The Internet Appendix of
"Corruption culture and corporate misconduct"

1. Surname matching

Insiders' country of ancestry is identified using their surnames similar to the methodology of Lauderdale and Kestenbaum (2000). While the use of names to classify populations into different ethnic groups has been around since the early 1900s (Rossiter, 1909), most recent efforts have been concentrated in the public health and population genetics literature (Mateos, 2007). Several recent studies (Kerr and Lincoln, 2010; Bengtsson and Hsu, 2015; Hegde and Tumlinson, 2013; Gompers, Mukharlyamov and Xuan, 2016) in the entrepreneurial finance literature also use surnames to identify the ethnic origin of inventors, venture capitalists, and entrepreneurial founders.

I use two main sources to identify the country of origin of surnames in a systematic way. First, I use U.S. Census records from 1850 to 1940. These records represent the complete set of Census records available to the public in which the respondents' names are disclosed since they are no longer subject to the 72-year confidentiality rule. For several of these datasets (1880, 1920, 1930, 1940), I acquired access to 100% of the records through the Minnesota Population Center. For the other years, only 1% of the records are currently available. To identify the country of origin of surnames, I restrict the dataset to first and second generation immigrants whose country of birth or father's country of birth is outside of the United States, which yields 54 million census records. I then link each unique surname from the Census records to its most frequently associated country of birth or father's country of birth. For instance, the surname "Wong" is linked to China because 97.2% of immigrants with the same surname are from China.

Second, I use the surname-ancestry country matching list from a commercial database. Origins Info Ltd., a well-known commercial vendor of name classification services, processed the list of surnames using its proprietary database constructed based on sources such as the American Dictionary of Family names and international telephone directories. The accuracy of Origins Info's matching has been validated in prior studies (Webber, 2007).

To create the final matching list, I do the following. First, I record matches where the most frequently associated country of birth from census records is the same country of origin identified by Origin Info. Second, I keep surnames for which the most frequently associated country of birth appears in more than 75% of the census records. Third, for surnames with different census and Origin Info country of origin, I hand-check their country of origin using sources such as ancestry.com, which provides a distribution of U.S. immigrants based on port entry records. Fourth, for the remaining unmatched surnames, I hand-check their country of origin using sources such as ancestry.com for 3,000 of the most common surnames. The procedure generates a list of over 1.5 million unique surnames and their associated country of origin.

I then merge the surname data with the list of officers and directors from Compact Disclosure from 1988 to 2006. Of the 1.87 million firm-year-insider observations, about 89% are matched to a country of origin.

2. First generation immigrants

One potential drawback of my corruption culture measure is that it may have low power since the impact of ancestry country culture tends to attenuate over time as the number of generations increases. While it is difficult to examine this issue directly without generational information on the insiders, there are several reasons to believe that this issue is unlikely to introduce significant biases that would alter the main results. First, the possibility that ancestry country culture attenuating over time creates a bias against finding culture to be significant. Second, many studies in the economics literature (e.g., Giuliano, 2007; Fernández and Fogli, 2009) use similar culture measures and document a significant impact of culture on individual behavior and economic outcomes.

To further investigate this issue empirically, I collect birth location information from *Marquis Who's Who* biographies to identify a sample of insiders who are first generation immigrants. I run the corporate misconduct regressions on this sample of foreign-born insiders. For this analysis, the corruption measure is the average Transparency International corruption index value from 1980 to 2009 in the insider's country of birth, where higher index values indicate more corruption. Unlike the main analysis, the corruption measure can only be constructed at the individual insider level rather than the firm level.

For option backdating and opportunistic insider trading, it is possible to link each event to a specific insider. Thus, these analyses are conducted at the firm-year-insider level. However, it is not possible to identify the specific individuals responsible for earnings management and accounting fraud. Thus, I run these regressions at the firm-year level using the sample of foreign-born CEOs and CFOs.

The results are presented in the Internet Appendix Table IA.1. Since the sample is small, I cannot include the full set of industry-year and county-year fixed effects. Instead, I control for time-varying industry, local, and market average misconduct rates, and time-varying industry, local, and market average corruption culture as in models (4) and (5), where these averages are calculated using the original sample.

Consistent with the main results, I find that insiders born in high corruption countries are more likely to commit corporate misconduct. In terms of economic significance, a one standard deviation (2.308) increase in the insider's corruption level is associated with an increase in the incidence of earnings management, accounting fraud, option backdating, and opportunistic insider trading of 6.1%, 6.4%, 10.4%, and 52.6%, respectively, compared to the means in the sample. These effects are larger than those from the main analysis, consistent with the idea that cultural influences are stronger for first-generation immigrants. I also rerun the main regressions excluding the foreign-born insiders and find results similar to the baseline case in Table 3. Overall, similar results are observed for both first and higher generation immigrants, providing additional support for the main culture measure.

References

Bengtsson, O., Hsu, D., 2015. Ethnic matching in the U.S. venture capital market. Journal of Business Venturing 30, 338–354.

Gompers, P., Mukharlyamov,V., Xuan, Y., 2016. The cost of friendship. Journal of Financial Economics 119, 626-644.

Hegde, D., Tumlinson, J., 2013. Can birds of a feather fly together? The payoffs of ethnic proximity in US venture capital. Unpublished working paper. New York University and University of Munich.

Kerr, W., Lincoln, W., 2010. The supply side of innovation: H-1B visa reforms and U.S. ethnic invention. Journal of Labor Economics 28, 473–508.

Mateos, P., 2007. A review of name-based ethnicity classification methods and their potential in population studies. Population Space and Place 13, 243–263.

Rossiter, W., 1909. A Century of Population Growth, from the First Census of the United States to the Twelfth, 1790–1900. U.S. Bureau of the Census.

Webber, R., 2007. Using names to segment customers by cultural, ethnic, or religious origin. Direct, Data and Digital Marketing Practice 8, 226–242.

Internet Appendix Table IA.1

Foreign-born insiders

Results from corporate misconduct regressions are reported, where the sample of foreign-born insiders is used. The sample in columns 1 and 2 consists of firm-year observations of publicly traded operating firms in the U.S from 1988 to 2006. The sample in column 3 consists of insider-grant date observations of publicly traded operating firms in the U.S from 1996 to 2006. The sample in columns 4 consists of insider-purchase date observations of publicly traded operating firms in the U.S from 1988 to 2006. Only foreign-born CEOs and CFOs are included in columns 1 and 2, whereas all foreign-born insiders are included in columns 3 and 4. The dependent variable in column 1 is earnings management, calculated as the absolute value of abnormal discretionary accruals scaled by total assets. The dependent variable in column 2 is a fraud dummy, which equals one (zero otherwise) if the firm-year is within a class action lawsuit period or has misstated earnings according to AAER or GAO. The dependent variable in column 3 is an insider backdating dummy, which equals one (zero otherwise) if the strike price of the insider's option grant is at the lowest price of the month. The dependent variable in column 4 is the price pattern ratio, computed as the ratio of the market-adjusted gross return over the 20 trading days following the insider purchase transaction to the market-adjusted gross return over the 20 trading days preceding the insider purchase transaction. Corruption is the average Transparency International corruption index value from 1980 to 2009 in the insider's country of birth, where higher index values indicate more corruption. The firm controls are defined in Appendix A. *t*-statistics or *z*-statistics (in parentheses) are computed using heteroskedasticity-consistent standard errors that are corrected for clustering at the firm level. All coefficients are multiplied by 100 for ease of exposition. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Earnings management	Fraud	Backdating	Price pattern
	(1)	(2)	(3)	(4)
Corruption	0.163*	0.079**	0.374**	3.611**
	(1.71)	(2.05)	(2.08)	(2.34)
Corruption culture$_{\text{county-year mean}}$	-0.143	0.164	0.238	5.138**
	(-0.89)	(1.59)	(0.73)	(1.98)
Corruption culture$_{\text{industry-year}}$	0.223	0.371**	0.239	7.030***
	(0.92)	(2.47)	(0.45)	(3.00)
Corruption culture$_{\text{market mean}}$	2.343	1.732	4.617	11.114***
	(1.04)	(1.41)	(0.92)	(2.71)
County-year mean	-0.018	0.024	-0.021	0.279
	(-0.95)	(1.40)	(-0.56)	(0.81)
Industry-year mean	0.309***	-0.057	0.072	-1.737*
	(4.62)	(-1.31)	(0.95)	(-1.87)
Market mean	0.263	0.300	0.245	1.659
	(1.54)	(1.60)	(1.36)	(1.17)
Ln(Assets)	-0.068	0.190*	-0.059	-0.092
	(-0.18)	(1.66)	(-0.19)	(-0.03)
Ln(1+Age)	-0.866	0.347	-2.105***	-12.637**
	(-1.25)	(1.15)	(-2.92)	(-2.28)
Market-to-book	0.302	-0.087	-0.970***	5.092*
	(0.84)	(-0.54)	(-2.90)	(1.85)
Leverage	-2.514	1.213	-3.827*	13.852
	(-0.86)	(1.52)	(-1.72)	(0.78)
Stock volatility	5.253*	2.470**	0.430	-22.815
	(1.83)	(2.54)	(0.15)	(-1.52)
ROA	-1.354	0.584	4.095	-180.367***
	(-0.32)	(0.33)	(0.85)	(-2.96)
Capital intensity	-2.194	-2.552**	0.827	54.429**
	(-1.36)	(-2.56)	(0.31)	(2.30)
R&D	-15.974	1.998	11.878	-108.029**
	(-1.26)	(0.68)	(1.62)	(-2.06)
High tech	0.315	-0.167	0.371	54.328***
	(0.43)	(-0.45)	(0.31)	(3.06)
Operating cycle	-0.041	-0.176	-	-
	(-0.14)	(-1.39)		
Loss percentage	2.719	-1.233*	-	-
	(1.45)	(-1.79)		
Sales growth	6.613	-0.061	-	-
	(1.15)	(-0.15)		
Sales volatility	0.523	-0.610	-	-
	(0.44)	(-1.18)		
Cash flow volatility	-1.533	0.807	-	-
	(-0.88)	(1.36)		

Ln(N. of options)	-	-	-0.506	-
			(-0.93)	
Shares traded	-	-	-	730.996
				(0.92)
Observations	2,881	2,741	4,188	2,757
N. of firms	479	487	629	273
R-squared	0.176	0.161	0.022	0.563
Sample	Foreign-born CEOs and CFOs		All foreign-born insiders	