Online Appendix for Wilcox, Ortiz-Bobea, Just 2021 (12/9/21)

Proofs of Propositions 1 and 2

The assumptions outlined in the main text for the separable agricultural household produce the following model and Lagrangian.

$$\begin{aligned} &\max_{F_s,\ A_s\ A_o} U(F_s) \\ &s.t.\ y = \left[P_o g(A_o) + P_s f(A_s) - C(A_o, A_s)\right] \geq P_s F_s, \\ &F_s,\ A_s,\ A_o \geq 0 \end{aligned} \tag{1}$$

$$L = U(F_s) + \lambda\left(P_o g(A_o) + P_s f(A_s) - C(A_o, A_s) - P_s F_s\right) \tag{2}$$

The resulting first order conditions and bordered Hessian with exogenous differential of interest for the separable model are as follows.

$$\mathscr{L} = U(F_s)\ + \lambda\left(P_o g(A_o) + P_s f(A_s) - C(A_o, A_s) - P_s F_s\right)$$

FOCs:

$$\begin{aligned} &\{\lambda\} : P_o g(A_o) + P_s f(A_s) - C(A_o, A_s) - P_s F_s = 0 \\ &\{F_s\} : U'_{F_s} - \lambda P_s = 0 \\ &\{A_o\} : \lambda\left(P_o g_o - C_o\right) = 0 \\ &\{A_s\} : \lambda\left(P_s f_s - C_s\right) = 0 \end{aligned}$$

$$\begin{pmatrix} 0 & -P_s & 0 & 0 \\ -P_s & U''_{F_s,F_s} & 0 & 0 \\ 0 & 0 & \lambda\left(P_o g_{oo} - C_{oo}\right) & -\left(\lambda C_{so}\right) \\ 0 & 0 & -\left(\lambda C_{so}\right) & \lambda\left(P_s f_{ss} - C_{ss}\right) \end{pmatrix} \cdot \begin{pmatrix} d\lambda \\ dF_s \\ dA_o \\ dA_s \end{pmatrix} = \begin{pmatrix} -\left(f(A_s) - F_s\right) \\ -(-\lambda) \\ 0 \\ -\lambda f_s \end{pmatrix} \cdot dP_s$$

We state that $dA_s/dP_s = \frac{f_s(P_o g_{oo} - C_{oo})}{\left[(C_{so})^2 - (P_o g_{oo} - C_{oo})(P_s f_{ss} - C_{ss})\right]} > 0$, which is a standard result.

<u>Proof of Proposition 1A.</u>

First we observe the following results for $\frac{dA_o}{dP_s}$ below, which follows since the sign of the denominator is negative by theorem (i.e. $sign(|H|) = (-1)^n$, where $n = 3$ is the number of choice variables in both models (1) and (3), thus $sign(|H|) = (-1)^3 < 0$, see Simon, Blume et al. 1994 theorem 19.6 or 19.8, or Chiang and Wainwright 2005 pg. 363).

$$\frac{dA_o}{dP_s} = \frac{f_s C_{so}}{\left[(C_{so})^2 - (P_o g_{oo} - C_{oo})(P_s f_{ss} - C_{ss})\right]}$$
$$\Rightarrow \frac{dA_o}{dP_s} > 0 \iff C_{so} < 0 \text{ (i.e. compliments)}$$

Now we use the expressions for $\frac{dA_s}{dP_s}$, $\frac{dA_o}{dP_s}$, and $\frac{dT}{dP_s}$, for the remaining proposition 1 results.

$$\rho_s = \frac{A_s}{T} \Rightarrow \frac{d\rho_s}{dP_s} = \frac{\frac{dA_s}{dP_s}T - A_s\frac{dT}{dP_s}}{T^2} \Rightarrow \eta_{\rho_{ss}} = \frac{P_s}{\rho_s}\frac{\frac{dA_s}{dP_s}T - A_s\frac{dT}{dP_s}}{T^2}$$
$$\Rightarrow \eta_{\rho_{ss}} = \frac{P_s}{\rho_s}\frac{\left[f_s(P_o g_{oo} - C_{oo})(T - A_s) - A_s f_s C_{so}\right]}{T^2\left[(C_{so})^2 - (P_o g_{oo} - C_{oo})(P_s f_{ss} - C_{ss})\right]}$$
$$\begin{aligned}\eta_{\rho_{ss}} > 0 \iff &(i)\ C_{so} < 0 \text{ (i.e. compliments), or}\\ &(ii)\ C_{so} > 0 \text{ and } f_s(P_o g_{oo} - C_{oo})(T - A_s) > |A_s f_s C_{so}| \text{ (i.e. substitutes).}\end{aligned}$$

$$\rho_o = \frac{A_o}{T} \Rightarrow \frac{d\rho_o}{dP_s} = \frac{\frac{dA_o}{dP_s}T - A_o\frac{dT}{dP_s}}{T^2} \Rightarrow \eta_{\rho_{os}} = \frac{P_s}{\rho_o}\frac{\frac{dA_o}{dP_s}T - A_o\frac{dT}{dP_s}}{T^2}$$
$$\Rightarrow \eta_{\rho_{os}} = \frac{P_s}{\rho_o}\frac{\left[f_s C_{so}(T - A_o) - A_o f_s(P_o g_{oo} - C_{oo})\right]}{T^2\left[(C_{so})^2 - (P_o g_{oo} - C_{oo})(P_s f_{ss} - C_{ss})\right]}$$
$$\eta_{\rho_{os}} > 0 \iff (i)\ C_{so} < 0 \text{ and } |f_s C_{so}(T - A_o)| > |A_o f_s(P_o g_{oo} - C_{oo})| \text{ (i.e. compliments)}$$

QED.

<u>Proof of Proposition 1B.</u>

Using the proceeding results for $\frac{dA_s}{dP_s}$ and $\frac{dA_o}{dP_s}$ we find the expression and conditions for $\frac{dT}{dP_s}$.

$$\frac{dT}{dP_s} = \frac{dA_s}{dP_s} + \frac{dA_o}{dP_s} \Rightarrow \eta_{T_s} = \frac{P_s}{T}\left(\frac{dA_s}{dP_s} + \frac{dA_o}{dP_s}\right)$$

$$\Rightarrow \eta_{T_s} = \frac{P_s}{T}\frac{f_s(P_o g_{oo} - C_{oo} + C_{so})}{\left[(C_{so})^2 - (P_o g_{oo} - C_{oo})(P_s f_{ss} - C_{ss})\right]}$$
$$\begin{aligned}\eta_{T_s} > 0 \iff &(i)\ C_{so} < 0 \text{ (i.e. compliments), or}\\ &(ii)\ C_{so} > 0 \text{ and } C_{so} < |P_o g_{oo} - C_{oo}| \text{ (i.e. substitutes)}\end{aligned}$$

QED.

Moving on to propositions 3, 4, and 5, the assumptions outlined in the main text produce the following model and Lagrangian.

$$
\begin{aligned}
&\max_{F_s,\, A_s\, A_o} U(F_s) \\
&\quad s.t.\ y = \left[P_o g(A_o) + P_s f(A_s) - C(A_o, A_s)\right] \geq P_s F_s, \\
&\quad \bar{Q}_s + \bar{S} \geq f(A_s) + F_s \\
&F_s,\ A_s,\ A_o \geq 0
\end{aligned} \tag{3}
$$

$$
L = U(F_s) + \lambda\left(P_o g(A_o) + P_s f(A_s) - C(A_o, A_s) - P_s F_s\right) + \theta(\bar{Q}_s + \bar{S} - f(A_s) - F_s) \tag{4}
$$

The first order conditions and bordered Hessian produced by the non-separable model) are then the following.

$$
\begin{aligned}
&\text{FOCs:} \\
&\{\lambda\} : P_o g(A_o) + P_s f(A_s) - C(A_o, A_s) - P_s F_s = 0 \\
&\{\theta\} : \bar{Q}_s + \bar{S} - f(A_s) - F_s = 0 \\
&\{F_s\} : U'_{F_s} - \lambda P_s - \theta = 0 \\
&\{A_o\} : \lambda\left(P_o g_o - C_o\right) = 0 \\
&\{A_s\} : \lambda\left(P_s f_s - C_s\right) - \theta f_s = 0
\end{aligned}
$$

$$
\begin{pmatrix}
0 & 0 & -P_s & 0 & \left(P_s f_s - C_s\right) \\
0 & 0 & -1 & 0 & -f_s \\
-P_s & -1 & U''_{F_s,F_s} & 0 & 0 \\
0 & 0 & 0 & \lambda\left(P_o g_{oo} - C_{oo}\right) & -\left(\lambda C_{so}\right) \\
\left(P_s f_s - C_s\right) & -f_s & 0 & -\left(\lambda C_{so}\right) & \lambda\left(P_s f_{ss} - C_{ss}\right)
\end{pmatrix}
\cdot
\begin{pmatrix}
d\lambda \\ d\theta \\ dF_s \\ dA_o \\ dA_s
\end{pmatrix}
=
\begin{pmatrix}
-\left(f(A_s) - F_s\right) \\ 0 \\ -(-\lambda) \\ 0 \\ -\lambda f_s
\end{pmatrix}
\cdot dP_s
$$

<u>Proof of Proposition 2A.</u>

We first show the result and conditions for $\frac{dA_s}{dP_s}$.

$$
\begin{aligned}
\frac{dA_s}{dP_s} &= \frac{(-1)^{1+5}(-\left(f(A_s) - F_s\right))|H_{15}| + (-1)^{3+5}(-(-\lambda))|H_{35}| + (-1)^{5+5}(-\lambda f_s)|H_{55}|}{|H|} \\
&= \frac{-MS_s \lambda\left(P_o g_{oo} - C_{oo}\right)\left[(2P_s f_s - C_s)\right]}{|H|},\ MS_s \text{ being staple marketable surplus} \\
&= \frac{-MS_s}{(2P_s f_s - C_s)}
\end{aligned}
$$

Therefore $\dfrac{dA_s}{dP_s} > 0 \iff$ the household is a net-buyer in the staple (i.e. since by FOC A_s, $P_s f_s > C_s$).

Using the expressions for $\frac{dA_s}{dP_s}$ and $\frac{dT}{dP_s}$ from model (3), we arrive at the following result and conditions for $\frac{d\rho_s}{dP_s}$.

$$\frac{d\rho_s}{dP_s} = \frac{\frac{dA_s}{dP_s}T - A_s\frac{dT}{dP_s}}{T^2} = \frac{\Big(-MS_s\lambda(2P_sf_s - C_s)\Big)\big(P_og_{oo} - C_{oo}\big)(T - A_s) + A_s\Big(MS_s(-\lambda C_{so})(2P_sf_s - C_s)\Big)}{T^2|H|}$$

$$= \frac{-MS_s\Big(A_o\big(P_og_{oo} - C_{oo}\big) + A_sC_{so}\Big)}{T^2(P_og_{oo} - C_{oo})(2P_sf_s - C_s)}$$

$\dfrac{d\rho_s}{dP_s} > 0 \iff (i)$ net-buyer and $C_{so} < 0$ (i.e. compliments), or

(ii) net-buyer, $C_{so} > 0$ (i.e. substitutes), and $|A_o(P_og_{oo} - C_{oo})| > |A_sC_{so}|$, or

(iii) net-seller, $C_{so} > 0$ (i.e. substitutes), and $|A_o(P_og_{oo} - C_{oo})| < |A_sC_{so}|$

QED.

Proof of Proposition 2B.

We first show the result and conditions for $\frac{dA_o}{dP_s}$.

$$\frac{dA_o}{dP_s} = \frac{(-1)^{1+4}(-\big(f(A_s) - F_s\big))|H_{14}| + (-1)^{3+4}(-(-\lambda))|H_{34}| + (-1)^{5+4}(-\lambda f_s)|H_{54}|}{|H|}$$

$$= \frac{MS_s(-\lambda C_{so})(2P_sf_s - C_s)}{\lambda(P_og_{oo} - C_{oo})(2P_sf_s - C_s)^2} \quad MS_s \text{ being staple marketable surplus}$$

$$= \frac{-MS_sC_{so}}{(P_og_{oo} - C_{oo})(2P_sf_s - C_s)}$$

Therefore $\dfrac{dA_o}{dP_s} > 0 \iff (i)$ the household is net-buyer in the staple and $C_{so} < 0$ (i.e. compliments), or

(ii) the household is net-seller in the staple and $C_{so} > 0$ (i.e. substitutes)

Using the expressions for $\frac{dA_o}{dP_s}$ and $\frac{dT}{dP_s}$ from model (3) we arrive at the following result and conditions for $\frac{d\rho_o}{dP_s}$.

$$
\begin{aligned}
\frac{d\rho_o}{dP_s} = \frac{\frac{dA_o}{dP_s}T - A_o\frac{dT}{dP_s}}{T^2} &= \frac{\Big(MS_s(-\lambda C_{so})(2P_sf_s - C_s)\Big)(T-A_o) - A_o\Big(-MS_s\lambda\left(P_og_{oo}-C_{oo}\right)(2P_sf_s-C_s)\Big)}{T^2|H|} \\
&= \frac{-MS_s\lambda\left(2P_sf_s - C_s\right)\Big(C_{so}(T-A_o) - A_o\big(P_og_{oo} - C_{oo}\big)\Big)}{T^2\lambda\left(P_og_{oo}-C_{oo}\right)(2P_sf_s - C_s)^2} \\
&= \frac{-MS_s\Big(A_sC_{so} - A_o\big(P_og_{oo} - C_{oo}\big)\Big)}{T^2(P_og_{oo}-C_{oo})(2P_sf_s-C_s)}
\end{aligned}
$$

$\frac{d\rho_o}{dP_s} > 0 \iff (i)$ household is net-seller and $C_{so} > 0$ (i.e. substitutes), or

(ii) household is net-buyer, $C_{so} < 0$ (i.e. compliments), and $|A_o(P_og_{oo} - C_{oo})| < |A_sC_{so}|$.

QED.

Proof of Proposition 2C.

Using the prior expression derived above for $\frac{dA_s}{dP_s}$ and $\frac{dA_o}{dP_s}$ we find the following result and conditions for $\frac{dT}{dP_s}$.

$$
\begin{aligned}
\frac{dT}{dP_s} = \frac{dA_s}{dP_s} + \frac{dA_o}{dP_s} &= \frac{-MS_s\lambda\left(P_og_{oo}-C_{oo}\right)\Big[(2P_sf_s - C_s)\Big]}{|H|} + \frac{MS_s(-\lambda C_{so})(2P_sf_s - C_s)}{|H|} \\
&= \frac{-MS_s\lambda(2P_sf_s - C_s)\big(P_og_{oo} - C_{oo} + C_{so}\big)}{\lambda(P_og_{oo}-C_{oo})(2P_sf_s-C_s)^2} \\
&= \frac{-MS_s\big(P_og_{oo} - C_{oo} + C_{so}\big)}{(P_og_{oo}-C_{oo})(2P_sf_s-C_s)}
\end{aligned}
$$

$\frac{dT}{dP_s} > 0 \iff (i)$ household is a net-buyer and $C_{so} < 0$ (i.e. compliments), or

(ii) household is a net-buyer, $C_{so} > 0$ (i.e. substitutes), and $|P_og_{oo} - C_{oo}| > |C_{so}|$, or

(iii) household is a net-seller, $C_{so} > 0$ (i.e. substitutes), and $|P_og_{oo} - C_{oo}| < |C_{so}|$

QED.

Discussion of Household-Level Econometric Models

As outlined in our main text, the nature of our household econometric models and corresponding data force some degree of compromise with estimators. If we had local price data, at say the village level, sufficient to provide contemporaneous and/or one-year lags of rice price moments to match with production data from 2004, 2007, and 2008 CSES rounds, then it would be possible to fully estimate the two-stage least squares model below,

$$P_{l,t} = \theta_1 Z_{l,t} + D'_{l,t-j}\theta_2 + X'_{i,l,t}\theta_3 + W'_{l,t-j}\theta_4 + \mathscr{L}_l + \mathscr{T}_t + u_{l,t} \tag{5}$$

$$A_{i,c,s,l,t+1} = \gamma_1 \widehat{P_{l,t}} + D'_{l,t-j}\gamma_2 + X'_{i,l,t}\gamma_3 + W'_{l,t-j}\gamma_4 + \mathscr{L}_l + \mathscr{T}_t + \varepsilon_{i,c,s,l,t+1}. \tag{6}$$

All variables in equations (5) and (6) are the same as the companion reduced-form with the exception of $P_{l,t}$ and $\widehat{P_{l,t}}$, which are the observed and predicted price variation, and $u_{l,t}$ and $\varepsilon_{i,c,s,l,t+1}$ which capture the error terms for the first and second stages. Unfortunately, the time-step of available recall-based production data forces one to drop most pre-price-shock observations because local price data is not available at a lagged or contemporaneous time-step. Specifically, in order to merge contemporaneous or lagged, local price data with household production data, a CSES survey round in the preceding year is needed. Since there were no CSES surveys in 2003, 2005 and 2006, local price data is not available to merge household production data from 2004 and 2007 – hence missing data for the endogenous variable of interest forces the researcher to omit data from these survey rounds to estimate respective instrumental variables specifications We operationalize this model with the district-level price data employed in our deforestation specifications.

By nature of the area-share data, the above set-up requires either estimation as a linear probability model, or a linearized regression model using an appropriate link function. A potentially significant caveat to either approach is less than ideal fidelity to the underlying data generating process – especially if there is sizable mass at either 0 or 1. If there is significant mass at either boundary point, the boundary observations must be dropped or censored with an arbitrary constant to ensure the data are bounded in $(0,1)$ and not $[0,1]$. In our setting, substantial mass is observed at boundaries resulting in a J-shaped distribution for rice and cash crop shares, thus implying non-trivial data censoring to operationalize standard fractional data estimators and/or a linearized approach with an appropriate link function (e.g. logit).

The alternative two-stage specification outlined below makes this non-linear reality more explicit via function $G(\cdot)$,

$$P_{l,t} = \theta_1 Z_{l,t} + D'_{l,t-j}\theta_2 + X'_{i,l,t}\theta_3 + W'_{l,t-j}\theta_4 + \mathscr{L}_l + \mathscr{T}_t + u_{l,t} \tag{7}$$

$$A_{i,c,s,l,t+1} = G\big(\gamma_1 \widehat{P_{l,t}} + D'_{l,t-j}\gamma_2 + X'_{i,l,t}\gamma_3 + W'_{l,t-j}\gamma_4 + \mathscr{L}_l + \mathscr{T}_t + \varepsilon_{i,c,s,l,t+1}\big). \tag{8}$$

Function $G(\cdot)$ is an unknown function that ensures the response variable stays within the unit interval, though the matter is theoretically deeper than just a mechanical relationship. As Ramalho, Ramalho, and Murteira (2011) point out, there are deeper underlying issues having to do with whether or not the model of interest is really a one- or two-part decision model (e.g. a binary discrete choice and continuous extensive margin choice). Our approach, as with our theoretical model, abstracts from these choices. Econometrically this could imply model mis-specification. In a footnote in our main text, we point out the there is some empirical evidence from Ramalho, Ramalho, and Murteira (2011) that linear approximations to these data generating processes can yield useful information in that they can correctly identify signs and approximate magnitudes, even if the underlying model being estimated represents a mis-specification of the data generating process.

Another estimation alternative with fractional data are the GMM estimators developed by

Ramalho and Ramalho (2017) and Ramalho, Ramalho, and Coelho (2018), referred to as exponential fractional-regression models. The starting point for operationalizing these estimators is, $y_{it} = G\left[exp\left(x_{it}\beta + \alpha_i + \varepsilon_{it}\right)\right]$ (equation 4 in Ramalho, Ramalho, and Coelho 2018). A key advantage of some of these estimators is that they reduce the need to trim boundary observations (i.e. they account for one of the boundaries), but many of them also permit modeling endogeneity with instruments and they also permit differencing-out of incidental parameters for fixed effects.

Although these estimators seem to provide particular advantages with true panel data, with repeated cross-sections the gains are not as clear. The applicable estimator from Ramalho, Ramalho, and Coelho (2018) is GMM_{pfe}, which is a pooled version of GMM_z developed by Ramalho and Ramalho (2017). We undertook testing with GMM_{pfe}, by implementing the GMM_z option in the applicable R package frmhet because the frmpd package only implements GMM_{pfe} for the version of this estimator that requires true panel data. Unfortunately, we encountered many local minima in our estimations, and computational challenges with the variance-covariance matrix required a bootstrap to estimate standard errors. It should also be noted that in contrast to most of the estimators pioneered by Ramalho, Ramalho, and Coelho (2018), GMM_{pfe} for repeated cross-sectional still requires trimming of boundary observations and it also requires estimating incidental parameters for fixed effects of interest because they cannot be differenced- or transformed-out.

In combination, this ended up being a very computationally intensive procedure, which for one estimation, required first optimizing starting values with a random number generator to avoid local minima (i.e. take the estimation with the lowest function value) all while estimating incidental parameters, and then implementing this procedure N times to estimate standard errors via a block-bootstrap that would sample with replacement down to the commune or district level. With at least seven different dependent variables of interest, this is a very burdensome undertaking. If we had suitable local price data for the entire set of cross-sections at, say, the commune or district level, this undertaking might be worth the effort[1].

With these complications and caveats in mind, we determined that the conservative and most tractable approach is to focus on the reduced-form as it allows us to study the full set of pre- and post-shock cross-sections and implement the available version of the two-stage least squares model for qualitative comparison. Clearly we would have a more ideal scenario with true panel data as we could leverage the state-of-the-art panel estimators from Ramalho, Ramalho, and Coelho (2018) for fractional data. However, under the circumstances, and compared with comparable literature in this space, we suggest that there is ample, novel value in implementing the models that we present as they are the most well-suited to the data we have, especially given the empirical support found in Ramalho, Ramalho, and Murteira (2011) for using linear or linearized models to identify signs and approximate magnitudes.

[1] We do not think province-level price variation is terribly meaningful at the household level, but one could implement such a specification. However, one still has data gaps for the earlier cross-sections when trying to merge lags of province level price variation with earlier cross-sections at the province level.

Supplementary Figures and Tables

Aggregate Production Data and Producer Prices from FAO

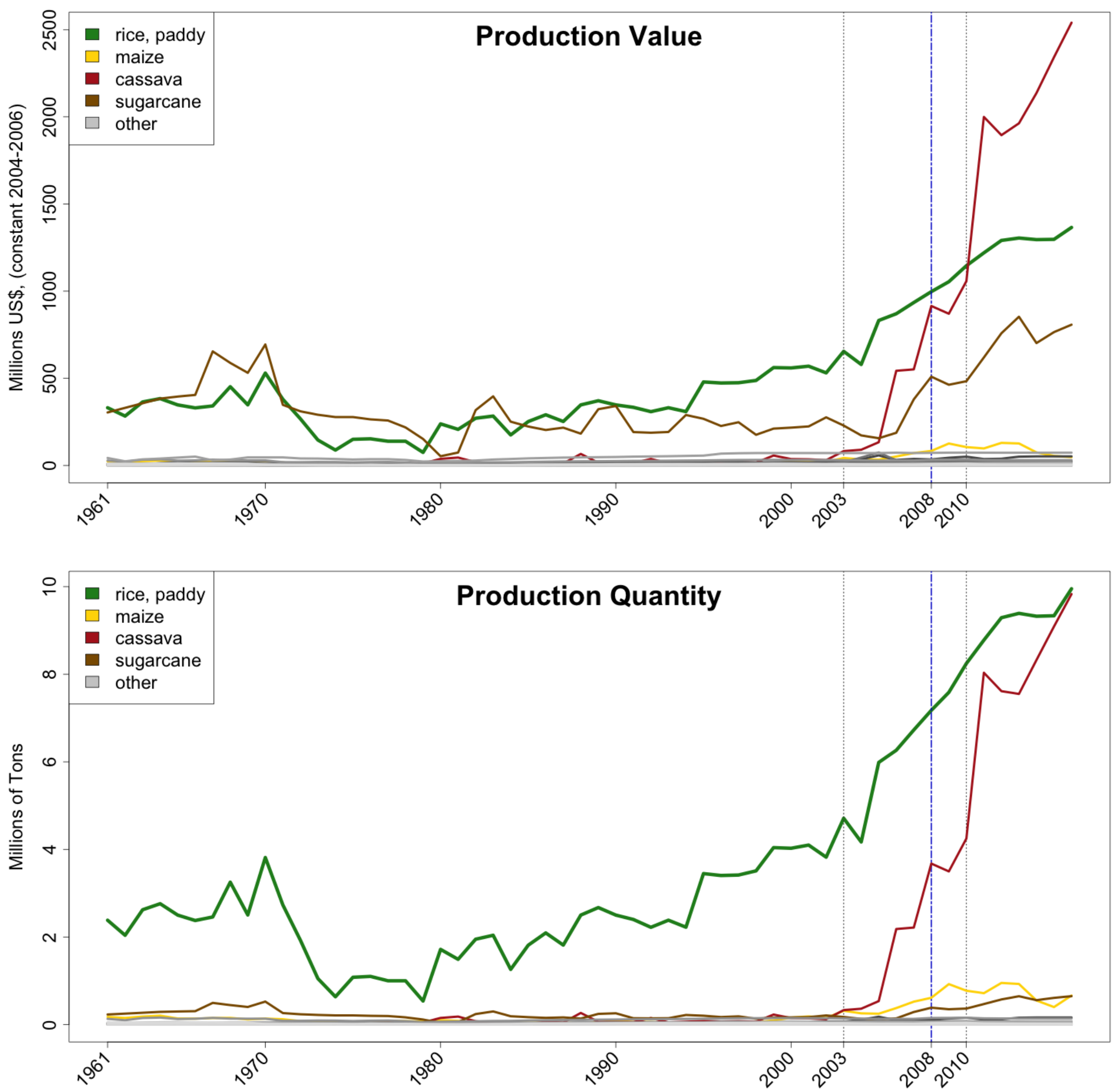

Figure 1: Production value and quantity 1961-2016. Authors' calculations using data from Cambodia Forestry Administration (2010).

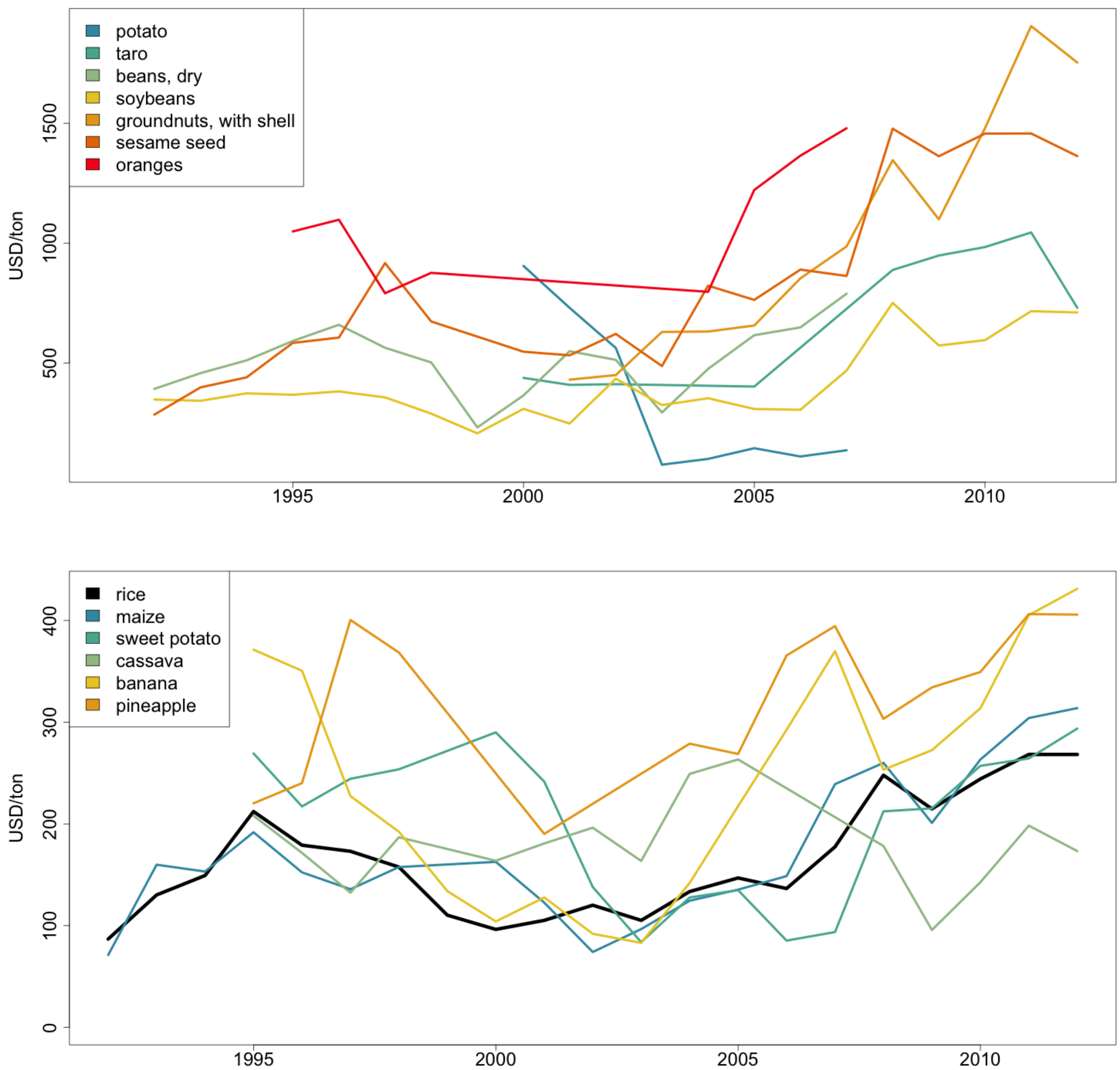

Figure 2: Selection of annual producer prices in Cambodia, 1992 - 2012. Authors' calculations using data from Cambodia Forestry Administration (2010).

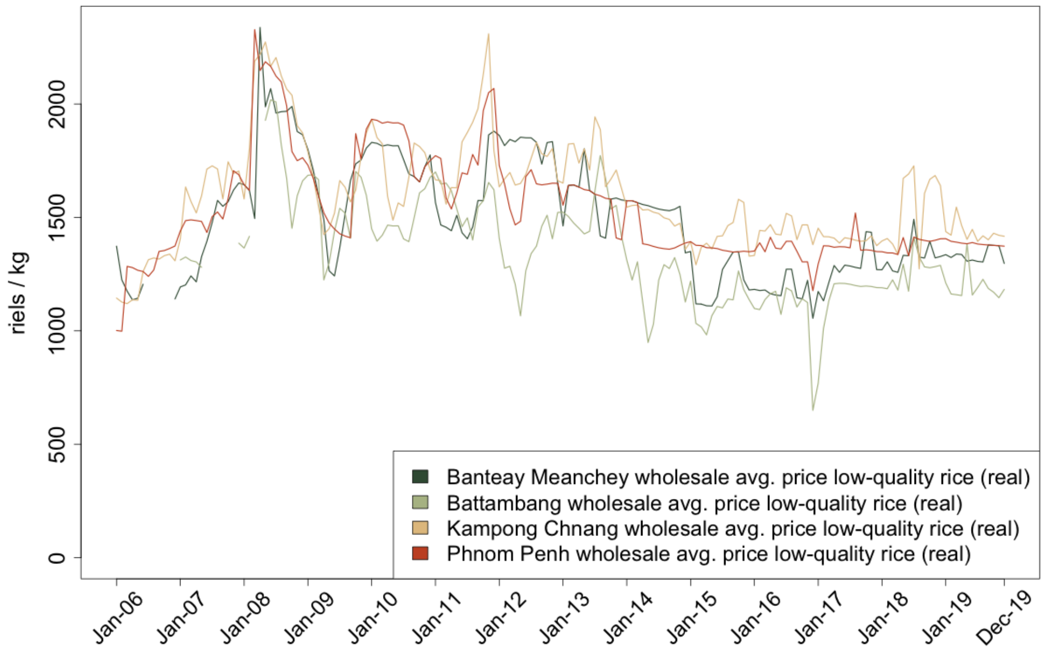

Figure 3: Average Price of Wholesale Low-Quality Rice, 2006-2019. Authors' calculations using data from FAO (2020).

Additional CSES Rice and Cash-Crop Price Data

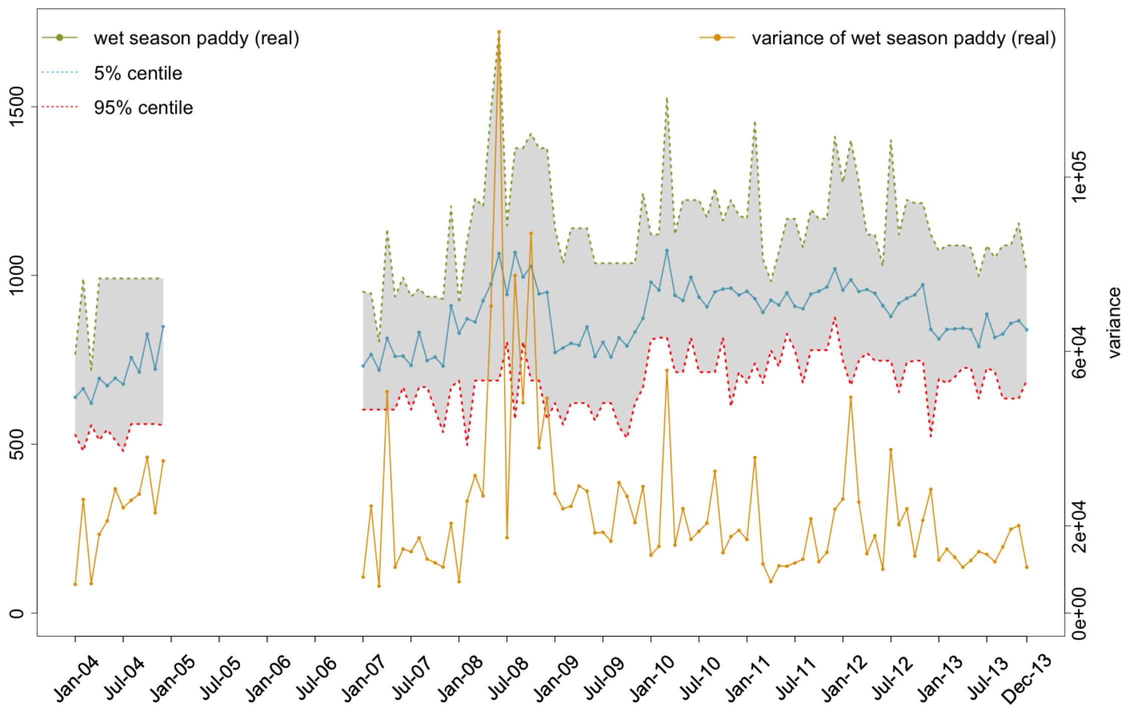

Figure 4: Wet season paddy farm-gate price distribution (n = 18,621) (author's calculations, CSES data). All observations are deflated using CPI data from Cambodia NIS; centiles above 99% or below 1% are assigned to the mean.

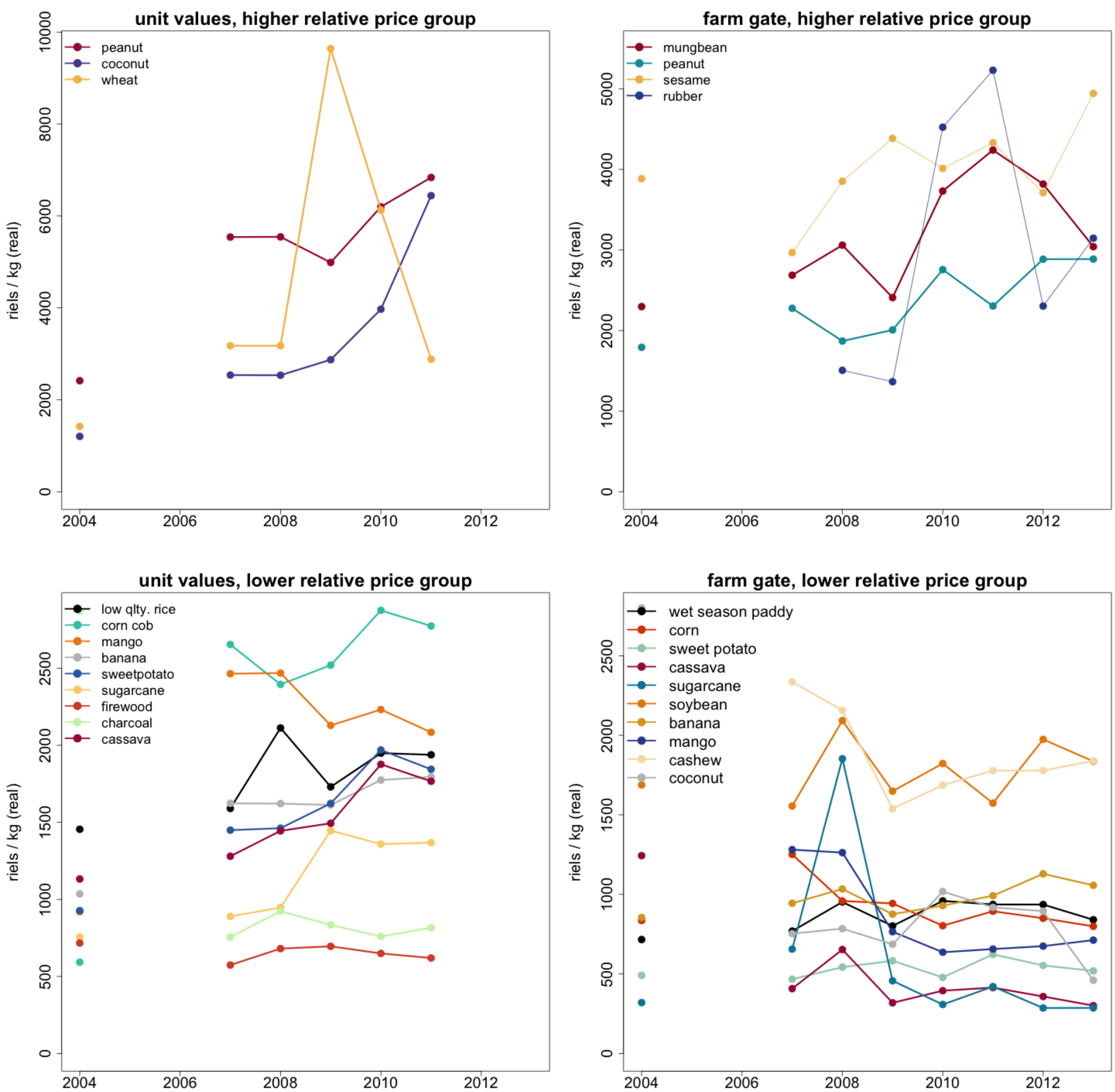

Figure 5: Selection of aggregate annual unit values and farm gate prices. Author's calculations using CSES data.

Household Summary Statistics and Approximate NBR Densities

Figure 6: Household (hh.) CSES summary statistics 2004, 2007-2013.

Variable name	All	Non-ag	Ag	Number of obs.
hh. size	4.76	4.63	4.83	(45969, 16202, 29767)
	(1.93)	(1.98)	(1.9)	
hh. size above 15 yrs.	3.23	3.26	3.22	(45968, 16201, 29767)
	(1.52)	(1.62)	(1.46)	
female head hh.	0.223	0.268	0.198	(45969, 16202, 29767)
	(0.416)	(0.443)	(0.399)	
years educ. head hh.	4.98	6.22	4.3	(45900, 16187, 29713)
	(4.04)	(4.53)	(3.58)	
male share above 15 yrs.	0.451	0.444	0.455	(45967, 16200, 29767)
	(0.191)	(0.203)	(0.185)	
weekly hrs. firewood prep	3.36	1.55	4.35	(45968, 16201, 29767)
	(5.55)	(4.21)	(5.94)	
hh. dietary diversity score	9.92	10.6	9.54	(45964, 16200, 29764)
	(2.35)	(2.38)	(2.25)	
total weeks starving	0.974	0.768	1.08	(42420, 14764, 27656)
	(4.03)	(3.83)	(4.12)	
approx. net rice sales	34400	-899	54100	(25363, 9056, 16307)
	(674000)	(797000)	(594000)	
approx. NBR	-0.00303	-0.0549	0.0258	(25363, 9056, 16307)
	(0.183)	(0.119)	(0.205)	
Engel value	0.802	0.807	0.8	(38214, 12731, 25483)
	(0.0931)	(0.0981)	(0.0904)	
rice share of food expnd.	0.217	0.102	0.275	(38214, 12731, 25483)
	(0.178)	(0.146)	(0.164)	
number of ag. plots			1.86	(NA, NA, 29767)
			(1.08)	
number of crops grown			1.42	(NA, NA, 27467)
			(0.794)	
hectares of ag. land			1.92	(NA, NA, 29767)
			(35.8)	
number plots irrigated rice			0.618	(NA, NA, 29767)
			(0.942)	
area share rice wet season			0.781	(NA, NA, 24722)
			(0.353)	
area share cash crops wet season			0.088	(NA, NA, 24722)
			(0.261)	
area share rice dry season			0.361	(NA, NA, 9533)
			(0.426)	
area share cash crops dry season			0.184	(NA, NA, 9533)
			(0.345)	

Note: Columns two through four present means and standard deviations for the entire sample (All), non-agricultural (Non-ag), and agricultural (Ag) households. The final three columns present the corresponding number of observations.

Note about Approximate NBR construction:

Following Deaton (1989), NBR variable construction is as follows. For household i, surveyed in month m, during year y,

$$NBR_{i,m,y} = \left[(\text{total ag. income}_{i,m,y}) - (\text{total value of rice consumption}_{i,m,y})\right] / (\text{total income}_{i,m,y}).$$

Since "total agricultural income month of survey" cannot be subset to rice-income only, the resulting values overstate the value of rice production and understate the density of net-buyers in rice (i.e. since, total ag. income $\geq$ total rice ag. income). We do, however, set agricultural income to zero for households who report not growing rice over the prior agricultural year to mitigate potential for non-zero agricultural income for households that produce crops, but not rice.

Below we provide annual non-parametric densities for agricultural households by pooling all households surveyed within a given year. 2004 data is not incorporated due to non-conformable questions with later year surveys. In particular, while income from agriculture is separated from income from forest related activities in later years of the CSES, in 2004 these respective income categories are not separated in the data.

The general shape and distribution of the densities down below is robust to choice of kernel and adaptive bandwidth selection method, with the exception of rule of thumb bandwidth (which appears to over-smooth the underlying densities). Rule of thumb bandwidth is generally only recommended for exploratory analysis (see Racine 2019 pg. 73).

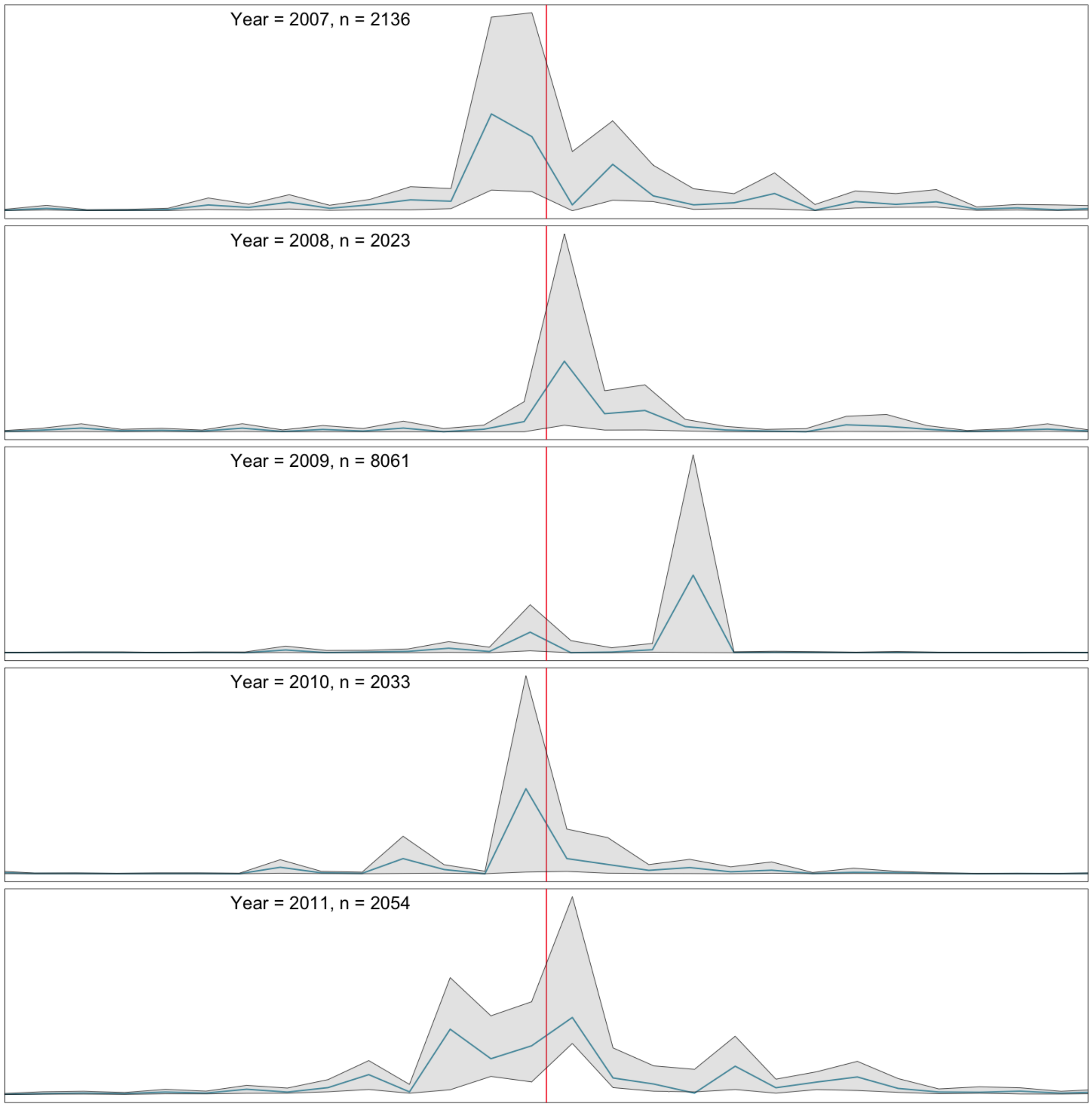

Figure 7: Non-parametric densities of approximate NBRs for agricultural households constructed using diary and crop production data. Specifications are: Epanechnikov kernel (order 2); optimal bandwidth via adaptive nearest-neighbor least squares cross-validation; and 95% bootstrapped confidence intervals (n= 1000, iid resample). The y-axis represents density, the x-axis is centered at zero (red line) spanning $[-0.5, 0.5]$ (a truncation of the theoretical $[-1, 1]$ NBR support).

Spatial Plots of Cambodia, Deforestation, and Rice Prices

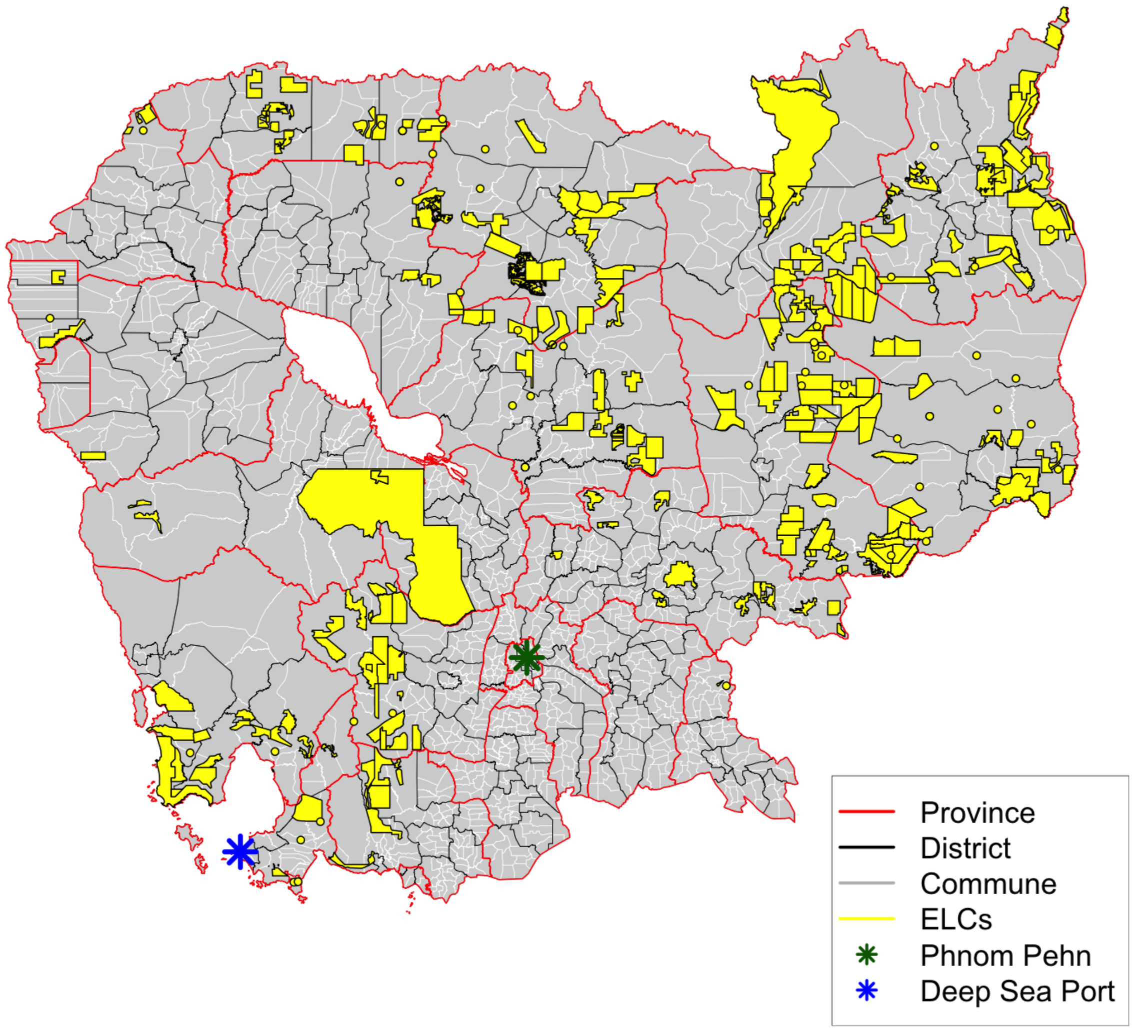

Figure 8: Map of Cambodia's administrative boundaries, economic land concession boundaries, and locations of Phnom Pehn and Sihanoukville deep sea port.

Below are several spatial plots showing deforestation and rice price variation. Note that CSES consumption module data were not gathered in 2012-2013, hence the associated plot for low-quality rice unit values extends to 2011 while our wet season paddy data extends through 2013.

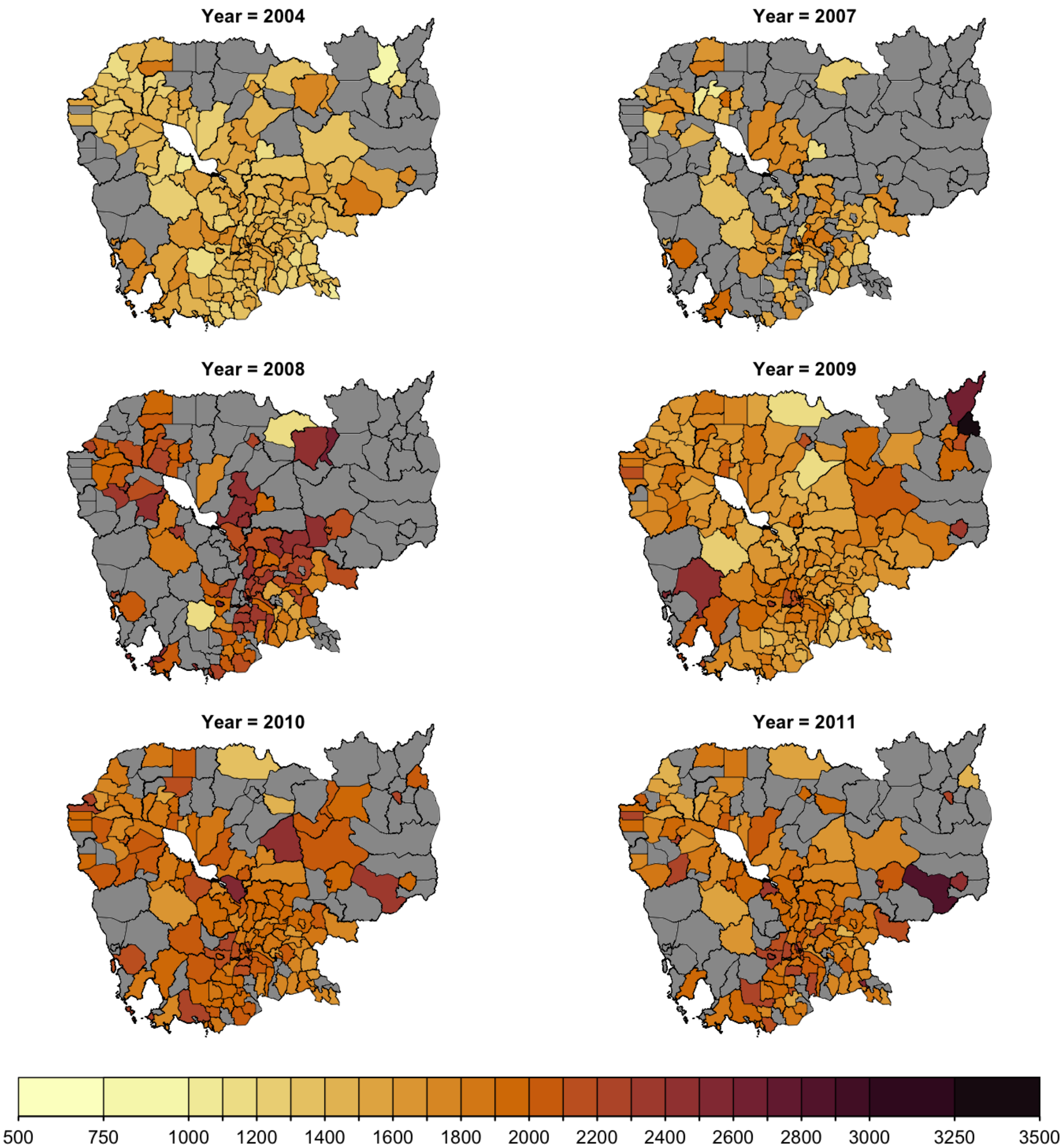

Figure 9: District-level average low-quality rice unit values by year. Values depicted are in riels (real). Grey districts were not surveyed or have missing data. Unit values for broken rice were obtained from CSES consumption diary. Extreme outliers (i.e. 99.9% or above, or 0.1% or below) were assigned to their respective annual mean.

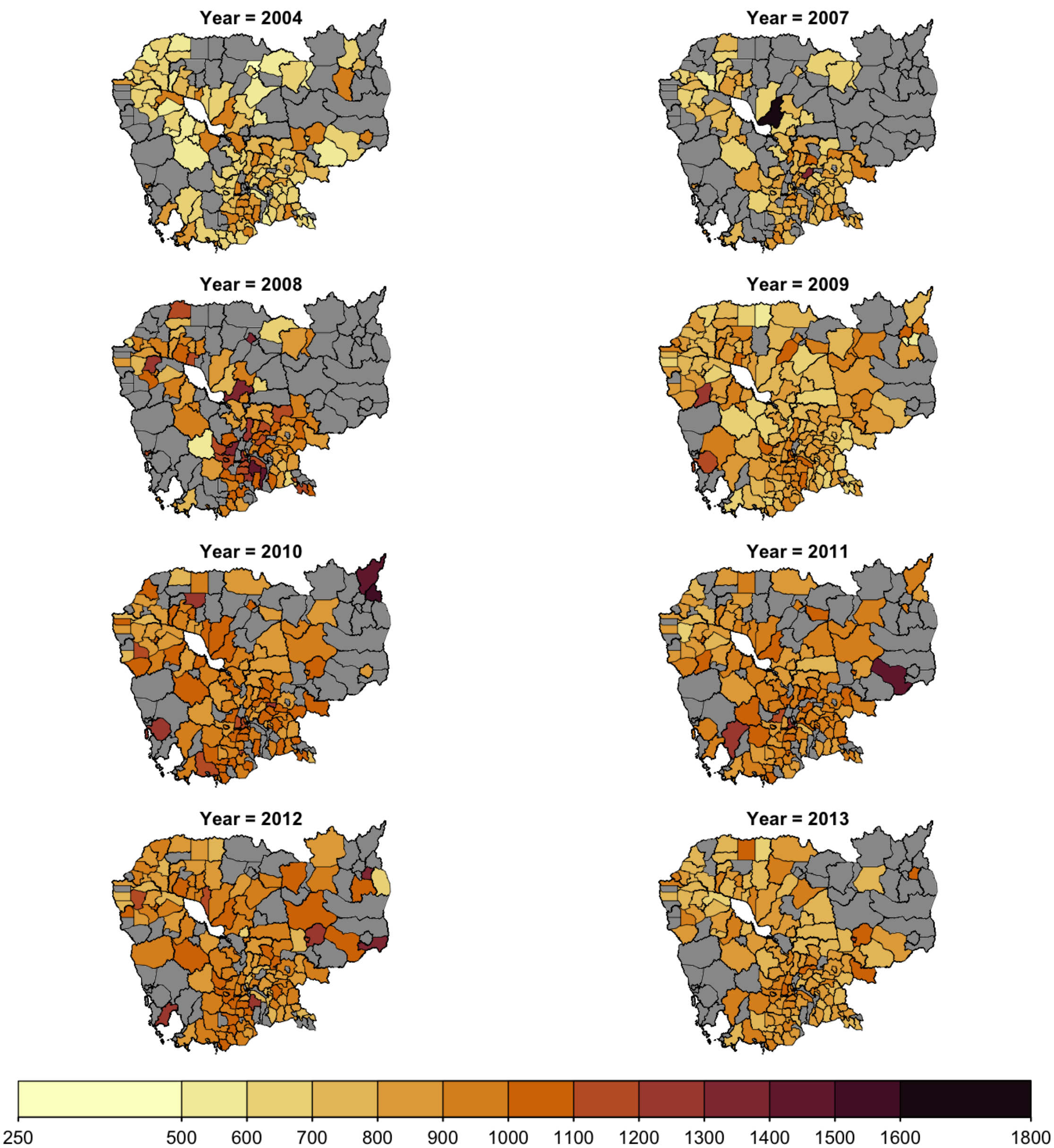

Figure 10: District-level average wet season paddy price by year. Values depicted are in riels (real). Grey districts were not surveyed or have missing data. Wet season paddy prices were obtained from CSES crop production data (section B). Extreme outliers (i.e. 99.9% or above, or 0.1% or below) were assigned to their respective annual mean.

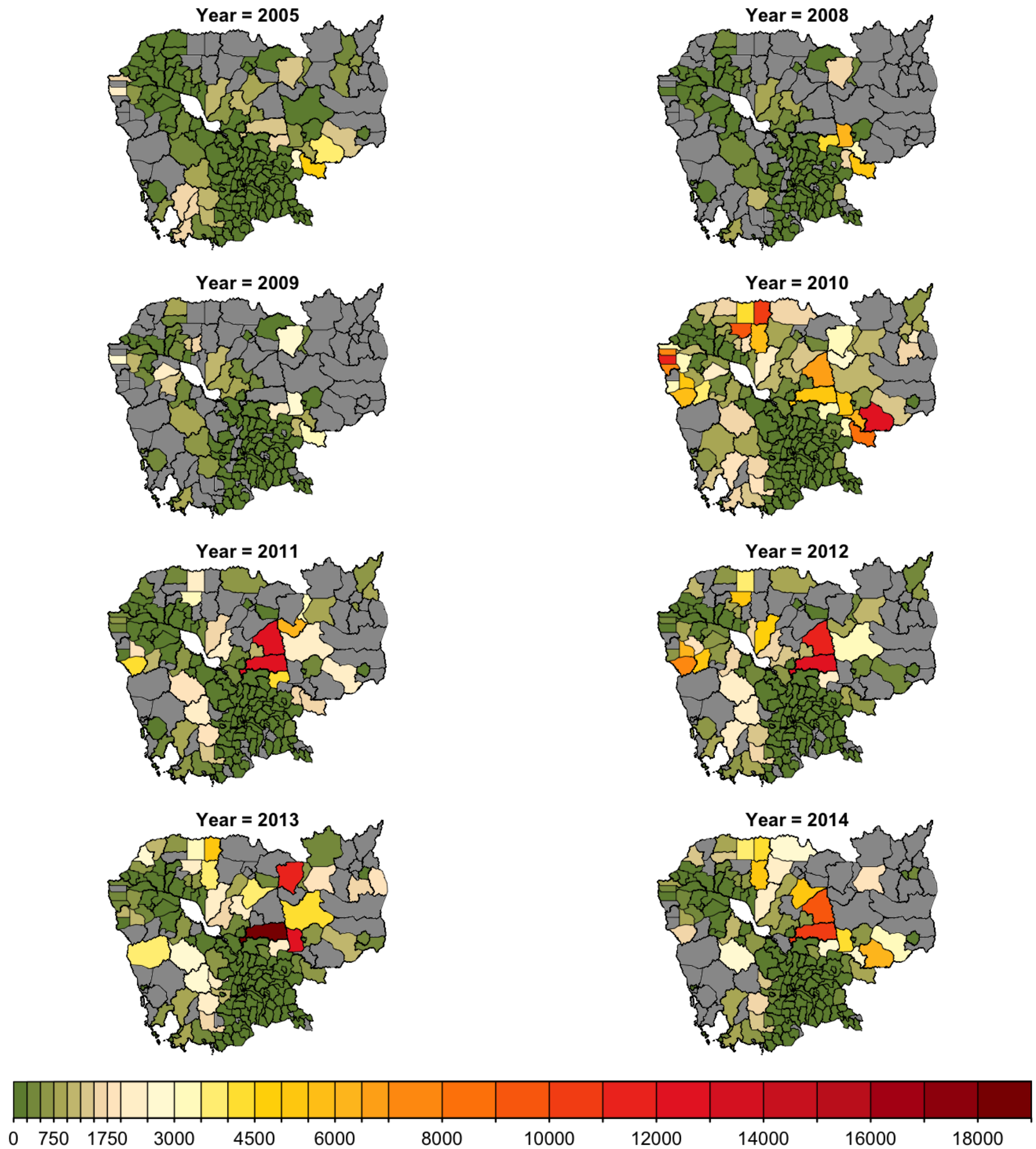

Figure 11: District-level non-ELC deforestation by year in hectares. Grey districts were not observed in CSES data for given *year* – 1. Deforestation data comes from Hansen et al. (2013) and ELC shapefile used to mask data in ELC areas comes from ODC (2019).

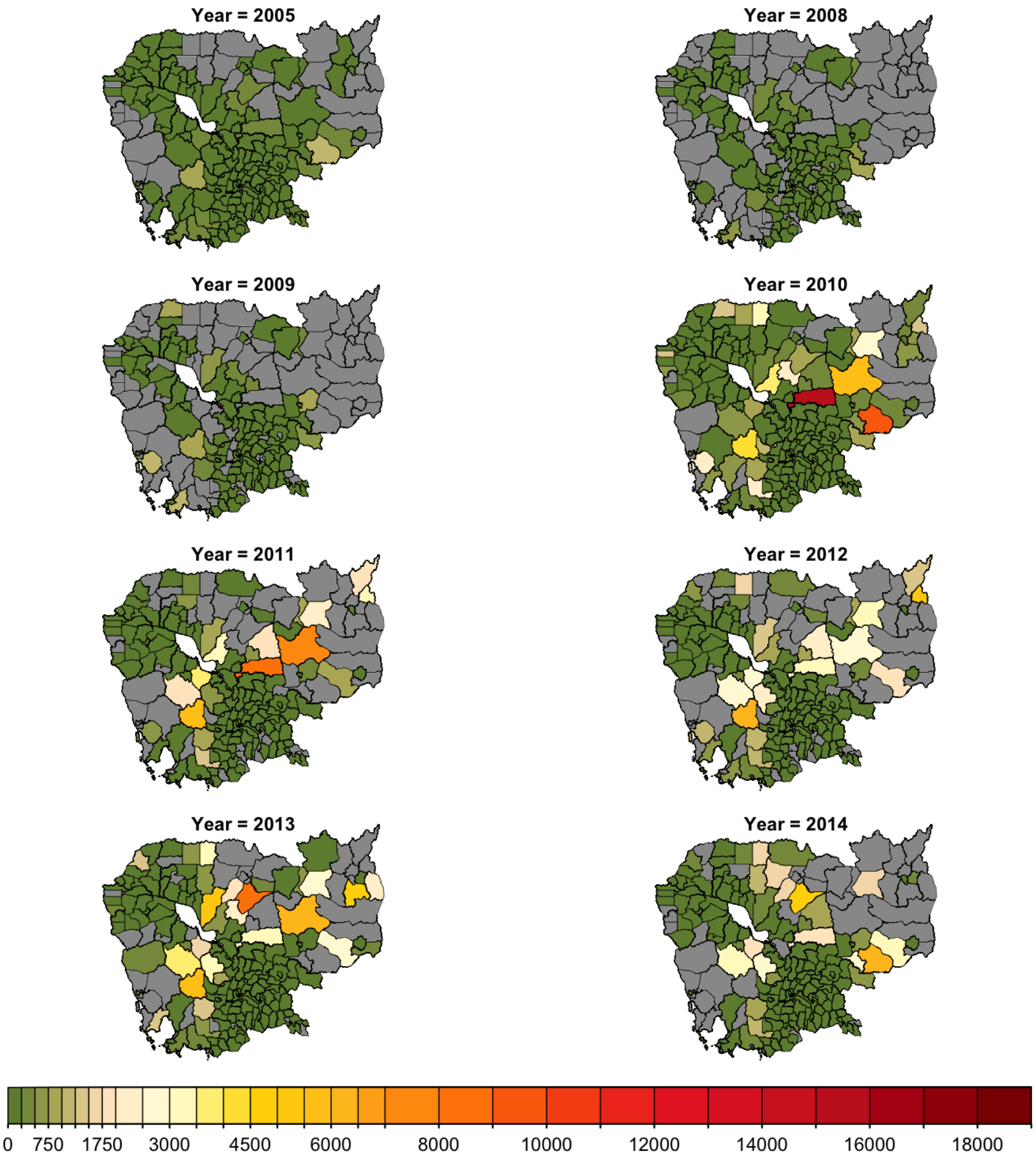

Figure 12: District-level ELC deforestation by year in hectares. Grey districts were not observed in CSES data for given $year - 1$. Deforestation data comes from Hansen et al. (2013) and ELC shapefile used to mask data in ELC areas comes from ODC (2019).

Regression Results for IV Correlation with Other Prices

Here we provide two figures showing point estimates and p-values for instrument Z when regressed on unit values or farm gate prices for major cash crops, firewood, and charcoal. Out of 60 farm gate price-focused regressions, 33% show Z is significant at 10% or less; 70% of this 33% are from "short" regressions with trends, Z, and location fixed effects (signs are largely negative, with an average of $n = 163.8$ observations per regression). Of the 33 cash-crop unit value-focused regressions, 30% show Z is significant at 10% or less; 81% of this 30% are “short” regressions with or without fixed effects (signs are largely negative, with an average of $n = 560.9$ observations per regression). Among the unit value regression tests, only mango and coconut unit values show signs of strong correlation with our instrument; cassava shows no correlation. Robustness checks in our deforestation specifications show that our results do not change when any of the price series with high correlation with our instruments are included as an instrumented endogenous regressor. Overall, these findings indicate that our instrument is not correlated with other commodity prices in any strong and meaningful way the changes our main results.

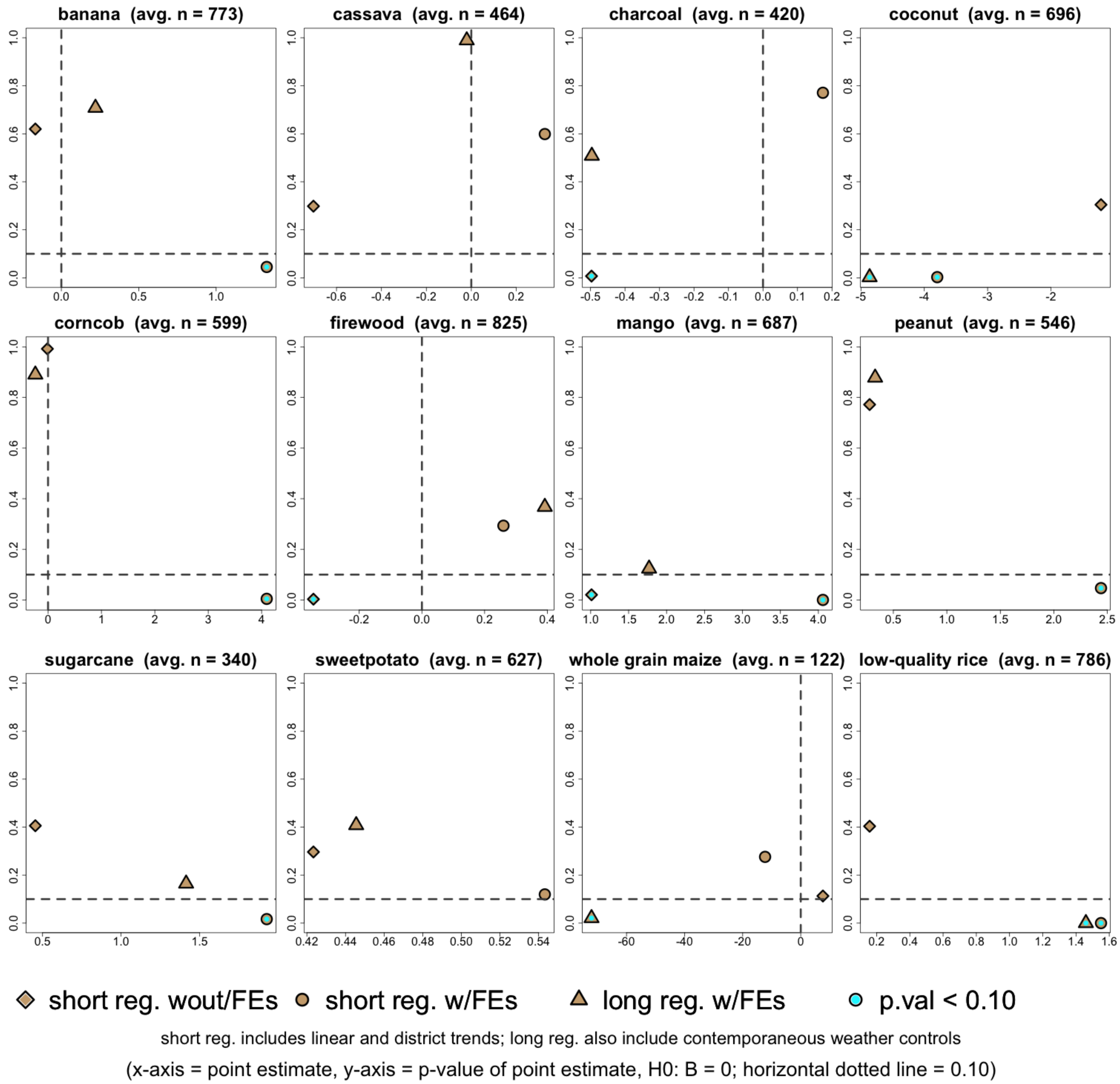

Figure 13: Point estimates of district-level regressions of various mean unit value cash crop prices in period t on contemporaneous instrument Z (interaction of centroid road distances to Cambodia's deep sea port with mean US based international rice price); points estimates are on the x-axis, corresponding p-values are on the y-axis; horizontal line is plotted at $y = 0.10$, vertical at 0. Far bottom right plot has low-quality rice unit values as the dependent variable for reference. All regressions include district level fixed effects and province and district level clustered standard errors: $n = 33$ total cash-crop focused regressions, 30% of which are significant at 10% or less (81% of these being 'short' regressions with and without district level fixed effects); overall mean of point estimate on Z is -1.78.

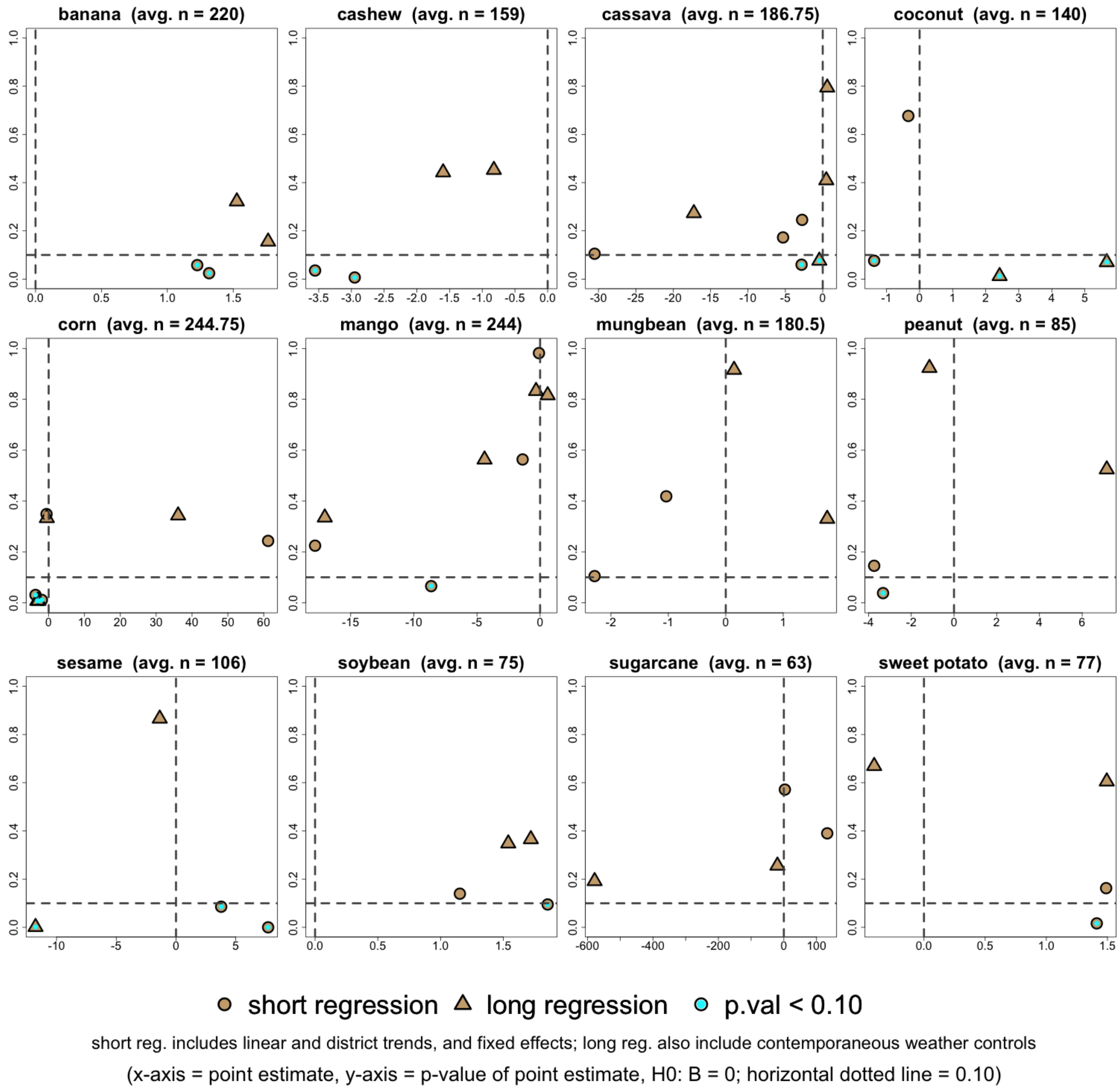

Figure 14: Point estimates of district-level regressions of various mean farm gate cash crop prices in period t on contemporaneous instrument Z (interaction of centroid road distances to Cambodia's deep sea port with mean US based international rice price); points estimates are on the x-axis, corresponding p-values are on the y-axis; horizontal line is plotted at $y = 0.10$, vertical at 0. Plots vary in having eight or four points estimates, capturing more or less adjustment, respectively, to censoring dependent variable outliers (light censoring to raw data); more points implies greater range of dependent variable adjustment. All regressions include district level fixed effects and province and district clustered standard errors: $n = 60$ total regressions, 33% of which are significant at 10% or less (70% of these being 'short' regressions); overall mean of point estimate on Z is -7.9.

Supplementary Regression Results for Deforestation Response

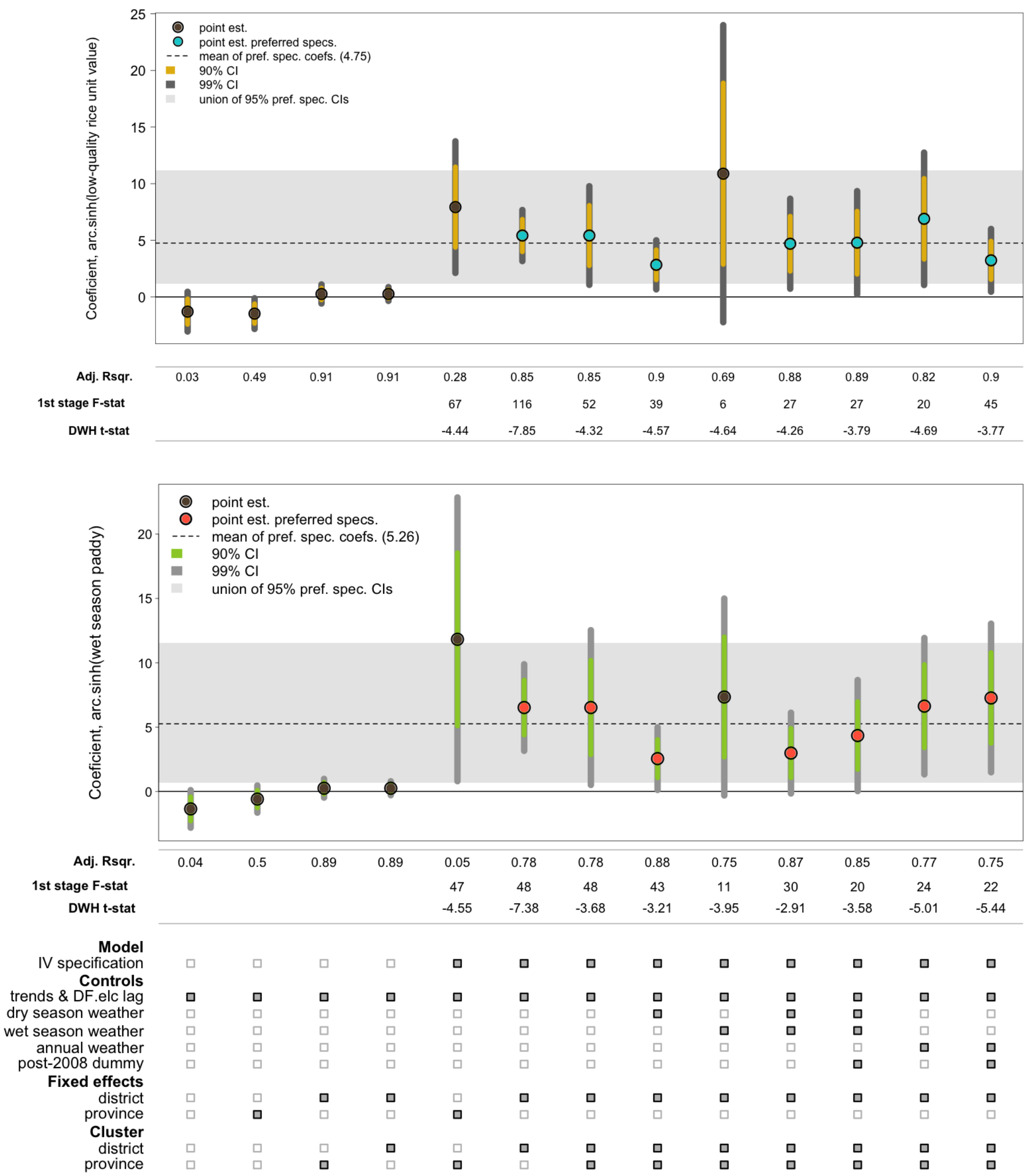

Figure 15: Side by side comparison of point estimates of the arcsinh of district-level mean rice prices in t regressed on arcsinh(district-level non-ELC hectares deforested) in $t+1$: top panel shows point estimates of arcsinh(low-quality rice unit values) ($N = 786$; $t \in (2004, 2007-2013)$; 172 districts, 24 provinces); bottom panel shows point estimates of arcsinh(wet season paddy) ($N = 988$; $t \in (2004, 2007-2013)$; 179 districts, 24 provinces).

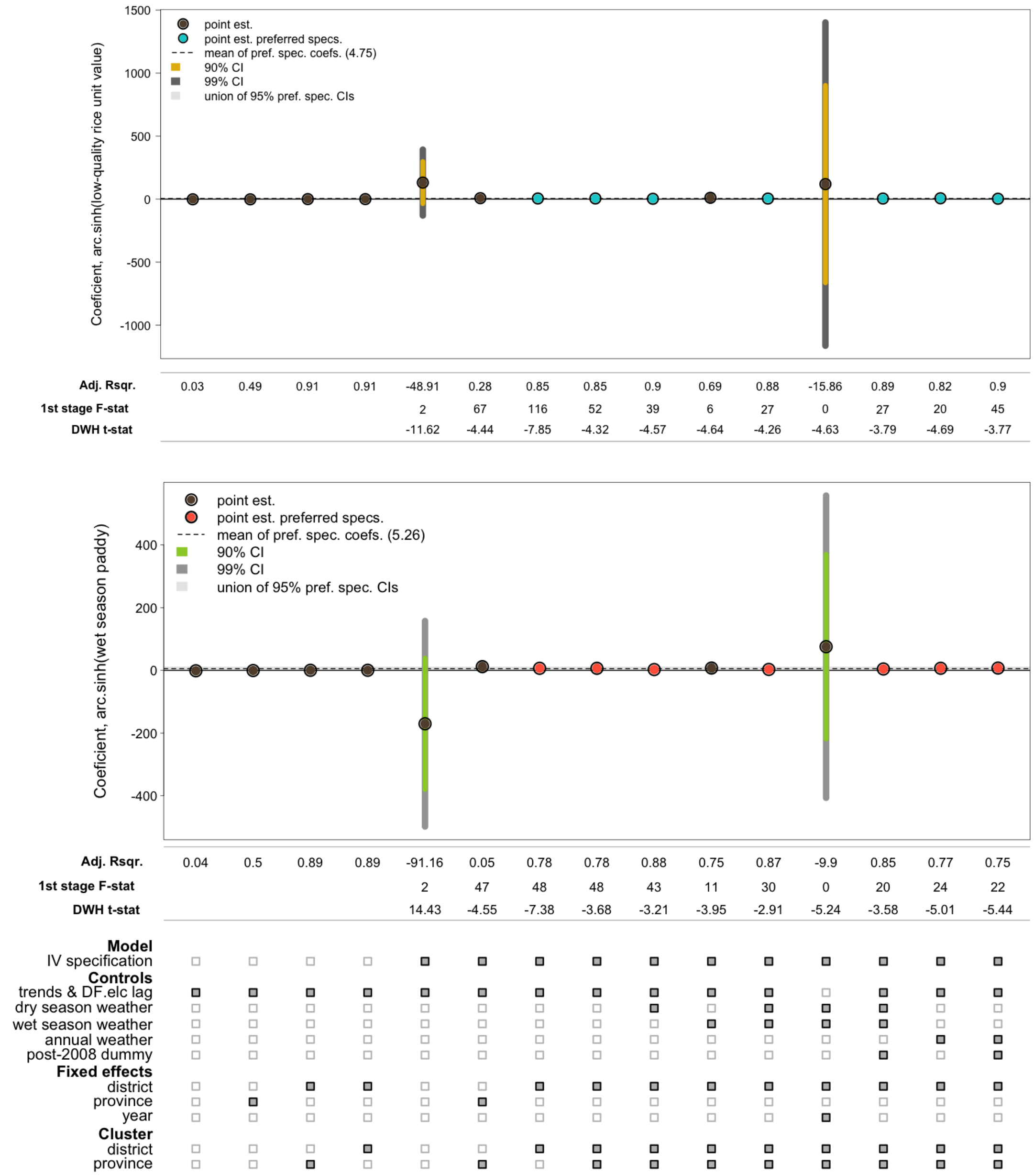

Figure 16: Identical side by side comparison of point estimates to the immediately preceding figure 15, except for columns 5 and 12, which are omitted from the main text for obvious scale reasons. Columns 5 shows the result of an IV specification without any fixed effects. Column 12 demonstrates what happens when year fixed effects are included.

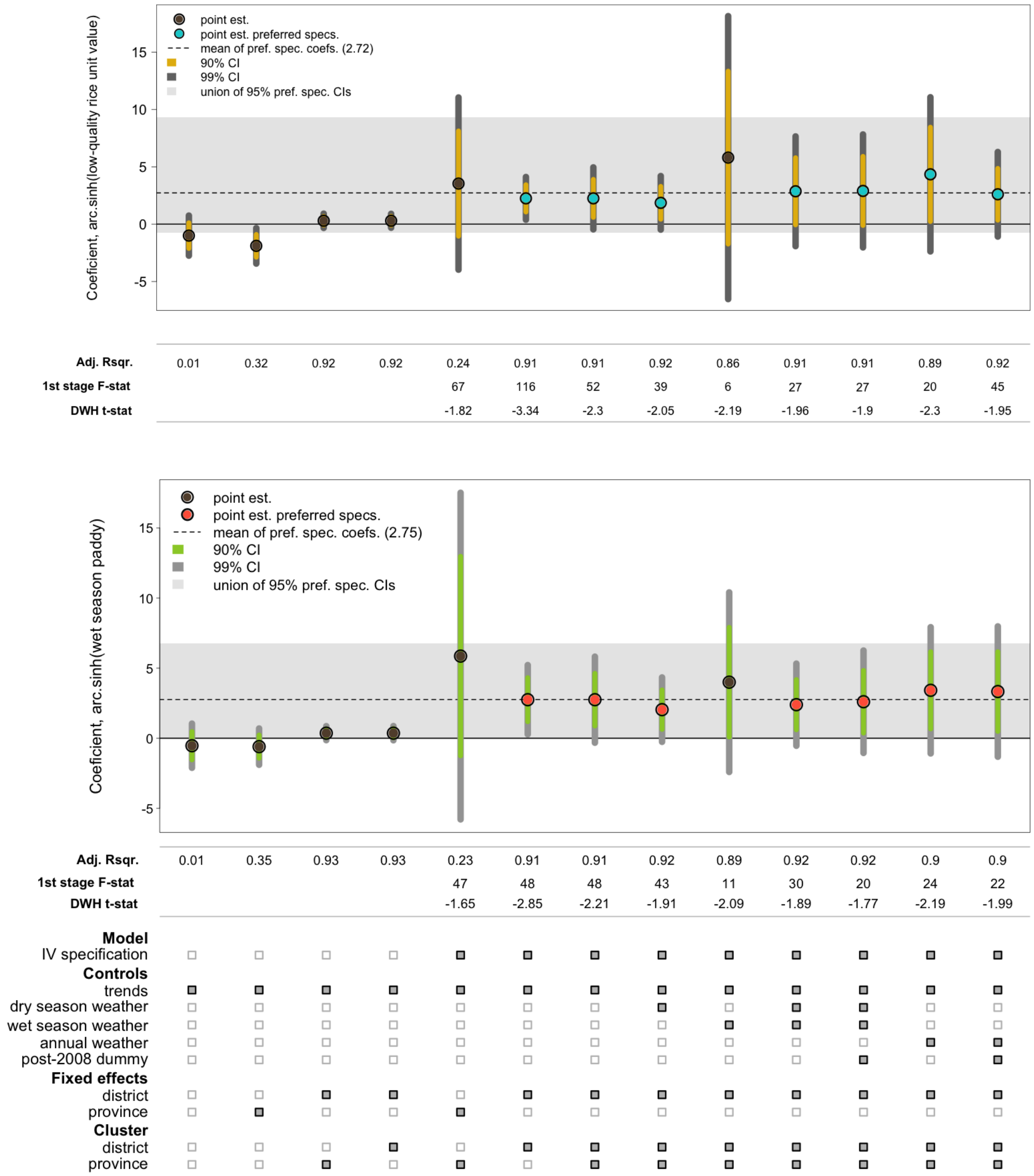

Figure 17: Side by side comparison of point estimates of the arcsinh of district-level mean rice prices in t regressed on arcsinh(district-level ELC hectares deforested) in $t+1$: top panel shows point estimates of arcsinh(low-quality rice unit values) ($N = 786$; $t \in (2004, 2007-2013)$; 172 districts, 24 provinces); bottom panel shows point estimates of arcsinh(wet season paddy) ($N = 988$; $t \in (2004, 2007-2013)$; 179 districts, 24 provinces).

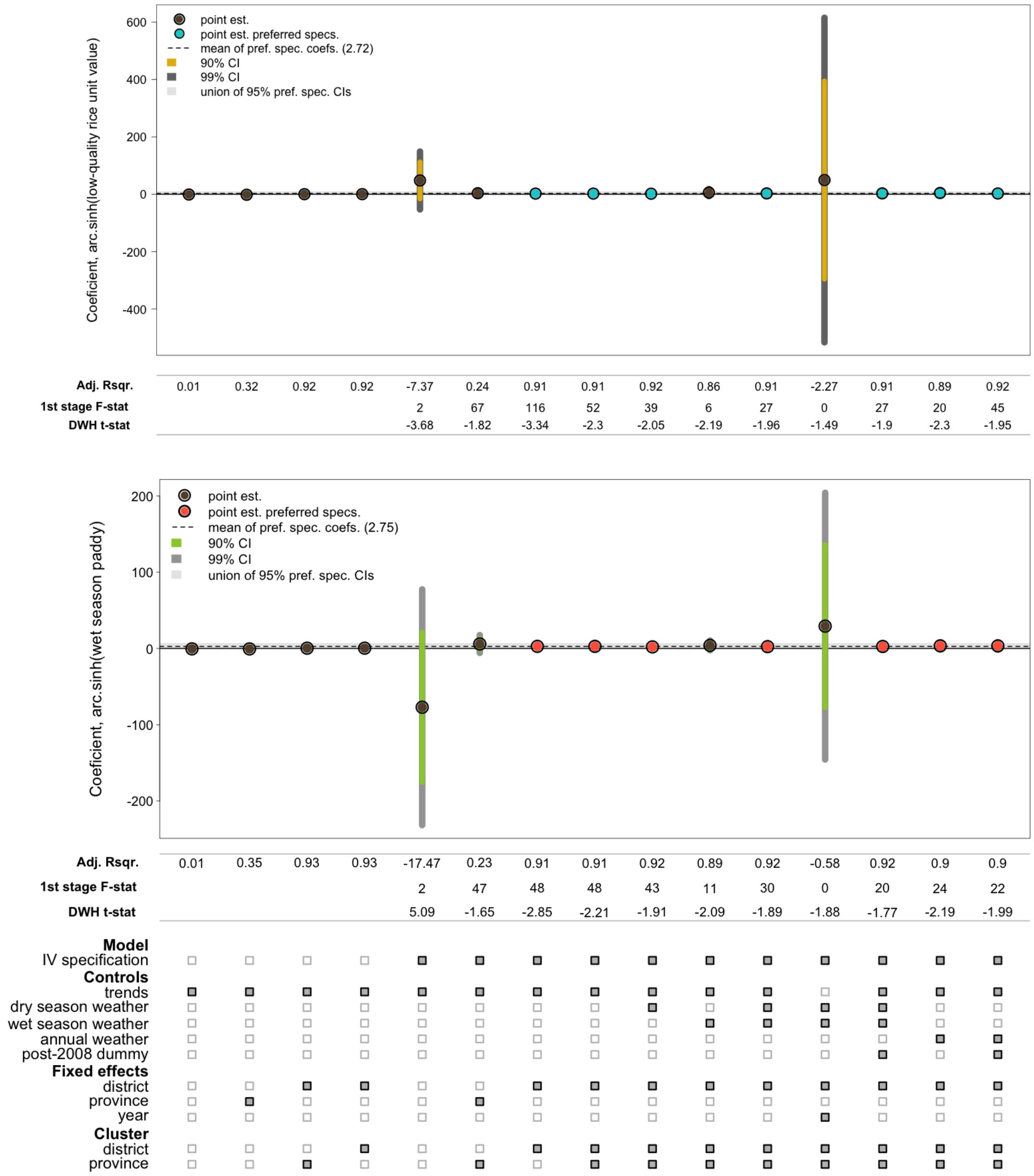

Figure 18: Identical side by side comparison of point estimates to the immediately preceding figure 17, except for columns 5 and 12, which are omitted from the main text for obvious scale reasons. Columns 5 shows the result of an IV specification without any fixed effects. Column 12 demonstrates what happens when year fixed effects are included.

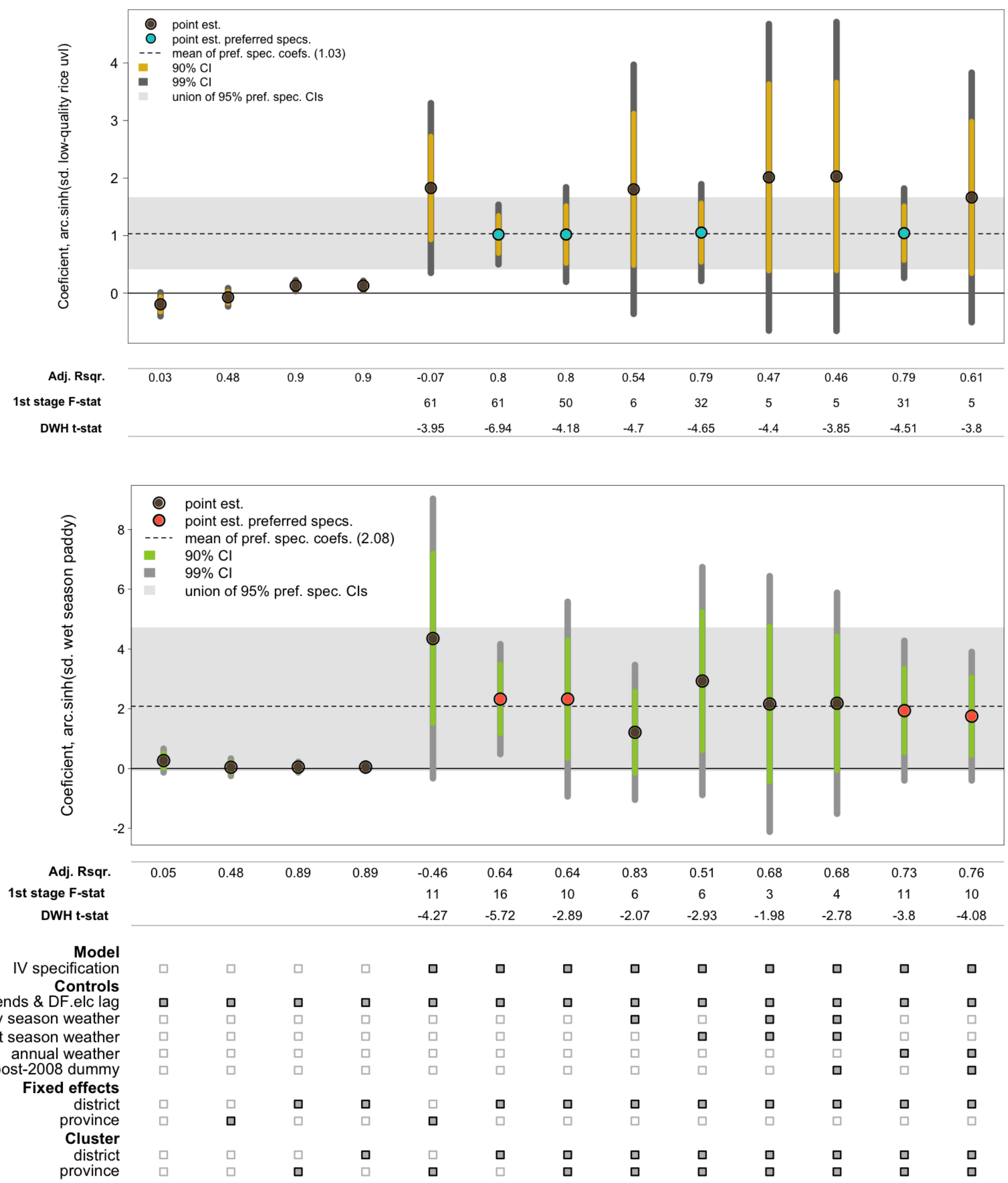

Figure 19: Side by side comparison of point estimates of the arcsinh of district-level rice price standard deviation (sd) in t regressed on arcsinh(district-level non-ELC hectares deforested) in $t+1$: top panel shows point estimates of arcsinh(sd low-quality rice unit values) ($N = 755$; $t \in (2004, 2007-2013)$; 172 districts, 24 provinces); bottom panel shows point estimates of arcsinh(sd wet season paddy) ($N = 810$; $t \in (2004, 2007-2013)$; 179 districts, 24 provinces).

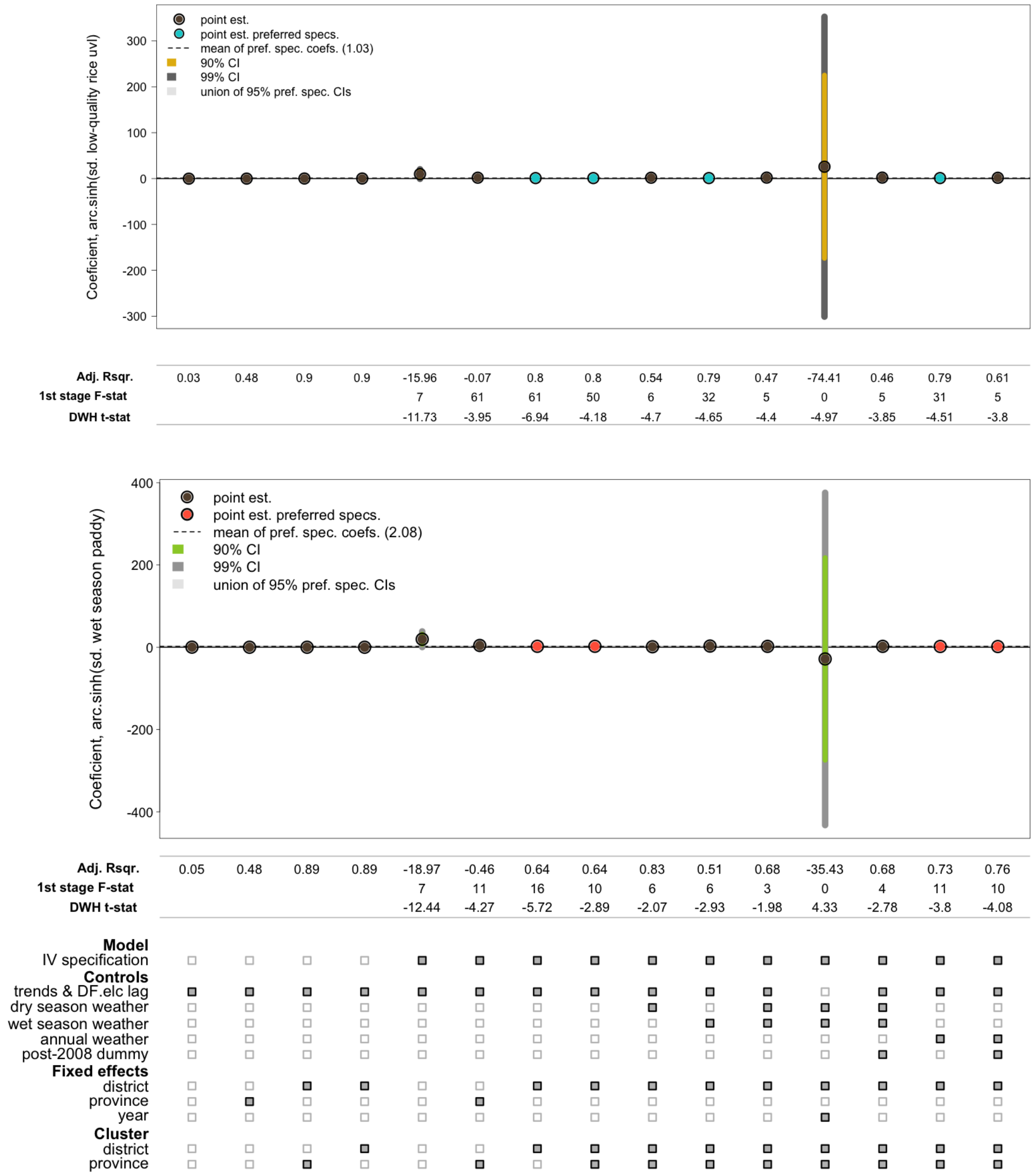

Figure 20: Identical side by side comparison of point estimates to the immediately preceding figure 19, except for columns 5 and 12, which are omitted from the main text for obvious scale reasons. Columns 5 shows the result of an IV specification without any fixed effects. Column 12 demonstrates what happens when year fixed effects are included.

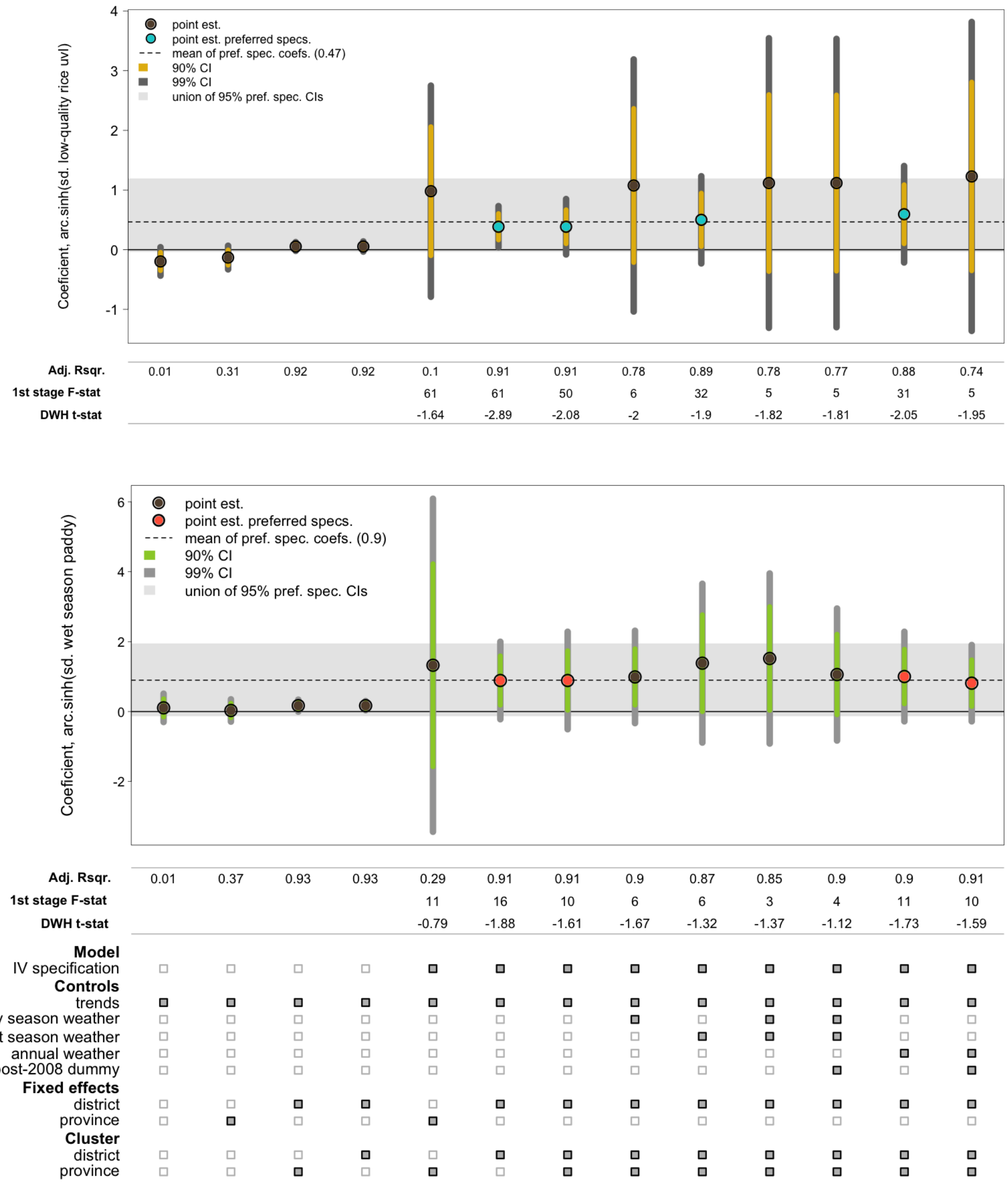

Figure 21: Side by side comparison of point estimates of the arcsinh of district-level rice price standard deviation (sd) in t regressed on arcsinh(district-level non-ELC hectares deforested) in $t+1$: top panel shows point estimates of arcsinh(sd low-quality rice unit values) ($N = 755$; $t \in (2004, 2007-2013)$; 172 districts, 24 provinces); bottom panel shows point estimates of arcsinh(sd wet season paddy) ($N = 810$; $t \in (2004, 2007-2013)$; 179 districts, 24 provinces).

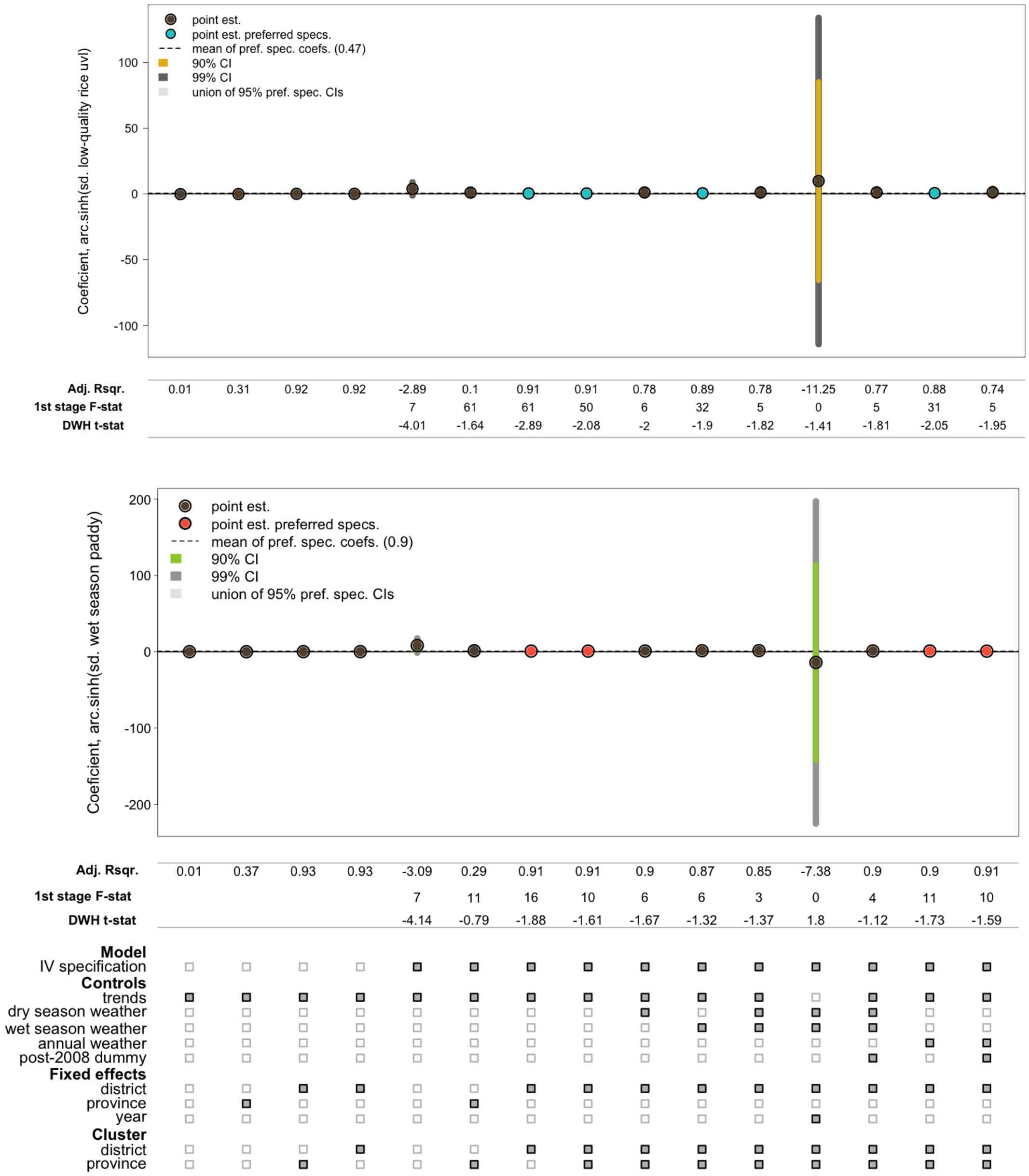

Figure 22: Identical side by side comparison of point estimates to the immediately preceding figure 21, except for columns 5 and 12, which are omitted from the main text for obvious scale reasons. Columns 5 shows the result of an IV specification without any fixed effects. Column 12 demonstrates what happens when year fixed effects are included.

Table 1: District-level non-ELC deforestation in $t+1$ on district-level mean low-quality rice unit-values in t. Instrument is the interaction between the average US-based international rice price in dollars per kg and district-level centroid road distances to Cambodia's deep sea port. N = 786 observations.

	(1)	(2)	(3)	(4)
linear trend	-0.12	-1.15***	-0.21*	-0.42**
	(0.09)	(0.28)	(0.11)	(0.15)
district trends	-0.0000	0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
arcsinh(ELC deforestation in t)	0.12**	0.12*	0.11*	0.12**
	(0.05)	(0.06)	(0.06)	(0.06)
avg. wet season temp (c) in t	1.86***	-3.89***		
	(0.63)	(0.95)		
avg. wet season rainfall (mm) in t	0.005*	-0.002		
	(0.002)	(0.002)		
avg. dry season temp (c) in t	-0.96***	0.20		
	(0.26)	(0.23)		
avg. dry season rainfall (mm) in t	0.002	-0.04***		
	(0.005)	(0.01)		
2009-2011 dummy		3.96***		1.38***
		(0.83)		(0.32)
avg. annual temp (c) in t			0.30	-0.74**
			(0.42)	(0.31)
avg. annual rainfall (mm) in t			0.01*	0.001
			(0.005)	(0.002)
arcsinh(avg.lq.rice.uvl in t)	4.49***	4.59***	6.65***	3.06***
	(1.35)	(1.58)	(1.99)	(0.96)
FEs?	D	D	D	D
Clust. S.E.?	P,D	P,D	P,D	P,D
iv1stg. F	27	27	20	44
Adjusted R^2	0.88	0.89	0.82	0.90

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2: District-level non-ELC deforestation in $t+1$ on district-level mean wet season paddy in in t. Instrument is the interaction between the average US-based international rice price in dollars per kg and district-level centroid road distances to Cambodia's deep sea port. N = 988 observations.

	(1)	(2)	(3)	(4)
linear trend	0.07	0.14	-0.02	-0.05
	(0.07)	(0.11)	(0.08)	(0.09)
district trends	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
arcsinh(ELC deforestation in t)	0.17^{***}	0.18^{***}	0.14^{**}	0.15^{***}
	(0.05)	(0.05)	(0.05)	(0.05)
avg. wet season temp (c) in t	0.51	1.16		
	(0.34)	(0.73)		
avg. wet season rainfall (mm) in t	0.001	0.002		
	(0.001)	(0.001)		
avg. dry season temp (c) in t	-0.37^{***}	-0.57^{**}		
	(0.11)	(0.22)		
avg. dry season rainfall (mm) in t	0.01^{**}	0.01^{***}		
	(0.004)	(0.004)		
2009-2013 dummy		-0.37		0.45^{**}
		(0.32)		(0.17)
avg. annual temp (c) in t			-0.01	-0.17
			(0.32)	(0.30)
avg. annual rainfall (mm) in t			0.01	0.004
			(0.003)	(0.003)
arcsinh(avg.rice.wsp in t)	2.73^{**}	3.07^{**}	6.51^{***}	4.90^{***}
	(1.09)	(1.12)	(1.84)	(1.22)
FEs?	D	D	D	D
Clust. S.E.?	P,D	P,D	P,D	P,D
iv1stg. F	30	24	25	44
Adjusted R^2	0.88	0.87	0.78	0.83

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: Sample regression specifications with year fixed effects.

	(1)	(2)	(3)	(4)	(5)	(6)
linear trend	0.19***		0.10***		-1.15***	0.14
	(0.06)		(0.04)		(0.28)	(0.11)
district trends	-0.0002***	-0.0002***	-0.0000	-0.0003	0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.001)	(0.0000)	(0.0000)
arcsinh(ELC deforest in *t*)	0.45***	0.39***	0.14***	0.18	0.12*	0.18***
	(0.03)	(0.03)	(0.05)	(0.51)	(0.06)	(0.05)
avg. wet season temp (c) in *t*	-0.81***	-0.68*	-0.15	0.36	-3.89***	1.16
	(0.25)	(0.37)	(0.27)	(34.27)	(0.95)	(0.73)
avg. wet season rainfall (mm) in *t*	0.001	0.003***	-0.002**	-0.03	-0.002	0.002
	(0.001)	(0.001)	(0.001)	(0.09)	(0.002)	(0.001)
avg. dry season temp (c) in *t*	-1.00***	-1.23***	-0.44***	-13.53	0.20	-0.57**
	(0.15)	(0.28)	(0.11)	(39.68)	(0.23)	(0.22)
avg. dry season rainfall (mm) in *t*	-0.02***	-0.03***	0.0002	0.17	-0.04***	0.01***
	(0.004)	(0.01)	(0.003)	(0.77)	(0.01)	(0.004)
2009-2011 dummy					3.96***	
					(0.83)	
arcsinh(avg.lq.rice.uvl in *t*)	-1.13**	-1.76***	-0.17	116.63	4.59***	
	(0.47)	(0.56)	(0.31)	(442.53)	(1.58)	
2009-2013 dummy						-0.37
						(0.32)
arcsinh(avg.rice.wsp in *t*)						3.07**
						(1.12)
Constant	63.16***					
	(6.44)					
FEs?	N	N	D	D	D	D
year FEs?	N	Y	N	Y	N	N
Clust. SEs?	N	N	N	P,D	P,D	P,D
1st stage F_R	NA	NA	NA	0	27	24
Observations	786	786	786	786	786	988
Adjusted R^2	0.52	0.54	0.91	-15.13	0.89	0.87

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4: Reduced form (column 1) and example specifications of first stage contrasting effects of including year fixed effects. N = 786 observations. The dependent variable is low quality rice unit values.

	(1)	(2)	(3)	(4)	(5)	(6)
linear trend	0.04207	0.04976***			0.03658***	0.12402***
	(0.05428)	(0.00611)			(0.00646)	(0.01844)
district trends	-0.000002	-0.000002	-0.000005	0.000003	0.000003	0.000001
	(0.00003)	(0.000005)	(0.000003)	(0.000004)	(0.000005)	(0.000004)
arcsinh(ELC deforestation in *t*)	0.11974**	-0.00626**	-0.00509*	-0.00054	-0.00084	-0.00015
	(0.05541)	(0.00247)	(0.00288)	(0.00392)	(0.00425)	(0.00524)
avg. wet season temp (c) in *t*	0.28532	-0.06153	0.09059	-0.04006	-0.35054***	0.14138
	(0.18168)	(0.05176)	(0.05630)	(0.28472)	(0.04819)	(0.09711)
avg. wet season rainfall (mm) in *t*	0.00006	-0.00004	0.00045***	0.00020	-0.00104***	-0.00047***
	(0.00097)	(0.00015)	(0.00009)	(0.00028)	(0.00017)	(0.00013)
avg. dry season temp (c) in *t*	-0.34446*	0.03395	-0.08221**	0.11566	0.13697***	0.03757*
	(0.19371)	(0.02717)	(0.03721)	(0.25952)	(0.01747)	(0.02157)
avg. dry season rainfall (mm) in *t*	-0.00340	0.00080**	-0.00145	-0.00154	-0.00128***	0.00231**
	(0.00393)	(0.00038)	(0.00090)	(0.00129)	(0.00045)	(0.00087)
Z	0.00384***	0.00016	-0.00039**	0.00006	0.00085***	0.00085***
	(0.00095)	(0.00012)	(0.00016)	(0.00024)	(0.00016)	(0.00016)
2009-2011 dummy						-0.33708***
						(0.06889)
Constant		8.79105***				
		(0.90132)				
FEs?	D	N	N	D	D	D
Year FEs?	N	N	Y	Y	N	N
Clust. S.E.?	P,D	P,D	P,D	P,D	P,D	P,D
Adjusted R^2	0.92	0.28	0.53	0.69	0.65	0.67

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5: Reduced form (column 1) and example specifications of first stage contrasting effects of including year fixed effects. N = 988 observations. The dependent variable is wet season paddy farm gate prices.

	(1)	(2)	(3)	(4)	(5)	(6)
linear trend	0.12334*	0.03456***			0.01875***	0.00456
	(0.06904)	(0.00407)			(0.00512)	(0.00938)
district trends	-0.00002	0.0000004	0.000001	0.000003	0.00001*	0.00001*
	(0.00004)	(0.000003)	(0.000002)	(0.000003)	(0.000003)	(0.000004)
arcsinh(ELC deforestation in *t*)	0.18048***	-0.00606***	-0.00422**	0.00246	0.00203	0.00115
	(0.05156)	(0.00222)	(0.00189)	(0.00638)	(0.00585)	(0.00588)
avg. wet season temp (c) in *t*	0.07199	0.00250	0.04370	0.03610	-0.15904***	-0.27333**
	(0.19299)	(0.02621)	(0.03201)	(0.20117)	(0.05489)	(0.10877)
avg. wet season rainfall (mm) in *t*	-0.00033	0.00002	0.00036***	0.00032	-0.00033*	-0.00050*
	(0.00073)	(0.00014)	(0.00010)	(0.00037)	(0.00019)	(0.00026)
avg. dry season temp (c) in *t*	-0.23738*	-0.04146***	-0.03987*	-0.02400	0.05011**	0.08400**
	(0.12229)	(0.01366)	(0.02380)	(0.09201)	(0.01841)	(0.03195)
avg. dry season rainfall (mm) in *t*	0.00115	-0.00130***	-0.00163***	-0.00139*	-0.00342***	-0.00372***
	(0.00208)	(0.00023)	(0.00040)	(0.00074)	(0.00032)	(0.00043)
Z	0.00264**	-0.00027***	-0.00045***	0.00011	0.00097***	0.00091***
	(0.00100)	(0.00006)	(0.00010)	(0.00025)	(0.00018)	(0.00018)
2009-2013 dummy						0.06989
						(0.04224)
Constant		8.52620***				
		(0.48363)				
FEs?	D	N	N	D	D	D
Year FEs?	N	N	Y	Y	N	N
Clust. S.E.?	P,D	P,D	P,D	P,D	P,D	P,D
Adjusted R^2	0.90	0.14	0.36	0.46	0.38	0.38

Note: *p<0.1; **p<0.05; ***p<0.01

Note on weather variation effects in the first stage:

In the above tables a variety of statistically significant relationships are exhibited between price variation and contemporaneous prices. For mean prices in the first stage, generally negative and significant effects of wet season temperature and dry and wet season rainfall are apparent, but positive and significant effects effects of dry season temperature are also clear. For price variance, somewhat weaker relationships with weather overall are observed, but some changes in sign with dry season temperature and wet season rainfall being negative and wet season temperature and dry season rainfall being positive[2].

These weather covariate signs are noteworthy and can be rationalized to a good degree. First, we expect that contemporaneous weather will have an effect on prices through expectations and within-year yield impacts. Second, most rainfall in Cambodia is rain-fed, hence increased rain on average should shift supply upward resulting in downward pressure on mean prices. Third, the relevance of maximum and minimum temperatures to farm-level rice yield in Asia has been established by Welch et al. (2010) who find that higher maximum temperatures (beneath a threshold of 35 degrees C) increase yield and higher minimum temperatures decrease yield. Our finding of a negative coefficient on wet season average maximum temperature in the first stage for mean prices is consistent with the expected response (i.e. prices decline with increased supply). The positive sign on dry season temperature could relate to declining yields for the dry season crop, which could put upward pressure on prices.

The fact that within season temperature and rainfall would have opposing effects on prices variance, and that across seasons the signs would flip between rainfall and temperature respectively is puzzle, which we have little existing information to aid interpretation.

Note on Correlation of Z with Mango, Coconut Unit Values, Other Prices:

The two tables below present our results to address the issue of possible correlation of our primary instrument Z with mango and coconut unit values and potential for these price changes to be driving our results. We instrument for low quality rice unit values, and either mango or coconut unit values, with wet season average maximum temperature[3] in $t-1$ and our triple interaction instrument Z. These models confirm our general results with significance at the 1% level.

Diagnostic tests for instrumental variable models with multiple endogenous variables are not well developed, particularly for nonhomeskedastic settings. To assess instrument strength we follow Sanderson and Windmeijer (2016) who develop conditional F statistics, $F_{i|j}$, for each first stage conditional on the other first stages[4] , which we compare with standard robust F statistics.

[2] First stage results for price variance are omitted from the appendix but can be provided upon request.

[3] Welch et al. (2010) find that higher maximum temperatures (beneath a threshold of 35 degrees C) increase yield and higher minimum temperatures decrease yield. Caution is needed in extrapolating farm-level results to aggregate relationships. Nevertheless, our finding of a negative coefficient on wet season average maximum temperature in $t-1$ in the first stage is consistent with the expected response (i.e. prices decline with increased supply).

[4] Sanderson and Windmeijer (2016) build upon an analogous approach suggested by Angrist and Pischke (2009). With two endogenous variables Sanderson and Windmeijer (2016) establish that $F_{i|j}$ is equivalent to the Cragg and Donald (1993) statistic, unless there is no correlation between predicted endogenous variables $\hat{x}_j$ and the other endogenous variables (i.e. when $\delta = 0$ in $\hat{\delta} = (\hat{X}_j'\hat{X}_j)^{-1}\hat{X}_j'X_i$), in which case $F_{i|j}$ provides more information above the Cragg and Donald (1993) (i.e. the Cragg and Donald (1993) statistic is global for all first stages and $F_{i|j}$ is not).

Instrument strength generally seems to remain in the safe range for our regressors of interest (i.e. $F_{rice|coconut}$ and $F_{rice|mango}$).

In homoskedastic settings Sanderson and Windmeijer (2016) establish that Stock and Yogo (2005) critical values can be used to relate instrument strength $F_{i|j}$ to bias and size distortions for each first stage[5]. They also establish that in homoskedastic settings an endogenous coefficient of interest can be consistently estimated if its respective first stage shows strong instruments, even if the first stage for the other endogenous regressor suffers from weak instrument bias. Although we are not in a homoskedastic setting, there are no other available tests in a multiple endogenous variable setting (Andrews, Stock, and Sun 2019), thus these statistics can be seen as an approximation.

These caveats notwithstanding, in each specification we see that coefficients on non-rice unit values turn negative and magnitudes and significance for low-quality rice unit values are consistent in sign and magnitude with preceding results. Instrument strength is in the safe range for the respective first stages for rice unit values, and for mango unit values, but somewhat weak for coconut unit values. Regression based tests for exogeneity (following Wooldridge 2010) reject exogeneity for rice unit values, but not for coconut of mango unit values (the result does not change if uninstrumented versions of the latter two unit values are included). These findings lend further support for our identification strategy and results.

Figure 23: District level specifications of non-ELC deforestation on instrumented mango-uvl and wet season paddy. Specification uses shift-share instrument Z and wet season average maximum temperature in t-1.

arcsinh(mango.uvl in *t*)	-0.66
	(0.72)
arcsinh(avg.lq.uvl in *t*)	5.44^{***}
	(1.87)
FEs?	D
Clust. S.E.?	P,D
1st stage F_R mango	21
1st stage $F_{\text{mango\|rice}}$	9
1st stage F_R rice	11
1st stage $F_{\text{rice\|mango}}$	15
DWH t-stat. rice	-3.69
DWH t-stat. mango	1.3
Observations	651
Adjusted R^2	0.87

Note: *p<0.1; **p<0.05; ***p<0.01

[5] Applicable tables are for one endogenous variable with the critical value corresponding to the total number of instruments minus the number of first stages.

Figure 24: District level specifications of non-ELC deforestation on instrumented coconut-uvl and wet season paddy. Specification uses shift-share instrument Z and wet season average maximum temperature in t-1.

arcsinh(coconut.uvl in t)	-2.52^{*}
	(1.44)
arcsinh(avg.lq.uvl in t)	3.42^{***}
	(1.18)
FEs?	D
Clust. S.E.?	P,D
1st stage F_R coconut	4
1st stage $F_{\text{coconut}\mid\text{rice}}$	8
1st stage F_R rice	7
1st stage $F_{\text{rice}\mid\text{coconut}}$	46
DWH t-stat. rice	-4.13
DWH t-stat. coconut	2.57
Observations	657
Adjusted R^2	0.73

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Deforestation Robustness and Extensions: placebo test & impulse response, leave-k-out & permutation tests

Figures (25) and (26) below show combined results for distributions of respective statistics of interest based on dropping districts (i.e. the main cluster variable, which is districts in our case). In each regression the same specification is run. Covariates include wet and dry season average temperature and precipitation, lagged ELC deforestation t, a linear trend, district specific trends, and a post-2008 dummy. District level fixed effects are included and standard errors are clustered at the province and district levels. These figures focus on district-level specifications with mean rice prices as the regressor of interest in period t, regressed on non-ELC deforestation in $t+1$; figures focused on different dependent variables or price moments can be provided upon request. For a detailed study of these issues see Young (2019) who focuses largely on the value of such tests to assess sensitivity to highly leveraged data in instrumental variables estimation. Also see further discussion of the relevance of these tests in Christian and Barrett (2021) who argue for the value of such tests in panel shift-share IV settings to assess the extent to which spurious time-series correlations may be driving results.

In each panel within figures (25) and (26), the observed estimates from our main results are plotted, as well as the resulting distributions of these statistics when one either: leaves-exactly-one district out; leaves two randomly selected districts out over 5000 iterations; leaves four randomly selected districts out over 5000 iterations. These tests easily pass criteria suggested by Young (2019) such as the bootstrap-c test (i.e. "...the tail probability of the squared coefficient deviation from the null hypothesis", or, more formally, $(\hat{\beta}^i_{out} - \hat{\beta}_{obs.})^2 > (\beta - 0)^2$) since all estimated coefficients of interest are positive. In all cases, statistics remain quite precisely estimated, though commensurate with losses of degrees of freedom as more districts are dropped, imprecision increases, though main results continue to hold. Results of other statistics of interest can be provided upon request. Overall, the results presented in figures (25) and (26) do not support the hypothesis that our results stem from our data be over-leveraged.

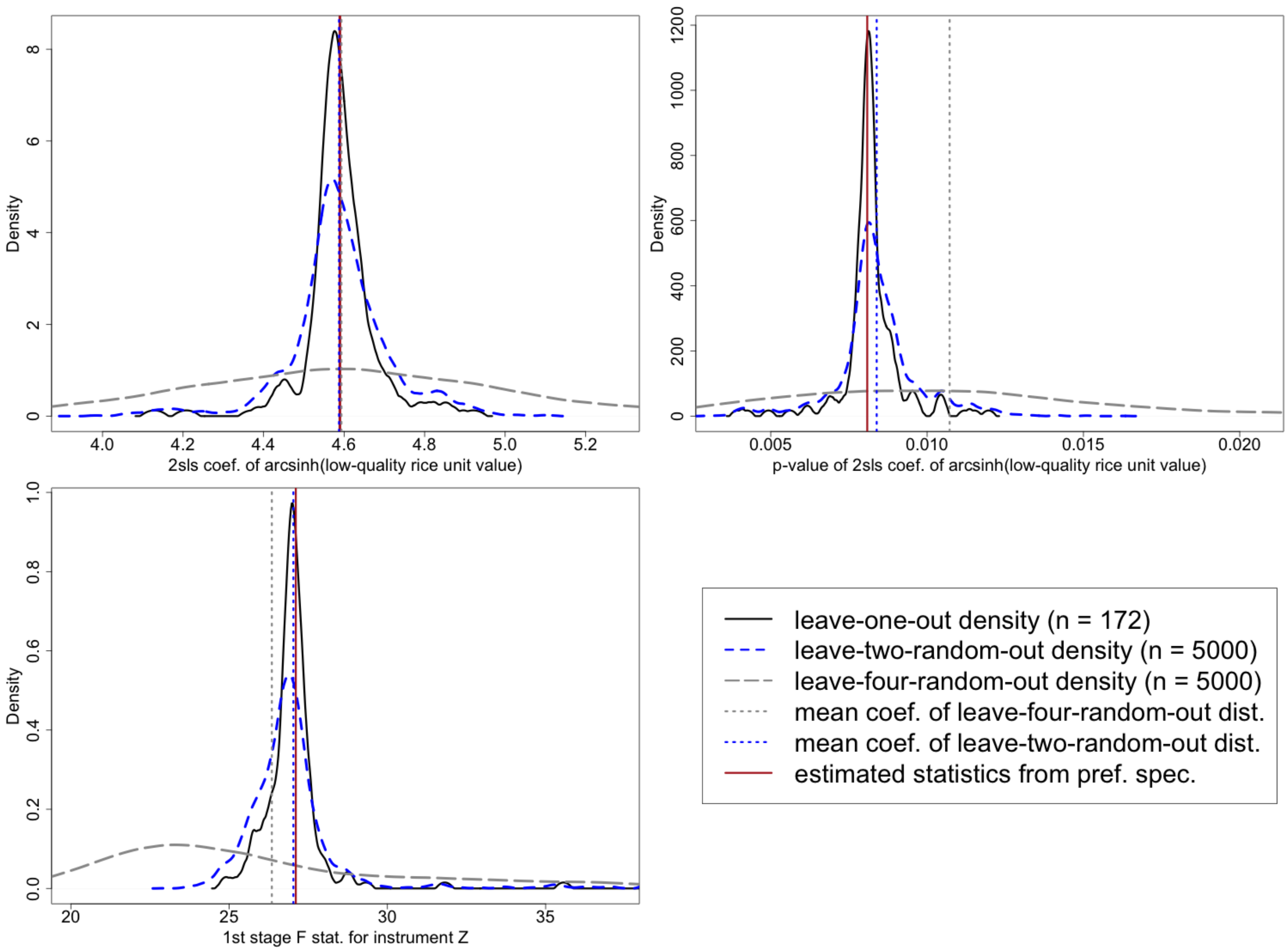

Figure 25: District-level leave-several-out tests for arcsinh(mean low-quality rice unit value) in t regressed on arcsinh(non-ELC deforestation hectares) in $t+1$.

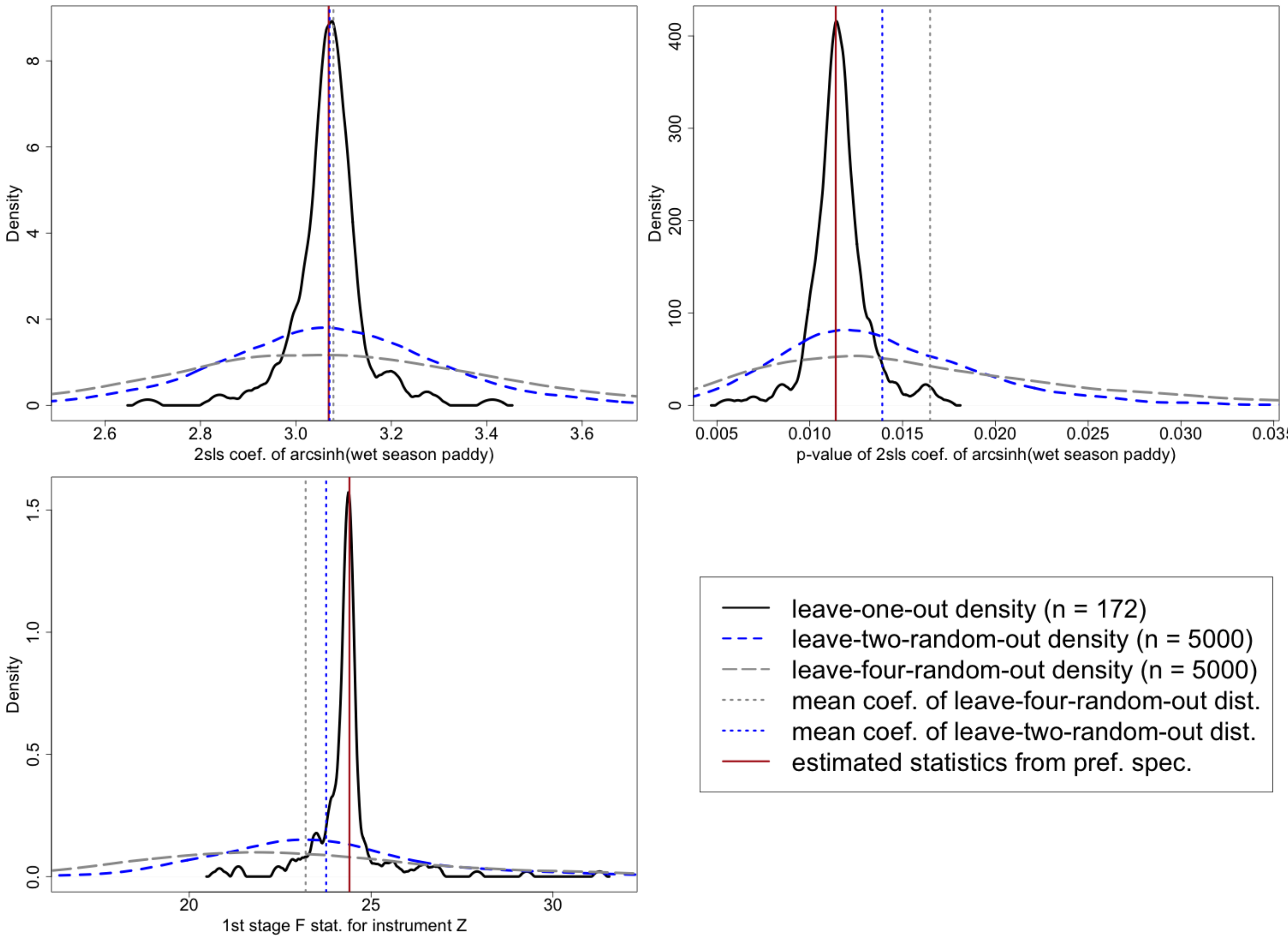

Figure 26: District-level leave-several-out tests for arcsinh(mean wet season paddy) in t regressed on arcsinh(non-ELC deforestation hectares) in $t+1$.

Figures (27) and (28) below show resulting distributions for statistics of interest based on randomly re-assigning the respective endogenous variable (low-quality rice unit values, or wet season paddy in period t) within two-stage least squares estimation. As above, these figures focus on mean rice prices in t as the regressor of interest in an instrumental variables regression with non-ELC deforestation in period $t+1$ as the dependent variable. Christian and Barrett (2021) argue for the value of such tests in panel shift-share IV research designs as part of robustness checks to guard against spurious time series correlation contaminating estimations of interest.

In the setting focused on by Christian and Barrett (2021), the relationship of interest is the causal effect of US food aid shipments on conflict, focusing in particular on the case of Nunn and Qian (2014) who use a shift-share research design to address this question. Christian and Barrett (2021) make compelling arguments that Nunn and Qian (2014), among others, are vulnerable to being affected by spurious time-series correlations, which may lead to type I error. Christian and Barrett (2021) make a variety of useful recommendations for panel shift-share IV research designs, particularly for long panels, to increase transparency and elevate the rigor that such research designs should undergo. We find Christian and Barrett (2021) convincing and agree on the value of applying applicable tests, such as leave-one-out estimation and permutation tests. However, there are limits to which we can apply all recommended tests from Christian and Barrett (2021). Moreover, in our setting we suggest that some of their recommended tests/steps are not viable, and/or would need to be applied with greater flexibility to more comprehensively implement the implied test.

The caveats our study raises are as follows. First, Christian and Barrett (2021) suggest applying tests for time series coeintegration, and, if tests come back in the affirmative, resolve the issue by taking first differences. While this approach might be viable in long panels, such Nunn and Qian (2014) with over 20 years of data, in comparatively short panels such as ours (6 years for low-quality rice unit value models, 8 years for wet season paddy models), which are also imbalanced and not uniformly contiguous in time, first differences may cause as many problems as they resolve. In our cases, the implied complications includes serious losses to degrees of freedom, and losing critical pre- and post- rice price shock variation. For this reason, we do not implement such tests or differencing.

Second, while we agree with Christian and Barrett (2021) that permutation tests that systematically seek to break the causal links of interest by scrambling the endogenous variable of interest are very important and should be more widely adopted, we suggest these tests should be approached with some flexibility and caution. In applying a permutation test to Nunn and Qian (2014), and in making a general recommendation to researchers, Christian and Barrett (2021) recommend randomizing within respective time periods and sampling without replacement. In the setting of Nunn and Qian (2014), this approach makes good sense, particularly because the units of observation are countries, between which there is less likely to be pervasive spatial correlation for the endogenous variable of interest (i.e. US food aid shipments to different countries, across oceans and continents, are highly unlikely to be systematically, spatially correlated). However, for research questions that have a narrow geographic focus, which are also focused on endogenous variables with high cross-sectional spatial correlation, restricting randomization to be within each time period and without replacement may not be the most appropriate test. The reason being that if the endogenous variable within cross-section is highly spatially correlated, randomization may largely retain very close variation to the observed, leading to very similar results overall. This might lead one to conclude that the main findings cannot therefore be genuine, even when identification is clean and the causal-relationship(s) of study does indeed exist. Rejecting the validity of estimation based on such tests in applicable settings would be equivalent to a type II error: one would reject the null that $\beta = 0$ from the second stage, but then find reason to question the validity of that test result because it fails the permutation test.

We submit that our research presents just such a setting: prices for staple foods, particularly within cross-section because the cross-section is contiguous in space, are likely to be highly spatially correlated. Because of this, a within-year test that randomizes without replacement, may produce findings with a randomized endogenous variable that show significant results, even commensurate in magnitude and significance of the observed results. For this reason, we argue that a wider battery or permutation tests are likely to be a wise in such settings, and that increasing the scope of tests applied should add further rigor, and also effectively guard against un-intended type II errors.

In particular, in the figures and panels below we present results of the following: (i) permutation tests that randomize the endogenous variable within cross-section and without replacement, randomize within period t and $t+1$ without replacement, and randomize across all time periods without replacement; (ii) for each test, study the resulting second stage coefficients of interest, the first stage relationship of interest, and standard IV test statistics (e.g. Durban-Wu-Hausman (DWH) tests, 1st stage F statistics, etc.). With this array of comparisons we can then compare all distributions, means of distributions, and compare these results against the respective observed statistics that are estimated from the models of interest.

Our results for this battery of tests are presented in the figures below. As before, note also that all specifications estimated are identical with the exception of the endogenous variable, which may be low-quality rice unit values or wet season paddy. Covariates include wet and dry season average temperature and precipitation, lagged ELC deforestation t, a linear trend, district specific trends, and a post-2008 dummy. Each model also includes district fixed effects and clustered standard errors at the province and district levels.

Figure 27 shows permutation test results for specifications with low-quality unit values as the regressor of interest in t and non-ELC deforestation as the outcome of interest in $t+1$, and 28 shows the corresponding results with the same response variable when wet season paddy is the regressor of interest. One overall result that stands out is the concern about cross-sectional spatial correlation in rice prices. Indeed, we suggest that our results imply this is a well-founded concern in that we can see that although most tests show estimated parameters stand out as well-outside observed distributions, the same is not the case for low-quality unit values, at least uniformly. We suggest that a simple explanation for this phenomena may be the overall nature of the differences between the two rice products. In particular, since low-quality rice is a very common consumption good it is not likely to vary substantially across the country, hence differences in price in cross-section are more likely to be smaller. In contrast, wet season paddy is more likely to have greater variation over space – perhaps for agro-ecological reasons, access to irrigation, production technology, plant materials used, etc. For example, rice in regions with high quality rice, such as Battambang or Takeo, may receive a consistently much higher price than that from, say, areas of Ratanakiri. If this was indeed the case, then in cross-section one might expect greater spatial variation for the respective farm gate price than that for the retail price of low-quality rice unit values.

The most clear example of how contrasting spatial correlation seems to affect these tests, is the top left panel of each figure 27 and 28, which show estimated second stage coefficient distributions against the corresponding estimate from the actual data. Within cross-sectional randomization, the low-quality rice unit value distribution exhibits similar magnitudes from the observed estimate, but is generally larger than the observed value. Wet season paddy distributions exhibit similar properties. This may seem a problem as it may suggest that our results are spurious. However, it is also clear that as soon as randomization is allowed to extend into t and $t+1$, the relationship breaks as shown by the imprecise underlying density in blue within the top left panel of both figures, particularly for wet season paddy. Moreover, as can be gleaned from the other panels within each figure, there are plenty of other results that lend confidence in accepting our results: significance of either instrumented rice price in the second stage drops (top right panel in each figure); the instrument becomes steadily less significant to insignificant in the first stage (center-left panel); the instrument becomes substantially weaker and very imprecisely estimated (center-right panel); and, the observed DWH statistic in randomized distributions is quite precisely centered over zero, even within cross-sectional randomization.

In combination, we suggest that cross-sectional spatial correlation is the driver behind some of the less-convincing permutation tests shown in these figures, but that the additional results – even within cross-sectional randomization – broadly do not support the notion that our results are driven spurious time series correlation, or some other form of spurious correlation.

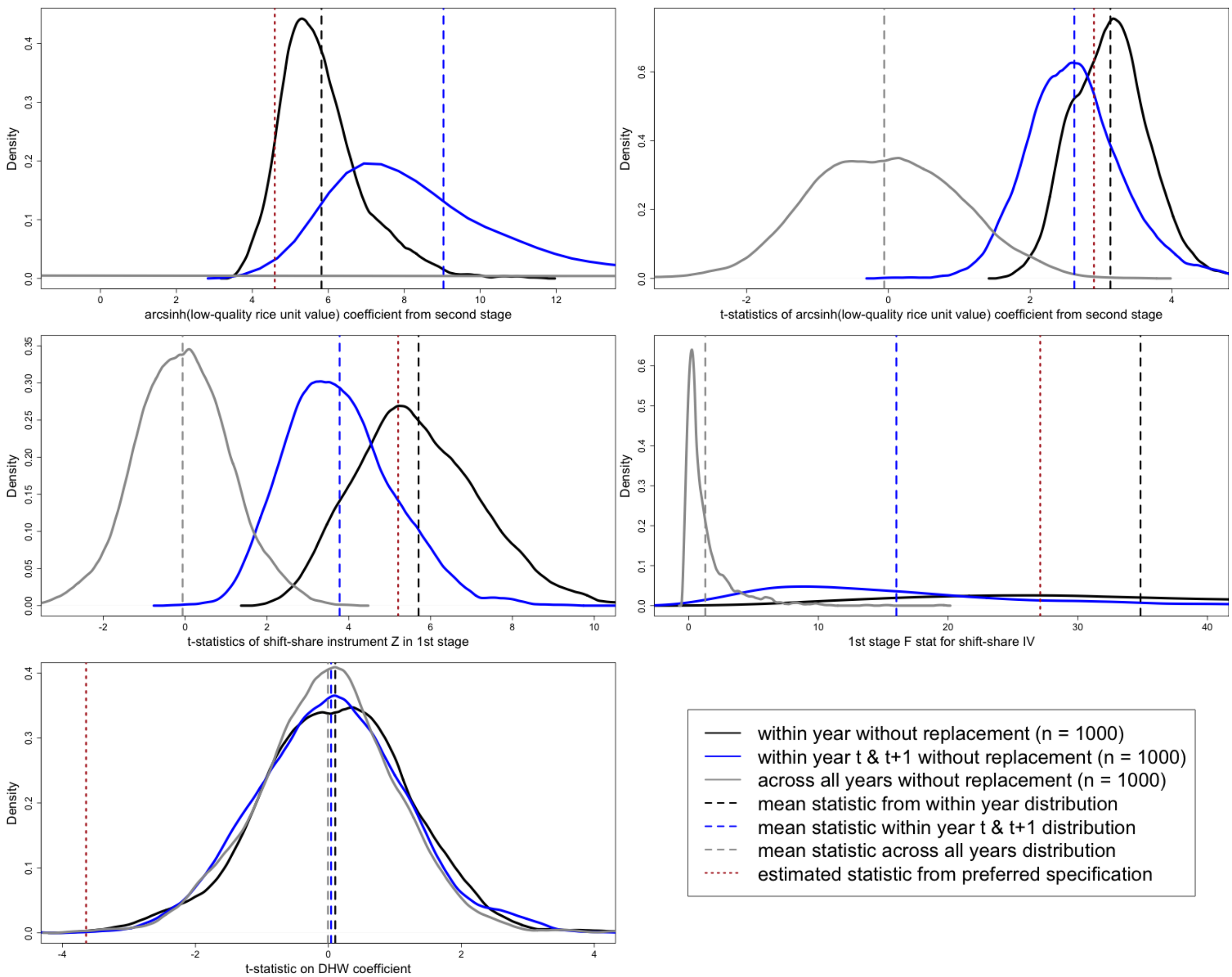

Figure 27: Permutation tests for regressions of mean acrsinh(low-quality rice unit value) in t on arcsinh(non-ELC deforestation hectares) in $t+1$. Each panel shows distributions and means of respective statistics under randomization of the endogenous variable within year and without replacement, within year t and year $t+1$ without replacement, and across all years without replacement.

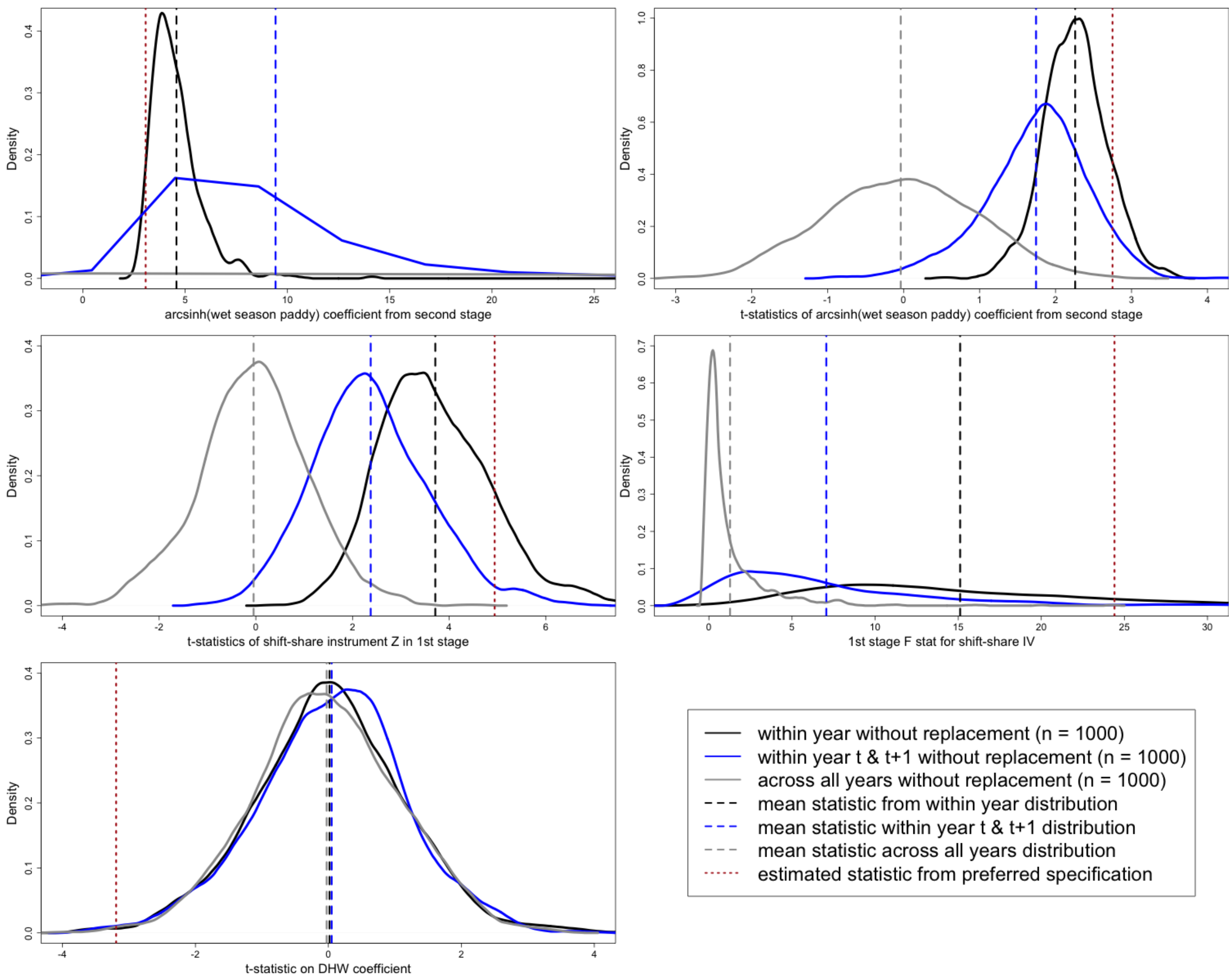

Figure 28: Permutation tests for regressions of mean acrsinh(wet season paddy) in t on arcsinh(non-ELC deforestation hectares) in $t+1$. Each panel shows distributions and means of respective statistics under randomization of the endogenous variable within year and without replacement, within year t and year $t+1$ without replacement, and across all years without replacement.

An additional battery of district-level tests are shown in figures 29 and 30, which respectively show placebo tests and tests for dynamic impacts beyond the base specifications that we focus on with rice prices in t and deforestation in $t+1$. Our dynamic estimation relies on the local projections method studied by Jordà (2005), sometimes referred to as impulse response estimation. Figure 29 shows respective tests for deforestation in non-ELC areas and figure 30 shows respective tests for deforestation in ELC areas. The x-axis in each panel represents different leads and lags $t+j$ that deforestation is set to with respect to rice price variation in t. Therefore, all lags $t-3$, $t-2$, and $t-1$ are placebo tests; t is a test for contemporaneous effects (plausible, but less likely); $t+1$ is our base model and $t+2$ up to $t+6$ test the extent to which further dynamic land use impacts are present from the rice price shock beyond those measured at $t+1$. For these figures, specifications differ mainly by whether or not the first or second moment of rice prices is at focus. For all specifications focused on mean prices, every specification has a linear trend, district trend, wet and dry season weather covariates in t, a lag of ELC deforestation in t (for non-ELC specifications), a post-2008 dummy, district level fixed effects, and clustered standard errors at the province and district levels. For all specifications focused on the standard deviation of prices, specifications differ slightly, given the weak IV issues that arise with seasonal weather covariates, and include a linear trend, district trend, a lag of ELC deforestation in t (for non-ELC specifications), and annual weather covariates in t.

An important first finding to note is that the vast majority of the placebo tests are not significant at any level, and sign is inconsistent. A few instances of significance do arise in placebo tests, though mostly for specifications focused on standard deviation and ELC-based deforestation (a couple models focused on mean wet season paddy and ELC-deforestation are also marginally significant in placebo test). Some spurious correlation is not surprising owing to serial correlation. Another important finding is that contemporaneous coefficient estimates are positive, generally smaller than respective $t+1$ coefficients, and inconsistently significant. Moving to our base specifications, we see further confirmation of our deforestation results. For non-ELC areas we see positive and large elasticities that are highly significant for mean rice prices, and much smaller but still positive and significant for our measure of price variance (standard deviation). For ELC areas, the story for our baseline model is similar, but magnitudes of point estimates are approximately half of the magnitude of the non-ELC focused point estimates, and statistical significance is somewhat weaker.

Tests for the presence of further dynamic impacts beyond those observed on average for the $t+1$ period are largely insignificant and inconsistent in sign. However, there is a clear trend in the point estimates and a distinct difference in arc of non-ELC versus ELC-focused dynamic tests. For non-ELC focused tests, results are negative and insignificant for $t+2$, but then become positive through $t+3$ and $t+4$, and even showing some significant at $t+4$, after which point estimates largely turn negative and insignificant. For ELC-focused tests, point estimates are positive at t and stay relatively stable with some notable significance between $t+1$ and $t+4$ and then decline and become negative and insignificant.

Overall, these further tests easily pass placebo tests for the most part and our primary results are confirmed. We also find some evidence for extended impacts up to 4 years after the rice price shock, though the evidence is not as statistically significant.

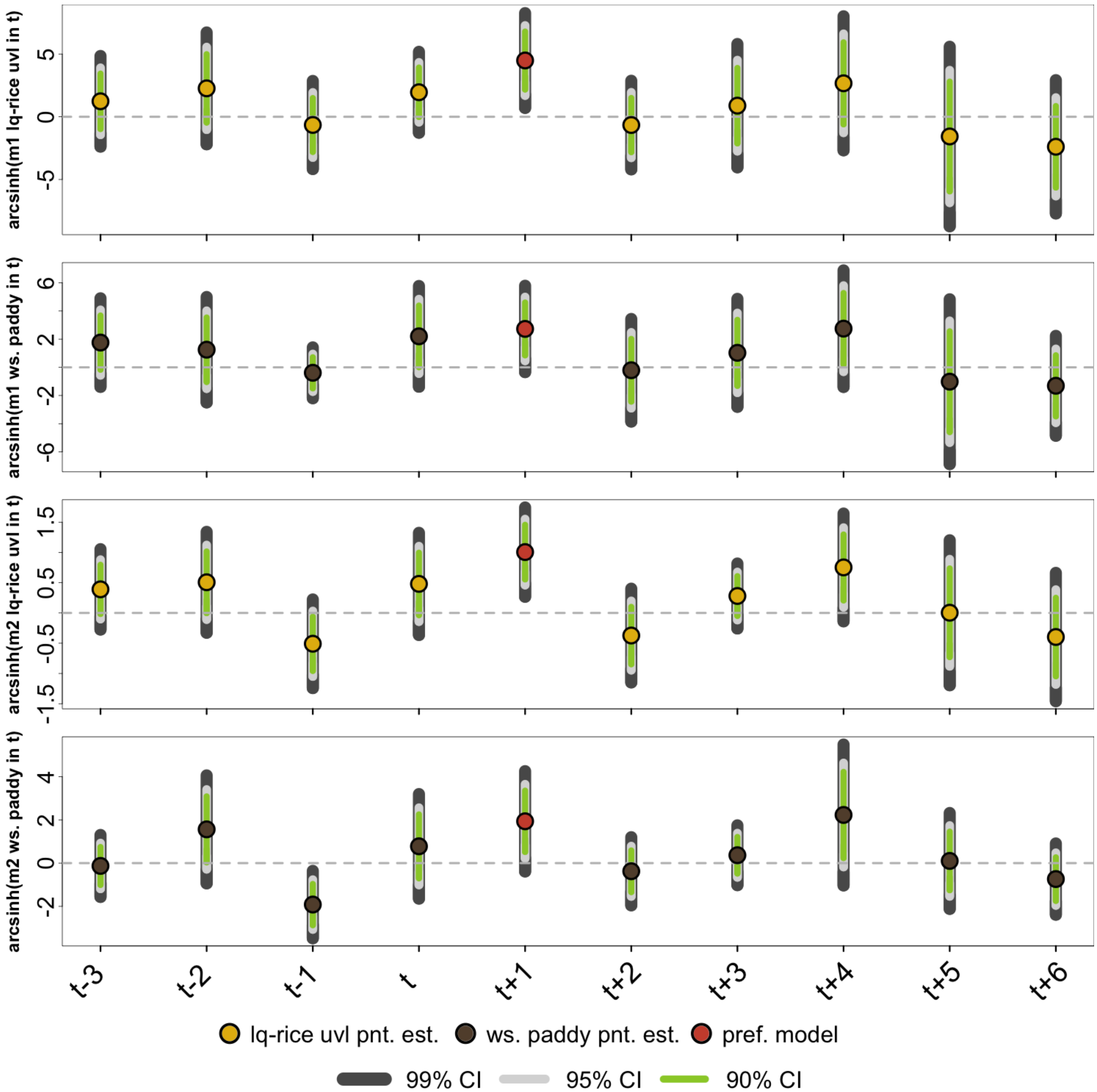

Figure 29: District-level placebo and dynamic effect tests on non-ELC-based deforestation. The y-axis represents point estimates and confidence intervals for second stage coefficients of our rice price variables of interest, where m1 refers to the mean of the respective rice price variable, and m2 refers to the variance, which we convert to standard deviation. The x-axis captures different leads and lags $t+j$ that arcsinh(non-ELC deforestation in hectares) is set to rice price variation in t. Our base model sets deforestation to $t+1$; deforestation in t is less likely to be responsive to rice prices in t, but it is plausible. Lags of $t-1$, $t-2$, and $t-3$ represent different placebo tests (i.e. it is not possible for future prices to affect past deforestation). Leads from $t+2$ to $t+6$ test the extent to which the effects of the rice price shock measured in our base model extend beyond $t+1$.

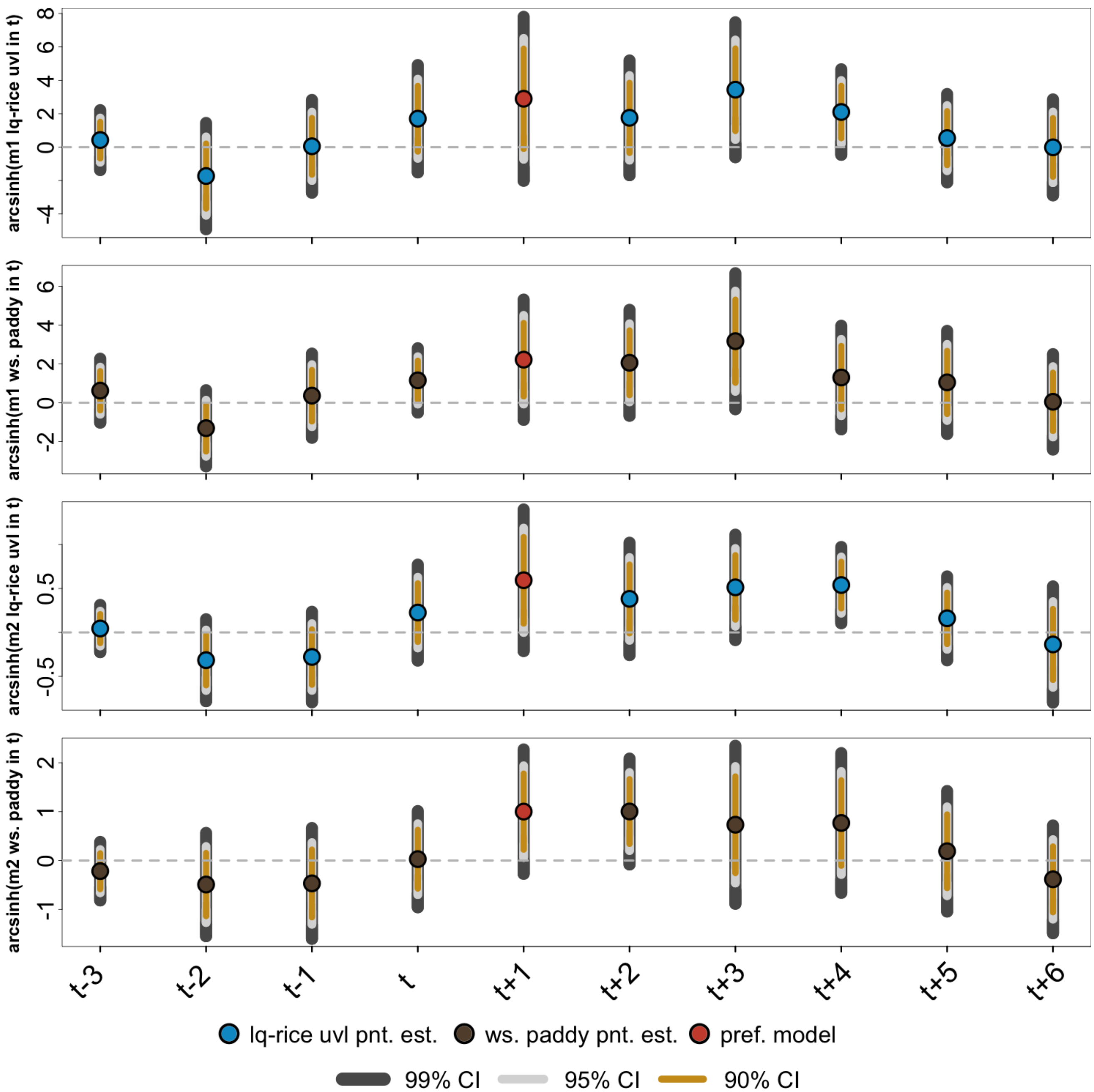

Figure 30: District-level placebo and dynamic effect tests on ELC-based deforestation. The y-axis represents point estimates and confidence intervals for second stage coefficients of our rice price variables of interest, where m1 refers to the mean of the respective rice price variable, and m2 refers to the variance, which we convert to standard deviation. The x-axis captures different leads and lags $t+j$ that arcsinh(ELC deforestation in hectares) is set to rice price variation in t. Our base model sets deforestation to $t+1$; deforestation in t is less likely to be responsive to rice prices in t, but it is plausible. Lags of $t-1$, $t-2$, and $t-3$ represent different placebo tests (i.e. it is not possible for future prices to affect past deforestation). Leads from $t+2$ to $t+6$ test the extent to which the effects of the rice price shock measured in our base model extend beyond $t+1$.

Supplementary Regression Results Household-level Specifications

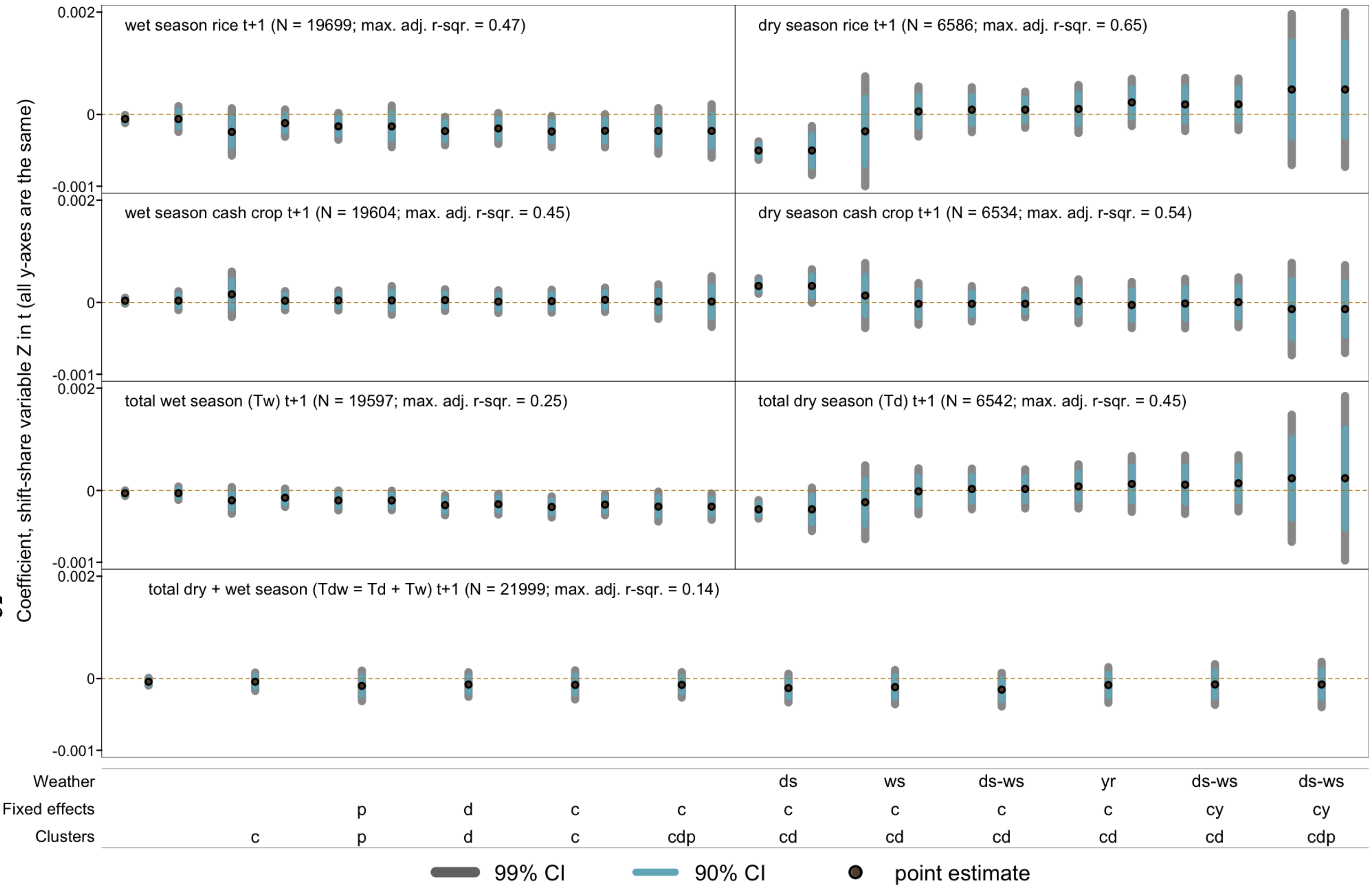

Figure 31: Point estimates of commune-level shift-share variable Z in t regressed on household-level area-shares in $t+1$. Wet and dry season allocations in rice, cash crops, and their sum (T_w and T_d) are the dependent variables in the top six panels; total rice and cash crops shares seasons, T_{dw}, is the dependent variable in the bottom panel. All models include linear trends, commune-specific trends, a parsimonious set of household controls (see footnote in the main text), and cumulative district-level deforestation (ELC and non-ELC) over t to $t-1$. Seasonal weather controls in t and fixed effects and clusters for standard errors are denoted in abbreviations in the bottom table: wet season (ws), dry season (ds), annual (yr), wet and dry season (ds-ws), commune (c), district (d), province (p), year (y).

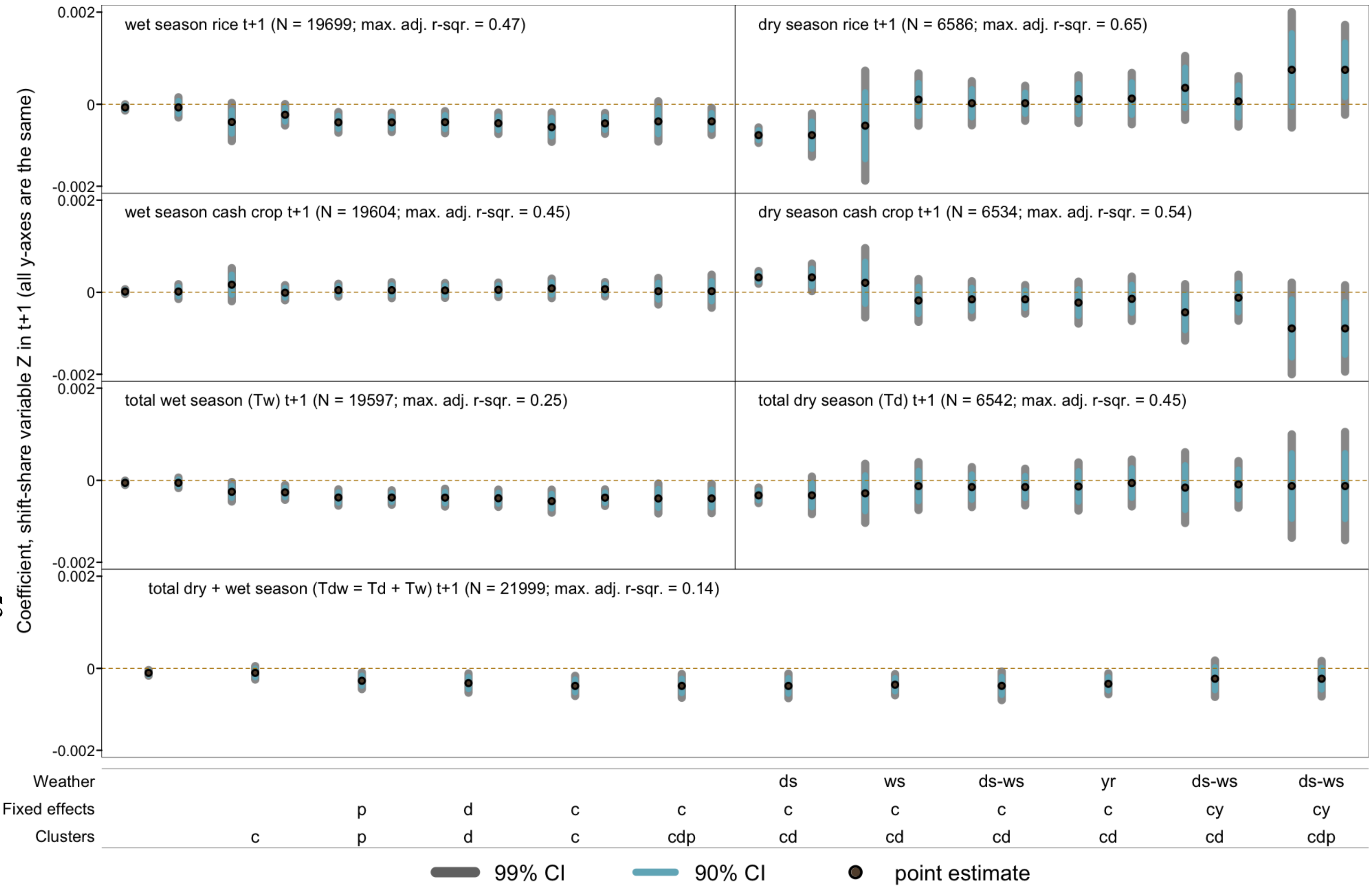

Figure 32: Point estimates of commune-level shift-share variable Z in $t+1$ regressed on household-level area-shares in $t+1$. Wet and dry season allocations in rice, cash crops, and their sum (T_w and T_d) are the dependent variables in the top six panels; total rice and cash crops shares seasons, T_{dw}, is the dependent variable in the bottom panel. All models include linear trends, commune-specific trends, a parsimonious set of household controls (see footnote in the main text), and cumulative district-level deforestation (ELC and non-ELC) over t to $t-1$. Seasonal weather controls in t and fixed effects and clusters for standard errors are denoted in abbreviations in the bottom table: wet season (ws), dry season (ds), annual (yr), wet and dry season (ds-ws), commune (c), district (d), province (p), year (y).

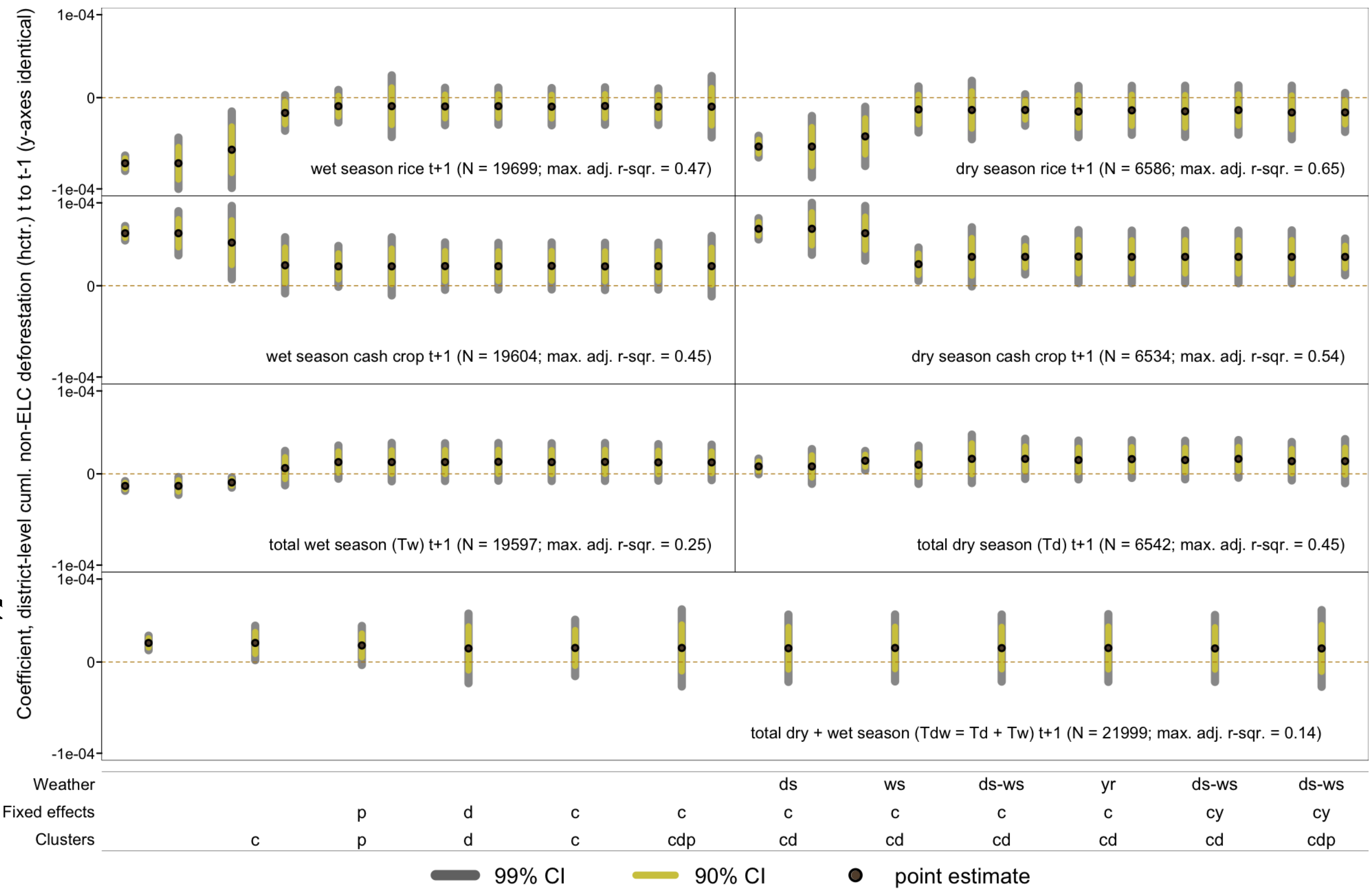

Figure 33: Point estimates of cumulative district-level non-ELC deforestation (hectares) in $t-1$ regressed on household-level area-shares in $t+1$. Wet and dry season allocations in rice, cash crops, and their sum (T_w and T_d) are the dependent variables in the top six panels; total rice and cash crops shares seasons, T_{dw}, is the dependent variable in the bottom panel. All models include linear trends, commune-specific trends, a small set of household controls (see footnote in the main text), cumulative district-level ELC deforestation over t to $t-1$, and shift-share variable Z in $t+1$. Seasonal weather controls in t and fixed effects and clusters for standard errors are denoted in abbreviations in the bottom table: wet season (ws), dry season (ds), annual (yr), wet and dry season (ds-ws), commune (c), district (d), province (p), year (y).

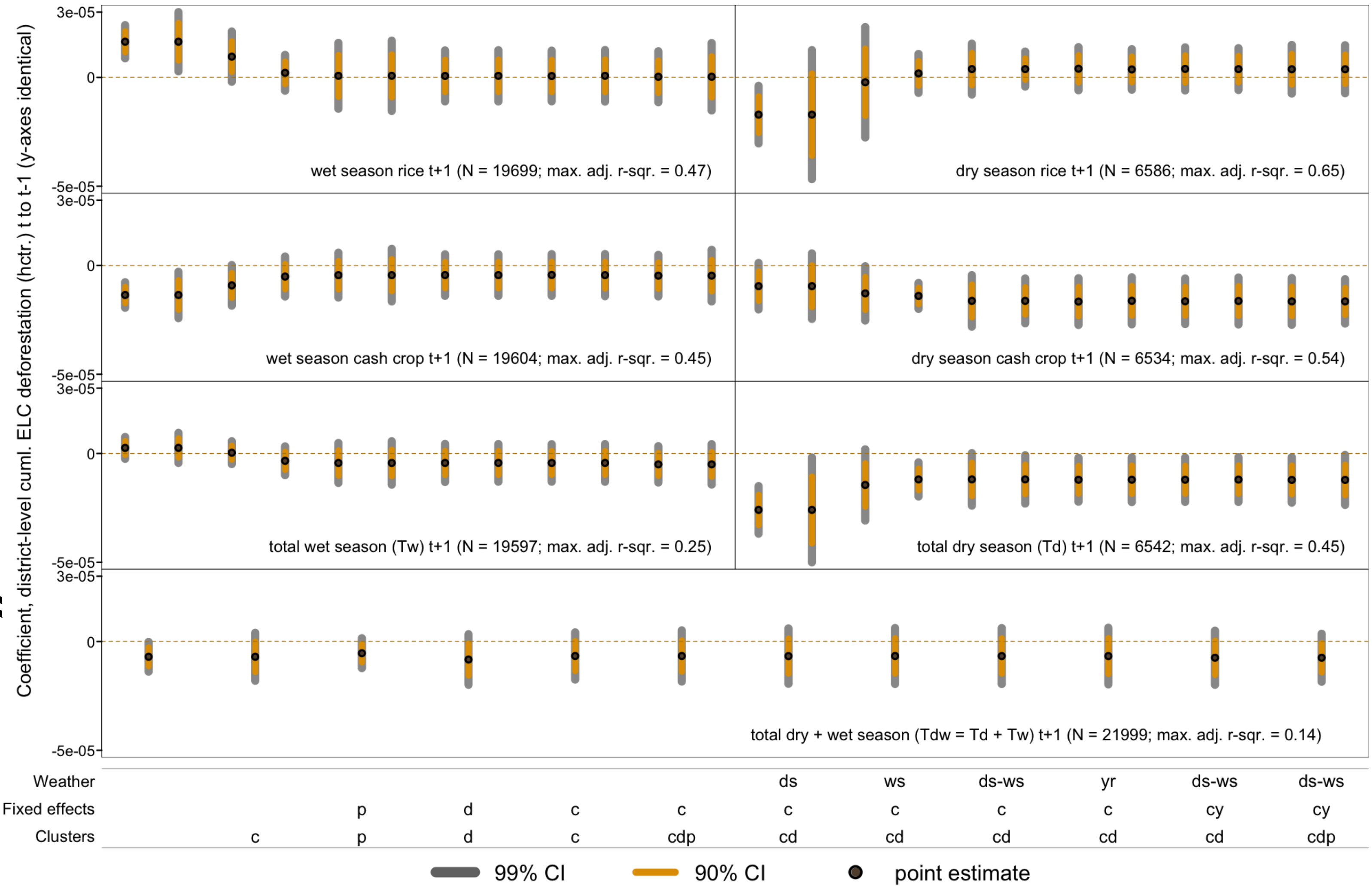

Figure 34: Point estimates of cumulative district-level ELC deforestation (hectares) over t to $t-1$ regressed on household-level area-shares in $t+1$. Wet and dry season allocations in rice, cash crops, and their sum (T_w and T_d) are the dependent variables in the top six panels; total rice and cash crops shares seasons, T_{dw}, is the dependent variable in the bottom panel. All models include linear trends, commune-specific trends, a small set of household controls (see footnote in the main text), cumulative district-level non-ELC deforestation in $t-1$, and shift-share variable Z in $t+1$. Seasonal weather controls in t and fixed effects and clusters for standard errors are denoted in the bottom table: wet season (ws), dry season (ds), annual (yr), wet and dry season (ds-ws), commune (c), district (d), province (p), year (y).

Figure 35: Sample household-level reduced-form regressions.

	ws-rice	ws-cc	T_w	ds-rice	ds-cc	T_d	T_{dw}
linear trend	0.003	-0.0005	0.002	0.002	0.001	0.003	0.0003
	(0.005)	(0.004)	(0.003)	(0.01)	(0.01)	(0.01)	(0.005)
commune trends	0.0000	-0.0000	0.00	0.0000	0.00	0.0000	0.00
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
post-2008 dummy	0.04*	0.005	0.05***	-0.05	0.04	-0.01	0.07***
	(0.02)	(0.01)	(0.02)	(0.04)	(0.04)	(0.04)	(0.02)
household (hh) size	-0.003***	0.0005	-0.003***	-0.003	-0.01***	-0.01***	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
male share hh. $>$ 15 yrs.	-0.02*	-0.01	-0.03***	0.06***	-0.01	0.05**	-0.02*
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
dummy female head-hh.	-0.01	0.002	-0.01	-0.03***	0.04***	0.01	-0.003
	(0.01)	(0.005)	(0.004)	(0.01)	(0.01)	(0.01)	(0.01)
yrs. educ. head-hh	-0.002**	0.001*	-0.001	-0.003**	0.003***	-0.0000	0.0000
	(0.001)	(0.001)	(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)
num. plots irgtd. rice	0.001	-0.02***	-0.02***	0.02**	-0.04***	-0.02***	-0.01***
	(0.004)	(0.003)	(0.003)	(0.01)	(0.01)	(0.005)	(0.002)
dis.lvl.no-ELC deforestation $t-1$	-0.0000*	0.0000*	0.0000	-0.0000	0.0000***	0.0000**	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
dis.lvl.ELC deforestation $t-1$	0.0000	-0.0000	0.0000	-0.0000	-0.0000***	-0.0000***	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Z_{t+1}	-0.0004***	0.0001	-0.0004***	0.0003	-0.0005**	-0.0002	-0.0003***
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0003)	(0.0001)
avg. com.level ws-tavg. (c) in t	-0.04	0.01	-0.03	0.08	-0.10**	-0.03	-0.004
	(0.03)	(0.02)	(0.02)	(0.05)	(0.04)	(0.06)	(0.03)
avg. com.level ws-rainfall (mm) in t	0.0000	-0.0001	-0.0001	0.0001	-0.0002	-0.0001	-0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)
avg. com.level ds-tavg. (c) in t	0.01	0.002	0.01	-0.05*	0.06**	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)	(0.02)
avg. com.level ds-rainfall (mm) in t	-0.0005	0.0003	-0.0001	-0.0004	-0.0004	-0.001*	-0.0001
	(0.0003)	(0.0003)	(0.0003)	(0.001)	(0.0004)	(0.0004)	(0.0003)
FEs?	C	C	C	C	C	C	C
Clust. SEs?	CD	CD	CD	CD	CD	CD	CD
Observations	19,699	19,604	19,597	6,586	6,534	6,542	21,999
Adjusted R^2	0.47	0.45	0.25	0.65	0.54	0.45	0.14

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

References

Andrews, I., J.H. Stock, and L. Sun. 2019. "Weak Instruments in Instrumental Variables Regression: Theory and Practice." *Annual Review of Economics* 11:727–753.

Angrist, J.D., and J.S. Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Cambodia Forestry Administration. 2010. "Cambodia Forestry Outlook Study." *FAO Working Paper Series*, pp. .

Chiang, A.C., and K. Wainwright. 2005. *Fundamental methods of mathematical economics*. McGraw-Hill.

Christian, P., and C. Barrett. 2021. "Spurious Regressions and Panel IV Estimation: Revisiting the Causes of Conflict." forthcoming, The Economic Journal.

Cragg, J.G., and S.G. Donald. 1993. "Testing identifiability and specification in instrumental variable models." *Econometric Theory* 9:222–240.

Deaton, A. 1989. "Rice prices and income distribution in Thailand: a non-parametric analysis." *The Economic Journal* 99:1–37.

FAO. 2020. "FAOSTATS, Crop Production Data, Area Harvested." http://www.fao.org/faostat/en//#data/QC, 2020-1-31.

Hansen, M.C., P.V. Potapov, R. Moore, M. Hancher, S. Turubanova, A. Tyukavina, D. Thau, S. Stehman, S. Goetz, T. Loveland, et al. 2013. "High-resolution global maps of 21st-century forest cover change." *Science* 342:850–853.

Jordà, Ò. 2005. "Estimation and inference of impulse responses by local projections." *American economic review* 95:161–182.

Nunn, N., and N. Qian. 2014. "US food aid and civil conflict." *American Economic Review* 104:1630–66.

ODC. 2019. "Economic Land Concession Shapefile." https://opendevelopmentcambodia.net/dataset/?id=economiclandconcessions, last accessed July 2019 under Creative Commons Attribution 4.0 (CC-BY-4.0).

Racine, J.S. 2019. *An Introduction to the Advanced Theory of Nonparametric Econometrics: A Replicable Approach Using R*. Cambridge University Press.

Ramalho, E.A., and J.J. Ramalho. 2017. "Moment-based estimation of nonlinear regression models with boundary outcomes and endogeneity, with applications to nonnegative and fractional responses." *Econometric Reviews* 36:397–420.

Ramalho, E.A., J.J. Ramalho, and L.M. Coelho. 2018. "Exponential regression of fractional-response fixed-effects models with an application to firm capital structure." *Journal of Econometric Methods* 7.

Ramalho, E.A., J.J. Ramalho, and J.M. Murteira. 2011. “Alternative estimating and testing empirical strategies for fractional regression models.” *Journal of Economic Surveys* 25:19–68.

Sanderson, E., and F. Windmeijer. 2016. “A weak instrument F-test in linear IV models with multiple endogenous variables.” *Journal of Econometrics* 190:212–221.

Simon, C.P., L. Blume, et al. 1994. *Mathematics for economists*, vol. 7. Norton New York.

Stock, J., and M. Yogo. 2005. *Testing for Weak Instruments in Linear IV Regression*, New York: Cambridge University Press. pp. 80–108.

Welch, J.R., J.R. Vincent, M. Auffhammer, P.F. Moya, A. Dobermann, and D. Dawe. 2010. “Rice yields in tropical/subtropical Asia exhibit large but opposing sensitivities to minimum and maximum temperatures.” *Proceedings of the National Academy of Sciences* 107:14562–14567.

Wooldridge, J.M. 2010. *Econometric analysis of cross section and panel data*. MIT press.

Young, A. 2019. “Consistency without inference: Instrumental variables in practical application.” Working paper, Working Paper.