Online Appendix: A Note on Instrumental Variables

Choice of Instruments For Incentive (INC):

Extant research in the field of information systems has pointed to the role of incentives in shaping the behavior of providers. Researchers have noted that the quantum of incentives can shape the behavior of the vendor in terms of the effort induced (Kirsch 1997) as well as the responsiveness of the vendor to the client's needs (Kirsch *et al*. 2002). Building on these studies researchers have shown that in offshore BPO, contractual service level agreements (SLAs) require that vendors random sample both the finished output and work in progress for errors, at pre-specified random sampling rates (Aron et al., 2008; Knowledge@Wharton 2005; Seshasai et al. 2010). For the 139 processes for which we have cross sectional data, the SLA specified sampling rates for finished output and work in process were extracted from the contract. We also obtained[1] the actual average sampling rates from the QA documents. Based on this we constructed two Instrumental Variables which measure *the extent to which the vendor* ***exceeded*** *the SLA specified sampling rate* in audit of the final output (termed as final audit) as well as the extent to which the vendor ***exceeded*** the SLA specified random sample rates for work in progress (termed as preliminary audit). The two ***excess*** *sampling rate* variables represent the extent to which the vendor ***exceeded*** the sampling efforts specified in the SLA. The variable Excess Sampling Rate is operationalized as follows:

Excess Sampling Rate of final output[2], given by ***ESR_F*** is defined as:

$$ESR\_F_i = QAF_i - SLAQF_i$$

[1] A scan of a vendor QA log was furnished in the technical appendix. The technical appendix will be made available on request.

[2] The excess sampling rate - ***ESR_P*** - of work in process is similarly defined.

Where QAF_i is the *actual fraction of finished output that was randomly sampled for QA and* - $SLAQF_i$ *is the SLA specified proportion of finished output to be sampled for QA purposes, for the* i^{th} *process* (see scan of a vendor QA log in the Technical Appendix).

Note that we are **not** interested in the SLA specified sampling rate per se; instead we use the ***excess*** *sampling rate* - *because greater the quantum of penalty (reward) faced by the vendor, the greater will be his incentive to* ***exceed*** *the SLA specified QA levels to identify possible errors*. The quantum of **this** **excess effort** is a direct result of the penalties (rewards) faced by the vendor and we expect that the *excess sampling rates will be correlated with the incentive variable*. Note too that unlike the actual SLA specified sampling rates, the **excess rate** – the extent to which the vendor **exceeds** the SLA specified rates – **is the vendor's decision alone and is a response to the quantum of incentives faced by the vendor**.

The correlation coefficients of the two variables with potentially endogenous regressor are provided below in Table 1.

Table 1: Correlations Between The IVs And The Incentive (INC) Variable

		INC (Incentive)
ESR_P	**Pearson Correlation Coefficient**	***0.631***
	R Standard Error	(0.004)
	t	9.520
	p-value	0.000
ESR_F	**Pearson Correlation Coefficient**	***0.504***
	R Standard Error	(0.005)
	t	6.824
	p-value	0.000

It can be seen that the two Instrumental Variables are significantly correlated with the potentially endogenous variable, Incentive (INC). This is an important requirement of instrumental variables (Cameron and Trivedi 2005, Greene 2011).

Choice of Instruments for Process Codifiability (COD):

We discuss below the two variables that we identified as instruments for the potentially endogenous regressor Process Codifiability (COD). We offer both conceptual arguments and reasons grounded in extant research in support of our choice of these variables in the discussion that follows.

Our definition of codifiability is identical to that that of Mithas and Whitaker (2007) from whose paper we adopt the definition of process codifiability as "the extent to which the tasks constituting a process can be described completely in a set of written instructions" (measured using 1-7 Likert Scale items). For each process there are a number of tasks that constitute the process. For instance, see the example of BPO discussed in Table 1A in the main paper where we have outlined the tasks associated with a process in our data set. It can be seen that there are 8 tasks that constitute the example process in Table 1A, each with its own task specification (description of task requirements). For the 139 processes in our cross sectional data set we *extracted the number of tasks associated with each process*[3]. We define this variable as *Process Task Index* (PTI).

$PTI_i = k_i$ where k_i is the number of tasks that constitute the i^{th} process.

Note that the actual set of tasks associated with a process defines the process. It is not a decision made by the vendor (or the client). When we refer to a process we mean a set of tasks (Mithas and Whitaker (2007)) that need to be accomplished. These tasks do not come into being at the point of offshoring. They were tasks that were being performed in-house by the client before the process was offshored to a vendor. Extant research suggests that a process with multiple clearly specified tasks must also be highly codifiable and vice versa (Langlois, 2002; Ulrich and

[3] In one of the sequential rounds of review, we furnished an alternate operationalization of process codifiability – used to test the robustness of the model – in response to a reviewer request. This operationalization was based on aggregating codifiability ratings at the level of each task in a process. We used the process-level task count from that operationalization.

Eppinger, 1999). We would therefore, expect that this variable be significantly correlated with the process codifiability (COD) variable.

We extracted a second instrumental variable for Process Codifiability (COD) based on the research of Malone *et al*. (1999) and Fettke *et al*. (2010) and Scheer *et al*. (2010). This variable is based on the techniques that are used to define business processes in the BPO process specifications. Malone *et al*. (1999) observe that it has become easier to define the activities that constitute a complex business process and codify it thanks to the emergence of a set of process specification tools. These tools identified by Malone *et al*. are based on 'process grammar' and description 'libraries' consisting of state transition diagrams, data flow diagrams, flow charts and process goal representation models. Recent research in business process design by Fettke *et al*. (2010) and Scheer *et al*. (2010) shows that these techniques first identified by Malone *et al*. have been combined into the domain of business process management (BPM) and are deployed as five principal BPM tools in specifying the nature of business processes (Fettke *et al*. 2010; Scheer *et al* 2010).

Five BPM tools used for process description are identified; (i) Data Flow Diagrams (DF), (ii) Work Flow Diagrams (aka Process Flow Diagrams), (iii) Data Dictionaries (iv) Task Rules and (v) Output metrics. For each process in our cross sectional database we extracted from the process specification the *number of BPM tools* used. If data flow diagrams existed (in the process specification) for a process it was coded as 1 and 0 if not. Similarly if data dictionaries existed for a process it was coded as 1 and 0 if not and so on. Each one of the five BPM tools above used to define a business process structure was coded as 1 if it existed and 0 if it did not. These scores were combined into a single variable *Business Process Definition Index* – ***BPDI***. The BPDI score could range from 0 if none of the above tools were used to 5 if all of the above tools were used.

$$BPDI_i = \sum_{j=1}^{5} (t_{ij}) : \quad t_j \in \{0,1\}$$

The actual scores ranged from 1 to 5. Our expectation is that the higher the score, the greater is the extent of process description and therefore, greater the process codifiability (Malone *et al*. 1999; Scheer *et al*. 2010). We would expect this variable too would be significantly correlated with process codifiability.

We would expect this variable to be significantly correlated with process codifiability. The correlations of the two IVs with the potentially endogenous regressors are furnished in Table 2 below.

Table 2: Correlations Between The IVs And The Codifiability (COD) Variable

		COD (Codifiability)
PTI	**Pearson Correlation Coefficient**	***0.52***
	R Standard Error	(0.005)
	t	7.127
	p-value	0.000
BPDI	**Pearson Correlation Coefficient**	***0.535***
	R Standard Error	(0.005)
	t	7.418
	p-value	0.000

It can be seen that the two instrument variables are indeed significantly correlated with the (potentially) endogenous regressor, codifiability (COD).

References

Aron, R., Bandopadyay, S., Jayanty, S., and Pathak, P.S. 2008. "Monitoring Process Quality in Offshore Outsourcing: A Model and Findings from Multi-Country Survey," *Journal of Operations Management*, Volume 26, Issue 2, March 2008, Pages 303-321.

Cameron, C.A., and Trivedi, P.K. "Microeconometrics: Methods and Applications, Cambridge University Press May 9, 2005.

Fettke, P., Houy, C., Loos, P. 2010. On the Relevance of Design Knowledge for Design-Oriented Business and Information Systems Engineering. *Business and Information Systems Engineerin*g 2, 347–358 (2010).

Greene, W.H. 2011. "Econometric Analysis" Prentice Hall; 7 edition, February 13, 2011.

Kirsch, J. L. 1997. "Portfolios of Control Modes and IS Project Management," *Information Systems Research*, (8:3), pp. 215-239.

Kirsch, J. L., Sambamurthy, V., Ko, D.G., and Purvis, L.R. 2002. "Controlling Information Systems Development Projects: The View from the Client," *Management Science*, (48:4), pp. 484-498.

Knowledge@Wharton 2005. "As the BPO Business Grows, There's a Greater Focus on Metrics and Measurement." http://knowledge.wharton.upenn.edu/article.cfm?articleid=1102

Langlois R.N., 2002. Modularity in technology and organization *Journal of Economic Behavior & Organization* Vol. 49 (2002) 19–37.
Malone, T. W., Crowston, J., Lee, B., Pentland, C., Dellarocas, G., Wyner, J., Quimby, C. S. Osborn, A., Bernstein, G., Herman, M., Klein, E., O'Donnell. 1999. Tools for inventing organizations: Toward a handbook of organizational processes. *Management Sci.* 45(3) 425–443.
Mithas, S. and Whitaker, J. 2007. "Is the World Flat or Spiky? Information Intensity, Skills, and Global Service Disaggregation," *Information Systems Research*, (18:3), pp. 237-259.
Scheer, A.W., Brabänder, E. 2010. "The Process of Business Process Management" Handbook on Business Process Management 2 - Strategic Alignment, Governance, People and Culture, pp. 239–265. Springer, Berlin (2010).
Seshasai, S., Gupta, A., Aron, R. 2010. "The 24-Hour Knowledge Factory: Work And Organizational Redesign and Associated Challenges" *Information Resources Management Journal*, (23: 4), pp. 40-56.
Ulrich K, Eppinger S. 1999. *Product Design and Development* (2nd edn). McGraw Hill: New York.