Research Statement

Youzhi Yang

My research interests cover areas of dynamic contract theory, macroeconomics, repeated games with private information, and the Chinese labor market.

Most of my research has been developed around the core question of how to design optimal dynamic contractual relationships in frictional labor and financial market environments, and how they generate outcomes that can better account for the data? In a specific model we study, for example, a long-term principal-agent relationship may not last forever, both parties can go through a frictional search process to find a new partner, either after the termination of the current contract or while still in the ongoing relationship. The optimal contract is designed as a fully history-dependent plan about not only how to compensate the agent and provide him with proper incentives (to exert effort or report truthfully), but also when to terminate the ongoing relationship. The issues raised include how termination arises as part of the optimal contract, as well as how the optimal contract, with endogenous termination, helps account for the distribution and flows of workers across all three states of the labor market (employment, unemployment and non-participation) and the wage dispersion in the data. In addressing these issues, I developed novel theoretical techniques especially for solving challenging dynamic contracting problems with private information and termination. As a macroeconomist, however, a significant part of my research also involves using sophisticated calibration and stimulation techniques for building a relationship between model and data.

My first project consists of three papers which study a problem of optimal contracting between a risk neutral principal and a risk averse agent where a stochastic outside opportunity (offer) arises each period for the agent. Termination of the ongoing relationship is costly. After termination, the principal goes back to an external market to find a new agent, while the agent pursues the outside opportunity received. From a dynamic point of view, outside opportunities to which a long-term relationship is exposed to dictate not only when the relationship is terminated, but also how the parties involved interact, before and after termination.

[1] **Outside Opportunities and Termination** (with Cheng Wang), *Games and Economic Behavior*, 91, 2015, 207-228
[2] **Optimal Self-Enforcement and Termination** (with Cheng Wang), Under Review
[3] **Optimal CEO Contracting with Moral Hazard and External Shocks** (with Cheng Wang)

The first paper assumes full commitment on the agent's part, the second paper assumes lack of commitment on the agent's part, and the third paper assumes both lack of commitment and private

information (specifically, moral hazard) on the agent's part.

The novelty is that the principal acts strategically, instead of passively, with respect to the agent's outside opportunities. That is, the principal has to choose a subset of the agent's outside opportunities in which the relationship continues, and outside which the relationship ends. The technical difficulty arises since the new choice variable is a set which we cannot take derivative with respect to. To overcome this, we introduce a probability function $I(\xi) \in [0,1]$ which denotes the probability with which the relationship continues conditional on the agent's outside opportunity being ξ. Therefore, we have a continuum of choice variables which we can take derivative with respect to, instead of one choice variable which we cannot take derivative with respect to. Furthermore, an index function is derived from the first order conditions, which is positive/negative if it is optimal to continue/terminate the relationship. Then, we show that the set of the agent's outside opportunities in which the relationship continues is always an interval, which leaves us only two choice variables which we can take derivative with respect to.

In [1], we show that even if the agent is able to fully commit to a long-term relationship, termination arises as part of the optimal contract. Two motives for termination are identified: First, the principal uses the agent's outside opportunity as an external means for compensating the agent. Second, termination of the ongoing relationship allows the principal to replace the agent with a potentially less expensive new agent. As a result, it is optimal to terminate the ongoing relationship if and only if the agent's outside opportunity is above a threshold. Thus when termination occurs depends on the outside opportunity the agent receives, his promised expected utility, the reservation expected utility of a new agent the principal can hire to replace him, and, of course, the costs the principal must incur in the process of termination.

The optimal contract generates both voluntary and involuntary terminations. But only in involuntary termination, severance compensation arises. Specifically, if there is a sufficiently strong motive for the principal to replace the agent with a less expensive new agent, then the principal would terminate the agent even when the agent's outside opportunity is not good enough to generate temptations for him to leave voluntarily. This is when involuntary termination occurs, which generates the need for consumption smoothing – across the states of retention and termination – by the way of severance compensation. We show that the optimal severance compensation depends positively on the agent's current compensation, but negatively on his outside opportunity.

Furthermore, the optimal contract dictates an inverted-U relationship between compensation and the probability of termination. To see this, take the agent's outside opportunity ξ as given, while increasing the agent's expected utility V. On one hand, the agent's marginal net gains from termination in terms of utility $\xi - V$ decreases, which makes termination less effective in lowering the amount of utility the principal must give to the agent in the states of retention. On the other hand,

as the agent's expected utility increases, his marginal utility of consumption decreases. In turn, this implies that any gains from termination, when measured in units of the agent's compensation, are worth more from the principal's perspective. Apparently, the first effect reduces termination while the second effect increases termination, and that explains the inverted-U relationship.

In [2], we study a problem of optimal contracting where a stochastic outside opportunity arises each period for the agent, which generates temptations for him to leave the ongoing relationship. That is, there is lack of commitment on the agent's part. Instead of imposing universal self-enforcement by designing the contract to be such that no parties have incentives to leave under any circumstances as commonly in the literature, we make self-enforcement endogenous, by choosing optimally a subset of the agent's outside opportunities to impose self-enforcement, while allowing the relationship to end in the states outside of it.

We show that if it is costless for the principal to replace the agent with a new agent, then lack of commitment on the agent's part is not a binding constraint. The first best is achieved. Otherwise, there exists a constant threshold (independent of the agent's expected utility) such that the principal retains the agent by matching his outside opportunity if it is above his expected utility but below this threshold, and terminates the agent if it is above this threshold. Hence, starting from a sufficiently low expected utility of the agent (so the self-enforcement constraint is binding, initially), the continuation of the optimal contract converges to Burdett (1978) where each period the agent quits whenever his outside opportunity is above the expected utility the current principal offers, and stays to receive the same constant expected utility otherwise. That is, all optimal dynamic contracts converge stochastically, but monotonically, to a single stationary contract in which the self-enforcement constraint is not binding.

In [3], we study a principal-agent relationship which is subject to moral hazard from inside, stochastic shocks from outside, and lack of commitment on the agent's part. The optimal contract is characterized in which termination is used (at the same time) as (a) an incentive device to punish the agent for bad outcomes, (b) a cost minimization device to make use of the agent's outside value as an external means for compensating him, and (c) a turnover device to replace the agent with a less expensive new agent. The optimal contract generates voluntary termination, involuntary termination, retention with counteroffer, as well as retention without counteroffer on the equilibrium path.

We first show that the contract can be decomposed into an incentive sub-contract and a risk-sharing sub-contract. In the incentive sub-contract, the principal makes the agent's continuation expected utility contingent on his outcome, which provides incentives for him to exert effort. And, in the risk-sharing contract, the principal designs a full state-contingent scheme about how to compensate the agent, as well as how to respond to the agent's outside opportunities, just as studied in [2].

We also show that the principal would terminate the agent when his outside value is either too low or too high. The logic goes as follows. Whenever the agent's outcome turns out to be bad, the principal would punish the agent by offering a lower compensation in the current period, and also a lower continuation expected utility. Hence, after producing a sequence of bad outcomes, the agent's expected utility would be pushed lower and lower. Then, the principal would have incentives to terminate the agent not only when his outside value is low as punishment, but also when his outside value is high by making use of it as an external means for compensating him. This explains the seemly surprising result.

Furthermore, there is a U-shape relationship between compensation and the probability of involuntary termination. There are actually two types of involuntary termination. One is that after producing a sequence of bad outcomes, the agent's expected utility would be pushed lower and lower. Then, the principal would terminate the agent when his outside value is low as punishment. Another is that after producing a sequence of good outcomes, the agent's expected utility would be pushed higher and higher. Then, the principal would terminate the agent in order to hire a less expensive new agent even when his outside value is not good enough to generate temptations for him to leave voluntarily. The first type of involuntary termination occurs when the agent's compensation is low, while the second type occurs when the agent's compensation is high. This explains the U-shape relationship.

My second project consists of two papers which incorporate dynamic contracts with termination into the classical search and matching model by Mortensen and Pissarides (1994), and the classical on-the-job search model by Burdett and Mortensen (1998) respectively. This project focuses on understanding the labor market phenomena, such as the distribution and flows of workers across all three states (employment, unemployment and non-participation) as well as wage dispersion, from a unified framework in which the dynamics in the internal labor market is characterized as a problem of optimal contracting, and the dynamics in the external labor market is characterized as a process of search and matching. The prominent feature of the optimal contracting problem is the interactions between the internal dynamics (driven by lack of commitment or/and private information) and the external dynamics (driven by stochastic outside shocks), in determining not only how to compensate the worker and provide incentives for the worker to exert effort (or be truthful) as commonly in the literature, but also when to terminate the worker. The optimal termination plays an important role in generating the flows from employment to unemployment/non-participation, as well as in generating (through history-dependent severance compensation) ex post heterogeneity among non-employed workers to account for the distinction between unemployment and non-participation.

[4] **Equilibrium Matching and Termination** (with Cheng Wang), *Journal of Monetary Economics*, 76, 2015, 208-229

[5] **On the Pure Theory of Contract Dispersion** (with Cheng Wang), Under Review

In [4], we introduce dynamic contracts with termination subject to moral hazard into the Mortensen-Pissarides (1994) model with risk averse workers. Specifically, we construct an equilibrium model of the labor market where workers are risk averse and employment relationships are subject to moral hazard. Jobs are dynamic contracts with endogenous termination. Vacant firms and unemployed workers are randomly matched in the labor market to bargain over the values of the dynamic contract for each party. Following termination, firms go back to the labor market to look for other workers. Non-employed workers make optimal consumption and saving decisions and must also decide whether or not to participate in the labor market. Finally, firms freely enter and exit the labor market to endogenously determine the total number of jobs in the economy.

We show that if there is no moral hazard, then the optimal contract is a fixed wage contract without endogenous termination. As a result, the wage distribution is degenerate. There is no flow from employment to unemployment such that all non-employed workers are new labor market entrants who are looking for a job, which implies that the labor participation rate is always one.

However, if there is moral hazard, then the optimal contract is designed as a fully state-contingent scheme addressing the tradeoff between risk sharing and incentive provision as follows.

On one hand, whenever the worker's output turns out to be low, the firm would punish the worker by offering a lower compensation in the current period, and a lower continuation expected utility. Hence, after producing a sequence of low outputs, the worker's expected utility would be pushed lower and lower. Eventually, the worker would lose incentives to exert effort, as in efficiency wage models such as Shapiro and Stiglitz (1984). As a result, the firm would terminate the worker without severance compensation, who then goes back to the labor marker to look for a new job immediately. This is how the flow from employment to unemployment is generated.

On the other hand, after producing a sequence of high outputs, the worker's expected utility would be pushed higher and higher, which implies that his marginal utility becomes lower and lower. Eventually, it becomes too expensive for the principal to provide incentives for the worker to exert effort. As a result, the firm would terminate the worker with severance compensation, who then stays out of the labor market at least temporarily. This is how the flow from employment to non-participation is generated.

Severance compensation, based on the worker's history of outputs, serves as the initial assets held by the non-employed worker, who then decides not only how much to consume/save, but also whether or not to participate the labor market by looking for a job. There are two options. One is to stay out of the labor market for the rest of his life, consuming the annuity of his assets each period. Another is to consume more than the annuity of his assets for some periods, and then rejoin the labor market

once his assets fall below a critical level. There is a tradeoff between the two plans. The first allows the worker to achieve perfect consumption smoothing across all periods in the rest of his life (his consumption will be constant in time), but it also requires the worker to forgo any opportunity in the labor market. With the second plan, the worker would experience less intertemporal consumption smoothing in the rest of his life but then he could expect to obtain the surplus a successful match would offer, after he rejoins the labor market. The loss in consumption smoothing comes from two sources. First, in order to make returning to the labor market beneficial (i.e., to generate a positive surplus from a potential match), the non-employed worker must over consume (to consume more than the annuity of his assets) to make himself sufficiently "poor" before returning to the labor market. Second, once he returns to the market, he runs the risk of not being matched with a vacant firm. Hence, we show that there is a threshold on the non-employed worker's assets above which the worker would stay out of the labor market permanently, and below which the worker would return to the labor market after some periods. This is how the flow from non-participation back to unemployment is generated.

Furthermore, the wage dispersion is generated as the worker's wage contingent on his history of outputs, which are stochastic, spread over time.

Quantitatively, the model is calibrated to the U.S. labor market to successfully account for the distribution and flows of workers across all three states, as well as the wage dispersion. As a specific application, we use the model to evaluate quantitatively the effects of the unemployment insurance (UI) system currently existing in the U.S. In the U.S., unemployed workers are eligible for UI benefits over a period of six months following layoff. We show that such a system increases the reservation utility of unemployed workers, inducing the optimal contract to generate faster and more termination into unemployment but slower flow into non-participation. This results in more steady state labor force participation. UI offers welfare gains for both the employed and unemployed workers, but welfare losses for those not in the labor force. With UI, wages are higher on average and with lower variability. Total output increases because of greater labor force participation, while the measure of jobs remains roughly constant, with or without UI.

In [5], we construct an equilibrium model of the labor market with on-the-job search where jobs are optimal dynamic contracts with endogenous termination. We first show that in the existing models of on-the-job search (e.g., Burdett and Mortensen, 1998; Burdett and Coles, 2003), if the contract is allowed to be contingent on the worker's public outside offers, then the distribution of the contracts offered in equilibrium is degenerate, in which all workers are paid the same monopsony wage. That is, when firms are allowed to compete with outside offers, by way of making counteroffers, workers get paid less, not more, in equilibrium. What happens, obviously, is that allowing for counteroffers discourages outside offers from being made in the first place and, through that, reduces competition in the labor market, rather than increase it. In the absence of counteroffers, any firm need only offer a

slightly higher wage, deviating from the monopsony wage, in order to steal workers from other firms. With counteroffers, such a deviation would be countered in time and thus is never used.

Perhaps even more surprisingly, the idea that allowing for counteroffers in the dynamic contract destroys the equilibrium dispersion and restores the monopsony wage goes farther than having force only with identical firms. Suppose that firms are not identical. Suppose some firms can make (identical) workers more productive, but not too much more productive, than other firms. Then the same logic applies and it continues to hold that the firm's ability to counter the worker's outside offers could deter the offers from being made in the first place, rendering an equilibrium of the labor market where only the monopsony wage is offered.

We then show that if the worker's outside offers are private, not observable to the firm that currently employs him, then a non-degenerate distribution of wage-tenure contracts would arise in equilibrium. That is, search and private information combined could support equilibrium contract dispersion. Furthermore, the model is calibrated to generate the shape of wage distribution observed in the data.

My third project consists of two papers which characterize the optimal strategies of both the principal and the agent in a repeated moral hazard game with private evaluation. Specifically, a theory is developed in accounting for some interesting empirical observations, such as why the principal overreports the agent's output instead of underreporting it in order to save on wages, and why the agent cares about the perceived 'fairness' of the principal's subjective evaluation.

[7] **Repeated Moral Hazard with Private Evaluation: Leniency Bias**
[8] **Repeated Moral Hazard with Private Evaluation: Why the Agent's Mixed Strategies Matter**

In [7], I consider the optimal perfect public equilibrium, which have been partially characterized by Levin (2003) and Fuchs (2007) under the constraints such as, the principal reports truthfully, and the agent is indifferent between shirking and exerting effort. I solve for the optimal perfect public equilibrium explicitly without any constraints. Specifically, I generalize the method of Radner, Myerson and Maskin (1986) to derive an upper bound on the maximum total payoff attainable by perfect public equilibria. Then, the optimal perfect public equilibrium attaining the upper bound is constructed, which consists of a static contract, a pure strategy for the agent, and a non-truth-reporting strategy for the principal.

The optimal contract is static instead of dynamic, which generalizes the static contract result of Levin (2003) without imposing the constraint that the principal reports truthfully. Moreover, the agent's optimal public strategy is pure instead of mixed, therefore his efforts are predictable by the principal. Furthermore, the principal does not reveal the low output with probability one. As suggested by Prendergast (1999), this phenomenon is well documented in empirical studies as leniency bias,

which implies that a supervisor tends to overstate a subordinate's performance. This reflects the fact that the low output occurs with positive probability, even when the agent exerts effort. In other words, the low output is not a perfect indicator for shirking. To summarize, I show that (i) the principal overreports the agent's output, known as leniency bias; (ii) as the discount factor goes to one, the probability with which the principal overreports the agent's output goes to one as well; (iii) more costly for the agent to exert effort, less likely for the principal to overreport the agent's output.

In [8], I study an infinitely repeated moral hazard problem in which the principal privately observes and publicly reports the agent's output, as in Fuchs (2007). The role of the agent's private strategies, which depend on the history of his private efforts, is examined in providing incentives for the principal to be truthful. I show that in order for his effort history to work as an incentive device, the agent has to use a mixed strategy, since otherwise his efforts are predictable by the principal and thus, in effect, public information. However, hiding the agent's efforts from the principal incurs a non-negligible efficiency loss, which may, or may not be justified by the efficiency gain from the use of the agent's private strategies. Moreover, the agent's optimal strategy is shown to be consistent with empirical studies on how employees respond to subjective performance evaluations.

I first show that if the agent uses a pure strategy as in Fuchs (2007), his efforts are predictable by the principal and thus, in effect, public information. Therefore, no efficiency is lost if we restrict the agent to public strategies, when it comes to weak perfect Bayesian equilibria, as defined in Mas-colell, Whinston and Green (1995). To see this, first notice that the principal perfectly predicts what the agent's effort is in the first period, as specified by the agent's pure strategy. And, the agent's effort in the second period is predicted as a deterministic function of the reported output and the agent's effort in the first period, and so on. Hence, the principal's belief about the agent's effort history is degenerate, and independent of the history of true outputs. Moreover, because the principal's belief is independent of the history of true outputs, she reports truthfully only if she is indifferent between reporting the low and high outputs. When it comes to sequential equilibria as considered in this paper, the agent's strategies are most likely to be private. The reason is that different private effort histories of the agent generate different beliefs, therefore different optimal continuation strategies. However, given that the agent uses a pure strategy, the set of allocations attainable by sequential equilibria is shown to be the same as the set of allocations attainable by weak perfect Bayesian equilibria in which the agent is restricted to public strategies. To summarize, the agent's use of mixed strategies is necessary for his prior efforts to be private, therefore potentially effective in providing incentives for the principal to be truthful. Moreover, I show that by using mixed strategies, the agent is able to provide stronger incentives in the sense that the principal strictly prefers reporting truthfully.

I show that there is only one static contract consistent with this class of sequential equilibria. Moreover, given this static contract and the principal's truth-reporting strategy, it can be shown that

the agent is always indifferent between shirking and exerting effort. Therefore, in order to construct an equilibrium, I need only make sure that given the agent's strategy, the principal has incentives to be truthful.

I first consider a subclass of sequential equilibria, in which the principal has incentives to report truthfully, regardless of her belief about whether the agent has shirked or exerted effort. These are belief-free equilibria according to Ely, Hörner and Olszewski (2005). Within this subclass, the agent's optimal strategy is shown to be public, which implies that the agent's effort in the prior period cannot work effectively as an incentive device. Then, given a belief-based equilibrium in which the principal's incentives depend non-trivially on her belief about whether the agent has shirked or exerted effort, it is shown that the agent's effort is always "private", in the sense that the principal is unable to infer it with certainty under any circumstances. In addition, the agent's output always contains the "right" message, in the sense that there exists a fixed threshold such that no matter what happened in the past, the low (high) output makes the principal believe that the probability of the agent having exerted effort is less (greater) than this threshold.

I further show that the agent's strategy in the optimal belief-free (also public as shown above) equilibrium, cannot be approximated by a sequence of his strategies in belief-based equilibria. The reason is that in the optimal belief-free equilibrium, the principal is indifferent between reporting the low and high outputs, instead of preferring strictly being truthful. Therefore, no matter how small the agent's deviation is from his strategy in the optimal belief-free equilibrium, the principal could lose incentives to be truthful.

Furthermore, I show that there exists $\varepsilon > 0$ such that in any belief-based equilibrium, the principal cannot expect the agent to exert effort with a probability greater than $1 - \varepsilon$ under any circumstances. On the contrary, in the optimal belief-free equilibrium, upon reporting the high output in the prior period, the principal expects the agent to exert effort with probability one. This is a non-negligible efficiency loss associated with the need of hiding the agent's efforts from the principal. Numerical analysis shows that this efficiency loss may, or may not be justified by the efficiency gain from using the agent's effort in the prior period as an incentive device.

My fourth project consists of one paper which uses the monthly employment history data provided by the China Urban Labor Survey 2001, 2005, 2009 to study the distribution and flows of workers in the Chinese labor market empirically. Each round of the China Urban Labor Survey provides a detailed 72-month employment history for a sample of local workers, as well as a sample of migrant workers. We use a fixed effect estimator in a dynamic multinomial logit model which can accommodate any form of unobserved heterogeneity in levels, as in Magnac (2000), to estimate the transition probabilities across employment, unemployment, and non-participation.

[6] **The Unemployment and Mobility of Rural Labor in the Urban Labor Market**, funded by National Natural Science Foundation of China (Grant NO.71403160).