



THE SIZE AND CHARACTER OF THE INFORMAL SECTOR AND ITS SHADOW ECONOMY IN MONGOLIA



In all countries some portion of economic activity remains unobserved by the government statistical agencies that measure Gross Domestic Product (GDP). This unobserved activity we call the *shadow economy* and it is defined in this report as the economic output that should be included in the national accounts, according to SNA-93 definitions, but that remains unrecorded (UN 1993). Previous research has identified two sources of the shadow economy (OECD 2002). The first, the *statistical shadow*, arises from issues related to data gathering and data compilation practices, and the second, the *economic shadow*, arises from the motivations of firms, households, and individuals who prefer to remain unobserved. The statistical shadow will arise, for example, when incomplete sampling frames are used for surveys or censuses so that some segments of the population are omitted from the data, or when procedural issues related to survey non-response are incorrectly handled, or more generally when limitations in capacity yield inadequate data collection or analysis.

The economic shadow is associated with the decisions of firms or households who misreport their activity in order to avoid observation. Those in the economic shadow may benefit materially, by evading tax or social insurance obligations, or by ignoring the requirements of costly regulations. Participation in the economic shadow will involve tradeoffs for these actors however, since there are also costs associated with shadow participation. These costs include potential penalties for non-compliance, as well as the possibility of reduced access to both public and private goods, such as pensions, police protection, and financing. Understanding how firms view these tradeoffs is fundamental to understanding the shadow economy's effect on the well-being of participants and its significance for the overall economy.

In addition to its economic and statistical sources, the shadow economy can be divided according to whether it is associated with formal or informal activity⁶. While the informal sector has been defined in numerous ways, often to accommodate local conditions, the distinction between the formal and informal need not lead to confusion in the measurement and study of the shadow economy. In this document we distinguish the sectors according to legal registry requirements so that informal sector shadow measurement can be made with precision⁷. As a result of this distinction quite different methodologies are required to study these sectors, and they will have very different characteristics. Since statistical tools are better developed in the formal sector, the formal sector shadow is likely to be associated primarily with economic causes and may be more responsive to policy changes. The informal sector presents the dual challenge of a less developed statistical infrastructure and economic motivations for remaining unobserved.

There is evidence that the informal sector shadow economy is particularly important in economies in transition from socialism to capitalism (Eilat and Zinnes 2002). The restructuring that occurs as government support of state run enterprises is eliminated typically displaces a large number of workers. In addition, institutions that support market functioning are often operating poorly at the outset of a transition. In this setting, the informal sector shadow economy can play several roles, perhaps most prominently as an employer of last resort, in which displaced workers can forestall poverty. The ease of entry in the informal sector shadow can also be a boon to workers with few assets, but who may be educated or otherwise skilled. Thus, despite the negative connotations of the economic shadow, its dynamism can lead it to play an important role harnessing

⁶ A third category of activity that remains mostly unobserved is the output of illegal goods (OECD 2002). In this report we do not address illegal outputs but will discuss illegal behaviors associated with noncompliance.

⁷ We recognize however that there are distinctions regarding types of enterprises within the informal sector that are often framed as degrees of formality. Examples include having paid labor versus only unpaid family labor, or differences in permanence of business location. In Mongolia an important distinction arises across enterprises with regard to their compliance with the Informal Sector Law. These degrees of formality are discussed further in Section 5 of this report.

entrepreneurial energy, buffering formal sector shocks, and serving as an incubator for small business.

Despite its potential benefits, particularly in the short-run, the effect of participating in the informal sector shadow can be a net negative in the long-run. Lack of access to capital markets and finance can distort investment and hiring decisions so that output is less capital intensive and horizons of shorter duration than would be optimal. These constraints can make it impossible for organizations to capitalize on economies of scale, leading to reductions in productive efficiency. Further, participation in the informal sector shadow may lead to changes in attitudes and social norms regarding the desirability of participating in the formal sector. As a result shadow participation in the informal sector may have enduring effects due to changes in the attitudes and capacities of participants as embodied in their physical, human, and social capital. Thus while formal firms may move into or out of the economic shadow relatively easily based on their perception of economic consequences, participation in the informal sector shadow may be more enduring, suggesting that policies to reduce participation in the informal sector shadow economy may be more difficult to implement.

The potential for enduring or *hysteretic* change is but one of several difficulties the informal sector shadow presents for policy-makers. Perhaps the most straightforward difficulty is the loss of tax revenue due to evasion, by both formal and informal participants, that limits the ability of government to pursue welfare enhancing policies and provide public goods. The low quality of public goods can result in a vicious cycle if firms enter the informal sector shadow when they do not receive fair value from their tax payments, thus further eroding tax receipts. The existence of the shadow also creates informational problems for decision makers. Since the true size and structure of the economy is not known, the allocation of resources across sectors and regions is likely to be inefficient. Further, at the level of the macro-economy, policy decisions regarding monetary and fiscal stimulus often depend on GDP and its growth rates. Inaccurate information on these variables implies that policies cannot be created with confidence and their effects are uncertain⁸.

The complex issues associated with the measurement of the size, the causes, and the economic and social consequences of the informal

sector shadow economy in Mongolia, and their policy implications are the subject of this report. The source of data on which conclusions are based was a collaborative project among the Open Society Forum of Mongolia (OSF), the National Statistical Office of Mongolia (NSO), the IRIS Center of the University of Maryland (IRIS) and the Economic Policy Reform and Competitiveness Project (EPRC) of the United States Agency for International Development (USAID). The main project activity was the implementation of a survey of 19,000 households in all regions of Mongolia, which took place in October and November 2004. Data on the financial status, and institutional environments of formal and informal enterprises was gathered as well as information on personal and household characteristics for these groups and for wage earners. Note that while data was collected on formal sector firms that appeared in the sample, the project was not designed to measure the size of the formal sector shadow economy.

The structure of the report is as follows. The present section introduces the scope and goals of the study, and the methodological approaches implemented to achieve them. Section 2 provides detailed discussion of the survey instrument and details of the methodological issues associated with the treatment design. Because the methodologies adopted in the final survey were novel in the Mongolian context, they required extensive pre-testing. Section 2.1 discusses the motivations for and the results of the two pretest exercises and an explanation of how the results affected the final methodological approach. Section 2.2 introduces the organization of the final survey instrument including details on the question types, question content, and on the use of variants of the survey instrument to create experimental treatments. Selecting a cost-effective yet representative sample for the informal sector survey presented some interesting challenges. Section 3 discusses how these issues were addressed, with sub-sections 3.1, 3.2, and 3.3 addressing respectively, sample size, sample selection, and the methodology of extrapolating of survey results to the broader population. Section 4 presents the core results on the measurement of the size of the informal sector shadow economy, and considers the results of the experimental treatments on informal sector shadow economy measurement. Section 5 analyzes the policy implications of the informal sector

⁸ Additional discussion of the policy relevant issues raised by the shadow economy is found in Eilat and Zinnes (2002).

survey. Sub-section 5.1 extends the measurement of shadow size in Section 4 by developing a picture of the shadow economy based on descriptive statistics of both formal and informal sector activity. Sub-sections 5.2 and 5.3 extend the descriptive analysis to examine the causes and consequences of informal sector shadow participation. Section 6 begins by summarizing the conclusions and policy implications of the study. It then identifies a number of weaknesses in the implementation of the instruments and analysis and then proposes low-cost refinements to overcome them. The section concludes by suggesting a number of follow-on activities that would help ensure the practicality of future informal sector surveys as well as extend this work to address other large parts of Mongolia's unmeasured shadow economy.

The primary activity of the collaborative project was the implementation of the informal sector household survey (ISHS). The ISHS was intended to increase the accuracy of the NSO's GDP measurement by providing a new baseline for current and future estimates of the informal sector's contribution to Mongolia's economic output. In this capacity the results of the ISHS replace those from survey conducted by USAID/EPSP in 1999 (Bikales *et al.* 2000). As a tool suitable for the measurement of GDP the ISHS addressed both the statistical and economic sources of the informal sector shadow economy. To address the statistical sources of the informal sector shadow economy the ISHS implemented a modification of the mixed household-enterprise survey in the Mongolian context. The mixed survey represents a research strategy that UN argues is "the most suitable approach" for collecting data on the informal sector as a whole (UN 2000; vol 1)⁹. The chief advantage of this approach is the relatively complete sampling frame derived from the listing of households, and as will be discussed in Section 2, this household frame was deemed preferable to a sampling frame based on the Informal Sector Law¹⁰. One difficulty with the mixed survey approach was identified as a result of previous work in Mongolia including the recently concluded Labor Force Survey (LFS). The LFS results indicated that only a small share of households would be of interest due to the large number that did not participate in the

informal sector activity. To address this issue the UN recommends a two-stage sampling process, identifying informal businesses in a brief initial interview and then returning to a sample of the business owners. Because this process would have proved prohibitively expensive and time consuming a single-stage modification was adopted that included the immediate dismissal of a large number of households without informal sector participation so that sample size requirements could be achieved at a reasonable cost. The implications of this dismissal protocol for the analysis are discussed in Section 3.3 and in Section 6.2 which discusses potential refinements.

The design of the ISHS also incorporated a number of techniques to address the economic sources of the informal sector shadow. Previous research in survey design has focused on how to reduce non-response, particularly to questions associated with income and other sensitive issues. Some of the findings suggest that quite simple measures can be effective in encouraging responses. Simple reminders about legal confidentiality requirements and the use of statements that encourage cooperation by reminding people of the importance of the data gathering processes have been shown to be effective (Moore *et al.* 2001). Additional methods aimed at insuring respondents that their anonymity was respected included the use of sealed envelopes so enumerator would not know the respondents answers, and the use of scrambled responses, that added statistical noise to individual's answers while allowing the aggregate results to remain interpretable. Notice that these two methods insure against two different respondent concerns. Using a sealed envelope insures the respondent against embarrassment that may arise from verbal admission to enumerators. The response is completely observable to the analyst however. The addition of statistical noise preserved anonymity more completely, since even the analyst was unaware of each individual's true response. Aggregate behaviors, however, remain interpretable. The procedures used to offer anonymity are discussed in more detail in Section 2.2 and their statistical properties are elaborated in Appendix 1.

⁹ Our adaptation of the household-enterprise survey design has the advantage of also yielding data on formal sector participants and being less costly than the design suggested by U.N. researchers. Some caveats regarding our approach are included in Section 6 of this report.

¹⁰ The Informal Sector Law applied to only a subset of activities in the informal sector shadow, and the pretest results indicated that even among these sectors there was significant non-compliance with registration requirements.

In addition to the measurement of the informal sector shadow's contribution to GDP at the national level, measures of size are developed within five regions, the West, Khangai, Central, East, and Ulaanbaatar, and within three locations, aimag center, soum center, and rural area. These measurement units provide comparability with other data collected by the NSO including the Household Income and Expenditure Survey (HHIES), and the LFS. Estimates of contributions to value added by sectors are also included, with the measurement at the level of aggregation of the single digit International Standard Industrial Classification (ISIC).

In addition to its measurement goals, the ISHS survey includes questions that identify the causes and social consequences of the informal sector shadow by measuring key aspects of the institutional setting in which the economic agents make their decision to participate in the formal or informal sector. Tax compliance, the prevalence of "pressure" from government agents and the frequency of bribery or

"gift-giving" is elicited, as are experiences and beliefs associated with governmental public goods such as the quality of the judiciary and the effect of regulatory regimes on profits. Experiences with private sector institutions such as financial intermediaries and business associations are also elicited. These questions allow us to understand the extent to which the sector retains its character as an employer of last resort, functioning primarily to alleviate poverty, or if other more dynamic entrepreneurial enterprises are forming, allowing the accumulation of wealth.¹¹ Finally, we analyze the consequences of the informal sector on attitudes regarding tax and regulatory "morality," to understand how informal institutions, such as norms of behavior related to bribery, tax evasion, smuggling and other activities interact with formal institutions. As a result of these inquiries we can draw conclusions on policy alternatives, and identifying those that may have positive, negative, or no effect on the well-being of Mongolian citizens. A summary of these results is found in Section 6.1.

¹¹ A related question that is explored is the extent to which the informal sector complements or substitutes for formal activity. As a complement to the formal sector, informal sector firms might serve as subcontractors, supplying inputs to production processes beyond their sophistication or capitalization. As a substitute to the formal sector we would expect to see direct competition with formal sector firms.

In this chapter we provide details on the content of the survey instrument, which was developed after two pretest exercises, conducted in April and June of 2004. The pretests were critical to the design of the final instrument since several of the survey techniques under consideration were experimental and the feasibility of using them in the larger survey needed to be assessed. In order to provide a clear understanding of how the final design was chosen, and to provide adequate information for those who may consider using these tools in future surveys, we begin our discussion of the survey design by discussing the pretests and the conclusions we drew from these exercises.

Step 1 of the pretest was a data collection exercise in Ulaanbaatar conducted on the weekend of April 13th and 14th of 2004, by a team of twenty NSO enumerators. A random sample of 240 households was selected and data was successfully collected from them all. There were several goals in the first step of the pretest. First, it was important to understand how important anonymity was to respondents. Since the measures to protect anonymity are costly it would not have made sense to use them if they did not reveal any additional sensitive information. Although the techniques have proven effective elsewhere for measurement of the informal sector (Sivasan 2003), it was necessary to test their effectiveness in the Mongolian context. A secondary question that, to our best knowledge, has not been studied elsewhere is the extent to which the need for anonymity varied across individuals with different personality characteristics associated with trust and risk. The investigation of social and personal attitudes is not only a methodological concern, however. We were also interested in learning, in the pretest and across

the course of the project, the extent to which measured characteristics are associated with shadow participation, household income, and other behaviors. The pretest also served the useful function of training a core team of experienced NSO enumerators in the new techniques. Many of this group would take on supervisory roles in the final survey. Their experience in Step 1 was instrumental in implementing modifications in Step 2 and in the final survey.

Step 1 of the pretest consisted of three treatments with each varying in an aspect of the respondent anonymity protocol. Respondents were randomly assigned to one of the three treatments. The main difference across treatments was the amount of statistical noise introduced into questions which required dichotomous (yes or no) responses to questions of varying degrees of sensitivity¹². The introduction of noise in two of the treatments is known as a *randomized response* protocol and was originally introduced by Warner (1965).¹³ Treatment 1 did not use the randomized response protocol but asked for the respondent to reply directly to the sensitive questions as is typical in surveys. The responses for Treatment 1 were sealed in an envelope, however, providing protection against embarrassment to the enumerator. Treatments 2 and 3 introduced protection for the respondent by including an additional non-sensitive or innocuous question and the respondent would answer either the sensitive or innocuous question based on the result of a die roll that was observable only to them. In Treatment 2 the respondent had a 75 percent chance of being asked to respond to the sensitive question. Treatment 3 provided additional protection for respondents since the probability of being asked to respond to sensitive question was reduced to 50 percent. As a result, in the randomized response treatments, neither the enumerator nor the survey analyst knew the response with precision.

¹² Section 2.2 below details the question protocol for the “two-question method” that was used in this pretest and in the final survey. Appendix 1 presents statistical properties of the randomized response protocol.

¹³ See also Chaudhuri and Mukherjee (1983), which contains a thorough discussion of a large number of randomized response techniques.

In addition to testing the randomized response protocol for dichotomous questions, the instrument for Step 1 included an experimental protocol for the elicitation of risk attitudes and questions on trusting behavior. The risk elicitation allowed respondents to choose among a number of different types of lotteries with cash payments, and the nature of their choices indicated the risk attitude¹⁴. The responses to these questions are of both substantive and methodological interest.

Step 1 of the pretest produced several important results that followed through to the final survey. First, with regard to the design of the “two-question” method, providing an intermediate level of protection in Treatment 2 generally yielded results very close to the less costly treatment 1 for the most sensitive questions. For example, with regard to the question “Did you register and pay the *patente* (informal sector tax) last month?”, 80 percent, 78 percent, and 44 percent responded “yes” in Treatments 1, 2, and 3 respectively. As a result, in Step 2 and in the final survey, treatments similar to Treatment 1, the least costly, and Treatment 3, the most informative, were used.

The pretest results also emphasized the fact that when designing the randomized response protocol the choice of the innocuous question has an important impact on the successful implementation of the method. The proportion answering “yes” to the innocuous question should not differ too greatly from the proportion answering “yes” to the sensitive question in order to give respondents security. Small differences between these two proportions are also necessary for the method to be computationally meaningful¹⁵. The design of the question pairings was aided by testing the questions on the team of

enumerators before the Step 1 pretest was fielded. A further finding, consistent with what has been discovered previously in the literature, was that the two-question method was not well suited to generating responses to numeric questions (Eichorn and Hayre 1983). Given the large number of numeric questions associated with the measurement of the informal sector shadow economy, the multiplier method, also discussed in Appendix 1, was tested in Step 2 and applied in the final survey.

With regard to the elicitation of attitudes towards risk and trust, we found that these characteristics were important both methodologically and substantively. With regard to substantive issues, we found that high trustors, defined as those scoring above the median score on three trust questions, were much less likely to have participated in bribing government officials. Risk aversion had a similar effect although the magnitude of the difference was much smaller. Methodologically, those exhibiting high levels of trust exhibited less variation in responses across the treatments.

Step 2 of the pretest extended the project in three important ways. First, the geographical scope was wider, with visits to *aimag* and *soum* centers and a small number of rural households along with additional enumeration in Ulaanbaatar. The addition of *soums* and outlying rural areas was deemed an essential part of the project in preliminary discussions with NSO principals, so that more valid measures of the size of the informal sector shadow economy would be obtained. To accommodate the geographical expansion the sample size for Step 2 was increased to 600. Table 2.1 presents the sample size for each location. Step 2 of the pretest survey was conducted from June 14th to June 18th 2004.

¹⁴ In both Steps 1 and 2 of the pretest the respondent’s compensation was tied to random outcomes associated with the risk game. This protocol was not used in the final survey since a very small payment to the respondent was possible, and the survey length was such that the possibility of such a small payment was inappropriate.

¹⁵ It is possible to generate a negative value from the computation if questions are poorly constructed. Appendix 1 discusses this in detail. Negative estimates of proportions occurred for one question in Step 1 but did not occur in Step 2 or in the final survey.

	Aimags center	Soum Center	Rural Area	Urban District
Dornogovi		Zamin Uud <i>n</i> =100		
Khovsgol	Murun <i>n</i> =143	Khatgal <i>n</i> =60	Alagardene <i>n</i> =17	
Selenge		Mandal <i>n</i> =120		
Ulaanbaatar				Bayanzurkh <i>n</i> =160

The second extension in Step 2 was to investigate alternative approaches to collecting information on value added of the self-employed businesses. Two types of income and balance sheets were developed, a short form which captured information relevant for value added in only five questions, and a long form which included extensive subcategories for different items contributing to value added. For example, with regard to expenses, the short form contained one question and the long form ten questions for distinct categories of expenses. It was not clear in advance if the complexity of the businesses justified such an extensive enumeration. Three treatments were implemented to determine the proper approach. Treatment 1 implemented the long form of the income and balance sheet with direct questions. Treatment 2 implemented the short form with direct questions and treatment 3 the short form with the randomized response protocols for both sensitive questions and numeric responses. Both Treatments 1 and 2, with direct questions, required that responses be sealed in envelopes. As a result of enumerator feedback from Step 1, refinements were made to two-question instructions, the risk elicitation and the business description section of the survey. Due to the importance of social attitudes in understanding responses, this section of the survey was expanded for Step 2¹⁶.

The critical result from Step 2 was associated with the design of the income and balance sheet section. The results showed that the short form would

lead to substantially higher estimates of the size of the informal sector shadow economy, overstating the size by 50 percent if treatment 1 is taken as a baseline. The higher estimates of shadow size resulted from dramatically lower expenses in the short form. It appears that the listing of expenses served to remind people of costs that would otherwise have been ignored¹⁷.

Step 2 of the pretest also led to a major restructuring of the survey format to insure that enumerators identified the proper respondents and their status in the formal or informal sector of the economy immediately. This insured that the personal characteristics elicited were for the relevant business owner as often as possible. This change also allowed for the introduction of a cost-saving protocol in the ISHS in which the interviews of a substantial portion of formal sector wage earners, who were a low priority for this survey, were ended immediately. The details of the final survey design used in the ISHS are the subject of the next section of the report.

As a final note, in spite of their small sample sizes the pretests provided evidence to suggest that income in Mongolia is not normally distributed. This is consistent with what is found in other countries and across time.¹⁸ This means that standard methods of hypothesis testing and sampling generally need to be modified. This carries important implications for the estimates developed in Section 4 and is discussed therein; we also return to this issue in Section 6.2.

¹⁶ All survey instruments are included in Volume II of the report.

¹⁷ The main effect of the long form was on expenses because firms had many categories of expenses but only a small number of sources of income.

¹⁸ While there is some disagreement as to the appropriate distribution, principal candidates are the Pareto and Lognormal distributions. See Cramer (1971; p. 46) for details.

The pretest work led to the development of a survey instrument for the ISHS that contained thirteen sections that can be segregated into five areas that include 1) identifying whether the household participated in the informal sector, 2) collecting basic demographic data as well as attitudes related to personal and social preferences such as trust, trustworthiness, and risk, 3) describing the business activity which, for both formal and informal sector owners, included a detailed income and balance sheet, 4) identifying attitudes, behaviors, and perceptions associated with the institutional setting in which business was conducted, and 5) generating feedback on the survey process. A more detailed look at the thirteen survey sections is presented in Table 2.2 which identifies the purpose of each section and clarifies who responds and what question types are used. Note that the question types will vary across treatments so that not all respondents will see all the question types. For example sections 8 and 10 include one or both of the randomized response questions identified here as *Q* for the two-question method and *M* for the multiplier method. Some respondents, however, received either direct questions (*D*) or direct questions with sealed envelope responses (*E*) in these sections. Note that methods *Q* and *M* also made use of the sealed envelope response method, except for a small test sample (Treatment 4) which responded to the *M* questions verbally. A detailed discussion of the treatment types, below, is associated with Table 2.3.

The different question types were designed to provide different levels of protection to the respondents. In developing the survey, we hypothesized that there are two primary reasons a respondent may want to preserve anonymity. In a face to face interaction a respondent may be reluctant to admit non-compliant behavior to the enumerator. Alternatively, the respondent may be primarily concerned about exposure of information to government officials at higher levels where enforcement may be a concern. These concerns may, of course, vary over question content, and we develop treatments that incorporate protection from one, both, or none of these concerns in order to gather information needed to further improve this

methodology. Details of the question types follow.

Direct (*D*) questions are standard survey questions which serve as baseline for our analysis of treatment effects. Respondents give a verbal reply and both the enumerator and the analyst know the actual response.

Envelope (*E*) questions replies are written by the respondent and placed in an envelope and so are never observed by the enumerator. These responses are known by the analyst and so can be used in an analysis just as direct survey questions are used.

Multiplier (*M*) questions use a randomizing device to introduce noise into the response and are used on questions eliciting a numerical response. The multiplier question was applied to the income and balance sheet questions. The technique asks respondents to multiply their true value by a number that results from the roll of a die, with the roll of the die hidden from the enumerator. With the multiplier question neither the enumerator nor the analyst knows the true response with certainty, thus preserving privacy. Population averages of the response are derived based on the statistical properties of the multiplier. There are two multiplier formats. For M^D the respondent reports the multiplied response verbally to the enumerator. For M^E the multiplier result is sealed in an envelope. Appendix 1 contains additional details on this protocol.

The Two-Question (*Q*) question style is used for sensitive dichotomous “yes/no” questions in the institutional setting section. With this approach, respondents randomly answer either the sensitive question or an innocuous question such as “Can you play a musical instrument?” We implement the *Q* questions with a 50 percent probability of the respondent being asked the sensitive question. As in the multiplier method, the die roll that determines the relevant question is concealed from the enumerator. After the die roll the answers are recorded and sealed in an envelope at the completion of the survey section. Appendix 1 contains additional details on this protocol.

Section	Purpose	Who Responds	Question Type
1. Introduction & Cover Sheet	Introduce enumerator and survey purpose.	All	D
2. Sorting	Determine self-employment status, sort to question type or dismiss from survey.	All	D
3. Socioeconomic Background	Analysis of behavioral responses to economic policy.	All remaining after dismissal	D
4. Social Attitudes	Analysis of behavioral responses to economic policy.	All remaining after dismissal	D, E
5. Risk Attitudes	Analysis of behavioral responses to economic policy	All remaining after dismissal	D
6. Innocuous Questions	Determine the sample replies for the “Two Question” method.	Treatments 1, 2, & 4 (Those not receiving two question method in section 10.)	D
7. Activity Description	Determine nature of business activity.	All employed. (Self-employed receive greater detail.)	D
8. Income and Balance Sheet	Financial data on primary self-employment activity.	Self-employed.	D, E, M
8a. Sector specific Questions	Additional detail on gold mining, tourism, and cashmere industry.	Those in the specified sectors	D, E, M
9. Sensitive Questions (Low)	Questions on business registry, tax, and policy issues	Self-employed.	D, E
10. Sensitive Questions (High)	Questions on bribery, smuggling, tax compliance, household income	All employed. (Self-employed receive greater detail.)	D, E, Q, M
11. Respondent Feedback	Comprehension and evaluation of instrument by respondent.	All remaining after dismissal	D, E
12. Respondent Earnings	Payment of participant and departure.	All remaining after dismissal	D
13. Enumerator Feedback	Evaluation of participant and completion of record-keeping.	All remaining after dismissal	D

D = Direct, E = Envelope, M = Multiplier, Q = Two Question. See Section 8, below, for detailed explanation.

We combined the question types into five treatments as shown in Table 2.3 below¹⁹. The treatments are identified by a two-letter description indicating the type of question used for the income and balance sheet and the type of question used for the sensitive questions on behavior. Treatment 1 (DD) serves as a no-privacy baseline with all responses made directly and verbally to the enumerator. Treatment 2 (EE) provides protection of privacy against the enumerator since the respondent places responses to sensitive questions directly in an envelope. Treatment 3 (EQ) provides additional privacy for the dichotomous sensitive questions. Treatment 4 (M^DE) incorporates the direct response version of the multiplier questions. Treatment 5 (M^EE) provides the maximum amount of privacy for respondents with all sensitive questions protected by a die roll and the answers sealed from enumerator view.

While our primary interest in developing the survey treatments is to understand the effectiveness of the different methods of preserving respondent anonymity, survey treatments are also appropriate in this study because of the flexibility in analyses that they permit. Because the randomized response yields aggregate values without linking responses directly to individuals it often outperforms direct questions for measurement of population values. Randomized response approaches are less useful however for some policy analyses, such as those in which individual behavior must be linked to the institutional environment. Thus the different treatments offer the opportunity to take advantage of different question types for different types of analyses, thereby balancing the constraints of precision, privacy protection, and sample size.

Treatment	Numeric	Dichotomous	Sample Size
1	Direct	Direct	300
2	Envelope	Envelope	2301
3	Envelope	Two-question	2001
4	Multiplier (M ^D)	Envelope	300
5	Multiplier (M ^E)	Two-question	8998

Combining the data from different treatments also allows us to generate high quality measures of informal sector shadow activity while retaining the flexibility to conduct policy analyses which require us

to link responses to sensitive questions to respondent characteristics. This type of policy analysis would not be possible if only the multiplier and two-question types were used.

¹⁹ The sample size by treatment in Table 2.3 is the figure remaining after the dismissals of 5,111 respondents (wage earners and unemployed) who did not contribute to measurement of the informal sector shadow economy. Some wage earners and unemployed were retained, however, in order to make comparative analyses. The dismissal protocol and its implication for sampling weights is discussed in Section 3.

In this section we provide details of the procedures through which the sample for the Informal Sector Household Survey (ISHS) was selected. The goal of the sampling plan is to ensure a sufficient number and distribution of household observations so as to be able to provide estimates of value added (i) for the five regions in Mongolia, (ii) at the national level by rural and urban locations, and (iii) for a small number of specific sectors.

To generate the size and distribution of the sample, we make use of income data from the Household Income and Expenditure Survey (HHIES) published in the *Monthly Bulletin of Statistics*, self-employment shares at the regional level from Labor Force Survey (LFS) of 2002-2003, the current list of *patente* holders at the *soum* level made available by the Ministry of Finance, the household sampling frame from 2002, and data from Step 2 of the pretest of the ISHS in June 2004.

One of the key issues addressed in the design of the ISHS was that a large number of households in any sample chosen from the household list are unlikely to be active in the small-scale shadow economy. We address this issue through the survey protocol by ensuring the rapid dismissal of most non-shadow-economy participants. The sampling plan also addresses this issue by generating selection probabilities that depend on estimates of the distribution of shadow activity based on the data in the Ministry of Finance's *patente* registry. A simple average of the number of individuals registered under the Informal Sector Law (*patente* holders) and the number of households is used as the relevant measure of size at the *aimag* and *soum* level. As discussed below, the dismissal protocol was extended so that the sample size could be expanded at minimal cost, and an adequate number of informal sector businesses be contacted.

The sampling plan was intended to assure an *ex ante* precision of plus or minus 5 percent for the non-

herding informal sector shadow size.²⁰ However, the appropriate sample size to achieve survey results with this level of precision depends on the variance of the underlying value added of the informal sector firms. The HHIES data, as published in the *Monthly Statistical Bulletin*, contain only the mean household incomes by location, not the variance. Unfortunately, variances of the HHIES income measure were not available. Without this information, we had to rely on our own estimates of variances based on data from *other* countries and on rules of thumb. This introduced considerable uncertainty regarding just how precise the results of the survey would be. To the extent our variance estimates were smaller than the true values then more imprecise results will obtain. If our estimates were larger we will achieve additional precision, although the survey would arguably be more costly than necessary.

Data was collected in the five regions of West, Khangai, Central, East, and Ulaanbaatar and in four locations: urban *aimag* center, urban Ulaanbaatar, *soum* center, and rural. Note that due to the way the NSO records information, Ulaanbaatar is both a "region" and a "location." A map of these areas is provided in Appendix 2.

Both the HHIES results and ISHS pretest data were used to independently develop estimates of sample size. The two approaches were in fairly close agreement for *soum* centers and *aimag* centers but diverged for rural areas and Ulaanbaatar. Due to the small sample size in the ISHS pretest—particularly in rural areas—we chose to adopt the results calculated from the HHIES data. Since the HHIES data is reported at the location level only, however, it was necessary to supplement this information to reflect regional variation. Regional variation in self-employment activity was captured with the LFS data. The use of this supplementary data resulted in the

²⁰ As discussed elsewhere in this report, the normality assumption underlying this estimate was not satisfied in the survey data. We return to this concern in Section 6.2

sample being distributed unequally across the regions. The distribution of the sample across the four locations is also derived from the LFS data and checked against the data from the ISHS pretest (on the contact rates with households with self-employment).

The LFS estimates of self-employment activity have a major impact on the structure of the sample due to the large amount of variability in self-employment activity that they reveal. A relatively large sample is needed in Ulaanbaatar in part because only 22 percent of employment is self-employment in the capital city. In contrast, the LFS reports self-employment between 55 percent and 78 percent of total employment in the other regions. The large variation in regional rates arises primarily from the differences in the urban-rural mix of the regions. Self-employment rates in rural areas are close to 95 percent according to LFS figures, while self-employment rates in the *aimag* centers are similar to, though somewhat greater than, those in Ulaanbaatar. We also assume that variance of income is smaller in the rural areas since there is more homogeneity of business type, primarily that associated with animal herding. Both the self-employment figures and the expected differences in variance of incomes lead the sample to be heavily weighted to *aimag* centers and Ulaanbaatar.

Our formula for determining the sample size needed in a particular region and location is given by

$$n_{ij} = z_{ij} * I * \frac{1}{c_{ij}}$$

with $i=1, \dots, 5$ and $j=1, \dots, 4$ representing regions and locations, respectively. The z_{ij} represent the raw sample size calculation based on our estimated variances of income and includes a finite-population correction. I is an inflation factor which addresses the noise incorporated by the random-response multiplier method and is equal to 1.3.²¹ The c_{ij} are the rates of contact with self-employed households. We use c_{ij} between 0.32 and 0.37 in urban areas, between 0.6 and 0.7 in the *soum* centers, and 0.89 in rural areas, with variability in the figures associated with the LFS results. Except in rural areas these numbers are larger than the self-employment rates in the LFS data. The numbers are inflated to reflect the fact that there are many households with more than one worker. About half way through the enumeration it became apparent that households with informal sector businesses were being found at a slower rate than anticipated. As a result, reserve units were sampled and the dismissal protocol was used so that survey resources were focused only on those with informal sector businesses. Table 3.1 displays the resulting sample across regions and locations.

Region	Location				Distribution by Region			
	Ulaan-Baatar	Aimag Center	Soum Center	Rural Area	Total	Share of Sample	Share of Patente	Share of Households
1. West	-	1,632	479	198	2,309	0.12	0.08	0.17
2. Khangai	-	2,910	520	265	3,695	0.20	0.16	0.23
3. Central	-	2,324	733	364	3,421	0.18	0.23	0.19
4. East	-	964	272	99	1,335	0.07	0.05	0.09
5. Ulaanbaatar	8,251	-	-	-	8,251	0.43	0.47	0.30
Total	8,251	7,830	2,004	926	19,011	1.00	1.00	1.00

The primary sampling units are the *aimags* in the regions and the districts in Ulaanbaatar. Within *aimags*,

the *aimag* centers, *soum* centers and rural areas are sampled in the amounts given in Table 3.1. The selection of *aimags* and *soums* and of districts in

²¹ In other words, the multiplier requires a 30-percent larger sample size than a direct question. Of course, we believe the direct question method results in biased estimates.

Ulaanbaatar is presented below. The final selection of households was done in the field after consulting updated household lists. There was some variation in how this was done. In Ulaanbaatar, systematic sampling was done across each district. In some *aimags* specific *bags* were selected with a probability proportional to size methodology before systematic sampling of the selected *bags*. When feasible geographically, systematic sampling across *bags* was conducted to reduce the concentration of observations in a *bag*.

Aimags selection

Each of the five regions is sampled with probability one. Within a region the number of *aimags* selected depends on the overall sample size as reported in Table 2.3. A total of nine *aimags* are selected: two in the West, three each in the Khangai and in the Central region, and one in the East. The selection probability for each *aimag* is the average of the share of *patente*

holders and the number of households. The regions and *aimags* selected are presented in Table 3.2 along with their probability of selection and the proportion of households in the *aimag*.

Aimags within a region are selected by drawing a random number that is associated with the selection probability after it is transformed to a cumulative probability within a region. To see how this works, let us use the probabilities from Region 1 in Table 3.2 as an example. Here the cumulative probability method would select Bayan-Olgii if the random number drawn were between 0 and 0.29 and would select Govi Altai if the random number drawn were greater than 0.29 but not greater than 0.45 (that is, $0.29 + 0.16$). The cumulative probabilities and the randomization device for choosing among *aimags* are provided in an attached Excel spreadsheet. Selections are made without replacement when there is more than one *aimag* to be selected in a region.

Region 1: West

Aimags	Aimags code	Number of Obs.	Selection Prob.	Selected
Bayan-Olgii	2	0	0.29	0
Govii-Altai	5	0	0.16	0
Khovd	16	0	0.11	0
Uvs	15	1000	0.21	1
Zavkhan	9	1000	0.23	1

Region 2: Khangai

Aimags	Aimags code	Number of Obs.	Selection Prob.	Selected
Arkhnagai	1	1000	0.17	1
Bayankhongor	3	0	0.13	0
Bulgan	4	0	0.12	0
Khovsgol	17	1000	0.15	1
Orkhon	21	1000	0.27	1
Ovorkhangai	10	0	0.17	0

Region 3: Central

Aimags	Aimags code	Number of Obs.	Selection Prob.	Selected
Darkhan-Uul	19	0	0.28	0
Dornogovi	6	933	0.20	1
Dundgovi	8	0	0.07	0

Govisumber	22	0	0.03	0
Omnogovi	11	933	0.09	1
Selenge	13	0	0.18	0
Tov	14	934	0.16	1

Region 4: East

Aimag	Aimag code	Number of Obs.	Selection Prob.	Selected
Dornod	7	0	0.30	0
Khentii	18	0	0.29	0
Sukhbaatar	12	1000	0.40	1

Soum selection

Within *aimags* we select respondent households from *aimag* centers, *soums*, and rural areas. *Aimag* centers are included in the sample with probability one and three additional *soums* are selected according to the PPS method with weights as described for the *aimag* selection. Table 3.3 presents the randomly selected *soums* for all regions. Outside of *aimag* centers the

probability weights for *soums* was the average of household and *patente* weights. The probability weight for *aimag* centers is always 1.00, conditional on the *aimag* being selected. For the selected *soums*, Table 3.3 also includes the total number of households, the size of the sample, and the selection probabilities.

Region = 1, aimag = Uvs

Soum Name	Sum Code	HH Total	Observations	Selection probability
Khovd	16	651	100	.0465958
Tes	15	1485	100	.1122324
Tsagaankhairkhan	18	763	100	.0728422
Ulaangom	19	5910	700	1

Region = 1, aimag = Zavkhan

Soum Name	Sum Code	HH Total	Observations	Selection probability
Tosontsenge	4	2073	133	.2403779
Ider	7	823	122	.0387218
Tsetsen-Uul	20	564	121	.0407083
Uliastai	18	4117	932	1

Region = 2, aimag = Arkhangai

Soum Name	mean(sum_cde)	mean(hhtotal)	mean(obs)	mean(s_prob)
Erdenebulgan	18	4150	700	1
Erdenemandal	6	1725	100	.1034546
Khashaat	13	1028	100	.0423621
Ulziit	11	986	100	.0388192

Region = 2, aimag = Khovsgol

Soum Name	Sum Code	HH Total	Observations	Selection probability
Murun	22	7919	955	1
Renchinkhumbé	9	1068	121	.0440513
Tarialan	10	1457	123	.0792718
Tsetserleg	18	1272	125	.0608465

Region = 2, aimag = Orkhon

Soum Name	Sum Code	HH Total	Observations	Selection probability
Bayan-Undur	1	18322	1265	1
Jargalant	2	821	106	.668911

Region = 3, aimag = Dornogovi

Soum Name	Sum Code	HH Total	Observations	Selection probability
Airag	1	881	100	.1021206
Erdene	12	660	100	.0638731
Sainshand	13	4552	634	1
Zamiin-Uud	14	1703	100	.3965035

Region = 3, aimag = Omnogovi

Soum Name	Sum Code	HH Total	Observations	Selection probability
Dalanzadgad	15	3417	830	1
Khanbogd	10	585	117	.1503803
Khankhongor	11	741	147	.1017098
Nomgon	7	755	142	Z.1025376

Region = 3, aimag = Tov

Soum Name	Sum Code	HH Total	Observations	Selection probability
Altanbulag	1	822	135	.0389175
Erdenesant	26	1425	123	.0588547
Zaamar	17	1389	133	.0898976
Zuunmod	27	3301	861	1

Region = 4, aimag = Sukhbaatar

Soum Name	Sum Code	HH Total	Observations	Selection probability
Baruun-Urt	13	3444	864	1
Erdenetsagaan	11	1339	122	.1217302
Menkhkhaan	4	1096	120	.1205302
Uulbayan	9	880	129	.0857333

Ulaanbaatar selection

In Ulaanbaatar we select districts and with the same probability proportional to size methodology used in the outlying areas. The total of 5,100 observations is

selected from six of the nine districts with 850 observations in each district. Systematic sampling over the updated household lists was implemented across the districts.

Sum name	Households	Observations	Prob. Weight	Selected
Bagakhangai	700	0	0.01	0
Baganuur	5.476	0	0.03	0
Bayangol	31.393	1.356	0.15	1
Bayanzurkh	35.682	1.388	0.19	1
Chingeltei	26.285	1.362	0.22	1
Khan-Uul	16.368	1.399	0.08	1
Nalaikh	5.469	0	0.02	0
Songinokhairkhan	35.208	1.384	0.17	1
Sukhbaatar	24.614	1.362	0.15	1

The data collected in the Informal Sector Household Survey (ISHS) is used to generate national and regional estimates of the size of small-scale shadow activity, as well as estimates in urban and rural areas. The selection probabilities discussed in the previous section are essential in making the extrapolation to population totals. In fact it is the inverse of the probability of household being selected that represents the weight of that observation in the population total. As we will see below the weight is equivalent to the number of households that the observation represents in the population as a whole.

The design approach

The underlying principle in the design method is the random selection of households for inclusion in the survey sample. This procedure leads to unbiased estimates of population values and their variances when all the households have a known, non-zero, probability of being selected. As a result of the randomization process the population values can be reported with known precision without imposing any additional assumptions.

The critical information for extrapolating from sample to population values are the selection probabilities. The inverse of the selection probability for a given unit in the sample yields a weight that can be thought of as the number of households that the

unit represents. Selection probabilities and weights for the ISHS are given according to the following formulas.

The primary sampling units in the ISHS are the *aimags* which are selected at random from within each region with their probability proportional to size. The probability of *aimag j* being selected in region *i* is given by

$$P_{ij} = k_i \frac{n_{ij}}{n_i}, \text{ with } n_i = \sum_{j=1}^J n_{ij}$$

and n_{ij} is a measure of the size of the *aimag* and k_i the number of *aimags* selected in a region, which varies from 1 to 3 as shown in Table 3.2, and J the number of *aimags* in the region. The measure of size chosen for the ISHS is the average of the proportion of households and proportion of registered *patente* holders in a selection unit. This measure of size is used to gain efficiency in survey implementation by increasing the probability of sampling in areas where informal activity is believed to be high. The size of an *aimag*, therefore, is given by

$$n_{ij} = \left(\frac{h_{ij}}{h_i} + \frac{r_{ij}}{r_i} \right) / 2$$

where h and r represent the number of households and registered *patente* holders in the relevant geographic areas.

Within *aimags* the *aimag* center is selected with probability 1. Their probability of selection is therefore P_{ij} , the same as the probability of selecting *aimag* j . Outlying *soums* are selected with the same probability proportional to size methodology as the *aimag* so that their probability of selection is

$$P_{ijk} = P_{ij} k_{ij} \frac{n_{ijk}}{n_{ij}}$$

which represents the probability of the *soum* being selected conditional on *aimag* selection. The number of *soums* selected, k_{ij} , is 3 except for Orkhon which is a compact (urban) *aimag* for which no outlying *soums* are selected.

Final selection probabilities for *baghs* and households depend on the number of *baghs* that are identified as being in the rural strata in each *soum* and on the number of households in the updated household listing. Two selection methods were used at the *bagh* and household level. Rural *baghs* were chosen either with probability proportional to size, where size is the number of households in the *bagh* as found in the updated household listing available to the enumerating team at the beginning of the enumeration. Following *bagh* selection, systematic sampling was carried out to obtain the prescribed number of households for enumeration. The alternate method was to use systematic sampling across the *baghs*, when distances between *baghs* were not large. An initial sample of 100 observations were planned for each *soum*. Sixty-seven of the observations are in the *soum* centers and 33 in the rural areas in six of the nine *aimags*. Additional sampling was heaviest in Ulaanbaatar, where informal activity was least dense.

Probability adjustments

Several adjustments to the selection probabilities are necessary in order to draw appropriate conclusions about population parameters. The need for adjustment arises both by design and through the need to make adjustments for non-sampling error. The adjustments that were expected, due to the survey design, are associated with the dismissal protocol implemented to reduce total costs. Since information is gathered

from only a portion of those with wage income and no recent history of self-employment experience, the weights of wage-earners retained in the survey need to be adjusted to yield correct figures for this population, adjusting for the dismissed group. During the ISHS, the initial sample consisted of 3,818 households that were wage-earners and 1,496 (39.18 percent) were dismissed immediately, while 2,322 were retained for further questioning. As a result the weight of each of the 2,322 households must be increased by approximately $\frac{3,818}{2,322} = 1.64$ since we know each of these households must represent roughly 1.64 households in the total sample²².

In addition to the dismissals, probability adjustments were made for survey and item non-response. After dismissing the 1,496 households, an additional 170 failed to complete the survey. In addition, there was a limited amount of non-response to specific questions. Survey weights were expanded so that the remaining households represented the population as a whole. The assumption underlying the technique applied assumes that the households not responding are not, as a group, different from those that do respond to the survey.

After these adjustments were made in accordance with the original survey design, weight adjustments were made to address the new sample size. Since all the dismissals in the secondary sample were either wage earners or unemployed, the responses of these groups in the original sample were expanded assuming that the two groups were represented in the same proportions as in the original sample.

A final adjustment to probability weights was made to address the issues associated with the comparison of treatments. To make treatment comparisons for population means and totals with survey weights it was necessary to expand the weights so that each treatment represented the entire population. This rescaling was achieved by multiplying the weights in each treatment by $\frac{19,011}{N_t}$, $t = 1, \dots, 5$, where N_t is the number of observations in treatment t .

²² In fact, because of the varying weights assigned to households, and slight differences in the average weights across the retained and dismissed subjects, a more accurate figure of 1.59 is used for statistics reported on the wage-earning group.

The measurement of the size of the shadow economy is one of the primary goals of the ISHS. In this section we report on estimates of the size of the informal sector shadow at the national, regional, and location level, as well as estimates for a small number of sectors as identified by the one digit ISIC. Additional results on the characteristics of the shadow economy are found in Section 5.1. Our approach to measuring the gross value added of the informal sector shadow economy is consistent with the production approach to national income accounting as detailed in the System of National Accounts (SNA) developed by the UN and applied by the NSO (UN 1993; NSO 2000). Our data collection goals were informed by UN research on measurement requirements for the informal sector (Becker 2000). Prior to implementation the questions were reviewed and approved by the methodology board of the NSO.

To calculate gross value added, questions on revenues (*rev*), starting and ending inventories of salable goods (*inv*) and of productive inputs (*invin*), as well as expenses (*exp*), were asked of all who reported self-employment income in the month of September. Respondents were instructed to include the value of their own consumption of business output along with revenues. In order to achieve the most complete enumeration of costs, respondents were asked for values in seven categories. These included materials purchased (1) for production and (2) for resale, (3) the cost of fuel and electricity, (4) rents paid for buildings, machinery and vehicles, (5) transportation costs, (6) professional services, and (7) other costs. Gross value added (*gval_add*) for each business unit for the month was calculated as

$$\begin{aligned} gval_add &= \text{output} - \text{intermediate consumption} \\ &= rev + inv - exp + invin \end{aligned} \quad (4.1)$$

where $\text{output} = rev + inv$ and intermediate

consumption = *exp-invin*. The annualized gross value added figures are reported in the tables below.

To generate the annualized figures on the size of the informal sector shadow economy seasonal adjustments were made to the responses for the month of September, 2004. To convert the monthly figures, the respondents were asked to directly compare current month results to those in other seasons (q110). For those with a positive gross value added in September, seasonal adjustments were made based on these responses. A subset of the sample, roughly 25 percent of those with current self-employment income reported that their gross value added for September was less than or equal to zero. For this group extrapolating the monthly result was not feasible, and so an indirect method was used to generate estimates for those reporting negative gross value added. We also applied the indirect method, described in detail below, to a second group of business reporting extremely high value-added, which we term the *upper-tail* group²³. These businesses were flagged because not only was their gross value-added large, but it was inconsistent with two other measures collected to provide a check on the gross value added figures from the income and balance sheet. The first measure was the annual income derived from the household income question. The second measure is the reported business income for the month, which was asked independently from the components of gross value added.

For the bulk of the respondents, the values from these three sources of data tracked very well. For example the average ratio of business income to gross value added for those with positive value added but not in the upper tail was approximately 1.10. Similarly, the ratio of value added to average monthly household income was roughly 1.18. We feel that given the different data generating processes (annual versus monthly) that these ratios are reasonably close

²³The upper-tail group was responsible for the highly non-normal distribution of the value added in the survey.

a result which suggests that the procedure for generating annual gross value added worked reasonably well. The observations in the upper tail of gross value added showed a different pattern, with the identical ratios at levels of .31 and 7.2 respectively. These results indicate that value added was disproportionately large and unadjusted would have led to an overstatement of the size of the informal sector shadow economy.

Adjustments to the upper tail and to those with negative gross value added were carried out by using an OLS regression to generate predicted values of gross value added for these two groups. This was done by estimating the relationship between the income derived from the gross value added statement and monthly business income as reported in q111.

Results of the OLS estimation for Treatment E and Treatment M are below. Note that a herder “dummy” variable was used since the relationship between the two measures of income was significantly different across the herders and all other sectors. The monthly income measure q111 was chosen as the most suitable independent variable due to the lack of aggregation and extrapolation it incorporated. In addition, a term interacting the income measure and the herding dummy was included, since not only the intercept but the slope of the regression differed for the herding sector. The dependent variable, *inc_new* is the income variable derived from the gross value added and is defined as *rev-exp-comp*, where *comp* is compensation for employees, excluding the business owner.

Envelope Treatment

Source	SS	df	MS	
Model	1.8917e+13	3	6.3057e+12	Number of obs = 1365
Residual	7.4177e+13	1361	5.4502e+10	F(3, 1361) = 115.70
Total	9.3094e+13	1364	6.8251e+10	Prob > F = 0.0000
				R-squared = 0.2032
				Adj R-squared = 0.2014
				Root MSE = 2.3e+05

<i>inc_new</i>	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_lherder1	-119717.2	21361.66	-5.60	0.000	-161622.6	-77811.88
q111m	.3180646	.0222857	14.27	0.000	.2743467	.3617826
_lherXq111~1	.6154829	.0819067	7.51	0.000	.4548058	.77616
_cons	140085.9	7898.873	17.73	0.000	124590.6	155581.2

Multiplier Treatment

Source	SS	df	MS	
Model	8.8647e+12	3	2.9549e+12	Number of obs = 2749
Residual	1.0847e+14	2745	3.9514e+10	F(3, 2745) = 74.78
Total	1.1733e+14	2748	4.2697e+10	Prob > F = 0.0000
				R-squared = 0.0756
				Adj R-squared = 0.0745
				Root MSE = 2.0e+05

<i>inc_new</i>	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_lherder_1	-98952.58	12658.83	-7.82	0.000	-123774.4	-74130.78
q111m	.1149619	.0109455	10.50	0.000	.0934996	.1364242
_lherXq111~1	.4823337	.0634559	7.60	0.000	.3579075	.6067598
_cons	132212.2	4334.041	30.51	0.000	123713.9	140710.5

All variables in the regressions for both treatments are significant and the pattern across treatments shows some consistency, giving us more confidence in the results. We observe that the coefficient on the herding sector dummy variable is highly negative in both treatments. This result indicates that herders are reporting lower measures of business income as derived from the gross value added protocol, after controlling for the reported level of business income in q111. The slope term for the herding sector arising from the interaction of q111 and the herder dummy is larger however than for the other sectors. Thus the herder's understatement in the income and balance relative to their q111 responses diminishes as q111 increases.

After using the regression results to generate the predicted value for income for those out of the regression sample, the compensation and inventory changes were added to *inc_new* to generate the gross value added for the upper tail and for those with negative gross value added. The original survey responses were used for those respondents in the middle group on which the estimation was conducted. The monthly numbers were then annualized using q110 which identified seasonal variation in earnings.

A final adjustment was made to account for those households with secondary self-employment income. Of the 7484 households that completed income and balance sheets, 10 percent (748) also reported the existence of an additional self-employment business. For these secondary firms, complete income balance sheet data was not collected. Instead estimates of the value added were developed based on the share of household income associated with that business. On average, the secondary firms in a household, when they existed, were one-third the size of the primary firms. Since 10 percent of the households reported secondary firms, the value added was approximately $0.10 \times 33 = 3.3$ percent of the value added by the primary firms. In the next section we present the results of these calculations.

We present two sets of figures for all of our results, one for the envelope Treatments 2 and 3, which we group together and one for Treatment 5 which used the multiplier method. In the tables and the text these are referred to as Treatment E and Treatment M,

respectively. In both Treatment E and Treatment M the respondent recorded their income and balance sheet responses on an answer sheet which was sealed in an envelope before being returned to the enumerator. In the multiplier treatments, the responses were scrambled as described in Section 2.2 before being sealed in the envelope.

Results on the informal sector shadow size presented in this section include aggregate national totals as well as a breakdown of the annualized measures by employment status, where employment status distinguishes the informal herders from all other self-employed business owners in the informal sector. Figures are reported at the national level as well as at regional and location levels, and by top level ISIC. To get a sense of the relative size of the activity uncovered in the ISHS, we report three ratios. The first compares the NSO estimate of the size of the herding sector in 2004, compiled from their annual census of herds, with that based on the economic values found in the ISHS. It is presented as

$$\text{Herd Ratio} = \frac{\text{Herd}^{\text{ISHS}}}{\text{Herd}^{\text{NSO}}}$$

Also informative is the share of non-herding activity as a share of overall activity, net of estimates of informal activity already incorporated in the GDP figures as a result of the early survey conducted by Bikales *et al.* (1999) The non-herding informal (NH) ratio is given by

$$\text{Informal Ratio} = \frac{\text{NH Informal}^{\text{ISHS}}}{\text{GDP} - \text{Informal}^{\text{BIK}}}$$

Finally we present a ratio that expresses the total additional value added observed in the ISHS as a share of that observed by the NSO. The additional value added (AVA) ratio is given by

$$\frac{\text{Additional Value Added Ratio}}{\text{Additional Value Added Ratio}} = \frac{\text{Informal}^{\text{ISHS}} - \text{Informal}^{\text{BIK}} - \text{Herd}^{\text{NSO}}}{\text{GDP}}$$

where the *Informal*^{ISHS} includes all informal sectors.

Three numbers generated by the NSO are used to create these ratios. The first is the overall estimate of GDP for 2004. The NSO's preliminary estimate of GDP is 1,807 billion togrogs²⁴. At the time of the

²⁴ In constant (1995) prices this represents growth from 2003 of 10.4 percent.

analysis the NSO had not yet estimated the share of this figure that was associated with the Bikales estimates of the informal sector shadow. As a result we estimate the $Informal^{Bik}$ figure using previous estimates of the Bikales values and total GDP for 2000 to 2003. $Informal^{Bik}$ has risen from 11.37 percent in 2000 to 13.73 percent in 2003 according to the NSO's results. An ordinary least squares regression of the time trend yields an estimate of the informal sector shadow economy that can be applied to the current NSO GDP estimates. This process yields a figure of 14.34 percent for 2004. Thus the baseline informal sector GDP of Mongolia is given by

$$0.1434 * GDP = 259 \text{ billion togrogs} \quad (4.2)$$

The NSO's estimate of the herding sector is approximately 18.27 percent of the total GDP for 2004 or 330 billion togrogs.

Table 4.2 shows that the mean estimate of the share of the shadow economy derived from the ISHS survey depends on the treatment with the mean value equal to 766 billion togrogs in Treatment E and 584 billion togrogs in Treatment M. The non-herding informal sector comprises roughly 63 percent (60 percent) for Treatment M (Treatment E).

Envelope treatment

Total	Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
informal		462658	36713.69	390661.4	534656.1	.8170159
inf_herd		302916	58317.46	188553.1	417280	5.918508
Total		765575	67974.99	632272.9	898877.6	2.204294

Multiplier treatment

Total	Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
informal		368776	61792.17	247633.9	489920	5.471972
inf_herd		214723	26795.97	162190.4	267257	3.874433
Total		583500	67089.85	451971.5	715029.7	5.308699

At the national level, the three ratios on shadow activity are presented in Table 4.3. In Treatment E, the ISHS measure of the herding sector, at 303 billion togrogs, is roughly 92 percent of the the NSO's figure. As in most sectors, Treatment M revealed a smaller amount of value added, 215 billion togrogs, equivalent to 65 percent of the NSO's estimate. The informal, non-herding, shadow size is estimated to be 30 percent and 24 percent of the observed GDP, respectively, in Treatment E and M, after excluding the $Informal^{Bik}$ values from the NSO estimates. This, we think is a very relevant statistic for estimating the size of the informal sector since it essentially serves as a replacement of the Bikales number. Note that the mean estimates are larger than those extrapolated from the Bikales survey, estimated at 14 percent for 2004, with the additional activity arising from both

the coverage of additional sectors and apparent growth relative to GDP in some important sectors. The sectoral distribution of value-added will be discussed in greater detail below in reference to Table 4.6.

The final ratio presented in Table 4.3 is the AVA ratio which measures the additional value added captured by the ISHS survey as a share of the newly estimated GDP. As shown in the previous tables and the previous ratios Treatment E reveals an additional 9.7 percent of economic activity while Treatment M suggests it is unchanged. As will be discussed below, the sectoral distribution of activity has changed dramatically, and the AVA ratio, relative to the NSO estimates, is based on a smaller herding sector and a larger non-herding informal sector.

Treatment	Herd Ratio	Shadow	AVA Ratio
Envelope	.9171901	.2987361	.0973704
Multiplier	.6501542	.2381171	-.0033354

Notes: Tables provide the ratios of activity. “Herd_ratio” is defined as $\text{Herd}^{\text{IRIS}}/\text{Herd}^{\text{NSO}}$. “Shadow” is defined as the Non-herding Informal/(GDP-Bikales). “AVA_ratio” is $(\text{Herd}^{\text{IRIS}} + \text{Non-herd Informal}^{\text{IRIS}} - \text{Informal}^{\text{Bikales}} - \text{Herd}^{\text{NSO}})/\text{GDP}$

Envelope treatment regional estimates

Total Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
West informal	67613.01	10656.59	46714.91	88511.11	.5082051
West inf_herd	64036.89	17538.14	29643.72	98430.06	2.261788
Khangai informal	72989.94	10143.54	53097.96	92881.92	.5949976
Khangai inf_herd	109675.1	50571.5	10501.9	208848.4	7.955758
Central informal	80732.64	21466.64	38635.5	122829.8	2.432345
Central inf_herd	87708.94	22584.53	43419.55	131998.3	4.460202
East informal	13243.54	5258.843	2930.688	23556.39	.8178075
East inf_herd	33991.43	7153.453	19963.16	48019.71	2.126052
Ulaanba informal	228079.6	26636.47	175844.1	280315	.646969
Ulaanba inf_herd	7504.15	2270.255	3052.067	11956.23	.5561744

Multiplier treatment regional estimates

Total Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
West informal	33616.65	6370.422	21127.48	46105.81	.7369472
West inf_herd	44515.95	11260.33	22440.16	66591.75	1.386683
Khangai informal	94409.63	60280.71	-23770.22	212589.5	8.795584
Khangai inf_herd	69999.98	12416.94	45656.67	94343.29	7.885287
Central informal	44090.66	6691.523	30971.98	57209.34	1.196376
Central inf_herd	70603.54	18232.55	34858.77	106348.3	6.687613
East informal	9990.853	1937.482	6192.435	13789.27	.6840574
East inf_herd	24581.44	10063.19	4852.639	44310.23	5.39251
Ulaanba informal	186669.1	10733.99	165625.2	207713	.5579795
Ulaanba inf_herd	5022.811	3291.728	-1430.596	11476.22	.8130742

Variation of shadow activity by region and location is presented in Table 4.4 and Table 4.5. For the non-herding informal sector, Ulaanbaatar dominates in value with mean estimates of the share of activity 49 percent in Treatment E and 51 percent in Treatment M. After Ulaanbaatar activity is greatest in Khangai, followed by the Central, Western, and

Eastern regions. There is some variability across treatments as to the distribution of the types of activity at the regional level. In particular, in the Khangai and Central regions, where the bulk of the herding was found, the multiplier treatment revealed significantly less activity, leading us to believe that the multiplier

treatment was not well received in the herding households. In addition, in the rural areas we did not locate many informal sector (non-herding) shadow

economy businesses. Note that when the herding sectors are included, the rural areas are second only to Ulaanbaatar in terms of total informal sector output.

Envelope treatment location estimates

Total	Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
	Ulaanbaa informal	228079.6	26636.47	175844.1	280315	.646969
	Ulaanbaa inf_herd	7504.15	2270.255	3052.067	11956.23	.5561744
	Aimag_ce informal	128707.2	12489.91	104213.9	153200.6	.3916935
	Aimag_ce inf_herd	14813.85	4764.767	5469.905	24157.8	.2891107
	Soum_cen informal	89999.1	22540.31	45796.44	134201.8	2.396606
	Soum_cen inf_herd	7339.272	4104.227	-709.3203	15387.86	1.653525
	Rural informal	15872.82	6087.233	3935.449	27810.18	6.630088
	Rural inf_herd	273259.3	57999.06	159520.2	386998.3	6.94573

Multiplier treatment location estimates

Total	Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
	Ulaanbaa informal	186669.1	10733.99	165625.2	207713	.5579795
	Ulaanbaa inf_herd	5022.811	3291.728	-1430.596	11476.22	.8130742
	Aimag_ce informal	85930.05	6391.349	73399.86	98460.24	.3685774
	Aimag_ce inf_herd	6344.091	2204.18	2022.813	10665.37	.2563898
	Soum_cen informal	83431.61	60451.44	-35082.96	201946.2	9.287993
	Soum_cen inf_herd	29653.56	16945.32	-3567.618	62874.73	2.842539
	Rural informal	12746.13	5206.918	2538.009	22954.25	4.305271
	Rural inf_herd	173703.3	20457.81	133595.9	213810.6	7.559049

As mentioned above with regard to the informal sector ratio in Table 4.3, the ISHS revealed substantially more value added than those extrapolated from the Bikales *et al.* (1999) estimates which covered trade, restaurant, and transportation sectors. The growth in the informal sector is not, however, due only to the additional sector coverage. In fact the trade sector, which is the largest of the sectors measured in the ISHS at 265 billion togrogs (208 billion togrogs) in Treatment E (Treatment M), is close in size to the total estimates extrapolated from the Bikales survey (259 billion togrogs). Direct comparisons with the restaurant sector indicate growth as well as additional coverage of sub-sectors

in the ISHS, which included estimates of informal lodging activity. The transport sector, which is the third largest sector measured by the ISHS, is smaller than the updated estimate of 102 billion togrogs derived from the measures of Bikales *et al.* (1999). The ISHS reports 48 billion togrogs in Treatment E and 35 billion in Treatment M. The smaller transport sector is surprising and indicates either a) under-coverage in the current survey, b) an overestimate in the previous informal sector survey, or c) a real decline in the informal sector transport sector which could arise from either a smaller sector or, more likely, an increase in the formality of the sector.

Envelope treatment location estimates

Total	Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
Ag		322714.3	58822.09	207361.2	438067.4	5.605484
Fishing		30.29535	30.29535	-29.11535	89.70606	0.4505077
Mining		3814.865	1512.79	848.2075	6781.523	1.413589
Manufact		79777.81	9709.726	60736.55	98819.07	0.8499624
Elec_gas		348.0964	270.7714	-182.9	879.0928	0.4146345
Construc		5107.539	1541.498	2084.584	8130.494	0.5577638
Trade		264640.1	32306.88	201284.7	327995.5	0.8748234
Hotel_Re		26013.19	5837.979	14564.62	37461.76	0.6886468
Transpor		48416.68	6866.617	34950.89	61882.46	0.4936939
Finance		3761.817	1932.469	-27.85063	7551.485	0.7386012
Real_Est		4316.387	1384.739	1600.844	7031.929	0.5843889
Educ		354.4417	224.8522	-86.50483	795.3882	0.4601628
Health		5322.366	4305.007	-3119.969	13764.7	0.359852
Other_Co		582.2341	9812.765	-18661.09	19825.56	0.6484286
Priv_hou		373.5442	332.4321	-278.3719	1025.46	0.7020484
Other		1.592725	1.592725	-1.530688	4.716137	0.0683262

Multiplier treatment location estimates

Total	Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
Ag		245468.7	32748	181266.5	309670.9	4.434961
Fishing		541.1902	458.6076	-357.9066	1440.287	7.998247
Mining		2516.155	804.1304	939.6639	4092.647	1.183033
Manufact		53779.98	5143.895	43695.41	63864.55	0.5944037
Elec_gas		468.8083	182.4192	111.1768	826.4397	0.5213296
Construc		4805.031	1202.577	2447.388	7162.674	0.5019692
Trade		208474	58344.49	94090.05	322857.9	6.244412
Hotel_Re		12381.85	2968.123	6562.867	18200.83	0.7866985
Transpor		34797.74	5470.339	24073.18	45522.3	0.7547262
Finance		2116.247	1435.573	-698.1834	4930.677	0.4598493
Real_Est		3713.169	1106.898	1543.104	5883.233	0.6124611
Educ		1789.043	1402.336	-960.2259	4538.313	0.6066508
Health		2653.418	1181.293	337.5012	4969.334	1.376959
Other_Co		9249.018	1385.108	6533.524	11964.51	0.6134121
Priv_hou		593.9755	367.5629	-126.6288	1314.58	0.6939072
Other		100.7402	50.02167	2.673126	198.8073	0.9539653

Two other results regarding sector size that are surprising are the relatively small size of the mining sector and a large manufacturing sector. Recent research suggests that gold mining in particular is a

significant component of informal activity (MBDA 2003). One possibility is that actors in the mining sector may have misreported the nature of their activity to retain secrecy. The large manufacturing and trade

sectors may be capturing some of this activity, although it would require additional work, such as random resurveying of respondents in these groups to determine the extent of a problem with misclassification. Alternatively, the distribution of mining activity may be such that it is not amenable to observation in a national level survey. Note that the level of mining activity in the ISHS is roughly comparable to those uncovered in the recent Labor Force Survey, with regards to the number of workers²⁵.

One further set of results is presented before discussing issues associated with the precision of the estimates. The figures in Table 4.7 present the shares of the components of activity by breaking the total supply of goods and services into mixed income (which includes indirect taxes), compensation, depreciation, and intermediate consumption. These results are fairly consistent across treatments with the exception of some smaller sectors such as

education and health in which coverage was sparse.

Accuracy and non-normality of the data. All estimates of the size of the informal economy derived from the ISHS enumeration are accompanied by confidence intervals that are a measure of the accuracy of the estimates. A 95-percent confidence interval implies that if the survey were conducted 100 times, with a new random sample of households drawn from the same sampling frame each time, we would expect the mean value calculated to be within the confidence interval 95 of those 100 times. While we would like this interval to be small, in most of the results presented above the confidence intervals are fairly large. One reason for this is that the confidence interval calculations are based on an assumption that the underlying values are normally distributed, whereas the actual distribution is not normal. Instead we find a much higher number of observations with extreme values increasing the variance of the estimates.

Envelope Method

ISIC at top level -

numeric	mean(mix~r)	mean(dep~r)	mean(comp~r)	mean(int~r)	mean(tot~r)
Ag	0.51	0.13	0.04	0.32	1.00
Fishing					
Mining	0.58	0.00	0.02	0.39	1.00
Manufact	0.58	0.01	0.04	0.37	1.00
Elec_gas_water	0.68	0.00	0.00	0.32	1.00
Construction	0.46	0.02	0.19	0.34	1.00
Trade	0.55	0.02	0.02	0.41	1.00
Hotel_Rest	0.39	0.07	0.15	0.39	1.01
Transport_Storage	0.56	0.06	0.02	0.36	1.00
Finance	0.65	0.10	0.00	0.24	1.00
Real_Estate	0.43	0.08	0.14	0.35	1.01
Educ	0.81	0.02	0.00	0.16	1.00
Health	0.05	0.12	0.33	0.50	1.01
Other_Comm_Soc_Pers	0.60	0.03	0.03	0.34	1.00
Priv_households	0.40	0.02	0.00	0.58	1.00
Total	0.52	0.05	0.07	0.36	1.00

²⁵ Tables on the number of households by sector are included in Section 5

Multiplier Method

ISIC at top level -

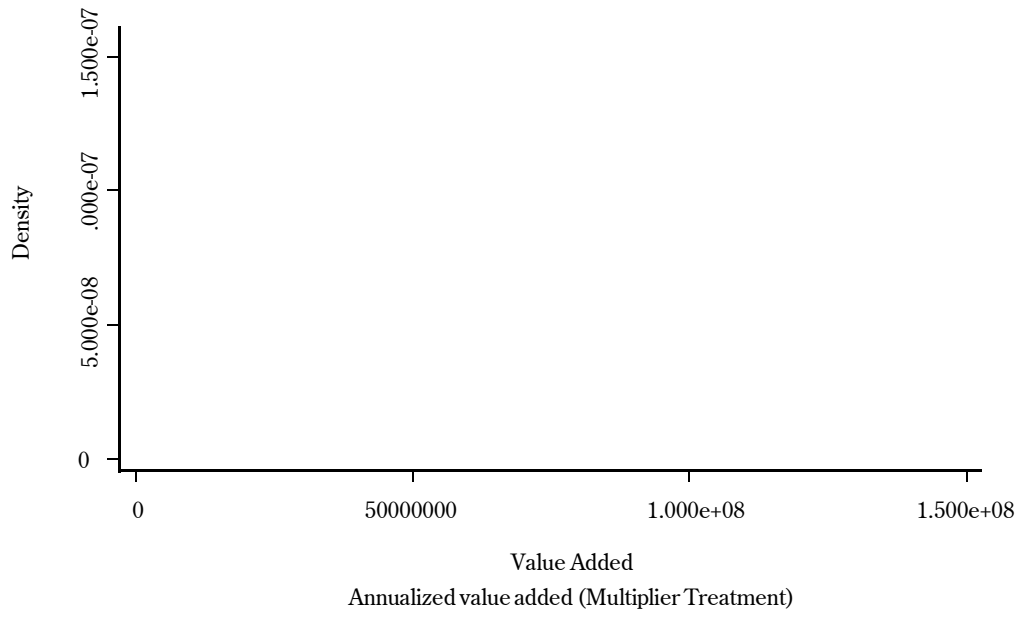
numeric	mean(mix_~r)	mean(dep_~r)	mean(comp~r)	mean(int_~r)	mean(tot_~r)
Ag	0.56	0.05	0.07	0.32	1.00
Fishing	0.84	0.01	0.00	0.16	1.00
Mining	0.37	0.11	0.26	0.26	0.92
Manufact	0.59	0.02	0.05	0.35	1.00
Elec_gas_water	0.45	0.09	0.07	0.39	1.01
Construction	0.42	0.03	0.27	0.28	1.00
Trade	0.60	0.02	0.02	0.37	1.00
Hotel_Rest	0.44	0.03	0.11	0.42	1.00
Transport_Storage	0.55	0.05	0.03	0.37	1.00
Financ	0.50	0.04	0.08	0.38	1.00
Real_Estate	0.47	0.07	0.09	0.37	1.01
Educ	0.10	0.09	0.29	0.52	1.02
Health	0.47	0.07	0.14	0.31	1.01
Other_Comm_Soc_Pers	0.51	0.03	0.07	0.39	1.00
Priv_households	0.55	0.07	0.00	0.38	1.00
Total	0.49	0.05	0.10	0.35	1.00

One indicator of the degree of non-normality of the value added data can be derived from the *interquartile range* of the distribution. The interquartile range identifies the width of the band between the 25th and 75th percentile, and a region of outliers derived from this measure that should comprise 0.7 percent of the values in a normal distribution contained more than 11 percent of the observations in both the envelope and multiplier treatments of the ISHS sample. Our tests for outliers indicate that these firms gave economically consistent results and so should not be excluded from the dataset. These outliers, however, comprise a large share of the value added, 34 percent in the envelope treatment and 45 percent in the multiplier treatment after the adjustment process described above. The outliers are surprisingly dense in the *soum* centers of the west and Khangai

regions, suggesting that refinements and updates of the informal sector measurements might sample more heavily from these areas. The potential for optimizing future efforts based on what we have learned about the variability of value added and survey costs is further discussed in Section 6.2

Figure 4.1 presents the density of value added for the informal sector (non-herding) in the multiplier treatment²⁶. Particularly noticeable is the rapid decline in the number of firms at relatively low levels of value added. A normal distribution would have significantly more firms in the area from zero to 50,000,000 togrogs, given the number of firms in the upper tail. The estimates of confidence bands implicitly assumes that those firms in fact do exist which creates the larger than expected confidence intervals.

²⁶ Using the envelope treatment or including the herding sector does not substantially change the graphic or the issues associated with the non-normality..



While the GDP measurement results reported in Section 4 provide important insights, in this section we extend the analysis in three stages in order to come to a fuller understanding of the informal sector shadow economy in Mongolia. Section 5.1 presents relevant descriptive statistics on business, household, and personal characteristics. Where relevant, comparisons are made to formal sector business owners, wage earners, and the unemployed. To provide the most informative picture of the informal shadow economy, descriptive statistics are presented at varying levels of disaggregation. At a minimum we break the informal sector into two parts, a herding sector (ISIC 012) and all other sectors. References to the herding sector will be explicit in both the text and tables and so throughout this section, whenever the informal sector is mentioned, the animal herding sector is excluded. Additional statistics by ISIC are presented both to understand the overall distribution of activity and because differences across ISIC, for example with regard to regulatory environment, are relevant to understanding causes or consequences of shadow participation in Sections 5.2 and 5.3.

While the discussion and tables below create a rich picture of the informal sector, one should keep in mind that it is but a descriptive picture based on bivariate correlations (relationships between two characteristics). In order to generate a more insightful and precise understanding additional multivariate analysis is recommended (see Section 6.2).²⁷ One such model, presented in Table 5.74, is used to assess how formal and informal sector business owners differ. This approach adds additional insight by allowing the analyst to control for personal characteristics, social attitudes, location and sector characteristics. This section does not attempt to pool treatment observations or sort out which elicitation (treatment) method is most appropriate for a particular question. Again, this important task must await the

next phase of analysis.

Finally, note that most tables in this section end with a row labeled “Total”. This refers to the average value of all observations related to the column’s entries and is also weighted by the characteristic used to disaggregate the column. Thus, the “Total” is not a simple average of the figures in the column.

In this section we lay the groundwork for understanding the causes and social consequences of the informal sector shadow economy by presenting data from the ISHS on personal, household and business characteristics. While our primary interest is in informal sector self-employment there is much to be learned from a comparative analysis. As a result we present data on the formal and animal herding sector alongside the informal sector results. When relevant we also present data collected from wage earners and the unemployed.

Household and personal characteristics

Based on data on household and personal characteristics in survey Sections 2, 3, 4, and 5 we compute descriptive statistics on comparisons across all respondent types. With the exception of the unemployed,²⁸ we find from Table 5.3 that there is little age variability across the different groups, although formal sector workers are slightly older than both informal workers and wage workers. Gender, as shown in Table 5.4 is fairly balanced across the different employment statuses with the exception of herding, where 66 percent of the respondents are male. Among the self-employed, women show a slight majority in the formal sector among primary earners, and predominate in all sectors among the secondary earners.

²⁷ Examples of such methods includes probit and multinomial logit analysis, analysis of variance, multiple regression (both single and multi-equation), to mention a few.

²⁸ The survey lumps the unemployed and retired together. Hence, below, the employment status of unemployed should be interpreted as including pensioners.

The distribution of all respondent types is given in Table 5.1. We see that enumerators were largely successful in achieving an important goal of the survey protocol, which was to reach the primary earner among the self-employed. Among the currently self-employed, 6,914 (92 percent) of the respondents were the primary earner and 592 (8 percent) the secondary earner.

For purposes of reporting our results respondent type is not the most informative characteristic. Instead we distinguish three self-employment categories: formal, informal, and informal herding sectors. With the exception of our discussion of shadow exit and entry, where we are interested in whether there has been previous self-employment, we group the wage earners together. The unemployed remain a separate category. The resulting five categories we term *employment status*. Table 5.2 presents the survey respondents categorized by employment status, and the estimate of the number of households associated with each employment status after the survey weights are applied—i.e., reflecting the country as a whole. Here we see that informal employment in the herding and non-herding sectors is associated with 48 percent of the households while formal sector self-employment is associated with 2.61 percent.

Table 5.5 shows that no major differences exist in marital status across employment status, though herders appear to be more conservative, eschewing divorce and separation. Regarding education in Table 5.6, non-herder informals in general have a higher level of education than herders and the formal sector in turn has a higher level of education than the non-herder informals. Finally, Table 5.7 indicates that while there are no big differences in family size across employment status, the unemployed have a family size about 10 percent lower than the others.

Importance of self-employment income to household

Importance	Freq.	Percent	Cum.
Largest self-employed	6.914	36.37	36.37
Secondary self-employed	592	3.11	39.48
Previous self-employment	290	1.53	41.01
Never had	7.068	37.16	78.18
unemployed/pension	3.940	20.72	98.91
refused/unavailable	207	1.01	100.00
Total	19.011	100.00	

In-sample distribution

status2_n	Freq.	Percent	Cum.
informal	5.695	29.56	29.56
inf_herd	1.106	5.82	35.77
formal	668	3.51	39.28
wage	7.382	38.82	78.12
unemp	3.953	20.78	98.91
Refus/unaval	207	1.01	100.00
Total	19.011	100.00	

Country-wide distribution

status2_n	Freq.	Percent	Cum.
informal	130.538.8	22.93	22.93
inf_herd	140.383.22	24.66	47.59
formal	14.885.886	2.61	50.21
wage	182.088.35	31.99	82.19
unemp	101.355.48	17.81	100.00
Total	569.251.73	100.00	

Mean	Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
informal		39.27087	.2139249	38.85154	39.6902	1.224964
inf_herd		40.05263	.5546379	38.96545	41.13981	5.365314
formal		40.97127	.4584415	40.07265	41.86989	.6926921
wage		39.5766	.3859168	38.82014	40.33306	6.261933
unemp		56.98567	.666832	55.67858	58.29277	4.267169

Question16

status2_n	Male	Female	Total
informal	.5047	.4953	1
inf_herd	.6567	.3433	1
formal	.4493	.5507	1
wage	.5026	.4974	1
unemp	.488	.512	1
Total	.5371	.4629	1

Key: row proportions

Question18

Employment							
Status	Never married	Married	Cohabit	Separated	Divorced	Widowed	Total
informal	.0806	.7845	.0307	.014	.0303	.0601	1
inf_herd	.096	.8163	.0145	3.2e-05	.006	.0672	1
formal	.0381	.8694	.016	.0148	.0193	.0424	1
wage	.0749	.7904	.0408	.0156	.0172	.0612	1
unemp	.0496	.6156	.0057	.0132	.0153	.3006	1
Total	.0759	.7664	.0251	.0109	.0171	.1045	1

Key: row proportions

Employment Status	Incomplete		Incomplete		Incomplete		Incomplete		Total
	Primary	Secondary	Primary	Secondary	Tech./vo	Univ.	Tech./vo	Univ.	
informal	.0028	.0396	.2014	.4113	.0685	.1434	.1285	.0045	1
inf_herd	.0366	.2325	.4348	.2065	.0251	.0456	.0167	.0022	1
formal	0	.0098	.0868	.2623	.0441	.2266	.3401	.0304	1
wage	.0051	.0178	.1671	.2814	.0597	.2033	.2441	.0215	1
unemp	.0216	.25	.3042	.2006	.0323	.1042	.0863	8.7e-04	1
Total	.0152	.1169	.2633	.2778	.0479	.1336	.1359	.0094	1

Key: row proportions

Employment status	Mean score
informal	4.432
inf_herd	4.552
formal	4.364
wage	4.439
unemp	3.986
Total	4.383

Social attitudes

In this section we report on the results of questions that provide insight into the social and personal attitudes that may play a role in the type of economic decisions that individuals make. These questions cover a range of issues including patience, trust—including trust of government, trustworthiness, fairness and honesty, and the willingness to take risks.²⁹ There are noticeable differences across groups for many of these questions, and in general the herding sector has quite different attitudes than those of the formal and informal sector business owners, who are often similar to each other. We also report on the responses of the wage earners and the unemployed in this section.

We see in Table 5.8 that the formal sector business owners exhibit significantly less patience than others, a result that is consistent across two question types.³⁰ Their willingness to do an unspecified but tedious task (q30) is significantly lower than for all the other respondents and they show less willingness to wait for service in a store (q36). At the other extreme are the herders who demonstrate the most patience with regard to waiting times. The three groups of the self-employed exhibit much less variability with regard to this characteristic when contrasted with the wage earners and the unemployed.

²⁹ Because of the sensitivity of these questions for the respondent, we report the descriptive statistics to these answers based on the treatment protocols that used the Envelope Method of elicitation. See Section 2.2 for further discussion of the issues involved.

³⁰ Note that the “average” yields a group response of less than 50 percent, indicating a bias or group myopia of one’s own relative position.

Self-declared willing to do an unspecified but tedious task, relative to others

Employment Status	Question30					Total
	Much higher	Higher	Average	Lower	Much low	
informal	.2062	.2725	.3332	.119	.0691	1
inf_herd	.1702	.2786	.3575	.1456	.0481	1
formal	.1108	.2333	.4044	.1687	.0828	1
wage	.1915	.2475	.3763	.1033	.0814	1
unemp	.1938	.2554	.3393	.1315	.0799	1
Total	.1879	.2619	.3559	.1241	.0701	1

Key: row proportions

Willing to wait for service at a shop:

Employment status	Estimate	Std. Err.	[95% Conf. Interval]		Deff
informal	10.03879	.262588	9.524073	10.5535	1.713189
inf_herd	13.47106	.4256569	12.63671	14.30542	5.439461
formal	8.284828	.3888062	7.522705	9.046951	.8119865
wage	11.89009	.582362	10.74857	13.03162	7.440421
unemployed	12.19621	.7241248	10.77681	13.61561	8.011775

Wait time in minutes

Table 5.9 shows that the herding sector reports much higher levels of trust than the other groups, with the formal sector indicating the least trust. Informal, wage, and unemployed respondents show increasing levels of trust, between these two extremes. With regard to whether people perceive

the respondent as trustworthy, there is little variability across groups. A slightly higher proportion of herders and the unemployed indicate higher trustworthiness than the other groups, although the differences are not statistically significant.

Employment Status	Question31			Total
	Most	About half	Few	
informal	.3468	.3907	.2625	1
inf_herd	.6226	.2297	.1477	1
formal	.358	.3015	.3405	1
wage	.438	.3127	.2492	1
unemp	.522	.2826	.1954	1
Total	.4755	.3045	.22	1

Key: row proportions

Employment	Question33					Total
	Much more	More	Same	Less	Much less	
informal	.0804	.1945	.6634	.0528	.0089	1
inf_herd	.1016	.2104	.6384	.0411	.0084	1
formal	.0649	.2422	.6641	.0261	.0027	1
wage	.0771	.1677	.7163	.0283	.0106	1
unemp	.1121	.2059	.6344	.0367	.011	1
Total	.0898	.1931	.669	.0385	.0095	1

Key: row proportions

Frequency of attendance at religious establishments is fairly low nationwide, with 73 percent attending either not at all or only once or twice a year. There is some variability across locations with more

attendance in urban areas. We cannot determine from this data, however if the difference results primarily from proximity to religious institutions.

Frequency	Ulaanbaar	Aimag Center	Soum Center	Rural	Total
Never	0.2446	0.2467	0.3235	0.3963	0.3088
Once/twi	0.3936	0.4193	0.4683	0.3924	0.4188
About ev	0.1771	0.1554	0.1276	0.0929	0.136
Monthly	0.0959	0.1053	0.0383	0.0657	0.0734
More tha	0.0889	0.0733	0.0423	0.0527	0.063
Total	1	1	1	1	1

Key: column proportions

The herders and unemployed indicate a greater commitment to honesty, when asked if there are ever good reasons to lie to someone else (q34). Self-employed business owners report the least honesty and wage earners intermediate values. There is little difference between formal and informal business owners with regards to this question. Similar results are found regarding character issues such as whether

people will take advantage of others, given the chance, or be fair (q35). The fairness question, perhaps because it is dichotomous, shows the biggest difference across groups, with 73 percent of herders believing people would be fair as compared to only 42 percent and 49 percent of the formal and informal self-employed.

Employment status	Strongly Agree	Mildly Agree	Unsure	Mildly Disagree	Strongly Disagree	Total
informal	.1563	.3707	.1782	.1345	.1603	1
inf_herd	.0815	.2672	.1573	.149	.345	1
formal	.1739	.3735	.1424	.1228	.1874	1
wage	.1166	.3958	.1573	.1247	.2057	1
unemp	.0812	.2221	.2612	.1236	.3119	1
Total	.1122	.3268	.1802	.1327	.2481	1

Key: row proportions

Employment Status	take adv	be fair	Total
informal	.5119	.4881	1
inf_herd	.2694	.7306	1
formal	.5812	.4188	1
wage	.4655	.5345	1
unemp	.2828	.7172	1
Total	.3983	.6017	1

Key: row proportions

With regard to risk attitudes the herding sector is the least risk averse based on two measures, 1) their willingness to pay for a risky lottery is significantly greater than all other groups, and 2) given a choice between a risky lottery (Game A in Table 5.15) and a safe one (Game B) the herders are most likely to choose the risky one. The differences among the other groups are not statistically significant. There seems to be an accurate self-perception among the herders that they are more risk seeking than others (q46).

Employment status	Willingness to play game (togrogs)	household income
informal	1816	1563116
informal herders	2519	864889
formal	2046	2637707
wage	1900	1481592
unemployed	1348	413549

Employment status	Game A	Game B	Total
informal	.3652	.6348	1
inf_herd	.4509	.5491	1
formal	.3757	.6243	1
wage	.4196	.5804	1
unemp	.39	.61	1
Total	.4084	.5916	1

Game A is the riskier game

Employment status	less willing	equally	more willing	Total
informal	.4169	.4918	.0913	1
inf_herd	.37	.4754	.1546	1
formal	.4018	.5053	.0929	1
wage	.3952	.5075	.0973	1
unemp	.5019	.401	.0971	1
Total	.4131	.477	.1099	1

Key: row proportions

Finally a question was asked about attitudes regarding competence and reliability of the government and was supplemented by a question to identify those who have had problematic interactions with government officials. The first, (q40) asked “How much of the time can you trust the government to do what is right?” and the second (q41) whether they had ever been unfairly treated by a government official. The responses to the first question, presented in, Table 5.17 are dispersed across employment status in a way that is similar to those we have observed in other attitudinal questions. The unemployed expressed the most confidence in government

activity followed closely by those in the herding sector, with business owners in the formal sector expressing the least. In the population as a whole about 60 percent believe that the government will often or always do what is right. This varies between 48 percent in the formal sector and 69 percent in the informal herding sector. The attitudinal responses were closely correlated with the information respondents provided on the quality of their actual contacts and resulting treatment by government officials (Table 5.18). While only 19 percent of herders had experienced unfair treatment by a government official, 47 percent in the formal sector reported this - a figure that falls to 32 percent in the informal sector.

Employment Status	Question40					Total
	Never	Rarely	Sometime	Often	Always	
informal	.0518	.1191	.3039	.355	.1702	1
inf_herd	.0246	.0457	.2411	.3563	.3322	1
formal	.0457	.139	.332	.3594	.124	1
wage	.0545	.0863	.2802	.3753	.2037	1
unemp	.0438	.0909	.2518	.3557	.2578	1
Total	.0444	.086	.2723	.3621	.2352	1

Key: row proportions

Question41

Employment Status	Never	Once/twice	More than twice	Total
informal	.6847	.2232	.0921	1
inf_herd	.8114	.1443	.0443	1
formal	.5291	.3232	.1477	1
wage	.6526	.262	.0855	1
unemp	.7394	.1932	.0674	1
Total	.7113	.2134	.0753	1

Key: row proportions

Status distribution of the informal sector economy

The survey indicates that a total of 285,808 households had current self-employment income in September 2004, with the largest number, 140,383, in the informal herding sector followed by informal non-herding, and a much smaller formal sector. The distribution by region (Table 5.20) and location (Table 5.21) shows variability as expected with Ulaanbaatar and the *aimag* centers in Khangai and Central regions having the bulk of the formal activity.

Status	Number of households
informal	130.539
inf_herd	140.383
formal	14.886
Total	285.808

Figures are based on primary self-employment activity

Employment status

region	informal	inf_herd	formal	Total
West	14.686	36.655	883	52.224
Khangai	29.445	52.106	3.925	85.476
Central	20.195	32.999	2.499	55.693
East	3.822	16.134	95	20.050
Ulaanbaatar	62.390	2.489	7.485	72.364
Total	130.539	140.383	14.886	285.808

Employment status

Location	informal	inf_herd	formal	Total
Ulaanbaatar	62.390	2.489	7.485	72.364
Aimag center	35.265	3.346	4.434	43.044
Soum center	25.924	7.201	2.418	35.543
Rural	6.960	127.347	550	134.856
Total	130.539	140.383	14.886	285.808

The proportion of households with self-employment activity is presented below, also with tabulations at the regional and location level. At the national level the informal makes up 46 percent, informal herding 49 percent, and the formal sector 5 percent of the self-employed households according to the NSO's classification methodology.

Empl. status	Est. Prop.	Std. Err.
informal	0.456736	0.011041
inf_herd	0.491180	0.011813
formal	0.052084	0.002759
Total	1.000000	

region= West

status	Est. Prop.	Std. Err.
informal	0.281215	0.021793
inf_herd	0.701879	0.022416
formal	0.016906	0.002700

region=Khangai

status	Est. Prop.	Std. Err
informal	0.344483	0.022003
inf_herd	0.609598	0.023356
formal	0.045919	0.005748

region=Central

status	Est. Prop.	Std. Err
informal	0.362619	0.022102
inf_herd	0.592518	0.023412
formal	0.044862	0.006587

region=East

status	Est. Prop.	Std. Err
informal	0.190609	0.021729
inf_herd	0.804656	0.022008
formal	0.004735	0.002391

region=Ulaanbaatar

status	Est. Prop.	Std. Err
informal	0.862171	0.006568
inf_herd	0.034401	0.003595
formal	0.103429	0.005747

location=Aimag center

status	Est. Prop.	Std. Err.
informal	0.819266	0.007743
inf_herd	0.077723	0.005536
formal	0.103012	0.006032

location=Soum center

status	Est. Prop.	Std. Err.
informal	0.729372	0.027167
inf_herd	0.202612	0.025695
formal	0.068016	0.011710

location=Rural

status	Est. Prop.	Std. Err.
informal	0.051607	0.009274
inf_herd	0.944316	0.009585
formal	0.004077	0.002532

Sectoral distribution of the informal sector shadow by ISIC

In this section we look more closely at how economic activity is distributed across ISIC codes. Figures on the formal sector are provided for comparison, and tables for national, regional, and location specific distributions are presented for the top level ISIC. We present two types of cross-tabulations at the national level, the first showing the share of activity within each ISIC by employment status, and the second showing the share across informal status, so that the

relative size of the informal and formal sectors can be more easily appreciated.

The first panel in Table 5.25 shows that trade activity, which includes wholesale, retail, and repairs, represents 50 percent of total economic activity in both the formal and informal sectors. The second panel in Table 5.25 shows that within the trade sector 90 percent of the households are in the informal sector.

Taking a closer look at trade by using the three-digit ISIC code, Table 5.26 indicates that there is relative specialization across the formal and informal sectors. The largest share of informal activity is in retail trade not in stores (ISIC 525 at 39 percent) followed by non-specialized trade in stores (ISIC 521 at 19.5 percent) and food, beverage and tobacco sales in stores (ISIC 522 at 16.5 percent). Formal sector activity is spread somewhat more evenly across the trade sector with four of the sub-sectors having at

least 15 percent of the firms. These subsectors include food beverage and tobacco specialty stores (ISIC 522 at 33.2 percent), non specialized retail (ISIC 521 22.9 percent), retail trade not in stores (ISIC 525 at 16.5 percent), and other specialized retail trade of new items (ISIC 523 at 15.0 percent).

While the formal and informal sectors have identical share of firms in trade, the informal sector has a greater concentration of businesses in manufacturing (ISIC D at 20.4 percent versus 10.4 percent for formal) and transport storage and communications (ISIC I at 11.4 percent versus 5.3 percent for formal). Formal sector activity is substantially higher in the hotel and restaurant sector (ISIC H at 4.5 percent for informal, 10.8 percent formal). Following the tables on proportions of activity is Table 5.27 with the number of businesses by ISIC at the national level.

Self-employment sector household distribution - across sectors

ISIC	informal	inf_herd	formal	Total
Ag	0.0441	1	0.0616	0.5145
Fishing	5.30E-04	0	0	2.40E-04
Mining	0.0092	0	0.0039	0.0044
Manufact	0.2044	0	0.1037	0.0987
Elec_gas	0.0019	0	0.0023	9.70E-04
Construc	0.0219	0	0.0216	0.0111
Trade	0.4994	0	0.5004	0.2542
Hotel_Re	0.0449	0	0.1077	0.0261
Transpor	0.114	0	0.0527	0.0548
Finance	0.0031	0	0.0296	0.003
Real_Est	0.0106	0	0.0199	0.0059
Educ	0.0021	0	0.0138	0.0017
Health	0.0042	0	0.0389	0.004
Other_Co	0.0382	0	0.0439	0.0197
Priv_hou	0.0016	0	0	7.30E-04
Total	1	1	1	1

Key: column proportions

Self-employment sector household distribution - within ISIC sector

ISIC	linformal	inf_herd	formal	Total
Ag	0.0391	0.9546	0.0062	1
Fishing	1	0	0	1
Mining	0.9541	0	0.0459	1
Manufact	0.9453	0	0.0547	1
Elec_gas	0.8795	0	0.1205	1
Construc	0.899	0	0.101	1
Trade	0.8975	0	0.1025	1
Hotel_Re	0.785	0	0.215	1
Transpor	0.9499	0	0.0501	1
Finance	0.4786	0	0.5214	1
Real_Est	0.8232	0	0.1768	1
Educ	0.5735	0	0.4265	1
Health	0.4889	0	0.5111	1
Other_Co	0.8841	0	0.1159	1
Priv_hou	1	0	0	1
Total	0.4567	0.4912	0.0521	1

Key: row proportions

ISIC3	informal	formal	Total	ISIC	informal	inf_herd	formal	Total
501	0.0036	0.0118	0.0044	Ag	5,756	140,383	917	147,056
502	0.033	0.0214	0.0318	Fishing	69			69
503	0.0054	0.0097	0.0059	Mining	1,197		58	1,254
504	0.0034	0.0036	0.0034	Manufact	26,679		1,543	28,222
505	0.0024	0.0182	0.004	Elec_gas_water	245		34	278
511	2.20E-04	0	2.00E-04	Construction	2,862		322	3,184
512	0.0609	0.0319	0.0579	Trade	65,191		7,449	72,640
513	0.0037	0	0.0033	Hotel_Rest	5,855		1,604	7,459
514	0.0047	0.0028	0.0045	Transport_Storage	14,880		785	15,665
515	0.003	0.0041	0.0031	Finance	404		440	845
519	0.0031	0	0.0027	Real_Estate	1,380		296	1,676
521	0.1947	0.2289	0.1982	Educ	277		206	483
522	0.1649	0.3327	0.1821	Health	554		580	1,134
523	0.0766	0.1505	0.0842	Other_Comm_				
524	0.0178	0.0071	0.0167	Soc_Pers	4,982		653	5,635
525	0.3853	0.1645	0.3626	Priv_households	208			208
526	0.0373	0.0126	0.0348					
Total	1	1	1	Total	130,539	140,383	14,886	285,808

Key: column proportions

Firm size

As shown in Table 5.28, formal sector businesses are larger than both animal herding and other informal sector businesses. The figures below include the business owner.

status	Number of employees
informal	2.30
inf_herd	2.91
formal	4.86
Total	2.74

Informal Mean	Estimate	Std. Err.	[95% Conf. Interval]		Deff
Household wage employees	0.033711	0.0045782	0.0247362	0.0426861	1.219944
Non-household wage empl.	0.200954	0.0124409	0.1765651	0.225343	1.059259
Unpaid household workers	0.421403	0.0170331	0.3880119	0.454795	1.465887
Av. hrs employee worked	44.58037	0.7320086	43.1448	46.01593	1.542626

(*) Some variables contain missing values.

Herding Mean	Estimate	Std. Err.	[95% Conf. Interval]		Deff
Household wage employees	0.007388	0.0035662	0.0003902	0.0143851	1.054602
Non-household wage empl.	0.048415	0.0083038	0.0321212	0.0647077	0.7777277
Unpaid household workers	1.116973	0.0528095	1.013353	1.220593	1.507635
Av. hrs employee worked	35.29553	1.882405	31.59908	38.99198	1.889963

(*) Some variables contain missing values.

Formal Mean	Estimate	Std. Err.	[95% Conf. Interval]		Deff
Household wage employees	0.076218	0.0170619	0.042714	0.1097223	1.566104
Non-household wage empl.	1.99694	0.2273022	1.550591	2.44329	0.9472176
Unpaid household workers	0.467783	0.0837656	0.3032942	0.6322725	0.8699874
Av. hrs employee worked	49.49582	1.312898	46.91419	52.07745	1.143062

Payment Method

Table 5.29 provides a window into household employee remuneration. Wage payments to household members are rare across all business statuses, with wage employees most prevalent in the formal sector as expected.³¹ Unpaid family workers are employed at roughly the same rate in the formal

and the non-herding informal sector, however among these informal businesses the unpaid family worker is the predominant type of employee. Employees work more and are better paid in the formal sector. Average hours worked by employees in the week prior to the interview rises with formality (q98). Animal

³¹ Here, the number .03, for example, means three households in one hundred have a household employee of the particular type.

herder employees worked 36 hours, informal sector workers 46 hours, and formal sector employees averaged 50 hours. Our estimate of the marginal product of labor is presented in Table 5.30 as the amount required to hire a worker for one additional

hour. This ranges from a negligible amount (47 togrogs) among the herders, to 387 togrogs among informals, and 666 togrogs per hour among the formal sector businesses.

Cost of extra hr.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
informal	386.54	33.77216	320.2695	452.8118	1.05095
inf_herd	47.33	25.31557	-2.343736	97.00982	2.794927
formal	666.18	137.6798	396.0176	936.356	1.110463

Intensity of activity

The difference in the amount of time spent in the formal and informal sector business activity by the owners shows little variation, although both groups do report significant breaks from business activity of between two and three months per year. Animal herders report the largest number of days per month (29 days, September report) and the most months per year of activity (11.5 months).

status	Days worked in September	Months work per year
informal	23.06	9.15
inf_herd	28.85	11.55
formal	25.46	9.95
Total	26.03	10.37

Association activity

Collective action in the private sector is often abetted through business associations. Associational activity is low in Mongolia, with more than 98 percent reporting that they are uninvolved in business associations. The formal sector shows significantly higher activity, with 8 percent of businesses in associations, and less than 2 percent in the informal sector. There is some variation by sector regarding strength of association. For example, in the informal sector both financial service providers and educators have relatively high rates of association (16 percent and 9 percent). In the formal sector, mining (38 percent), construction (29 percent), manufacturing (18 percent), and real estate (18 percent) have the highest rates of associational membership. Among those who do participate in associations, a majority (67 percent) pay association fees. There is variation across formality in the rate at which businesses pay for associational memberships, 22 percent in the herding sector, 69 percent among other informals, and 88 percent in the formal sector. See Table 5.32 and Table 5.33 for additional details.

status	Member of Association		
	Yes	No	Total
informal	0.0158	0.9842	1
inf_herd	0.0067	0.9933	1
formal	0.08	0.92	1
Total	0.0147	0.9853	1

Key: row proportions

status	Participated/Paid fee		
	Yes	No	Total
informal	0.6913	0.3087	1
inf_herd	0.2157	0.7843	1
formal	0.8807	0.1193	1
Total	0.6401	0.3599	1

Key: row proportions

<u>Non-herding Informal</u>				<u>Formal</u>			
Member of Association				Member of Association			
ISIC	Yes	No	Total	ISIC	Yes	No	Total
Ag	0.009	0.991	1	Ag	0.0508	0.9492	1
Fishing	0	1	1	Mining	0.381	0.619	1
Mining	0	1	1	Manufact	0.1818	0.8182	1
Manufact	0.0069	0.9931	1	Elec_gas	0	1	1
Elec_gas	0	1	1	Construc	0.2924	0.7076	1
Construc	0	1	1	Trade	0.0347	0.9653	1
Trade	0.0174	0.9826	1	Hotel_Re	0.0994	0.9006	1
Hotel_Re	0.0169	0.9831	1	Transpor	0.0592	0.9408	1
Transpor	0.0159	0.9841	1	Finance	0.1407	0.8593	1
Finance	0.1551	0.8449	1	Real_Est	0.1846	0.8154	1
Real_Est	0.0106	0.9894	1	Educ	0	1	1
Educ	0.0925	0.9075	1	Health	0.1637	0.8363	1
Health	0.0246	0.9754	1	Other_Co	0.1114	0.8886	1
Other_Co	0.0419	0.9581	1	Total	0.08	0.92	1
Priv_hou	0.1423	0.8577	1				
Total	0.0158	0.9842	1				

Key: row proportions

Business formality — recordkeeping, workplace permanence, ownership type

Our results so far have made use of definitions of formality associated with registration as legal entities. The adoption of business practices typically associated with formal sector activity provides another measure to assess the informal sector. Written accounts, permanence in the place of business, and ownership type are three ways we examine both within and across our registration definitions of formality to shed light on the nature of informal sector shadow activity. (See Table 5.34, Table 5.35, and Table 5.36 for details)

Formal firms are more than twice as likely to have written accounts (77 percent versus 36 percent). Forty-nine percent of the informals in the trade sector keep written accounts as do 86 and 59 percent in the finance and restaurant/hotel sectors. Among the larger informal sectors written records are scarce in manufacturing (20 percent), transport, storage and communications (11 percent), and construction (9 percent).

Business workplace type is another area we might expect differences across the formal and informal sector, and we have already reported some evidence of this with regard to trade, with 39 percent of informal trading firms outside of regular store location (ISIC 525). Interestingly, about the same proportion of businesses are based in the home in the formal (30 percent) and informal sectors (36 percent). When the workplace is outside the home however there is consistently more variability in workplace type in the informal sector, with 28 percent of informal businesses in a fixed shop as compared to 52 percent of formal sector firms. Informal sector businesses are roughly 3 times more likely to be located in market stalls, motor vehicles, or other variable locations than formal sector businesses.

With regards to ownership type the informal and formal sectors do not differ greatly with 64 percent and 60 percent reporting sole ownership respectively. The herding sector reports only 43 percent sole ownership and 57 percent family partnership.

Question93

<u>status</u>	<u>Sole own</u>	<u>Partner in family</u>	<u>Partner non-family</u>	<u>Total</u>
informal	0.6363	0.3017	0.062	1
inf_herd	0.4314	0.5646	0.0041	1
formal	0.611	0.2925	0.0965	1
Total	0.5343	0.4304	0.0353	1

Key: row proportions

<u>status</u>	<u>Yes</u>	<u>No</u>	<u>Total</u>
informal	0.3578	0.6422	1
inf_herd	0.3238	0.6762	1
formal	0.7742	0.2258	1
Total	0.3628	0.6372	1

Key: row proportions

<u>Question106</u>			
<u>ISIC</u>	<u>Yes</u>	<u>No</u>	<u>Total</u>
Ag	0.1327	0.8673	1
Fishing	0	1	1
Mining	0.0638	0.9362	1
Manufact	0.1958	0.8042	1
Elec_gas	0.5779	0.4221	1
Construc	0.0905	0.9095	1
Худалдаа	0.4941	0.5059	1
Hotel_Re	0.591	0.409	1
Transpor	0.1092	0.8908	1
Finance	0.8638	0.1362	1
Real_Est	0.3627	0.6373	1
Educ	0.427	0.573	1
Health	0.5548	0.4452	1
Other_Co	0.3312	0.6688	1
Priv_hou	0.1324	0.8676	1
Total	0.3578	0.6422	1

Key: row proportions

All self-employment

<u>status</u>	<u>Shop/plant at home</u>	<u>Shop/plant not at home</u>	<u>Market stall</u>	<u>Location varies</u>	<u>Motor vehicle</u>	<u>Grass</u>	<u>Other</u>	<u>Total</u>
informal	0.3554	0.2765	0.1195	0.1229	0.0875	0.0369	0.0014	1
inf_herd	0.0728	0.0698	5.70E-04	0.0638	0.0014	0.7917	0	1
formal	0.2961	0.5173	0.0377	0.0565	0.0331	0.0574	0.0018	1
Total	0.2135	0.1875	0.0568	0.0904	0.0424	0.4087	7.10E-04	1

Key: row proportions

Finance

One of the key ingredients of business success is its access to financing. For the present purposes we consider the sources and start-up financing and the informal business's experience with borrowing. The sources of start-up capital are given in Table 5.37. As seen, 64 percent (for formal small businesses) to 88 percent (herders) come from own (or family) sources. The rest is spread out among the other possible

sources, with banks being about as important as moneylenders or family friends. Nonetheless, Table 5.40 suggests that only about one-eighth of informal businesses (13 percent) have ever borrowed, compared to more than a quarter (27 percent) for formal small businesses. Perhaps the good news is that the table suggests a strong increase in borrowing for those who have engaged in it.

status	Family or own saving	Friend	Moneylender	Cooperative	Bank	Other	Total
informal	0.7591	0.0702	0.0613	0.0136	0.0725	0.0233	1
inf_herd	0.8813	0.0134	5.60E-04	0.0049	0.0093	0.0906	1
formal	0.6373	0.0639	0.1029	0.0347	0.1431	0.0181	1
Total	0.8128	0.042	0.0336	0.0104	0.0452	0.0561	1

Key: row proportions

Bank loan use is more widespread for continuing investment than for startup capital. Among the 9 percent of firms that have borrowed in the last 12 months, banks are the predominant source at 63 percent followed by friends/family and moneylenders

(15 percent each). The herding sector is the most likely to have made use of bank lending, with 86 percent, citing banks as a source of funds. This compares to 54 percent among the informal firms and 71 percent in the formal sector.

status	Family/f	Cooperative	Pawnshop	Bank	Supplier	Customer	Moneylender	Other	Total
informal	0.1969	0.0464	0.0206	0.539	0.0114	0.0058	0.1731	0.0068	1
inf_herd	0.0131	0.0077	0	0.8644	0	0	0.1051	0.0096	1
formal	0.1571	0.0263	0.0083	0.7068	0.0055	0	0.0905	0.0055	1
Total	0.1523	0.0354	0.0145	0.6318	0.0082	0.0037	0.1469	0.0072	1

Key: row proportions

The source of finance, both for startup and continuing operations varies considerably by the age of the business. As demonstrated in Table 5.39, older firms were more likely to get startup capital from family and friends, but are more likely at present to use bank financing. These facts along with some encouraging trends in collateral requirements, discussed below, suggest that there have been positive developments

in financing self-employed businesses in recent years. It is important to remember however that those borrowing for purposes other than startup constitute a small portion of the self-employed firms. We discuss the importance of credit constraints on the remaining firms, below in Section 5.3 on the social and economic consequences of informal activity.

Question 141

Age of Business	Family	Friend	Moneylender	Cooperative	Bank	Other	Total
1.5	0.7185	0.0733	0.0842	0.0194	0.0721	0.0326	1
3	0.7578	0.0706	0.0713	0.0087	0.071	0.0206	1
5	0.7933	0.0505	0.0501	0.0174	0.0612	0.0275	1
9	0.8282	0.0397	0.0272	0.0149	0.0519	0.0381	1
80	0.8561	0.0226	0.0052	0.0035	0.0213	0.0913	1
Total	0.8128	0.0419	0.0336	0.0104	0.0452	0.0561	1

Key: row proportions

Question 146

Age of business	Family or friend	Cooperative	Pawnshop	Bank	Supplier	Customer	Moneylender	Other	Total
1.5	.2137	.0618	.0257	.5248	.0059	.0039	.1558	.0085	1
3.0	.1696	.0298	.0318	.6418	0	0	.1271	0.0	1
5.0	.1442	.0353	.011	.5347	.0175	.0152	.2392	.0029	1
9.0	.1735	.0347	.0121	.615	.0082	0	.1485	.008	1
80	.0942	.0242	.003	.7811	.0067	0	.0775	.0134	1
Total	.1525	.0354	.0145	.6314	.0082	.0037	.1471	.0072	1

Key: row proportions

Table 5.41 presents loan size as a percent of collateral requirements so a high number means lower collateral requirements. These figures suggest that collateral requirements are on the order of 2 to 3 times the size of the loan. While this is common for countries with commercial law institutions at the level of development of Mongolia's, it still places a inordinate

risk on the business borrower and points to the failure of the financial system to fulfill its role as an intermediary for diversification and spreading of risk. Also note that collateral requirements appear to have risen in the formal sector, but show either a decline or stability in the non-herding informal sector.

status	Never	Yes, before 2001	Yes, in 2001/2	Yes, in 2003/4	Total
informal	0.8105	0.0303	0.025	0.1341	1
inf_herd	0.891	0.0194	0.0146	0.075	1
formal	0.6544	0.0321	0.0487	0.2648	1
Total	0.8442	0.0249	0.0208	0.1101	1

Key: row proportions

status	Year of borrowing		
	before 2001	2001/2002	2003/2004
inf_n_herd	43.7387	32.6525	40.9475
inf_herd	39.0447	35.2796	35.3053
formal	62.5321	41.2808	36.9984
Total	41.8424	34.4225	38.6747

Table 5.42 indicates that within the informal sector collateral requirements fell in almost all sectors where we can observe activity from 2001 to 2004. The sole exceptions were transport, storage and communications (9), and other community, social, and personal activities (15). The highest collateral requirements in the informal sector (lowest number in Table 5.42) are associated with mining, although we do not have time series data to identify trends in this sector.

ISIC	before 2001	2001/2002	2003/2004
Ag		33.0865	40.5803
Mining			26.1559
Manufact	34.616	29.5016	44.2032
Construction	30	30	39.6555
Trade	26.9576	32.7285	41.483
Hotel_Rest	50	31.0859	43.6439
Transport_Storage		41.4979	37.8039
Finance		45	57.8593
Real_Estate			37.4069
Educ			
Health			50.4558
Other_Comm_Soc_Pers		39.9917	39.4215
Total	32.2665	33.3835	41.7037

Key: row proportions

Dispute resolution

Table 5.43 and Table 5.44 allow us to investigate the informal sector's use of the country's legal infrastructure and institutions. We see in Table 5.43 that a large proportion of Mongolians (83 percent) exhibit a willingness to make use of the court system if other dispute resolution mechanisms have failed to achieve a result. Table 5.44 shows that this willingness is practically the same (79 percent) for those who have previously gone to court to resolve a dispute, although the skepticism about the efficacy of the

judicial system increases for this group. For example, 8 percent of those who have gone to court are concerned that their opposing party could bribe the court. Only 2 percent of those who have not gone to court share this concern. In addition there is more generalized mistrust of the court's ability to reliably serve justice among those with experience with 6 percent expressing this as a reason not to enter a claim. Only 1.5 percent of those without court experience share this view.

status	Yes	No, because...					Total
		Defendant could bribe	Court takes too long	it costs too much	courts too unreliable	Don't know how	
Informal	0.828	0.0274	0.018	0.0392	0.0263	0.0612	1
Inf_herd	0.8307	0.0183	0.0187	0.0162	0.0059	0.1102	1
Formal	0.8756	0.0117	0.0379	0.0256	0.0376	0.0116	1
Total	0.8318	0.0221	0.0194	0.0272	0.0168	0.0827	1

Key: row proportions

Past use of courts	Yes	No, because...					Total
		Defendant could bribe	Court takes too long	it costs too much	courts too unreliable	Don't know how	
Yes	0.7927	0.0772	0.0195	0.0338	0.0603	0.0165	1
No	0.8336	0.0195	0.0194	0.0269	0.0148	0.0858	1
Total	0.8318	0.0221	0.0194	0.0272	0.0168	0.0827	1

Key: row proportions

To understand the causes of shadow activity the interaction of personal and business characteristics and beliefs with the business environment are examined. We find that both perceptions of noncompliance and admissions of noncompliant behavior are fairly substantial. Further, a large majority of respondents believe that punishment for noncompliance will not be severe. Among the small group having had experience with the judicial system for any reason, compliance with business registration requirements is dramatically increased. The data suggests, also, that registration in the formal sector is driven, at least in part, by a desire to maintain banking relationships. To the extent that financing is increasingly available to informal sector businesses, this motivation to register may be reduced. Our discussion of results, below, focuses on three areas: beliefs regarding government enforcement of registration requirements, the impact of business

regulations, and beliefs and behaviors regarding corrupt activities including tax evasion. As will be seen, there appears to be much ignorance regarding compliance processes and consequences as well as considerable regional variation in compliance and beliefs.

Registration requirements

The business environment is shaped first by legal processes associated with registration of business entities. In addition to formal sector registration of corporations, we consider beliefs and behaviors associated with compliance with the Informal Sector Tax Law, also known as the *Patente* Registry. Note that the questions on beliefs were asked of all self-employed respondents, so that informal sector actors responded to questions regarding the implications of noncompliance with formal sector requirements, and the reverse was also the case.

Perceptions regarding the consequences of

failing to register in the formal sector differ moderately across the different statuses of businesses as shown in Table 5.45. Ten percent of formal sector respondents believe that there would be no negative consequences for failing to register, while 19 percent

of informal firms believe this is the case. While many have an unclear idea of what the negative consequences would be, the one of most concern, particularly to formal sector respondents, is the loss of ability to borrow funds.

status	1	2	3	4	5	6	7	8	9	10	Total
Informal	.1952	.3843	.2565	.0868	.0223	.0202	.0034	.0235	.0057	.0021	1
InfHerd	.1715	.5911	.1613	.0292	.0062	.0087	.0000	.0308	.0012	.0000	1
Formal	.0894	.1350	.4098	.2039	.0641	.0474	.0073	.0222	.0081	.0128	1
Total	.1784	.4539	.2270	.0705	.0184	.0172	.0023	.0264	.0040	.0019	1

Legend:

1 = None

2 = Don't know

3 = Don't know, but there would probably be other consequences with a negative impact for running my business

4 = It would be harder to get a loan from a bank or credit cooperative

5 = I would not be able to use the court system to resolve business disputes

6 = I would receive a lower level of police protection

7 = It would be harder to import or export internationally

8 = I would not be eligible to benefit from any government programs

9 = It would be harder to receive municipal services such as access to piped water, gas, electricity, and garbage collection

10 = Other

Key: row proportions

Perceived penalties of discovered non-compliance are shown in Table 5.46. Nine percent overall, and 11 percent in the formal sector believe that the enforcement mechanisms and penalties implemented by the government will have serious consequences for those failing to register for the formal sector. Fifty-two percent of those in the formal sector believe that they would be responsible for back taxes if discovered (a combination of choices 5 and

6), while only 25 percent and 9 percent in the informal and herding sector believe this is the case. In addition to paying back taxes, a much larger percentage of those in the formal sector believe that time lost interacting with the government during enforcement activities would be a significant consequence. A very small proportion (less than 1 percent) thought that time in jail was a likely consequence.

status	None or could bribe	Minor	Serious	Jail	Waste time/ pay back tax	Forced to Register	Other	Total
informal	0.0346	0.5936	0.0982	0.006	0.0548	0.1931	0.02	1
inf_herd	0.0766	0.7354	0.0759	0.008	0.0172	0.0702	0.0172	1
formal	0.0168	0.3221	0.1146	0.006	0.1587	0.3601	0.0223	1
Total	0.0543	0.6491	0.0881	0.007	0.0417	0.1414	0.0188	1

Key: row proportions

Beliefs regarding the probability of discovery of non-compliance with the Informal Sector Law also vary by employment status, as shown in Table 5.47. The proportion of those believing that their evasion of this law would probably or definitely be discovered varies from a low of 43 percent for the herders to 86 percent among formal sector workers. Sixty-eight percent of informal sector business workers share this belief. Among this group of informal sector

business owners we also present results on location (Table 5.48) and regional variability (Table 5.49) in the belief regarding detection of *patente* law noncompliance. Belief in the likelihood of detection is relatively low in rural areas, and extremely low in the western region where only 21 percent believe noncompliance would probably or definitely be discovered. By contrast, 81 percent share this belief in the Central region.

status	Probably No	Don't Not	know	Probably	Definitely	Total
informal	0.0498	0.044	0.2353	0.2112	0.4597	1
inf_herd	0.0398	0.1804	0.3554	0.1091	0.3153	1
formal	0.0314	0.0167	0.0843	0.146	0.7217	1
Total	0.0439	0.1096	0.2864	0.1576	0.4024	1

Key: row proportions

Location	No	Probably Not	Don't know	Probably	Definitely	Total
Ulaanbaa	0.0397	0.0437	0.2091	0.2564	0.4512	1
Aimag ce	0.0481	0.0403	0.2189	0.1919	0.5008	1
Soum cen	0.0615	0.0499	0.2375	0.1246	0.5264	1
Rural	0.04	0.1843	0.3637	0.1019	0.31	1
Total	0.0439	0.1096	0.2864	0.1576	0.4024	1

Key: row proportions

Region	No	Probably Not	Don't know	Probably	Definitely	Total
West	0.0642	0.4128	0.3154	0.0912	0.1164	1
Khangai	0.054	0.0741	0.4385	0.0769	0.3564	1
Central	0.0213	0.0092	0.1577	0.1981	0.6136	1
East	0.0277	0.0237	0.1933	0.2061	0.5492	1
Ulaanbaa	0.0397	0.0437	0.2091	0.2564	0.4512	1
Total	0.0439	0.1096	0.2864	0.1576	0.4024	1

Key: row proportions

With regard to the consequences associated with noncompliance with the Informal Sector Law, we see in Table 5.50 that the results are broadly similar to those for the business registry. There is much ignorance about the consequences of noncompliance, though a large proportion believes there will be some unspecified negative consequence. As previously, those in the formal sector express much less ignorance about potential consequences.

With regard to the penalties associated with noncompliance, the results shown in Table 5.51 are broadly similar with those associated with formal sector registry. Again the formal sector firms see the costs associated with time spent dealing with the government registration process as a much more severe consequence than do those in the informal sector.

status	1	2	3	4	5	6	7	8	9	10	Total
Informal	.1121	.3744	.3241	.0787	.0241	.0265	.0037	.0432	.0115	.0019	1
inf_herd	.1672	.5085	.2029	.0358	.0033	.0066	.0000	.0752	.0000	.0000	1
formal	.0651	.1807	.3862	.1554	.0836	.0467	.0144	.0452	.0167	.0060	1
Total	.1367	.4302	.2678	.0617	.0170	.0178	.0024	.0590	.0063	.0012	1

Legend:

1 = None

2 = Don't know

3 = Don't know, but there would probably be other consequences with a negative impact for running my business

4 = It would be harder to get a loan from a bank or credit cooperative

5 = I would not be able to use the court system to resolve business disputes

6 = I would receive a lower level of police protection

7 = It would be harder to import or export internationally

8 = I would not be eligible to benefit from any government programs

9 = It would be harder to receive municipal services such as access to piped water, gas, electricity, and garbage collection

10 =Other

Key: row proportions

status	None or could bribe	Minor	Serious	Jail	Waste time/ pay back tax	Forced to Register	Other	Total
informal	.0346	.5936	.0982	.0058	.0548	.1931	.0200	1
inf_herd	.0766	.7354	.0759	.0076	.0172	.0702	.0172	1
formal	.0168	.3221	.1146	.0055	.1587	.3601	.0223	1
Total	.0543	.6491	.0881	.0067	.0417	.1414	.0188	1

Key: row proportions

The next set of results compares beliefs with behaviors. To do this we restrict attention to those who believed they were required to register under the Informal Sector Law and divides the sample

according to whether the respondent acknowledged compliance³². As one would expect, Table 5.52 indicates that a larger share of those in compliance thought that non-compliance would definitely be

³² All respondents were in either Treatment 2 or 4 in which the envelope technique made their answers to the Question 161 observable.

detected (52 percent versus 40 percent of those who were not registered). An even larger difference between the two groups—and one with more implications for policy—is the much larger admission of ignorance with regard to both probability of detection and the consequences of non-registration. Across these dimensions 40 percent of the noncompliant group admitted ignorance regarding the probability of detection, compared to 20 percent of

the compliant informal sector responders. Table 5.53 indicates relatively similar views regarding the consequences, with large shares of both compliant and non-compliant respondents admitting ignorance about penalties. Perhaps most damning is the evidence in Table 5.54, which suggests that *even if a patente* tax were owed, two-thirds of the respondents thought that little would happen if it weren't paid—and this result is independent of whether the respondent is paying her tax.

Did respondent pay tax?						Total
	No	Probably Not	Don't know	Probably	Definitely	
No	0.0266	0.0145	0.3996	0.1638	0.3955	1
Yes	0.0152	0.0138	0.1946	0.2581	0.5183	1
Total	0.0204	0.0141	0.2887	0.2149	0.462	1

Key: row proportions

Did respondent pay tax?										Total
	1	2	3	4	5	6	7	8	9	
No	.0899	.4843	.2182	.0924	.0142	.0152	.0000	.0759	.0100	1
Yes	.1100	.3671	.3608	.0685	.0232	.0393	.0023	.0288	.0000	1
Total	.1007	.4211	.2951	.0795	.0191	.0282	.0012	.0504	.0046	1

Legend:

1 = None

2 = Don't know

3 = Don't know, but there would probably be other consequences with a negative impact for running my business

4 = It would be harder to get a loan from a bank or credit cooperative

5 = I would not be able to use the court system to resolve business disputes

6 = I would receive a lower level of police protection

7 = It would be harder to import or export internationally

8 = I would not be eligible to benefit from any government programs

9 = It would be harder to receive municipal services such as access to piped water, gas, electricity, and garbage collection

Key: row proportions

Did re- spondent pay tax?	None or could bribe	Minor	Serious	Jail	Waste time/ pay back tax	Forced to Register	Other	Total
No	0.0033	0.6672	0.056	0.0383	0.0493	0.1789	0.0071	1
Yes	0.0159	0.6701	0.0881	0.0126	0.0225	0.1907	0	1
Total	0.0101	0.6688	0.0733	0.0244	0.0349	0.1853	0.0033	1

Key: row proportions

Regulation

Table 5.55 suggests that the perception that regulation is a burden increases with the formality of the enterprise. Only 2 percent of herders are affected, but 18 percent of the informal sector and 34 percent of the formal sector feel that governmental regulations, procedures, and other rules are obstacles for their business operation. The perception of the size of the problem is similar in the informal and formal sectors; with the affected groups estimating, on average, that profits would be 24 percent higher without the obstacles.

status	Yes	No	Total
informal	0.1819	0.8181	1
inf_herd	0.0226	0.9774	1
formal	0.3359	0.6641	1
Total	0.1116	0.8884	1

Key: row proportions

status	Trade	Labor	Sector- specific	Operating /Construc- tion permit	Business registry permit	Real estate transact.	Environ- ment	Other	Total
informal	.1803	.1802	.2386	.1052	.1331	.0477	.0648	.0500	1
formal	.0988	.1128	.4072	.1255	.1732	.0106	.0185	.0534	1
Total	.1661	.1685	.2680	.1087	.1401	.0412	.0568	.0506	1

Key: row proportions

The distribution of regulatory concerns varies across the formal and informal sectors, as seen in Table 5.56. While sector specific regulations are paramount for both groups, they are the primary concern for 41 percent in the formal sector but only 24 percent in the informal sector. After the sector

specific regulations the primary concern in the informal sector is for customs and foreign trade regulations and labor regulations (18 percent each). Formal sector firms are more concerned with permits to open a business (17 percent) and permits for construction and business operation (13 percent).

Given the importance of sector specific regulations, in Table 5.57 we investigate by ISIC which sectors report regulatory obstacles, taking particular notice of sectors in which obstacles are high in the formal sector and low in the informal. The sectors which show the greatest degree of difference in this regard are manufacturing and construction, with 8 percent of informal and 31 percent of formal manufacturers reporting regulatory obstacles. In the construction sector, again 8 percent of the informal

firms report regulatory obstacles, while 55 percent of formal sector firms do. Trade and transportation are two other sectors with large informal components that share this pattern of higher obstacles in the formal sector although the differences are not as great. In the trade sector 23 percent of informal enterprises report obstacles and 34 percent of formal enterprises do. In the transportation sector 22 percent of the informal enterprises report problems and 40 percent of the formal enterprises do.

Non-herding Informal Sector

ISIC	Yes	No	Total
Ag	.1660	.8340	1
Fishing	0	1	1
Mining	.2105	.7895	1
Manufact	.0762	.9238	1
Elec_gas	.0509	.9491	1
Construc	.0753	.9247	1
Trade	.2246	.7754	1
Hotel_Re	.1994	.8006	1
Transpor	.2178	.7822	1
Finance	.2820	.7180	1
Real_Est	.1211	.8789	1
Educ	.1715	.8285	1
Health	.2064	.7936	1
Other_Co	.1567	.8433	1
Priv_hou	0	1	1
Total	.1819	.8181	1

Key: row proportions

Formal Sector

ISIC	Yes	No	Total
Ag	.1091	.8909	1
Mining	.3810	.6190	1
Manufact	.3068	.6932	1
Elec_gas	0	1	1
Construc	.5499	.4501	1
Trade	.3346	.6654	1
Hotel_Re	.4298	.5702	1
Transpor	.4030	.5970	1
Finance	.2941	.7059	1
Real_Est	.3804	.6196	1
Educ	.4011	.5989	1
Health	.3959	.6041	1
Other_Co	.2726	.7274	1
Total	.3359	.6641	1

Key: row proportions

Corruption and tax evasion: perception and behaviors³³

Large proportions of the population believe that corruption (Table 5.58) and tax evasion (Table 5.59) are prevalent in Mongolia. Herders differ greatly from the rest of the population perceiving much less corruption, although even in this group more than half (53 percent) believe corruption is considerable or widespread. The proportion among other self-

employed business owners and wage earners is about 80 percent, however. The unemployed are between these extremes with 66 percent reporting beliefs that high levels of corruption exist. Differences in perceptions across groups regarding tax evasion are similar, although the perceived levels are lower than for corruption, averaging 50 percent and ranging from 30 percent (herders) to 64 percent (formal sector business owners).

³³ Note that this section includes responses from wage earners and the unemployed whenever possible. These groups were excluded from previous questions on the business environment.

status2	None	Little	Somewhat	Considerable	Widespread	Total
informal	0.019	0.0862	0.1079	0.2177	0.5693	1
inf_herd	0.1266	0.1763	0.1749	0.1933	0.3288	1
formal	0.0364	0.0627	0.0564	0.2481	0.5964	1
wage	0.0228	0.0821	0.1005	0.3094	0.4852	1
unemp	0.065	0.1328	0.1443	0.2354	0.4225	1
Total	0.0554	0.1148	0.1272	0.2449	0.4577	1

Key: row proportions

status2	Not at all	Don't think so	Not sure	Think so	Strongly think so	Total
informal	0.0615	0.1223	0.2211	0.322	0.273	1
inf_herd	0.2471	0.1707	0.2796	0.202	0.1007	1
formal	0.065	0.1396	0.1554	0.3537	0.2863	1
wage	0.0584	0.1216	0.2117	0.3818	0.2265	1
unemp	0.1528	0.1398	0.2855	0.2698	0.1522	1
Total	0.1226	0.1376	0.2423	0.3031	0.1945	1

Key: row proportions

Corrupt behaviors and evasion are highly correlated with the respondents' views on the impact of regulation, as seen in the panels of Table 5.60. For customs bribes (q158) the rate of activity is more than four times greater among those who acknowledge that regulatory burdens are onerous (39 percent versus 9 percent). Bribe-paying in general (q162) shows a similar effect with 32 percent admitting bribes among those who find regulations an obstacle compared to 19 percent among those who

do not. A follow-up question to the bribery question on "gift-giving" shows the same pattern and also a larger proportion admitting this behavior (27 percent and 8 percent). Tax compliance is an anomaly with a slightly larger share of compliance among those who acknowledge regulatory obstacles (70 percent versus 62 percent). These results are take from the two-question method treatments, which showed the greatest admission of corrupt behaviors.

Are regulations Onerous?	Have you paid a bribe to reduce customs taxes?		
	No	Yes	Total
Yes	0.6058	0.3942	1
No	0.9121	0.0879	1
Total	0.8559	0.1441	1

Key: row proportions

Are regulations Onerous?	Have you given a gift to government official?		
	No	Yes	Total
Yes	0.7432	0.2568	1
No	0.9215	0.0785	1
Total	0.8894	0.1106	1

Key: row proportions

Are regulations Onerous?	Have you paid a bribe to a government official?		
	No	Yes	Total
Yes	0.68	0.32	1
No	0.8057	0.1943	1
Total	0.7856	0.2134	1

Key: row proportions

We have observed some of the impacts that participation in the informal sector shadow can have when identifying business characteristics in Section 5.1. In general those in the informal sector are smaller firms, with less productive workers, and significantly fewer concerns with regulatory burdens. In this section we investigate three critical issues associated with participation in the informal sector, associated with finance, well being and poverty, and the durability of informal participation.

Access to finance

Table 5.61 presents a view into the extent to which business owners have access to credit. The large

Are regulations Onerous?	Did you cheat on your taxes last year?		
	No	Yes	Total
Yes	0.3012	0.6988	1
No	0.3809	0.6191	1
Total	0.3665	0.6335	1

Key: row proportions

majority of the population (85 percent) did not use bank loans for investment. The unmet need for loans was greatest in the informal sector (46 percent) with respondents indicating constraints associated with collateral requirements (27 percent), information (10 percent), and interest rates (7 percent) the most pressing concern. The figures presented for collateral are the sum of the responses that collateral requirements are too high, or that the business does not have sufficient collateral. In the herding sector the lack of information dominated the concerns of those (18 percent) who had a need for additional capital. In the formal sector, collateral requirements (20 percent) and interest rates (15 percent) were the main concerns.

Employment status	Never needed to	I don't know how	No col- lateral	Collateral too high	No tax returns	No accounts	High interest	Other	Total
informal	.5440	.1003	.1934	.0820	.0024	.0054	.0687	.0039	1
inf_herd	.7049	.1799	.0635	.0236	.0004	.0025	.0247	.0005	1
formal	.6165	.0269	.0997	.0965	.0000	.0044	.1524	.0035	1
Total	.6325	.1400	.1207	.0513	.0012	.0038	.0483	.0021	1

Key: row proportions

From Table 5.62, the impact of credit constraints on the informal sector can be seen in both the present circumstances and future prospects of constrained individuals. The annualized value added of informal sector businesses who indicate no need for loans exceeds those have no collateral by 20 percent.

Several groups of respondents indicate value added that exceeds those who report being unconstrained. This suggests that there are diverse set of constraints operating for firms of different size. In particular, the group with the highest average value added is constrained by the level of interest rates.

Reason for not Borrowing	Estimate	Std. Err.	[95% Conf. Interval]		Deff
No_need	2462811	537471.6	1409083	3516538	8.393145
I_don't_how	3254682	1066934	1162929	5346435	11.96088
No_collateral	2054669	134423.5	1791128	2318210	1.163809
Collateral high	2195917	766293.2	693578	3698255	1.211481
No_tax_returns	3669469	999229.3	1710453	5628485	1.125073
No_account recs	2805757	832182	1174242	4437273	0.925691
High_interest	3878612	643837.2	2616352	5140872	1.126792
Other	1886879	814698	289641.6	3484117	1.198243

With regard to future prospects, Table 5.63 indicates that 77 percent of the group without credit constraints believes growth prospects for their businesses are positive in the next three years, with 16 percent predicting stagnation and only 8 percent a decline. Those constrained due to lack of collateral or lack of tax records are less optimistic, with 57 and 55

percent believing prospects are positive, respectively. Those without tax records are by far the most pessimistic, with thirty-one percent believing that business growth will be negative in the next three years, a contrast with the consensus estimate of 9 percent.

Reason for not borrowing from bank	Business prospects			
	Decline	Stagnant	Growth	Total
No need	0.077	0.1579	0.765	1
I don't know how	0.0758	0.2331	0.6911	1
No collateral	0.1268	0.3033	0.5698	1
Collateral Too high	0.1638	0.1992	0.6369	1
No tax records	0.3067	0.1467	0.5466	1
No written accounts	0.146	0.0564	0.7976	1
High interest rates	0.092	0.201	0.7069	1
Other	0.3251	0.331	0.3439	1
Total	0.089	0.1904	0.7206	1

Key: row proportions

Constraints due to lack of collateral are particularly severe in the non-herding agriculture sector with 30 percent constrained for this reason. Lack of collateral is also severe in mining and manufacturing sectors where 28 percent and 23 percent cite this issue. The hotel and restaurant sector is also relatively constrained with the lack of collateral (16 percent) and high interest rates (15 percent) the primary causes.

Income, assets and poverty

One of the critical issues associated with informal sector activity is its contribution to the well-being of

household members. We next assess the contribution of the informal sector to household income and to its assets.

First, in Table 5.64 we examine the contribution of the informal sector and compare income levels to the minimum per capita living standards for the country developed by the NSO. The table below presents three numbers for each region and household status. The first is the annual household income, the second the mean monthly household income per person, and the third the NSO minimum living standard per person per month. Notice that the NSO standards vary by region.

region	informal	inf_herd	formal	wage	unemp	Total
West	874,460	486,764	1,166,772	1,207,437	254,037	750,770
	19,797	10,118	21,051	26,240	5,541	16,262
	20,200	20,200	20,200	20,200	20,200	20,200
Khangai	1,019,558	687,135	2,282,220	1,799,410	260,638	1,018,159
	22,542	12,698	47,826	34,072	7,713	20,279
	20,600	20,600	20,600	20,600	20,600	20,600
Central	1,195,244	1,254,313	1,229,450	1,320,473	280,654	1,104,405
	25,028	29,761	24,107	28,089	6,930	24,204
	20,600	20,600	20,600	20,600	20,600	20,600
East	3,845,717	889,479		1,124,935	195,495	1,145,571
	49,350	21,908		23,558	6,131	22,698
	21,200	21,200		21,200	21,200	21,200
Ulaanbaatar	2,042,721	1,947,525	2,866,844	1,011,102	439,836	1,396,081
	43,240	35,164	53,875	23,780	12,057	30,515
	26,500	26,500	26,500	26,500	26,500	26,500
Total	1,576,394	800,459	2,241,811	1,345,623	325,193	1,105,847
	33,010	16,925	43,804	28,163	8,765	23,458
	23,371	20,662	23,093	21,960	22,664	22,118

Note: Income information derives from the Envelope Method of elicitation (see Section 2.2)

The figures in the bottom right corner of the table shows that the mean income per capita does exceed the minimum standard at the national level, though just barely. Reported income is 23,458 togrogs and the weighted national average figure for minimum

living standards is 22,118 togrogs. The unemployed and herding sectors on the whole do not achieve this level, with the unemployed on average at 39 percent and the herding sector at 82 percent of the minimum standard. Wage earners (128 percent), informal

households (141 percent) and formal sector households (190 percent) exceed this threshold. In Table 5.65 the same data sorted by location indicates that while households with formal sector firms exceed

the minimum standards in all locations informal sector firms do so only in the *aimag* centers and in Ulaanbaatar.

Location	informal	inf_herd	formal	wage	unemp	Total
Ulaanbaatar	2,042,721	1,947,525	2,866,844	1,011,102	439,836	1,396,081
	43,240	35,164	53,875	23,780	12,057	30,515
	26,500	26,500	26,500	26,500	26,500	26,500
Aimag center	1,310,331	716,046	2,377,582	2,396,856	381,759	1,495,245
	29,251	14,334	55,802	45,490	10,595	31,011
	20,560	20,772	20,551	20,536	20,566	20,559
Soum center	1,170,220	966,839	418,034	1,117,929	169,920	881,270
	19,575	18,730	7,446	24,560	4,427	18,534
	20,528	20,580	20,554	20,580	20,599	20,577
Rural	313,865	773,728	2,474,339	562,699	345,542	755,568
	7,079	16,582	42,357	9,956	5,157	15,838
	20,397	20,563	20,600	20,494	20,316	20,546
Total	1,576,394	800,459	2,241,811	1,345,623	325,193	1,105,847
	33,010	16,925	43,804	28,163	8,765	23,458
	23,371	20,662	23,093	21,960	22,664	22,118

Note: Income information derives from the Envelope Method of elicitation (see Section 2.2).

Turning to assets, we divide these into two types, animal and plant/equipment (non-animal). Table 5.66 presents the balance sheet, the declared assets (q119) and liabilities (q120) of respondents. Here, again, the Envelope Treatment (*Env*) reveals significantly more assets than the Multiplier Treatment, with the difference more extreme among the herding households. Interestingly, liabilities are an anomaly when considering treatment effects; they

are much higher for the herders in the multiplier treatment. This result implies that one should consider the possibility that strategic responses are driving the difference between envelope and multiplier questions, and that individuals may behave more strategically in the Multiplier Treatment. Animal assets are for both herders and non-herder informal sector businesses in Table 5.67.

Total	Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
Assets						
Envelop	Informal	7.86E+10	1.30E+10	5.30E+10	1.04E+11	0.4233722
Multiplier	Informal	1.20E+11	8.34E+09	1.04E+11	1.36E+11	0.6972588
Envelop	Herding	1.42E+11	4.90E+10	4.59E+10	2.38E+11	6.119897
Multiplier	Herding	1.11E+11	1.70E+10	7.80E+10	1.45E+11	7.372394
Liabilities						
Envelop	Informal	8.88E+09	2.94E+09	3.11E+09	1.47E+10	0.4967668
Multiplier	Informal	1.20E+10	1.48E+09	9.08E+09	1.49E+10	0.7671556
Envelop	Herding	3.22E+08	2.21E+08	-1.12E+08	7.56E+08	5.142009
Multiplier	Herding	4.28E+09	2.07E+09	2.18E+08	8.34E+09	7.741012

Status	Average size	Total for status
informal	13	1.640.144
inf_herd	167	23.171.692
formal	50	746.161
wage	17	3.179.243
unemp	12	1.241.466
Total	53	29.978.706

Foreign remittances

The figures on household income reported above include income from all sources. Here we examine more closely a component of household income—that derived from foreign remittances. Survey estimates imply that only a small number of households (12,497 or 2.2 percent) have income from foreign remittances, but among those who do have this source of income it represents an important share, on average 31 percent of the total. As shown in Table 5.68 remittances are extremely important among the unemployed with recipients reporting the value of foreign remittances as 48 percent of the annual

household income. The large share among the unemployed is not due to larger remittances but to lower household incomes. In fact, the mean value of annual remittances for the unemployed, at 344,850 togrogs is the lowest of all employment statuses. The share of households receiving remittances and the mean value of remittance received is lowest in the herding sector (0.9 percent); however, the herding sector reports the largest value of remittances at 609,380 togrogs. The formal sector has the highest rate of remittance receipt at 5.2 percent, and a value that is second to the herders at 565,487 togrogs. Households receiving remittances on average report income 277,641 togrogs greater than those without remittances, a difference that is not statistically significant.

Table 5.69 and Table 5.70 reveal the variability in remittances by region and location. Note that the bulk of the households receiving remittances are in Ulaanbaatar, with very few in the west and in general in rural areas. With the exception of a small number of households in the West, the bulk of the value of remittances goes mostly to households in Ulaanbaatar and the *aimag* centers, with very little in the *soums*, rural areas, and in particular little to the Eastern Region.

Status	Percent		
	Value	of income	N(households)
informal	457,477	26	2,948
inf_herd	609,380	14	1,261
formal	565,487	25	777
wage	392,395	26	4,317
unemp	344,850	48	3,193
Total	428,263	31	12,497

Remittance value is in togrogs; remittance share is percent of household income as foreign remittances. N(households) is the number of households receiving foreign remittances.

Status	Percent		
	Value	of income	N(households)
Ulaanbaatar	494,177	34	7,269
Aimag center	563,162	31	1,544
Soum center	322,850	29	2,675
Rural	26,362	8	1,009
Total	428,263	31	12,497

See notes for Table 5.70.

Status	Percent		
	Value	of income	N(households)
West	1,026,079	29	732
Khangai	276,515	29	2,693
Central	228,666	24	934
East	57,379	13	868
Ulaanbaatar	494,177	34	7,269
Total	428,263	31	12,497

See notes for Table 5.70.

Durability and hysteric change

The economic consequences of the informal sector shadow also depend on the extent to which there is exit and entry from the sector. Do people acquire skills in this sector that enable them to be more productive elsewhere? We find little evidence for this conjecture given the results presented in Table 5.71. Household incomes among the wage earning group are substantially lower at present for those who previously had informal self-employment income as compared to those who never participated in this sector. Nevertheless we observe, in Table 5.72 we observe frustrated entry into self-employment from among wage earners, with credit constraints mentioned by 49 percent as the reason for not entering the sector. It appears that 31 percent are satisfied with their wage earning and have no desire to enter self-employment.

Any self-employment income since 2001	Estimate				
	Estimate	Std. Err.	[95% Conf. Interval]		Deff
Previous	1108051	132136.7	849038.9	1367063	0.833752
Never	1508464	112165.7	1288599	1728329	1.055862

status2	Have fulltime job	Can't raise the funds	Taxes too high	Don't have skills	Family duties	Other	Total
Wage	0.3124	0.4925	0.035	0.055	0.0801	0.025	1

Among those who maintain informal self-employment businesses we find evidence in Table 5.73 that behavior changes depending on the length of time in the informal sector. Compliance with the requirements of the Informal Sector Law through registry and payment of the *patente*, for example declines fairly consistently with the number of years in the informal sector. Those in the quintile of shortest duration of self-employment (1.5 years or less) admit a 67 percent compliance rate, which falls to 38 percent for those with more than nine years in the sector. This behavior indicates a diminishing concern with compliance and with relationships with officials among the most experienced of the informal sector self-employed.

Years in Current business	Owed and paid patente last month		Total
	No	Yes	
1.5	0.3273	0.6727	1
3	0.3769	0.6231	1
5	0.3416	0.6584	1
9	0.4702	0.5298	1
60	0.623	0.377	1
Total	0.4195	0.5805	1

Key: row proportions

Each row represents 20 percent of businesses with the "Years in informal business" representing the upper range of a category that represents 20 percent of the business population.

Example of the multivariate approach: the relative cause of informal participation

In this final section of Chapter 5 we present an example of the sort of multivariate statistical analysis possible with the survey data. It is different from the descriptive statistics used in this report up to this point in that it simultaneously takes into account several sources of variation and permits their relative importance to be more rigorously compared.

In the present example, we use a multivariate probit model to investigate the differences between formal sector and informal sector self-employed business owners. The model generates a predicted probability of being in the informal sector based on the location and sector of the business and on personal and household characteristics. Many of the results discussed previously in the presentation of tables that compared personal characteristics across employment status carry through to this analysis. The power of this approach is greater, however, since we can observe the impact of each variable of interest, while holding the value of all other variables constant. The results should be interpreted with caution. There has been no attempt to address issues associated with the direction of causality of effects, which requires multi-equation modeling. For example, by one measure we find that trustworthy people are more likely to be in the formal sector. The results here cannot identify whether trustworthy people sort themselves into the sector or whether they become more trustworthy as a result of their participation in the formal sector.

In the model itself, dummy variables, with the value of either 1 or 0 are inserted for each region, location, and sector in order to measure their different effects. To estimate the model one of each of the dummy variables is dropped and the dropped value becomes the region, location, and sector to which the others are compared. In this model, rural agriculture in the west is the baseline to which others

are compared.³⁴ Characteristics and attitudes that are in the model include *gender*, age (*q17age*), education (*q20*), household size (*q25*), patience (*q30* and *q36*), trust (*q31*), trustworthiness (*q39*), fairness (*q35*), trust in government (*q40*), a measure of unfair treatment by a government official (*q41*), and risk (*q44*). We also include dummy variables regarding attitudes towards enforcement of business registration requirements (*q137*). This question has 7 possible responses and response 1 (“no negative consequences of non-registration, I could pay a bribe and solve the problem”) serves as the baseline for this question. Additional questions on perceptions of regulatory obstacles (*q148*), and perceptions regarding the extent of corruption (*q155*) and tax evasion (*q156*) are also included.

The results of the estimation are presented in Table 5.74. The information of primary interest is in the column labeled *delta Prob* which identifies the marginal change in the probability of being in the informal sector with a change in the variable identified in Column 1.³⁵ A positive *delta Prob* implies a greater likelihood of being in the informal sector. Both the size of the marginal effect and its statistical significance are important in understanding the results. To assess the statistical significance we consider those variables in which the p-value, identified by the column $P > |z|$, is less than 0.10 to identify effects that are not due to statistical noise. Thus, for example, the age variable *q17age* is significant ($p=0.076$), but the effect is small. A change in age changes the probability of being in the informal sector by -0.06 percent indicating that there is a small effect with younger workers more likely to be informal sector business owners. By contrast, the gender variable (coded as male=1 and female=0) has a statistically insignificant (p-value of 0.137) marginal effect of 1.04 percent, leading us to conclude that there are no gender differences in shadow participation after controlling for other influences.

The largest impact on the likelihood of informal participation is associated with perceptions regarding consequences of non-registration. Recall that the baseline for this response is that there would be no negative consequences aside from the need to pay a

bribe. Those responding (3), (5), and (6) to this question are much more likely to be in the formal sector with marginal effects of -10.20 percent, -15.62 percent, and -8.62 percent, respectively, and all with strongly significant p-values. These responses are associated with costs due to noncompliance that are either uncertain but high (3), with bureaucratic inefficiency (5), and the need to pay back taxes (6).

Among the personal characteristics education (*q20*) and trustworthiness (*q39b*) have the largest effects of variables that are statistically significant. For both, higher levels are associated with participation in the formal sector. The perception of unfair treatment by a government official (*q41*) is also strongly associated with formal sector participation. These variable have smaller effects than the enforcement question (-2.57 percent for *q20*, -2.30 percent for *q39b*, and -2.63 percent for *q41*) but still have an important impact on behavior. The perception of regulatory obstacles is also important. Here, those perceiving regulations to be significant obstacles are 2.8 percent more likely to be in the informal sector than those not perceiving regulations to be so.

There are some significant regional and sectoral effects as well, with the Khangai (-4.18 percent) and Central (-3.11 percent) regions more likely to be formal, and the East Region (5.97 percent) less formal. Ulaanbaatar is not significantly different than the baseline, which we took as the Western Region ($p=0.30$).

Now we turn to the influence of the economic sector in the decision to be informal. For this, we use the agricultural sector as the comparison. Of the larger sectors, manufacturing (Sector 4) and transport (Sector 9) show significantly more formality than the agricultural sector (4.30 percent and 4.59 percent, respectively) although the largest sector, trade (Sector 7), is not significantly different from agriculture ($p=0.287$). Hotels and restaurants (Sector 8) are more formal than agriculture (-4.54 percent). Finally, several of the smaller sectors, including finance, education, and health are, unsurprisingly, substantially more likely to be in the formal sector than agriculture as well (-12.79 percent, -12.84 percent, and -18.43 percent, respectively).

³⁴ With regard to locations, Ulaanbaatar (Location 1) is dropped along with the rural (Location 4) because it is already included as a region (Region 5).

³⁵ The marginal effects are calculated at the mean of the value of the variable, these values are in the column \bar{x} in the table of results.

Probit estimates

Number of obs = 6318

LR chi2(40) = 662.68

Prob > chi2 = 0.0000

Log likelihood = -1794.5266

Pseudo R2 = 0.1559

shadow2	delta Prob	Std. Err.	z	P> z	x-bar	[95% C.I.]	
gender*	0.0103859	0.0069841	1.49	0.137	0.50918	-0.003303	0.024074
q17age	-0.0005919	0.0003336	-1.77	0.076	39.7395	-0.001246	0.000062
q20	-0.0257137	0.0022578	-11.35	0	4.61048	-0.030139	-0.02129
q25	-0.0003224	0.0021867	-0.15	0.883	4.41722	-0.004608	0.003963
q30	-0.0122474	0.0030134	-4.06	0	2.65369	-0.018154	-0.00634
q31	-0.0018129	0.0044272	-0.41	0.682	1.94302	-0.01049	0.006864
q33	0.0072829	0.0046116	1.58	0.114	2.71921	-0.001756	0.016321
q35	0.0076088	0.0069116	1.1	0.271	1.47657	-0.005938	0.021155
q36	0.0007979	0.0004203	1.9	0.058	9.67569	-0.000026	0.001622
q39b*	-0.0229277	0.0072715	-2.96	0.003	0.753245	-0.03718	-0.00868
q40	-0.003819	0.0033222	-1.15	0.251	3.43906	-0.01033	0.002692
q41	-0.0263176	0.0049345	-5.33	0	1.42039	-0.035989	-0.01665
q44	-2.14E-06	1.32E-06	-1.62	0.106	1807.07	-4.70E-06	4.50E-07
_Iq137_2*	-0.0231716	0.0216261	-1.06	0.291	0.563944	-0.065558	0.019215
_Iq137_3*	-0.102005	0.0404939	-3.19	0.001	0.092909	-0.181371	-0.02264
_Iq137_4*	-0.041845	0.0713811	-0.68	0.494	0.006648	-0.181749	0.098059
_Iq137_5*	-0.1561566	0.0480348	-4.37	0	0.072175	-0.250303	-0.06201
_Iq137_6*	-0.0872593	0.0332851	-3.13	0.002	0.207186	-0.152497	-0.02202
_Iq137_7*	-0.04861	0.0459582	-1.25	0.212	0.019152	-0.138686	0.041466
q148	0.0285538	0.0077724	3.68	0	1.80073	0.01332	0.043787
q155	-0.0012807	0.0035477	-0.36	0.718	4.26242	-0.008234	0.005673
q156	0.0054081	0.0030138	1.79	0.073	3.67854	-0.000499	0.011315
_Iregi~2*	-0.0417587	0.0153167	-3.02	0.003	0.216049	-0.071779	-0.01174
_Iregi~3*	-0.0311075	0.0160931	-2.12	0.034	0.154004	-0.06265	0.000434
_Iregi~4*	0.0597119	0.0099069	3.36	0.001	0.04416	0.040295	0.079129
_Iregi~5*	0.0396875	0.0382784	1.03	0.303	0.468186	-0.035337	0.114712
_Iloca~2*	0.0514993	0.0359014	1.41	0.159	0.440013	-0.018866	0.121865
_Iloca~3*	0.0436495	0.0232569	1.43	0.153	0.084362	-0.001933	0.089232
_Isec~_3*	0.0398966	0.0270329	1.07	0.285	0.006964	-0.013087	0.09288
_Isec~_4*	0.0429267	0.0125999	2.84	0.005	0.176638	0.018231	0.067622
_Isec~_5*	-0.0429917	0.0897777	-0.56	0.575	0.002374	-0.218953	0.132969
_Isec~_6*	0.021471	0.0210012	0.9	0.369	0.023583	-0.01969	0.062633
_Isect~7*	0.0172487	0.0162658	1.07	0.287	0.520734	-0.014632	0.049129
_Isect~8*	-0.0453994	0.0278379	-1.9	0.057	0.043526	-0.099961	0.009162
_Isect~9*	0.0458564	0.0118277	2.98	0.003	0.112219	0.022675	0.069038
_Isect~10*	-0.1278727	0.0664144	-2.59	0.01	0.006015	-0.258043	0.002297
_Isect~11*	-0.0113675	0.0324756	-0.37	0.712	0.012504	-0.075018	0.052283
_Isect~13*	-0.1283704	0.078915	-2.19	0.029	0.003957	-0.283041	0.0263
_Isect~14*	-0.1843417	0.0699129	-3.69	0	0.007756	-0.321368	-0.04732
_Isect~15*	-0.0136218	0.02416	-0.6	0.55	0.040203	-0.060975	0.033731
obs. P	0.8947452						
pred. P	0.9254726	(at x-bar)					

(*) delta Prob is for discrete change of dummy variable from 0 to 1

z and P>|z| are the test of the underlying coefficient being 0

e are now in a position to summarize what we have now learned about the informal sector and its components as well as what still remains to be done with the wealth of information collected in the household survey. We then end by suggesting next steps in implementing the remediation plan (also developed in the current project) so as to ensure the sustainability and comprehensiveness of the NSO's efforts to measure shadow economic activity in Mongolia on a regular and cost-effective basis.

Results can be summarized according to (i) measurement and distribution, (ii) informal sector characteristics, (iii) causes and consequences and (iv) treatment effects. Recall that we break the informal sector into two parts, a herding sector (ISIC 012) and all other sectors.

Also recall that while the discussion and tables of the previous sections create a rich picture of the informal sector, two caveats are in order. First, what emerges is primarily a descriptive picture based on bivariate correlations (relationships between two characteristics). In order to generate a more insightful and precise understanding additional multivariate analysis is recommended. Second, the results need to be interpreted in light of the various elicitation methods used in the survey and these still remain to be synthesized. Until then, conclusions must remain tentative since each elicitation method generates different statistics, especially concerning those issues involving the "sensitive" (potentially incriminating) questions. Rectifying both is discussed in Section 6.2, below.

Measurement and distribution

The NSO's preliminary estimate of GDP for 2004 is 1,807 billion togrogs. Of this amount we estimate from NSO data that 259 billion togrogs or 14.34 percent of the formal activity is associated with earlier attempts to measure the informal sector shadow economy (Bikales *et al.* 2000). Mean estimates from the ISHS suggest that a better estimate of the informal sector shadow, not including the animal herding sector, is

24 percent (multiplier method) or 30 percent (envelope method). This expansion in the informal sector shadow arises for both economic and statistical reasons. The evidence suggests growth in several sectors including the trade sector, relative to overall GDP growth. In addition, the ISHS covers additional sectors with the most significant being manufacturing.

The ISHS yielded results with relatively wide confidence bands due to the existence within the sample of a small number of firms with extremely large value added. Our analysis suggests that these firms do exist in the population and that the results were not due to errors of enumeration or data entry, which in general was of high quality. The fact that the informal sector has these two distinct components; a large mass of smaller firms, alongside a few larger ones is a result that deserves additional scrutiny, since policy concerns and prescriptions differ significantly for the two groups. Additional analysis that segregates the two would likely find that poverty is a more serious concern than the aggregated analysis revealed.

Characteristics of the informal sector

The survey enumeration succeeded in its goal of interviewing primary earners (92 percent). These are the individuals whose characteristics are most relevant for studying compliance decisions and other economic behavior. Across the self-employed groups no great differences exist regarding gender, marital status, family size, and age, though two-thirds of the herders are male compared to 53 percent for the self-employed in general. *Viz.* education, non-herder informals in general have a higher level of education than herders and the formal sector in turn has a higher level of education than the non-herder informals.

The survey design was based on the insight from other IRIS work that social and personal attitudes play an important role in the type of economic decisions that individuals make. As discussed in Section 6.2, below these characteristics can help to adjust responses for several biases as well as to help combine sets of observations gathered from different treatments. Hence the survey posed questions covering a range of issues including patience, trust—

including trust of government, trustworthiness, fairness and honesty, and the willingness to take risks. We find, in general, the herding sector has quite different attitudes than those of the formal and informal sector business owners, who are often similar to each other.

In particular, we find formal sector owners the least patient and trusting of all groups; herders showed the highest level of trust. No significant differences were found in trustworthiness. Only a quarter of respondents indicated they regularly attend religious establishments. *Viz.* self-assessed honesty and a sense of fairness, herders and the unemployed showed the highest levels, with self-employed business owners the least. No big differences were found in the latter between the formal and informal sector. Regarding risk attitudes, the herders were substantially more willing to take risks, a result all the more surprising given their lower levels of income. Finally, the views of respondents regarding trust in and fairness of the government fall with formality of the business, and this fall is associated with negative experiences that increase with business formality.

We estimate that approximately 286,000 households had current self-employment income in September 2004, distributed across the herding sector (49 percent), the non-herding informal sector (46 percent), and the formal sector (5 percent). Regional variation of self-employment showed the non-herder informal share to range from 19 percent in the Eastern Region to 86 percent in Ulaanbaatar. Location differences were also great. Self-employment showed the non-herder informal share to range from 5 percent in the rural areas to 73 percent in *soum* centers to 82 percent in *aimag* centers. (Note that here we have been discussing the *number* of self-employed households, whereas above, when discussing measurement, we refer to the distribution of *value* of self-employment activity.)

Turning to the composition of self-employment by economic sector, we find the trade activity, which includes wholesale, retail, and repairs, to contain 50 percent of the total of self-employment in both the formal and informal sectors. However, within the trade sector itself, 90 percent of self-employed households are in the informal sector, compared to just 10 percent for formal businesses. Following trade, the economic sectors that involve the largest share of the informal sector households are manufacturing (20 percent) and transport, storage and communications (11 percent). Finally, formal self-employed firms, with 4.9 workers (including the owner), are on average more

than twice as large as the informal sector firms, which average 2.3 workers.

Wage payments to household members are rare across all business status, with wage employees highest in the formal sector as expected. Unpaid family workers are the predominant type of employee in the non-herding informal sector. Employees work considerably more and are better paid in the formal sector. Our estimate of the marginal product of labor, the amount required to hire an additional worker for one hour ranged from a negligible amount (47 togrogs) among the herders, to 387 togrogs among non-herding informals, and 666 togrogs among the formal sector businesses. Finally, formal and non-herding informal sectors report breaks in business activity of 2 to 3 months per year, while herders report essentially constant activity.

Collective action in the private sector is often abetted through business associations. Associational activity is low in Mongolia, with more than 98 percent reporting that they are uninvolved in business associations.

Formal firms are more than twice as likely to have written accounts (77 percent versus 36 percent). In the large informal sectors written records are scarce in manufacturing (20 percent), construction (9 percent), and transport, storage and communications (11 percent). Regarding the location of the business workplace, a slightly larger share of informal businesses are home-based (36 percent versus 30 percent). Only 40 percent of non-herding informals have stable non-home-based businesses locations; the figure for the formal sector is 56 percent. With regards to ownership type the formal and informal (non-herding) sectors do not differ greatly with approximately 60 percent reporting sole ownership in both sectors.

Viz. sources of financing, two-thirds (for formal small businesses) to seven-eighths (herders) come from own (or family) sources. The rest is spread out among the other possible sources, with banks being about as important as moneylenders or family friends. Only about one-eighth of informal businesses have ever borrowed, compared to one-third for formal small businesses.

Finally, with regard to dispute resolution there seems to be a consensus that the courts are viable, with only 10 percent citing deficiencies associated with capacity, corruption, or court costs as reasons for not using the judicial system. A further 7 percent would not use them because they are unsure of how to do

so, but roughly 83 percent of the population expressed a willingness to use this mechanism. These results were similar across employment statuses. There was a small decline in willingness to use the courts among the group who has used them previously, however the difference was not statistically significant. Among this group however, more raised concerns regarding the potential for bribery to affect outcomes and for the general unreliability of the system.

Causes and consequences

Due to the massive amount of information generated by the survey, this report was barely able to scratch the surface of the full implications for policy. We return to this point in Section 6.2. What we can say is that it appears that regulatory concerns and the perception that there is a lack of serious consequences for noncompliance are two factors associated with many of the illicit behaviors that respondents admitted. Formal businesses perceive sector-specific regulations and the business registry process itself as their largest obstacles. Informal workers are concerned with sector-specific regulations, customs regulations, and labor regulations. There are differences across the formal and informal employment statuses with regard to which sectors' regulations are a concern. This suggests that avoidance of formal sector regulation may be motivating informal activity. The construction, manufacturing, hotel/restaurants, and transportation sectors have the largest differences in proportions reporting regulatory obstacles across the formal and informal statuses. In finance, mining, and agriculture, the differences in perceived regulatory burdens are relatively similar.

With regard to registration requirements, the consensus is that noncompliance leads to minor penalties, such as being forced to register. However the likelihood of discovery is believed to be high and this leads to compliance rates with the informal sector law at a rate of roughly 60 percent.

Treatment Effects

In the discussion of survey results relating to corrupt practices and tax evasion the results reported above have made use of responses derived from the *two-question* method. As discussed in section 2 the two-

question method was the most secure question style implemented for "yes-no" questions since both the envelope and a scrambling device were used. Thus the respondent's anonymity was protected both during and after the interview. In this section we present descriptive statistics showing that, in general, more sensitive information was revealed through the use of the two-question method. In addition Table 6.1 reveals a number of other interesting and important effects.

In Table 6.1 Treatment 1 is the direct-question method. Treatments 2 and 4 use the envelope technique, and Treatments 3 and 5 use the two-question method. Treatment 2 and Treatment 4 differ in that during the income and balance sheet section, Treatment 2 used the envelope while Treatment 4 used the multiplier. Treatments 3 and Treatment 5 differ in the same way. In both cases the treatment with the lower number used the less secure method for the income and balance sheet questions. Our results show that, while the effect is small, those who received more security during the income and balance sheet questions were consistently more willing to reveal sensitive information in the latter part of the survey. For example in Question 158 8.5 percent admitted export related bribes in Treatment 2 while 11.5 percent did so in Treatment 4.

These differences suggest that not only is the question style for a particular topic important, but that the overall tone set by the survey with regard to protecting the anonymity of respondents has an effect. Efforts made earlier in the survey to reassure respondents by using secure question styles had a positive impact later in the survey even when the question style did not differ. This effect is consistent with results discussed earlier suggesting that reassurances of confidentiality can be important in generating truthful responses (Moore and Loomis 2001). While concern with the overall impression of the survey should not be neglected, we see much more dramatic effects within the sensitive question section when comparing the envelope and two-question methods directly. In Table 6.2 we combine observations from (envelope) Treatments 2 and 4 and (two-question) Treatments 3 and 5, and for comparison also present the direct-question approach (Treatment 1).

Q158-Q163 were asked of self-employed.

Q164 was asked of full sample.

Q158 Have you paid customs officials?

Treatment	Share answering "Yes"	Number receiving the treatment
Treatment-1	0.1011905	168,0
Treatment-2	0.0853081	1,055
Treatment-3	0.1666525	1,253
Treatment-4	0.1153846	156
Treatment-5	0.1817627	4,848
Total	0.1624333	7,480

Q159 Have you smuggled goods into Mongolia?

Treatment	Share answering "Yes"	Number receiving the treatment
Treatment-1	0.1011905	168,0
Treatment-2	0.1090047	1,055
Treatment-3	0.1686605	1,253
Treatment-4	0.1217949	156
Treatment-5	0.1975703	4,848
Total	0.176491	7,480

Q160 Do you believe you are required to pay the Informal Sector tax?

Treatment	Share answering "Yes"	Number receiving the treatment
Treatment-1	0.6369048	168
Treatment-2	0.5905213	1,055
Treatment-3	0.6907012	1,251
Treatment-4	0.6474359	156
Treatment-5	0.6929926	4,847
Total	0.6759399	7,477

Q161 Did you pay the Informal Sector tax

Treatment	Share answering "Yes"	Number receiving the treatment
Treatment-1	0.5963303	109
Treatment-2	0.6305344	655
Treatment-3	0.5909976	898
Treatment-4	0.5471698	106
Treatment-5	0.561185	3,478
Total	0.5753941	5,246

Q162 Have you paid money to government officials?

Treatment	Share answering “Yes”	Number receiving the treatment
Treatment-1	0.1666667	168
Treatment-2	0.1488152	1,055
Treatment-3	0.2675456	1,251
Treatment-4	0.1730769	156
Treatment-5	0.2921428	4,848
Total	0.2625044	7,478

Q163 Have you presented gifts to government officials?

Treatment	Share answering “Yes”	Number receiving the treatment
Treatment-1	0.2678571	168
Treatment-2	0.2018957	1,055
Treatment-3	0.1268526	1,252
Treatment-4	0.2115385	156
Treatment-5	0.1934035	4,848
Total	0.1855114	7,479

Q164 Did you pay taxes in full amount required by law?

Treatment	Share answering “Yes”	Number receiving the treatment
Treatment-1	0.581749	263
Treatment-2	0.5993091	1,737
Treatment-3	0.5919496	2,048
Treatment-4	0.5402299	261
Treatment-5	0.5673019	7,855
Total	0.5757537	12,164

With one exception we see that the two-question method is effective. With regard to admitting bribes to customs officials in q158 double the number (18 percent versus 9 percent) admit this behavior when provided the additional security of the two question method. The result is similar with regard to smuggling activity, 19 percent admit this behavior with the two-question approach versus 11 percent with the envelope method. These results are statistically significant at a p-value less than 0.01. The direct responses are not significantly different than those derived from the envelope method.

The questions on the *patente* eligibility and compliance also are more revealing due to the use of the two-question method. The first question (q160)

was used to identify only those for whom it was relevant to ask the *patente* compliance question. While this was not deemed overly sensitive, there is a significantly greater number admitting a need to comply with the two-question method (69 percent versus 60 percent). This suggests people may have anticipated the follow-up question on compliance or in general were less willing to address the *patente* issue without the additional security. There is also a relatively small difference in the compliance rates with 57-percent compliant under the two-question method and 62-percent compliant under the envelope method. As mentioned above, however, it appears a larger number of the non-compliant did not answer this question in the envelope treatment. Results with both

methodologies suggest *patente* compliance is an important policy issue.

The results regarding bribery in general provide a similar pattern of evidence as those regarding border activity. Twenty-nine percent admit bribery with the two-question technique as opposed to 15 percent with the envelope and 17 percent with the direct method—a value which does not differ, statistically, from the envelope result. Following up the bribery question is a less sensitive question regarding giving non-monetary gifts. Interestingly, the results are

reversed for this question with the two-question method revealing the least gift-giving (18 percent) and the direct question the most (26 percent). It seems that both monetary bribes and the gift-giving are fairly widespread but, not surprisingly, people substitute the less sensitive behavior for the more sensitive one when asked directly. The final question on tax-evasion does not show a statistically significant difference across question types leading us to believe that this question is not particularly sensitive.

Treatment	mean(q158)	Number
Direct	0.1011905	168
Envelop	0.0891825	1.211
Two-q	0.1786594	6.101
Total	0.1624333	7.480

Treatment	mean(q158)	Number
Direct	0.1011905	168
Envelop	0.1106523	1.211
Two-q	0.1916329	6.101
Total	0.176491	7.480

Treatment	mean(q158)	Number
Direct	0.6369048	168
Envelop	0.597853	1.211
Two-q	0.6925225	6.098
Total	0.6759399	7.477

Treatment	mean(q158)	Number
Direct	0.5963303	109
Envelop	0.6189225	761
Two-q	0.5673029	4.376
Total	0.5753942	5.246

Treatment	mean(q158)	Number
Direct	0.1666667	168
Envelop	0.1519405	1.211
Two-q	0.2870975	6.099
Total	0.2625044	7.478

Treatment	mean(q158)	Number
Direct	0.2678571	168
Envelop	0.2031379	1.211
Two-q	0.1797442	6.100
Total	0.1855114	7.479

Treatment	mean(q158)	Number
Direct	0.581749	263
Envelop	0.5915916	1.998
Two-q	0.5723991	9.903
Total	0.5757537	12.16

The analysis of the ISHS presented in the previous sections does not provide a complete analysis of the data collected in the survey and in this section we discuss several areas where additional work could be done to refine the analysis and glean more useful results from the study. Some of these refinements require the use of data external to the survey while others simply apply additional concepts or techniques to the ISHS data itself. Furthermore, during the implementation and analysis of the ISHS it became clear that some aspects of the implementation could be improved if a similar survey were to be implemented in the future. In this section we consider both of these issues, identifying potential refinements to the survey instrument, the analysis, as well as some issues associated with the selection of the survey sample.

Analytical refinements

Analytical refinements with regard to both shadow economy measurement and policy implications are possible and we believe could be informative. With regard to the overall measurement of the informal sector shadow economy we have reported design based estimates that are based on survey weights derived from the sampling plan. An alternative to the design based approach that would serve to check the robustness of the design-based estimates is to model the size of the informal sector shadow econometrically by making use of additional data from outside of the ISHS. The most critical data for a model-based estimation is the NSO's data from the business registry on the formal sector, and the Ministry of Finance's (MOF) data on the extent of *patente* registry. In addition to providing a check on the design-based measures, the model-based approaches, by

providing a link between ISHS data and data that is available in all *aimags* and *soums* would allow us to generate estimates of the size and sectoral distribution of informal shadow activity in areas of the country that were not surveyed in the ISHS.

Modeling approaches could be implemented with varying degrees of comprehensiveness depending on an assessment of priorities. For example, a model that included formal sector information from the business register could also be supplemented by formal sector results from the ISHS. Design-based estimates suggest that the ISHS picked up households representing approximately 80 percent of the firms in the business registry. Better modeling of the informal sector could be obtained by examining the relationship between these two sources of formal sector data.

Refinements of the model based measurements presented earlier in this volume are also possible. One area in which the survey measures could be improved is in assessing the regional distribution of the informal sector shadow size. Budget limitations required that a limited number of *aimags* be sampled in each region. This restriction made it difficult to fully measure the variability from north to south in each region. As a result the regional measures in activity may be biased. The survey as a whole, however, did contain a reasonable distribution of *aimags* in the north and south. We therefore can use “post-stratification” techniques that infer regional activity by using results from other regions, to improve estimates of the regional distribution of activity. Further work on the precision of the estimates is also warranted. There are a number of avenues to consider here. First the non-normality of the dataset implies that a small number of firms have a large impact on the overall shadow size and estimates of its variance. This arises from the small number of firms that have very high incomes leading to a long tail on the upper end of the annualized value added distribution. We believe it is incorrect to remove these observations. Additional insights into the causes and consequences of shadow activity, as well as interesting measurement results could be attained by analyzing the upper tail and the 90 percent of the observations that make up the “body” of the survey responses independently.

The distribution of activity across economic sectors could also be addressed with model-based analytical approaches. As discussed with regards to the design-based measurement results, the size of several sectors differed from expectations. While the design-based measures provide an unbiased estimate

of sector size due to the random selection of households, it is possible that areas rich in activity in a specific sector are missed. A model based approach would allow inferences regarding the size of sectors to be made in areas that were not part of the ISHS sample.

Another set of ways to improve the current estimates that remains to be implemented is to take advantage of the two survey sections containing respondent and enumerator feedback information, which was explicitly designed for this purpose. These questions contain information regarding the seriousness and conscientiousness of the respondent, an assessment of her understanding of the questions (and games) asked, and the enumerator’s assessment of the respondent’s relative standard of living. This information can be used to weigh certain observations more than others in the econometric estimation procedures as well as to conduct verification and even adjustment of the responses themselves.

In addition to using the formal sector data in the ISHS for modeling as discussed above, the formal sector data may be used to generate a rough estimate of the formal sector shadow contribution from these smaller formal sector businesses. Since we believe that the shadow economy may contain a substantial formal component, this extension may provide initial estimates that though provisional, due to the auxiliary assumptions needed to make such an estimate, may be a useful start. At the least, the results may be informative regarding the types of resources it would be worthwhile expending on investigating formal sector activity.

Another aspect of the ISHS analysis that could bear further development is the linking of characteristics to behaviors. While the tabulations generated for the discussion in Section 5 provide insight into how those with different employment statuses differ, a further analysis that examines participation by status conditional on characteristics would have more power to reveal causal links. Moreover, using attitudes and characteristics data would allow the pooling of responses from the various treatments, thereby effectively increasing sample size and thus narrowing confidence bands.

Survey refinement

The survey was implemented in five versions, or treatments, and one of the surprising results was the significantly lower estimates both with regard to value added and household income, that were revealed by

the multiplier treatment. While we continue to believe that this method is appropriate for protecting individuals' anonymity, the large differences between this protocol and the envelope method in rural areas suggests that this protocol should not be used in its present form in these areas. We do think that the multiplier method could be modified slightly to address a question that was not included in the current implementation of the ISHS. Although we asked about participation in corrupt activities, we did not attempt to gather more specific information on their value. We believe that a variant of the multiplier method, which was sensitive to the fact that any non-zero amount was revealing sensitive information, could be developed for future survey work so that values of bribes paid could be estimated.

We suggest that two options be considered before using the income and balance sheet again. First, it would be useful to provide additional options when asking individuals to provide estimates of seasonal variability of their income. This would smooth calculations for seasonal adjustments, although econometric methods could also be used to achieve this objective. In addition it would be useful, if several treatments are implemented, to implement one that provided detailed information on the components of value added rather than the summed aggregates which were collected in the ISHS. While the aggregation was intended to provide the respondents with additional security, some information is lost with this procedure, and additional policy implications could be drawn with knowledge of the details of the income and balance sheet.

Further thought needs to be given, as well, to the operational transformation of the survey if it is used to provide updated measures of the size of the informal sector shadow economy in years to come. An alternative to the mixed household-enterprise survey that was considered for the current project was the attachment of the survey to an ongoing NSO data collection effort, with the Household Income and Expenditure Survey (HHIES) the best candidate for this approach. While it was decided that the independent mixed survey was more appropriate for meeting the current goals, updates created by attaching the income and balance sheet component of the ISHS, and other selected sections to the HHIES may be feasible. One issue to address in any future implementation is to consider whether a fuller

enumeration of self-employment activity for each household member should be attempted. Given the relatively greater time taken to complete the HHIES as compared with the ISHS this might be feasible and would give a better understanding of additional, though smaller self-employment activities in the household.

Other issues associated with specific questions also arose. Question 111 on current income, inserted as a check on the balance sheet results, proved to be important for estimating value added due to the large number firms reporting current losses. Question 111 itself could be modified to more exactly mirror the income derived from the income and balance sheet by asking the respondent to report other income streams. The income and balance sheet should be adjusted to accommodate losses both in original recording of results and in extrapolation of the seasonal adjustments.

We think there are two important ways that additional useful information could be extracted from the ISHS project with a view toward improving future implementations. A first step would involve conducting an analysis on how survey costs and the sampling plan interacted so that an optimal design can be constructed for future efforts. While estimates of travel costs and other survey costs were well understood, the distribution of the sample made use of very limited information on the variability of the underlying values that were to be measured. The knowledge of the variability of these figures can now be used to fine tune the sampling plan for cost effectiveness.

A final issue to consider is whether the overall framework, which made use of a variant of the mixed household-enterprise survey, could be improved upon. One survey structure that was considered but not implemented was a mixed household-enterprise survey with two-stage sampling. Two-stage sampling involves an initial set of brief interviews from a random sample of the household frame to determine if individuals are of interest for the informal sector survey. This initial interview should be done in a relatively unclustered fashion, perhaps by making use of local NSO staff. The results of the first stage would then be used to stratify the sample for the final survey. The critical advantage of this method is that stratification could be done by sector, insuring appropriate representation of all sectors in the survey.

This note outlines two approaches to providing protection to survey respondents who are asked to reveal sensitive information about their economic activity. Randomized response methods provide protection by introducing noise into the data collection process so that it is impossible for the interviewer or the final users of the data to know how any specific individual has responded to a sensitive question, allaying fears of exposure. The methods provide population data, however, on the subjects of interest.

Because the randomized response yields aggregate values without linking responses directly to individuals it often outperforms direct questions for measurement of population values. Randomized response approaches are less useful however for some policy analyses, such as those making use of regressions in which individual characteristics and individual behavior must be linked. As a result we expect that it will be necessary to use both randomized response and direct question methods in order to get the most informative data from the shadow economy survey.

Randomized response methods have been used in a variety of contexts in both developing and developed countries beginning in the mid 1960's after the groundbreaking work of Warner (1965). Most early implementations of randomized response techniques investigated sensitive issues such as drug and alcohol use and sexual behavior. Applications with more direct economic content have become more prevalent. Chaudhuri (1983) investigated input use in a large agricultural survey using randomized response methods as a preliminary step to estimating production functions for crop agriculture in India. More recently randomized response techniques have been used to study tax evasion (Houston and Tran 2001) and the size of the unobserved economy in Turkey (Savasan 2003).

A large number of randomized response techniques have been developed and the literature suggests that the suitability of a method to specific problem and to a specific location deserves careful

study. In general there are clear advantages to randomized response methods both for reducing non-response rates and for getting respondents to reveal sensitive information. Two types of randomized response techniques are proposed for the shadow economy survey, one for dichotomous, yes/no questions, and one for questions with numeric responses. For dichotomous responses we implement a technique we call the *two-question method*. For numeric responses the technique will be called the *multiplier method*.

The two-question method presents two unrelated questions, one which is sensitive such as "Did you cheat on your taxes this year?" and a harmless question that people should not have a problem answering such as "Do you like basketball?" The outcome of a randomizing device, such as a die roll or coin flip, is observed only by the respondent. In the discussion that follows we assume that the sensitive and harmless questions are each asked with a 50% probability.

Let $\hat{\lambda}$ be the observed proportion responding "yes" to the two-question method question. The proportion of interest is π_S , the proportion responding "yes" to the sensitive question. To find π_S , the observed response can be decomposed such that $\hat{\lambda} = P * \pi_S + (1 - P)\pi_H$ where $P = .5$ is the probability determined by design, and π_H is the proportion responding "yes" to the harmless question. An estimate of the proportion with the sensitive characteristic is therefore given by

$$\hat{\pi}_S = \frac{(\hat{\lambda} - (1 - P)\pi_H)}{P} = \frac{\hat{\lambda}}{P} - \pi_H = 2\hat{\lambda} - \pi_H$$

Since P is chosen by design and $\hat{\lambda}$ is observed in the data, we need only have an estimate of to estimate the proportion with the sensitive attribute. The proportion is determined by splitting the respondent sample and asking another group who does not receive the two-question question to answer the harmless questions. Note that if $\hat{\lambda}$ is too large it is possible to infer that π_S is greater than 1, a nonsensical result. Care must be taken in

the design and through pre-testing to avoid this result.

The probability of being asked the sensitive question, P , has an important impact on the level of protection that the method provides because it represents the amount of noise introduced. The variance of the proportion of interest in the two question method is approximated by $\text{var}(\pi_s) = \frac{\text{var}(\hat{\lambda})}{P^2}$.

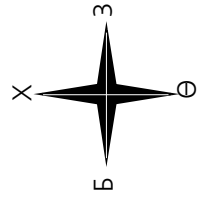
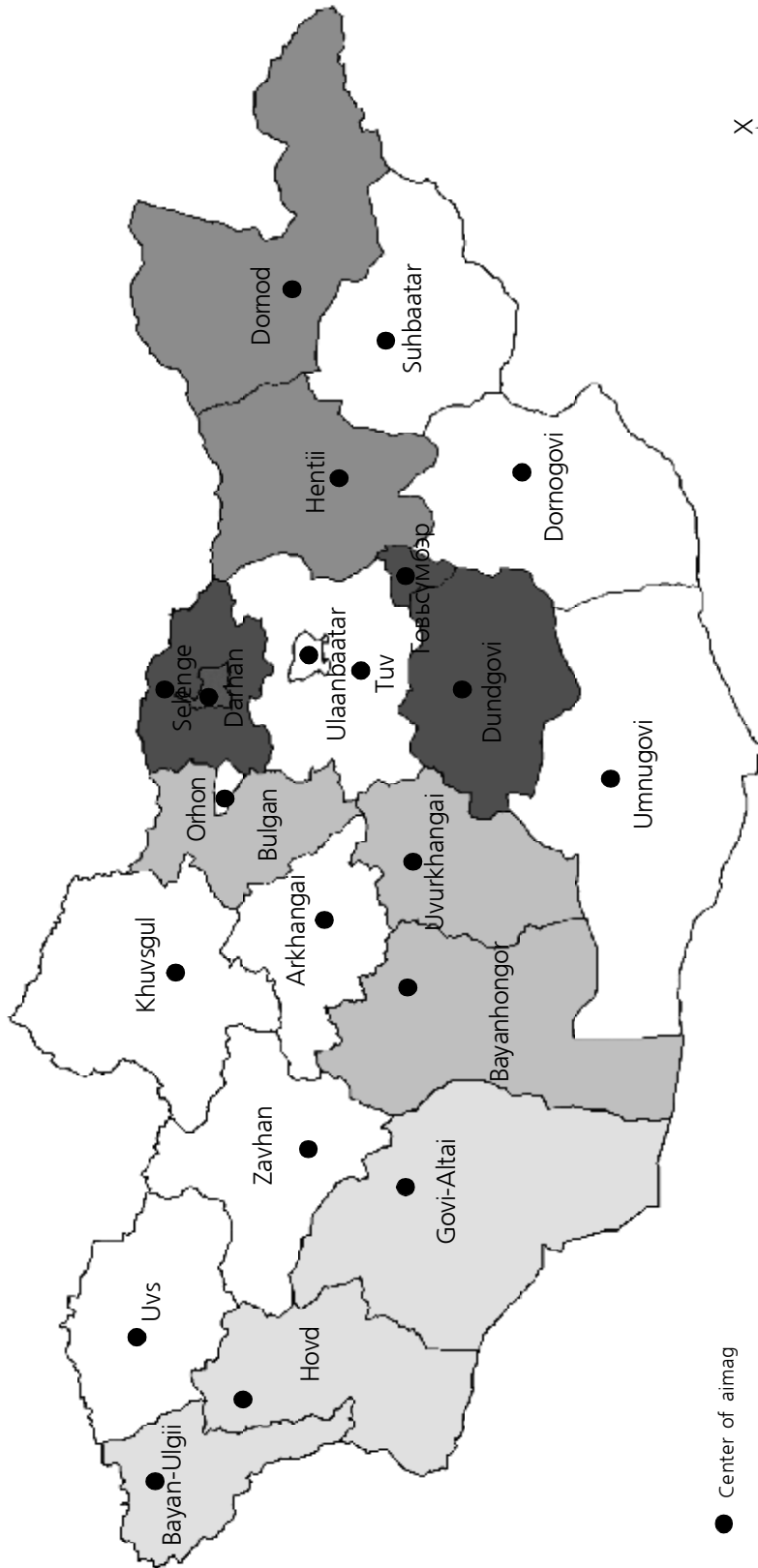
The multiplier method generates estimates for numeric data such as income or value added by asking individuals to multiply their truthful response to a question and multiply it by the roll of a die that only they observe. To generate different amounts of

confidentiality die of different sizes can be used. Let z be the observed response to the question. The parameter in the population is given by $Z = XY$, where X is the value of interest and Y is the scrambling component that results from the die roll³⁶. An unbiased estimate of the parameter of interest is denoted $\hat{\mu}_x$, with $\hat{\mu}_x = \frac{\bar{z}}{\mu_y}$. Here $\bar{z} = \frac{\sum_{i=1}^n z_i}{n}$ and

μ_y depends on the choice of scrambling device. A four-sided die for instance would yield

$$\mu_{y14} = \frac{1+2+3+4}{4} = 2.5. \text{ The variance of the estimate is given by } \text{var}(\mu_x) = \frac{\sigma_z^2}{n\mu_{y14}^2}.$$

³⁶ Additive approaches, with $Z=X+Y$, may also prove useful. These will provide protection when a response of zero may be revealing. The additive method poses some additional difficulties however since information on the magnitude of X is needed to insure that the distribution of Y provides protection for the respondent.



Open Society Forum

Stephen D. Vance, Executive Director, Mongolian
Foundation for Open Society

Erdenejargal Perenlei, Executive Director

Dorjdari Namkhajantsan, Manager

Munkhsoyol Baatarjav, Manager

Economic Policy Reform and Competitiveness Project (USAID, Chemonics)

Fernando Bertoli, Chief of Party

Janusz Szyrmer, Project Advisor

Doljinsuren Jambal, Economic Analyst

Consultancy Team, University of Maryland, Institutional Reform and the Informal Sector (IRIS)

Clifford Zinnes, Team Leader

Jonathan Alevy, Economist

Ch. Narantuya, Team Member

Brandie Sasser, Team Assistant

National Statistical Office

P. Byambatseren, Chairman

A. Demberel, Head of MEBSD

B. Badamtsetseg, Deputy Head of MEBSD

B. Lkhagvajargal, Officer of MEBSD

D. Pagmasuren, Officer of MEBSD

Enumerators/Enumerator Team Leaders

A. Zolboo

B. Lkhagvajargal

B. Bayarmaa

B. Bayanmandal

B. Naranzul

B. Khongorzul

B. Batbayar

B. Oyuntsetseg

B. Otgontsetseg

B. Tsevelsuren

B. Altangerel

B. Oyuntuya

B. Bolormaa

B. Enkhchimeg

B. Badambayar

B. Uranbaigal

B. Ankhtuya

B. Bulgantamir

G. Dolgorsuren

G. Sainkhuu

G. Davaasuren

G. Enkhbayar

G. Jiidiimaa

G. Damdinragchaa

G. Kherlen

G. Oyunbat

G. Enkhtungalag

D. Enkhtaivan

D. Badral

D. Suvdaa

D. Batchimeg

D. Enkhsaikhan

D. Odmaa

D. Enkhbayasgal

D. Narantsetseg

D. Otgoncimeg

D. Sarantuya

D. Enkhbold

J. Terbish

J. Odgiiv

J. Batsukh

J. Khulan

J. Batbold

J. Sarantsetseg

Z. Munkhjargal

Z. Munkhzul

L. Myagmarsuren

L. Altantsetseg

L. Unurnasan

L. Tuul

M. Oyun

M. Oyunchimeg

M. Munkh-Och

N. Munkhzaya

N. Batsukh

N. Javzanjargal

N. Otgonbayar

N. Enkhtuya

N. Altantsetseg

N. Ariyasuren

N. Bayarmaa

N. Dagvadorj

N. Byambasuren

O. Orkhonbaatar

P. Byambadorj

P. Gantsooj

P. Orgil

P. Unenbat

P. Batnasan

P. Battsetseg

R. Oidovdanzan

R. Erdenemandal

S. Oyuntsetseg

S. Enkh-Otgon

S. Narangerel

S. Ariunbold

T. Bodigerel

T. Nyam-Ochir

T. Burenjargal

U. Batjargal

Ts. Uranbileg

Ts. Tuya

Ts. Munkhtsatsral

Ts. Badrakh

Ts. Ulziibat

Ts. Dulamdorj

Ts. Oyunchimeg

- Anderson, J. (1998).** *"The Size, Origins, and Character of Mongolia's Informal Sector During the Transition"*, Policy Research Working Paper, no. 1916, The World Bank, May.
- Becker, B. (2000).** Minimum Data Set On the Informal Sector for National Accounting Purposes. *Household Accounting: Experience in Concepts and Compilation*. United Nations Department of Economic and Social Affairs: Statistics Division, 143-52. Vol. 1. New York: United Nations.
- Bikales, B., C. Khurelbaatar, K. Schelzig (2000).** *"The Mongolian Informal Sector: Survey Results and Analysis"*, USAID and DAI (EPSP).
- Eilat, Y. and C. Zinnes (2002).** *"The Evolution of the Shadow Economy in Transition Countries: the Consequences for Economic Growth and Donor Assistance"*, *World Development*, July..
- Chaudhuri, A. (1983).** *"RR technique to determine input crop estimation"*, *Bulletin of Calcutta Statistical Association* 32: 208-10.
- Cramer, J. S. (1971).** *Empirical Econometrics*, Amsterdam: North-Holland Publishing Company.
- Eichorn, B. H. and L. S. Hayre (1983).** *"Scrambled Randomized Response Methods for Obtaining Sensitive Quantitative Information"*, *Journal of Statistical Planning and Inference* 7: 307-16.
- Houston, J. and A Tran (2001).** *"A Survey of Tax Evasion using Randomized Response Techniques"*, *Advances in Taxation*, Vol. 13. New York: Elsevier.
- Kish, L. (1965).** *Survey Sampling*, New York: John Wiley & Sons, Inc.
- Moore, J. C., and L. S. Loomis (2001).** *Using Alternative Question Strategies to Reduce Income Non-response*. U.S. Census Research Report Series. Washington, DC: U.S. Census.
- National Statistical Office of Mongolia (2000).** *"Methodology to Estimate GDP and GNI at Current Prices by the Production, Income, and Expenditure Approaches."* Ulaanbaatar, Mongolia.
- OECD (2002).** *Measuring the Non-Observed Economy*, Paris: OECD.
- Savasan, F. (2003).** Modeling the Underground Economy in Turkey: *Randomized Response and MIMIC models*. *Journal of Economics* 29, no. 1: 49-76.
- Warner, S. L. (1965).** *"RR: A Survey Technique for Eliminating Evasive Answer Bias."* *Journal of the American Statistical Association* 60 (1965): 63-69.
- United Nations (1993).** *System of National Accounts*. New York, N.Y.: United Nations.