comparative-advantage-reversal threshold in definition 1 to be expressed in terms of exogenous parameters, and sharp predictions to be made about the caloric impacts of marginal reductions in trade costs (proposition 2 and example 1).

The period utility function is

$$U(c_{rt}, c_{wt}, \theta_t) = \theta_t u(c_{rt}) + (1-\theta_t)u(c_{wt}), \tag{9}$$

where $u(c_{gt})$ is given by the isoelastic utility function

$$u(c_{gt}) = \frac{c_{gt}^{1-\frac{1}{\epsilon}}}{1-\frac{1}{\epsilon}} \text{ if } \epsilon \neq 1 \text{ and } \epsilon > 0,$$

$$u(c_{gt}) = \ln c_{gt} \text{ if } \epsilon = 1.$$

This utility function implies that both the relative consumption of rice and the budget share spent on rice are independent of food expenditure:

$$\frac{c_{rt}}{c_{wt}} = p_t^{-\epsilon}\left(\frac{\theta_t}{1-\theta_t}\right)^{\epsilon},$$

$$s_t = \frac{1}{1+p_t^{\epsilon-1}(\frac{1-\theta_t}{\theta_t})^{\epsilon}}$$

The current taste stock depends on past consumption through the following relationship:

$$\theta_{t+1} = \frac{1}{1+(\frac{c_{rt}}{c_{wt}})^{-\nu}},\ \theta_1 = \frac{1}{2}.$$

Accordingly, $h(csh_t;\nu) = \frac{csh_t^{\nu}}{csh_t^{\nu}+(1-csh_t)^{\nu}}$ using the caloric share definition of habit formation in assumption 3.

I assume production functions take the Cobb-Douglas form where the two production technologies are equally labor intensive:

$$Q_{rt} = L_{rt}^{1-\alpha} V_r^{\alpha},$$

$$Q_{wt} = L_{wt}^{1-\alpha} V_w^{\alpha},$$

$$0 < \alpha < 1.$$

Now profit maximization and market clearing leads to the following labor allocation, wages (ω_t) and land rental prices (π_{rt} and π_{wt}):

$$(L_{rt}, L_{wt}) = \left(\frac{1}{1 + p_t^{-\frac{1}{\alpha}} \frac{V_w}{V_r}} L, \frac{p_t^{-\frac{1}{\alpha}} \frac{V_w}{V_r}}{1 + p_t^{-\frac{1}{\alpha}} \frac{V_w}{V_r}} L \right), \tag{10}$$

$$\omega_t = p_{wt}(1-\alpha) \left(V_w + V_r p_t^{\frac{1}{\alpha}} \right)^{\alpha} L^{-\alpha} = p_{rt}(1-\alpha) \left(V_r + V_w p_t^{-\frac{1}{\alpha}} \right)^{\alpha} L^{-\alpha}, \tag{11}$$

$$\pi_{rt} = p_{rt}\alpha(V_r + p_t^{-\frac{1}{\alpha}} V_w)^{\alpha-1} L^{1-\alpha},\ \pi_{wt} = p_{wt}\alpha(V_w + V_r p_t^{\frac{1}{\alpha}})^{\alpha-1} L^{1-\alpha}. \tag{12}$$

Relative production of rice is a function of prices and relative factor endowments:

$$\frac{Q_{rt}}{Q_{wt}} = p_t^{\frac{1-\alpha}{\alpha}} \frac{V_r}{V_w}.$$

Lemma 1 implies that if the economy is trading,

$$p_t = \frac{1}{\tau}, p_t^* = \tau.$$

Region H will only trade in period 1 if τ is sufficiently low that at this price, relative consumption of rice is lower than relative production:

$$p_t^{-\epsilon}(\frac{\theta_1}{1-\theta_1})^{\epsilon} < p_t^{\frac{1-\alpha}{\alpha}} \frac{V_r}{V_w},$$

$$\tau < (\frac{V_r}{V_w}(\frac{\theta_1}{1-\theta_1})^{-\epsilon})^{\frac{1}{\gamma+\epsilon}} = (\frac{V_r}{V_w})^{\frac{1}{\gamma+\epsilon}},$$

where $\gamma = \frac{1-\alpha}{\alpha} > 0$. Otherwise, region H will be in autarky, and market clearing implies that the relative consumption and production of rice are equal. Autarky value are denoted with an A superscript:

$$p_t^{A-\epsilon}(\frac{\theta_t}{1-\theta_t})^{\epsilon} = p_t^{A\frac{1-\alpha}{\alpha}} \frac{V_r}{V_w},$$

$$p_t^A = (\frac{V_w}{V_r})^{\frac{1}{\gamma+\epsilon}} (\frac{\theta_t}{1-\theta_t})^{\frac{\epsilon}{\gamma+\epsilon}}.$$

Now that prices in any pre-liberalization period are known, I study the equation of motion for tastes:

$$\theta_{t+1} = \frac{1}{1 + (p_t^\epsilon(\frac{\theta_t}{1-\theta_t})^{-\epsilon})^\nu}. \tag{13}$$

First I deal with the case where the two regions are initially trading in period 1:

$$\theta_{t+1} = f(\theta_t) = \frac{1}{1 + \frac{1}{\tau}^{\epsilon\nu}(\frac{\theta_t}{1-\theta_t})^{-\epsilon\nu}}, \tag{14}$$

$$\theta_s = \frac{1}{1 + \frac{1}{\tau}^{\frac{\epsilon\nu}{1-\epsilon\nu}}},$$

where the subscript s denotes the steady state. As $f'(\theta_s) = \epsilon\nu$, this steady state is stable if $f'(\theta_s) < 1$ or $\nu < \frac{1}{\epsilon}$. Therefore, regional tastes will converge towards θ_s from below. The steady state will remain a trading equilibria where rice is exported as long as $\tau < (\frac{V_r}{V_w})^{\frac{1}{\frac{\epsilon}{(1-\epsilon\nu)}+\gamma}}$.

The second possibility is that the region is initially in autarky in period 1 ($\tau \geq (\frac{V_r}{V_w})^{\frac{1}{\gamma+\epsilon}}$,):

$$\theta_{t+1} = f(\theta_t) = \frac{1}{1 + (\frac{V_w}{V_r})^{\frac{\epsilon\nu}{\gamma+\epsilon}}(\frac{\theta_t}{1-\theta_t})^{-\frac{\gamma\epsilon\nu}{\gamma+\epsilon}}}, \tag{15}$$

$$\theta_s^A = \frac{1}{1 + (\frac{V_w}{V_r})^{(\frac{\epsilon\nu}{\epsilon+\gamma(1-\epsilon\nu)})}}.$$

As $f'(\theta_s^A) = \frac{\gamma\epsilon\nu}{\gamma+\epsilon}$, this steady state is stable if $f'(\theta_s^A) < 1$ or $\nu - \gamma^{-1} < \frac{1}{\epsilon}$. The autarky price at the steady state is:

$$p_s^A = (\frac{V_w}{V_r})^{\frac{1}{\frac{\epsilon}{(1-\epsilon\nu)}+\gamma}}.$$

It remains to show that the steady state θ_s is globally stable in both the trade and autarky cases. From the equations of motion for tastes, equations 14 and 15,

$$f'(\theta_t) = \frac{c(b\frac{1-\theta_t}{\theta_t})^{\epsilon\nu-1}}{\left(1 + (b\frac{1-\theta_t}{\theta_t})^c\right)^2}[\frac{1}{\theta^2}b] > 0\,\forall\theta_t \in (0,1),$$

where b and c are positive constants.[1] Therefore, for all $\theta_t \in (0, \theta_s)$,

$$\theta_{t+1} - \theta_s = f(\theta_t) - \theta_s = -\int_{\theta_t}^{\theta_s} f'(\theta_t)d\theta < 0.$$

In the trade case, if $v\epsilon < 1$, it holds that:

$$\frac{\theta_{t+1} - \theta_t}{\theta_t} = \frac{\frac{1}{1+\frac{1}{\tau}^{\epsilon\nu}(\frac{\theta_t}{1-\theta_t})^{-\epsilon\nu}} - \theta_t}{\theta_t} = \frac{1}{\theta_t(1 + \frac{1}{\tau}^{\epsilon\nu}(\frac{\theta_t}{1-\theta_t})^{-\epsilon\nu})} - 1 > 0$$

where the inequality comes from $\theta_s = \frac{1}{1+\frac{1}{\tau}^{\frac{\epsilon v}{1-\epsilon v}}} > \theta_t$. Thus for all $\theta_t \in (0, \theta_s)$, $\theta_{t+1} \in (\theta_t, \theta_s)$. The exact same logic can be applied to show that for all $\theta_t \in (\theta_s, 1)$, $\theta_{t+1} \in (\theta_s, \theta_t)$ in the trade case.

[1] For the trade case, $b = \frac{1}{\tau}$, $c = \epsilon\nu$, while for the autarky case $b = (\frac{V_w}{V_r})^{\frac{1}{\gamma}}$, $c = \frac{\epsilon\nu}{1+\frac{\epsilon}{\gamma}}$.

In the autarky case, if $\nu - \frac{1}{\gamma} < \frac{1}{\epsilon}$, it holds that:

$$\frac{\theta_{t+1} - \theta_t}{\theta_t} = \frac{\frac{1}{1+(\frac{V_w}{V_r})^{\frac{\epsilon\nu}{\gamma+\epsilon}}(\frac{\theta_t}{1-\theta_t})^{-\frac{\gamma\epsilon\nu}{\gamma+\epsilon}}} - \theta_t}{\theta_t} = \frac{1}{\theta_t(1+(\frac{V_w}{V_r})^{\frac{\epsilon\nu}{\gamma+\epsilon}}(\frac{\theta_t}{1-\theta_t})^{-\frac{\gamma\epsilon\nu}{\gamma+\epsilon}})} - 1 > 0$$

where the inequality comes from $\theta_s = \frac{1}{1+(\frac{V_w}{V_r})^{(\frac{\epsilon\nu}{\epsilon+\gamma(1-\epsilon\nu)})}} > \theta_t$. Thus for all $\theta_t \in (0, \theta_s)$, $\theta_{t+1} \in (\theta_t, \theta_s)$. The exact same logic can be applied to show that for all $\theta_t \in (\theta_s, 1)$, $\theta_{t+1} \in (\theta_s, \theta_t)$ in the autarky case. Combining the results shows that the steady state in the autarky case is globally stable.

As long as the steady states are stable ($\nu < \frac{1}{\epsilon}$), the various possible outcomes are neatly summarized in the phase diagram shown in figure 3. If transport costs in period 1 are less than or equal to $(\frac{V_r}{V_w})^{\frac{1}{\frac{\epsilon}{(1-\epsilon\nu)}+\gamma}}$, tastes for rice will rise over time, converging towards the steady state value θ_s. If $(\frac{V_r}{V_w})^{\frac{1}{\frac{\epsilon}{(1-\epsilon\nu)}+\gamma}} < \tau \leq (\frac{V_r}{V_w})^{\frac{1}{\epsilon+\gamma}}$, the two regions trade in period 1. However as θ_t rises through habit formation, region H has less desire to import rice and the two regions enter autarky, with tastes eventually converging to θ_s^A. Finally, if $\tau > (\frac{V_r}{V_w})^{\frac{1}{\epsilon+\gamma}}$, high transport costs choke off trade in period 1 and all subsequent periods prior to liberalization with tastes converging towards θ_s^A.

The simple functional forms now allow me to revisit definition 1, proposition 2 and example 1 and make more precise statements.

B.1.1 Conditions for No Comparative Advantage Reversal (Definition 1)

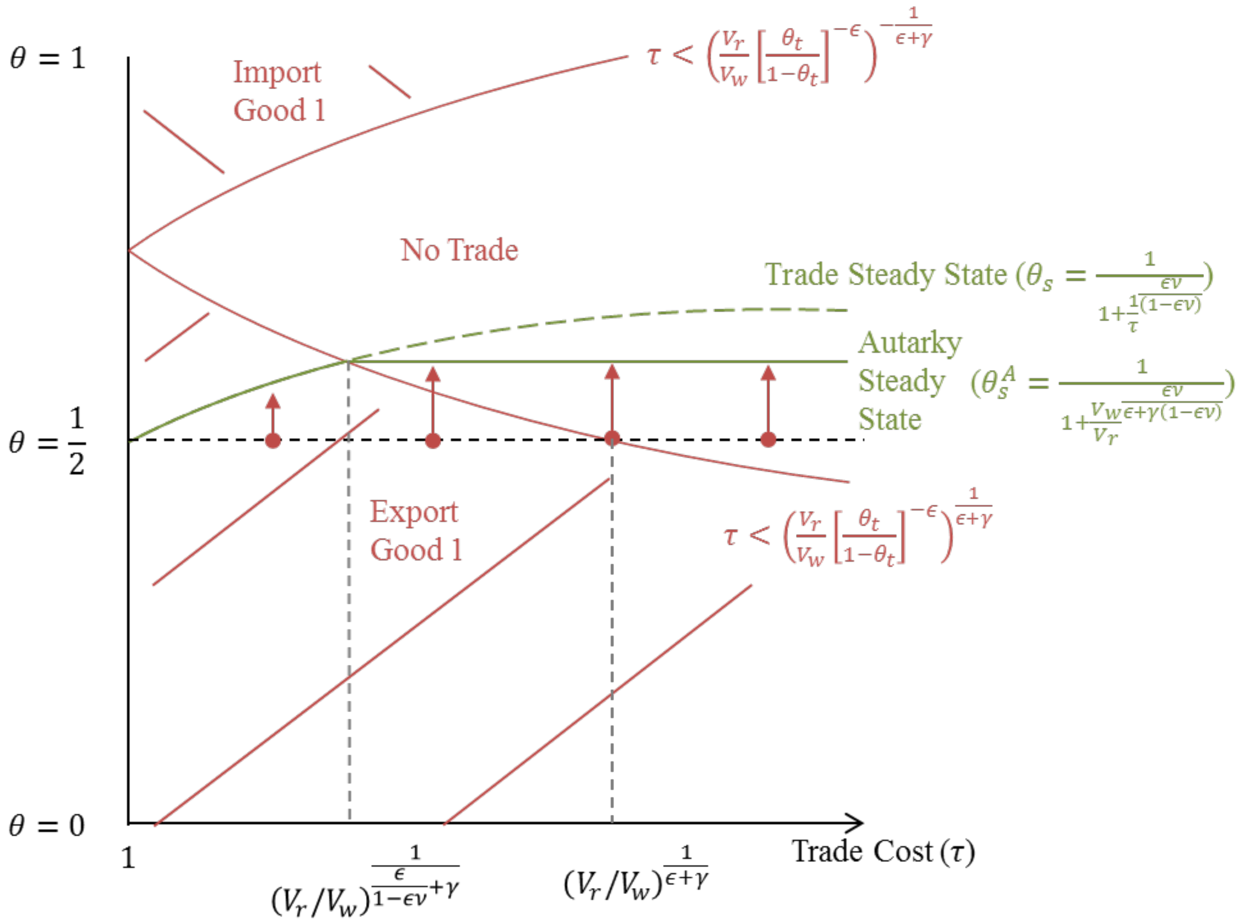
Region H maintains its comparative advantage if ν is less than the smallest value $\tilde{\nu}$ at which the steady state autarky price is equal to 1 (definition 1).

$$p_s^A = (\frac{V_w}{V_r})^{\frac{1}{\frac{\epsilon}{(1-\epsilon\tilde{\nu})}+\gamma}} = 1,$$
$$\tilde{\nu} = \frac{1}{\epsilon}.$$

A necessary and sufficient condition for $p_s^A < 1$ is that $\nu < \frac{1}{\epsilon}$. This condition is also sufficient to ensure stability of the steady state under either autarky or trade and corresponds to habits not being so strong that they overpower the love of variety that is indexed by $\frac{1}{\epsilon}$ in the isoelastic utility.[2]

[2] p_s^A is greater than 1 if $\frac{1}{\epsilon} < \nu < \frac{1}{\epsilon} + \frac{1}{\gamma}$ (the region for which the autarky steady state is stable but the trade steady state is not). p_s^A equal 1 once more if $\nu = \frac{1}{\epsilon} + \frac{1}{\gamma}$ but this steady state is never stable.

Figure 3: The Evolution of Tastes



B.1.2 Conditions for the Wealth Effect to be Greater than the Reallocation Effect (Proposition 2)

I now evaluate the impact of a marginal reduction in trade costs that occurs after many periods when the region is arbitrarily close to its steady state value of tastes, θ_s. I focus on the case where the region is trading in period T as otherwise a marginal reduction in trade costs will have no effect.

The combination of proposition 1 and proposition 2 implies that the immediate caloric change from a marginal reduction in trade costs in period T will be decreasing with the strength of habit formation if the increase in the size of the wealth effect necessarily dominates any changes in the reallocation effect.

This condition can be easily evaluated in case where trade costs are sufficiently low that both regions trade at the steady state. The change in caloric intake with a marginal reduction in τ is:

$$\frac{dK_s}{K_s}/\frac{-d\tau}{\tau}=\frac{dm_s}{m_s}/\frac{-d\tau}{\tau}-\frac{(1-\epsilon)\tau^{\frac{-(\epsilon\nu-\epsilon-1)}{\epsilon\nu-1}}+\tau^{-1}+\epsilon\tau^{\frac{\epsilon}{\epsilon\nu-1}}}{(\tau^{-1}+\tau^{\frac{\epsilon}{\epsilon\nu-1}}+\tau^{\frac{-(\epsilon\nu-\epsilon-1)}{\epsilon\nu-1}}+\tau^{\frac{2\epsilon}{\epsilon\nu-1}})}, \tag{16}$$

where $\frac{dm_s}{m_s}/\frac{-d\tau}{\tau}=\frac{\tau^{-\frac{1}{\alpha}}V_r}{\tau^{-\frac{1}{\alpha}}V_r+V_w}$ is the factor income effect if factors are evenly distributed across the population. The factor income effect is independent of the strength of habits. Hence, $\frac{dK_s}{K_s}/\frac{d\tau}{\tau}$ is smaller in the presence of habit formation, $\frac{d\frac{dK_s}{K_s}/\frac{-d\tau}{\tau}}{d\nu}<0$, if:

$$\epsilon<\tilde{\epsilon}\equiv\frac{1+\tau^{2\frac{\epsilon}{(1-\epsilon\nu)}}+2\tau^{\frac{\epsilon}{(1-\epsilon\nu)}}}{1+\tau^{2\frac{\epsilon}{(1-\epsilon\nu)}}-\tau-\tau^{2\frac{\epsilon}{(1-\epsilon\nu)}-1}},$$

where $\tilde{\epsilon}>1$.

B.1.3 Absolute Caloric Losses for Laborers with Habit Formation (Example 1)

I assume that some proportion of the population own only l units of labor each, and receive only wage income $\omega_t l$. I will now show that it is possible for these "landless laborers" to maintain their caloric intake, $K_t^L=\omega_t l(\frac{s_t+p_t(1-s_t)}{p_t})$, with trade liberalization in the absence of habit formation, yet suffer caloric losses at the time of trade liberalization in the presence of habit formation.

Once more, I consider a region that is trading in period T and tastes have converged to their steady state values, $\theta_s=\frac{1}{1+\frac{1}{\tau}^{\frac{\epsilon\nu}{1-\epsilon\nu}}}$. As preferences are homothetic, the expression for $\frac{dK_t^L}{K_t^L}/\frac{-d\tau}{\tau}$ is identical to equation 16 except I must replace $\frac{dm_s}{m_s}/\frac{-d\tau}{\tau}$ with $\frac{d\omega_s}{\omega_s}/\frac{-d\tau}{\tau}$. However, with since production is Cobb-Douglas, labors' share of income is constant and hence $\frac{d\omega_s}{\omega_s}/\frac{-d\tau}{\tau}=\frac{dm_s}{m_s}/\frac{-d\tau}{\tau}$. Accordingly, $\frac{dK_t^L}{K_t^L}/\frac{-d\tau}{\tau}=0$ if

$$\frac{V_w}{V_r}=\tau^{-\gamma-\frac{\epsilon}{(1-\epsilon\nu)}}\frac{1+\tau^{\frac{\epsilon}{(1-\epsilon\nu)}}-\tau^{\frac{\epsilon}{(1-\epsilon\nu)}-1}\epsilon(\tau-1)}{1+\tau^{\frac{\epsilon}{(1-\epsilon\nu)}}+\epsilon(\tau-1)}. \tag{17}$$

Regions will always trade at the steady state if $\frac{V_w}{V_r}<\tau^{-(\gamma+\frac{\epsilon}{(1-\epsilon\nu)})}$. Hence, the endowment ratio at which the gains from trade are zero is always a valid trading steady state.

Therefore, for parameter sets that satisfy $\epsilon<\tilde{\epsilon}$, the caloric change at the time of trade liberalization is declining with the strength of habits, $\frac{d(\frac{dK_T}{K_T}/\frac{-d\tau}{\tau})}{d\nu}<0$. If equation 17 holds for $\nu=0$, $\frac{dK_T}{K_T}/\frac{-d\tau}{\tau}=0$, laborers will see no change in caloric intake with trade liberalization in the abscence of habits. For parametrizations that satisfy both of these conditions, laborers must see an absolute decline in caloric intake at the time of trade liberalization in the presence of habit formation but would have seen no such loss in the absence of habits.

For example, take the parameter set where $\tau = \frac{3}{2}, \epsilon = \frac{1}{2}$ and $\gamma = \frac{1-\alpha}{\alpha} = 1$. Clearly $\epsilon < \tilde{\epsilon}$ in this case, and equation 17 reduces to $\frac{V_w}{V_r} = \frac{\frac{3}{2}^{-\frac{3}{2}}(1+\frac{3}{2}^{\frac{1}{2}}-\frac{3}{2}^{-\frac{1}{2}}\frac{1}{2}(\frac{3}{2}-1))}{1+\frac{3}{2}^{\frac{1}{2}}+\frac{1}{2}(\frac{3}{2}-1)} = \frac{4}{9}$ if $\nu = 0$ (no habits). This relative endowment satisfies the condition for a trading equilibrium, $\frac{V_w}{V_r} < (\frac{3}{2})^{-\frac{3}{2}}$. Thus, for these parameter values the caloric gains from trade liberalization are zero for landless laborers in the absence of habits, and negative with habits.

B.2 A Forward-Looking Two-Period Model

In this section I show that my main results carry through to a two-period model of forward-looking dynastic households in two always-trading regions. Additionally, I simulate a variant of the model where I allow for many periods and the regions to be in autarky prior to trade liberalization. The simulation highlights the fact that the magnitude of the reduction in the caloric gains from trade in the presence of habit formation can actually grow larger if households are forward-looking.

The two-period analysis proceeds as follows. I rule out the possibility that forward-looking dynastic households adjust their consumption in anticipation of a forthcoming trade liberalization in such a way that habit formation no longer leads to preferences at the time of liberalization that favor the endowment-comparative-advantage food (lemma 4). Therefore, a discrete (habits versus no habits) analogue of proposition 1 still holds in the presence of forward looking adults. Once more, I must rule out that habit formation is so strong that the resource comparative advantage is reversed (the threshold is defined in definition 2). As long as this easily verifiable scenario does not occur, I can provide a discrete analogue to lemma 2: habit formation reduces the change in the caloric intake at the time of full trade liberalization.

I consider a two-period model with forward-looking dynastic households. In the first period there are iceberg trade costs $\tau > 1$ and I restrict attention to the case where these trade costs are not sufficiently large to choke off trade entirely. I focus on the most extreme case, where a trade liberalization is expected with probability 1 an instant after the start of period 2.[3] As in the myopic case, I explore the effect of habit formation in altering the caloric change at the time of trade liberalization for the households alive at that time. I compare the caloric intake for period 2 households at the very start of that period when $\tau > 1$ with the caloric intake for period 2 households after a full trade liberalization when $\tau' = 1$. Period 2 subscripts on quantities, prices and incomes will indicate values after the trade liberalization.

[3] For example, if the trade liberalization is expected to occur only at the end of period 2 or with a very low probability, then households will be less likely to adjust their consumption in period 1 in order to benefit from the new free-trade relative prices.

A dynastic household comes into being at the beginning of the first period with unbiased tastes, $\theta_1 = \frac{1}{2}$. The household's consumption in period 1 determines the tastes of the household in period 2 through the process of habit formation. The first period household is forward looking and cares about the welfare of the second period household with some discount factor $0 \leq \beta \leq 1$.[4] Period 1 households solve the following problem:

$$\max_{c_{r1}, c_{w1}, c_{r2}, c_{w2}} \quad U(c_{r1}, c_{w1}, \theta_1) + \beta U(c_{r2}, c_{w2}, \theta_2)$$

subject to:

$$\theta_2 = h(\frac{c_{r1}}{c_{w1} + c_{r1}}, \nu),$$

$$p_1 c_{r1} + c_{w1} = m_1,$$

$$p_2 c_{r2} + c_{w2} = m_2.$$

Lemma 4. *Consider a symmetric two-period model with two always-trading regions and forward-looking dynastic households satisfying assumptions 1-3 and A1-A2. Households anticipate a reduction in trade costs from $\tau > 1$ to $\tau' = 1$ occurring an instant after the start of period 2. In the presence of habit formation ($\nu > 0$), it can never be optimal for households in region H (where $V_r > V_w$) to choose a bundle (c_{r1}, c_{w1}) such that $c_{w1} \geq c_{r1}$.*

Proof. By contradiction. I will deal separately with the cases $c_{w1} = c_{r1}$ and $c_{w1} > c_{r1}$.

First, take any optimal bundle in period 1, (c_{r1}, c_{w1}), such that $c_{w1} = c_{r1}$. This implies that $\theta_2 = h(\frac{c_{r1}}{c_{w1}+c_{r1}}, \nu) = \frac{1}{2}$ by assumption 3. The first order condition for period 2 households is $p_2 \frac{\partial U(c_{r2}, c_{w2}, \theta_2)}{\partial c_{w2}} = \frac{\partial U(c_{r2}, c_{w2}, \theta_2)}{\partial c_{r2}}$ which implies that $c_{r2} = c_{w2}$ as $p_2 = \frac{1}{\tau'} = 1$ for two symmetric regions trading freely. The five first-order conditions for period 1 households can be combined

[4]This dynastic household model perfectly maps into a 2 period variant of the model presented in the main paper in two scenarios: (1) Forward-looking parents derive utility from their child's current utility and believe (either correctly or incorrectly) that their newborn children possess the same preferences that they do. (2) Adults only care about their own utility and the utility of their adult children, yet all household consumption is shared between adults and children. An alternative strategy would be to assume that children are born with unbiased preferences and parents know this. Even if parents prepare separate meals (and presumably this is costly), there will still be a trade off between feeding them more food today and providing them with "better" preferences in future.

into the following expression:

$$\frac{\partial U(c_{r1}, c_{w1}, \theta_1)/\partial c_{r1}}{\partial U(c_{r1}, c_{w1}, \theta_1)/\partial c_{w1}} - p_1 = -\beta \frac{\partial U(c_{r2}, c_{w2}, \theta_2)}{\partial \theta_2} \frac{\partial h(\frac{c_{r1}}{c_{w1}+c_{r1}}, \nu)/\partial \frac{c_{r1}}{c_{w1}}}{\partial U(c_{r1}, c_{w1}, \theta_1)/\partial c_{w1}} \frac{1}{c_{w1}} (\frac{c_{r1}}{c_{w1}} p_1 + 1).$$

The symmetry of the utility function implies that $\frac{\partial U(c_{r2}, c_{w2}, \theta_2)}{\partial \theta_2} = 0$ when $\theta_2 = \frac{1}{2}$ and $c_{r2} = c_{w2}$. Therefore, $c_{w1} = c_{r1}$ would only be chosen if $p_1 = 1$. This outcome cannot be a symmetric equilibrium for both regions, as it would imply that both regions would have no reason to trade. Since relative rice demand is equal to 1 with p_1=1 and $\theta_1 = \frac{1}{2}$ due to assumption 2, the relative rice supply must also equal 1, $x(1, V_r, V_w, L) = 1$. However, $V_w > V_r$ implies that $x(1, V_r, V_w, L) > 1$ and so $c_{w1} \neq c_{r1}$.

Second, take any optimal bundle in period 1, (c_{r1}, c_{w1}), such that $c_{w1} > c_{r1}$. This implies $\theta_2 = h(\frac{c_{r1}}{c_{w1}+c_{r1}}, \nu) = \frac{1}{2} - \epsilon$ where $\epsilon > 0$, and utility $U(c_{r1}, c_{w1}, \frac{1}{2})$. The first order condition in period 2 is $p_2 \frac{\partial U(c_{r2}, c_{w2}, \theta_2)}{\partial c_{w2}} = \frac{\partial U(c_{r2}, c_{w2}, \theta_2)}{\partial c_{r2}}$. By assumptions 1-2 and A1-A2 and the fact that $p_2 = \frac{1}{\tau'} = 1$, the optimal bundle in period 2, (c_{r2}, c_{w2}), must be such that $c_{w2} > c_{r2}$. By non-satiation $p_1 c_{r1} + c_{w1} = m_1$ and $p_2 c_{r2} + c_{w2} = m_2$. Now consider the bundle $(\tilde{c_{r1}}, \tilde{c_{w1}})$ with $\tilde{c_{r1}} = c_{w1}$ and $\tilde{c_{w1}} = c_{r1}$. By the "symmetry" of the habit formation function this bundle implies $\tilde{\theta}_2 = \frac{1}{2} + \epsilon$. By symmetry of the utility function and $p_2 = 1$ this implies a bundle $(\tilde{c_{r2}}, \tilde{c_{w2}})$, such that $\tilde{c_{r2}} = c_{w2}$ and $\tilde{c_{w2}} = c_{r2}$. Thus, the utility derived from the two bundles is identical: $U(c_{r2}, c_{w2}, \theta_2) = U(\tilde{c_{r2}}, \tilde{c_{w2}}, \tilde{\theta}_2)$. If $c_{w1} > c_{r1}$, then $\frac{c_{r1}}{c_{w1}} < 1 < x(1, V_r, V_w, L)$ and region H must export rice with sufficiently low τ, implying that $p_1 = \frac{1}{\tau}$. At this price, it is always true that $p_1 \tilde{c_{r1}} + \tilde{c_{w1}} < m_1$ if $p_1 c_{r1} + c_{w1} = m_1$. Therefore, (c_{r1}, c_{w1}) such that $c_{w1} > c_{r1}$ cannot be an optimal bundle for an individually rational consumer as consumers could obtain higher utility by purchasing $(\tilde{c_{r1}} + \varepsilon, \tilde{c_{w1}} + \varepsilon)$ which is affordable in period 1 with $p_1 = \frac{1}{\tau}$.

There is one final case that must be ruled out. I showed above that $(\tilde{c_{r1}}+\varepsilon, \tilde{c_{w1}}+\varepsilon)$ is preferred and affordable if all other consumers are choosing (c_{r1}, c_{w1}) and so $p_1 = \frac{1}{\tau}$. However, consumers may anticipate that if everyone chooses the bundle $(\tilde{c_{r1}}+\varepsilon, \tilde{c_{w1}}+\varepsilon)$, region H may import rice and $\tilde{p_1} = \tau$ (this price change will occur only if $\frac{\tilde{c_{r1}}}{\tilde{c_{w1}}} > x(1, V_r, V_w, L)$). As incomes are functions of prices and endowments, $m_1 = m(p_{r1}, p_{w1}, V_r, V_w, L)$, the bundle $(\tilde{c_{r1}}+\varepsilon, \tilde{c_{w1}}+\varepsilon)$ is only affordable in this scenario if $\tau \tilde{c_{r1}} + \tilde{c_{w1}} < m(\tau, 1, .)$. As $\tau \tilde{c_{r1}} + \tilde{c_{w1}} = \tau c_{w1} + c_{r1} = \tau m(\frac{1}{\tau}, 1, .)$, a sufficient condition for (c_{r1}, c_{w1}) to not be an optimal bundle even if consumers anticipate their aggregate effects is that $\tau m(\frac{1}{\tau}, 1, .) < m(\tau, 1, .)$. This inequality can be rewritten as $m(\tau, 1, .) > m(1, \tau, .)$ as Lm_t is revenue and the revenue (or GDP) function is homogenous of degree 1 in prices. The inequality must hold as $\frac{dm(p_r, p_w, .)}{dp_r} = \frac{Q_r}{L}$, $\frac{dm(p_r, p_w, .)}{dp_w} = \frac{Q_w}{L}$ and $x(1, V_r, V_w, L) = \frac{Q_r}{Q_w} > 1$ when $V_r > V_w$. Hence, (c_{r1}, c_{w1}) where $c_{w1} > c_{r1}$ cannot be an optimal bundle. □

Lemma 4 implies that a discrete (habits versus no habits) analogue of proposition 1 still holds in the presence of forward looking households:

Proposition 3. *Consider a symmetric two-period model with two always trading regions and forward-looking dynastic households satisfying assumptions 1-3 and A1-A2. Households anticipate a reduction in trade costs from $\tau > 1$ to $\tau' = 1$ occurring an instant after the start of period 2. Habit formation ($\nu > 0$) raises household tastes for the endowment-comparative-advantage food in period 2: $\theta_2|_{\nu>0} > \theta_2|_{\nu=0}$ if $V_r > V_w$ and $\theta_2|_{\nu>0} < \theta_2|_{\nu=0}$ if $V_r < V_w$.*

Proof. Lemma 4 implies that $c_{r1} > c_{w1}$. Therefore, $\theta_2 > \theta_1 = \frac{1}{2}$ if $\nu > 0$ and $\theta_2 = \theta_1 = \frac{1}{2}$ if $\nu = 0$ from the definition of habit formation (assumption 3). □

As in the model with myopic households, habit formation can be so strong that a region's resource-driven comparative advantage is reversed at the start of the second period. Additionally, as in the rational addiction literature, forward-looking households may actually wish to increase their relative consumption of the locally-abundant food in the first period in the presence of habit formation (see section B.2.1 for a simulated example where this occurs). Therefore, I restrict attention to the empirically verifiable case where a region maintains its resource comparative advantage at the time of trade liberalization.

Definition 2. No-Comparative-Advantage-Reversal Threshold: Define $\check{\nu}_t$ as the smallest ν at which relative rice supply equals relative rice demand in period t with a relative price of $p_t = \frac{1}{\tau}$: For period 1, $x(\frac{1}{\tau}, V_r, V_w, L) = \check{y}_1$, where $\check{y}_1 = y_1(\frac{1}{\tau}, m(\frac{1}{\tau}, .), \frac{1}{2}, \frac{1}{\tau}, m(\frac{1}{\tau}, .), \check{\theta}_2)$ is the relative consumption chosen by forward-looking households in period 1 and $\check{\theta}_2 = h(\check{y}_1, \check{\nu}_1)$. For period 2, $x(\frac{1}{\tau}, V_r, V_w, L) = y(\frac{1}{\tau}, m(\frac{1}{\tau}, .), \check{\theta}_2)$ where $\check{\theta}_2 = h(\check{y}_1, \check{\nu}_2)$ and $\check{y}_1$ is defined as previously.

If assumption 2 is satisfied, region H, for whom $V_r > V_w$, maintains the rice comparative advantage derived from its endowments in both period 1 and 2. Hence, combining proposition 3 with the assumption that habits are moderate (weaker than the threshold defined in definition 2) produces implication 4, $(\theta_2 - \theta_2^*)(p_2 - p_2^*) < 0$ iff $0 < \nu < min(\check{\nu}_1, \check{\nu}_2)$.

Implication 6. *Assume that assumptions 1-3 and A1-A2 hold, and that households anticipate a reduction in trade costs from $\tau > 1$ to $\tau' = 1$ occurring an instant after the start of period 2. Household tastes at the start of period 2 are biased towards the food for which a region has a relatively low price compared to the other region, $(\theta_2 - \theta_2^*)(p_2 - p_2^*) < 0$, iff $0 < \nu < min(\check{\nu}_1, \check{\nu}_2)$.*

Proposition 3 is sufficient for the proof of a discrete (habits versus no habits) analogue of lemma 2 for the two-trading-region subcase.

Lemma 5. *Consider a symmetric two-period model with two always-trading regions and forward-looking dynastic households satisfying assumptions assumptions 1-3 and A1-A2, and $\nu < min(\check{\nu}_1, \check{\nu}_2)$. For any period 2 household, the proportional gain in the caloric intake, $\frac{K_2' - K_2}{K_2}$, that accompanies a reduction in trade costs from $\tau > 1$ to $\tau' = 1$ will be smaller in the presence of habit formation: $\frac{K_2' - K_2}{K_2}|_{\nu=0} > \frac{K_2' - K_2}{K_2}|_{\nu>0}$.*

Proof. I will show that caloric intake at the start of period 2 and prior to liberalization, K_2, is larger under habit formation, $K_2|_{\nu>0} > K_2|_{\nu=0}$. The symmetry of the model implies that the post-liberalization price $p_2' = \frac{1}{\tau'} = 1$. As the price per calorie is 1 for both foods, caloric intake post liberalization, $K_2' = m(p_2', .)$, is independent of habits, $K_2'|_{\nu>0} = K_2'|_{\nu=0}$. Therefore, $\frac{K_2' - K_2}{K_2}|_{\nu=0} > \frac{K_2' - K_2}{K_2}|_{\nu>0}$.

As the two regions trade at the start of period 2, $p_2 = \frac{1}{\tau}$ by the assumption that $\nu < min(\check{\nu}_1, \check{\nu}_2)$ and definition 2. The total caloric intake for any household at the start of period 2 is equal to $K_2 = \frac{s_2}{p_2} m(p_2, .) + (1 - s_2) m(p_2, .)$. $K_2|_{\nu>0} - K_2|_{\nu=0} = (s(\theta_2|_{\nu>0}, p_2, m(p_2, .)) - s(\theta_2|_{\nu=0}, p_2, m(p_2, .)))(\frac{m(p_2,.)}{p_2} - m(p_2, .)) > 0$ by proposition 1 and assumption 1 (and assumption A1 if preferences are non-homothetic).

□

B.2.1 A Simulated Multi-period Model with Forward Looking Consumers

I now extend the two-period model above to a multi-period case and solve it numerically. This analysis provides a deeper understanding of the effects of forward-looking consumers in the model and highlights the fact that the reduction in the caloric gains from trade in the presence of habit formation can be even larger in this framework.

I explore the case of two regions that are initially in autarky prior to trade liberalization. I assume that the economy moves from autarky to free trade shortly after the start of period $T = 9$ and that households have perfect foresight (e.g. this liberalization is perfectly anticipated).

Households solve the following infinite horizon problem:

$$\max_{\{c_{rt},c_{wt}\}_{t=1}^{\infty}} \sum_{t=0}^{\infty} \beta^t U(c_{rt}, c_{wt}, \theta_t)$$

subject to:

$$\theta_{t+1} = h(\frac{c_{rt}}{c_{rt} + c_{wt}}, \nu),$$

$$p_t c_{rt} + c_{wt} = m_t \forall t.$$

I impose the isoelastic utility function and a Cobb Douglas production function both described in Appendix B.1 and solve the model with a shooting algorithm in Matlab.

I assume that land is distributed equally across all households such that m_t is the income of a representative agent in period t,

$$m_t = \omega_t + \frac{\pi_{rt} V_r}{L} + \frac{\pi_{wt} V_w}{L}.$$

I set $L = 1$ and normalize $p_{wt} = 1$ in appendix equations 11 and 12:

$$m_t = (1-\alpha)\left(V_w + V_r p_t^{\frac{1}{\alpha}}\right)^{\alpha} + p_t \alpha (V_r + p_t^{-\frac{1}{\alpha}} V_w)^{\alpha-1} + \alpha (V_w + V_r p_t^{\frac{1}{\alpha}})^{\alpha-1}.$$

Since $x_t = c_{rt}/c_{wt}$ in autarky and there are no savings, the budget constraint implies:

$$(c_{rt}, c_{wt}) = \left(\frac{x_t m_t}{1 + p_t x_t}, \frac{m_t}{1 + p_t x_t}\right).$$

The dynamic problem is therefore:

$$\max \frac{1}{1-\frac{1}{e}} \sum_{t=0}^{\infty} \beta^t \left(\frac{m_t}{1+p_t x_t} \right)^{1-\frac{1}{e}} \left(\theta_t x_t^{1-\frac{1}{e}} + (1-\theta_t) \right)$$

$$\text{subject to:}\, \theta_{t+1} = h(\frac{1}{1+x_t^{-1}}, \nu),\ m_t \text{defined as above} \, \forall t.$$

I solve the above problem numerically with the values $\alpha = 1/2$, $\beta = 0.95$, $\sigma = 4$, $\nu = 0, 0.11., \ldots, 0.89, 1$, $V_r = 4$, $V_w = 1$. Figure 4 shows the results for relative rice prices, tastes, utility and caloric intake at the beginning of each period. Figure 5 compares the proportional utility and caloric gains at the instant of trade liberalization in period $T = 9$ for a range of different habit strengths in both the myopia and the perfect foresight models.

The caloric gains from trade are decreasing in the strength of habits (panel 2 of figure 5) in the perfect foresight case. In fact, the decline in the caloric gains with the strength of habits is larger in the perfect foresight case. As in the rational addiction literature (Becker and Murphy, 1988), households anticipate that consumption will be biased towards rice in future periods and so wish to further raise the tastes for rice to make this inevitable outcome more enjoyable. Hence, tastes for rice are higher prior to trade liberalization, pre-trade caloric intake is larger, and the caloric gain with trade liberalization is smaller.

Figure 4: Perfectly anticipated trade liberalization

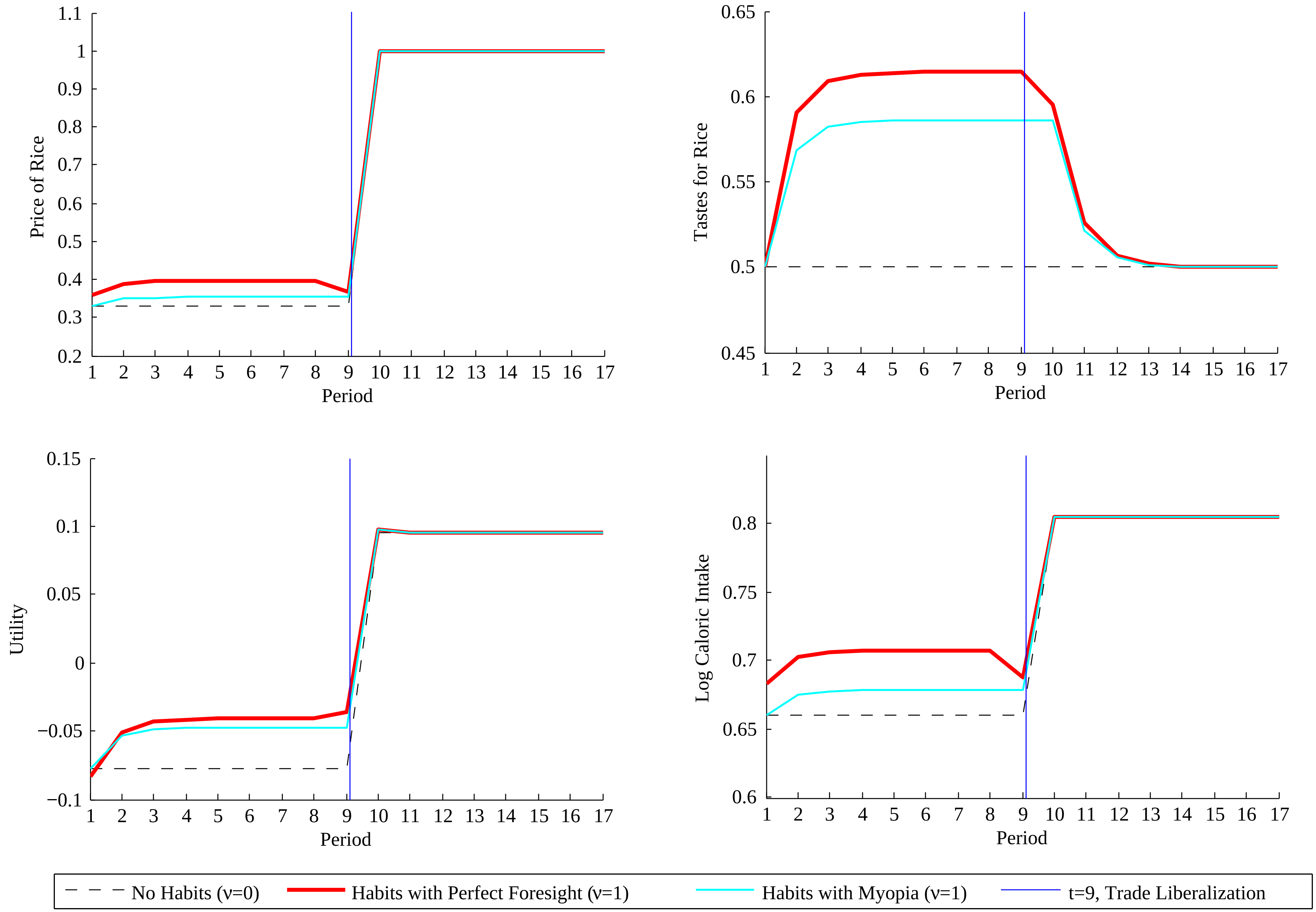
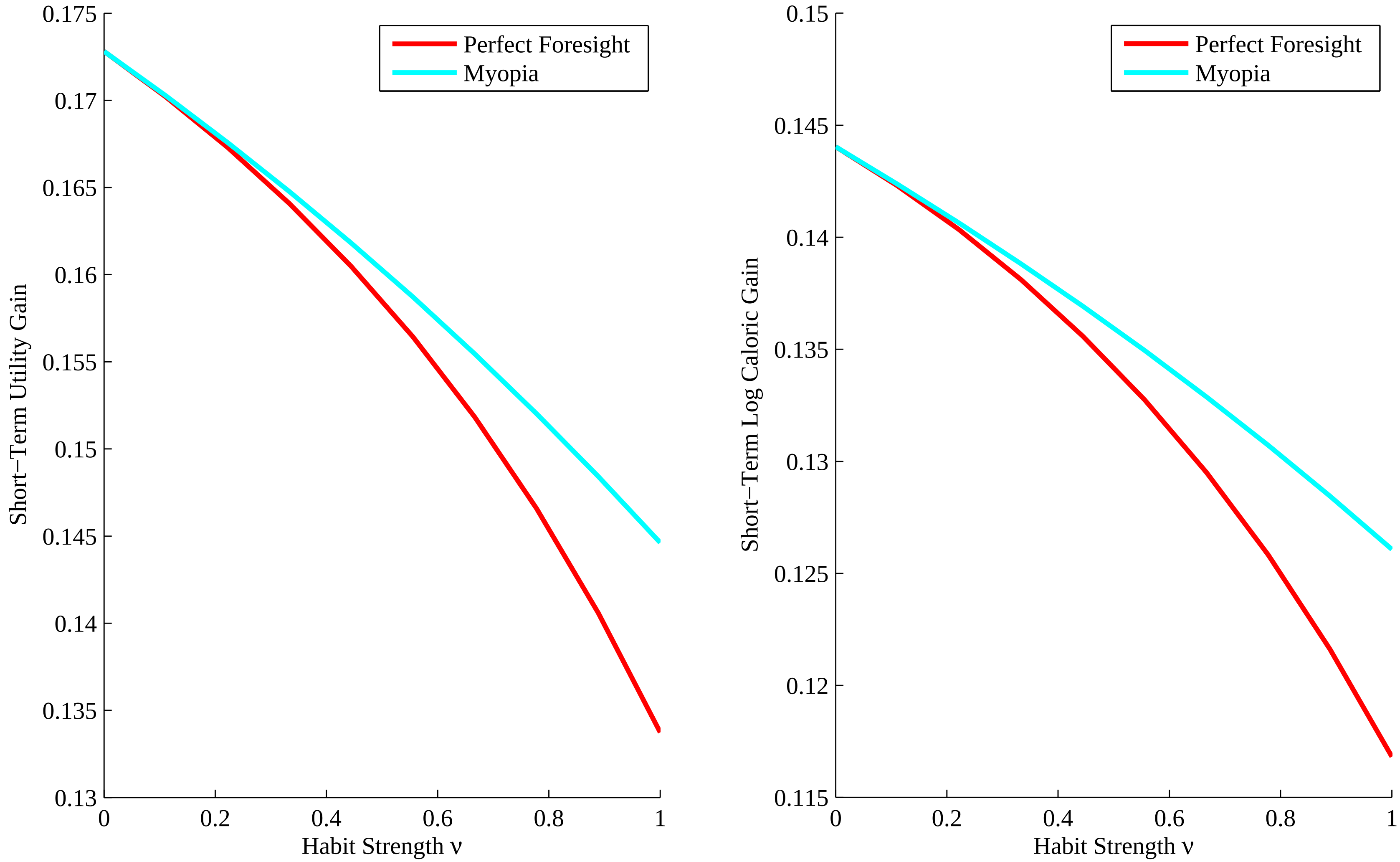


Figure 5: Habit strength and the utility and caloric gains from trade



B.3 A Simulated Model Using AIDS Demands, Asymmetric Countries and Initially Biased Preferences

In this appendix, I simulate a two-period, two-region model using the very general AIDS demand structure that I use in the empirical section. I do not impose symmetry between rice and wheat in the expenditure function. In particular, I use the Indian survey data to parametrize the AIDS utility function, which results in a universal bias towards rice, and non-homothetic demand. Additionally, I allow for two types of agent in the economy, laborers, L, and landowners, V, and allow endowments to be asymmetric across the two regions.

I explore the case of a reduction in trade costs from $\tau > 1$ to $\tau = 1$ in a particular year during period 2 for two regions that are initially trading in period 1. I find that even when there are initially biased preferences and non-homotheticities, the caloric gains from trade are decreasing in the strength of habits.

Rice demand for agent $i = \{L, V\}$ come from the Linear Approximate AIDS used in the empirical analysis:

$$s_{it} = \theta_t + \gamma_{11} \ln p_{rt} + \gamma_{12} \ln p_{wt} + b \ln m_{it} - b(s_{it} \ln p_{rt} + (1 - s_{it}) \ln p_{wt}).$$

I use the Indian survey data from 1987-1988 and the same techniques as in section 3.3 for estimating tastes, except now in a two-good specification. I find the following parameter values using my IV strategy: $\hat{b} = -0.014$, $\hat{\gamma_{11}} = -0.355$ and $\hat{\gamma_{12}} = 0.355$.

I assume production functions take the Cobb-Douglas form where the two production technologies are equally labor intensive and $0 < \alpha < 1$:

$$Q_{rt} = L_{rt}^{1-\alpha} V_{rt}^{\alpha},$$

$$Q_{wt} = L_{wt}^{1-\alpha} V_{wt}^{\alpha},$$

Vollrath (2011) reviews the estimates of the labor share in agriculture and reports values of between 0.35 and 0.40 for wheat and an average share of around 0.55 for rice. In order to abstract from technological differences across crops, I assume that these labor shares are equal across the two crops and choose a value of $\alpha = 0.5$.

With trade and region H importing wheat, world relative supply has to equal world relative demand:

$$
\begin{aligned}
c_{wt} + c^*_{wt} =& Q_{wt} + c^*_{wt} + \frac{(Q^*_{wt} - c^*_{wt})}{\tau}, \\
c_{rt} + c^*_{rt} =& Q^*_{rt} + c_{rt} + \frac{(Q_{rt} - c_{rt})}{\tau}.
\end{aligned}
$$

I assume the following function for habit formation in the presence of an initial universal rice bias, $\wp \in (0,1)$:

$$
\theta_{t+1,i} = \frac{1}{1 + (y_{t,i})^{-\nu}\wp},
$$

where $y_{t,i}$ is the relative rice consumption for group $i = \{L, V\}$ in period t. I set the bias $\wp$ such that in the case without habit formation, $\nu = 0$, tastes for rice are equal to the initial tastes θ_1. Therefore, $\wp = \frac{1}{\theta_1} - 1$. In the symmetric case without an initial bias, $\theta_1 = \frac{1}{2}$ and $\wp = 1$.

I now turn to the distribution of factors in the region. I assume everyone owns one unit of labor. Some fraction of the population μ own all the land in the region, equally divided up among the μL landowners. Therefore,

$$
\begin{aligned}
y_t =& \left(p_t(\frac{1}{\phi_L s_{tL} + \phi_T s_{tT}} - 1)\right)^{-1}, \\
\phi_L =& \frac{(1-\mu)L\omega_{it}}{(L\omega_{it} + V_r\pi_{rt} + V_w\pi_{wt})},\ \phi_T = \frac{\mu L\omega_{it} + V_r\pi_{rt} + V_w\pi_{wt}}{(L\omega_{it} + V_r\pi_{rt} + V_w\pi_{wt})},
\end{aligned}
$$

where ϕ_L is landless labor's share of total income and ϕ_T is the landowners share. Around 44 percent of households in rural areas report their primary occupation as being self employed in agriculture, and so I set $\mu = 0.44$ in both regions for my simulations.

I simulate a trade liberalization episode by lowering the iceberg trade costs during period 2. I choose an initial value of τ such that both regions are initially trading, and restrict attention to scenarios where habit formation is not so strong such that comparative advantage is reversed. This setup mimics the situation in India, where regions' locally-abundant foods are relatively cheap and where costly trade currently occurs. I explore the caloric gains from trade that result from a reduction in iceberg trade costs from $\tau = 1.25$ to $\tau = 1$, free trade.

Figure 6 shows the results of the simulation for several sets of land endowments and initial foods biases. I plot the relationship between the strength of habits, ν, and both the tastes for rice in period 2 and the short-run caloric gains from trade liberalization. Due to the non-homothetic demands, landless laborers and landowners are affected differently, and results for both of these groups are shown.[5]

Panel 1 of figure 6 shows the baseline case, where endowments are symmetric and tastes are initially unbiased ($\theta_1 = 0.50,\ L = 1,\ V_r = 1.5,\ V_w = 0.5,\ \theta_1^* = 0.50,\ L^* = 1,\ V_r^* = 0.5,\ V_w^* = 1.5$). Panel 2 of figure 6 allows for an initial bias towards rice in both countries ($\theta_1 = \theta_1^* = 0.60$). Panel 3 of figure 6 allows for an initially asymmetric distribution of endowments, with region H having a more extreme allocation of its abundant land ($V_r = 1.75,\ V_w = 0.25,\ V_r^* = 0.75,\ V_w^* = 1.25$).

For both landless labor and landowners in both regions and all cases, proposition 1 and lemma 2 are satisfied: habit formation raises tastes for the endowment-comparative-advantage food ($\frac{d\theta_2}{d\nu} > 0$ where $V_r > V_w$, and $\frac{d\theta_2^*}{d\nu} < 0$ where $V_r^* < V_w^*$), and habit formation reduces the caloric gains from trade ($d(\frac{K_T' - K_T}{K_T})/d\nu < 0$). As I assume no-comparative advantage reversal (by restricting the range of ν allowed in the simulation), proposition 4 is necessarily satisfied.

In these parametrizations, rice is an inferior good in both regions ($b < 0$). Therefore, landless laborers in both regions consume relatively more rice in period 1, and have relatively stronger tastes for rice in period 2, than do landowners. As region H has a comparative advantage in rice, the relative rice price rises with trade in region H and the caloric gains from trade are smaller for landless laborers than they are for landowners. In region F, the reverse is true.[6] This is most clearly seen in the first panel of figure 6, where endowments are symmetric and tastes are initially unbiased.

As preferences are no longer perfectly symmetric between rice and wheat, the world price with free trade is no longer equal to 1. Therefore, habit formation shrinks both the production and consumption side gains from trade, as was the case in the move from autarky to free trade in my theoretical model.

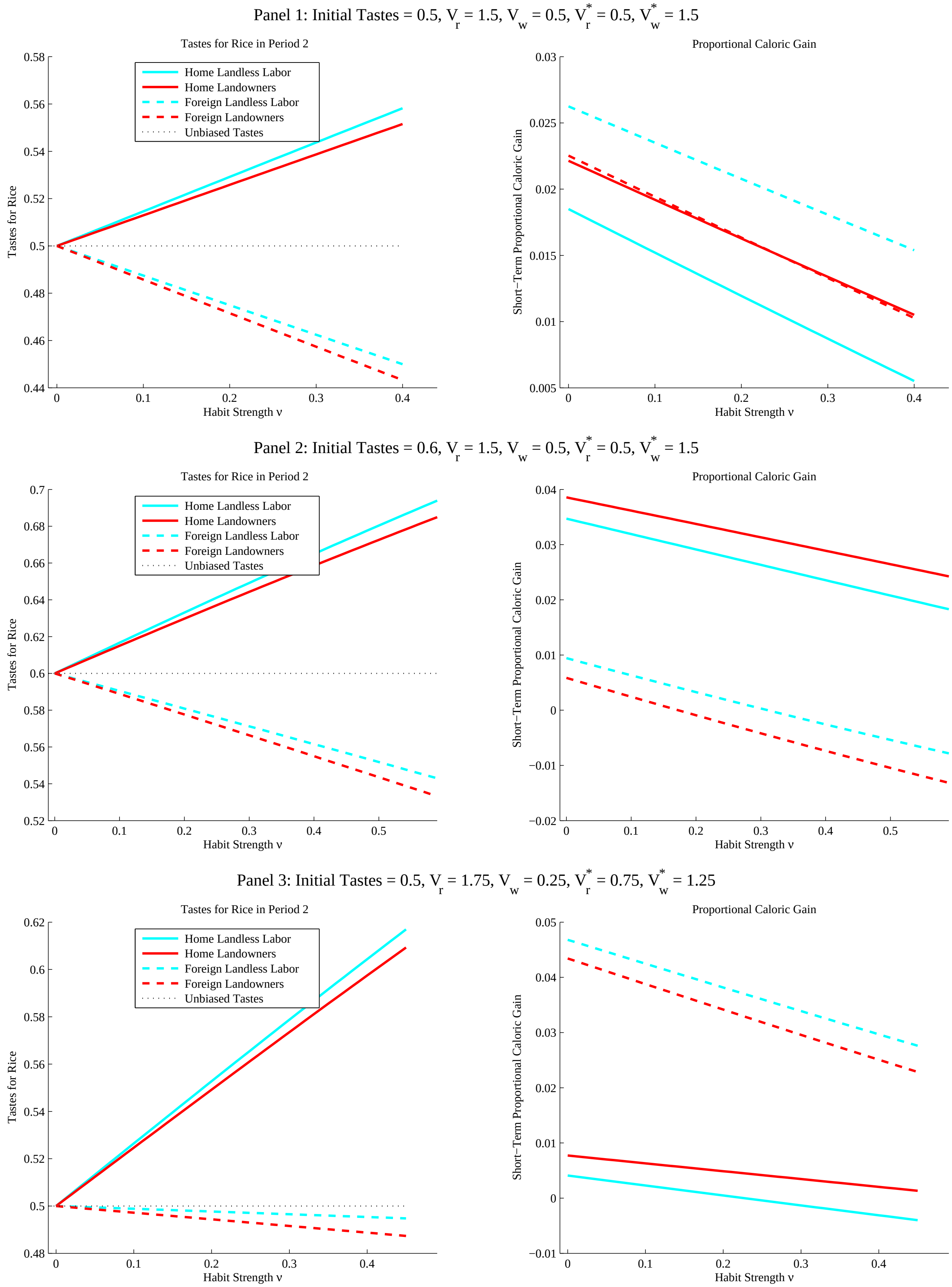
Even with asymmetric endowments, initial biases in tastes, and the non symmetric demand structure implied by the AIDS, the negative relationship between the strength of habits and

[5]The short-run caloric gains are equal to the proportional change in caloric intake for period 2 households at the time of trade liberalization.

[6]If habit formation alters the income elasticity of demand, as is possible in my theoretical model, landless laborers in both regions could experience smaller caloric gains from trade compared to landowners.

the caloric gains from trade remains (lemma 2). With sufficiently strong habit formation, both groups in region F suffer caloric losses at the time of trade liberalization in the case of initially biased tastes (panel 2 of figure 6) and landless laborers in region H suffer caloric losses at the time of trade liberalization in the case of asymmetric endowments (panel 3 of figure 6).

Figure 6: Asymmetric Endowments, Universal Taste Biases and AIDS Demands



B.4 A Multi-Good Small-Region Model

In this section, I briefly lay out a small-region variant of the main model that more closely matches the multi-region empirical exercise. As is well known, with N+1 factors and N goods (a many-good specific factors model) it is not possible to establish clear systematic relationship between factor endowments and comparative advantage (Dixit and Norman, 1980). Therefore, this stripped-down version of the model is agnostic about the source of comparative advantage. I simply take as given that some foods are exported and others imported. However, in section B.4.1, I reintroduce the symmetry assumptions that I relax in this section and provide a chain of three correlation-like results which are strongly suggestive of a multi-good analog of implication 3 that links endowments and tastes at the time of trade liberalization.

As in the model detailed in the main text, I assume that there are iceberg trade costs τ and these costs are reduced during a trade liberalization episode in period $T > 1$. Unlike the model in the main text, I assume that the region is small in the sense that the aggregate choices of households in the region have no effect on the vector world prices $\mathbf{p}_t^W$, comprised of prices p_{gt}^W for each food g, where W superscripts denote world values. The vector of domestic prices, $\mathbf{p}_t$, contains prices p_{gt}.

I assume perfect competition and constant returns to scale. Domestic output of good g, $Q_{gt} = Q_g(\mathbf{p}_t, \mathbf{A})$, is a function of prices and some vector of fixed technologies and endowments $\mathbf{A}$, which are shared equally amongst identical households and differ from the world vector $\mathbf{A}^W$ in the following sense: $\mathbf{A}^W \neq \psi \mathbf{A}$ where ψ is any scalar. I assume that output of good g is increasing in the price of good g and decreasing in the price of all other goods, $\frac{\partial Q_g}{\partial p_g} > 0$ and $\frac{\partial Q_g}{\partial p_{g'}} < 0 \; \forall g' \neq g$.

I make the following preference assumptions which are multi-good analogues of assumptions 1-2 and A1-A2:

Assumption 1*: Higher relative tastes for good g raise the proportional increase in expenditure required to maintain utility u_t with a rise in price p_{gt}: for $g \neq g'$, $\frac{\partial^2 \ln e(u_t, \mathbf{p}_t; \Theta_t)}{\partial \ln p_{gt} \partial \theta_{gt}} > 0$ and $\frac{\partial^2 \ln e(u_t, \mathbf{p}_t; \Theta_t)}{\partial \ln p_{gt} \partial \theta_{g't}} = 0$ where Θ_t is a vector of θ_{gt} tastes and $\sum_g \theta_{gt} = 1$.

Assumption 2*: Adult tastes in generation 1 are identical to world tastes in generation 1:

$\Theta_1 = \Theta_1^W$, where Θ_t^W is a vector of θ_{gt}^W tastes.

Assumption A1*: $\frac{\partial^2 \ln e(u,p;\theta)}{\partial \ln p \partial u}$ must be close enough to zero such that $s_g(u_t, \mathbf{p}_t, \Theta) > s_g(u_t, \mathbf{p}_t, \tilde{\Theta})$ implies $s_g(v(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta), \mathbf{p}_t, \Theta) > s_g(v(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \tilde{\Theta}), \mathbf{p}_t, \tilde{\Theta})$, $s_g(u_t, \mathbf{p}_t, \Theta) < s_g(u_t, \mathbf{p}_t, \tilde{\Theta})$ implies $s_g(v(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta), \mathbf{p}_t, \Theta) < s_g(v(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \tilde{\Theta}), \mathbf{p}_t, \tilde{\Theta})$ and $s_g(u_t, \mathbf{p}_t, \Theta) = s_g(u_t, \mathbf{p}_t, \tilde{\Theta})$ implies $s_g(v(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta), \mathbf{p}_t, \Theta) = s_g(v(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \tilde{\Theta}), \mathbf{p}_t, \tilde{\Theta})$, where s_g is the budget share spent on good g.

Assumption A2*: Every good is a strict gross substitute for each other, $\frac{dc_{gt}}{dp_{gt}} < 0$ and $\frac{dc_{gt}}{dp_{g't}} > 0$ if $g \neq g'$, where $c_{gt} = c_g(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta_t)$ is the consumption of good g.

Three comments are in order. First, the AIDS expenditure function used in the paper clearly satisfies assumption 1* as $\frac{\partial \ln e(u_t, \mathbf{p}_t; \Theta_t)}{\partial \ln p_{gt}} = \theta_{gt} + \sum_{g'} \gamma_{gg'} \ln p_{g't} + \beta_g u_t \beta_0 \prod_{g'} p_{g'}^{\beta_{g'}}$. Second, in assumption 2* I no longer restrict all foods to be equally favored by period 1 households. Similarly, the expenditure function need not be symmetric in every food. Therefore, I allow global biases towards certain foods. For example every household in the world can have stronger tastes for meat compared to rice in period 1. Third, the main propositions are still likely to carry through in the aggregate without assumptions A1* and A2*, however, sharp general equilibrium predictions are not possible without restrictions of this type. In the empirical analysis, I do not force income elasticities to be small or goods to be gross substitutes and still find support for the main propositions.

Assumption 3*. ***Habit Formation:*** *adult tastes for food g are increasing with the relative consumption of food g,* $Y_{gt} \equiv \frac{c_{gt}}{\overline{c_g}} = Y_g(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta_t, \overline{c_g})$, *as a child,*

$$\theta_{gt+1} = h_g(Y_{gt}; \nu), \quad \textit{with } \frac{\partial h_g(Y_{gt}; \nu)}{\partial Y_{gt}} \geq 0 \textit{ and } \frac{\partial^2 h_g(Y_{gt}; \nu)}{\partial Y_{gt} \partial \nu} > 0, \tag{18}$$

where $\nu \geq 0$ *parametrizes the strength of habit formation,* $\overline{c_g} = c_g(\mathbf{p}_1^W, m(\mathbf{p}_1^W, \mathbf{A}), \Theta_1)$, *and* $h_g(Y_{gt}; 0) = h_g(1; \nu) = \theta_{g1}$.

Implication 1*. *Under assumptions 1*, 3* and A1*-A2*, rice tastes decrease with past rice prices,* $\frac{d\theta_{gt}}{dp_{gt-n}} < 0 \; \forall n > 0$, *iff* $\nu > 0$.

Proof. Fixing $p_{g,t'} \; \forall t' \neq t-n$, $\frac{d\theta_{g,t}}{dp_{g,t-n}} = \frac{d\theta_{g,t}}{dc_{g,t-1}} \frac{dc_{g,t-1}}{d\theta_{g,t-1}} \ldots \frac{d\theta_{g,t-n+1}}{dc_{g,t-n}} \frac{dc_{g,t-n}}{dp_{g,t-n}} < 0$ if $\nu > 0$ and $\frac{d\theta_{g,t}}{dp_{g,t-n}} = \frac{d\theta_{g,t}}{dc_{g,t-1}} \frac{dc_{g,t-1}}{d\theta_{g,t-1}} \ldots \frac{d\theta_{g,t-n+1}}{dc_{g,t-n}} \frac{dc_{g,t-n}}{dp_{g,t-n}} = 0$ if $\nu = 0$ from assumptions 1*, 3*, A1* and A2*. □

Implication 2*. *Under assumptions 1*, 3* and A1*-A2*, the budget share spent on good g,*

$s(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta_t, \overline{c_g})$, depends on past prices for good g' after conditioning on current prices and incomes, $\frac{ds(\mathbf{p}_t,m(\mathbf{p}_t,\mathbf{A}),\Theta_t,\overline{c_g})}{dp_{g't-n}} = \frac{\partial s(\mathbf{p}_t,m(\mathbf{p}_t,\mathbf{A}),\Theta_t,\overline{c_g})}{\partial \theta_{gt}} \frac{d\theta_{gt}}{dp_{g't-n}} \neq 0 \ \forall n > 0$, g and g', iff $\nu > 0$.

Proof. $\frac{ds(\mathbf{p}_t,m(\mathbf{p}_t,\mathbf{A}),\Theta_t,\overline{c_g})}{dp_{g't-n}} = \frac{\partial s(\mathbf{p}_t,m(\mathbf{p}_t,\mathbf{A}),\Theta_t,\overline{c_g})}{\partial \theta_{gt}} \frac{d\theta_{gt}}{dp_{g't-n}}$ from assumption A1*. $\frac{\partial s(\mathbf{p}_t,m(\mathbf{p}_t,\mathbf{A}),\Theta_t,\overline{c_g})}{\partial \theta_{gt}} > 0$ from 1* and A1*. $\frac{d\theta_{gt}}{dp_{gt-n}} < 0 \ \forall n > 0$ if $\nu > 0$ and $\frac{d\theta_{gt}}{dp_{gt-n}} = 0 \ \forall n > 0$ if $\nu = 0$ from the proof of implication 1*. Finally, $\frac{d\theta_{gt}}{dp_{g't-n}} < 0 \ \forall n > 0$ if $\nu > 0$ and $\frac{d\theta_{gt}}{dp_{g't-n}} = 0 \ \forall n > 0$ if $\nu = 0$ and $g \neq g'$ following the same steps as in the proof of implication 1* but replacing $p_{g,t-n}$ by $p_{g',t-n}$. □

In the original assumption 3, the tastes for rice were increasing in the share of rice in the total consumption of calories, $\frac{c_{rt}}{c_{rt}+c_{wt}}$, with adult tastes unbiased ($\theta_t = \frac{1}{2}$) if the childhood consumption of rice and wheat was exactly equal. As assumption 2* admits intrinsic global biases in tastes and there are now many foods, I require a more general definition of habit formation. I assume that adult tastes depend on Y_{gt}, the consumption of good g as a child relative to a level of "unbiased" consumption for that food $\overline{c_g}$. Adult tastes for a food remain unbiased in the presence of habit formation only if the adult consumed precisely this level of consumption as a child. As richer regions will consume more of most foods, I require a $\overline{c_g}$ that is region specific and hence depends on A to ensure that tastes for food g remain relative preference measures.[7]

The proofs of implication 4* and implication 5* require a choice of $(\overline{c_g}, \overline{c_g}^W)$ such that the relative consumption of good g, $\frac{c_{gt}}{\overline{c_g}}$, is larger in the location where that food is cheaper, $p_{gt} \lesseqgtr p_{gt}^W \implies \frac{c_g(\mathbf{p}_t,m(\mathbf{p}_t,\mathbf{A}),\Theta_t)}{\overline{c_g}} \gtreqless \frac{c_g(\mathbf{p}_t^W,m(\mathbf{p}_t^W,\mathbf{A}^W),\Theta_t)}{\overline{c_g}^W}$ for any Θ_t.[8] I choose to define $(\overline{c_g}, \overline{c_g}^W)$ so

[7]An alternative approach is for tastes to depend on the past caloric share, $\theta_{gt+1} = h_g(\frac{c_{gt}}{\sum_{g'} c_{g't}}; \nu)$ with $h_g(\overline{c_g}; \nu) = \theta_{g1}$. In this case $\overline{c_g}$ would not need to depend on A. However, with more than two foods the gross substitutes property would not be sufficient to ensure that lemma 7* holds and stronger restrictions on preferences would be needed.

[8]To see implication 4* under this assumption note that $\frac{c_g(\mathbf{p}_1,m(\mathbf{p}_1,\mathbf{A}),\Theta_1)}{\overline{c_g}} > \frac{c_g(\mathbf{p}_1^W,m(\mathbf{p}_1^W,\mathbf{A}^W),\Theta_1)}{\overline{c_g}^W}$ if $p_{g1} < p_{g1}^W$ hence $\theta_{g2} > \theta_{g2}^W$ by assumption 3*. Similarly $\frac{c_g(\mathbf{p}_2,m(\mathbf{p}_2,\mathbf{A}),\Theta_2)}{\overline{c_g}} > \frac{c_g(\mathbf{p}_2^W,m(\mathbf{p}_2^W,\mathbf{A}^W),\Theta_2)}{\overline{c_g}^W} > \frac{c_g(\mathbf{p}_2^W,m(\mathbf{p}_2^W,\mathbf{A}^W),\Theta_2^W)}{\overline{c_g}^W}$ if $p_{g2} < p_{g2}^W$ and $\theta_{g2} > \theta_{g2}^W$ because of assumption A1*. Following this logic, $\theta_{gT} > \theta_{gT}^W$ if there is no comparative advantage reversal (habits are weaker than the threshold defined in definition 3*). The proof of implication 5* exactly follows the proofs below except that in this scenario, world tastes in period T are not equal to world tastes in the absence of habit formation ($\Theta_T^W \neq \Theta_1^W$). Therefore, although the wealth effect with habit formation will be smaller than if domestic tastes were equal to world tastes ($\Theta_T = \Theta_T^W$), it may be larger than if there was no habit formation ($\Theta_T = \Theta_1^W$) for any particular small region. On aggregate however, the wealth effect is likely to be decline with habit formation. For example, imagine that unbiased tastes for a food are very low. However, as the food is globally abundant, tastes for that food are above the unbiased tastes in every region. Therefore, in the presence of habit formation the wealth effect declines for small exporting regions and increases for small importing regions through this channel. In contrast, the fact that tastes are higher for a particular food in the

that this condition holds without any further assumptions on preferences.

I set $\overline{c_g}$ to be equal to the quantity demanded by domestic consumers with endowments/technologies $\mathbf{A}$ in the presence of generation 1 (unbiased) tastes $\Theta_1 = \Theta_1^W$ and some price vector. In particular, I pick the world price vector prevailing in generation 1, and define $\overline{c_g} = c_g(\mathbf{p}_1^W, m(\mathbf{p}_1^W, \mathbf{A}), \Theta_1)$ and $\overline{c_g}^W = c_g(\mathbf{p}_1^W, m(\mathbf{p}_1^W, \mathbf{A}^W), \Theta_1^W)$. The choice of the world price vector is driven in part by convenience, as it ensures that the world is always at a steady state, with world prices constant in every period, simplifying the small-region analysis.

Lemma 6*. *Assumption A3* implies that world prices and the tastes of world households are fixed and unbiased in all periods,* $\Theta_1^W = \Theta_t^W$ *and* $\mathbf{p}_t^W = \mathbf{p}_1^W$ $\forall t$.

Proof. As $c_{g1}^W = c_g(\mathbf{p}_1^W, m(\mathbf{p}_1^W, \mathbf{A}^W), \Theta_1^W) = \overline{c_g}^W$, assumption 3* implies that $\theta_{g2}^W = \theta_{g1}^W$ for all goods. Therefore, the world economy in period 2 is identical to that in period 1 and $p_{g2}^W = p_{g1}^W$ for all goods. By the same logic, $\Theta_t^W = \Theta_1^W$ and $\mathbf{p}_t^W = \mathbf{p}_1^W$ $\forall t$. □

The world economy is at a steady state. Such an outcome naturally arises from the two large region model in the main text. In that model, if the two regions trade freely for many generations, tastes eventually return to their unbiased values and remain there indefinitely.[9]

Without loss of generality, I partition foods in the home economy into three groups: Export foods g^X for which the domestic price in period 1, $p_{g^X 1}$, is sufficiently less than the world price such that the food is exported, $p_{g^X 1} = \frac{1}{\tau} p_{g^X 1}^W$. Import foods g^M for which the domestic price $p_{g^M 1}$ is sufficiently greater than the world price such that the food is imported, $p_{g^M 1} = \tau p_{g^M 1}^W$. Non-traded foods g^N for which $\frac{1}{\tau} p_{g^N 1}^W < p_{g^N 1} < \tau p_{g^N 1}^W$ and there is no trade.

For simplicity, I restrict the analysis to focusing on economies where all goods are traded.

Assumption 4*. ***All Goods Traded****: I assume that the vectors of endowments and technologies* $\mathbf{A}$ *and* $\mathbf{A}^W$ *are sufficiently different and that* $\tau > 1$ *is sufficiently low such that the set of foods where there is no trade, the set of* g *for which* $\frac{1}{\tau} p_{gt}^W < p_{gt} < \tau p_{gt}^W$*, is empty.*

Assumption 4* implies that the domestic price vector in period 1 comprises g^X goods priced

regions that export that food compared to the regions that import it reduces the wealth effect in every region.

[9] As the world relative price post trade is 1, eventually rice tastes will return to $\theta = \frac{1}{2}$ in both regions.

at $\frac{1}{\tau}p^W_{g^X 1}$ and g^M goods priced at $\tau p^W_{g^M 1}$, $\mathbf{p}_1 = \begin{bmatrix} \mathbf{p}_1^X \\ \mathbf{p}_1^M \end{bmatrix} = \begin{bmatrix} \frac{1}{\tau}\mathbf{p}_1^{XW} \\ \tau \mathbf{p}_1^{MW} \end{bmatrix}$.

I now provide three lemmas that will be used in the proofs of the main implications.

Lemma 7*. *Assumptions A2* and 4* imply that* $c_g(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta_t) > c_g(\mathbf{p}_t^W, m(\mathbf{p}_t^W, \mathbf{A}), \Theta_t)$ *if* $p_{gt} < p^W_{gt}$ *, and* $c_g(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta_t) < c_g(\mathbf{p}_t^W, m(\mathbf{p}_t^W, \mathbf{A}), \Theta_t)$ *if* $p_{gt} > p^W_{gt}$.

Proof. Due to assumption 4*, I can partition all goods into two sets, the set g^Y if $p_{gt} < p^W_{gt}$ and g^Z if $p_{gt} > p^W_{gt}$: $\mathbf{p}_t = \begin{bmatrix} \frac{1}{\tau}\mathbf{p}_t^{YW} \\ \tau \mathbf{p}_t^{ZW} \end{bmatrix}$ where $\mathbf{p}_t^{YW}$ is the vector of world prices for goods in the set g^Y and similarly for g^Z. Define $\tilde{\mathbf{p}}_\mathbf{t} = \tau \mathbf{p}_t = \begin{bmatrix} \mathbf{p}_t^{YW} \\ \tau^2 \mathbf{p}_t^{ZW} \end{bmatrix}$. Consider moving from $\mathbf{p}_t^W$ to $\tilde{\mathbf{p}}_\mathbf{t}$ by increasing the price of the goods in the set g^Z one at a time. By repeatedly applying A2*, the demand for any good g^Y strictly increases: $c_{g^Y}(\tilde{\mathbf{p}}_\mathbf{t}, m(\tilde{\mathbf{p}}_\mathbf{t}, \mathbf{A}), \Theta_t) > c_{g^Y}(\mathbf{p}_t^W, m(\mathbf{p}_t^W, \mathbf{A}), \Theta_t)$. As relative demand is homogenous of degree 0 in prices and income, and income (e.g. the revenue function) is homogenous of degree 1 in prices, $c_{g^Y}(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta_t) > c_{g^Y}(\mathbf{p}_t^W, m(\mathbf{p}_t^W, \mathbf{A}), \Theta_t)$. The exact same logic can be applied for goods in the set g^Y by considering the vector $\tilde{\tilde{\mathbf{p}}}_\mathbf{t} = \frac{1}{\tau}\mathbf{p}_t$ and lowering the price of export goods. □

Lemma 8*. *Assumption 1* and A1* imply that* $c_g(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta) > c_g(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \tilde{\Theta})$ *if* $\theta_g > \tilde{\theta}_g$*, and* $c_g(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta) < c_g(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \tilde{\Theta})$ *if* $\theta_g < \tilde{\theta}_g$.

Proof. Assumption 1* implies that $s_g(u_t, \mathbf{p}_t, \Theta) > s_g(u_t, \mathbf{p}_t, \tilde{\Theta})$ if $\theta_g > \tilde{\theta}_g$. Assumption implies that $c_g(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta) > c_g(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \tilde{\Theta})$ if $\theta_g > \tilde{\theta}_g$, where $c_g(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta) = \frac{s_g(v(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta), \mathbf{p}_t, \Theta) m(\mathbf{p}_t, \mathbf{A})}{p_{gt}}$. Similarly, $c_g(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \Theta) < c_g(\mathbf{p}_t, m(\mathbf{p}_t, \mathbf{A}), \tilde{\Theta})$ if $\theta_g < \tilde{\theta}_g$. □

Lemma 9*. *The assumption on production,* $\frac{\partial Q_g}{\partial p_g} > 0$ *and* $\frac{\partial Q_g}{\partial p_{g'}} < 0 \ \forall g' \neq g$*, implies that* $Q_g(\mathbf{p}_t, \mathbf{A}) < Q_g(\mathbf{p}_t^W, \mathbf{A})$ *if* $p_{gt} < p^W_{gt}$ *, and* $Q_g(\mathbf{p}_t, \mathbf{A}) > Q_g(\mathbf{p}_t^W, \mathbf{A})$ *if* $p_{gt} > p^W_{gt}$.

Proof. Due to assumption 4*, I can partition all goods into two sets, the set g^Y if $p_{gt} < p^W_{gt}$ and g^Z if $p_{gt} > p^W_{gt}$: $\mathbf{p}_t = \begin{bmatrix} \frac{1}{\tau}\mathbf{p}_t^{YW} \\ \tau \mathbf{p}_t^{ZW} \end{bmatrix}$ where $\mathbf{p}_t^{YW}$ is the vector of world prices for goods in the set

g^Y and similarly for g^Z. Define $\tilde{\mathbf{p}}_{\mathbf{t}} = \tau \mathbf{p}_t = \begin{bmatrix} \mathbf{p}_t^{YW} \\ \tau^2 \mathbf{p}_t^{ZW} \end{bmatrix}$. Consider moving from $\mathbf{p}_t^W$ to $\tilde{\mathbf{p}}_{\mathbf{t}}$ by increasing the price of the goods in the set g^Z one at a time. By repeatedly applying $\frac{\partial Q_g}{\partial p_{g'}} < 0$, the quantity of g^Y produced strictly decreases: $Q_g(\tilde{\mathbf{p}}_{\mathbf{t}}, \mathbf{A}) < Q_g(\mathbf{p}_t^W, \mathbf{A})$. As the GDP function is homogenous of degree 1 in output prices, output supplies must be homogenous of degree zero in output prices, $Q_g(\mathbf{p}_t, \mathbf{A}) = Q_g(\tilde{\mathbf{p}}_{\mathbf{t}}, \mathbf{A}) < Q_g(\mathbf{p}_t^W, \mathbf{A})$. The exact same logic can be applied for goods in the set g^Y by considering the vector $\tilde{\tilde{\mathbf{p}}}_{\mathbf{t}} = \frac{1}{\tau}\mathbf{p}_t$ and lowering the price of export goods. □

As in the main text, I now define a set of habit strengths for which there is no comparative advantage reversal.

Definition 3*. ***No-Comparative-Advantage-Reversal Set****:*

Define $\tilde{\nu_t}$ as a set of habit strengths, $\nu_t > 0$, for which the domestic supply is less than or equal to domestic the demand for at least one good $g \in g^X$ in period t under the price vector $\mathbf{p}_1 = \begin{bmatrix} \frac{1}{\tau}\mathbf{p}_1^{XW} \\ \tau \mathbf{p}_1^{MW} \end{bmatrix}$, or domestic supply is greater than or equal to domestic demand for at least one good $g \in g^M$ in period t under the price vector $\mathbf{p}_1$: $Q_{g^X}(\mathbf{p}_1, \mathbf{A}) \leq c_{g^X}(\mathbf{p}_1, m(\mathbf{p}_1, \mathbf{A}), \tilde{\Theta}_t)$ and $Q_{g^M}(\mathbf{p}_1, \mathbf{A}) \geq c_{g^M}(\mathbf{p}_1, m(\mathbf{p}_1, \mathbf{A}), \tilde{\Theta}_t)$ where $\tilde{\Theta}_t = H_t(\mathbf{p}_1, \mathbf{A}, \Theta_1, \nu_t)$ through recursive substitution.

Assumption 4* implies that every good is traded in every period. Trade in each good always flows in the same direction if $\nu \notin \tilde{\nu_t}\ \forall t$. In this scenario, the domestic price vector $\mathbf{p_t}$ at the start of any pre-trade period is equal to $\mathbf{p_1}$ defined above.

Implication 4*. *Under assumptions 1*-4* and A1*-A2*, at the start of period T, household tastes are biased towards the foods for which a region has a relatively low price compared to the world price, $(\theta_{gT} - \theta_{gT}^W)(p_{gT} - p_{gT}^W) < 0\ \forall g$, if and only if $\nu > 0$ and $\nu \notin \tilde{\nu_T}$. There is no such relationship in the absence of habit formation, $(\theta_{gT} - \theta_{gT}^W)(p_{gT} - p_{gT}^W) = 0\ \forall g$ if $\nu = 0$.*

Proof. First, I address the case where $\nu > 0$ and $\nu \notin \tilde{\nu_T}$. As $\nu \notin \tilde{\nu_T}$, $\mathbf{p_1} = \mathbf{p_T}$ from definition 3*. It suffices to show that $\theta_{g^X T} > \theta_{g^X 1}$ and $\theta_{g^M T} < \theta_{g^M 1}$. Without loss of generality, I consider an initially exported good, $g' \in g^X$, and show that $\theta_{g'T} > \theta_{g'1}$ by exploring the two possible cases.

Case 1: No reversal in any period ($p_{g't} < p^W_{g'1}\ \forall t$). In this case $c_{g'}(\mathbf{p}_1, m(\mathbf{p}_1, \mathbf{A}), \Theta_1) > c_{g'}(\mathbf{p}^W_1, m(\mathbf{p}^W_1, \mathbf{A}), \Theta_1)$ by lemma 7*. Thus, $\theta_{g'2} > \theta_{g'1}$ by assumption 3*. Higher habits in period 2 ensures that period 2 consumption remains above the threshold $\overline{c_{g'}}$, $c_{g'}(\mathbf{p}_2, m(\mathbf{p}_2, \mathbf{A}), \Theta_2) > c_{g'}(\mathbf{p}_2, m(\mathbf{p}_2, \mathbf{A}), \Theta_1) > c_{g'}(\mathbf{p}^W_1, m(\mathbf{p}^W_1, \mathbf{A}), \Theta_1) = \overline{c_{g'}}$ by lemma 8*, lemma 7* and $p_{g't} < p^W_{g'1}$. Thus, $\theta_{g'3} > \theta_{g'1}$ by assumption 3*. By repeatedly applying this logic, $\theta_{g'T} > \theta_{g'1}$.

Case 2: Last reversal in period $T-n$ where $T-2 > n > 0$, no reversal thereafter ($p_{g'T-n} > p^W_{g'1}$ by assumption 4*, $p_{g'T-j} < p^W_{g'1}$ for $0 \leq j < n$). In this case, good g' is imported in period $T-n$ but is exported in period 1: $Q_{g'}(\mathbf{p_{T-n}}, \mathbf{A}) < c_{g'}(\mathbf{p_{T-n}}, m(\mathbf{p_{T-n}}, \mathbf{A}), \Theta_{T-n})$ and $Q_{g'}(\mathbf{p_1}, \mathbf{A}) > c_{g'}(\mathbf{p_1}, m(\mathbf{p_1}, \mathbf{A}), \Theta_1)$. Therefore, $c_{g'}(\mathbf{p_{T-n}}, m(\mathbf{p_{T-n}}, \mathbf{A}), \Theta_{T-n}) > c_{g'}(\mathbf{p_1}, m(\mathbf{p_1}, \mathbf{A}), \Theta_1)$ by lemma 9*. Hence $\theta_{g'T-n+1} > \theta_{g'1}$ by assumption 3* and the fact $c_{g'}(\mathbf{p}_1, m(\mathbf{p}_1, \mathbf{A}), \Theta_1) > c_{g'}(\mathbf{p}^W_1, m(\mathbf{p}^W_1, \mathbf{A}), \Theta_1)$ via lemma 7* and $p_{g'1} < p^W_{g'1}\ \forall\ g \in g^X$. Combining these inequalities, $c_{g'}(\mathbf{p}_{T-n+1}, m(\mathbf{p}_{T-n+1}, \mathbf{A}), \Theta_{T-n+1}) > c_{g'}(\mathbf{p}_{T-n+1}, m(\mathbf{p}_{T-n+1}, \mathbf{A}), \Theta_1) > c_{g'}(\mathbf{p}^W_1, m(\mathbf{p}^W_1, \mathbf{A}), \Theta_1)$ by lemma 8*, lemma 7* and $p_{g'T-j} < p^W_{g'1}$ for $0 \leq j < n$. Thus, $\theta_{g'T-n+2} > \theta_{g'1}$ by assumption 3*. By repeatedly applying this logic, $\theta_{g'T} > \theta_{g'1}$.

Second, I address the case where $\nu > 0$ and $\nu \in \tilde{\nu_T}$. As $\nu \in \tilde{\nu_T}$, $\mathbf{p}_1 \neq \mathbf{p_T}$. I prove that $(\theta_{gT} - \theta_{g1})(p_{gT} - p^W_{gT}) \geq 0$ for some g by contradiction. Without loss of generality, I consider one member g' of the set g^X for which $p_{g1} \neq p_{gT}$, hence $p_{g'1} < p^W_{g't} < p_{g'T}$ by assumption 4*. Therefore, in period T the good is imported but was exported in period 1: $Q_{g'}(\mathbf{p_T}, \mathbf{A}) < c_{g'}(\mathbf{p_T}, m(\mathbf{p_T}, \mathbf{A}), \Theta_T)$ and $Q_{g'}(\mathbf{p_1}, \mathbf{A}) > c_{g'}(\mathbf{p_1}, m(\mathbf{p_1}, \mathbf{A}), \Theta_1)$. Therefore, $c_{g'}(\mathbf{p_T}, m(\mathbf{p_T}, \mathbf{A}), \Theta_T) > c_{g'}(\mathbf{p_1}, m(\mathbf{p_1}, \mathbf{A}), \Theta_1)$ by lemma 9*. Hence, $c_{g'}(\mathbf{p_T}, m(\mathbf{p_T}, \mathbf{A}), \Theta_T) > c_{g'}(\mathbf{p_1}, m(\mathbf{p_1}, \mathbf{A}), \Theta_1) > c_{g'}(\mathbf{p_T}, m(\mathbf{p_T}, \mathbf{A}), \Theta_1)$ by lemma 7* and $p^W_{g't} < p_{g'T}$. If $(\theta_{gT} - \theta_{g1})(p_{gT} - p^W_{gT}) < 0\ \forall g$, then $\theta_{g'T} < \theta_{g1}$ as $p_{g'T} > p^W_{g't}$. Thus, $c_{g'}(\mathbf{p_T}, m(\mathbf{p_T}, \mathbf{A}), \Theta_T) < c_{g'}(\mathbf{p_T}, m(\mathbf{p_T}, \mathbf{A}), \Theta_1)$ by lemma 8, a contradiction. Therefore, $(\theta_{gT} - \theta_{g1})(p_{gT} - p^W_{gT}) \geq 0$ for some g.

Third, I address the case where $\nu = 0$. Assumption 3* implies that $\theta_{gT} = \theta_{g1}$ and so $(\theta_{gT} - \theta_{g1})(p_{gT} - p^W_{gT}) = 0\ \forall g$.

Finally, assumption 2* and lemma 6* imply that $\theta^W_{gT} = \theta^W_{g1} = \theta_{g1}\ \forall g$, hence, the inequality can be rewritten as $(\theta_{gT}-\theta_{gT}^W)(p_{gT}-p_{gT}^W) < 0\ \forall g$ iff $\nu > 0$ and $\nu \notin \tilde{\nu_T}$, and $(\theta_{gT}-\theta_{gT}^W)(p_{gT}-p_{gT}^W) = 0$ $\forall g$ if $\nu = 0$.. □

I now evaluate the caloric impact of a marginal reduction in τ at time T. Recall equation 1 from the main text:

$$\frac{dK_T}{K_T} = \underbrace{-\sum_g csh_{gT}\frac{dp_{gT}}{p_{gT}}}_{\text{wealth effect } W} + \underbrace{\frac{dm_T}{m_T}}_{\text{factor income effect } F} + \underbrace{\sum_g csh_{gT}\frac{ds_{gT}}{s_{gT}}}_{\text{reallocation effect } R}, \tag{19}$$

where $csh_{gT} = \frac{\frac{s_{gT}}{p_{gT}}}{\sum_{g'}\frac{s_{g'T}}{p_{g'T}}}$.

Implication 5*. *Under assumptions 1*-4* and A1*-A2*, and if* $\nu \notin \tilde{\nu_T}$, $W_T/\frac{-d\tau}{\tau}$ *is more negative in the scenario where* $\nu > 0$ *and tastes favor the comparative advantage food compared to the scenario where* $\nu = 0$ *and tastes are unbiased and identical across regions and equal to the tastes of world households.* $F_T/\frac{-d\tau}{\tau}$ *is unchanged across the two scenarios. Therefore,* $\frac{dK_T}{K_T}/\frac{-d\tau}{\tau}|_{(\theta_{gT}-\theta_{gT}^W)(p_{gT}-p_{gT}^W)<0\forall g} < \frac{dK_T}{K_T}/\frac{-d\tau}{\tau}|_{\theta_{gT}=\theta_{gT}^W=\theta_{g1}^W\forall g}$ *if* $\frac{W_T}{\frac{-d\tau}{\tau}}|_{\sum_g(\theta_{gT}-\theta_{gT}^W)(p_{gT}-p_{gT}^W))<0\forall g} - \frac{W_T}{\frac{-d\tau}{\tau}}|_{\theta_{gT}=\theta_{gT}^W=\theta_{g1}^W\forall g} < -\frac{R_T}{\frac{-d\tau}{\tau}}|_{\sum_g(\theta_{gT}-\theta_{gT}^W)(p_{gT}-p_{gT}^W)<0\forall g} + \frac{R_T}{\frac{-d\tau}{\tau}}|_{\theta_{gT}=\theta_{gT}^W=\theta_{g1}^W\forall g}$.

Proof. At the time of liberalization, $\frac{dp_{g^XT}}{p_{g^XT}}/\frac{-d\tau}{\tau} = 1$ and $\frac{dp_{g^MT}}{p_{g^MT}}/\frac{-d\tau}{\tau} = -1$. Hence, $W = \frac{-\sum_{g^X}\frac{s_{g^XT}}{p_{g^XT}}+\sum_{g^M}\frac{s_{g^MT}}{p_{g^MT}}}{\sum_{g^X}\frac{s_{g^XT}}{p_{g^XT}}+\sum_{g^M}\frac{s_{g^MT}}{p_{g^MT}}} \equiv \frac{-A+B}{A+B}$. If $(\theta_{gT}-\theta_{gT}^W)(p_{gT}-p_{gT}^W) < 0\ \forall g$, $\theta_{gT} > \theta_{gT}^W$ for g^X goods and $\theta_{gT} < \theta_{gT}^W$ for g^M goods. Assumption 1* and A1* imply that:

$s_{g^X}(v(\mathbf{p}_T, m(\mathbf{p}_T,\mathbf{A}),\Theta_T),\mathbf{p}_T,\Theta_T) > s_{g^X}(v(\mathbf{p}_T, m(\mathbf{p}_T,\mathbf{A}),\Theta_T^W),\mathbf{p}_T,\Theta_T^W)$ and

$s_{g^M}(v(\mathbf{p}_T, m(\mathbf{p}_T,\mathbf{A}),\Theta_T),\mathbf{p}_T,\Theta_T) < s_{g^M}(v(\mathbf{p}_T, m(\mathbf{p}_T,\mathbf{A}),\Theta_T^W),\mathbf{p}_T,\Theta_T^W)$.

Therefore, $A|_{\theta_{g^XT}>\theta_{g^XT}^W\forall g^X} > A|_{\theta_{g^XT}=\theta_{g^XT}^W\forall g^X}$ and $B|_{\theta_{g^MT}>\theta_{g^MT}^W\forall g^M} < B|_{\theta_{g^MT}=\theta_{g^MT}^W\forall g^M}$, which in turn implies that $W/\frac{-d\tau}{\tau}|_{(\theta_{gT}-\theta_{gT}^W)(p_{gT}-p_{gT}^W)<0\forall g} < W/\frac{-d\tau}{\tau}|_{\theta_{gT}=\theta_{gT}^W=\theta_{g1}^W\forall g}$.

Therefore, the implication follows directly from equation 19 and the fact that $\frac{dm(\mathbf{p}_T,\mathbf{A})}{m(\mathbf{p}_T,\mathbf{A})}$ is independent of θ_{gT} in the case of strict no comparative advantage reversal and hence $F_T/\frac{-d\tau}{\tau}$ is unchanged in the two scenarios. □

Note that as the region is small, the global average prices and tastes are simply equal to world prices and tastes: $\overline{p}_{gT} = p_{gT}^W$ and $\overline{\theta}_{gT} = \theta_{gT}^W = \theta_{g1}^W\ \forall g$. Replacing world values with average values generates implications 1-2 and 4-5 which I test in the empirical section of the main paper.

B.4.1 The Relationship Between Tastes and Endowments (Propositions 1.1*, 1.2* and 1.3*)

The multi-good model laid out above was agnostic about the source of comparative advantage. Without further assumptions on technologies, endowments and preferences it is not possible to make strong statements about the relationship between autarky price differences and factor endowments even in generation 1 when tastes are identical everywhere (Dixit and Norman, 1980). Therefore, in this section, I impose some additional restrictions on preferences and endowments in order to motivate the multi-good empirical counterpart to implication 3.

A correlation-like relationship between period 1 autarky prices and factor endowments can be derived in a multi-good world if I follow the two good model and both restrict preferences to be homothetic and initially symmetric across foods and impose some symmetry on the distribution of factor endowments around the world (proposition 1.1*). Furthermore, the generalized law of comparative advantage (Deardorff, 1980) implies that there will be a correlation-like relationship between relative autarky prices and net trade flows in the presence of non-prohibitive transport costs (proposition 1.2*). Finally, habit formation implies that there is a positive relationship between net trade flows and relative tastes (proposition 1.3*). Combining these three correlation-like results suggests the multi-good empirical counterpart to implication 3.

As in the two-good model in the main paper, I assume that technologies are identical across countries and goods, and that the production function f for good g, $Q_{gt} = f(V_g, L_{gt})$, is increasing, concave and homogeneous of degree one in L and V. The specific land endowments for the G goods and labor are stacked in the vector $\mathbf{A}^T = \begin{pmatrix} V_1 & \cdots & V_G & L \end{pmatrix}$.

I reintroduce four assumptions present in the two-country two-good model in order to prove proposition 1.1*: i) I assume that preferences are homothetic. ii) I assume a multi-good assumption 2, that the expenditure function is symmetric in all goods, and the first generation of adults has unbiased preferences that are equal for all foods. iii) I assume the world is comprised of N regions that are of identical size (in terms of population and total land area) as the small region under study. iv) I assume the world is endowed with the same total amount of each specific land factor but each region has an uneven distribution of factors.

Proposition 1.1*. *Under assumptions 1*-2*, and i)-iv), on average, regions which have a particularly high proportion of cropland that is suitable for growing a food compared to the world will have lower period 1 autarky prices p^a_{g1} for that food compared to the world,* $\sum_g (p^a_{g1} - p^{aW}_{g1})(\frac{V_g}{\sum_g V_g} - \frac{V^W_g}{\sum_g V^W_g}) \leq 0$.

Proof. In period 1, tastes are identical across the world (assumption 2*) and preferences are homothetic. This is the standard preference assumption analyzed in the international trade literature and the first half of the argument follows Dixit and Norman (1980, pp. 96-98) exactly.

As preferences are homothetic, the expenditure functions in period 1 can be written as $e(u_1, \mathbf{p}_1; \Theta_1) = u_1 \bar{e}(\mathbf{p}_1; \Theta_1)$. Therefore, the autarky price vector $\mathbf{p}^a_1$ is determined in the small region and in the world by equating expenditure with the revenue function of the economy, $r(\mathbf{p}_1, \mathbf{A})$.

$$u_1 \bar{e}(\mathbf{p}^a_1; \Theta_1) = r(\mathbf{p}^a_1, \mathbf{A}),$$

$$u^W_1 \bar{e}(\mathbf{p}^{aW}_1; \Theta_1) = r(\mathbf{p}^{aW}_1, \mathbf{A}^W).$$

I choose a numeraire in each country such that $\bar{e}(\mathbf{p}^{aW}_1; \Theta_1) = \bar{e}(\mathbf{p}^a_1; \Theta_1) = 1$. As I assumed that the small region has an uneven distribution of factors, the region's autarky price vector differs from the world autarky price vector in generation 1 when preferences are identical and symmetric. As free trade is always preferred to autarky, any other vector of prices is preferable to the autarky price vector if indifference curves are smooth.[10] Therefore,

$$r(\mathbf{p}^{aW}_1, \mathbf{A}) > r(\mathbf{p}^a_1, \mathbf{A}),$$

$$r(\mathbf{p}^{aW}_1, \mathbf{A}^W) < r(\mathbf{p}^a_1, \mathbf{A}^W),$$

$$[r(\mathbf{p}^a_1, \mathbf{A}) - r(\mathbf{p}^{aW}_1, \mathbf{A})] - [r(\mathbf{p}^a_1, \mathbf{A}^W) - r(\mathbf{p}^{aW}_1, \mathbf{A}^W)] < 0. \tag{20}$$

Equation 20 can be thought of as a second difference of r between the points $(\mathbf{p}^a_1, \mathbf{A})$ and $(\mathbf{p}^{aW}_1, \mathbf{A}^W)$. If these points are sufficiently close together, equation 20 can be approximated by the second-order terms of a Taylor expansion of the revenue function:

$$(\mathbf{p}^a_1 - \mathbf{p}^{aW}_1)^T r^{aW}_{\mathbf{pA}} (\mathbf{A} - \mathbf{A}^W) < 0, \tag{21}$$

[10]See Dixit and Norman (1980, pp. 74) for a formal proof.

where $r^{aW}_{\mathbf{pA}}$ is the matrix of cross-derivatives of the revenue function evaluated at $(\mathbf{p}^{aW}_1, \mathbf{A}^W)$. In the specific factors model, the cross-derivatives can be easily signed:

$$r_{\mathbf{pA}} = \begin{bmatrix} \frac{\partial^2 r(\mathbf{p},\mathbf{A})}{\partial p_1 \partial V_1} & \cdots & \frac{\partial^2 r(\mathbf{p},\mathbf{A})}{\partial p_1 \partial V_G} & \frac{\partial^2 r(\mathbf{p},\mathbf{A})}{\partial p_1 \partial L} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial^2 r(\mathbf{p},\mathbf{A})}{\partial p_G \partial V_1} & \cdots & \frac{\partial^2 r(\mathbf{p},\mathbf{A})}{\partial p_G \partial V_G} & \frac{\partial^2 r(\mathbf{p},\mathbf{A})}{\partial p_G \partial L} \end{bmatrix},$$

where $\frac{\partial^2 r(\mathbf{p},\mathbf{A})}{\partial p_g \partial L} > 0$, $\frac{\partial^2 r(\mathbf{p},\mathbf{A})}{\partial p_g \partial V_g} > 0$ and $\frac{\partial^2 r(\mathbf{p},\mathbf{A})}{\partial p_g \partial V_{g'}} < 0$ if $g \neq g'$.

As the revenue function is homogenous of degree 1 in endowments, $\mathbf{A}$ and $\mathbf{A}^W$ in equation 20 can be replaced by $\tilde{\mathbf{A}} = \frac{\mathbf{A}}{\sum_g V_g}$ and $\tilde{\mathbf{A}}^W = \frac{\mathbf{A}^W}{\sum_g V_g^W}$. Hence equation 21 becomes:

$$(\mathbf{p}^a_1 - \mathbf{p}^{aW}_1) r^W_{\mathbf{pA}} (\tilde{\mathbf{A}} - \tilde{\mathbf{A}}^W) < 0. \tag{22}$$

Furthermore, as the world is comprised of N regions with identical land area and population to the home region:

$$\tilde{\mathbf{A}} - \tilde{\mathbf{A}}^W = \begin{bmatrix} \frac{V_1}{\sum_g V_g} - \frac{V_1^W}{\sum_g V_g^W} \\ \vdots \\ \frac{V_1}{\sum_g V_g} - \frac{V_1^W}{\sum_g V_g^W} \\ 0 \end{bmatrix}.$$

Expanding equation 22 implies that:

$$\sum_g (p^a_{g1} - p^{aW}_{g1}) \frac{\partial^2 r(\mathbf{p}^{aW}_1, \mathbf{A}^W)}{\partial p_{g1} \partial V_g} (\frac{V_g}{\sum_g V_g} - \frac{V_g^W}{\sum_g V_g^W}) + \sum_g \sum_{g' \neq g} (p^a_{g1} - p^{aW}_{g1}) \frac{\partial^2 r(\mathbf{p}^{aW}_1, \mathbf{A}^W)}{\partial p_{g1} \partial V_{g'}} (\frac{V_{g'}}{\sum_g V_g} - \frac{V_{g'}^W}{\sum_g V_g^W}) < 0.$$

The fact that the world has an equal quantity of each specific land endowment and preferences and technologies are symmetric in period 1 across foods implies that the world price of each food is equal and hence $\frac{\partial^2 r(\mathbf{p}^{aW}_1, \mathbf{A}^W)}{\partial p_{g1} \partial V_{g'}} < 0$ is the same for all $g \neq g'$, and $\frac{\partial^2 r(\mathbf{p}^{aW}_1, \mathbf{A}^W)}{\partial p_{g1} \partial V_{g'}} > 0$ is the same for all $g = g'$. Therefore,

$$(\frac{\partial^2 r(\mathbf{p}^{aW}_1, \mathbf{A}^W)}{\partial p_{g1} \partial V_g} - \frac{\partial^2 r(\mathbf{p}^{aW}_1, \mathbf{A}^W)}{\partial p_{g1} \partial V_{g'}}) \sum_g (p^a_{g1} - p^{aW}_{g1}) (\frac{V_g}{\sum_g V_g} - \frac{V_g^W}{\sum_g V_g^W}) < 0,$$

$$\sum_g (p^a_{g1} - p^{aW}_{g1}) (\frac{V_g}{\sum_g V_g} - \frac{V_g^W}{\sum_g V_g^W}) < 0.$$

□

Proposition 1.2*. *Under assumptions 1*-2* and i)-iv), on average, if trade costs* $\tau > 1$ *are sufficiently low such that trade occurs in period 1, regions will export the foods for which the autarky price for that food is relatively inexpensive compared to the world autarky price,* $\sum_g (p_{g1}^a - p_{g1}^{aW})(Q_{g1} - c_{g1}) < 0$.

Proof. This is the weak form of the Law of Comparative Advantage. The proof is contained in Deardorff (1980, pp. 943-952) and carries through under much more general assumptions than the model in this section. □

Proposition 1.3*. *Under assumptions 1*-4*, i)-iv) and* $\nu \notin \tilde{\nu_T}$*, at the start of period* T*, household tastes are biased towards the foods which a region exported in period 1,* $(\theta_{gT} - \theta_{gT}^W)(Q_{g1} - c_{g1}) > 0 \ \forall g$*, if and only if there is habit formation (*$\nu > 0$*).*

Proof. Implication 4* states that $(\theta_{gT} - \theta_{gT}^W)(p_{gT} - p_{gT}^W) < 0 \ \forall g$ if $\nu > 0$ and $\nu \notin \tilde{\nu_T}$. $p_{gT} < p_{gT}^W$ implies that good g is exported and $Q_{gT} > c_{gT}$, while $p_{gT} > p_{gT}^W$ implies $Q_{gT} < c_{gT}$. Therefore, $(\theta_{gT} - \theta_{gT}^W)(Q_{gT} - c_{gT}) > 0$ if $\nu > 0$. Finally, the fact that habits are below the no-comparative-advantage-reversal threshold (definition 3*) implies that $sign(Q_{g1} - c_{g1}) = sign(Q_{gT} - c_{gT})$, hence $(\theta_{gT} - \theta_{gT}^W)(Q_{g1} - c_{g1}) > 0$ if $\nu > 0$. □

In the two-good model, implication 3 can be rewritten as $\sum_g (\theta_{gT} - \theta_{gT}^*)(\frac{V_g}{\sum_g V_g} - \frac{V_g^*}{\sum_g V_g^*}) > 0$ iff $\nu > 0$, where asterisks denote foreign region variables. In the many-good case, I have three distinct relationships linking tastes with endowments: $(\theta_{gT} - \theta_{gT}^W)(Q_{g1} - c_{g1}) > 0 \ \forall g$ if $\nu > 0$ and $\nu \notin \tilde{\nu_T}$, $\sum_g (Q_{g1} - c_{g1})(p_{g1}^a - p_{g1}^{aW}) < 0$ and $\sum_g (p_{g1}^a - p_{g1}^{aW})(\frac{V_g}{\sum_g V_g} - \frac{V_g^W}{\sum_g V_g^W}) < 0$ (propositions 1.3*, 1.2* and 1.1*). In words, habits favor export foods, exported foods are on average relatively inexpensive under autarky, and goods that are relatively inexpensive under autarky are on average intensive in the relatively abundant endowments. The latter two relationships are correlation-like results. Since $corr(X, Y) < 0$ and $corr(Y, Z) < 0$ do not necessarily imply that $corr(X, Z) > 0$, the three relationships are not sufficient to prove that $\sum_g (\theta_{gT} - \theta_{gT}^W)(\frac{V_g}{\sum_g V_g} - \frac{V_g^W}{\sum_g V_g^W}) > 0$. However, this chain of results is strongly suggestive of a positive empirical correlation between regional taste deviations, $\theta_{gT} - \theta_{gT}^W$, and relative endowment differences, $\frac{V_g}{\sum_g V_g} - \frac{V_g^W}{\sum_g V_g^W}$, in the multi-good case.

Accordingly, in the empirical section 4.2, I explore whether the most natural multi-good extension of implication 3, $(\theta_{gT} - \theta_{gT}^W)(\frac{V_g}{\sum_g V_g} - \frac{V_g^W}{\sum_g V_g^W}) > 0$ if $\nu > 0$ and $\nu \notin \tilde{\nu_T}$, holds on average across regions and goods: $\sum_g(\theta_{gT} - \theta_{gT}^W)(\frac{V_g}{\sum_g V_g} - \frac{V_g^W}{\sum_g V_g^W}) > 0$.

C Background on Agricultural Trade in India

I briefly review the current state of Indian agricultural trade before assessing the potential impact of domestic liberalization. Despite wide ranging economic reforms over the last two decades, India's agricultural sector remains highly restricted. While there has been new legislation at the national (Union) level to liberalize domestic markets, these measures have been applied erratically at best because agricultural policy is under the exclusive constitutional remit of state governments.[11]

Interventionist food policies were initially enacted in response to the perceived failures of private trade in the Bengal famine of 1943. The Essential Commodities Act (1955) entitles both governments and states to impose restrictions on "trade and commerce in, and the production, supply and distribution of foodstuffs."[12] Other agricultural acts control to whom farmers and traders are allowed to sell and at what price. All traders require licenses, have restricted access to credit and must follow over 400 rules that govern food trade (Planning Commisson of India, 2001).

Internal trade is further restrained through state tariffs and district-level entry taxes, Octroi, collected at often corrupt checkpoints (Das-Gupta, 2006). This is in addition to the extremely poor transport infrastructure across India, which is perhaps the biggest hindrance to trading bulky agricultural goods within the country. State governments are also directly involved in the purchase and sale of food. The Commission on Agricultural Costs and Prices sets minimum support prices for farmers that are only available in certain regions, while state levies require private mills to supply grain at a fixed price, which is then sold to the poor through the Public Distribution System at prices chosen by each state. Jha et al. (2005) discuss these numerous restrictions in more detail, and show that as a result wholesale rice markets across India are not integrated. The lack of integration is evident in the NSS data, in which the dispersion of regional prices actually increased between 1987-88 and 2004-05.[13]

[11]For example, the Agricultural Produce Marketing Acts was amended in 2003 to allow farmers to sell their produce directly to buyers for the first time. Only about half of the states have so far incorporated the amendment and in most cases with substantial changes.

[12]FAO (2005) details some of the numerous state-level and even district-level restrictions that remain.

[13]The average over 52 foods of the cross-regional coefficients of variation of rural median food prices rose from 0.51 in 1987-88 to 0.53 in 2004-05. Similarly, the average pairwise correlation between the median prices

Although there has been little progress reforming the domestic market, if India had fully liberalized all external trade, the domestic agricultural market would have become integrated. However, external agricultural trade has only seen limited reform in the years following India's 1991 liberalization. The initial tariff reductions did not cover agricultural goods at all. The impetus for agricultural liberalization came from the Agreement on Agriculture, which India committed to as a founding member of the WTO. This agreement required the conversion of all non-tariff barriers and quantitative restrictions into tariffs by 2002, but left domestic support untouched. However, tariff levels were set sufficiently high to choke imports in all but pulses and oilseeds.[14] As a result, the FAO (2003) reports that there was little impact from the liberalization of agricultural trade under the Agreement on Agriculture between 1997 and 2002.[15]

India still maintains high tariffs, agricultural import monopolies, state trading enterprises and export restrictions that maintain a "highly interventionist agricultural development policy regime" (Athukorala, Prema-chandra, 2005). Accordingly, alongside the domestic restraints detailed above, agricultural trade within India remains highly restricted, and internal markets are far from integrated.[16]

I provide two empirical tests in support of the hypothesis that internal markets are far from inetgrated. First, in the absence of barriers to trade, the possibility of arbitrage ensures that prices are equalized across regions, yet substantial price differences persist. There is sizeable price dispersion, with an average log price difference of 0.49 between village median prices for the same food in the same season in different regions. The distribution is shown in appendix figure 7.

Second, in the absence of barriers to trade, abnormal weather conditions in a particular region should affect prices equally in all regions, a hypothesis that is easily rejected by the data. As shown in appendix table 6, regional prices respond significantly to regional weather

of the 52 foods in any two regions declined from 0.85 to 0.83 between the two surveys.

[14] In these two categories India is not self-sufficient and the government itself controls a substantial portion of imports via government agencies. According to (Gulati, 1998), the Indian Government followed the following rule: "Allow imports if there was a net deficit and allow exports if there was a comfortable surplus."

[15] Agricultural exports did, however, respond positively to the 20 percent devaluation of the rupee in 1991.

[16] Therefore, my theoretical mechanism cannot explain the decline in caloric intake that has occurred across India in the last 20 years. In fact relative prices across regions have moved in the opposite direction to that suggested by relative endowments. For example rice was already relatively cheap in large rice growing areas, and has become more so over the reform period.

deviations after controlling for national trends.

D Testing the Assumptions Behind the Identification Strategy

In the main paper, I highlight two key assumptions required to identify the regional tastes implicitly defined by the demand equation, equation 3. First, there must be price variation within each region in order to identify the common price, income and demographic effects. Second, this within-region price variation must be driven by temporary local supply shocks, such as abnormal local rainfall. If within-region price differences are driven by permanent factors, such as local endowment variation, the model implies that idiosyncratic village tastes would develop through habits and village prices will be correlated with the error term.

In order to clarify the identification assumptions, I separate the price difference for good g between village a and b in quarter q of survey period t into two components: a mean-zero temporary supply shock, ϵ_{gabqt}, and a permanent difference, x_{gabq}, $\ln p_{gaqt} - \ln p_{gbqt} = x_{gabq} + \epsilon_{gabqt}$.[17] I provide three pieces of evidence justifying the reasonableness of the identifying assumptions that there is price variation within a region, this variation is driven by supply shocks and is temporary.

First, unlike the simplified model of section 2, there is substantial price variation within regions, $x_{gabq} + \epsilon_{gabqt} \neq 0$ if $r_a = r_b$, where r_v denotes the region in which location v is situated. Figure 7 shows a kernel density plot of all log price differences between village median prices recorded in the same survey round during the same quarter (season) of the year.[18] If $x_{gabq} = 0$ within regions, there should be substantially less price dispersion within than between regions as I find in figure 7.[19] Figure 8 shows the distribution of village median prices by food and region. Again, substantial price variation within regions is apparent. Additionally, there is a large amount of price overlap, with a large range of prices observed in most regions.

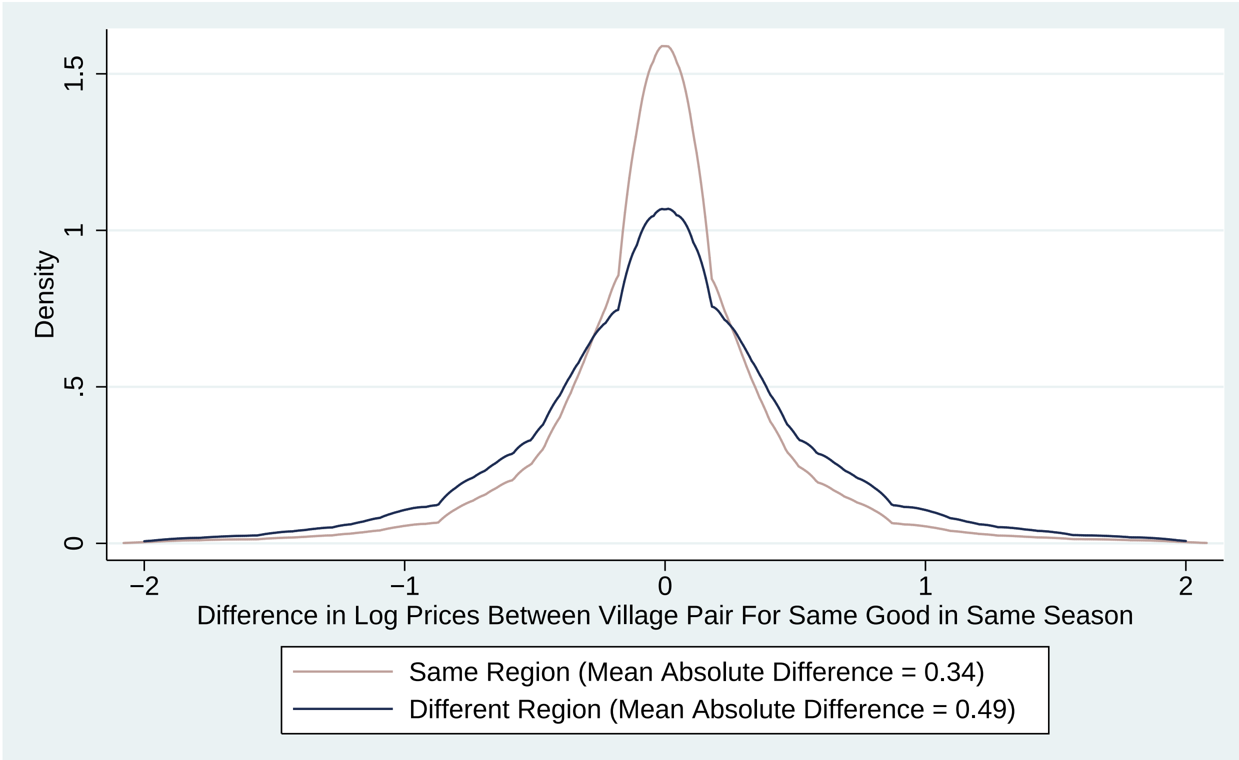
Second, using price data from all four thick survey rounds that contain district identifiers (1987-88, 1993-94, 1999-2000 and 2004-05), I find that temporary local weather shocks do alter

[17]For now I will assume that these temporary shocks are supply shocks, although my instrumentation strategy will be robust to the presence of temporary village demand shocks if these shocks are uncorrelated across space.

[18]The NSS surveys take place over one full year, with the quarter in which that particular village was surveyed recorded in the dataset.

[19]This finding also holds when I follow the border effects literature by measuring price dispersion as the standard deviation of log price differences between two villages over all 52 goods. There is substantially less price dispersion within regions than across, even once I include a quadratic distance control.

Figure 7: Dispersion of Village Median Prices (1987-1988)



local food prices after I flexibly control for regional price trends. My units of observation are quarterly (q) prices and weather deviations at the lowest geographic identifier in the survey, the district (d).[20] I match the (weighted) median unit values with weather data from Willmott and Matsuura (2001). I then regress the log price on weather shocks using a flexible specification.[21] I interact deviations from long run district means of precipitation and temperature with all of my food items, allowing each crop to be affected in a different manner:

$$\ln p_{gdqt} = \alpha_{grqt} + \beta_g(Rain_{dqt} - \overline{Rain}_{dq}) + \delta_g(Temp_{dqt} - \overline{Temp}_{dq}) + \varepsilon_{gdqt}. \tag{23}$$

Item-Region-Quarter-Year fixed effects, α_{grqt}, pick up regional price trends by food item. Therefore, the coefficients on the weather deviations, β_g and δ_g, should be zero only if within-region price variation is unaffected by local (district) weather shocks. As a further robustness check, I include Item-District fixed effects to control for permanent price differences at the district level. The same specification is also run at the region rather than district level to provide evidence that Indian regions are partially autarkic in section 3.2.

[20] I form a panel of districts as defined by the 422 district boundaries in the first sample round, 1987-88.

[21] Weather shocks are the deviations from the mean monthly temperature and precipitation 1955-2006 at the district level. These deviations are then averaged over the three month quarters in which the surveys were collected.

The district results are reported in columns 3 and 4 of table 6. The F-statistics on all the weather deviations terms are highly significant, suggesting that local weather shocks do indeed drive within-region price variation as required for my strategy. Unfortunately, I cannot use the weather shocks themselves as instruments as the lack of village identifiers means that the instrument only varies at the district level. As I require substantial price variation within regions to identify tastes separately from price effects, this instrumentation strategy is very weak and therefore unsuitable as it removes most of this within-region price variation in the data.

Third, I use multiple survey rounds to investigate how permanent price differences are between regions. The simplest methodology is to estimate the mean price difference between two districts across the survey rounds, x_{gabq}, with deviations from this mean providing estimates of ϵ_{gabqt} at the district level.[22] The means of x_{gabq} and ϵ_{gabqt} are reported in columns 1 and 3 of table 8. The absolute value of x_{gabq} is smaller within regions compared to across (0.230 compared to 0.412), suggesting that within region price differences are less permanent than temporary price differences. However, although smaller, the mean value of x_{gabq} for within-region price variation is non-zero.

Alternatively, I can perform a Dickey-Fuller type test by regressing the change in the log price difference between two districts across different survey rounds (t, $t-1$ etc.) on the level of the log price difference in the previous survey round:

$$\Delta_t(\ln p_{gaqt} - \ln p_{gbqt}) = \beta_1(\ln p_{gaq,t-1} - \ln p_{gbq,t-1}) + \beta_2(\ln p_{gaq,t-1} - \ln p_{gbq,t-1}) \times 1[r_a = r_b] + u_{gabqt}.$$

I allow the coefficient on the lagged price difference to vary depending on whether the two districts are in the same region, $r_a = r_b$. As $\hat{\beta} = -var(\epsilon_{gabqt})/(E(x^2_{gabq}) + var(\epsilon_{gabqt}))$ in the bivariate case where there is no region interaction, the coefficient on the initial price difference estimates the relative importance of the temporary and permanent components. β should be -1 if price differences are entirely temporary ($x_{gabq} = 0$), and 0 if they are entirely permanent ($var(\epsilon_{gabqt}) = 0$).

Column 1 of table 7 reports these results. The coefficient on the initial price difference is -0.536 for price differences between two districts that are situated in different regions, and -0.872 for two districts situated in the same region. Within-region price deviations are significantly

[22] I only compare prices in the same quarter of the year across different survey rounds to avoid confounding district and seasonal price differences. This provides me with four observations per district in each survey round.

more temporary than across-region price deviations and are close to being entirely temporary. However, $\beta_1 + \beta_2 = -0.872$ is still significantly less negative than -1.

Columns 2, 3 and 4 of table 7 report additional specifications. As districts in the same region are on average closer together than districts in different regions, column 2 includes additional controls for the distance between a and b as well as the interaction between $\ln distance_{ab}$ and $\ln p_{gaq,t-1} - \ln p_{gbq,t-1}$.[23] Columns 3 and 4 include district fixed effects and district pair (a, b) fixed effects respectively. Results are broadly similar in the additional specifications.

Although the above evidence generally supports the validity of the basic identifying assumptions, the small permanent component of within-region price differences is likely to generate idiosyncratic village tastes through habit formation. Hence, village prices will be endogenous in the demand system.[24] In this scenario, in order to estimate the mean regional tastes, I require an instrument that is correlated with prices but uncorrelated with the permanent idiosyncratic village tastes. Following Hausman (1994), the price in a nearby village will provide such an instrument if supply shocks are correlated spatially within regions but the idiosyncratic village tastes are not.

For this IV strategy to be valid, I require that temporary supply shocks are spatially correlated within regions (instrument relevance) but that permanent supply differences and hence idiosyncratic village tastes are not (instrument exogeneity). The relevance condition seems reasonable in this context given the finding above that local weather shocks partially drive within-region price variation and the fact that weather shocks are spatially correlated. The usual concern with this IV strategy regarding the exogeneity condition is that demand shocks are also spatially correlated due to promotions and national advertising. However, these issues are less worrisome in rural India as all my sample foods are unbranded commodities sold at village markets.

I now test these identification assumptions more formally. If temporary within-region price differences are spatially correlated but any permanent within-region price differences are not, I expect to find that $E[distance_{ab}\epsilon_{abqt}] \neq 0$ and $E[distance_{ab}x_{abq}] = 0$ if $r_a = r_b$.

[23] These are great circle distances between the centroids of the districts.

[24] For example, idiosyncratic village agro-climatic endowments can lower local prices and in later generations raise demand through habit formation.. Alternatively, the arrival of an immigrant who introduces a new food or recipe to the village can raise both local prices and demand.

In contrast, as agro-climatic conditions are generally similar in adjacent regions, I expect permanent across-region price differences to be spatially correlated: $E[distance_{ab}\epsilon_{abqt}] \neq 0$ and $E[distance_{ab}x_{abq}] \neq 0$ if $r_a \neq r_b$.

I test these hypotheses by regressing the absolute values of the estimates of x_{abq} and the residual, ϵ_{abqt}, on the distance between districts a and b. I allow the coefficient on distance to vary depending on whether the two districts are in the same region:

$$
\begin{aligned}
Abs\left\{E_t[\ln(p_{gaqt}/p_{gbqt})]\right\} = \alpha_1 1[r_a \neq r_b] + \alpha_2 1[r_a = r_b] + \beta_1 \ln distance_{ab} \\
+ \beta_2 \ln distance_{ab} \times 1[r_a = r_b] + u_{gabqt},
\end{aligned} \tag{24}
$$

$$
\begin{aligned}
Abs\left\{\ln(p_{gaqt}/p_{gbqt}) - E_t[\ln(p_{gaqt}/p_{gbqt})]\right\} = \alpha_3 1[r_a \neq r_b] + \alpha_4 1[r_a = r_b] \\
+ \beta_3 \ln distance_{ab} + \beta_4 \ln distance_{ab} \times 1[r_a = r_b] + u_{gabqt}.
\end{aligned} \tag{25}
$$

These regression results are shown in columns 2 and 4 of table 8. The permanent component of district price differences increases with distance ($\beta_1 > 0$), but substantially less so when the two districts are within the same region ($\beta_2 < 0$). However, there is still a small correlation between distance and the permanent component as $\beta_1 + \beta_2 \neq 0$. In contrast, the temporary component also increases with distance ($\beta_3 > 0$), but substantially more so when the two districts are within the same region ($\beta_4 > 0$). Therefore, the temporary component of within-region price differences is strongly spatially correlated, supporting the conjecture that supply shocks are spatially correlated as the IV strategy requires.

In conclusion, I find broadly supportive evidence for my identification assumptions. However, as with almost all demand estimates, the instrumentation strategy is imperfect. In this case, I find that permanent within-region price differences are weakly correlated across space. Hence, there are likely to be spatially correlated idiosyncratic tastes within regions due to habits. In defense of the strategy, this correlation is relatively small, at least in comparison to the across-region correlation. Additionally, the Dickey-Fuller type regressions suggest that the component of within-region price differences that is attributable to permanent factors is small in total magnitude.

Figure 8: Village Median Prices by Good and by Region (Staples, 1987-1988)

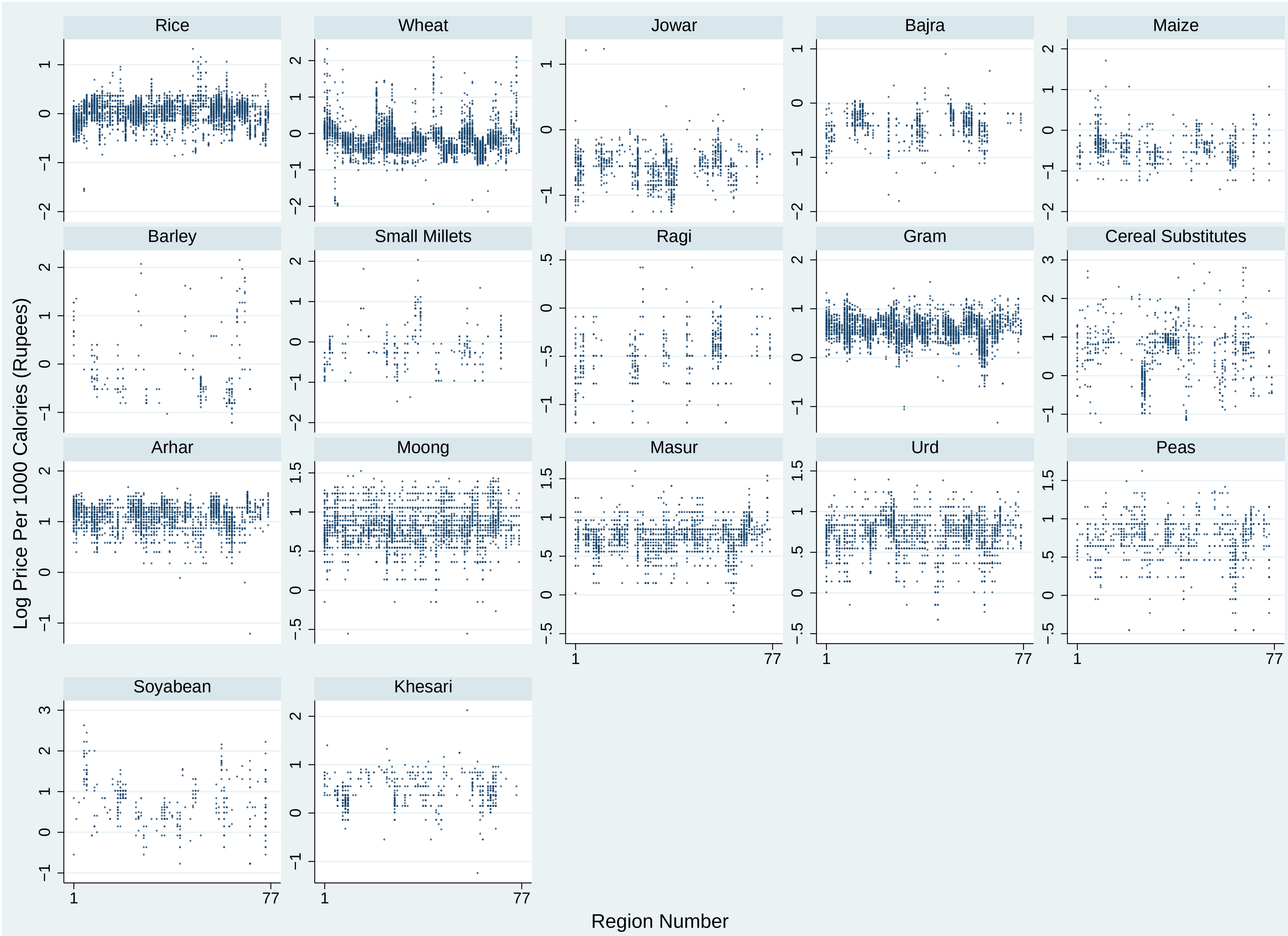


Table 6: Price Responses to Weather Shocks

	(1)	(2)	(3)	(4)
	$\ln p_{grqt}$ (regional prices)		$\ln p_{gdqt}$ (district prices)	
F-Test on Flexible Weather Deviations	470.9***	116.1***	80.08***	64.18***
Item-Quarter-Year FE	Yes	Yes		
Item-Region FE	No	Yes		
Item-Region-Quarter-Year FE			Yes	Yes
Item-District FE			No	Yes
Observations	51,882	51,882	229,200	229,200
Number of Quarters	16	16	16	16
R^2	0.949	0.976	0.975	0.980

Note: Dependent variable is the log of the weighted median price for good g in region r or district d for each quarter q of each survey period t. Independent variables are deviations from the mean monthly temperature and precipitation 1955-2006 at the region or district level. These deviations are then averaged over the three-month quarters in which the surveys were collected. Weather shocks are separately interacted with every food item. Item-Quarter-Year and Item-Region-Quarter-Year fixed effects pick up national and regional price trends by item. Item-Region and Item-District fixed effects control for permanent price differences at the region or district level. The F-test is a joint significance test on the 208 coefficients (52 good $\times$ 2 weather metrics $\times$ 2 sign interactions). Standard errors are clustered at the region level for the region regressions, and the district level for the district regressions. * significant at 10 percent, ** 5, *** 1.

Table 7: The Persistence of Price Deviations Across and Within Regions

	(1)	(2)	(3)	(4)
		$\Delta_t(\ln p_{gaqt} - \ln p_{gbqt})$		
$\ln p_{gaq,t-1} - \ln p_{gbq,t-1}$	-0.536***	-0.574***	-0.561***	-1.270***
	(0.000820)	(0.000821)	(0.00172)	(0.00837)
$(\ln p_{gaq,t-1} - \ln p_{gbq,t-1}) \times 1[region_a = region_b]$	-0.336***	-0.318***	-0.346***	-0.105***
	(0.00593)	(0.00600)	(0.0133)	(0.00622)
$1[region_a = region_b]$				0.0161***
				(0.00311)
$1[region_a \neq region_b]$				0.0177***
				(0.00365)
$\ln distance_{ab}$				-0.00123**
				(0.000564)
$(\ln p_{gaq,t-1} - \ln p_{gbq,t-1}) \times \ln distance_{ab}$				0.113***
				(0.00129)
Item, Quarter and Year FE	No	Yes	Yes	Yes
District FE	No	No	Yes	No
District-Pair FE	No	Yes	No	No
Observations	23,186,294	23,186,294	2,215,286	23,186,294
R^2	0.290	0.318	0.309	0.296

Note: Dependent variable is the change over survey rounds in the log price difference between two districts for good g in the same quarter q of the survey period t, $\Delta_t(\ln p_{gaqt} - \ln p_{gbqt})$. Independent variables are the price difference in the previous survey period, and the same variable interacted with an indicator for whether the two districts are in the same region of India. Additional controls include indicator variables for whether the two districts are in the same region, log distance and log distance interacted with the previous periods price difference, as well as various fixed effects. Column 3 uses a 30 percent subsample of district pairs in order to reduce the computational requirements. Standard errors clustered at at the district-pair level. * significant at 10 percent, ** 5, *** 1.

Table 8: Spatial Correlation of Temporary and Permanent Components of Price Differences

	(1)	(2)	(3)	(4)
	Permanent Price Difference		Temporary Price Difference	
	Abs$E_t[\ln(p_{gaqt}/p_{gbqt})]$		Abs$\{\ln(p_{gaqt}/p_{gbqt}) - E_t[\ln(p_{gaqt}/p_{gbqt})]\}$	
$1[region_a = region_b]$	0.230***	-0.156***	0.234***	0.0498***
	(0.000813)	(0.00636)	(0.000699)	(0.00471)
$1[region_a \neq region_b]$	0.412***	-0.390***	0.267***	0.0776***
	(0.000151)	(0.00131)	(9.62e-05)	(0.000803)
$\ln distance_{ab}$		0.0881***		0.00975***
		(0.000198)		(0.000122)
$\ln distance_{ab}$		-0.0544***		0.00440***
$\times 1[region_a = region_b]$		(0.00146)		(0.00107)
Item, Quarter and Year FE	No	Yes	No	Yes
Observations	12,837,529	12,837,529	38,312,119	38,312,119
R^2	0.369	0.531	0.327	0.476

Note: Dependent variable in columns 1 and 2 is the absolute value of the mean log price difference between two districts for good g in the same quarter q of the year over the multiple survey rounds t. Dependent variable in columns 3 and 4 is the absolute value of deviations from the mean log price difference. Independent variables are indicator variables for whether the two districts are in the same region or not, log distance and log distance interacted with the same region indicator variable as well as various fixed effects. Standard errors clustered at at the district-pair level. * significant at 10 percent, ** 5, *** 1.

Table 9: Endowment Differences Across and Within Regions

	(1)	(2)	(3)
	Mean Endowment Difference $E[\text{Abs}(\ln AC_a - \ln AC_b)]$		
$1[region_a = region_b]$	-0.407***	-0.139***	-0.0732***
	(0.00454)	(0.00612)	(0.00630)
$\ln distance_{ab}$			0.139***
			(0.00118)
$1[state_a = state_b]$		-0.281***	
		(0.00438)	
Constant	0.616***	0.629***	0.393***
	(0.00109)	(0.00111)	(0.0279)
District FE	No	No	Yes
Observations	89,676	89,676	89,676
R^2	0.024	0.056	0.768

Note: Dependent variable is the mean of the absolute value of the log difference in 9 agro-climatic variables AC between two districts a and b. These variables are the mean, standard deviation, minimum and maximum values of both monthly rainfall and precipitation as well as altitude as described in section 4.2. The independent variables are indicators for whether the two districts are in the same region or state. Column 3 contains additional controls (the log distance between the two districts and district fixed effects). Standard errors clustered at at the district-pair level. * significant at 10 percent, ** 5, *** 1.

E An Explicit Test to Reject Explanations that Involve Identical Preferences Across Regions

One concern is that identical preferences coupled with Engel's law imply that poor consumers would purchase large amounts of the cheapest staple that was available locally. Since regions with land relatively suitable for rice cultivation will have relatively cheap rice, I may spuriously attribute the high rice consumption in these areas to regional taste differences if preferences are identical across regions but price and income effects are highly nonlinear.

In this section, I set out to explicitly dismiss this alternative explanation for my findings. Standard preferences (preferences that are identical across regions and satisfy Engel's law) predict that the relative consumption of rice and wheat should depend only on current relative prices and incomes. If there is misspecification of the demand system, there may appear to be abnormally high relative demand for rice in locations where rice is the cheaper calorie source. In contrast, a habit formation explanation predicts that areas with a habit stock that favors rice should have higher relative rice consumption compared to areas with a habit stock that favors wheat, conditional on current relative prices and real income. This relationship should hold even in areas with habit stocks that favor rice, but where wheat is currently a less expensive calorie source. (These are the unusual regions where either the strength of habits is above the no-comparative-advantage-reversal threshold, or where there have been recent shocks to relative prices that fall outside of the model. These regions provide the most convincing evidence for habits that are driven by historic relative prices. Of course, if all regions had habit stocks that favored the relatively expensive food sources, I would not find evidence for proposition 4.)

I focus on consumption of only two goods, wheat and rice, for three reasons: 1) these two goods are the dominant staple cereals in India and so we should expect similar income elasticities of demand in the standard case where preferences are identical across India. 2) I can obtain better measures of land suitability for these two crops. 3) There is substantial overlap in relative prices between areas with good wheat-growing land and areas with good rice-growing land.

I regress the caloric share from rice on functions of prices and incomes, as well as a proxy for the habit stock:

$$\frac{c_{ri}}{c_{ri}+c_{wi}} = \alpha_1 1[p_{ri} \geq p_{wi}]1[f(\frac{E_{ri}}{E_{wi}}) \geq E] + \alpha_2 1[p_{ri} < p_{wi}]1[f(\frac{E_{ri}}{E_{wi}}) \geq E] + \beta_1 1[p_{ri} \geq p_{wi}]1[f(\frac{E_{ri}}{E_{wi}}) < E] + \beta_2 1[p_{ri} < p_{wi}]1[f(\frac{E_{ri}}{E_{wi}}) < E] + z(p_{ri}, p_{wi}, \frac{m_i}{P_i}, Z_i) + \varepsilon_i, \tag{26}$$

where c_{ri} and c_{wi} are the caloric intake from rice and wheat respectively and $1[x \geq y]$ denotes an indicator variable that takes the value of 1 if $x \geq y$ and 0 otherwise. As in section 3.3 of the original draft, i indexes rural households, p_{ri} and p_{wi} are the village median prices per calorie for rice and wheat, m is per capita household expenditure, $\ln P$ is a region-level Stone Price index over all food purchases and Z are household characteristic controls. Finally $f(\frac{E_{ri}}{E_{wi}})$ is a measure of the relative suitability of the region for growing rice vis a vis wheat. This measure aims to proxy for the whole past history of relative prices going back many generations and hence the current habit stock.

I use a discrete measure of relative endowments. I code a region as a rice loving region if the relative endowment measure is above some cutoff E. Standard preferences with misspecification predict that $\alpha_1 < \alpha_2$, $\alpha_1 = \beta_1$, $\alpha_1 < \beta_2$, $\alpha_2 > \beta_1$, $\alpha_2 = \beta_2$ and $\beta_1 < \beta_2$ (relatively cheap calorie sources appear to be overconsumed). Habit formation predicts that $\alpha_1 = \alpha_2$, $\alpha_1 > \beta_1$, $\alpha_1 > \beta_2$, $\alpha_2 > \beta_1$, $\alpha_2 > \beta_2$ and $\beta_1 = \beta_2$ (foods with a relatively high habit stock tend to be overconsumed even if they are not relatively inexpensive in the current period). Table 10 summarizes these predictions.

I utilize a measure of relative suitability $f(\frac{E_{ri}}{E_{wi}})$ produced by the FAO and the IISA as part of the Global Agro-Ecological Zones project (GAEZ). The GAEZ data contain measures of the relative suitability of each State in India for growing both rice and wheat. The particular measure I use is the "crop suitability index" for rain-fed agriculture using intermediate input usage. The index ranges from 0 (Not suitable) to 1 (very high suitability). I compute the simple difference between the index for rice and the index for wheat in the State as my measure of $f(\frac{E_{ri}}{E_{wi}})$.[25]

[25] These measures are obtained from crop suitability models and detailed agro-climatic data: "Soil suitability classifications are based on knowledge of crop requirements, of prevailing soil conditions, and of applied soil management. In other words, soil suitability procedures quantify to what extent soil conditions match crop requirements under defined input and management circumstances." The GAEZ website

Table 11 reports the regression results. I show results for two values of the suitability cutoff E, and two functional forms for the price and income controls. Column 1 sets E equal to the mean of the relative endowment measures (0.15), and uses the same functional form for $z(p_{ri}, p_{wi}, \frac{m_i}{P_i})$ as in the main paper (log prices and log real income plus household characteristic controls). Column 2 uses the same specification as column 1 but sets E equal to zero. Column 3 uses the same specification as column 1 but includes six additional interactions between the three price and income terms (including quadratic terms). Column 4 uses the same specification as column 1 but replaces the caloric share with the household expenditure share on rice $\frac{p_r c_r}{p_r c_r + p_w c_w}$.

Table 10 reports the 6 pairwise hypothesis tests that the coefficients α_1, α_2, β_1 and β_2 are equal to each other. Across all four specifications, and all 6 pairwise tests, I reject the hypothesis that the coefficients are equal to each other (with one exception,$\beta_1 = \beta_2$ in column 3). In all four specifications, the ordering of the coefficients on wheat and rice regions is inconsistent with all 4 of the standard preference inequalities. In contrast the 4 habit formation inequalities are satisfied in each specification. The sharpest evidence against a misspecification story comes from the sign of $\alpha_1 - \beta_2$. I find that after conditioning on prices and income, relative rice consumption is higher in rice-suitable regions than wheat suitable-regions even when rice is the relatively expensive calorie source in the rice-suitable regions and the relatively cheap calorie source in the wheat-suitable regions.

In conclusion, foods for which a region has a habit stock for due to historic comparative advantage tend to be overconsumed even when they are no longer relatively inexpensive. This finding is summarized in figure 9. I plots the locally weighted polynomial of relative rice consumption against relative rice prices for regions above and below the mean relative rice endowment. There is substantial overlap in relative prices across the two curves. Households in States with relatively high suitability for rice cultivation have their demand curves shifted upwards as predicted, even in the range where $p_{ri} \geq p_{wi}$.

As in section 5.2 of the main body of the paper, I also aggregate over regions and explore

---

http://www.iiasa.ac.at/Research/LUC/GAEZv3.0/ contains further details. For rice, my suitability measure is the maximum of the two state-level index values for wetland and dryland rice cultivation.

the caloric impacts of price changes. Standard preferences (preferences that are identical across regions and satisfy Engel's law) predict that the caloric change due to a rise in the relative price of rice will be smaller if rice is the relatively cheap calorie source. A habit formation story predicts that the caloric change due to a rise in the relative price of rice will be smaller if there is a habit stock that favors rice.

I regress the proportional change in the total caloric intake from rice and wheat on relative price changes and changes in real income. In a similar manner to the specification above, I interact the relative price change with indicator variables for rice being the cheaper calorie source initially and the region being relatively suitable for rice cultivation. As in the main paper, I look at the caloric change across the regions of India and replace all household level variables with regional weighted averages:

$$\begin{aligned}\Delta \ln(c_r + c_w) = {} & \alpha_1 1[p_{ri} \geq p_{wi}] 1[f(\frac{E_{ri}}{E_{wi}}) \geq E] \Delta \ln \frac{p_r}{p_w} + \alpha_2 1[p_{ri} < p_{wi}] 1[f(\frac{E_{ri}}{E_{wi}}) \geq E] \Delta \ln \frac{p_r}{p_w} \\ & + \beta_1 1[p_{ri} \geq p_{wi}] 1[f(\frac{E_{ri}}{E_{wi}}) < E] \Delta \ln \frac{p_r}{p_w} + \beta_2 1[p_{ri} < p_{wi}] 1[f(\frac{E_{ri}}{E_{wi}}) < E] \Delta \ln \frac{p_r}{p_w} \\ & + a_0 + \Delta \ln \frac{m}{P} + \varepsilon_i. \qquad (27)\end{aligned}$$

Standard preferences with misspecification predict that $\alpha_1 > \alpha_2$, $\alpha_1 = \beta_1$, $\alpha_1 > \beta_2$, $\alpha_2 < \beta_1$, $\alpha_2 = \beta_2$ and $\beta_1 > \beta_2$ (caloric intake increases less when relatively cheap calorie sources rise in relative price). Habit formation predicts that $\alpha_1 = \alpha_2$, $\alpha_1 < \beta_1$, $\alpha_1 < \beta_2$, $\alpha_2 < \beta_1$, $\alpha_2 < \beta_2$ (caloric intake increases less when foods with high habit stocks rise in relative price). Once again, the hypothesis tests on the differences between coefficients are shown in table 12, and the regression coefficients are reported in table 13.

In the main paper, I explore caloric changes between 1987-88 and 2004-05. As I am attempting estimate the coefficient on relative prices for four mutually-exclusive subsamples with only 76 regional observations per survey, I incorporate multiple survey rounds in this analysis. Columns 1 through 4 show the results with the caloric change over 1 survey period, 2 survey periods, 3 survey periods and 4 survey periods respectively. In all cases I use the same threshold for the relative endowment as in the baseline specification above.[26]

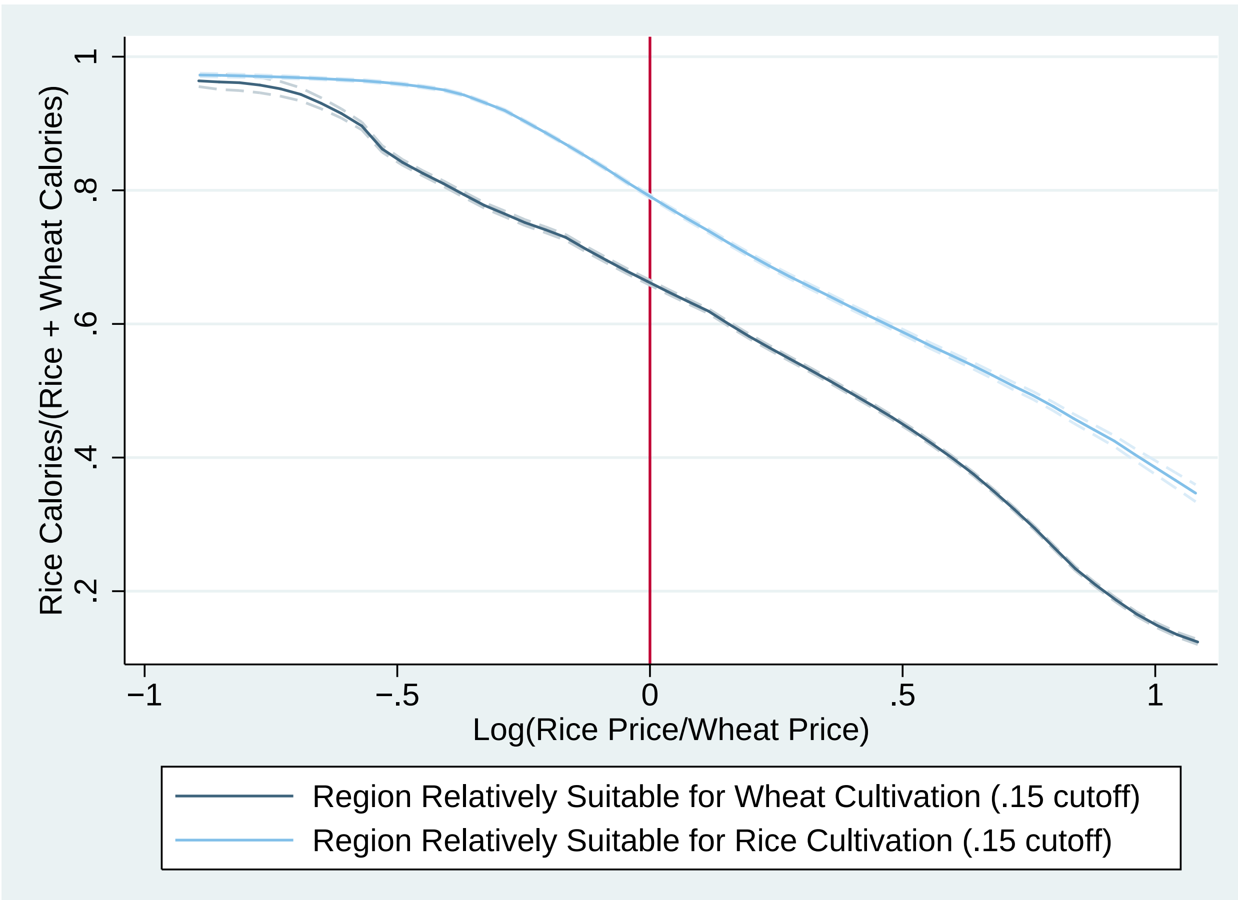
[26] I use $E = 0.15$, the mean value of the relative endowment from the household analysis. I cannot use

Once more, the ordering of the coefficients is supportive of a habit formation story rather than a standard preferences with misspecification story. There are 3 inequalities unique to the misspecification story, and 3 unique to the habit formation story. Across the 4 specifications, 11 out of the 12 unique inequalities hold for the habit formation story, and only 5 out of 12 hold for the misspecification story. The 4 inequalities common across both stories are also present in the data. Unfortunately, consistent with the small number of regions, many of these differences are not significant at the 5 percent level. However, the prediction that $\alpha_1 < \beta_1$ is inconsistent with a standard preference story with misspecification and is significant in 3 of 4 specifications.

In conclusion, I do not find evidence that misspecification of the demand system combined with Engel's law can explain my findings that there are regional taste differences related to historic endowments, and that these taste differences have caloric impacts at the time of price changes.

a value of $E = 0$ as there are no regions with $[f(\frac{E_{ri}}{E_{wi}}) < 0]$ and $p_{ri} < p_{wi}$.

Figure 9: Relative Rice Demand, Relative Rice Suitability and Relative Rice Prices



Note: Locally weighted polynomial of rice calorie share and relative rice price. Epanechnikov kernel with a bandwidth of 0.25. 95% confidence intervals shown by dashed lines. Relative suitability $f(\frac{E_{ri}}{E_{wi}})$ is defined as the GAEZ rice suitability index minus the GAEZ wheat suitability index at the state level. Top and bottom 1 percent of relative prices not shown.

Table 10: Relative Rice Demand, Relative Rice Suitability and Relative Rice Prices – Coefficient Tests

			(1)		(2)		(3)		(4)	
Difference	Engel Mis-specification	Habit Formation	Cal Share, $E = 0.15$		Cal Share, $E = 0$		Cal Share, $E = 0.15$		Exp Share, $E = 0.15$	
			Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value
$\alpha_1 - \alpha_2$	<0	0	0.044***	0.000	0.004	0.714	-0.036***	0.007	0.065***	0.000
$\alpha_1 - \beta_1$	0	>0	0.092***	0.000	0.332***	0.000	0.077***	0.000	0.099***	0.000
$\alpha_1 - \beta_2$	<0	>0	0.139***	0.000	0.654***	0.000	0.075***	0.000	0.159***	0.000
$\alpha_2 - \beta_1$	>0	>0	0.048***	0.000	0.328***	0.000	0.113***	0.000	0.034***	0.007
$\alpha_2 - \beta_2$	0	>0	0.095***	0.000	0.650***	0.000	0.111***	0.000	0.094***	0.000
$\beta_1 - \beta_2$	<0	0	0.047**	0.019	0.322***	0.000	-0.002	0.917	0.060***	0.002

Note: Summary of hypothesis tests using coefficients reported in table 11. * significant at 10 percent, ** 5, *** 1.

Table 11: Relative Rice Demand, Relative Rice Suitability and Relative Rice Prices – Regression Results

(1987-88 Cross Section)	(1) Caloric Share $E = 0.15$	(2) Caloric Share $E = 0$	(3) Caloric Share, $E = 0.15$	(4) Exp. Share, $E = 0.15$
$1[p_{ri} \geq p_{wi}]1[f(\frac{E_{ri}}{E_{wi}}) \geq E]$	0.957***	1.047***	1.023***	0.937***
	(0.0215)	(0.0210)	(0.0577)	(0.0212)
$1[p_{ri} < p_{wi}]1[f(\frac{E_{ri}}{E_{wi}}) \geq E]$	0.913***	1.043***	1.058***	0.872***
	(0.0223)	(0.0223)	(0.0570)	(0.0220)
$1[p_{ri} \geq p_{wi}]1[f(\frac{E_{ri}}{E_{wi}}) < E]$	0.865***	0.715***	0.946***	0.838***
	(0.0222)	(0.0213)	(0.0587)	(0.0219)
$1[p_{ri} < p_{wi}]1[f(\frac{E_{ri}}{E_{wi}}) < E]$	0.818***	0.393***	0.948***	0.778***
	(0.0269)	(0.0351)	(0.0588)	(0.0266)
$\ln p_{ri}$	-0.483***	-0.459***	0.201***	-0.465***
	(0.0187)	(0.0185)	(0.0773)	(0.0186)
$\ln p_{wi}$	0.601***	0.564***	0.401***	0.554***
	(0.0225)	(0.0218)	(0.0698)	(0.0219)
$\ln \frac{m_i}{P_i}$	-0.0493***	-0.0647***	-0.102***	-0.0361***
	(0.00519)	(0.00502)	(0.0335)	(0.00515)
Observations	80,414	80,414	80,414	80,414
R^2	0.819	0.824	0.830	0.832
Add p and $\frac{m}{P}$ interactions	No	No	Yes	No
Hhold characteristic controls	Yes	Yes	Yes	Yes

Note: Dependent variable is the caloric share from rice (columns 1-3) or the expenditure share from rice (column 4) for rural Indian households in 1987-88. Dependent variables are interactions of indicator variables for whether the village median wheat price is lower than the village median rice price and indicator variables for whether the region has agro-climatic conditions relatively suited to rice or wheat cultivation. Relative suitability $f(\frac{E_{ri}}{E_{wi}})$ is defined as the GAEZ rice suitability index minus the GAEZ wheat suitability index at the state level. Price and income controls as well as household characteristics are as in the taste estimation specification. Real income is deflated by a regional Stone Price index over the 52 foods. Additional p and $\frac{m}{P}$ interactions are $\ln p_{ri} \ln p_{ri}$, $\ln p_{ri} \ln p_{wi}$, $\ln p_{wi} \ln p_{wi}$, $\ln \frac{m_i}{P_i} \ln \frac{m_i}{P_i}$, $\ln \frac{m_i}{P_i} \ln p_{ri}$ and $\ln \frac{m_i}{P_i} \ln p_{wi}$. Regressions weighted by household survey weights in 1987-88. Robust standard errors. * significant at 10 percent, ** 5, *** 1.

Table 12: Caloric Changes, Relative Rice Suitability and Relative Rice Prices – Coefficient Tests

Difference	Engel Mis-specification	Habit Formation	(1) 1 Period Changes Estimate	(1) 1 Period Changes P-Value	(2) 2 Period Changes Estimate	(2) 2 Period Changes P-Value	(3) 3 Period Changes Estimate	(3) 3 Period Changes P-Value	(4) 4 Period Changes Estimate	(4) 4 Period Changes P-Value
$\alpha_1 - \alpha_2$	>0	0	-0.02	0.89	-0.01	0.96	0.06	0.82	-0.36	0.49
$\alpha_1 - \beta_1$	0	<0	-0.07	0.58	-0.32***	0.01	-0.51***	0.00	-0.47***	0.01
$\alpha_1 - \beta_2$	>0	<0	-0.06	0.72	-0.27	0.25	-0.15	0.86	-0.20	0.79
$\alpha_2 - \beta_1$	<0	<0	-0.05	0.73	-0.31**	0.05	-0.57**	0.04	-0.11	0.84
$\alpha_2 - \beta_2$	0	<0	-0.04	0.82	-0.26	0.30	-0.21	0.81	0.16	0.86
$\beta_1 - \beta_2$	>0	0	0.01	0.96	0.05	0.85	0.36	0.66	0.26	0.73

Note: Summary of hypothesis tests using coefficients reported in table 13. * significant at 10 percent, ** 5, *** 1.

Table 13: Caloric Changes, Relative Rice Suitability and Relative Rice Prices – Regression Results

	(1) 1 Period Δ $\Delta\ln(c_r+c_w)$	(2) 2 Period Δ $\Delta\ln(c_r+c_w)$	(3) 3 Period Δ $\Delta\ln(c_r+c_w)$	(4) 4 Period Δ $\Delta\ln(c_r+c_w)$
$1[p_{ri} \geq p_{wi}]1[f(\frac{E_{ri}}{E_{wi}}) \geq E]\Delta\ln\frac{p_r}{p_w}$	-0.0116	0.0246	0.0270	0.0235
	(0.0982)	(0.0842)	(0.0973)	(0.113)
$1[p_{ri} < p_{wi}]1[f(\frac{E_{ri}}{E_{wi}}) \geq E]\Delta\ln\frac{p_r}{p_w}$	0.00874	0.0313	-0.0343	0.385
	(0.117)	(0.130)	(0.256)	(0.530)
$1[p_{ri} \geq p_{wi}]1[f(\frac{E_{ri}}{E_{wi}}) < E]\Delta\ln\frac{p_r}{p_w}$	0.0592	0.341***	0.534***	0.490***
	(0.0922)	(0.0978)	(0.133)	(0.164)
$1[p_{ri} < p_{wi}]1[f(\frac{E_{ri}}{E_{wi}}) < E]\Delta\ln\frac{p_r}{p_w}$	0.0501	0.295	0.175	0.228
	(0.142)	(0.225)	(0.817)	(0.749)
$\Delta\ln\frac{m_i}{P_i}$	0.0583***	0.0361***	0.0633***	-0.00105
	(0.00971)	(0.00921)	(0.0179)	(0.0415)
Constant	0.120***	0.173***	0.437***	0.0882
	(0.0204)	(0.0390)	(0.110)	(0.317)
Observations	299	222	147	71
R^2	0.116	0.130	0.176	0.128

Note: Dependent variable is the log change in the regional caloric intake per capita from rice and wheat. Dependent variables are interactions of indicator variables for whether the initial regional wheat price is lower than the initial regional rice price and indicator variables for whether the region has agro-climatic conditions relatively suited to rice or wheat cultivation. Relative suitability $f(\frac{E_{ri}}{E_{wi}})$ is defined as the GAEZ rice suitability index minus the GAEZ wheat suitability index at the state level. Five sample periods included (1983, 1987-88, 1993-94, 1999-2000 and 2004-05). Additional control included for the log change in regional expenditure per capita deflated by a regional Stone Price index over the 52 foods. Regressions weighted by a region's initial total survey weight Robust standard errors. * significant at 10 percent, ** 5, *** 1.

F Data Sources

The NSS data used in both empirical sections of the paper are described in section 3.1. The full set of 52 foods in the large sample are: rice, wheat, jowar, bajra, maize, barley, small millets, ragi, gram, cereal substitutes, arhar, moong, masur, urd, peas, soyabean, khesari, milk products, vanaspati margarine, mustard oil, groundnut oil, coconut oil, other oil, meat, chicken and eggs, fish, potato, onion, other vegetables, other fruit, sugar, other spices, other nuts, other pulses, sweet potato, garlic, ginger, chillis, turmeric, black pepper, coconuts, banana, mango, pan and supari, oranges, cauliflower, cabbage, brinjal, lady finger, tomato, lemon and guava. The larger grouping only omits processed foods and beverages that constitute less than 2 percent of caloric intake. For these goods, it is impossible to match the good to endowment measures, or to obtain accurate quantity or caloric data. However, results are robust to including these goods and using recorded quantities and NSS calorie approximations.

To measure agricultural endowments, I use district-level agricultural data from Indian Harvest produced by the Centre for Monitoring Indian Economy, aggregated up to NSS regions. Further regional data come from the Indian District Database (Vanneman and Barnes, 2000) and the India Agriculture and Climate Data Set (Sanghi et al., 1998), while weather data come from Willmott and Matsuura (2001). Finally, as discussed in section E, I obtain the relative suitability of each State in India for growing 11 of the 17 staple foods from GAEZ data collected by the FAO.

G Robustness Results and Additional Tables

G.1 Robustness of Regional Caloric Change Regression

There are several concerns regarding the parameters estimated by running regression 8 in the main paper. If households reduce non-food expenditure in response to rising prices for more favored foods, the caloric decline will be tempered. Table 27 shows the results of rerunning regression 8, but replacing $\Delta \ln food_r$ with the change in total expenditure on all goods, $\Delta \ln expenditure_r$. The magnitude of the caloric reduction coming from tastes correlating with price changes declines by about half as expenditure is partially reallocated towards food. However, conditional upon total expenditure, caloric intake still declines with the correlation between tastes and price changes.

As a further robustness check, I instrument for $\Delta \ln food_r$ with the log change in non-food expenditure, $\Delta \ln non\ food_r$. A shock that increases the demand for calories, such as changing work patterns, will also affect food expenditure and result in a positive correlation between $\Delta \ln food_r$ and the error term, biasing b_4 upwards. However, there will be a negative or zero correlation with $\Delta \ln non\ food_r$, and the true value of b_4 will be bounded between the instrumented and uninstrumented estimates. These results are also shown in table 27, and b_1 is essentially unchanged in the two specifications, implying that the endogeneity of food expenditure is not a major problem.

Finally, I replace $\sum_g (\theta_{gr} - \overline{\theta}_g) \Delta \ln p_{gr}$ with the correlation between $\Delta \ln p_{gr}$ and tastes that have been normalized across either goods or regions, or the rank of the taste coefficient across goods or regions.

Figure 10: Counterfactual Results for Different Hypothetical Free Trade Prices
(All-India Means for Staple Foods (2004-2005), comparable to Table 4 columns 1-3)

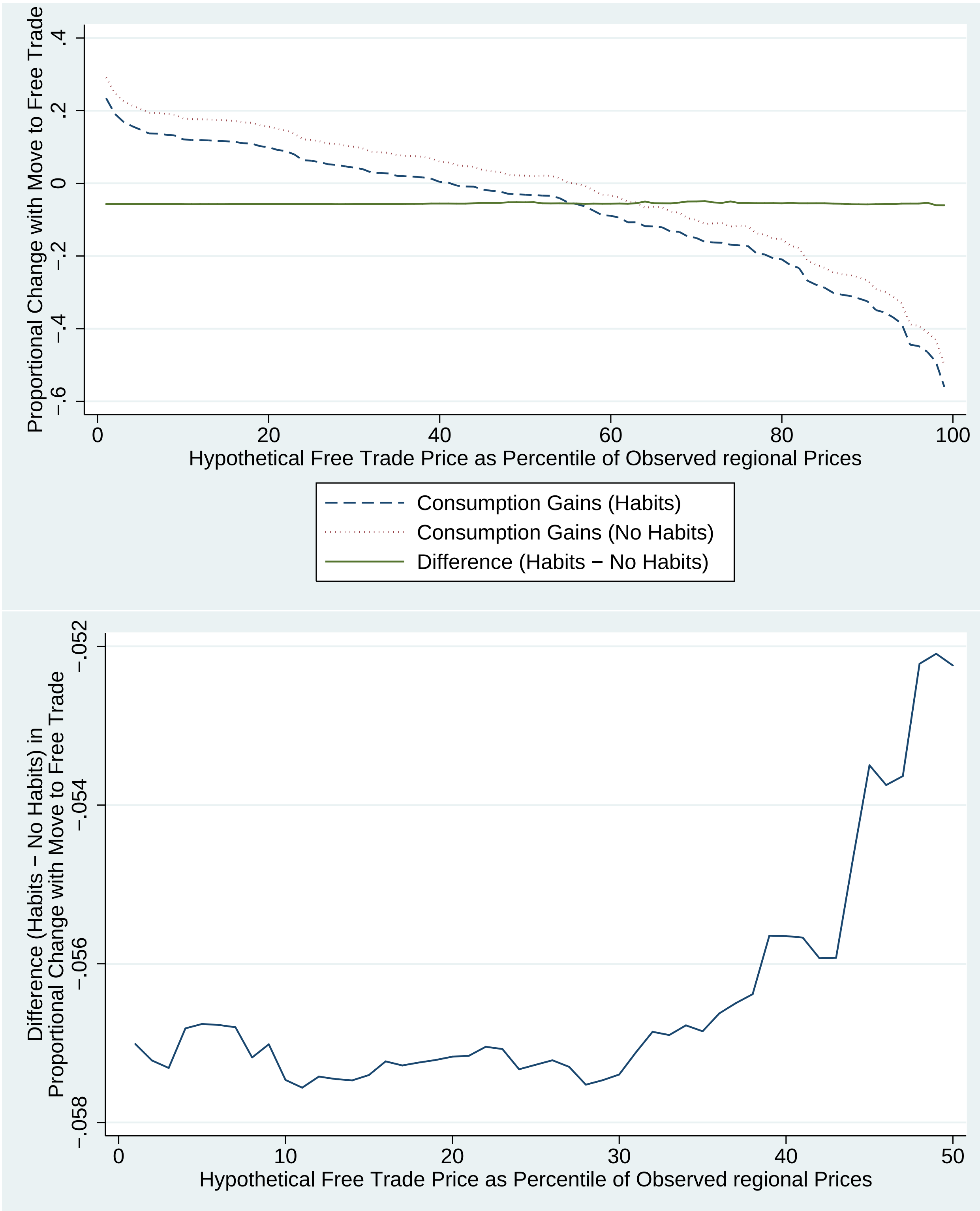


Table 14: The Significance of Lagged Prices in the Demand System (Staple Foods)

F-Test that all	2 Lagged Regional Price Terms				3 Lagged Regional Price Terms					
lagged price	$t-1$ Region Prices		$t-2$ Region Prices		$t-1$ Region Prices		$t-2$ Region Prices		$t-3$ Region Prices	
terms equal zero	F-stat	P-value	F-stat	P-value	F-stat	P-value	F-stat	P-value	F-stat	P-value
Rice	3.747	0.000	7.932	0.000	2.524	0.001	1.940	0.011	3.277	0.000
Wheat	7.990	0.000	4.468	0.000	4.900	0.000	4.402	0.000	3.403	0.000
Jowar	9.295	0.000	5.887	0.000	3.888	0.000	4.436	0.000	3.123	0.000
Bajra	3.348	0.000	3.245	0.000	3.002	0.000	2.427	0.001	1.706	0.035
Maize	3.025	0.000	4.406	0.000	4.309	0.000	2.195	0.003	2.924	0.000
Barley	1.762	0.027	1.599	0.056	1.059	0.389	0.751	0.752	1.225	0.234
Small Millets	1.666	0.041	0.798	0.697	2.585	0.000	3.259	0.000	3.392	0.000
Ragi	3.411	0.000	4.196	0.000	3.307	0.000	4.547	0.000	4.322	0.000
Gram	14.740	0.000	6.875	0.000	3.978	0.000	3.918	0.000	5.843	0.000
Cereal Substitutes	3.760	0.000	5.298	0.000	2.731	0.000	3.086	0.000	1.858	0.017
Arhar	10.884	0.000	9.823	0.000	4.425	0.000	4.203	0.000	3.209	0.000
Moong	7.465	0.000	5.824	0.000	4.946	0.000	5.207	0.000	4.332	0.000
Masur	12.074	0.000	10.194	0.000	6.344	0.000	4.669	0.000	5.552	0.000
Urd	11.379	0.000	17.714	0.000	4.363	0.000	2.903	0.000	2.749	0.000
Peas	17.016	0.000	10.627	0.000	10.948	0.000	7.331	0.000	6.083	0.000
Soyabean	4.039	0.000	6.982	0.000	4.932	0.000	5.750	0.000	5.210	0.000
Khesari	5.527	0.000	4.105	0.000	3.979	0.000	3.865	0.000	4.636	0.000

Note: The F-tests are for the null that historic region prices do not predict current budget shares once contemporaneous prices have been controlled for in the demand system shown in equation 5 of the main paper. Region prices $\ln p_{gr}$ are logs of weighted regional means of village median unit values. t denotes consecutive NSS thick survey rounds 1983, 1987-88, 1993-94, 1999-2000 and 2004-05. The three lag term specification estimates demands in 1999-2000 and 2004-05, while the two lag specification also includes 1993-94 demands. F-tests distributed F(17,19351) and F(17,12533).

Table 15: The Significance of Lagged Prices in the Demand System (All Foods)

	2 Lagged Price Terms				3 Lagged Price Terms					
	$t-1$ Prices		$t-2$ Prices		$t-1$ Prices		$t-2$ Prices		$t-3$ Prices	
Price coefs=0	F-stat	P-val	F-stat	P-val	F-stat	P-val	F-stat	P-val	F-stat	P-val
Rice	8.07	0.00	13.87	0.00	34.93	0.00	71.57	0.00	33.05	0.00
Wheat	9.80	0.00	9.19	0.00	29.47	0.00	20.17	0.00	24.53	0.00
Jowar	5.26	0.00	5.79	0.00	12.07	0.00	15.93	0.00	10.92	0.00
Bajra	3.24	0.00	3.37	0.00	7.19	0.00	7.05	0.00	6.10	0.00
Maize	4.36	0.00	4.54	0.00	8.37	0.00	8.19	0.00	5.93	0.00
Barley	1.02	0.43	1.16	0.20	1.01	0.45	0.91	0.63	0.95	0.56
Small Millets	1.49	0.01	1.43	0.02	2.91	0.00	2.87	0.00	2.81	0.00
Ragi	4.64	0.00	3.90	0.00	26.92	0.00	22.08	0.00	27.05	0.00
Gram	9.15	0.00	7.65	0.00	10.38	0.00	14.23	0.00	9.69	0.00
Cereal Subs.	2.69	0.00	3.32	0.00	7.05	0.00	8.36	0.00	6.17	0.00
Arhar	4.60	0.00	5.21	0.00	25.13	0.00	30.54	0.00	18.40	0.00
Moong	4.29	0.00	4.88	0.00	9.28	0.00	12.17	0.00	9.35	0.00
Masur	5.59	0.00	8.23	0.00	27.59	0.00	15.33	0.00	14.72	0.00
Urd	4.86	0.00	6.56	0.00	8.11	0.00	13.84	0.00	13.39	0.00
Peas	8.62	0.00	8.16	0.00	8.05	0.00	10.02	0.00	9.97	0.00
Soyabean	3.11	0.00	4.15	0.00	7.98	0.00	8.27	0.00	7.93	0.00
Khesari	4.51	0.00	4.54	0.00	4.78	0.00	7.40	0.00	4.94	0.00
Milk Products	17.15	0.00	20.40	0.00	16.41	0.00	40.45	0.00	27.51	0.00
Vanaspati	7.26	0.00	6.65	0.00	13.34	0.00	13.09	0.00	13.19	0.00
Mustard Oil	137.08	0.00	135.42	0.00	42.93	0.00	33.24	0.00	36.71	0.00
Groundnut Oil	81.30	0.00	73.48	0.00	20.30	0.00	31.37	0.00	19.74	0.00
Coconut Oil	236.90	0.00	220.78	0.00	51.62	0.00	58.59	0.00	49.49	0.00
Other Oil	67.77	0.00	57.59	0.00	27.87	0.00	18.31	0.00	28.71	0.00
Meat	6.61	0.00	6.72	0.00	11.54	0.00	11.51	0.00	13.42	0.00
Chicken/Eggs	8.77	0.00	7.13	0.00	10.86	0.00	15.97	0.00	12.03	0.00
Fish	5.93	0.00	5.77	0.00	31.41	0.00	34.27	0.00	32.79	0.00

	2 Lagged Price Terms				3 Lagged Price Terms					
	$t-1$ Prices		$t-2$ Prices		$t-1$ Prices		$t-2$ Prices		$t-3$ Prices	
Price coefs=0	F-stat	P-val	F-stat	P-val	F-stat	P-val	F-stat	P-val	F-stat	P-val
Potato	9.29	0.00	7.97	0.00	17.61	0.00	24.05	0.00	22.29	0.00
Onion	6.46	0.00	6.64	0.00	8.45	0.00	6.35	0.00	6.33	0.00
Other Veg	6.14	0.00	6.62	0.00	7.60	0.00	9.67	0.00	7.31	0.00
Other Fruit	4.45	0.00	3.72	0.00	3.81	0.00	3.18	0.00	3.23	0.00
Sugar	6.89	0.00	6.22	0.00	20.59	0.00	23.70	0.00	16.88	0.00
Other Spices	6.91	0.00	8.14	0.00	17.38	0.00	14.16	0.00	10.47	0.00
Other Nuts	3.84	0.00	4.29	0.00	5.81	0.00	8.90	0.00	7.28	0.00
Other Pulses	15.33	0.00	16.55	0.00	6.59	0.00	7.21	0.00	6.60	0.00
Sweet Potato	2.01	0.00	1.94	0.00	3.87	0.00	3.58	0.00	3.86	0.00
Garlic	6.71	0.00	6.75	0.00	6.77	0.00	7.28	0.00	6.88	0.00
Ginger	6.20	0.00	7.40	0.00	5.51	0.00	11.61	0.00	6.00	0.00
Chili	9.01	0.00	6.20	0.00	12.44	0.00	20.89	0.00	13.58	0.00
Turmeric	9.27	0.00	6.45	0.00	13.94	0.00	10.24	0.00	9.22	0.00
Black Pepper	6.07	0.00	8.05	0.00	9.60	0.00	8.33	0.00	9.51	0.00
Coconuts	5.03	0.00	4.46	0.00	62.00	0.00	89.48	0.00	63.49	0.00
Banana	3.59	0.00	3.85	0.00	3.04	0.00	4.96	0.00	4.49	0.00
Mango	1.76	0.00	1.98	0.00	1.29	0.11	1.24	0.16	0.92	0.61
Pan/Supari	5.27	0.00	5.61	0.00	23.20	0.00	29.06	0.00	19.39	0.00
Oranges	2.51	0.00	2.94	0.00	2.09	0.00	2.18	0.00	1.94	0.00
Cauliflower	2.65	0.00	2.60	0.00	3.48	0.00	5.13	0.00	3.28	0.00
Cabbage	3.08	0.00	4.27	0.00	5.56	0.00	4.50	0.00	5.80	0.00
Brinjal	3.05	0.00	3.52	0.00	9.50	0.00	13.06	0.00	9.21	0.00
Lady Finger	2.59	0.00	2.77	0.00	4.72	0.00	2.62	0.00	3.71	0.00
Tomato	4.24	0.00	6.87	0.00	6.33	0.00	5.95	0.00	8.61	0.00
Lemon	5.08	0.00	5.05	0.00	4.90	0.00	9.11	0.00	5.00	0.00
Guava	2.11	0.00	2.09	0.00	2.27	0.00	2.06	0.00	3.10	0.00

Note: The F-tests are for the null that historic region prices do not predict current budget shares once contemporaneous prices have been controlled for in the demand system shown in equation 5 of the main paper. Region prices $\ln p_{gr}$ are logs of weighted regional means of village median unit values. t denotes consecutive NSS thick survey rounds 1983, 1987-88, 1993-94, 1999-2000 and 2004-05. The three lag term specification estimates demands in 1999-2000 and 2004-05, while the two lag specification also includes 1993-94 demands. F-tests distributed F(52,19316) and F(52,12512).

Table 16: Contemporary Tastes and Past Prices: Alternative Specifications

LHS: θ_{grt}	(1) Alternative Spec.	(2) Unweighted Staple Foods	(3) Urban	(4) Only Cereals	(5) Top 20 Foods	(6) Random $\frac{1}{2}$ of Foods	(7) Other $\frac{1}{2}$ of Foods
$\ln p_{gr,t}$	0.00785*	0.00307	-0.00430	-0.0315*	-0.0982***	0.00745	-0.0218***
	(0.00439)	(0.00432)	(0.00782)	(0.0181)	(0.0138)	(0.00697)	(0.00648)
$\ln p_{gr,t-1}$	0.00833**	0.00659*	-0.0203**	-0.0434***	-0.118***	-0.0364***	-0.0269***
	(0.00375)	(0.00391)	(0.00848)	(0.0139)	(0.0105)	(0.00689)	(0.00383)
$\ln p_{gr,t-2}$	-0.0140***	-0.0167***	0.00856	-0.0446***	-0.0534***	-0.000235	-0.00702***
	(0.00395)	(0.00404)	(0.00693)	(0.0132)	(0.0146)	(0.00759)	(0.00152)
$z_g(\mathbf{p}_{r,t-1}, \frac{food_{r,t-1}}{P^*_{r,t-1}}, Z_{r,t-1})$	0.0355						
	(0.0449)						
$z_g(\mathbf{p}_{r,t-2}, \frac{food_{r,t-2}}{P^*_{r,t-2}}, Z_{r,t-2})$	0.107**						
	(0.0514)						
$\theta_{gr,t-2}$	0.824***						
	(0.0214)						
Region-Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-Good FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,757	3,774	2,125	2,160	4,320	5,400	5,832
R^2	0.655	0.798	0.736	0.757	0.888	0.784	0.553

Note: Dependent variable θ_{grt} is the taste coefficient, estimated using unexplained regional variation in food budget shares. $\ln p_{gr,t}$ are weighted means of median village prices. Column 1 includes three additional regressors that replace the region-good fixed effect in an alternative specification to the Cobb-Douglas parametrization highlighted in footnote 28: the explained component of the budget share in the two previous periods and the taste stock in the earliest period. Regressions weighted by survey population weights for the 76 regions of India except column 2. Urban sample in column 3 includes only 53 regions where surveys administered in every round. Columns 4 to 7 use alternative subsamples of foods. Robust standard errors clustered at the region-good level. * significant at 10 percent, ** 5, *** 1.

Table 17: Contemporary Tastes and Past Prices: Alternative Taste Estimates

LHS: θ_{grt} (Staple Foods)	(1) OLS Prices	(2) Non-Food IV	(3) HHold P Interactions	(4) Quadratic P and m	(5) No Caste /Religion
$\ln p_{gr,t}$	0.00847**	-0.00834	-0.345	-0.0320**	-0.0169
	(0.00381)	(0.0134)	(0.343)	(0.0138)	(0.0128)
$\ln p_{gr,t-1}$	0.00175	-0.0198*	0.208	-0.0384***	-0.0298***
	(0.00408)	(0.0108)	(0.332)	(0.0117)	(0.0110)
$\ln p_{gr,t-2}$	-0.00859**	-0.0177	-0.575*	-0.0466***	-0.0402***
	(0.00437)	(0.0117)	(0.306)	(0.0115)	(0.0101)
Region-Time & Region Good FE	Yes	Yes	Yes	Yes	Yes
Observations	3,672	3,672	3,672	3,672	3,672
R^2	0.947	0.829	0.422	0.565	0.660

Note: Dependent variable θ_{grt} is the taste coefficient, estimated using unexplained regional variation in food budget shares. $\ln p_{gr,t}$ are weighted means of median village prices. Regressions weighted by survey population weights for the 76 regions of India. Different columns represent different specifications for taste estimates. Column 1 does not use nearby village prices as instrument for local prices. Column 2 instruments food expenditure with other expenditures. Column 3 interacts own prices with household characteristics. Column 4 include quadratic price and food expenditure terms. Column 5 excludes caste and religion controls. Robust standard errors clustered at the region-good level. * significant at 10 percent, ** 5, *** 1.

Table 18: Contemporary Tastes and Past Prices: Alternative Price Measures

LHS: θ_{grt} (Staple Foods)	(1) Mean Prices	(2) 25th Percentile	(3) 75th Percentile	(4) Transport Cost 1	(5) Transport Cost 2
$\ln p_{gr,t}$	-0.0125	0.00831	-0.0390**	-0.0101	-0.0208*
	(0.00989)	-0.00599	(0.0167)	(0.0121)	(0.0119)
$\ln p_{gr,t-1}$	-0.0196**	0.00398	-0.0362**	-0.0303***	-0.0367***
	(0.00847)	-0.0059	(0.0148)	(0.0105)	(0.0105)
$\ln p_{gr,t-2}$	-0.0191**	-0.0216***	-0.0344**	-0.0422***	-0.0477***
	(0.00878)	-0.00643	(0.0173)	(0.00966)	(0.00940)
Region-Time & Region Good FE	Yes	Yes	Yes	Yes	Yes
Observations	3,774	3,774	3,774	3,774	3,774
R^2	0.853	0.610	0.891	0.685	0.666

Note: Dependent variable θ_{grt} is the taste coefficient, estimated using unexplained regional variation in food budget shares. $\ln p_{gr,t}$ are regional prices. Regressions weighted by survey population weights. Various village prices are used instead of the median price that is used in the main specification (mean, 25th percentile and 75th percentile of the reported unit values, a unit price including a 5 percent ad-valorem transport cost when a good is not available locally and so a nearby price used instead, and a unit price including an ad-valorem transport cost based on sugar prices differences). Regional prices are weighted means of these village prices. Robust standard errors clustered at the region-good level. * significant at 10 percent, ** 5, *** 1.

Table 19: Mean Tastes and Relative Endowments Across 76 Regions (1987-88, Staples only)

	Estimated Tastes		Relative Endowment		Log Prices	
Food Item	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Rice	0.2538	0.0335	0.3492	0.0354	0.0449	0.0177
Wheat	0.2252	0.0235	0.1073	0.0165	-0.1432	0.0367
Jowar	0.1531	0.0126	0.0714	0.0120	-0.2242	0.0656
Bajra	-0.0312	0.0078	0.0451	0.0096	-0.4520	0.0472
Maize	0.0667	0.0058	0.0581	0.0130	-0.3649	0.0306
Barley	0.0081	0.0008	0.0092	0.0019	-0.0457	0.0785
Small Millets	0.0227	0.0016	0.0105	0.0036	-0.0111	0.0573
Ragi	0.0267	0.0058	0.0242	0.0065	-0.4415	0.0223
Gram	0.0407	0.0027	0.0347	0.0069	0.6167	0.0151
Cereal Substitutes	0.0045	0.0014	0.0049	0.0027	0.8746	0.0691
Arhar	0.0544	0.0059	0.0131	0.0021	1.0944	0.0179
Moong	0.0253	0.0018	0.0067	0.0018	0.8235	0.0115
Masur	0.0812	0.0028	0.0027	0.0011	0.7812	0.0147
Urd	0.0437	0.0022	0.0094	0.0025	0.7421	0.0158
Peas	0.0061	0.0005	0.0007	0.0003	0.7645	0.0196
Soyabean	-0.0001	0.0003	0.0017	0.0009	0.6690	0.0513
Khesari	0.0192	0.0006	0.0016	0.0012	0.6445	0.0338

Note: Tastes estimated using unexplained regional variation in food budget shares in 1987-88, instrumenting village prices with those in a nearby village. Relative endowment is the portion of regional cropland planted with a particular crop using endowment data from the 1970's. Unweighted means across 76 regions.

Table 20: Tastes, Relative Resource Endowments and Prices: Alternative Instrument Sets and Dependent Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Limited Information Maximum Likelihood									
LHS: θ_{grt} (Staple Foods)	Standard Instruments	Monthly Instruments	Pop. & Wage Instruments	GAEZ Instruments	GAEZ×Item Instruments	Only Precip. Instruments	Mean T/P Instruments	Mean Precip. Instruments	Value of Prod.	Tons of Prod.
$V_{gr}/\sum_{g'} V_{g'r}$	1.218***	1.247***	1.185***	0.670***	0.982***	1.205***	1.161***	1.144***	1.744***	2.148***
	(0.131)	(0.148)	(0.110)	(0.175)	(0.180)	(0.205)	(0.171)	(0.165)	(0.506)	(0.557)
Good FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,275	1,275	901	792	792	1,275	1,275	1,275	1,275	1,292
R^2	0.770	0.761	0.788	0.785	0.814	0.774	0.804	0.789	0.439	0.179
First Stage F	11.81	9.701	11.07	10.67	12.36	11.00	14.39	25.41	8.976	6.210
No. of Instruments	153	136	187	1	11	68	34	17	153	153
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
	Two-Stage Least Squares Estimates									
LHS: θ_{grt} (Staple Foods)	Standard Instruments	Monthly Instruments	Pop. & Wage Instruments	GAEZ Instruments	GAEZ×Item Instruments	Only Precip. Instruments	Mean T/P Instruments	Mean Precip. Instruments	Value of Prod.	Tons of Prod.
$V_{gr}/\sum_{g'} V_{g'r}$	1.010***	1.031***	1.015***	0.645***	0.976***	1.012***	1.064***	1.077***	1.006***	1.209***
	(0.0601)	(0.0653)	(0.0577)	(0.189)	(0.163)	(0.0839)	(0.116)	(0.125)	(0.0761)	(0.0957)
Good FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,275	1,275	901	792	792	1,275	1,275	1,275	1,275	1,292
R^2	0.811	0.809	0.815	0.778	0.814	0.811	0.804	0.802	0.767	0.693
First Stage F	11.81	9.701	11.07	10.67	12.36	11.00	14.39	25.41	8.976	6.210
No. of Instruments	136	136	187	1	11	68	34	17	153	153

Note: Dependent variable, tastes, estimated using unexplained regional variation in food budget shares in 1987-88. Relative endowment is the portion of regional cropland planted with a particular crop using endowment data from the 1970's. Prices are regional weighted means of village median unit values. Endowment columns generally use predicted values of relative endowment from regressing crop shares on 8 crop-specific rainfall and temperature variables and altitude and use limited information maximum likelihood. Column 2 uses 7 monthly rainfall measures suggested by Dev and Evenson (2003) as an alternative (the mean temperature in January, April, July and October and mean rainfall for June, July and August). Column 3 includes two additional endowments, the real agricultural wage and the population density for the subset of regions covered by the India Agriculture and Climate Data Set. Column 4 uses crop suitability index for rain-fed agriculture using intermediate input usage calculated by GAEZ and the level of Indian states. Data only covers 11 of 17 staple foods. Column 5 uses the same instruments but interacts the GAEZ measures with item-specific fixed effects. Column 6 uses only the rainfall measures from my main specification as instruments. Column 7 uses only the mean temperature and mean rainfall measures from my main specification as instruments. Column 8 uses only the mean rainfall measure from my main specification as instruments. Column 9 uses total value instead of area planted. Column 10 uses total production in tons rather than are planted for relative endowment in first stage. Columns 11-19 repeat the specifications above using two stage least squares. Regressions weighted by survey population weights for the 76 regions of India. Robust standard errors clustered at the region level. Constant not reported. * significant at 10 percent, ** 5, *** 1.

Table 21: Tastes, Relative Resource Endowments and Prices: Alternative Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LHS: θ_{grt} (Staple Foods)	Unweighted		No Fixed Effects		Region Fixed Effects		Urban	
$V_{gr}/\sum_{g'} V_{g'r}$	1.057***		1.168***		1.217***		1.276*	
	(0.142)		(0.205)		(0.131)		(0.681)	
$\ln p_{gr}$		-0.034***		-0.043***		-0.027***		-0.017***
		(0.00385)		(0.00292)		(0.00500)		(0.00342)
Good FE	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Observations	1,275	1,309	1,275	1,309	1,275	1,309	1,275	1,309
R^2	0.721	0.612	0.590	0.051	0.776	0.577	0.511	0.598
First Stage F	8.104		26.11		11.63		14.18	
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
LHS: θ_{grt} (Staple Foods)	Only Cereals		Top 20 Cereals		Random $\frac{1}{2}$ Foods		Other $\frac{1}{2}$ Foods	
$V_{gr}/\sum_{g'} V_{g'r}$	1.355***		0.704***		1.207**		0.746***	
	(0.147)		(0.0496)		(0.515)		(0.0659)	
$\ln p_{gr}$		-0.030***		-0.065***		-0.017***		-0.011***
		(0.00519)		(0.00739)		(0.00306)		(0.00242)
Good FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	750	770	1,200	1,540	1,725	1,925	1,800	2,079
R^2	0.804	0.864	0.859	0.882	0.673	0.795	0.888	0.612
First Stage F	12.00		11.90		12.12		10.76	

Note: Dependent variable, tastes, estimated using unexplained regional variation in food budget shares in 1987-88 (odd numbered columns) and 2004-05 (even numbered columns). Relative endowment is the portion of regional cropland planted with a particular crop using endowment data from the 1970's. Prices are regional weighted means of village median unit values. Endowment columns generally use predicted values of relative endowment from regressing crop shares on 8 crop-specific rainfall and temperature variables and altitude and use LIML. Regressions weighted by survey population weights for the 76 regions of India except columns 1 and 2. Columns 3 and 4 do not include good fixed effects. Columns 5 and 6 include region fixed effects as well as good fixed effects. The urban sample is used in columns 7 and 8. Columns 9 to 16 use alternative subsamples of foods. Robust standard errors clustered at the region level. Constant not reported. * significant at 10 percent, ** 5, *** 1.

Table 22: Tastes, Relative Resource Endowments and Prices: Alternative Taste Measures

	(1)	(2)	(3)	(4)	(5)	(6)
LHS: θ_{grt}(Staple Foods)	OLS Prices		Non Food IV		HHold P Interactions	
$V_{gr}/\sum_{g'} V_{g'r}$	1.472***		1.158***		1.237***	
	(0.238)		(0.118)		(0.134)	
$\ln p_{gr}$		-0.0280***		-0.0224***		-0.0237***
		(0.00499)		(0.00384)		(0.00399)
Good FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,275	1,309	1,275	1,309	1,275	1,309
R^2	0.681	0.380	0.938	0.926	0.781	0.823
First Stage F	11.81		11.81		11.81	

	(7)	(8)	(9)	(10)	(11)	(12)
LHS: θ_{grt}(Staple Foods)	Quadratic P and m		No Caste/Religion		Regional $\gamma_{gg'r}$ & β_{gr}	
$V_{gr}/\sum_{g'} V_{g'r}$	1.230***		1.221***		2.174**	
	(0.138)		(0.131)		(1.038)	
$\ln p_{gr}$		-0.0212***		-0.0238***		-0.0247*
		(0.00357)		(0.00410)		(0.0130)
Good FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,275	1,309	1,275	1,309	1,275	3,910
R^2	0.823	0.975	0.769	0.844	0.015	0.024
First Stage F	11.81		11.81		11.81	

Note: Dependent variable, tastes, estimated using unexplained regional variation in food budget shares in 1987-88 (odd numbered columns) and 2004-05 (even numbered columns) except columns 11 and 12. Relative endowment is the portion of regional cropland planted with a particular crop using endowment data from the 1970's. Prices are regional weighted means of village median unit values. All odd numbered columns use predicted values of relative endowment from regressing crop shares on 8 crop-specific rainfall and temperature variables and altitude. LIML estimation. Regressions weighted by survey population weights for the 76 regions of India. Different columns represent different specifications for taste estimates. Columns 1 and 2 do not use nearby village prices as instrument for local prices. Columns 3 and 4 instrument food expenditure with other expenditures. Column 5 and 6 interact own prices with household characteristics. Columns 7 and 8 include quadratic price and food expenditure terms. Columns 9 and 10 excludes caste and religion controls. Column 11 allows price and income coefficients to vary by region in the taste estimation and draw on additional survey data from 1983 and 1993-94, column 12 draws on additional survey data from 1993-94 and 1999-2000. Accordingly prices from all three survey rounds are included in column 12 alongside round fixed effects. In column 12, observations are weighted by the regional population in each survey. Robust standard errors clustered at the region level. Constant not reported. * significant at 10 percent, ** 5, *** 1.

Table 23: Tastes, Relative Resource Endowments and Prices: Alternative Price Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
LHS: θ_{grt}(Staple Foods)	Mean Prices		25th Percentile		75th Percentile		Transport Cost 1		Transport Cost 2	
$V_{gr}/\sum_{g'} V_{g'r}$	1.190***		1.108***		1.356***		1.135***		1.195***	
	(0.136)		(0.122)		(0.171)		(0.136)		(0.127)	
$\ln p_{gr}$		-0.0228***		-0.0213***		-0.0255***		-0.0249***		-0.0250***
		(0.00381)		(0.00458)		(0.00366)		(0.00448)		(0.00457)
Good FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,275	1,309	1,275	1,309	1,275	1,309	1,275	1,309	1,275	1,309
R^2	0.760	0.825	0.788	0.620	0.709	0.912	0.782	0.825	0.774	0.828
First Stage F	11.81		11.81		11.81		11.81		11.81	

Note: Dependent variable, tastes, estimated using unexplained regional variation in food budget shares in 1987-88 (odd numbered columns) and 2004-05 (even numbered columns). Relative endowment is the portion of regional cropland planted with a particular crop using endowment data from the 1970's. All odd numbered columns use predicted values of relative endowment from regressing crop shares on 8 crop-specific rainfall and temperature variables and altitude. LIML estimation. Various village prices are used instead of the median price that is used in the main specification (mean, 25th percentile and 75th percentile of the reported unit values, a unit price including a 5 percent ad-valorem transport cost when a good is not available locally and so a nearby price used instead, and a unit price including an ad-valorem transport cost based on sugar prices differences). Regional prices are weighted means of these village prices. Robust standard errors clustered at the region level. Constant not reported. * significant at 10 percent, ** 5, *** 1.

Table 24: Tastes, Relative Resource Endowments and Prices: Alternative Survey Rounds

	(1)	(2)	(3)	(4)	(5)	(4)	(5)	(6)	(7)	(8)
LHS: θ_{grt}(Staple Foods)	1983		1987-88		1993-94		1999-2000		2004-05	
$V_{gr}/\sum_{g'} V_{g'r}$	1.107***		1.218***		1.075***		1.171***		0.908***	
	(0.131)		(0.131)		(0.0847)		(0.140)		(0.117)	
$\ln p_{gr}$		-0.0450***		-0.0261***		-0.0211***		-0.0100**		-0.0248***
		(0.0100)		(0.00404)		(0.00347)		(0.00461)		(0.00444)
Good FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,190	1,207	1,275	1,292	1,275	1,292	1,292	1,309	1,292	1,309
R^2	0.834	0.394	0.770	0.436	0.799	0.492	0.769	0.501	0.819	0.576
First Stage F	11.01		11.81		15.50		12.52		12.67	

Note: Dependent variable, tastes, estimated using unexplained regional variation in food budget shares calculated for various survey rounds. Relative endowment is the portion of regional cropland planted with a particular crop using endowment data from the 1970's. All odd numbered columns use predicted values of relative endowment from regressing crop shares on 8 crop-specific rainfall and temperature variables and altitude. LIML estimation. Prices are regional weighted means of village median unit values for the appropriate survey round. Robust standard errors clustered at the region level. Constant not reported. * significant at 10 percent, ** 5, *** 1.

Table 25: Caloric Change and the Correlation of Tastes with Temporal Price Changes: Alternative Specifications

	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta \ln K_r$ 1987-88 to 2004-05 (Staple Foods)					
	Total Expenditure	Food Instrumented	g Normed Taste Corr.	r Normed Taste Corr.	g Rank Taste Corr.	r Rank Taste Corr.
$\sum_g(\theta_{gr} - \overline{\theta}_g)\Delta \ln p_{gr}$	-0.375***	-1.015***				
	(0.0834)	(0.0923)				
$Corr_g(\theta_{gr}, \Delta \ln p_{gr})$			-0.0999***	-0.0872***	-0.100***	-0.0798***
			(0.0111)	(0.0141)	(0.0109)	(0.0159)
$\sum_g(\overline{\theta}_g + z_g(.,.,.))\Delta \ln p_{gr}$	-0.438***	-1.003***	-0.695***		-0.700***	
	(0.139)	(0.111)	(0.0704)		(0.0738)	
$\sum_g z_g(.,.,.)\Delta \ln p_{gr}$				-0.365***		-0.412***
				(0.103)		(0.116)
$\sum_g(\bar{p}_r/p_{gr} - \overline{J}_r)\Delta \ln p_{gr}$	0.0177***	0.00354	0.0143**	0.0122**	0.0140**	0.0145**
	(0.00539)	(0.00314)	(0.00587)	(0.00519)	(0.00603)	(0.00560)
$\Delta \ln food_r$		0.983***	0.656***	0.507***	0.639***	0.461***
		(0.108)	(0.0520)	(0.0684)	(0.0511)	(0.0631)
$\Delta \ln m_r$	0.513***					
	(0.0756)					
$\sum_g s_{gr}(\bar{p}_r/p_{gr})\Delta \ln s_{gr}$	0.121	0.0753**	-0.0326	-0.0167	-0.0278	0.000903
	(0.0881)	(0.0358)	(0.0642)	(0.0618)	(0.0665)	(0.0702)
Constant	-0.334***	0.0238	-0.0249	-0.678***	0.00169	-0.634***
	(0.117)	(0.0511)	(0.0650)	(0.0670)	(0.0656)	(0.0616)
Observations	76	76	76	76	76	76
R^2	0.463	0.873	0.766	0.633	0.756	0.579

Note: Dependent variable is log change in caloric intake per person between 1987-88 and 2004-05. Independent variables come from log linearizing caloric intake. Tastes estimated using unexplained regional variation in food budget shares. Column 1 replaces the expenditure on the sample foods with total expenditure on all goods. Column 2 instruments food expenditure with expenditure on other goods, bounding any bias from food expenditure being endogenous. Columns 3 and 4 use a correlation between $\Delta \ln p_{gr}$ and tastes normalized mean 0 s.d. 1 by good and region respectively, with the correlation weighted by national food budget shares for each good. Columns 5 and 6 use a similar correlation with the rank of tastes over goods and regions respectively. Regressions weighted by survey population weights for the 76 regions of India. Robust standard errors. * significant at 10 percent, ** 5, *** 1.

Table 26: Caloric Change and the Correlation of Tastes with Temporal Price Changes: Alternative Samples

	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta \ln K_r$ 1987-88 to 2004-05					
	Unweighted (Staple Foods)	Urban (Staple Foods)	Only Cereals	Top 20 Foods	Random $\frac{1}{2}$ of Foods	Other $\frac{1}{2}$ of Foods
$\sum_g(\theta_{gr} - \overline{\theta}_g)\Delta \ln p_{gr}$	-0.838***	-0.811***	-0.975***	-0.998***	-0.327***	-0.968***
	(0.0617)	(0.0806)	(0.0597)	(0.0871)	(0.115)	(0.169)
$\sum_g(\overline{\theta}_g + z_g(.,.,.))\Delta \ln p_{gr}$	-0.861***	-0.869***	-0.785***	-1.027***	-0.825***	-1.255***
	(0.0616)	(0.105)	(0.0470)	(0.151)	(0.135)	(0.152)
$\sum_g(\bar{p}_r/p_{gr} - \overline{J}_r)\Delta \ln p_{gr}$	0.00710**	0.00284	0.00282	0.00638	0.00252	0.00290
	(0.00271)	(0.00334)	(0.00358)	(0.00400)	(0.00280)	(0.00241)
$\Delta \ln food_r$	0.827***	0.802***	0.826***	0.853***	1.169***	1.128***
	(0.0400)	(0.0807)	(0.0356)	(0.0691)	(0.0346)	(0.0702)
$\sum_g s_{gr}(\bar{p}_r/p_{gr})\Delta \ln s_{gr}$	0.0422	0.253***	0.0108	0.176*	0.578***	0.368***
	(0.0417)	(0.0807)	(0.0320)	(0.0929)	(0.0696)	(0.0932)
Constant	0.00779	0.0549	-0.0749*	0.141	-0.386**	0.145
	(0.0484)	(0.0663)	(0.0387)	(0.139)	(0.159)	(0.129)
Observations	76	76	76	76	76	76
R^2	0.903	0.807	0.912	0.778	0.966	0.877

Note: Dependent variable is log change in caloric intake per person between 1987-88 and 2004-05. Independent variables come from log linearizing caloric intake. Tastes estimated using unexplained regional variation in food budget shares. Regressions weighted by survey population weights for the 76 regions of India except column 1. Urban sample in column 2. Columns 3 to 6 use alternative subsamples of foods. Robust standard errors. * significant at 10 percent, ** 5, *** 1.

Table 27: Caloric Change and the Correlation of Tastes with Temporal Price Changes: Alternative Taste Measures

	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta \ln K_r$ 1987-88 to 2004-05 (Staple Foods)					
	OLS Prices	Non-Food IV	HHold P Interactions	Quadratic P and m	No Caste /Religion	Regional $\gamma_{gg'r}$ & β_{gr}
$\sum_g (\theta_{gr} - \overline{\theta}_g)\Delta \ln p_{gr}$	-0.900***	-0.894***	-0.899***	-0.891***	-0.904***	-0.877***
	(0.0624)	(0.0667)	(0.0630)	(0.0637)	(0.0652)	(0.0528)
$\sum_g (\overline{\theta}_g + z_g(.,.,.))\Delta \ln p_{gr}$	-0.834***	-0.855***	-0.842***	-0.857***	-0.839***	-0.881***
	(0.0516)	(0.0525)	(0.0529)	(0.0536)	(0.0512)	(0.0516)
$\sum_g (\bar{p}_r/p_{gr} - \overline{J}_r)\Delta \ln p_{gr}$	0.00475	0.00509	0.00487	0.00529*	0.00474	0.00545*
	(0.00329)	(0.00332)	(0.00328)	(0.00307)	(0.00334)	(0.00274)
$\Delta \ln food_r$	0.819***	0.823***	0.819***	0.823***	0.819***	0.822***
	(0.0428)	(0.0429)	(0.0432)	(0.0434)	(0.0427)	(0.0438)
$\sum_g s_{gr}(\bar{p}_r/p_{gr})\Delta \ln s_{gr}$	0.0641	0.0617	0.0623	0.0611	0.0629	0.0619
	(0.0386)	(0.0398)	(0.0387)	(0.0401)	(0.0390)	(0.0393)
Constant	-0.0112	0.00746	-0.00302	0.00964	-0.00636	0.0375
	(0.0387)	(0.0392)	(0.0379)	(0.0391)	(0.0377)	(0.0388)
Observations	76	76	76	76	76	76
R^2	0.903	0.902	0.902	0.901	0.903	0.902

Note: Dependent variable is log change in caloric intake per person between 1987-88 and 2004-05. Independent variables come from log linearizing caloric intake. Tastes estimated using unexplained regional variation in food budget shares. Different columns represent different specifications for taste estimates. Column 1 does not use nearby village prices as instrument for local prices. Column 2 instruments food expenditure with other expenditures. Column 3 interacts own prices with household characteristics. Column 4 include quadratic price and food expenditure terms. Column 5 excludes caste and religion controls. Column 6 allows price and income coefficients to vary by region in the taste estimation. Regressions weighted by a region's total survey weight in 1987-88. Robust standard errors. * significant at 10 percent, ** 5, *** 1.

Table 28: Caloric Change and the Correlation of Tastes with Temporal Price Changes: Alternative Price Measures

	(1)	(2)	(3)	(4)	(5)
	$\Delta \ln K_r$ 1987-88 to 2004-05 (Staple Foods)				
	Mean Prices	25th Percentile	75th Percentile	Transport Cost 1	Transport Cost 2
$\sum_g (\theta_{gr} - \overline{\theta}_g)\Delta \ln p_{gr}$	-0.966***	-0.754***	-1.058***	-0.919***	-0.916***
	(0.0618)	(0.0537)	(0.0824)	(0.0623)	(0.0674)
$\sum_g (\overline{\theta}_g + z_g(.,.,.))\Delta \ln p_{gr}$	-0.950***	-0.674***	-0.930***	-0.826***	-0.831***
	(0.0512)	(0.0648)	(0.0664)	(0.0560)	(0.0578)
$\sum_g (\bar{p}_r/p_{gr} - \overline{J}_r)\Delta \ln p_{gr}$	0.00560	0.00783**	0.000504	0.00589*	0.00453
	(0.00400)	(0.00361)	(0.00401)	(0.00335)	(0.00344)
$\Delta \ln food_r$	0.862***	0.789***	0.838***	0.831***	0.834***
	(0.0416)	(0.0533)	(0.0449)	(0.0455)	(0.0462)
$\sum_g s_{gr}(\bar{p}_r/p_{gr})\Delta \ln s_{gr}$	0.0251	0.135**	-0.0362	0.0704*	0.0783*
	(0.0328)	(0.0571)	(0.0348)	(0.0409)	(0.0400)
Constant	0.0686*	-0.171***	0.0796	-0.0336	-0.0278
	(0.0355)	(0.0495)	(0.0558)	(0.0407)	(0.0439)
Observations	76	76	76	76	76
R^2	0.917	0.862	0.883	0.906	0.898

Note: Dependent variable is log change in caloric intake per person between 1987-88 and 2004-05. Independent variables come from log linearizing caloric intake. Tastes estimated using unexplained regional variation in food budget shares. Various village prices are used instead of the median price that is used in the main specification (mean, 25th percentile and 75th percentile of the reported unit values, a unit price including a 5 percent ad-valorem transport cost when a good is not available locally and so a nearby price used instead, and a unit price including an ad-valorem transport cost based on sugar prices differences). Regressions weighted by a region's total survey weight in 1987-88. Robust standard errors. * significant at 10 percent, ** 5, *** 1.

Table 29: Caloric Change and the Correlation of Tastes with Temporal Price Changes: Other Price Changes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	$\Delta \ln K_r$ (Staple Foods)									
Initial Survey Round	1983	1987-88	1993-94	1999-00	1983	1987-88	1993-94	1983	1987-88	1983
Final Survey Round	2004-05	2004-05	2004-05	2004-05	1999-00	1999-00	1999-00	1993-94	1993-94	1987-88
$\sum_g(\theta_{gr} - \overline{\theta}_g)\Delta \ln p_{gr}$	-0.86***	-0.90***	-0.93***	-0.81***	-0.85***	-0.85***	-0.89***	-0.92***	-0.83***	-0.58***
	(0.050)	(0.065)	(0.083)	(0.057)	(0.058)	(0.093)	(0.14)	(0.079)	(0.097)	(0.11)
$\sum_g(\overline{\theta}_g + z_g(.,.,.))\Delta \ln p_{gr}$	-0.63***	-0.84***	-0.85***	-0.80***	-0.61***	-0.80***	-0.73***	-0.78***	-0.93***	-0.35***
	(0.071)	(0.052)	(0.070)	(0.076)	(0.066)	(0.077)	(0.095)	(0.078)	(0.058)	(0.13)
$\sum_g(\bar{p}_r/p_{gr} - \overline{J}_r)\Delta \ln p_{gr}$	-0.000012	0.0048	-0.0035*	-0.000045	-0.014**	-0.0022	-0.0082***	-0.00083	0.0024	0.013
	(0.0071)	(0.0033)	(0.0020)	(0.000052)	(0.0058)	(0.0033)	(0.0022)	(0.0071)	(0.0020)	(0.0093)
$\Delta \ln food_r$	0.71***	0.82***	0.81***	0.81***	0.74***	0.81***	0.78***	0.95***	0.87***	0.68***
	(0.058)	(0.043)	(0.051)	(0.038)	(0.062)	(0.065)	(0.060)	(0.067)	(0.046)	(0.085)
$\sum_g s_{gr}(\bar{p}_r/p_{gr})\Delta \ln s_{gr}$	0.047	0.063	0.33***	0.33*	0.096***	0.10**	0.13	0.24***	0.17**	0.024
	(0.033)	(0.039)	(0.064)	(0.19)	(0.027)	(0.048)	(0.083)	(0.043)	(0.080)	(0.082)
Constant	-0.17**	-0.0045	0.027	0.0071	-0.20***	-0.025	-0.023	-0.15**	0.022	-0.073***
	(0.071)	(0.038)	(0.023)	(0.0046)	(0.063)	(0.055)	(0.035)	(0.058)	(0.027)	(0.019)
Observations	71	76	76	77	71	76	76	70	75	71
R^2	0.847	0.902	0.868	0.883	0.841	0.804	0.806	0.841	0.883	0.714

Note: Dependent variable is log change in caloric intake per person between various survey rounds. Independent variables come from log linearizing caloric intake. Tastes estimated using unexplained regional variation in food budget shares. Regressions weighted by a region's total survey weight in the initial year. Robust standard errors. * significant at 10 percent, ** 5, *** 1.

Table 30: Predicted Impact of Internal Trade Liberalization With and Without Habit Formation

	(1)	(2)	(3)	(4)	(5)	(6)
	All Foods		Excluding Farmers, Staple Foods			
	Full Sample		Full Sample		<2000 Calories	
All-India Predicted Means	Habits Region θ_{gr}	No Habits Identical θ_g	Habits Region θ_{gr}	No Habits Identical θ_g	Habits Region θ_{gr}	No Habits Identical θ_g
$b_1 \sum_g \widehat{(\theta_{gr} - \overline{\theta}_g)} \Delta \ln p_{gr}$	-0.045***	0	-0.057***	0	-0.045***	0
	(0.0049)	(0)	(0.009)	(0)	(0.0074)	(0)
$b_2 \sum_g (\overline{\theta}_g + \widehat{z_g(.,.,.)}) \Delta \ln p_{gr}$	-0.0053	-0.0061	-0.012	-0.011	-0.0093	-0.0089
	(0.0090)	(0.0090)	(0.015)	(0.015)	(0.012)	(0.012)
$b_4 \sum_g (\bar{p}_r/\widehat{p_{gr} - \overline{J}_r}) \Delta \ln p_{gr}$	0.016***	0.017***	0.016***	0.017***	0.011***	0.012***
	(0.0011)	(0.0011)	(0.001)	(0.0011)	(0.00070)	(0.00076)
$b_5 \sum_g s_{gr} \widehat{(\bar{p}_r/p_{gr})} \Delta \ln s_{gr}$	-0.0057***	-0.018***	-0.0036***	-0.010***	-0.00045***	-0.0013***
	(0.0017)	(0.0021)	(0.0011)	(0.0015)	(0.00015)	(0.00019)
Total Effect ($\widehat{\Delta \ln K_r}$)	-0.040***	-0.0077	-0.056***	-0.0042	-0.044***	0.0015
	(0.0067)	(0.0084)	(0.013)	(0.015)	(0.011)	(0.012)
$\Delta \ln food$ to avoid K loss	-0.051***	-0.0100	-0.070***	-0.0052	-0.066***	0.0023
	(0.0087)	(0.011)	(0.016)	(0.019)	(0.016)	(0.018)

Note: 77 observations weighted by a region's total survey weight, or a samples total survey weight for the two subsamples. $\widehat{\Delta \ln K_r}$ is the predicted log change in calories holding food expenditure constant if 2004-2005 regional prices are equalized. Other variables are the components of $\widehat{\Delta \ln K_r}$ from equation 8 in the main paper. Predicted means use coefficients from estimating equation 8 and predicted values of s_{gr}, Δs_{gr} and $\bar{p_r}/p_{gr}$ from the AIDS demand estimates in section 3.3. Habits columns uses regional taste estimates, no habits columns sets all these regional tastes equal to the all-India average taste for each good. Columns 3 to 6 exclude households who report self-employment in agriculture as their primary activity. Additionally, columns 5 and 6 exclude households that consume more than 2000 calories per person per day. Robust standard errors for means. * significant at 10 percent, ** 5, *** 1.

Appendix References

Athukorala, Prema-chandra, "Agricultural Trade Policy Reforms in India," *South Asia Economic Journal*, 2005, *6* (1), 23–26.

Becker, Gary S. and Kevin M. Murphy, "A Theory of Rational Addiction," *The Journal of Political Economy*, 1988, *96* (4), 675–700.

Das-Gupta, Arindam, "Internal Trade Barriers in India: Fiscal Check-posts," *South Asia Economic Journal*, 2006, *7* (2), 231.

Deardorff, Alan V., "The General Validity of the Law of Comparative Advantage," *The Journal of Political Economy*, 1980, *88* (5), 941–957.

Dev, S. Mahendra and Robert E. Evenson, "Rural Development in India: Rural, Non-farm, and Migration," SCID Working Paper 187 2003.

Dixit, Avinash K. and Victor D. Norman, *Theory of International Trade: A Dual, General Equilibrium Approach*, Cambridge University Press, 1980.

FAO, "WTO Agreement on Agriculture: The Implementation Experience - Developing Country Case Studies," Report, Food and Agriculture Organization 2003.

FAO, "Towards An Indian Common Market: Removal of Restrictions on Internal Trade in Agriculture Commodities," Report, Food and Agriculture Organization 2005.

Gulati, Ashok, "Indian Agriculture in an Open Economy: Will it Prosper?," *India's Economic Reforms and Development: Essays for Manmohan Singh*, 1998, pp. 122–46.

Hausman, Jerry A., "Valuation of New Goods under Perfect and Imperfect Competition," NBER Working Paper 4970 1994.

Jha, Raghbendra, K. V. Bhanu Murthy, and Anurag Sharma, "Market Integration in Wholesale Rice Markets in India," ASARC Working Papers 2005-03 2005.

Planning Commisson of India, "Internal Trade," in "Tenth Five Year plan (2002–2007)," New Delhi, 2001, chapter 7.8.

Sanghi, Apurva, K. S. Kavi Kumar, and James W. McKinsey, "India Agriculture and Climate Data Set," 1998. http://chd.ucla.edu/dev_data.

Vanneman, Reeve and Douglas Barnes, "Indian District Data, 1961-1991: Machine-readable Data File and Codebook," 2000. http://www.inform.umd.edu/districts/index.html.

Vollrath, Dietrich, "The Agricultural Basis of Comparative Development," Technical Report, University of Houston 2011.

Willmott, Cort J. and Kenji Matsuura, "Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series 1950–1999," 2001. http://climate.geog.udel.edu/~climate.