Journal of Applied Econometrics

WILEY

Journal of APPLIED ECONOMETRICS

REPLICATION OPEN ACCESS

Are Princelings Truly Busted? Evaluating Transaction Discounts in China's Land Market

Julia Manso

University of Oxford, Department of Statistics and Nuffield College, Oxford, UK

Correspondence: Julia Manso (jumanso@stats.ox.ac.uk)

Received: 25 April 2025 | **Revised:** 3 February 2026 | **Accepted:** 20 February 2026

Keywords: Chinese real estate | data reproducibility | empirical validation | institutional reform

ABSTRACT

This paper replicates Chen and Kung's 2019 analysis (*The Quarterly Journal of Economics* 134(1): 185–226). Inspecting the data reveals that nearly one-third of transactions (388,903 out of 1,208,621) are perfect duplicates of other rows, excluding the transaction number. The analysis of the data sans duplicates replicates their statistically significant princeling effect, robust across various specifications. Further analysis reveals a disagreement between Chen and Kung's text and code: The paper's "logarithm of area" is actually area (m^2) divided by one million. This therefore necessitates a reinterpretation of the estimation results, revealing that the princeling effect is extremely large.

1 | Introduction

The recent implosion of several leading Chinese real estate developers, such as Evergrande and Country Garden, has kindled interest in the mechanisms of the Chinese real estate market. Unique not only for its highly leveraged financing practices, the market is also a nexus of interactions between state-owned and private enterprises and has historically been plagued by corruption and cronyism, particularly on the land sale side. One often-cited paper on the topic is Chen and Kung (2019), titled "Busting the 'Princelings': The Campaign against Corruption in China's Primary Land Market," which explores the unique mechanism of Chinese cronyism in land sale. Specifically, they examine how "princeling" firms—those firms tied to family members of China's top governing body, the Politburo—receive substantial price discounts on land parcels after controlling for several location-based and transaction-level variables (Chen and Kung 2019). They also have significant findings relating to princeling power, promotion likelihood, and magnitude of discount: More powerful princelings obtain larger discounts, and provincial party secretaries providing discounts were more likely to be promoted, with the likelihood increasing by the magnitude of the discount and the area of discounted land sold.

Between containing detailed information about officials' promotions and establishing firms' state-ownership and princeling statuses, the paper's replication data is also of popular interest and has over 12,200 downloads from its archive on the Harvard Dataverse.[1]

This paper replicates the results of (Chen and Kung (2019), hereafter "CK"). A previous investigation by Wiebe (2024) diagnosed issues with CK's variable on mayoral promotion. Yet, Wiebe found that the result was unchanged even when a corrected promotion variable was used. My replication identifies more extensive issues: An initial inspection of CK's data reveals that nearly one-third of the transactions (388,903 out of 1,208,621) are perfect duplicates of other rows, excluding the transaction number. These duplicates suggest, for instance, that the same buyer purchased two (or more) identical parcels of land with the same quality rating, area, and usage in the same city, month, and year, at the same price. Examining records of land sale transactions

suggests that some of these duplicates correctly refer to identical yet distinct parcels sold at the same time, but many other duplicates erroneously reference the same parcel of land multiple times. Importantly, over one-half of the princeling parcels in the full dataset (9815 out of 19,812) are duplicates. Replicating the analysis on the data, sans duplicates, yields similar results to CK's original findings: While the discount magnitude of princelings is slightly smaller than that found using the original dataset, the princeling effect is still large and statistically significant across the regression results.

Yet, further inspection reveals that when CK aggregate parcel areas to group parcels (for instance, summing transaction data to calculate the total area of land sold to each firm in each year), their text and code differ: The "logarithm of area" in CK's paper is actually area (m^2) divided by 1,000,000. I thus reinterpret CK's originally calculated coefficients as scaled area values rather than logged area values, with and without duplicates. On this scale, the magnitude of the princeling effect is quite large, suggesting that princelings on average purchase tens of thousands of additional square meters of land per year compared to their nonprinceling peers—areas double, triple, or even quadruple the median parcel size, depending on the specification. In the online appendix, I include calculations utilizing the logarithm of area purchased; these results are strange and somewhat unstable, with the princeling effect not only changing sign and direction between tables, but also dropping sharply prior to the anticorruption campaign.

2 | Apparent Duplicates in the Data

In their paper, CK construct a set of "princeling" firms from several online sources and match them to land transaction data obtained from the Ministry of Land and Resources. I focus on this latter data, extracted from the Land Transaction Monitoring System (http://www.landchina.com/). As each municipality's "bureau of land and resources is required to report each land transaction in their jurisdiction electronically on this website," per the Law of Land Management (p. 199), the dataset captures all land transactions.[2] The Land Transaction Monitoring System provides comprehensive information on land transactions, relaying everything from transaction details like total payment amount, sale method, buyer name, and total area of land sold, to building restrictions, like floor area ratio and greening ratio, to precise location details on the parcel.[3]

Examining CK's dataset reveals several cases of apparent duplicates: Specifically, after removing the transaction identifier, which was likely added by CK in data processing, 388,903 land transactions appear to be perfect duplicates, suggesting that a single buyer purchased identical parcels in the same month, year, and city for the same price and use. Table 1 shows the extent and distribution of these duplicates. Of the 1,208,621 land transactions in CK's data, 819,718 are unique, and 299,167 of these have at least one duplicate. While most transactions have only one duplicate, others have several—as many as 158. Importantly, of the 19,812 transactions that were conducted by princelings in the original data, over half of these are duplicates, with a significant share having three or more seemingly identical transactions. Duplicates are especially prevalent after 2012—the year that the anticorruption campaign began.

In light of these results, there are a couple of possibilities: First, that there are truly identical neighboring parcels that are sold at the same time and price to a single buyer for specific usages. Second, it is possible that the duplicate rows are erroneous and that the database itself has multiple land transactions that refer to the same underlying land parcel. Examining CK's replication data is not enough to confirm either possibility, as the data do not include each parcel's specific address. Yet, the transaction details can be matched with the original data from the Ministry of Land's database, and this additional information can be used to determine if multiple transaction listings refer to the same parcel. Specifically, after being identified in the database, each land transaction can be linked with a parcel number via the transaction's original sale announcement (the "Transfer Announcement"), and one can confirm how many parcels of each size are

TABLE 1 | Summary of duplicates by frequency of occurrence and type.

	Total observations			Observations by time period	
Frequency of occurrence	**All**	**State-owned**	**Princeling**	**2012 and earlier**	**After 2012**
1	520,551	181,571	4414	253,322	267,120
2	218,001	75,640	3479	191,718	26,283
3	79,366	27,473	1554	79,200	166
4	581	193	81	518	63
5–10	966	319	201	870	96
11–20	167	59	48	134	33
21–30	48	10	21	34	14
31–158	38	15	17	29	9
Number of unique rows	819,718	285,280	9815	525,825	293,784
Number of total rows in Chen and Kung	1,208,621	420,004	19,812	886,116	322,396
Number of duplicate rows	388,903	134,724	9997	360,291	28,612
Number of unique rows with duplicates	299,167	103,709	5401	272,503	26,664

up for sale in the area at the time. With this information, identical land transaction listings that refer to the same underlying parcel can be confirmed as erroneous duplicates.

Looking at the detailed transaction records indicates that both hypotheses are true: While identical parcels were sometimes sold to a single buyer, many other duplicates refer to one underlying parcel and are thus erroneous. Some of these are quite easy to recognize as duplicate listings of the same land transaction when referencing the Ministry of Land's database, as they may even have the same electronic lookup number—a "unique" identifier—for multiple transactions.[4] The Transfer Announcements also confirm there should only be one listing for the parcel in the Ministry of Land's database—and in CK's data.

Other cases are less clear but strongly indicative of incorrect duplicates. For instance, two transactions have different electronic lookup numbers, but the Transfer Announcement suggests that both listings refer to the same underlying parcel.[5] Cases with extremely high numbers of duplicates appear to largely be erroneous; for instance, for one entry with 71 apparent duplicates in CK's data, there is no record of even one transaction occurring with the given parcel features.

Additional cases appear to be partially correct, partially incorrect duplicates. For instance, the Ministry of Land's database reveals that two identical parcels were sold at the same time to a single buyer at the same price. Yet, CK's dataset has three transactions that match these features rather than two.[6] A sample of several such cases is included in the online appendix.

Further, some transactions appear identical except for the location dummy variables (`near500` and `near1500`) created by CK. These are dummy variables that reflect whether a given parcel is within 500 or 1500 meters, respectively, of a parcel purchased by a princeling firm. Presumably, CK used the specific location data (not included in their replication files) to determine these binary values, and thus, having different binaries for these values implies that the parcels are indeed distinct. These rows are therefore not classified as duplicates in this paper.[7]

Thus, a matched comparison with the Ministry of Land's database, tracing parcel features and Transfer Announcements, suggests a significant portion of the duplicate rows erroneously refer to the same underlying transaction. Yet, without parcel addresses, it is not possible to examine the exact extent of incorrect duplicates versus those that refer to different but identical parcels.

2.1 | Replicating Results Without Duplicates: Tables V and X

When replicating CK's regression results, I drop all duplicate rows but recognize that this path excludes duplicate entries that refer to different but identical parcels. However, as highlighted in Table 2, the regression results are largely still similar for their main specification (CK's Table V),

$$Price_{ickst} = \beta_0 + \beta_1 PrincelingFirms_{ikt} + \gamma X_i + T_{cst} + \nu_{ickst}. \quad (1)$$

As detailed by CK, the dependent variable $Price_{ickst}$ "is the log of price (yuan per square meter) for land parcel i sold by local government c to firm k for usage s in month-year t. The key explanatory variable, $PrincelingFirms_{ikt}$, is a dummy variable equal to 1 if a firm k is connected to a princeling" and 0 otherwise (p. 203). They control for several transaction-level variables, including land quality (as evaluated by the municipal government, on a scale of 1–20), state-ownership status, sale method, firm size, and log(area/100), where area is the land area sold in square meters. They also have city-year-usage, industry, and month fixed effects, with standard errors clustered at the province and firm levels and singletons included.

Column (1) illustrates regression (1) run on the full dataset, while columns (2) and (3) are conducted on the matched sample of parcels within a specified radius (1500 or 500 meters, respectively) from a princeling parcel. Columns (4), (5), and (6) repeat these regressions but now include princeling firms interacted with the dummy variable $PSCM$ (Politburo Standing Committee Member), aiming to explore whether firms connected to these committee members received larger discounts. Likewise, columns (7), (8), and (9) include princeling firms interacted with a dummy variable that reflects whether a transaction occurred after the connected official was retired. In all of these cases, the princeling effect is strongly negative with small standard errors—as was the case in CK's original results—suggesting that the princeling effect is robust despite the apparent duplicates. Interestingly, when duplicates are removed, the princeling effect itself appears slightly smaller; for instance, the price advantage for princeling-connected firms, based on results for the 500-meter-radius sample, is now a 55.3% $((1 - \exp(-0.804)) \times 100)$ discount compared to CK's original 57.0% $((1 - \exp(-0.844)) \times 100)$ discount. However, the interacted terms (princeling firms with Politburo Standing Committee connections and princeling firms connected to retired officials) are consistently larger than the original data's results.[8]

CK also uses this granular transaction data in their Table X, "Princeling Purchases and Land Prices after Xi Took Office," which examines whether President Xi's anticorruption campaign impacted the price discount given to princeling firms. I regenerate this table without duplicate rows (Table 3), and the results are again quite similar to CK's original findings: For the 500-meter matched sample, CK found that princeling firms had previously obtained an average price discount of 60.1% $((1 - \exp(-0.920)) \times 100)$ that dropped by 11.7% $((1 - \exp(-0.920) - [1 - \exp(-0.920 + 0.257)]) \times 100)$ after the introduction of President Xi's anticorruption campaign (column (2)); without duplicates, the 60.2% $((1 - \exp(-0.923)) \times 100)$ discount before the anticorruption campaign drops by 13.8% $((1 - \exp(-0.923) - [1 - \exp(-0.923 + 0.298)]) \times 100)$ after the campaign begins (Table 3, column (2)). The other columns are equally similar and again suggest the anticorruption campaign strongly impacts the magnitude of discounts obtained by princelings; these remaining columns explore changes in princeling advantage in the presence of interacted dummy variables involving different dimensions of the anticorruption campaign (e.g., whether a central inspection occurred in the province at the time of sale, whether the transaction occurred after the anticorruption campaign began, etc.). As with CK's Table V, these results are robust even in the presence of potentially erroneous duplicate rows.

TABLE 2 | CK's Table V.

	Log of land price								
	All	≤ 1500 meters	≤ 500 meters	All	≤ 1500 meters	≤ 500 meters	All	≤ 1500 meters	≤ 500 meters
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Original data									
Princeling firms	−0.808***	−0.904***	−0.844***	−0.545***	−0.666***	−0.620***	−0.808***	−0.894***	−0.835***
	(0.025)	(0.034)	(0.033)	(0.035)	(0.043)	(0.043)	(0.030)	(0.040)	(0.038)
Princeling firms*PSCM				−0.442***	−0.420***	−0.396***			
				(0.037)	(0.048)	(0.049)			
Princeling firms*Retired							−0.001	−0.051	−0.044
							(0.056)	(0.063)	(0.058)
Number of observations	1,144,507	334,232	191,896	1,144,507	334,232	191,896	1,144,507	334,232	191,896
Adjusted R^2	0.692	0.727	0.755	0.692	0.728	0.756	0.692	0.727	0.755
Data without duplicates									
Princeling firms	−0.750***	−0.849***	−0.804***	−0.485***	−0.602***	−0.574***	−0.737***	−0.830***	−0.789***
	(0.027)	(0.033)	(0.033)	(0.035)	(0.039)	(0.041)	(0.033)	(0.038)	(0.037)
Princeling firms*PSCM				−0.488***	−0.470***	−0.436***			
				(0.040)	(0.048)	(0.048)			
Princeling firms*Retired							−0.063	−0.090	−0.069
							(0.057)	(0.062)	(0.056)
Number of observations	779,372	248,627	136,480	779,372	248,627	136,480	779,372	248,627	136,480
Adjusted R^2	0.693	0.716	0.736	0.694	0.717	0.737	0.693	0.716	0.736

Note: Following CK, all columns include city-year-usage, month, and industry fixed effects, as well as the following control variables: Land quality (on a scale of 1–20), the logarithm of land area sold, firm size (on a scale of 0–3), state-ownership status, and dummy variables for sale method. Standard errors are robust to two-way clustering by firm and province and are in parentheses (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

3 | Misspecifying Area

For their later tables, CK filter and transform the land transaction records to group by relevant features. When they aggregate transaction records at the firm and year level—such that they list the total area of all of firm k's transactions in a given year—the value that they define as the natural logarithm of the area purchased in their paper is actually the area purchased (m^2) divided by 1,000,000, per their code. However, CK interpret these area/1,000,000 values as logarithms; I include specific examples in the online appendix. Thus, I re-interpret their coefficients for the tables impacted by this transformation error, for cases with and without duplicates. This error extends to CK's Tables VI, VIII, IX, XI, and XII as well as Figure VI.[9]

3.1 | Replicating CK's Table VI

CK's Table VI "Quantity of Land Purchased by Princeling Firms, 2004–2016" regresses the area of land purchased in the primary land market, divided by 1,000,000, on the dummy indicator signaling a princeling connection, replicated in Panel A of Table 4.[10] While CK originally find that princeling-connected firms purchased 0.2% ($(\exp(0.002)-1)\times 100$) more land each year than their nonprinceling counterparts, the regression coefficient actually suggests that princeling-connected firms purchased an additional 2000 (0.002×1,000,000) square meters of land annually compared to their nonprinceling peers, across all land transactions combined. Contextually, 90.8% of rows have a total annual area purchased of zero square meters, as few firms (only 14%) purchase land in multiple years between 2004 and 2016. Excluding firms that purchase zero area, the median area of land purchased annually by a firm is 26,668 square meters; the mean is 68,193 square meters, as a small number of firms have very high yearly transaction amounts (up to 14.67 million square meters). I thus find that the median more accurately reflects the average firm's yearly transaction volume. When the rows with zero area are included, the median is 0 square meters, and the mean is 6250 square meters. Excluding zeros, the average number of parcels purchased annually per firm is 1, although the maximum is 826 parcel purchases.

The 2,000 square meter advantage is therefore a 7.5% (2000/26,668) increase over the median parcel size (excluding rows with zero area), much more than CK's original 0.2%. In column (2), firms with PSCM connections purchase 32,000 square meters ($(0.001+0.031)\times 1{,}000{,}000$) of additional land per year compared to nonprincelings—more than doubling the median parcel size; this is far above the 3% that CK originally obtained (column (2), $(\exp(0.001+0.031)-1)\times 100$)). Column (3) indicates that retirement has a negligible effect on the quantity of land purchased.

TABLE 3 | CK's Table X.

	Log of land price									
	All	≤ 500 meters	All	≤ 500 meters	All	≤ 500 meters	All	≤ 500 meters	All	≤ 500 meters
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Original data										
Princeling firms	−0.907***	−0.920***	−0.825***	−0.858***	−0.870***	−0.896***	−0.907***	−0.920***	−0.818***	−0.847***
	(0.029)	(0.040)	(0.024)	(0.032)	(0.028)	(0.035)	(0.029)	(0.040)	(0.023)	(0.028)
Princeling firms	0.318***	0.257***					0.140*	0.093		
∗Transaction after 2012	(0.047)	(0.058)					(0.052)	(0.054)		
Princeling firms			0.819***	0.695***			0.504***	0.420***		
∗Central inspection			(0.124)	(0.139)			(0.079)	(0.096)		
Princeling firms					0.614***	0.572***	0.449***	0.447***		
∗Xi-appointed officials					(0.055)	(0.051)	(0.064)	(0.059)		
Princeling firms									0.109	0.037
∗Pre-2012 inspection									(0.074)	(0.070)
Number of observations	1,144,507	191,896	1,144,507	191,896	1,144,507	191,896	1,144,507	191,896	1,144,507	191,896
Adjusted R^2	0.692	0.755	0.692	0.755	0.692	0.755	0.692	0.756	0.692	0.755
Data without duplicates										
Princeling firms	−0.899***	−0.923***	−0.774***	−0.823***	−0.844***	−0.884***	−0.900***	−0.923***	−0.752***	−0.802***
	(0.033)	(0.041)	(0.025)	(0.031)	(0.034)	(0.037)	(0.033)	(0.041)	(0.025)	(0.028)
Princeling firms	0.351***	0.298***					0.151**	0.113*		
∗Transaction after 2012	(0.047)	(0.059)					(0.051)	(0.053)		
Princeling firms			0.766***	0.675***			0.488***	0.415***		
∗Central inspection			(0.115)	(0.135)			(0.081)	(0.099)		
Princeling firms					0.595***	0.571***	0.441***	0.445***		
∗Xi-appointed officials					(0.055)	(0.054)	(0.065)	(0.060)		
Princeling firms									0.028	−0.025
∗Pre-2012 inspection									(0.077)	(0.073)
Number of observations	779,372	136,480	779,372	136,480	779,372	136,480	779,372	136,480	779,372	136,480
Adjusted R^2	0.694	0.736	0.693	0.736	0.694	0.737	0.694	0.737	0.693	0.736

Note: Following CK, all columns include city-year-usage, month, and industry fixed effects. The controls are land quality, logarithm of land area sold, firm size, sale method, and state-ownership status and "its interaction with transactions after 2012, central inspection, Xi-appointed officials, and pre-2012 inspection" (p. 216). Standard errors (in parentheses) are robust to two-way clustering by firm and province (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

TABLE 4 | CK's Table VI.

	Area of land purchased (m^2)/1,000,000					
	Panel A: Original data			Panel B: Data without duplicates		
	(1)	(2)	(3)	(1)	(2)	(3)
Princeling firms	0.002***	0.001**	0.002***	0.002***	−0.000	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Princeling firms*PSCM		0.031***			0.041***	
		(0.004)			(0.004)	
Princeling firms*Retired			−0.001			0.058***
			(0.001)			(0.004)
Adjusted R^2	0.015	0.016	0.015	0.009	0.010	0.010

Note: Following CK, all columns include year fixed effects, as well as the control variables of firm size and state-ownership status. Standard errors are robust to clustering by firm and are in parentheses (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). There are 5,690,984 observations in each column.

TABLE 5 | CK's Table XI.

	Area of land purchased (m^2)/1,000,000							
	Panel A: Original data				Panel B: Data without duplicates			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Princeling firms	0.078***	0.073***	0.075***	0.078***	0.048***	0.050***	0.050***	0.048***
	(0.008)	(0.007)	(0.007)	(0.008)	(0.005)	(0.005)	(0.005)	(0.005)
Princeling firms	−0.022**			−0.015	0.006			0.013
∗Transactions after 2012	(0.006)			(0.007)	(0.005)			(0.007)
Central inspection		0.038***		0.025***		0.038***		0.025***
		(0.002)		(0.004)		(0.001)		(0.004)
Princeling firms		−0.053***		−0.025***		−0.032***		−0.024**
∗Central inspection		(0.007)		(0.007)		(0.006)		(0.007)
Xi-appointed officials			0.036***	0.028***			0.036***	0.029***
			(0.002)	(0.002)			(0.002)	(0.003)
Princeling firms			−0.056***	−0.036***			−0.033***	−0.035***
∗Xi-appointed officials			(0.006)	(0.006)			(0.005)	(0.006)
Adjusted R^2	0.032	0.047	0.052	0.057	0.021	0.038	0.044	0.050

Note: Following CK, all columns include province and year fixed effects, as well as the control variables of firm size and state-ownership status. Standard errors are robust to two-way clustering by firm and province and are in parentheses (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). All columns have 11,516,622 observations.

When Table VI is recalculated without duplicates (Table 4, Panel B), the result of column (1) does not change, but column (2) now suggests that princelings with PSCM connections purchase 41,000 ($0.041 \times 1{,}000{,}000$) additional square meters of land a year than their nonprinceling peers. Retired official connections (column (3)) are now associated with a significant increase in the area of land purchased by princelings, with princeling firms connected to retired officials purchasing, on average, an additional 59,000 ($(0.001 + 0.058) \times 1{,}000{,}000$) square meters of land relative to nonprincelings—bringing their total purchases to more than three times the median parcel size.[11]

3.2 | Tables VIII, IX, and XII

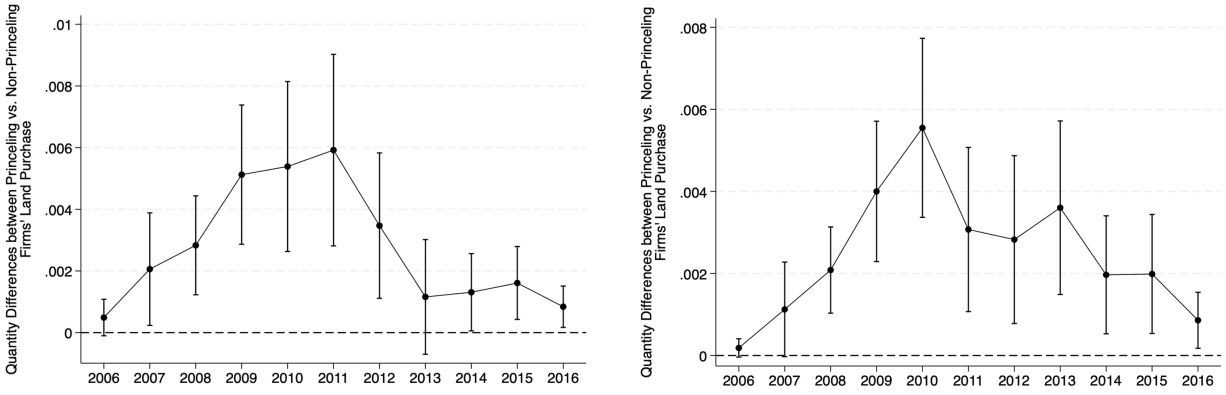
For Tables VIII, IX, and XII, I am unable to replicate CK's aggregated area measures. These specifications focus on whether there is an association between officials selling land to princelings for discounted prices and their likelihood of promotion. As such, they regress political turnover for provincial officials on different measures of princeling connections (Table VIII), political turnover for municipal officials (Table IX), and then political turnover for both types of officials (Table XII); they thus aggregate the total land sold to firms under each official's purview by year and province/municipality, respectively. CK likely has another variable that connects a transaction to an official at the province/municipality level, but they exclude this linking variable from the replication data, preventing recalculation.[12] Yet, in all cases, the logarithm of the area sold seems extremely small, suggesting that this "logarithm of area sold" is again area (m^2) /1,000,000.[13]

3.3 | Table XI and Figure VI

CK's Table XI investigates whether the quantity of land purchased by princeling firms changed after President Xi took office and began the anticorruption campaign.[14] Table XI thus regresses the land area purchased (m^2) /1,000,000 on whether a transaction occurred after 2012 (column (1)), whether a central inspection took place (column (2)), and whether the transaction occurs in a province wherein Xi replaced the party secretary (column (3)); these results are illustrated in Table 5. CK originally found that princeling firms purchased more land than their nonprinceling counterparts by approximately 8.1% ($(\exp(0.078)-1) \times 100$), but this advantage dropped by 2.3% ($(\exp(0.078)-\exp(0.078-0.022))\times 100$) after 2012. Reinterpreting these coefficients yields that princeling firms annually purchased 78,000 ($0.078\times 1{,}000{,}000$) more square meters of land than their nonprinceling counterparts, on average, before 2012, dropping to 56,000 ($(0.078-0.022)\times 1{,}000{,}000$) square meters afterwards. With the median total firm purchase in a year/province being 26,114 square meters, these purchase advantages reflect a sizeable princelings' benefit, suggesting that princelings purchased almost quadruple the median land purchase volume before 2012 and then only triple the median after 2012.[15]

In column (4), CK originally find that the binary variable reflecting whether a transaction occurs after 2012 becomes insignificant when the dummy variables for central inspection and Xi-appointed officials are included. They further note that "given that the sum of the coefficients in column (4) is not significantly different from zero, these two measures together effectively eliminate the advantage of the princeling firms in purchasing a larger quantity of land" (p. 219).[16] This result remains in the absence of the intended logarithm.

When duplicates are removed (Table 5, Panel B), the magnitude of the princeling effect drops across all specifications but remains statistically significant: Column (1), for instance, indicates that princeling firms purchased 48,000 ($0.048\times 1{,}000{,}000$) square meters more land than nonprincelings annually. After 2012, this effect does not change in a statistically significant way.



(a) CK's Figure VI (b) CK's Figure VI, no duplicates

FIGURE 1 | Comparison of CK's Figure VI with and without duplicates. (a) CK's Figure VI, (b) CK's Figure VI, no duplicates. *Note:* Bars reflect 95% confidence intervals.

For column (4), as above, "transactions after 2012" is insignificant when "central inspection" and "Xi-appointed officials" are included, the coefficient changing sign now that duplicates have been removed, and the sum of the princeling dummy variable and its interacted terms is not statistically different from zero. "Transactions after 2012" is thus no longer significant in any specification, suggesting that the effect of the anticorruption campaign may not be as clear-cut.

Finally, the misinterpreted logarithm of area sold is also carried into CK's Figure VI. Originally, CK find a consistent positive difference in the quantity of land purchased between princeling and nonprinceling firms, with the 95% confidence interval excluding 0 from 2007 to 2012, as is apparent in Figure 1a. The quantity difference then "shrunk noticeably" after 2012, with the effect being statistically insignificant in 2013 and only marginally significant in 2015 and 2016 (p. 219). Reinterpreting the y-axis yields a smaller range of differences between princeling and nonprinceling land purchases, with the differences ranging from slightly below 0 to nearly 8000 square meters when the outer bounds of the confidence interval are included.

In Figure VI without duplicates (illustrated in Figure 1b), the effect of the anticorruption campaign is less clear: The quantity difference in land purchased between princeling and nonprinceling firms drops in 2011 prior to the start of the anticorruption campaign, although the confidence intervals remain large. Further, the 95% confidence interval of the quantity difference between princelings and nonprincelings does not drop fully towards zero as it did in Figure 1a, with the quantity difference remaining statistically significant after 2007. The effect is ultimately much less clean, again limiting arguments about the anticorruption campaign's causal impact.

4 | Discussion

The presence of duplicates and the misinterpreted logarithm of area raise important implications about CK's original results: While the princeling effect often remained statistically significant but shrunk when duplicates were removed (as apparent in Table 5), the economic significance is not fully clear. Given the significant variation in firm parcel purchase volumes and sizes, a benefit of around 78,000 square meters (Table 5, column (1)) is extremely significant for a firm whose yearly purchase total per province is around the median of 26,114 square meters, but less so for one whose yearly purchase volume is over 4 million square meters (the year-level maximum).

Importantly, these results also raise questions about underlying data quality; with the panel data used in CK's Tables VI and XI having approximately 91% and 95% zeros, respectively, in the area column, effects are estimated from a small number of nonzero coefficients and are sensitive to duplicates. Given these consistency issues, taking the logarithm of the area would certainly afford more meaningful insight. Yet, when the logarithm of area is used (as described in the online appendix), the results are unstable, changing sign and direction across specifications.[17]

Further, the causal implications of the anticorruption campaign are limited when duplicates are removed, with Figure 1b highlighting that the difference in princeling and nonprinceling land purchase volumes dropped before 2012. Similarly, Table 5 confirms that the coefficient for a transaction occurring after 2012 changes sign and is no longer statistically significant—although the variables representing central inspections and Xi-appointed officials remain significant—perhaps suggesting that the anticorruption campaign depends more on regional enforcement and signaling. Given these results as well as the instability of the analysis when the logarithm is used (per the online appendix), further investigation should be done to determine whether the anticorruption campaign truly plays a causal role in reducing princeling discounts.

Acknowledgments

I thank Frank Windmeijer, Steve Bond, and David Steinsaltz for their advice and feedback on this paper.

Funding

Revisions to this manuscript were completed while the author was supported by Nuffield College, University of Oxford.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in the Journal Data Archive at https://doi.org/10.15456/jae.2026097.0234860238.

Endnotes

[1]The paper's replication data can be found at https://doi.org/10.7910/DVN/XW6OJT.

[2]China eliminated the secondary market for land use rights (LURs) in 2004, meaning the State is the sole seller of land, doing so at the municipal level (Li et al. 2019; Gyourko et al. 2022).

[3]CK note that the system also provides the address and postal code of the parcel, "a two-digit code of land usage . . . [and] a three-digit industry code of the buyer's firm" (p. 200). These fields are no longer included on the Land Transaction Monitoring System website, nor are they included in CK's replication data, which makes determining the extent of duplicates more complicated; it is possible that these information fields were removed from the database sometime after CK collected the data. Alternatively, CK may have obtained a slightly more comprehensive version of the website's data directly from the Ministry.

[4]The electronic lookup number is the database's main identifier/search key for unique land transactions. Having two "separate" transactions with the same electronic lookup number is thus impossible and indicates an error.

[5]In many cases, examining the full transaction details of the two listings reveals missing information in one listing. These omissions are non-essential details (e.g., greening ratio, floor area ratio) rather than key identifiers (e.g., buyer, area, quality). A second listing of the same parcel may thus exist simply to complete the initial record.

[6]This may reflect an erroneous duplicate entry being removed from the website after CK obtained their data.

[7]There are approximately 100,000 of these rows in the data. I was unable to verify whether these parcels are distinct using the Ministry of Land's database alone; the specific address information would be needed to ensure these transactions reflect separate parcels. Even if one wanted to remove these rows as duplicates, it is not apparent which observation contains the correct `near500` and `near1500` values without the parcel addresses. I thus leave these quasi-duplicates in, but note that their status as unique parcels should be verified further. I also conduct robustness checks that exclude these duplicates, as discussed in footnote 10.

[8]As a robustness check, I also replicate these main regression specifications using a stronger definition of duplicates—that is, when defining a unique row, I exclude all variables added in data processing, like `near500` and `near1500`. Under this definition, there are 727,127 unique rows. The regression results shrink slightly more towards zero but remain largely the same. Standard errors likewise remain similar.

[9]The previously discussed Tables V and X do not suffer from these issues because they do not contain aggregated measures of the logarithm of area, remaining at the parcel-level. The remaining tables (I, II, III, IV, and VII) offer summary information and are not impacted by the misinterpretation of CK's figures; only Figure VI is impacted.

[10]While CK state that industry fixed effects are included in Table VI, they are omitted from their code.

[11]Note that a relatively small number of firms (668) are connected to retired officials, and the data suggests that among purchases made by firms with connections to retired officials, smaller parcels are disproportionately represented among the duplicates, causing this large increase in the coefficient. Of the unique transactions involving retired officials, 46% have at least one duplicate, with some having as many as 22 duplicates. Indeed, the mean and median yearly parcel purchase areas are significantly higher for firms connected to retired officials, while those for nonprincelings and for princelings without this connection are roughly equal to each other.

[12]Wiebe (2024) was able to replicate part of CK's Table IX, specifically the coefficient on GDP growth in columns (7) and (8); Wiebe noted that he received additional data from one of the authors.

[13]For Table XII, CK utilize the logarithm of area as the independent variable for a set of regressions, calling it $Areaof\ Land\ Purchased_{it}$, yet as this variable is described as measuring the "quantity of land purchased," it is not clear whether they mean to interpret it as the logarithm of area or as area in square meters (p. 220). Regardless, as CK does not offer a detailed interpretation of these coefficients, the substantive implications do not change: The effect of $Areaof\ Land\ Purchased_{it}$ on the promotion of party secretaries drops for transactions occurring after 2012 for provincial and municipal secretaries and in regions with central inspections for provincial party secretaries.

[14]One of Xi's first initiatives as party leader was the anticorruption campaign: He announced at the 18th National Congress (the top meeting for setting agendas within the party) that he would work to catch the "tigers" (high-ranking corrupt officials) and "flies" (low-ranking corrupt officials) to ensure the party's survival (Wong 2012; Xi 2013; Yuen 2014).

[15]The panel data used in Table XI lists each firm's parcel purchases grouped by each province and year between 1991 and 2016. Given few firms make frequent land purchases, 95.4% of the observations have zero square meters of land purchased in a given province/year. The reported median includes only those nonzero transactions, as the median would be 0 otherwise. The mean is 3076 square meters when including rows with zero transactions per year/province, and 66,423 square meters when these zeros are excluded.

[16]Here, CK sum all coefficients including $Princeling Firms$—that is, the dummy variable and its interactions.

[17]In the online appendix, the logarithm is calculated as log(y+1), as this is the method that CK utilize in their paper. Given that log(y+1) assigns an arbitrary weight to the extensive margin effect, I also estimate these specifications using Poisson regression and still find unstable effect estimates. I further examine the extensive margin effect via logistic regression and a linear probability model.

References

Chen, T., and J. K.-s. Kung. 2019. "Busting the Princelings: The Campaign Against Corruption in China's Primary Land Market." *Quarterly Journal of Economics* 134, no. 1: 185–226. https://doi.org/10.1093/qje/qjy027.

Gyourko, J., Y. Shen, J. Wu, and R. Zhang. 2022. "Land Finance in China: Analysis and Review." *China Economic Review* 76: 101868. https://doi.org/10.1016/j.chieco.2022.101868.

Li, C., H. Bao, and G. Robinson. 2019. "Urban Land Marketization in China: A Supply Side Analysis." https://www.landecon.cam.ac.uk/sites/default/files/2023-05/rerc-wp-19-10-li.pdf.

Wiebe, M. 2024. "Replicating the Literature on Prefecture-Level Meritocratic Promotion in China." *Research & Politics* 11: 1. https://doi.org/10.1177/20531680241229875.

Wong, E. 2012. "New Communist Party Chief in China Denounces Corruption in Speech." *New York Times*. https://www.nytimes.com/2012/11/20/world/asia/new-communist-party-chief-in-china-denounces-corruption.html.

Xi, J. 2013. "Power Must Be 'Caged' by the System." *Qiushi*. http://en.qstheory.cn/2021-03/10/c_607648.htm.

Yuen, S. 2014. "Disciplining the Party: Xi Jinping's Anti-Corruption Campaign and Its Lim-Its." *China Perspectives* 2014, no. 3: 41–47. https://doi.org/10.4000/chinaperspectives.6542.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1**: Online Appendix.